? Water makes life possible: The cells of your body could not function without water in which to dissolve essential biological molecules. Water is such a good solvent because its molecules (i) have zero net charge; (ii) have zero net charge, but the positive and negative charges are separated; (iii) have nonzero net charge; (iv) do not respond to electric forces; (v) exert repulsive electric forces on each other.

21 ELECTRIC CHARGE AND ELECTRIC FIELD

LEARNING GOALS

Looking forward at …

21.1 The nature of electric charge, and how we know that electric charge is conserved.

21.2 How objects become electrically charged.

21.3 How to use Coulomb's law to calculate the electric force between charges.

21.4 The distinction between electric force and electric field.

21.5 How to calculate the electric field due to a collection of charges.

21.6 How to use the idea of electric field lines to visualize and interpret electric fields.

21.7 How to calculate the properties of electric dipoles.

Looking back at …

1.7–1.10 Vector algebra, including the scalar (dot) product and the vector (cross) product.

4.3 Newton's second law.

7.5 Stable and unstable equilibria.

12.5 Streamlines in fluid flow.

In Chapter 5 we mentioned the four kinds of fundamental forces. To this point the only one of these forces that we have examined in any detail is gravity. Now we are ready to examine the force of *electromagnetism,* which encompasses both electricity and magnetism.

Electromagnetic interactions involve particles that have *electric charge,* an attribute that is as fundamental as mass. Just as objects with mass are accelerated by gravitational forces, so electrically charged objects are accelerated by electric forces. The shock you feel when you scuff your shoes across a carpet and then reach for a metal doorknob is due to charged particles leaping between your finger and the doorknob. Electric currents are simply streams of charged particles flowing within wires in response to electric forces. Even the forces that hold atoms together to form solid matter, and that keep the atoms of solid objects from passing through each other, are fundamentally due to electric interactions between the charged particles within atoms.

We begin our study of electromagnetism in this chapter by examining the nature of electric charge. We'll find that charge is quantized and obeys a conservation principle. When charges are at rest in our frame of reference, they exert *electrostatic* forces on each other. These forces are of tremendous importance in chemistry and biology and have many technological applications. Electrostatic forces are governed by a simple relationship known as *Coulomb's law* and are most conveniently described by using the concept of *electric field.* In later chapters we'll expand our discussion to include electric charges in motion. This will lead us to an understanding of magnetism and, remarkably, of the nature of light.

While the key ideas of electromagnetism are conceptually simple, applying them to practical problems will make use of many of your mathematical skills, especially your knowledge of geometry and integral calculus. For this reason you may find this chapter and those that follow to be more mathematically demanding than earlier chapters. The reward for your extra effort will be a deeper understanding of principles that are at the heart of modern physics and technology.

21.1 ELECTRIC CHARGE

The ancient Greeks discovered as early as 600 B.C. that after they rubbed amber with wool, the amber could attract other objects. Today we say that the amber has acquired a net **electric charge,** or has become *charged.* The word "electric" is derived from the Greek word *elektron,* meaning amber. When you scuff your shoes across a nylon carpet, you become electrically charged, and you can charge a comb by passing it through dry hair.

Plastic rods and fur (real or fake) are particularly good for demonstrating **electrostatics,** the interactions between electric charges that are at rest (or nearly so). After we charge both plastic rods in **Fig. 21.1a** by rubbing them with the piece of fur, we find that the rods repel each other.

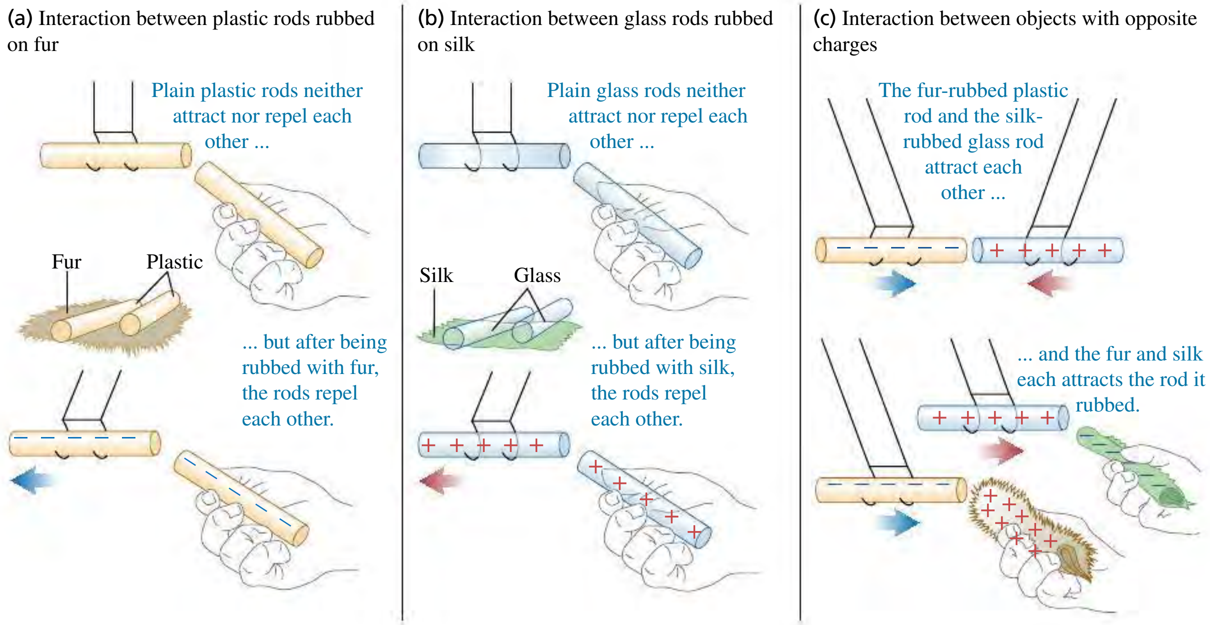
When we rub glass rods with silk, the glass rods also become charged and repel each other (Fig. 21.1b). But a charged plastic rod *attracts* a charged glass rod; furthermore, the plastic rod and the fur attract each other, and the glass rod and the silk attract each other (Fig. 21.1c).

These experiments and many others like them have shown that there are exactly two kinds of electric charge: the kind on the plastic rod rubbed with fur and the kind on the glass rod rubbed with silk. Benjamin Franklin (1706–1790) suggested calling these two kinds of charge *negative* and *positive,* respectively, and these names are still used. The plastic rod and the silk have negative charge; the glass rod and the fur have positive charge.

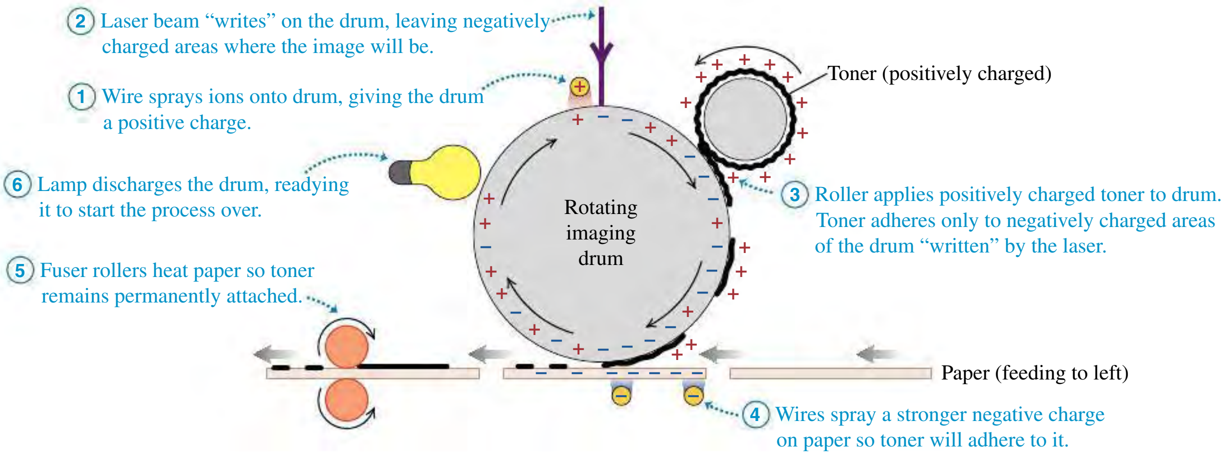
Two positive charges or two negative charges repel each other. A positive charge and a negative charge attract each other.

CAUTION **Electric attraction and repulsion** The attraction and repulsion of two charged objects are sometimes summarized as "Like charges repel, and opposite charges attract." But "like charges" does *not* mean that the two charges are exactly identical, only that both charges have the same algebraic *sign* (both positive or both negative). "Opposite charges" means that both objects have an electric charge, and those charges have different signs (one positive and the other negative).

21.1 Experiments in electrostatics. (a) Negatively charged objects repel each other. (b) Positively charged objects repel each other. (c) Positively charged objects and negatively charged objects attract each other.



21.2 Schematic diagram of the operation of a laser printer.



A laser printer (**Fig. 21.2**) utilizes the forces between charged bodies. The printer's light-sensitive imaging drum is given a positive charge. As the drum rotates, a laser beam shines on selected areas of the drum, leaving those areas with a *negative* charge. Positively charged particles of toner adhere only to the areas of the drum "written" by the laser. When a piece of paper is placed in contact with the drum, the toner particles stick to the paper and form an image.

Electric Charge and the Structure of Matter

When you charge a rod by rubbing it with fur or silk as in Fig. 21.1, there is no visible change in the appearance of the rod. What, then, actually happens to the rod when you charge it? To answer this question, we must look more closely at the structure of atoms, the building blocks of ordinary matter.

The structure of atoms can be described in terms of three particles: the negatively charged **electron**, the positively charged **proton,** and the uncharged **neutron** (**Fig. 21.3**). The proton and neutron are combinations of other entities called *quarks,* which have charges of $\pm\frac{1}{3}$ and $\pm\frac{2}{3}$ times the electron charge. Isolated quarks have not been observed, and there are theoretical reasons to believe that it is impossible in principle to observe a quark in isolation.

The protons and neutrons in an atom make up a small, very dense core called the **nucleus,** with dimensions of the order of 10^{-15} m. Surrounding the nucleus are the electrons, extending out to distances of the order of 10^{-10} m from the nucleus. If an atom were a few kilometers across, its nucleus would be the size of a tennis ball. The negatively charged electrons are held within the atom by the attractive electric forces exerted on them by the positively charged nucleus. (The protons and neutrons are held within stable atomic nuclei by an attractive interaction, called the *strong nuclear force,* that overcomes the electric repulsion of the protons. The strong nuclear force has a short range, and its effects do not extend far beyond the nucleus.)

The masses of the individual particles, to the precision that they are presently known, are

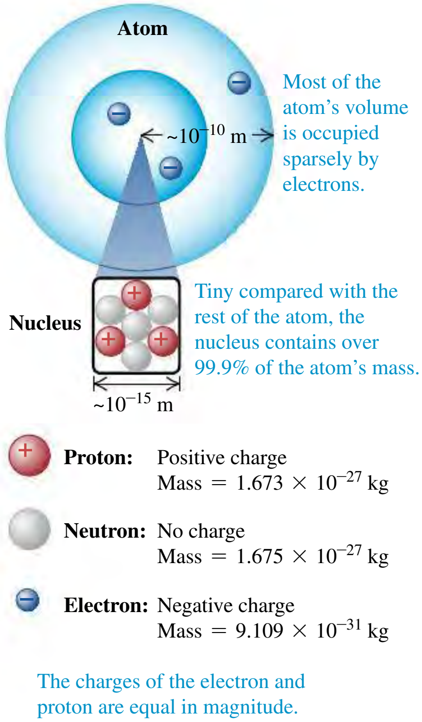
$$\text{Mass of electron} = m_e = 9.10938291(40) \times 10^{-31}\ \text{kg}$$

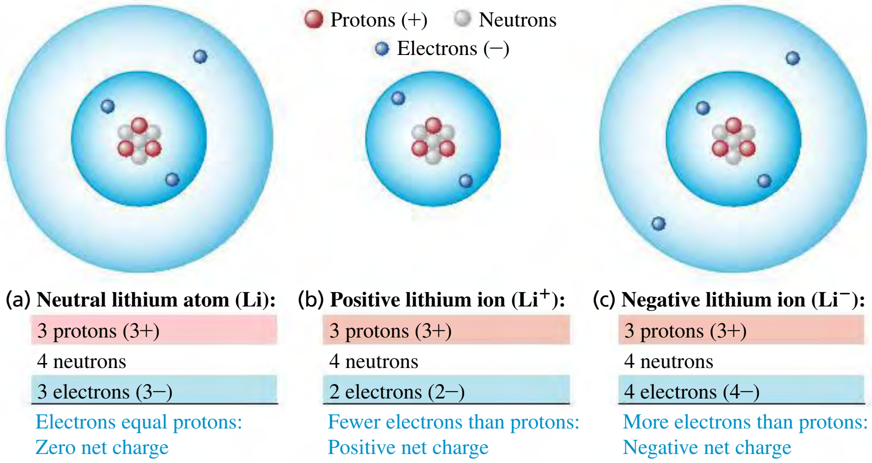
$$\text{Mass of proton} = m_p = 1.672621777(74) \times 10^{-27}\ \text{kg}$$

$$\text{Mass of neutron} = m_n = 1.674927351(74) \times 10^{-27}\ \text{kg}$$

The numbers in parentheses are the uncertainties in the last two digits. Note that the masses of the proton and neutron are nearly equal and are roughly 2000 times the mass of the electron. Over 99.9% of the mass of any atom is concentrated in its nucleus.

21.3 The structure of an atom. The particular atom depicted here is lithium (see Fig. 21.4a).





21.4 (a) A neutral atom has as many electrons as it does protons. (b) A positive ion has a deficit of electrons. (c) A negative ion has an excess of electrons. (The electron "shells" are a schematic representation of the actual electron distribution, a diffuse cloud many times larger than the nucleus.)

The negative charge of the electron has (within experimental error) *exactly* the same magnitude as the positive charge of the proton. In a neutral atom the number of electrons equals the number of protons in the nucleus, and the net electric charge (the algebraic sum of all the charges) is exactly zero (**Fig. 21.4a**). The number of protons or electrons in a neutral atom of an element is called the **atomic number** of the element. If one or more electrons are removed from an atom, what remains is called a **positive ion** (Fig. 21.4b). A **negative ion** is an atom that has *gained* one or more electrons (Fig. 21.4c). This gain or loss of electrons is called **ionization.**

When the total number of protons in a macroscopic body equals the total number of electrons, the total charge is zero and the body as a whole is electrically neutral. To give a body an excess negative charge, we may either *add negative* charges to a neutral body or *remove positive* charges from that body. Similarly, we can create an excess positive charge by either *adding positive* charge or *removing negative* charge. In most cases, negatively charged (and highly mobile) electrons are added or removed, and a "positively charged body" is one that has lost some of its normal complement of electrons. When we speak of the charge of a body, we always mean its *net* charge. The net charge is always a very small fraction (typically no more than 10^{-12}) of the total positive charge or negative charge in the body.

Electric Charge Is Conserved

Implicit in the foregoing discussion are two very important principles. First is the **principle of conservation of charge:**

The algebraic sum of all the electric charges in any closed system is constant.

If we rub together a plastic rod and a piece of fur, both initially uncharged, the rod acquires a negative charge (since it takes electrons from the fur) and the fur acquires a positive charge of the *same* magnitude (since it has lost as many electrons as the rod has gained). Hence the total electric charge on the two bodies together does not change. In any charging process, charge is not created or destroyed; it is merely *transferred* from one body to another.

Conservation of charge is thought to be a *universal* conservation law. No experimental evidence for any violation of this principle has ever been observed. Even in high-energy interactions in which particles are created and destroyed, such as the creation of electron–positron pairs, the total charge of any closed system is exactly constant.

The second important principle is:

The magnitude of charge of the electron or proton is a natural unit of charge.

Every observable amount of electric charge is always an integer multiple of this basic unit. We say that charge is *quantized*. A familiar example of quantization is money. When you pay cash for an item in a store, you have to do it in one-cent increments. Cash can't be divided into amounts smaller than one cent, and electric charge can't be divided into amounts smaller than the charge of one electron or proton. (The quark charges, $\pm\frac{1}{3}$ and $\pm\frac{2}{3}$ of the electron charge, are probably not observable as isolated charges.) Thus the charge on any macroscopic body is always zero or an integer multiple (negative or positive) of the electron charge.

Understanding the electric nature of matter gives us insight into many aspects of the physical world (**Fig. 21.5**). The chemical bonds that hold atoms together to form molecules are due to electric interactions between the atoms. They include the strong ionic bonds that hold sodium and chlorine atoms together to make table salt and the relatively weak bonds between the strands of DNA that record your body's genetic code. When you stand, the normal force exerted on you by the floor arises from electric forces between charged particles in the atoms of your shoes and the atoms of the floor. The tension force in a stretched string and the adhesive force of glue are likewise due to electric interactions of atoms.

21.5 Most of the forces on this water skier are electric. Electric interactions between adjacent molecules give rise to the force of the water on the ski, the tension in the tow rope, and the resistance of the air on the skier's body. Electric interactions also hold the atoms of the skier's body together. Only one wholly nonelectric force acts on the skier: the force of gravity.

TEST YOUR UNDERSTANDING OF SECTION 21.1 Two charged objects repel each other through the electric force. The charges on the objects are (i) one positive and one negative; (ii) both positive; (iii) both negative; (iv) either (ii) or (iii); (v) any of (i), (ii), or (iii). ❙

21.2 CONDUCTORS, INSULATORS, AND INDUCED CHARGES

MP

PhET: Balloons and Static Electricity
PhET: John Travoltage

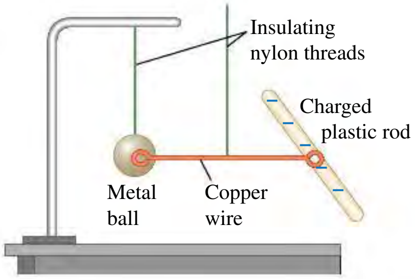
Some materials permit electric charge to move easily from one region of the material to another, while others do not. For example, **Fig. 21.6a** (next page) shows a copper wire supported by a nylon thread. Suppose you touch one end of the wire to a charged plastic rod and attach the other end to a metal ball that is initially uncharged; you then remove the charged rod and the wire. When you bring another charged body up close to the ball (Figs. 21.6b and 21.6c), the ball is attracted or repelled, showing that the ball has become electrically charged. Electric charge has been transferred through the copper wire between the ball and the surface of the plastic rod.

The copper wire is called a **conductor** of electricity. If you repeat the experiment using a rubber band or nylon thread in place of the wire, you find that *no* charge is transferred to the ball. These materials are called **insulators.** Conductors permit the easy movement of charge through them, while insulators do not. (The supporting nylon threads shown in Fig. 21.6 are insulators, which prevents charge from leaving the metal ball and copper wire.)

As an example, carpet fibers on a dry day are good insulators. As you walk across a carpet, the rubbing of your shoes against the fibers causes charge to build up on you, and this charge remains on you because it can't flow through the insulating fibers. If you then touch a conducting object such as a doorknob, a rapid charge transfer takes place between your finger and the doorknob, and you feel a shock. One way to prevent this is to wind some of the carpet fibers around conducting cores so that any charge that builds up on you can be transferred harmlessly to the carpet. Another solution is to coat the carpet fibers with an antistatic layer that does not easily transfer electrons to or from your shoes; this prevents any charge from building up on you in the first place.

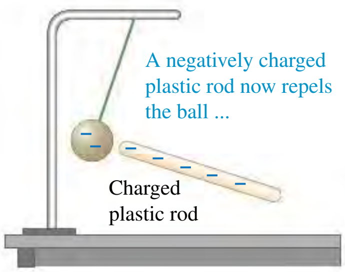
21.6 Copper is a good conductor of electricity; nylon is a good insulator. (a) The copper wire conducts charge between the metal ball and the charged plastic rod to charge the ball negatively. Afterward, the metal ball is (b) repelled by a negatively charged plastic rod and (c) attracted to a positively charged glass rod.

(a)

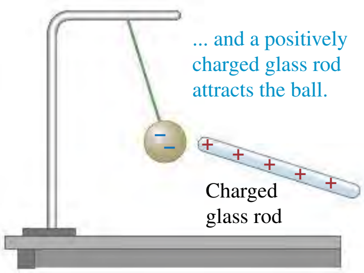


The wire conducts charge from the negatively charged plastic rod to the metal ball.

(b)



(c)



Most metals are good conductors, while most nonmetals are insulators. Within a solid metal such as copper, one or more outer electrons in each atom become detached and can move freely throughout the material, just as the molecules of a gas can move through the spaces between the grains in a bucket of sand. The other electrons remain bound to the positively charged nuclei, which themselves are bound in nearly fixed positions within the material. In an insulator there are no, or very few, free electrons, and electric charge cannot move freely through the material. Some materials called *semiconductors* are intermediate in their properties between good conductors and good insulators.

Charging by Induction

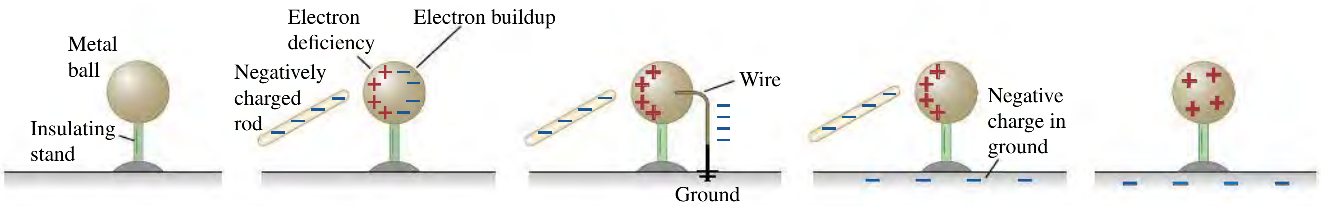
We can charge a metal ball by using a copper wire and an electrically charged plastic rod, as in Fig. 21.6a. In this process, some of the excess electrons on the rod are transferred from it to the ball, leaving the rod with a smaller negative charge. But there is a different technique in which the plastic rod can give another body a charge of *opposite* sign without losing any of its own charge. This process is called charging by **induction.**

Figure 21.7 shows an example of charging by induction. An uncharged metal ball is supported on an insulating stand (Fig. 21.7a). When you bring a negatively charged rod near it, without actually touching it (Fig. 21.7b), the free electrons in the metal ball are repelled by the excess electrons on the rod, and they shift toward the right, away from the rod. They cannot escape from the ball because the supporting stand and the surrounding air are insulators. So we get excess negative charge at the right surface of the ball and a deficiency of negative charge (that is, a net positive charge) at the left surface. These excess charges are called **induced charges.**

Not all of the free electrons move to the right surface of the ball. As soon as any induced charge develops, it exerts forces toward the *left* on the other free electrons. These electrons are repelled by the negative induced charge on the right and attracted toward the positive induced charge on the left. The system reaches an equilibrium state in which the force toward the right on an electron, due to the charged rod, is just balanced by the force toward the left due to the induced charge. If we remove the charged rod, the free electrons shift back to the left, and the original neutral condition is restored.

What happens if, while the plastic rod is nearby, you touch one end of a conducting wire to the right surface of the ball and the other end to the earth (Fig. 21.7c)? The earth is a conductor, and it is so large that it can act as a practically infinite source of extra electrons or sink of unwanted electrons. Some of the negative charge flows through the wire to the earth. Now suppose you disconnect the wire (Fig. 21.7d) and then remove the rod (Fig. 21.7e); a net positive charge is left on the ball. The charge on the negatively charged rod has not changed during this process. The earth acquires a negative charge that is equal in magnitude to the induced positive charge remaining on the ball.

21.7 Charging a metal ball by induction.



(a) Uncharged metal ball

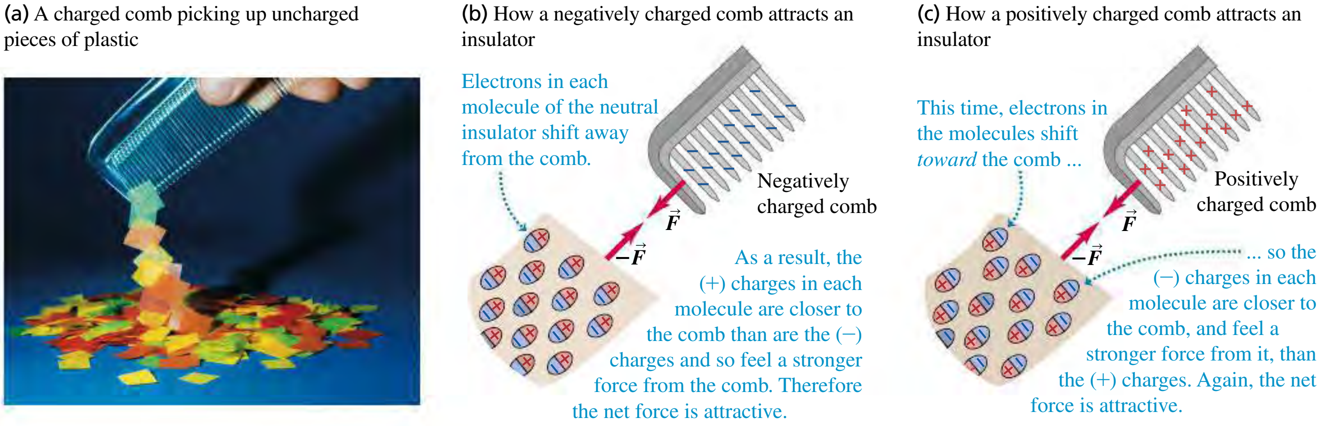
(b) Negative charge on rod repels electrons, creating zones of negative and positive **induced charge**.

(c) Wire lets electron buildup (induced negative charge) flow into ground.

(d) Wire removed; ball now has only an electron-deficient region of positive charge.

(e) Rod removed; electrons rearrange themselves, ball has overall electron deficiency (net positive charge).

21.8 The charges within the molecules of an insulating material can shift slightly. As a result, a comb with either sign of charge attracts a neutral insulator. By Newton's third law the neutral insulator exerts an equal-magnitude attractive force on the comb.



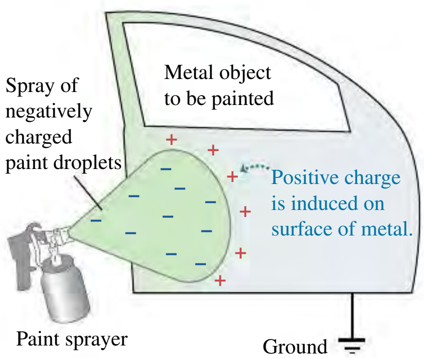
Electric Forces on Uncharged Objects

Finally, we note that a charged body can exert forces even on objects that are *not* charged themselves. If you rub a balloon on the rug and then hold the balloon against the ceiling, it sticks, even though the ceiling has no net electric charge. After you electrify a comb by running it through your hair, you can pick up uncharged bits of paper or plastic with it (**Fig. 21.8a**). How is this possible?

DEMO

This interaction is an induced-charge effect. Even in an insulator, electric charges can shift back and forth a little when there is charge nearby. This is shown in Fig. 21.8b; the negatively charged plastic comb causes a slight shifting of charge within the molecules of the neutral insulator, an effect called *polarization*. The positive and negative charges in the material are present in equal amounts, but the positive charges are closer to the plastic comb and so feel an attraction that is stronger than the repulsion felt by the negative charges, giving a net attractive force. (In Section 21.3 we will study how electric forces depend on distance.) Note that a neutral insulator is also attracted to a *positively* charged comb (Fig. 21.8c). Now the charges in the insulator shift in the opposite direction; the negative charges in the insulator are closer to the comb and feel an attractive force that is stronger than the repulsion felt by the positive charges in the insulator. Hence a charged object of *either* sign exerts an attractive force on an uncharged insulator. **Figure 21.9** shows an industrial application of this effect.

TEST YOUR UNDERSTANDING OF SECTION 21.2 You have two lightweight metal spheres, each hanging from an insulating nylon thread. One of the spheres has a net negative charge, while the other sphere has no net charge. (a) If the spheres are close together but do not touch, will they (i) attract each other, (ii) repel each other, or (iii) exert no force on each other? (b) You now allow the two spheres to touch. Once they have touched, will the two spheres (i) attract each other, (ii) repel each other, or (iii) exert no force on each other? ❙



21.9 The electrostatic painting process (compare Figs. 21.7b and 21.7c). A metal object to be painted is connected to the earth ("ground"), and the paint droplets are given an electric charge as they exit the sprayer nozzle. Induced charges of the opposite sign appear in the object as the droplets approach, just as in Fig. 21.7b, and they attract the droplets to the surface. This process minimizes overspray from clouds of stray paint particles and gives a particularly smooth finish.

BIO Application Electric Forces, Sweat, and Cystic Fibrosis One way to test for the genetic disease cystic fibrosis (CF) is to measure the salt content of a person's sweat. Sweat is a mixture of water and ions, including the sodium (Na^+) and chloride (Cl^-) ions that make up ordinary salt (NaCl). When sweat is secreted by epithelial cells, some of the Cl^- ions flow from the sweat back into these cells (a process called reabsorption). The electric attraction between negative and positive charges pulls Na^+ ions along with the Cl^-. Water molecules cannot flow back into the epithelial cells, so sweat on the skin has a low salt content. However, in persons with CF the reabsorption of Cl^- ions is blocked. Hence the sweat of persons with CF is unusually salty, with up to four times the normal concentration of Cl^- and Na^+.

21.3 COULOMB'S LAW

Charles Augustin de Coulomb (1736–1806) studied the interaction forces of charged particles in detail in 1784. He used a torsion balance (**Fig. 21.10a**) similar to the one used 13 years later by Cavendish to study the much weaker gravitational interaction, as we discussed in Section 13.1. For **point charges,** charged bodies that are very small in comparison with the distance r between them, Coulomb found that the electric force is proportional to $1/r^2$. That is, when the distance r doubles, the force decreases to one-quarter of its initial value; when the distance is halved, the force increases to four times its initial value.

The electric force between two point charges also depends on the quantity of charge on each body, which we will denote by q or Q. To explore this dependence, Coulomb divided a charge into two equal parts by placing a small charged spherical conductor into contact with an identical but uncharged sphere; by symmetry, the charge is shared equally between the two spheres. (Note the essential role of the principle of conservation of charge in this procedure.) Thus he could obtain one-half, one-quarter, and so on, of any initial charge. He found that the forces that two point charges q_1 and q_2 exert on each other are proportional to each charge and therefore are proportional to the *product* $q_1 q_2$ of the two charges.

Thus Coulomb established what we now call **Coulomb's law:**

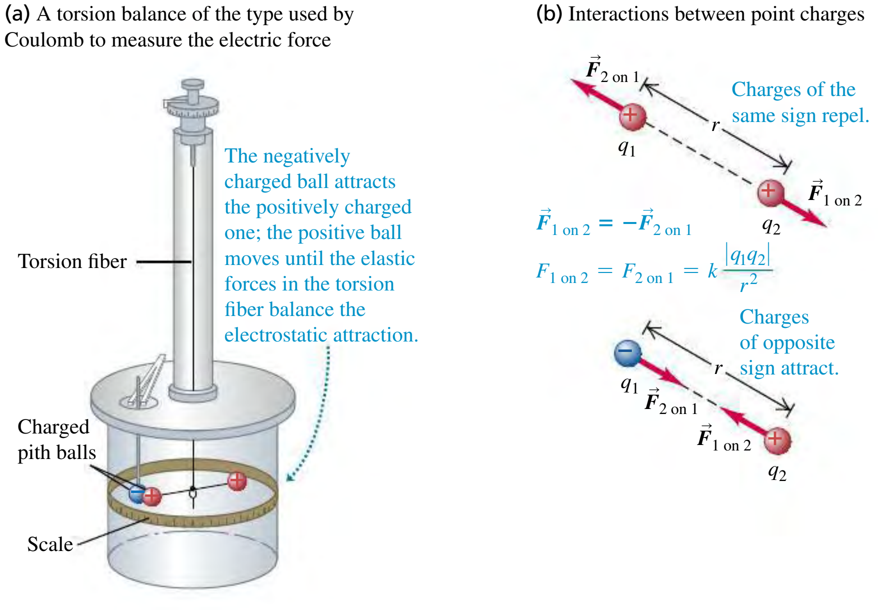
The magnitude of the electric force between two point charges is directly proportional to the product of the charges and inversely proportional to the square of the distance between them.

In mathematical terms, the magnitude F of the force that each of two point charges q_1 and q_2 a distance r apart exerts on the other can be expressed as

$$F = k\frac{|q_1 q_2|}{r^2} \tag{21.1}$$

where k is a proportionality constant whose numerical value depends on the system of units used. The absolute value bars are used in Eq. (21.1) because the charges q_1 and q_2 can be either positive or negative, while the force magnitude F is always positive.

The directions of the forces the two charges exert on each other are always along the line joining them. When the charges q_1 and q_2 have the same sign, either both positive or both negative, the forces are repulsive; when the charges



21.10 (a) Measuring the electric force between point charges. (b) The electric forces between point charges obey Newton's third law: $\vec{F}_{1\text{ on }2} = -\vec{F}_{2\text{ on }1}$.

have opposite signs, the forces are attractive (Fig. 21.10b). The two forces obey Newton's third law; they are always equal in magnitude and opposite in direction, even when the charges are not equal in magnitude.

The proportionality of the electric force to $1/r^2$ has been verified with great precision. There is no reason to suspect that the exponent is different from precisely 2. Thus the form of Eq. (21.1) is the same as that of the law of gravitation. But electric and gravitational interactions are two distinct classes of phenomena. Electric interactions depend on electric charges and can be either attractive or repulsive, while gravitational interactions depend on mass and are always attractive (because there is no such thing as negative mass).

Fundamental Electric Constants

The value of the proportionality constant k in Coulomb's law depends on the system of units used. In our study of electricity and magnetism we will use SI units exclusively. The SI electric units include most of the familiar units such as the volt, the ampere, the ohm, and the watt. (There is *no* British system of electric units.) The SI unit of electric charge is called one **coulomb** (1 C). In SI units the constant k in Eq. (21.1) is

$$k = 8.987551787 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2 \cong 8.988 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$$

The value of k is known to such a large number of significant figures because this value is closely related to the speed of light in vacuum. (We will show this in Chapter 32 when we study electromagnetic radiation.) As we discussed in Section 1.3, this speed is *defined* to be exactly $c = 2.99792458 \times 10^8$ m/s. The numerical value of k is defined in terms of c to be precisely

$$k = (10^{-7} \text{ N} \cdot \text{s}^2/\text{C}^2)c^2$$

You should check this expression to confirm that k has the right units.

In principle we can measure the electric force F between two equal charges q at a measured distance r and use Coulomb's law to determine the charge. Thus we could regard the value of k as an operational definition of the coulomb. For reasons of experimental precision it is better to define the coulomb instead in terms of a unit of electric *current* (charge per unit time), the *ampere,* equal to 1 coulomb per second. We will return to this definition in Chapter 28.

In SI units we usually write the constant k in Eq. (21.1) as $1/4\pi\epsilon_0$, where ϵ_0 ("epsilon-nought" or "epsilon-zero") is called the **electric constant.** This shorthand simplifies many formulas that we will encounter in later chapters. From now on, we will usually write Coulomb's law as

Coulomb's law:
Magnitude of electric force between two point charges

$$F = \frac{1}{4\pi\epsilon_0}\frac{|q_1 q_2|}{r^2} \qquad (21.2)$$

Electric constant; Values of the two charges; Distance between the two charges

The constants in Eq. (21.2) are approximately

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2 \quad \text{and} \quad \frac{1}{4\pi\epsilon_0} = k = 8.988 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$$

In examples and problems we will often use the approximate value

$$\frac{1}{4\pi\epsilon_0} = 9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$$

As we mentioned in Section 21.1, the most fundamental unit of charge is the magnitude of the charge of an electron or a proton, which is denoted by e. The most precise value available as of the writing of this book is

$$e = 1.602176565(35) \times 10^{-19} \text{ C}$$

One coulomb represents the negative of the total charge of about 6×10^{18} electrons. For comparison, a copper cube 1 cm on a side contains about 2.4×10^{24} electrons. About 10^{19} electrons pass through the glowing filament of a flashlight bulb every second.

In electrostatics problems (problems that involve charges at rest), it's very unusual to encounter charges as large as 1 coulomb. Two 1-C charges separated by 1 m would exert forces on each other of magnitude 9×10^9 N (about 1 million tons)! The total charge of all the electrons in a copper one-cent coin is even greater, about 1.4×10^5 C, which shows that we can't disturb electric neutrality very much without using enormous forces. More typical values of charge range from about a microcoulomb ($1\ \mu\text{C} = 10^{-6}\ \text{C}$) to about a nanocoulomb ($1\ \text{nC} = 10^{-9}\ \text{C}$).

EXAMPLE 21.1 ELECTRIC FORCE VERSUS GRAVITATIONAL FORCE

SOLUTION

An α particle (the nucleus of a helium atom) has mass $m = 6.64 \times 10^{-27}$ kg and charge $q = +2e = 3.2 \times 10^{-19}$ C. Compare the magnitude of the electric repulsion between two α ("alpha") particles with that of the gravitational attraction between them.

SOLUTION

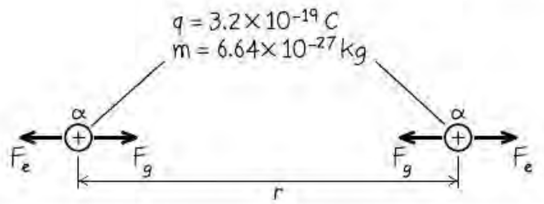
IDENTIFY and SET UP: This problem involves Newton's law for the gravitational force F_g between particles (see Section 13.1) and Coulomb's law for the electric force F_e between point charges. To compare these forces, we make our target variable the ratio F_e/F_g. We use Eq. (21.2) for F_e and Eq. (13.1) for F_g.

EXECUTE: **Figure 21.11** shows our sketch. From Eqs. (21.2) and (13.1),

$$F_e = \frac{1}{4\pi\epsilon_0}\frac{q^2}{r^2} \qquad F_g = G\frac{m^2}{r^2}$$

These are both inverse-square forces, so the r^2 factors cancel when we take the ratio:

21.11 Our sketch for this problem.



$$\frac{F_e}{F_g} = \frac{1}{4\pi\epsilon_0 G}\frac{q^2}{m^2}$$

$$= \frac{9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2}{6.67 \times 10^{-11}\ \text{N}\cdot\text{m}^2/\text{kg}^2}\frac{(3.2 \times 10^{-19}\ \text{C})^2}{(6.64 \times 10^{-27}\ \text{kg})^2} = 3.1 \times 10^{35}$$

EVALUATE: This astonishingly large number shows that the gravitational force in this situation is completely negligible in comparison to the electric force. This is always true for interactions of atomic and subnuclear particles. But within objects the size of a person or a planet, the positive and negative charges are nearly equal in magnitude, and the net electric force is usually much *smaller* than the gravitational force.

Superposition of Forces

Coulomb's law as we have stated it describes only the interaction of two *point* charges. Experiments show that when two charges exert forces simultaneously on a third charge, the total force acting on that charge is the *vector sum* of the forces that the two charges would exert individually. This important property, called the **principle of superposition of forces,** holds for any number of charges. By using this principle, we can apply Coulomb's law to *any* collection of charges. Two of the examples at the end of this section use the superposition principle.

Strictly speaking, Coulomb's law as we have stated it should be used only for point charges *in a vacuum*. If matter is present in the space between the charges, the net force acting on each charge is altered because charges are induced in the molecules of the intervening material. We will describe this effect later. As a practical matter, though, we can use Coulomb's law unaltered for point charges in air. At normal atmospheric pressure, the presence of air changes the electric force from its vacuum value by only about one part in 2000.

PROBLEM-SOLVING STRATEGY 21.1 COULOMB'S LAW

IDENTIFY *the relevant concepts:* Coulomb's law describes the electric force between charged particles.

SET UP *the problem* using the following steps:
1. Sketch the locations of the charged particles and label each particle with its charge.
2. If the charges do not all lie on a single line, set up an *xy*-coordinate system.
3. The problem will ask you to find the electric force on one or more particles. Identify which these are.

EXECUTE *the solution* as follows:
1. For each particle that exerts an electric force on a given particle of interest, use Eq. (21.2) to calculate the magnitude of that force.
2. Using those magnitudes, sketch a free-body diagram showing the electric-force vectors acting on each particle of interest. The force exerted by particle 1 on particle 2 points from particle 2 toward particle 1 if the charges have opposite signs, but points from particle 2 directly away from particle 1 if the charges have the same sign.
3. Use the principle of superposition to calculate the total electric force—a *vector* sum—on each particle of interest. (Review the vector algebra in Sections 1.7 through 1.9. The method of components is often helpful.)
4. Use consistent units; SI units are completely consistent. With $1/4\pi\epsilon_0 = 9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2}$, distances must be in meters, charges in coulombs, and forces in newtons.
5. Some examples and problems in this and later chapters involve *continuous* distributions of charge along a line, over a surface, or throughout a volume. In these cases the vector sum in step 3 becomes a vector *integral*. We divide the charge distribution into infinitesimal pieces, use Coulomb's law for each piece, and integrate to find the vector sum. Sometimes this can be done without actual integration.
6. Exploit any symmetries in the charge distribution to simplify your problem solving. For example, two identical charges q exert zero net electric force on a charge Q midway between them, because the forces on Q have equal magnitude and opposite direction.

EVALUATE *your answer:* Check whether your numerical results are reasonable. Confirm that the direction of the net electric force agrees with the principle that charges of the same sign repel and charges of opposite sign attract.

EXAMPLE 21.2 FORCE BETWEEN TWO POINT CHARGES

SOLUTION

Two point charges, $q_1 = +25\ \mathrm{nC}$ and $q_2 = -75\ \mathrm{nC}$, are separated by a distance $r = 3.0\ \mathrm{cm}$ (**Fig. 21.12a**). Find the magnitude and direction of the electric force (a) that q_1 exerts on q_2 and (b) that q_2 exerts on q_1.

SOLUTION

IDENTIFY and SET UP: This problem asks for the electric forces that two charges exert on each other. We use Coulomb's law, Eq. (21.2), to calculate the magnitudes of the forces. The signs of the charges will determine the directions of the forces.

EXECUTE: (a) After converting the units of r to meters and the units of q_1 and q_2 to coulombs, Eq. (21.2) gives us

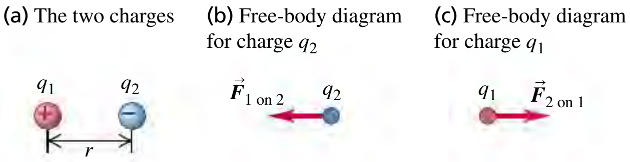
$$F_{1\ \mathrm{on}\ 2} = \frac{1}{4\pi\epsilon_0}\frac{|q_1 q_2|}{r^2}$$

$$= (9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2})\frac{|(+25 \times 10^{-9}\ \mathrm{C})(-75 \times 10^{-9}\ \mathrm{C})|}{(0.030\ \mathrm{m})^2}$$

$$= 0.019\ \mathrm{N}$$

The charges have opposite signs, so the force is attractive (to the left in Fig. 21.12b); that is, the force that acts on q_2 is directed toward q_1 along the line joining the two charges.

21.12 What force does q_1 exert on q_2, and what force does q_2 exert on q_1? Gravitational forces are negligible.



(b) Proceeding as in part (a), we have

$$F_{2\ \mathrm{on}\ 1} = \frac{1}{4\pi\epsilon_0}\frac{|q_2 q_1|}{r^2} = F_{1\ \mathrm{on}\ 2} = 0.019\ \mathrm{N}$$

The attractive force that acts on q_1 is to the right, toward q_2 (Fig. 21.12c).

EVALUATE: Newton's third law applies to the electric force. Even though the charges have different magnitudes, the magnitude of the force that q_2 exerts on q_1 is the same as the magnitude of the force that q_1 exerts on q_2, and these two forces are in opposite directions.

EXAMPLE 21.3 VECTOR ADDITION OF ELECTRIC FORCES ON A LINE

SOLUTION

Two point charges are located on the x-axis of a coordinate system: $q_1 = 1.0\text{ nC}$ is at $x = +2.0\text{ cm}$, and $q_2 = -3.0\text{ nC}$ is at $x = +4.0\text{ cm}$. What is the total electric force exerted by q_1 and q_2 on a charge $q_3 = 5.0\text{ nC}$ at $x = 0$?

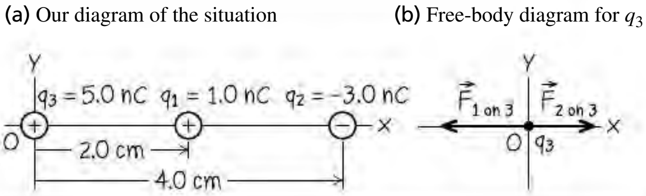
SOLUTION

IDENTIFY and SET UP: **Figure 21.13a** shows the situation. To find the total force on q_3, our target variable, we find the vector sum of the two electric forces on it.

EXECUTE: Figure 21.13b is a free-body diagram for q_3, which is repelled by q_1 (which has the same sign) and attracted to q_2 (which has the opposite sign): $\vec{F}_{1\text{ on }3}$ is in the $-x$-direction and $\vec{F}_{2\text{ on }3}$ is in the $+x$-direction. After unit conversions, we have from Eq. (21.2)

$$\begin{aligned} F_{1\text{ on }3} &= \frac{1}{4\pi\epsilon_0}\frac{|q_1q_3|}{r_{13}^2} \\ &= (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{(1.0 \times 10^{-9}\ \text{C})(5.0 \times 10^{-9}\ \text{C})}{(0.020\ \text{m})^2} \\ &= 1.12 \times 10^{-4}\ \text{N} = 112\ \mu\text{N} \end{aligned}$$

21.13 Our sketches for this problem.



In the same way you can show that $F_{2\text{ on }3} = 84\ \mu\text{N}$. We thus have $\vec{F}_{1\text{ on }3} = (-112\ \mu\text{N})\hat{\imath}$ and $\vec{F}_{2\text{ on }3} = (84\ \mu\text{N})\hat{\imath}$. The net force on q_3 is

$$\vec{F}_3 = \vec{F}_{1\text{ on }3} + \vec{F}_{2\text{ on }3} = (-112\ \mu\text{N})\hat{\imath} + (84\ \mu\text{N})\hat{\imath} = (-28\ \mu\text{N})\hat{\imath}$$

EVALUATE: As a check, note that the magnitude of q_2 is three times that of q_1, but q_2 is twice as far from q_3 as q_1. Equation (21.2) then says that $F_{2\text{ on }3}$ must be $3/2^2 = 3/4 = 0.75$ as large as $F_{1\text{ on }3}$. This agrees with our calculated values: $F_{2\text{ on }3}/F_{1\text{ on }3} = (84\ \mu\text{N})/(112\ \mu\text{N}) = 0.75$. Because $F_{2\text{ on }3}$ is the weaker force, the direction of the net force is that of $\vec{F}_{1\text{ on }3}$—that is, in the negative x-direction.

EXAMPLE 21.4 VECTOR ADDITION OF ELECTRIC FORCES IN A PLANE

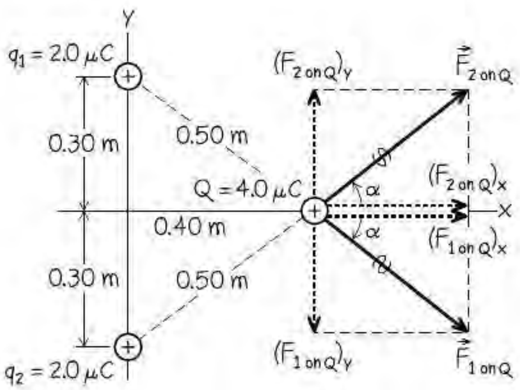
SOLUTION

Two equal positive charges $q_1 = q_2 = 2.0\ \mu\text{C}$ are located at $x = 0$, $y = 0.30$ m and $x = 0$, $y = -0.30$ m, respectively. What are the magnitude and direction of the total electric force that q_1 and q_2 exert on a third charge $Q = 4.0\ \mu\text{C}$ at $x = 0.40$ m, $y = 0$?

SOLUTION

IDENTIFY and SET UP: As in Example 21.3, we must compute the force that each charge exerts on Q and then find the vector sum of those forces. **Figure 21.14** shows the situation. Since the three charges do not all lie on a line, the best way to calculate the forces is to use components.

21.14 Our sketch for this problem.



EXECUTE: Figure 21.14 shows the forces $\vec{F}_{1\text{ on }Q}$ and $\vec{F}_{2\text{ on }Q}$ due to the identical charges q_1 and q_2, which are at equal distances from Q. From Coulomb's law, *both* forces have magnitude

$$F_{1\text{ or }2\text{ on }Q} = (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2) \times \frac{(4.0 \times 10^{-6}\ \text{C})(2.0 \times 10^{-6}\ \text{C})}{(0.50\ \text{m})^2} = 0.29\ \text{N}$$

The x-components of the two forces are equal:

$$(F_{1\text{ or }2\text{ on }Q})_x = (F_{1\text{ or }2\text{ on }Q})\cos\alpha = (0.29\ \text{N})\frac{0.40\ \text{m}}{0.50\ \text{m}} = 0.23\ \text{N}$$

From symmetry we see that the y-components of the two forces are equal and opposite. Hence their sum is zero and the total force $\vec{F}$ on Q has only an x-component $F_x = 0.23\ \text{N} + 0.23\ \text{N} = 0.46\ \text{N}$. The total force on Q is in the $+x$-direction, with magnitude 0.46 N.

EVALUATE: The total force on Q points neither directly away from q_1 nor directly away from q_2. Rather, this direction is a compromise that points away from the *system* of charges q_1 and q_2. Can you see that the total force would *not* be in the $+x$-direction if q_1 and q_2 were not equal or if the geometrical arrangement of the changes were not so symmetric?

TEST YOUR UNDERSTANDING OF SECTION 21.3 Suppose that charge q_2 in Example 21.4 were $-2.0\ \mu C$. In this case, the total electric force on Q would be (i) in the positive x-direction; (ii) in the negative x-direction; (iii) in the positive y-direction; (iv) in the negative y-direction; (v) zero; (vi) none of these. ❙

21.4 ELECTRIC FIELD AND ELECTRIC FORCES

When two electrically charged particles in empty space interact, how does each one know the other is there? We can begin to answer this question, and at the same time reformulate Coulomb's law in a very useful way, by using the concept of *electric field*.

Electric Field

To introduce this concept, let's look at the mutual repulsion of two positively charged bodies A and B (**Fig. 21.15a**). Suppose B has charge q_0, and let $\vec{F}_0$ be the electric force of A on B. One way to think about this force is as an "action-at-a-distance" force—that is, as a force that acts across empty space without needing physical contact between A and B. (Gravity can also be thought of as an "action-at-a-distance" force.) But a more fruitful way to visualize the repulsion between A and B is as a two-stage process. We first envision that body A, as a result of the charge that it carries, somehow *modifies the properties of the space around it*. Then body B, as a result of the charge that *it* carries, senses how space has been modified at its position. The response of body B is to experience the force $\vec{F}_0$.

To clarify how this two-stage process occurs, we first consider body A by itself: We remove body B and label its former position as point P (Fig. 21.15b). We say that the charged body A produces or causes an **electric field** at point P (and at all other points in the neighborhood). This electric field is present at P even if there is no charge at P; it is a consequence of the charge on body A only. If a point charge q_0 is then placed at point P, it experiences the force $\vec{F}_0$. We take the point of view that this force is exerted on q_0 *by the field* at P (Fig. 21.15c). Thus the electric field is the intermediary through which A communicates its presence to q_0. Because the point charge q_0 would experience a force at *any* point in the neighborhood of A, the electric field that A produces exists at all points in the region around A.

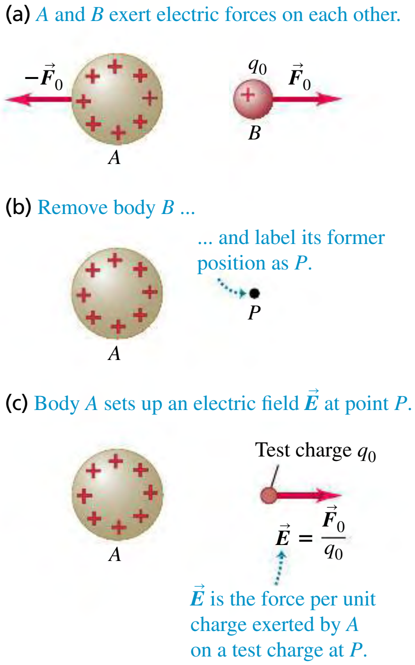
We can likewise say that the point charge q_0 produces an electric field in the space around it and that this electric field exerts the force $-\vec{F}_0$ on body A. For each force (the force of A on q_0 and the force of q_0 on A), one charge sets up an electric field that exerts a force on the second charge. We emphasize that this is an *interaction* between *two* charged bodies. A single charge produces an electric field in the surrounding space, but this electric field cannot exert a net force on the charge that created it; as we discussed in Section 4.3, a body cannot exert a net force on itself. (If this wasn't true, you would be able to lift yourself to the ceiling by pulling up on your belt!)

The electric force on a charged body is exerted by the electric field created by *other* charged bodies.

To find out experimentally whether there is an electric field at a particular point, we place a small charged body, which we call a **test charge,** at the point (Fig. 21.15c). If the test charge experiences an electric force, then there is an electric field at that point. This field is produced by charges other than q_0.

Force is a vector quantity, so electric field is also a vector quantity. (Note the use of vector signs as well as boldface letters and plus, minus, and equals signs in the following discussion.) We define the *electric field* $\vec{E}$ at a point as the electric

21.15 A charged body creates an electric field in the space around it.



BIO Application Sharks and the "Sixth Sense" Sharks have the ability to locate prey (such as flounder and other bottom-dwelling fish) that are completely hidden beneath the sand at the bottom of the ocean. They do this by sensing the weak electric fields produced by muscle contractions in their prey. Sharks derive their sensitivity to electric fields (a "sixth sense") from jelly-filled canals in their bodies. These canals end in pores on the shark's skin (shown in this photograph). An electric field as weak as 5×10^{-7} N/C causes charge flow within the canals and triggers a signal in the shark's nervous system. Because the shark has canals with different orientations, it can measure different components of the electric-field vector and hence determine the direction of the field.

force $\vec{F}_0$ experienced by a test charge q_0 at the point, divided by the charge q_0. That is, the electric field at a certain point is equal to the *electric force per unit charge* experienced by a charge at that point:

Electric field = electric force per unit charge

$$\vec{E} = \frac{\vec{F}_0}{q_0} \tag{21.3}$$

Electric force on a test charge q_0 due to other charges

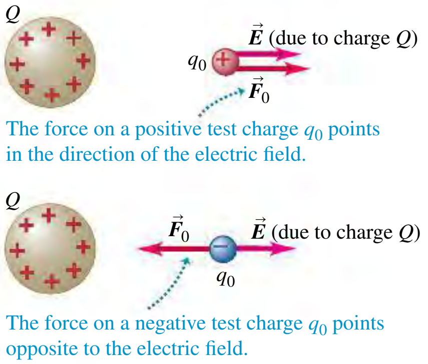
Value of test charge

In SI units, in which the unit of force is 1 N and the unit of charge is 1 C, the unit of electric-field magnitude is 1 newton per coulomb (1 N/C).

If the field $\vec{E}$ at a certain point is known, rearranging Eq. (21.3) gives the force $\vec{F}_0$ experienced by a point charge q_0 placed at that point. This force is just equal to the electric field $\vec{E}$ produced at that point by charges other than q_0, multiplied by the charge q_0:

$$\vec{F}_0 = q_0\vec{E} \quad \text{(force exerted on a point charge } q_0 \text{ by an electric field } \vec{E}\text{)} \tag{21.4}$$

21.16 The force $\vec{F}_0 = q_0\vec{E}$ exerted on a point charge q_0 placed in an electric field $\vec{E}$.



The charge q_0 can be either positive or negative. If q_0 is *positive,* the force $\vec{F}_0$ experienced by the charge is in the same direction as $\vec{E}$; if q_0 is *negative,* $\vec{F}_0$ and $\vec{E}$ are in opposite directions (**Fig. 21.16**).

While the electric field concept may be new to you, the basic idea—that one body sets up a field in the space around it and a second body responds to that field—is one that you've actually used before. Compare Eq. (21.4) to the familiar expression for the gravitational force $\vec{F}_g$ that the earth exerts on a mass m_0:

$$\vec{F}_g = m_0\vec{g} \tag{21.5}$$

In this expression, $\vec{g}$ is the acceleration due to gravity. If we divide both sides of Eq. (21.5) by the mass m_0, we obtain

$$\vec{g} = \frac{\vec{F}_g}{m_0}$$

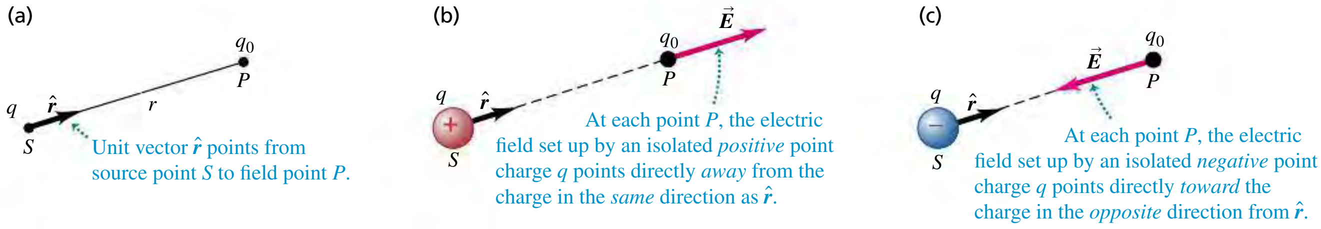
Thus $\vec{g}$ can be regarded as the gravitational force per unit mass. By analogy to Eq. (21.3), we can interpret $\vec{g}$ as the *gravitational field.* Thus we treat the gravitational interaction between the earth and the mass m_0 as a two-stage process: The earth sets up a gravitational field $\vec{g}$ in the space around it, and this gravitational field exerts a force given by Eq. (21.5) on the mass m_0 (which we can regard as a *test mass*). The gravitational field $\vec{g}$, or gravitational force per unit mass, is a useful concept because it does not depend on the mass of the body on which the gravitational force is exerted; likewise, the electric field $\vec{E}$, or electric force per unit charge, is useful because it does not depend on the charge of the body on which the electric force is exerted.

CAUTION **$\vec{F}_0 = q_0\vec{E}$ is for *point* test charges only** The electric force experienced by a test charge q_0 can vary from point to point, so the electric field can also be different at different points. For this reason, use Eq. (21.4) to find the electric force on a *point* charge only. If a charged body is large enough in size, the electric field $\vec{E}$ may be noticeably different in magnitude and direction at different points on the body, and calculating the net electric force on it can be complicated.

Electric Field of a Point Charge

If the source distribution is a point charge q, it is easy to find the electric field that it produces. We call the location of the charge the **source point,** and we call the point P where we are determining the field the **field point.** It is also useful to introduce a *unit vector* $\hat{r}$ that points along the line from source point

21.17 The electric field $\vec{E}$ produced at point P by an isolated point charge q at S. Note that in both (b) and (c), $\vec{E}$ is *produced* by q [see Eq. (21.7)] but *acts* on the charge q_0 at point P [see Eq. (21.4)].



to field point (**Fig. 21.17a**). This unit vector is equal to the displacement vector $\vec{r}$ from the source point to the field point, divided by the distance $r = |\vec{r}|$ between these two points; that is, $\hat{r} = \vec{r}/r$. If we place a small test charge q_0 at the field point P, at a distance r from the source point, the magnitude F_0 of the force is given by Coulomb's law, Eq. (21.2):

$$F_0 = \frac{1}{4\pi\epsilon_0}\frac{|qq_0|}{r^2}$$

From Eq. (21.3) the magnitude E of the electric field at P is

$$E = \frac{1}{4\pi\epsilon_0}\frac{|q|}{r^2} \quad \text{(magnitude of electric field of a point charge)} \qquad (21.6)$$

Using the unit vector $\hat{r}$, we can write a *vector* equation that gives both the magnitude and direction of the electric field $\vec{E}$:

Electric field due to a point charge

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r} \qquad (21.7)$$

Value of point charge
Unit vector from point charge toward where field is measured
Distance from point charge to where field is measured
Electric constant

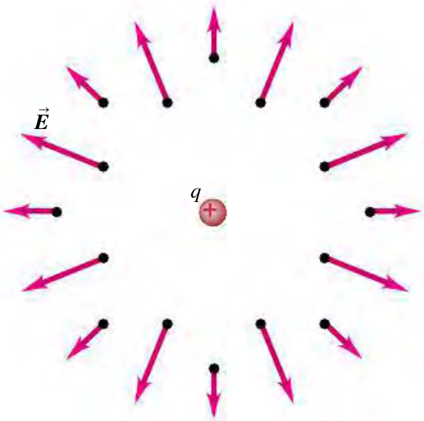
By definition, the electric field of a point charge always points *away from* a positive charge (that is, in the same direction as $\hat{r}$; see Fig. 21.17b) but *toward* a negative charge (that is, in the direction opposite $\hat{r}$; see Fig. 21.17c).

We have emphasized calculating the electric field $\vec{E}$ at a certain point. But since $\vec{E}$ can vary from point to point, it is not a single vector quantity but rather an *infinite* set of vector quantities, one associated with each point in space. This is an example of a **vector field. Figure 21.18** shows a number of the field vectors produced by a positive or negative point charge. If we use a rectangular (x, y, z) coordinate system, each component of $\vec{E}$ at any point is in general a function of the coordinates (x, y, z) of the point. We can represent the functions as $E_x(x, y, z)$, $E_y(x, y, z)$, and $E_z(x, y, z)$. Another example of a vector field is the velocity $\vec{v}$ of wind currents; the magnitude and direction of $\vec{v}$, and hence its vector components, vary from point to point in the atmosphere.

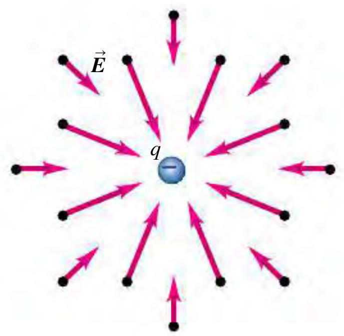
In some situations the magnitude and direction of the field (and hence its vector components) have the same values everywhere throughout a certain region; we then say that the field is *uniform* in this region. An important example of this is the electric field inside a *conductor.* If there is an electric field within a conductor, the field exerts a force on every charge in the conductor, giving the free charges a net motion. By definition an electrostatic situation is one in which the charges have *no* net motion. We conclude that *in electrostatics the electric field at every point within the material of a conductor must be zero.* (Note that we are not saying that the field is necessarily zero in a *hole* inside a conductor.)

21.18 A point charge q produces an electric field $\vec{E}$ at *all* points in space. The field strength decreases with increasing distance.

(a) The field produced by a positive point charge points *away from* the charge.



(b) The field produced by a negative point charge points *toward* the charge.



In summary, our description of electric interactions has two parts. First, a given charge distribution acts as a source of electric field. Second, the electric field exerts a force on any charge that is present in the field. Our analysis often has two corresponding steps: first, calculating the field caused by a source charge distribution; second, looking at the effect of the field in terms of force and motion. The second step often involves Newton's laws as well as the principles of electric interactions. In the next section we show how to calculate fields caused by various source distributions, but first here are three examples of calculating the field due to a point charge and of finding the force on a charge due to a given field $\vec{E}$.

EXAMPLE 21.5 ELECTRIC-FIELD MAGNITUDE FOR A POINT CHARGE

SOLUTION

What is the magnitude of the electric field $\vec{E}$ at a field point 2.0 m from a point charge $q = 4.0\ \text{nC}$?

SOLUTION

IDENTIFY and SET UP: This problem concerns the electric field due to a point charge. We are given the magnitude of the charge and the distance from the charge to the field point, so we use Eq. (21.6) to calculate the field magnitude E.

EXECUTE: From Eq. (21.6),

$$E = \frac{1}{4\pi\epsilon_0}\frac{|q|}{r^2} = (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{4.0 \times 10^{-9}\ \text{C}}{(2.0\ \text{m})^2}$$
$$= 9.0\ \text{N/C}$$

EVALUATE: Our result $E = 9.0\ \text{N/C}$ means that if we placed a 1.0-C charge at a point 2.0 m from q, it would experience a 9.0-N force. The force on a 2.0-C charge at that point would be $(2.0\ \text{C})(9.0\ \text{N/C}) = 18\ \text{N}$, and so on.

EXAMPLE 21.6 ELECTRIC-FIELD VECTOR FOR A POINT CHARGE

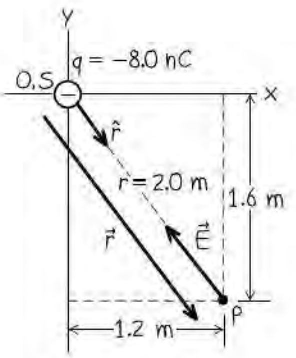
SOLUTION

A point charge $q = -8.0\ \text{nC}$ is located at the origin. Find the electric-field vector at the field point $x = 1.2\ \text{m}$, $y = -1.6\ \text{m}$.

SOLUTION

IDENTIFY and SET UP: We must find the electric-field *vector* $\vec{E}$ due to a point charge. **Figure 21.19** shows the situation. We use Eq. (21.7); to do this, we must find the distance r from the source point S (the position of the charge q, which in this example is at the origin O) to the field point P, and we must obtain an expression for the unit vector $\hat{r} = \vec{r}/r$ that points from S to P.

21.19 Our sketch for this problem.



EXECUTE: The distance from S to P is

$$r = \sqrt{x^2 + y^2} = \sqrt{(1.2\ \text{m})^2 + (-1.6\ \text{m})^2} = 2.0\ \text{m}$$

The unit vector $\hat{r}$ is then

$$\hat{r} = \frac{\vec{r}}{r} = \frac{x\hat{\imath} + y\hat{\jmath}}{r}$$
$$= \frac{(1.2\ \text{m})\hat{\imath} + (-1.6\ \text{m})\hat{\jmath}}{2.0\ \text{m}} = 0.60\hat{\imath} - 0.80\hat{\jmath}$$

Then, from Eq. (21.7),

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$$
$$= (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{(-8.0 \times 10^{-9}\ \text{C})}{(2.0\ \text{m})^2}(0.60\hat{\imath} - 0.80\hat{\jmath})$$
$$= (-11\ \text{N/C})\hat{\imath} + (14\ \text{N/C})\hat{\jmath}$$

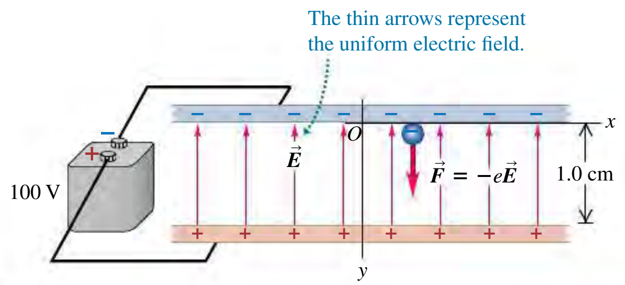
EVALUATE: Since q is negative, $\vec{E}$ points from the field point to the charge (the source point), in the direction opposite to $\hat{r}$ (compare Fig. 21.17c). We leave the calculation of the magnitude and direction of $\vec{E}$ to you (see Exercise 21.30).

EXAMPLE 21.7 ELECTRON IN A UNIFORM FIELD

When the terminals of a battery are connected to two parallel conducting plates with a small gap between them, the resulting charges on the plates produce a nearly uniform electric field $\vec{E}$ between the plates. (In the next section we'll see why this is.) If the plates are 1.0 cm apart and are connected to a 100-volt battery as shown in **Fig. 21.20**, the field is vertically upward and has magnitude $E = 1.00 \times 10^4$ N/C. (a) If an electron (charge $-e = -1.60 \times 10^{-9}$ C, mass $m = 9.11 \times 10^{-31}$ kg) is released from rest at the upper plate, what is its acceleration? (b) What speed and kinetic energy does it acquire while traveling 1.0 cm to the lower plate? (c) How long does it take to travel this distance?

SOLUTION

IDENTIFY and SET UP: This example involves the relationship between electric field and electric force. It also involves the relationship between force and acceleration, the definition of kinetic energy, and the kinematic relationships among acceleration, distance, velocity, and time. Figure 21.20 shows our coordinate system. We are given the electric field, so we use Eq. (21.4) to find the force on the electron and Newton's second law to find its acceleration. Because the field is uniform, the force is constant and we can use the constant-acceleration formulas from Chapter 2 to find the electron's velocity and travel time. We find the kinetic energy from $K = \frac{1}{2}mv^2$.



21.20 A uniform electric field between two parallel conducting plates connected to a 100-volt battery. (The separation of the plates is exaggerated in this figure relative to the dimensions of the plates.)

EXECUTE: (a) Although $\vec{E}$ is upward (in the $+y$-direction), $\vec{F}$ is downward (because the electron's charge is negative) and so F_y is negative. Because F_y is constant, the electron's acceleration is constant:

$$a_y = \frac{F_y}{m} = \frac{-eE}{m} = \frac{(-1.60 \times 10^{-19}\ \text{C})(1.00 \times 10^4\ \text{N/C})}{9.11 \times 10^{-31}\ \text{kg}}$$
$$= -1.76 \times 10^{15}\ \text{m/s}^2$$

(b) The electron starts from rest, so its motion is in the y-direction only (the direction of the acceleration). We can find the electron's speed at any position y from the constant-acceleration Eq. (2.13), $v_y^2 = v_{0y}^2 + 2a_y(y - y_0)$. We have $v_{0y} = 0$ and $y_0 = 0$, so at $y = -1.0$ cm $= -1.0 \times 10^{-2}$ m we have

$$|v_y| = \sqrt{2a_y y} = \sqrt{2(-1.76 \times 10^{15}\ \text{m/s}^2)(-1.0 \times 10^{-2}\ \text{m})}$$
$$= 5.9 \times 10^6\ \text{m/s}$$

The velocity is downward, so $v_y = -5.9 \times 10^6$ m/s. The electron's kinetic energy is

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}(9.11 \times 10^{-31}\ \text{kg})(5.9 \times 10^6\ \text{m/s})^2$$
$$= 1.6 \times 10^{-17}\ \text{J}$$

(c) From Eq. (2.8) for constant acceleration, $v_y = v_{0y} + a_y t$,

$$t = \frac{v_y - v_{0y}}{a_y} = \frac{(-5.9 \times 10^6\ \text{m/s}) - (0\ \text{m/s})}{-1.76 \times 10^{15}\ \text{m/s}^2}$$
$$= 3.4 \times 10^{-9}\ \text{s}$$

EVALUATE: Our results show that in problems concerning subatomic particles such as electrons, many quantities—including acceleration, speed, kinetic energy, and time—will have *very* different values from those typical of everyday objects such as baseballs and automobiles.

TEST YOUR UNDERSTANDING OF SECTION 21.4 (a) A negative point charge moves along a straight-line path directly toward a stationary positive point charge. Which aspect(s) of the electric force on the negative point charge will remain constant as it moves? (i) Magnitude; (ii) direction; (iii) both magnitude and direction; (iv) neither magnitude nor direction. (b) A negative point charge moves along a circular orbit around a positive point charge. Which aspect(s) of the electric force on the negative point charge will remain constant as it moves? (i) Magnitude; (ii) direction; (iii) both magnitude and direction; (iv) neither magnitude nor direction. ❙

21.5 ELECTRIC-FIELD CALCULATIONS

Equation (21.7) gives the electric field caused by a single point charge. But in most realistic situations that involve electric fields and forces, we encounter charge that is *distributed* over space. The charged plastic and glass rods in Fig. 21.1 have electric charge distributed over their surfaces, as does the imaging drum of a laser printer (Fig. 21.2). In this section we'll learn to calculate electric fields caused by various distributions of electric charge. Calculations of this kind are of tremendous importance for technological applications of electric forces.

DATA *SPEAKS*

Electric Force and Electric Field

When students were given a problem involving electric force and electric field, more than 28% gave an incorrect response. Common errors:

- Forgetting that the electric field $\vec{E}$ experienced by a point charge does not depend on the value of that point charge. The value of $\vec{E}$ is determined by the charges that produce the field, not the charge that experiences it.
- Forgetting that $\vec{E}$ is a vector. If the field $\vec{E}$ at point P is due to two or more point charges, $\vec{E}$ is the vector sum of the fields due to the individual charges. In general, this is not the sum of the magnitudes of these fields.

To determine the trajectories of atomic nuclei in an accelerator for cancer radiotherapy or of charged particles in a semiconductor electronic device, you have to know the detailed nature of the electric field acting on the charges.

The Superposition of Electric Fields

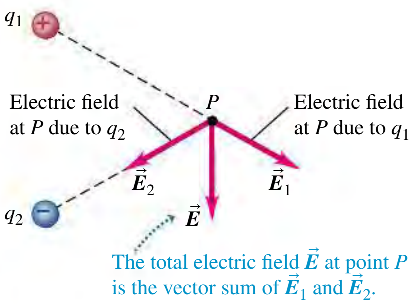
To find the field caused by a charge distribution, we imagine the distribution to be made up of many point charges $q_1, q_2, q_3, \ldots$. (This is actually quite a realistic description, since we have seen that charge is carried by electrons and protons that are so small as to be almost pointlike.) At any given point P, each point charge produces its own electric field $\vec{E}_1, \vec{E}_2, \vec{E}_3, \ldots$, so a test charge q_0 placed at P experiences a force $\vec{F}_1 = q_0\vec{E}_1$ from charge q_1, a force $\vec{F}_2 = q_0\vec{E}_2$ from charge q_2, and so on. From the principle of superposition of forces discussed in Section 21.3, the *total* force $\vec{F}_0$ that the charge distribution exerts on q_0 is the vector sum of these individual forces:

$$\vec{F}_0 = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \cdots = q_0\vec{E}_1 + q_0\vec{E}_2 + q_0\vec{E}_3 + \cdots$$

The combined effect of all the charges in the distribution is described by the *total* electric field $\vec{E}$ at point P. From the definition of electric field, Eq. (21.3), this is

$$\vec{E} = \frac{\vec{F}_0}{q_0} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \cdots$$

21.21 Illustrating the principle of superposition of electric fields.



The total electric field at P is the vector sum of the fields at P due to each point charge in the charge distribution (**Fig. 21.21**). This statement is the **principle of superposition of electric fields.**

When charge is distributed along a line, over a surface, or through a volume, a few additional terms are useful. For a line charge distribution (such as a long, thin, charged plastic rod), we use λ (the Greek letter lambda) to represent the **linear charge density** (charge per unit length, measured in C/m). When charge is distributed over a surface (such as the surface of the imaging drum of a laser printer), we use σ (sigma) to represent the **surface charge density** (charge per unit area, measured in C/m^2). And when charge is distributed through a volume, we use ρ (rho) to represent the **volume charge density** (charge per unit volume, C/m^3).

Some of the calculations in the following examples may look complex. After you've worked through the examples one step at a time, the process will seem less formidable. We will use many of the calculational techniques in these examples in Chapter 28 to calculate the *magnetic* fields caused by charges in motion.

PROBLEM-SOLVING STRATEGY 21.2 ELECTRIC-FIELD CALCULATIONS

IDENTIFY *the relevant concepts:* Use the principle of superposition to calculate the electric field due to a discrete or continuous charge distribution.

SET UP *the problem* using the following steps:

1. Make a drawing showing the locations of the charges and your choice of coordinate axes.
2. On your drawing, indicate the position of the *field point P* (the point at which you want to calculate the electric field $\vec{E}$).

EXECUTE *the solution* as follows:

1. Use consistent units. Distances must be in meters and charge must be in coulombs. If you are given centimeters or nanocoulombs, don't forget to convert.
2. Distinguish between the source point S and the field point P. The field produced by a point charge always points from S to P if the charge is positive, and from P to S if the charge is negative.
3. Use *vector* addition when applying the principle of superposition; review the treatment of vector addition in Chapter 1 if necessary.
4. Simplify your calculations by exploiting any symmetries in the charge distribution.
5. If the charge distribution is continuous, define a small element of charge that can be considered as a point, find its electric field at P, and find a way to add the fields of all the charge elements by doing an integral. Usually it is easiest to do this for each component of $\vec{E}$ separately, so you may need to evaluate more than one integral. Ensure that the limits on your integrals are correct; especially when the situation has symmetry, don't count a charge twice.

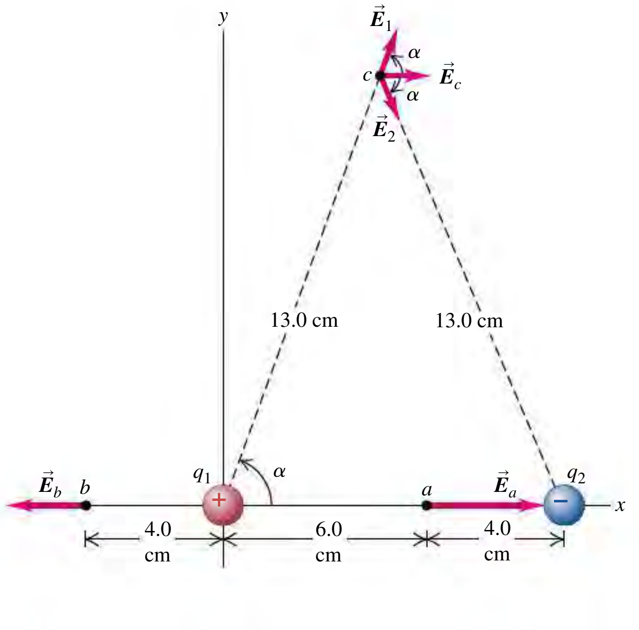
EVALUATE *your answer:* Check that the direction of $\vec{E}$ is reasonable. If your result for the electric-field magnitude E is a function of position (say, the coordinate x), check your result in any limits for which you know what the magnitude should be. When possible, check your answer by calculating it in a different way.

SOLUTION

EXAMPLE 21.8 FIELD OF AN ELECTRIC DIPOLE

Point charges $q_1 = +12$ nC and $q_2 = -12$ nC are 0.100 m apart (**Fig. 21.22**). (Such pairs of point charges with equal magnitude and opposite sign are called *electric dipoles*.) Compute the electric field caused by q_1, the field caused by q_2, and the total field (a) at point a; (b) at point b; and (c) at point c.

21.22 Electric field at three points, a, b, and c, set up by charges q_1 and q_2, which form an electric dipole.



SOLUTION

IDENTIFY and SET UP: We must find the total electric field at various points due to two point charges. We use the principle of superposition: $\vec{E} = \vec{E}_1 + \vec{E}_2$. Figure 21.22 shows the coordinate system and the locations of the field points a, b, and c.

EXECUTE: At each field point, $\vec{E}$ depends on $\vec{E}_1$ and $\vec{E}_2$ there; we first calculate the magnitudes E_1 and E_2 at each field point. At a the magnitude of the field $\vec{E}_{1a}$ caused by q_1 is

$$\begin{aligned} E_{1a} &= \frac{1}{4\pi\epsilon_0}\frac{|q_1|}{r^2} \\ &= (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{12 \times 10^{-9}\ \text{C}}{(0.060\ \text{m})^2} \\ &= 3.0 \times 10^4\ \text{N/C} \end{aligned}$$

We calculate the other field magnitudes in a similar way. The results are

$$\begin{aligned} E_{1a} &= 3.0 \times 10^4\ \text{N/C} \\ E_{1b} &= 6.8 \times 10^4\ \text{N/C} \\ E_{1c} &= 6.39 \times 10^3\ \text{N/C} \\ E_{2a} &= 6.8 \times 10^4\ \text{N/C} \\ E_{2b} &= 0.55 \times 10^4\ \text{N/C} \\ E_{2c} &= E_{1c} = 6.39 \times 10^3\ \text{N/C} \end{aligned}$$

The *directions* of the corresponding fields are in all cases *away* from the positive charge q_1 and *toward* the negative charge q_2.

(a) At a, $\vec{E}_{1a}$ and $\vec{E}_{2a}$ are both directed to the right, so

$$\vec{E}_a = E_{1a}\hat{\imath} + E_{2a}\hat{\imath} = (9.8 \times 10^4\ \text{N/C})\hat{\imath}$$

(b) At b, $\vec{E}_{1b}$ is directed to the left and $\vec{E}_{2b}$ is directed to the right, so

$$\vec{E}_b = -E_{1b}\hat{\imath} + E_{2b}\hat{\imath} = (-6.2 \times 10^4\ \text{N/C})\hat{\imath}$$

(c) Figure 21.22 shows the directions of $\vec{E}_1$ and $\vec{E}_2$ at c. Both vectors have the same x-component:

$$\begin{aligned} E_{1cx} = E_{2cx} = E_{1c}\cos\alpha &= (6.39 \times 10^3\ \text{N/C})\left(\tfrac{5}{13}\right) \\ &= 2.46 \times 10^3\ \text{N/C} \end{aligned}$$

From symmetry, E_{1y} and E_{2y} are equal and opposite, so their sum is zero. Hence

$$\vec{E}_c = 2(2.46 \times 10^3\ \text{N/C})\hat{\imath} = (4.9 \times 10^3\ \text{N/C})\hat{\imath}$$

EVALUATE: We can also find $\vec{E}_c$ by using Eq. (21.7) for the field of a point charge. The displacement vector $\vec{r}_1$ from q_1 to point c is $\vec{r}_1 = r\cos\alpha\hat{\imath} + r\sin\alpha\hat{\jmath}$. Hence the unit vector that points from q_1 to point c is $\hat{r}_1 = \vec{r}_1/r = \cos\alpha\hat{\imath} + \sin\alpha\hat{\jmath}$. By symmetry, the unit vector that points from q_2 to point c has the opposite x-component but the same y-component: $\hat{r}_2 = -\cos\alpha\hat{\imath} + \sin\alpha\hat{\jmath}$. We can now use Eq. (21.7) to write the fields $\vec{E}_{1c}$ and $\vec{E}_{2c}$ at c in vector form, then find their sum. Since $q_2 = -q_1$ and the distance r to c is the same for both charges,

$$\begin{aligned} \vec{E}_c = \vec{E}_{1c} + \vec{E}_{2c} &= \frac{1}{4\pi\epsilon_0}\frac{q_1}{r^2}\hat{r}_1 + \frac{1}{4\pi\epsilon_0}\frac{q_2}{r^2}\hat{r}_2 \\ &= \frac{1}{4\pi\epsilon_0 r^2}(q_1\hat{r}_1 + q_2\hat{r}_2) \\ &= \frac{q_1}{4\pi\epsilon_0 r^2}(\hat{r}_1 - \hat{r}_2) \\ &= \frac{1}{4\pi\epsilon_0}\frac{q_1}{r^2}(2\cos\alpha\hat{\imath}) \\ &= 2(9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{12 \times 10^{-9}\ \text{C}}{(0.13\ \text{m})^2}\left(\tfrac{5}{13}\right)\hat{\imath} \\ &= (4.9 \times 10^3\ \text{N/C})\hat{\imath} \end{aligned}$$

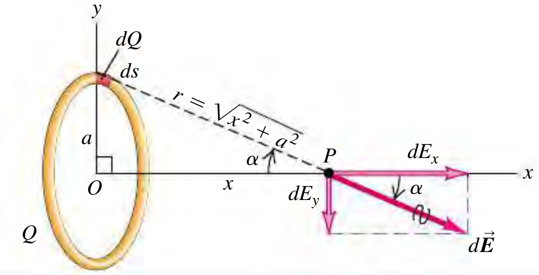
This is the same as we calculated in part (c).

SOLUTION

EXAMPLE 21.9 FIELD OF A RING OF CHARGE

Charge Q is uniformly distributed around a conducting ring of radius a (**Fig. 21.23**). Find the electric field at a point P on the ring axis at a distance x from its center.

21.23 Calculating the electric field on the axis of a ring of charge. In this figure, the charge is assumed to be positive.



SOLUTION

IDENTIFY and SET UP: This is a problem in the superposition of electric fields. Each bit of charge around the ring produces an electric field at an arbitrary point on the x-axis; our target variable is the total field at this point due to all such bits of charge.

EXECUTE: We divide the ring into infinitesimal segments ds as shown in Fig. 21.23. In terms of the linear charge density $\lambda = Q/2\pi a$, the charge in a segment of length ds is $dQ = \lambda\, ds$. Consider two identical segments, one as shown in the figure at $y = a$ and another halfway around the ring at $y = -a$. From Example 21.4, we see that the net force $d\vec{F}$ they exert on a point test charge at P, and thus their net field $d\vec{E}$, are directed along the x-axis. The same is true for any such pair of segments around the ring, so the *net* field at P is along the x-axis: $\vec{E} = E_x\hat{\imath}$.

To calculate E_x, note that the square of the distance r from a single ring segment to the point P is $r^2 = x^2 + a^2$. Hence the magnitude of this segment's contribution $d\vec{E}$ to the electric field at P is

$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{x^2 + a^2}$$

The x-component of this field is $dE_x = dE\cos\alpha$. We know $dQ = \lambda\, ds$ and Fig. 21.23 shows that $\cos\alpha = x/r = x/(x^2 + a^2)^{1/2}$, so

$$dE_x = dE\cos\alpha = \frac{1}{4\pi\epsilon_0}\frac{dQ}{x^2 + a^2}\frac{x}{\sqrt{x^2 + a^2}}$$

$$= \frac{1}{4\pi\epsilon_0}\frac{\lambda x}{(x^2 + a^2)^{3/2}}ds$$

To find E_x we integrate this expression over the entire ring—that is, for s from 0 to $2\pi a$ (the circumference of the ring). The integrand has the same value for all points on the ring, so it can be taken outside the integral. Hence we get

$$E_x = \int dE_x = \frac{1}{4\pi\epsilon_0}\frac{\lambda x}{(x^2 + a^2)^{3/2}}\int_0^{2\pi a} ds$$

$$= \frac{1}{4\pi\epsilon_0}\frac{\lambda x}{(x^2 + a^2)^{3/2}}(2\pi a)$$

$$\vec{E} = E_x\hat{\imath} = \frac{1}{4\pi\epsilon_0}\frac{Qx}{(x^2 + a^2)^{3/2}}\hat{\imath} \qquad (21.8)$$

EVALUATE: Equation (21.8) shows that $\vec{E} = \mathbf{0}$ at the center of the ring ($x = 0$). This makes sense; charges on opposite sides of the ring push in opposite directions on a test charge at the center, and the vector sum of each such pair of forces is zero. When the field point P is much farther from the ring than the ring's radius, we have $x \gg a$ and the denominator in Eq. (21.8) becomes approximately equal to x^3. In this limit the electric field at P is

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{Q}{x^2}\hat{\imath}$$

That is, when the ring is so far away that its radius is negligible in comparison to the distance x, its field is the same as that of a point charge.

SOLUTION

EXAMPLE 21.10 FIELD OF A CHARGED LINE SEGMENT

Positive charge Q is distributed uniformly along the y-axis between $y = -a$ and $y = +a$. Find the electric field at point P on the x-axis at a distance x from the origin.

SOLUTION

IDENTIFY and SET UP: Figure 21.24 shows the situation. As in Example 21.9, we must find the electric field due to a continuous distribution of charge. Our target variable is an expression for the electric field at P as a function of x. The x-axis is a perpendicular bisector of the segment, so we can use a symmetry argument.

EXECUTE: We divide the line charge of length $2a$ into infinitesimal segments of length dy. The linear charge density is $\lambda = Q/2a$, and the charge in a segment is $dQ = \lambda\, dy = (Q/2a)dy$. The distance r from a segment at height y to the field point P is $r = (x^2 + y^2)^{1/2}$, so the magnitude of the field at P due to the segment at height y is

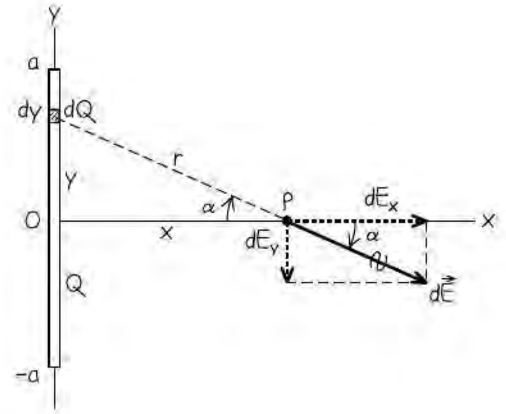
$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\frac{dy}{(x^2 + y^2)}$$

Figure 21.24 shows that the x- and y-components of this field are $dE_x = dE\cos\alpha$ and $dE_y = -dE\sin\alpha$, where $\cos\alpha = x/r$ and $\sin\alpha = y/r$. Hence

$$dE_x = \frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\frac{x\,dy}{(x^2 + y^2)^{3/2}}$$

$$dE_y = -\frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\frac{y\,dy}{(x^2 + y^2)^{3/2}}$$

21.24 Our sketch for this problem.



To find the total field at P, we must sum the fields from all segments along the line—that is, we must integrate from $y = -a$ to $y = +a$. You should work out the details of the integration (a table of integrals will help). The results are

$$E_x = \frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\int_{-a}^{+a}\frac{x\,dy}{(x^2 + y^2)^{3/2}} = \frac{Q}{4\pi\epsilon_0}\frac{1}{x\sqrt{x^2 + a^2}}$$

$$E_y = -\frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\int_{-a}^{+a}\frac{y\,dy}{(x^2 + y^2)^{3/2}} = 0$$

or, in vector form,

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{Q}{x\sqrt{x^2 + a^2}}\hat{\imath} \tag{21.9}$$

$\vec{E}$ points away from the line of charge if λ is positive and toward the line of charge if λ is negative.

EVALUATE: Using a symmetry argument as in Example 21.9, we could have guessed that E_y would be zero; if we place a positive test charge at P, the upper half of the line of charge pushes downward on it, and the lower half pushes up with equal magnitude. Symmetry also tells us that the upper and lower halves of the segment contribute equally to the total field at P.

If the segment is very *short* (or the field point is very far from the segment) so that $x \gg a$, we can ignore a in the denominator of Eq. (21.9). Then the field becomes that of a point charge, just as in Example 21.9:

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{Q}{x^2}\hat{\imath}$$

To see what happens if the segment is very *long* (or the field point is very close to it) so that $a \gg x$, we first rewrite Eq. (21.9) slightly:

$$\vec{E} = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{x\sqrt{(x^2/a^2) + 1}}\hat{\imath} \tag{21.10}$$

In the limit $a \gg x$ we can ignore x^2/a^2 in the denominator of Eq. (21.10), so

$$\vec{E} = \frac{\lambda}{2\pi\epsilon_0 x}\hat{\imath}$$

This is the field of an *infinitely long* line of charge. At any point P at a perpendicular distance r from the line in *any* direction, $\vec{E}$ has magnitude

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \qquad \text{(infinite line of charge)}$$

Note that this field is proportional to $1/r$ rather than to $1/r^2$ as for a point charge.

There's really no such thing in nature as an infinite line of charge. But when the field point is close enough to the line, there's very little difference between the result for an infinite line and the real-life finite case. For example, if the distance r of the field point from the center of the line is 1% of the length of the line, the value of E differs from the infinite-length value by less than 0.02%.

EXAMPLE 21.11 FIELD OF A UNIFORMLY CHARGED DISK

SOLUTION

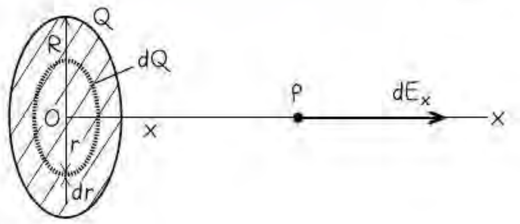
A nonconducting disk of radius R has a uniform positive surface charge density σ. Find the electric field at a point along the axis of the disk a distance x from its center. Assume that x is positive.

SOLUTION

IDENTIFY and SET UP: **Figure 21.25** shows the situation. We represent the charge distribution as a collection of concentric rings of charge dQ. In Example 21.9 we obtained Eq. (21.8) for the field on the axis of a single uniformly charged ring, so all we need do here is integrate the contributions of our rings.

EXECUTE: A typical ring has charge dQ, inner radius r, and outer radius $r + dr$. Its area is approximately equal to its width dr times its circumference $2\pi r$, or $dA = 2\pi r\,dr$. The charge per unit area is $\sigma = dQ/dA$, so the charge of the ring is $dQ = \sigma\,dA = 2\pi\sigma r\,dr$. We use dQ in place of Q in Eq. (21.8), the expression for the field due to a ring that we found in Example 21.9, and replace the ring radius a with r. Then the field component dE_x at point P due to this ring is

21.25 Our sketch for this problem.



$$dE_x = \frac{1}{4\pi\epsilon_0}\frac{2\pi\sigma rx\,dr}{(x^2 + r^2)^{3/2}}$$

Continued

To find the total field due to all the rings, we integrate dE_x over r from $r = 0$ to $r = R$ (*not* from $-R$ to R):

$$E_x = \int_0^R \frac{1}{4\pi\epsilon_0} \frac{(2\pi\sigma r\, dr)x}{(x^2 + r^2)^{3/2}} = \frac{\sigma x}{4\epsilon_0} \int_0^R \frac{2r\, dr}{(x^2 + r^2)^{3/2}}$$

You can evaluate this integral by making the substitution $t = x^2 + r^2$ (which yields $dt = 2r\, dr$); you can work out the details. The result is

$$E_x = \frac{\sigma x}{2\epsilon_0}\left[-\frac{1}{\sqrt{x^2 + R^2}} + \frac{1}{x}\right]$$

$$= \frac{\sigma}{2\epsilon_0}\left[1 - \frac{1}{\sqrt{(R^2/x^2) + 1}}\right] \tag{21.11}$$

EVALUATE: If the disk is very large (or if we are very close to it), so that $R \gg x$, the term $1/\sqrt{(R^2/x^2) + 1}$ in Eq. (21.11) is very much less than 1. Then Eq. (21.11) becomes

$$E = \frac{\sigma}{2\epsilon_0} \tag{21.12}$$

Our final result does not contain the distance x from the plane. Hence the electric field produced by an *infinite* plane sheet of charge is *independent of the distance from the sheet.* The field direction is everywhere perpendicular to the sheet, away from it. There is no such thing as an infinite sheet of charge, but if the dimensions of the sheet are much larger than the distance x of the field point P from the sheet, the field is very nearly given by Eq. (21.12).

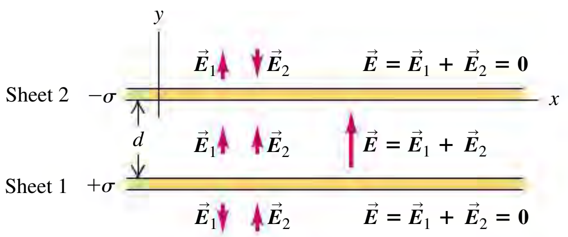
If P is to the *left* of the plane $(x < 0)$, the result is the same except that the direction of $\vec{E}$ is to the left instead of the right. If the surface charge density is negative, the directions of the fields on both sides of the plane are toward it rather than away from it.

EXAMPLE 21.12 FIELD OF TWO OPPOSITELY CHARGED INFINITE SHEETS

SOLUTION

Two infinite plane sheets with uniform surface charge densities $+\sigma$ and $-\sigma$ are placed parallel to each other with separation d (**Fig. 21.26**). Find the electric field between the sheets, above the upper sheet, and below the lower sheet.

21.26 Finding the electric field due to two oppositely charged infinite sheets. The sheets are seen edge-on; only a portion of the infinite sheets can be shown!



SOLUTION

IDENTIFY and SET UP: Equation (21.12) gives the electric field due to a single infinite plane sheet of charge. To find the field due to *two* such sheets, we combine the fields by using the principle of superposition (Fig. 21.26).

EXECUTE: From Eq. (21.12), both $\vec{E}_1$ and $\vec{E}_2$ have the same magnitude at all points, independent of distance from either sheet:

$$E_1 = E_2 = \frac{\sigma}{2\epsilon_0}$$

From Example 21.11, $\vec{E}_1$ is everywhere directed away from sheet 1, and $\vec{E}_2$ is everywhere directed toward sheet 2.

Between the sheets, $\vec{E}_1$ and $\vec{E}_2$ reinforce each other; above the upper sheet and below the lower sheet, they cancel each other. Thus the total field is

$$\vec{E} = \vec{E}_1 + \vec{E}_2 = \begin{cases} \mathbf{0} & \text{above the upper sheet} \\ \dfrac{\sigma}{\epsilon_0}\hat{\jmath} & \text{between the sheets} \\ \mathbf{0} & \text{below the lower sheet} \end{cases}$$

EVALUATE: Because we considered the sheets to be infinite, our result does not depend on the separation d. Our result shows that the field between oppositely charged plates is essentially uniform if the plate separation is much smaller than the dimensions of the plates. We actually used this result in Example 21.7 (Section 21.4).

CAUTION Electric fields are not "flows" You may have thought that the field $\vec{E}_1$ of sheet 1 would be unable to "penetrate" sheet 2, and that field $\vec{E}_2$ caused by sheet 2 would be unable to "penetrate" sheet 1. You might conclude this if you think of the electric field as some kind of physical substance that "flows" into or out of charges. But there is no such substance, and the electric fields $\vec{E}_1$ and $\vec{E}_2$ depend on only the individual charge distributions that create them. The *total* field at every point is just the vector sum of $\vec{E}_1$ and $\vec{E}_2$.

TEST YOUR UNDERSTANDING OF SECTION 21.5 Suppose that the line of charge in Fig. 21.24 (Example 21.10) had charge $+Q$ distributed uniformly between $y = 0$ and $y = +a$ and had charge $-Q$ distributed uniformly between $y = 0$ and $y = -a$. In this situation, the electric field at P would be (i) in the positive x-direction; (ii) in the negative x-direction; (iii) in the positive y-direction; (iv) in the negative y-direction; (v) zero; (vi) none of these.

21.6 ELECTRIC FIELD LINES

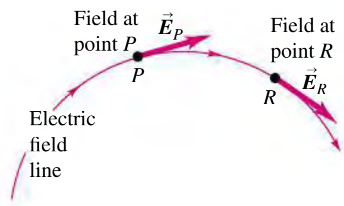
The concept of an electric field can be a little elusive because you can't see an electric field directly. Electric field *lines* can be a big help for visualizing electric fields and making them seem more real. An **electric field line** is an imaginary line or curve drawn through a region of space so that its tangent at any point is in the direction of the electric-field vector at that point. **Figure 21.27** shows the basic idea. (We used a similar concept in our discussion of fluid flow in Section 12.5. A *streamline* is a line or curve whose tangent at any point is in the direction of the velocity of the fluid at that point. However, the similarity between electric field lines and fluid streamlines is a mathematical one only; there is nothing "flowing" in an electric field.) The English scientist Michael Faraday (1791–1867) first introduced the concept of field lines. He called them "lines of force," but the term "field lines" is preferable.

Electric field lines show the direction of $\vec{E}$ at each point, and their spacing gives a general idea of the *magnitude* of $\vec{E}$ at each point. Where $\vec{E}$ is strong, we draw lines close together; where $\vec{E}$ is weaker, they are farther apart. At any particular point, the electric field has a unique direction, so only one field line can pass through each point of the field. In other words, *field lines never intersect.*

Figure 21.28 shows some of the electric field lines in a plane containing (a) a single positive charge; (b) two equal-magnitude charges, one positive and one negative (a dipole); and (c) two equal positive charges. Such diagrams are called *field maps;* they are cross sections of the actual three-dimensional patterns. The direction of the total electric field at every point in each diagram is along the tangent to the electric field line passing through the point. Arrowheads indicate the direction of the $\vec{E}$-field vector along each field line. The actual field vectors have been drawn at several points in each pattern. Notice that in general, the magnitude of the electric field is different at different points on a given field line; a field line is *not* a curve of constant electric-field magnitude!

Figure 21.28 shows that field lines are directed *away* from positive charges (since close to a positive point charge, $\vec{E}$ points away from the charge) and *toward* negative charges (since close to a negative point charge, $\vec{E}$ points toward the charge). In regions where the field magnitude is large, such as between the positive and negative charges in Fig. 21.28b, the field lines are drawn close together. In regions where the field magnitude is small, such as between the two positive charges in Fig. 21.28c, the lines are widely separated. In a *uniform* field, the field lines are straight, parallel, and uniformly spaced, as in Fig. 21.20.

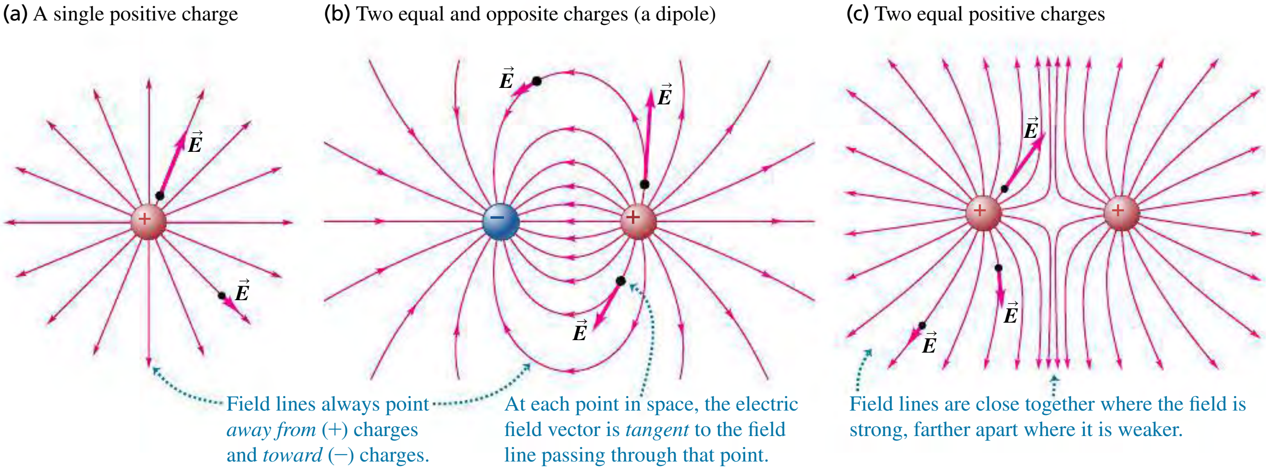
21.27 The direction of the electric field at any point is tangent to the field line through that point.



MP

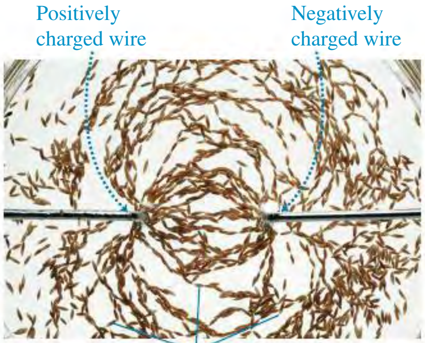
PhET: Charges and Fields
PhET: Electric Field of Dreams
PhET: Electric Field Hockey

21.28 Electric field lines for three different charge distributions. In general, the magnitude of $\vec{E}$ is different at different points along a given field line.



21.29 (a) Electric field lines produced by two opposite charges. The pattern is formed by grass seeds floating in mineral oil; the charges are on two wires whose tips are inserted into the oil. Compare this pattern with Fig. 21.28b. (b) The electric field causes polarization of the grass seeds, which in turn causes the seeds to align with the field.

(a)



(b)

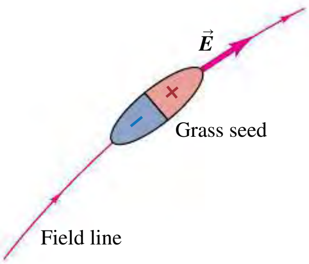


Figure 21.29a is a view from above of a demonstration setup for visualizing electric field lines. In the arrangement shown here, the tips of two positively charged wires are inserted in a container of insulating liquid, and some grass seeds are floated on the liquid. The grass seeds are electrically neutral insulators, but the electric field of the two charged wires causes *polarization* of the grass seeds; there is a slight shifting of the positive and negative charges within the molecules of each seed, like that shown in Fig. 21.8. The positively charged end of each grass seed is pulled in the direction of $\vec{E}$ and the negatively charged end is pulled opposite $\vec{E}$. Hence the long axis of each grass seed tends to orient parallel to the electric field, in the direction of the field line that passes through the position of the seed (Fig. 21.29b).

CAUTION **Electric field lines are not trajectories** It's a common misconception that if a particle of charge q is in motion where there is an electric field, the particle must move along an electric field line. Because $\vec{E}$ at any point is tangent to the field line that passes through that point, it is true that the *force* $\vec{F} = q\vec{E}$ on the particle, and hence the particle's acceleration, are tangent to the field line. But we learned in Chapter 3 that when a particle moves on a curved path, its acceleration *cannot* be tangent to the path. In general, the trajectory of a charged particle is *not* the same as a field line.

TEST YOUR UNDERSTANDING OF SECTION 21.6 Suppose the electric field lines in a region of space are straight lines. If a charged particle is released from rest in that region, will the trajectory of the particle be along a field line?

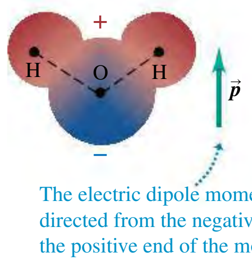
21.7 ELECTRIC DIPOLES

An **electric dipole** is a pair of point charges with equal magnitude and opposite sign (a positive charge q and a negative charge $-q$) separated by a distance d. We introduced electric dipoles in Example 21.8 (Section 21.5); the concept is worth exploring further because many physical systems, from molecules to TV antennas, can be described as electric dipoles. We will also use this concept extensively in our discussion of dielectrics in Chapter 24.

Figure 21.30a shows a molecule of water (H_2O), which in many ways behaves like an electric dipole. The water molecule as a whole is electrically neutral, but the chemical bonds within the molecule cause a displacement of charge; the result is a net negative charge on the oxygen end of the molecule and a net positive charge on the hydrogen end, forming an electric dipole. The effect is equivalent to shifting one electron only about 4×10^{-11} m (about the radius of a hydrogen atom), but the consequences of this shift are profound. Water is an excellent solvent for ionic substances such as table salt (sodium chloride, NaCl) precisely because the water molecule is an electric dipole (Fig. 21.30b).

21.30 (a) A water molecule is an example of an electric dipole. (b) Each test tube contains a solution of a different substance in water. The large electric dipole moment of water makes it an excellent solvent.

(a) A water molecule, showing positive charge as red and negative charge as blue



(b) Various substances dissolved in water

When dissolved in water, salt dissociates into a positive sodium ion (Na^+) and a negative chlorine ion (Cl^-), which tend to be attracted to the negative and positive ends, respectively, of water molecules; this holds the ions in solution. If water molecules were not electric dipoles, water would be a poor solvent, and almost all of the chemistry that occurs in aqueous solutions would be impossible. This includes all of the biochemical reactions that occur in all of the life on earth. In a very real sense, your existence as a living being depends on electric dipoles!

We examine two questions about electric dipoles. First, what forces and torques does an electric dipole experience when placed in an external electric field (that is, a field set up by charges outside the dipole)? Second, what electric field does an electric dipole itself produce?

Force and Torque on an Electric Dipole

To start with the first question, let's place an electric dipole in a *uniform* external electric field $\vec{E}$, as shown in **Fig. 21.31**. Both forces $\vec{F}_+$ and $\vec{F}_-$ on the two charges have magnitude qE, but their directions are opposite, and they add to zero. *The net force on an electric dipole in a uniform external electric field is zero.*

However, the two forces don't act along the same line, so their *torques* don't add to zero. We calculate torques with respect to the center of the dipole. Let the angle between the electric field $\vec{E}$ and the dipole axis be ϕ; then the lever arm for both $\vec{F}_+$ and $\vec{F}_-$ is $(d/2)\sin\phi$. The torque of $\vec{F}_+$ and the torque of $\vec{F}_-$ both have the same magnitude of $(qE)(d/2)\sin\phi$, and both torques tend to rotate the dipole clockwise (that is, $\vec{\tau}$ is directed into the page in Fig. 21.31). Hence the magnitude of the net torque is twice the magnitude of either individual torque:

$$\tau = (qE)(d\sin\phi) \tag{21.13}$$

where $d\sin\phi$ is the perpendicular distance between the lines of action of the two forces.

The product of the charge q and the separation d is the magnitude of a quantity called the **electric dipole moment,** denoted by p:

$$p = qd \qquad \text{(magnitude of electric dipole moment)} \tag{21.14}$$

The units of p are charge times distance $(C \cdot m)$. For example, the magnitude of the electric dipole moment of a water molecule is $p = 6.13 \times 10^{-30}\ C \cdot m$.

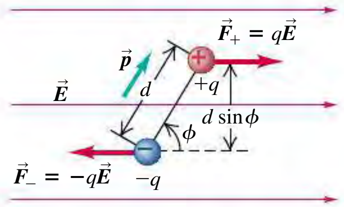
CAUTION **The symbol p has multiple meanings** Do not confuse dipole moment with momentum or pressure. There aren't as many letters in the alphabet as there are physical quantities, so some letters are used several times. The context usually makes it clear what we mean, but be careful.

We further define the electric dipole moment to be a *vector* quantity $\vec{p}$. The magnitude of $\vec{p}$ is given by Eq. (21.14), and its direction is along the dipole axis from the negative charge to the positive charge as shown in Fig. 21.31.

In terms of p, Eq. (21.13) for the magnitude τ of the torque exerted by the field becomes

$$\underbrace{\tau}_{\text{Magnitude of torque on an electric dipole}} = \underbrace{p}_{\text{Magnitude of electric dipole moment } \vec{p}}\ \underbrace{E}_{\text{Magnitude of electric field } \vec{E}}\ \sin\underbrace{\phi}_{\text{Angle between } \vec{p} \text{ and } \vec{E}} \tag{21.15}$$

Since the angle ϕ in Fig. 21.31 is the angle between the directions of the vectors $\vec{p}$ and $\vec{E}$, this is reminiscent of the expression for the magnitude of the *vector product*



21.31 The net force on this electric dipole is zero, but there is a torque directed into the page that tends to rotate the dipole clockwise.

discussed in Section 1.10. (You may want to review that discussion.) Hence we can write the torque on the dipole in vector form as

Vector torque on an electric dipole

$$\vec{\tau} = \vec{p} \times \vec{E} \qquad (21.16)$$

Electric dipole moment: $\vec{p}$; Electric field: $\vec{E}$

You can use the right-hand rule for the vector product to verify that in the situation shown in Fig. 21.31, $\vec{\tau}$ is directed into the page. The torque is greatest when $\vec{p}$ and $\vec{E}$ are perpendicular and is zero when they are parallel or antiparallel. The torque always tends to turn $\vec{p}$ to line it up with $\vec{E}$. The position $\phi = 0$, with $\vec{p}$ parallel to $\vec{E}$, is a position of stable equilibrium, and the position $\phi = \pi$, with $\vec{p}$ and $\vec{E}$ antiparallel, is a position of unstable equilibrium. The polarization of a grass seed in the apparatus of Fig. 21.29b gives it an electric dipole moment; the torque exerted by $\vec{E}$ then causes the seed to align with $\vec{E}$ and hence with the field lines.

MP

PhET: Microwaves

Potential Energy of an Electric Dipole

When a dipole changes direction in an electric field, the electric-field torque does *work* on it, with a corresponding change in potential energy. The work dW done by a torque τ during an infinitesimal displacement $d\phi$ is given by Eq. (10.19): $dW = \tau\, d\phi$. Because the torque is in the direction of decreasing ϕ, we must write the torque as $\tau = -pE\sin\phi$, and

$$dW = \tau\, d\phi = -pE\sin\phi\, d\phi$$

In a finite displacement from ϕ_1 to ϕ_2 the total work done on the dipole is

$$W = \int_{\phi_1}^{\phi_2} (-pE\sin\phi)\, d\phi$$

$$= pE\cos\phi_2 - pE\cos\phi_1$$

The work is the negative of the change of potential energy, just as in Chapter 7: $W = U_1 - U_2$. So a suitable definition of potential energy U for this system is

$$U(\phi) = -pE\cos\phi \qquad (21.17)$$

In this expression we recognize the *scalar product* $\vec{p} \cdot \vec{E} = pE\cos\phi$, so we can also write

Potential energy for an electric dipole in an electric field

$$U = -\vec{p} \cdot \vec{E} \qquad (21.18)$$

Electric field: $\vec{E}$; Electric dipole moment: $\vec{p}$

The potential energy has its minimum (most negative) value $U = -pE$ at the stable equilibrium position, where $\phi = 0$ and $\vec{p}$ is parallel to $\vec{E}$. The potential energy is maximum when $\phi = \pi$ and $\vec{p}$ is antiparallel to $\vec{E}$; then $U = +pE$. At $\phi = \pi/2$, where $\vec{p}$ is perpendicular to $\vec{E}$, U is zero. We could define U differently so that it is zero at some other orientation of $\vec{p}$, but our definition is simplest.

Equation (21.18) gives us another way to look at the effect illustrated in Fig. 21.29. The electric field $\vec{E}$ gives each grass seed an electric dipole moment, and the grass seed then aligns itself with $\vec{E}$ to minimize the potential energy.

SOLUTION

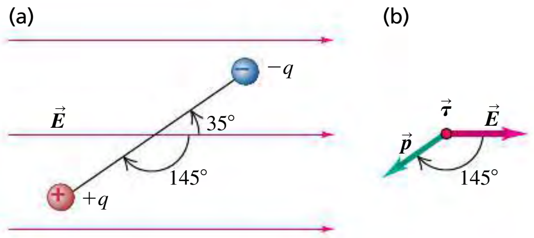
EXAMPLE 21.13 FORCE AND TORQUE ON AN ELECTRIC DIPOLE

Figure 21.32a shows an electric dipole in a uniform electric field of magnitude 5.0×10^5 N/C that is directed parallel to the plane of the figure. The charges are $\pm 1.6 \times 10^{-19}$ C; both lie in the plane and are separated by 0.125 nm $= 0.125 \times 10^{-9}$ m. Find (a) the net force exerted by the field on the dipole; (b) the magnitude and direction of the electric dipole moment; (c) the magnitude and direction of the torque; (d) the potential energy of the system in the position shown.

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of this section about an electric dipole placed in an electric field. We use the relationship $\vec{F} = q\vec{E}$ for each point charge to find the force on the dipole as a whole. Equation (21.14) gives the dipole moment, Eq. (21.16) gives the torque on the dipole, and Eq. (21.18) gives the potential energy of the system.

21.32 (a) An electric dipole. (b) Directions of the electric dipole moment, electric field, and torque ($\vec{\tau}$ points out of the page).



EXECUTE: (a) The field is uniform, so the forces on the two charges are equal and opposite. Hence the total force on the dipole is zero.

(b) The magnitude p of the electric dipole moment $\vec{p}$ is

$$p = qd = (1.6 \times 10^{-19}\ \text{C})(0.125 \times 10^{-9}\ \text{m})$$
$$= 2.0 \times 10^{-29}\ \text{C} \cdot \text{m}$$

The direction of $\vec{p}$ is from the negative to the positive charge, 145° clockwise from the electric-field direction (Fig. 21.32b).

(c) The magnitude of the torque is

$$\tau = pE\sin\phi = (2.0 \times 10^{-29}\ \text{C} \cdot \text{m})(5.0 \times 10^5\ \text{N/C})(\sin 145°)$$
$$= 5.7 \times 10^{-24}\ \text{N} \cdot \text{m}$$

From the right-hand rule for vector products (see Section 1.10), the direction of the torque $\vec{\tau} = \vec{p} \times \vec{E}$ is out of the page. This corresponds to a counterclockwise torque that tends to align $\vec{p}$ with $\vec{E}$.

(d) The potential energy

$$U = -pE\cos\phi$$
$$= -(2.0 \times 10^{-29}\ \text{C} \cdot \text{m})(5.0 \times 10^5\ \text{N/C})(\cos 145°)$$
$$= 8.2 \times 10^{-24}\ \text{J}$$

EVALUATE: The charge magnitude, the distance between the charges, the dipole moment, and the potential energy are all very small, but are all typical of molecules.

In this discussion we have assumed that $\vec{E}$ is uniform, so there is no net force on the dipole. If $\vec{E}$ is not uniform, the forces at the ends may not cancel completely, and the net force may not be zero. Thus a body with zero net charge but an electric dipole moment can experience a net force in a nonuniform electric field. As we mentioned in Section 21.1, an uncharged body can be polarized by an electric field, giving rise to a separation of charge and an electric dipole moment. This is how uncharged bodies can experience electrostatic forces (see Fig. 21.8).

BIO Application A Fish with an Electric Dipole Moment Unlike the tiger shark (Section 21.4), which senses the electric fields produced by its prey, the African knifefish *Gymnarchus niloticus*—which is nocturnal and has poor vision—hunts other fish by generating its own electric field. It can make its tail negatively charged relative to its head, producing a dipole-like field similar to that in Fig. 21.28b. When a smaller fish ventures into the field, its body alters the field pattern and alerts *G. niloticus* that a meal is present.

Field of an Electric Dipole

Now let's think of an electric dipole as a *source* of electric field. Figure 21.28b shows the general shape of the field due to a dipole. At each point in the pattern the total $\vec{E}$ field is the vector sum of the fields from the two individual charges, as in Example 21.8 (Section 21.5). Try drawing diagrams showing this vector sum for several points.

To get quantitative information about the field of an electric dipole, we have to do some calculating, as illustrated in the next example. Notice the use of the principle of superposition of electric fields to add up the contributions to the field of the individual charges. Also notice that we need to use approximation techniques even for the relatively simple case of a field due to two charges. Field calculations often become very complicated, and computer analysis is typically used to determine the field due to an arbitrary charge distribution.

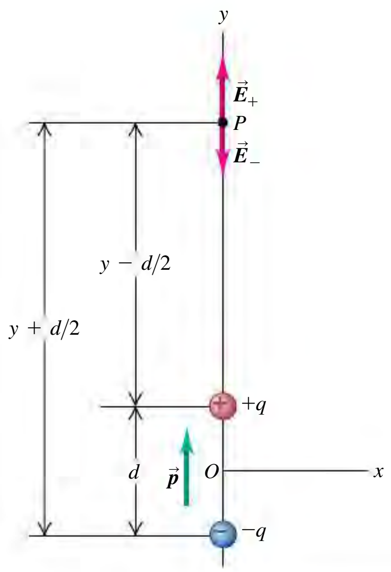
EXAMPLE 21.14 FIELD OF AN ELECTRIC DIPOLE, REVISITED

An electric dipole is centered at the origin, with $\vec{p}$ in the direction of the $+y$-axis (**Fig. 21.33**). Derive an approximate expression for the electric field at a point P on the y-axis for which y is much larger than d. To do this, use the binomial expansion $(1 + x)^n \cong 1 + nx + n(n - 1)x^2/2 + \cdots$ (valid for the case $|x| < 1$).

SOLUTION

IDENTIFY and SET UP: We use the principle of superposition: The total electric field is the vector sum of the field produced by the positive charge and the field produced by the negative charge. At the field point P shown in Fig. 21.33, the field $\vec{E}_+$ of the positive charge has a positive (upward) y-component and the field $\vec{E}_-$ of the negative charge has a negative (downward) y-component. We add these components to find the total field and then apply the approximation that y is much greater than d.

21.33 Finding the electric field of an electric dipole at a point on its axis.



EXECUTE: The total y-component E_y of electric field from the two charges is

$$E_y = \frac{q}{4\pi\epsilon_0}\left[\frac{1}{(y - d/2)^2} - \frac{1}{(y + d/2)^2}\right]$$

$$= \frac{q}{4\pi\epsilon_0 y^2}\left[\left(1 - \frac{d}{2y}\right)^{-2} - \left(1 + \frac{d}{2y}\right)^{-2}\right]$$

We used this same approach in Example 21.8 (Section 21.5). Now the approximation: When we are far from the dipole compared to its size, so $y \gg d$, we have $d/2y \ll 1$. With $n = -2$ and with $d/2y$ replacing x in the binomial expansion, we keep only the first two terms (the terms we discard are much smaller). We then have

$$\left(1 - \frac{d}{2y}\right)^{-2} \cong 1 + \frac{d}{y} \quad \text{and} \quad \left(1 + \frac{d}{2y}\right)^{-2} \cong 1 - \frac{d}{y}$$

Hence E_y is given approximately by

$$E_y \cong \frac{q}{4\pi\epsilon_0 y^2}\left[1 + \frac{d}{y} - \left(1 - \frac{d}{y}\right)\right] = \frac{qd}{2\pi\epsilon_0 y^3} = \frac{p}{2\pi\epsilon_0 y^3}$$

EVALUATE: An alternative route to this result is to put the fractions in the first expression for E_y over a common denominator, add, and then approximate the denominator $(y - d/2)^2(y + d/2)^2$ as y^4. We leave the details to you (see Exercise 21.58).

For points P off the coordinate axes, the expressions are more complicated, but at *all* points far away from the dipole (in any direction) the field drops off as $1/r^3$. We can compare this with the $1/r^2$ behavior of a point charge, the $1/r$ behavior of a long line charge, and the independence of r for a large sheet of charge. There are charge distributions for which the field drops off even more quickly. At large distances, the field of an *electric quadrupole,* which consists of two equal dipoles with opposite orientation, separated by a small distance, drops off as $1/r^4$.

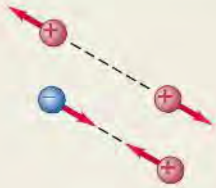
TEST YOUR UNDERSTANDING OF SECTION 21.7 An electric dipole is placed in a region of uniform electric field $\vec{E}$, with the electric dipole moment $\vec{p}$ pointing in the direction opposite to $\vec{E}$. Is the dipole (i) in stable equilibrium, (ii) in unstable equilibrium, or (iii) neither? (*Hint:* You many want to review Section 7.5.)

CHAPTER 21 SUMMARY

Electric charge, conductors, and insulators: The fundamental quantity in electrostatics is electric charge. There are two kinds of charge, positive and negative. Charges of the same sign repel each other; charges of opposite sign attract. Charge is conserved; the total charge in an isolated system is constant.

All ordinary matter is made of protons, neutrons, and electrons. The positive protons and electrically neutral neutrons in the nucleus of an atom are bound together by the nuclear force; the negative electrons surround the nucleus at distances much greater than the nuclear size. Electric interactions are chiefly responsible for the structure of atoms, molecules, and solids.

Conductors are materials in which charge moves easily; in insulators, charge does not move easily. Most metals are good conductors; most nonmetals are insulators.

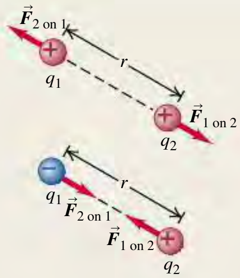


Coulomb's law: For charges q_1 and q_2 separated by a distance r, the magnitude of the electric force on either charge is proportional to the product $q_1 q_2$ and inversely proportional to r^2. The force on each charge is along the line joining the two charges—repulsive if q_1 and q_2 have the same sign, attractive if they have opposite signs. In SI units the unit of electric charge is the coulomb, abbreviated C. (See Examples 21.1 and 21.2.)

When two or more charges each exert a force on a charge, the total force on that charge is the vector sum of the forces exerted by the individual charges. (See Examples 21.3 and 21.4.)

$$F = \frac{1}{4\pi\epsilon_0}\frac{|q_1 q_2|}{r^2} \qquad (21.2)$$

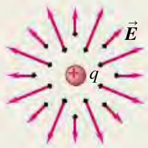
$$\frac{1}{4\pi\epsilon_0} = 8.988 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2$$



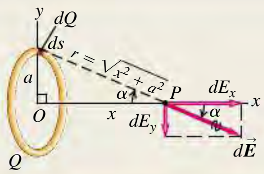
Electric field: Electric field $\vec{E}$, a vector quantity, is the force per unit charge exerted on a test charge at any point. The electric field produced by a point charge is directed radially away from or toward the charge. (See Examples 21.5–21.7.)

$$\vec{E} = \frac{\vec{F}_0}{q_0} \qquad (21.3)$$

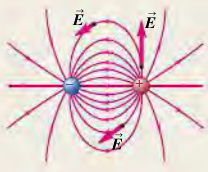
$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r} \qquad (21.7)$$



Superposition of electric fields: The electric field $\vec{E}$ of any combination of charges is the vector sum of the fields caused by the individual charges. To calculate the electric field caused by a continuous distribution of charge, divide the distribution into small elements, calculate the field caused by each element, and then carry out the vector sum, usually by integrating. Charge distributions are described by linear charge density λ, surface charge density σ, and volume charge density ρ. (See Examples 21.8–21.12.)



Electric field lines: Field lines provide a graphical representation of electric fields. At any point on a field line, the tangent to the line is in the direction of $\vec{E}$ at that point. The number of lines per unit area (perpendicular to their direction) is proportional to the magnitude of $\vec{E}$ at the point.

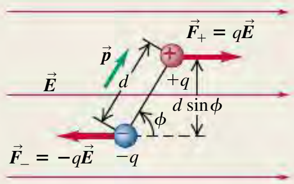


Electric dipoles: An electric dipole is a pair of electric charges of equal magnitude q but opposite sign, separated by a distance d. The electric dipole moment $\vec{p}$ has magnitude $p = qd$. The direction of $\vec{p}$ is from negative toward positive charge. An electric dipole in an electric field $\vec{E}$ experiences a torque $\vec{\tau}$ equal to the vector product of $\vec{p}$ and $\vec{E}$. The magnitude of the torque depends on the angle ϕ between $\vec{p}$ and $\vec{E}$. The potential energy U for an electric dipole in an electric field also depends on the relative orientation of $\vec{p}$ and $\vec{E}$. (See Examples 21.13 and 21.14.)

$$\tau = pE\sin\phi \qquad (21.15)$$

$$\vec{\tau} = \vec{p} \times \vec{E} \qquad (21.16)$$

$$U = -\vec{p}\cdot\vec{E} \qquad (21.18)$$



BRIDGING PROBLEM CALCULATING ELECTRIC FIELD: HALF A RING OF CHARGE

SOLUTION

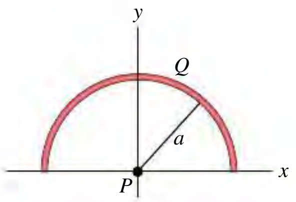
Positive charge Q is uniformly distributed around a semicircle of radius a as shown in **Fig. 21.34**. Find the magnitude and direction of the resulting electric field at point P, the center of curvature of the semicircle.

SOLUTION GUIDE

IDENTIFY and SET UP

1. The target variables are the components of the electric field at P.
2. Divide the semicircle into infinitesimal segments, each of which is a short circular arc of radius a and angle $d\theta$. What is the length of such a segment? How much charge is on a segment?

21.34 Charge uniformly distributed around a semicircle.



3. Consider an infinitesimal segment located at an angular position θ on the semicircle, measured from the lower right corner of the semicircle at $x = a$, $y = 0$. (Thus $\theta = \pi/2$ at $x = 0$, $y = a$ and $\theta = \pi$ at $x = -a$, $y = 0$.) What are the x- and y-components of the electric field at P (dE_x and dE_y) produced by just this segment?

EXECUTE

4. Integrate your expressions for dE_x and dE_y from $\theta = 0$ to $\theta = \pi$. The results will be the x-component and y-component of the electric field at P.
5. Use your results from step 4 to find the magnitude and direction of the field at P.

EVALUATE

6. Does your result for the electric-field magnitude have the correct units?
7. Explain how you could have found the x-component of the electric field by using a symmetry argument.
8. What would be the electric field at P if the semicircle were extended to a full circle centered at P?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q21.1 If you peel two strips of transparent tape off the same roll and immediately let them hang near each other, they will repel each other. If you then stick the sticky side of one to the shiny side of the other and rip them apart, they will attract each other. Give a plausible explanation, involving transfer of electrons between the strips of tape, for this sequence of events.

Q21.2 Two metal spheres are hanging from nylon threads. When you bring the spheres close to each other, they tend to attract. Based on this information alone, discuss all the possible ways that the spheres could be charged. Is it possible that after the spheres touch, they will cling together? Explain.

Q21.3 The electric force between two charged particles becomes weaker with increasing distance. Suppose instead that the electric force were *independent* of distance. In this case, would a charged comb still cause a neutral insulator to become polarized as in Fig. 21.8? Why or why not? Would the neutral insulator still be attracted to the comb? Again, why or why not?

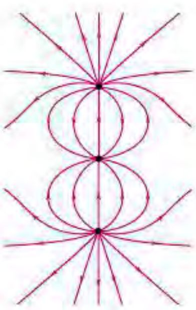
Q21.4 Your clothing tends to cling together after going through the dryer. Why? Would you expect more or less clinging if all your clothing were made of the same material (say, cotton) than if you dried different kinds of clothing together? Again, why? (You may want to experiment with your next load of laundry.)

Q21.5 An uncharged metal sphere hangs from a nylon thread. When a positively charged glass rod is brought close to the metal sphere, the sphere is drawn toward the rod. But if the sphere touches the rod, it suddenly flies away from the rod. Explain why the sphere is first attracted and then repelled.

Q21.6 **BIO** Estimate how many electrons there are in your body. Make any assumptions you feel are necessary, but clearly state what they are. (*Hint:* Most of the atoms in your body have equal numbers of electrons, protons, and neutrons.) What is the combined charge of all these electrons?

Q21.7 **Figure Q21.7** shows some of the electric field lines due to three point charges arranged along the vertical axis. All three charges have the same magnitude. (a) What are the signs of the three charges? Explain your reasoning. (b) At what point(s) is the magnitude of the electric field the smallest? Explain your reasoning. Explain how the fields produced by each individual point charge combine to give a small net field at this point or points.

Figure **Q21.7**



Q21.8 Good conductors of electricity, such as metals, are typically good conductors of heat; insulators, such as wood, are typically poor conductors of heat. Explain why there is a relationship between conduction of electricity and conduction of heat in these materials.

Q21.9 Suppose that the charge shown in Fig. 21.28a is fixed in position. A small, positively charged particle is then placed at some location and released. Will the trajectory of the particle follow an electric field line? Why or why not? Suppose instead that the particle is placed at some point in Fig. 21.28b and released (the positive and negative charges shown are fixed in position). Will its trajectory follow an electric field line? Again, why or why not? Explain any differences between your answers for the two situations.

Q21.10 Two identical metal objects are mounted on insulating stands. Describe how you could place charges of opposite sign but exactly equal magnitude on the two objects.

Q21.11 Because the charges on the electron and proton have the same absolute value, atoms are electrically neutral. Suppose that this is not precisely true, and the absolute value of the charge of the electron is less than the charge of the proton by 0.00100%. Estimate what the net charge of this textbook would be under these circumstances. Make any assumptions you feel are justified, but state clearly what they are. (*Hint:* Most of the atoms in this textbook have equal numbers of electrons, protons, and neutrons.) What would be the magnitude of the electric force between two textbooks placed 5.0 m apart? Would this force be attractive or repulsive? Discuss how the fact that ordinary matter is stable shows that the absolute values of the charges on the electron and proton must be identical to a *very* high level of accuracy.

Q21.12 If you walk across a nylon rug and then touch a large metal object such as a doorknob, you may get a spark and a shock. Why does this tend to happen more on dry days than on humid days? (*Hint:* See Fig. 21.30.) Why are you less likely to get a shock if you touch a *small* metal object, such as a paper clip?

Q21.13 You have a negatively charged object. How can you use it to place a net negative charge on an insulated metal sphere? To place a net positive charge on the sphere?

Q21.14 When two point charges of equal mass and charge are released on a frictionless table, each has an initial acceleration (magnitude) a_0. If instead you keep one fixed and release the other one, what will be its initial acceleration: a_0, $2a_0$, or $a_0/2$? Explain.

Q21.15 A point charge of mass m and charge Q and another point charge of mass m but charge $2Q$ are released on a frictionless table. If the charge Q has an initial acceleration a_0, what will be the acceleration of $2Q$: a_0, $2a_0$, $4a_0$, $a_0/2$, or $a_0/4$? Explain.

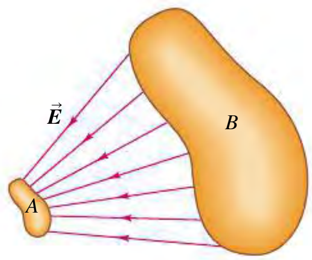
Q21.16 A proton is placed in a uniform electric field and then released. Then an electron is placed at this same point and released. Do these two particles experience the same force? The same acceleration? Do they move in the same direction when released?

Q21.17 In Example 21.1 (Section 21.3) we saw that the electric force between two α particles is of the order of 10^{35} times as strong as the gravitational force. So why do we readily feel the gravity of the earth but no electric force from it?

Q21.18 What similarities do electric forces have with gravitational forces? What are the most significant differences?

Q21.19 Two irregular objects A and B carry charges of opposite sign. **Figure Q21.19** shows the electric field lines near each of these objects. (a) Which object is positive, A or B? How do you know? (b) Where is the electric field stronger, close to A or close to B? How do you know?

Figure **Q21.19**

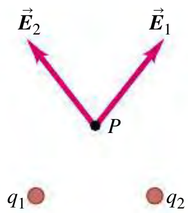


Q21.20 Atomic nuclei are made of protons and neutrons. This shows that there must be another kind of interaction in addition to gravitational and electric forces. Explain.

Q21.21 Sufficiently strong electric fields can cause atoms to become positively ionized—that is, to lose one or more electrons. Explain how this can happen. What determines how strong the field must be to make this happen?

Q21.22 The electric fields at point P due to the positive charges q_1 and q_2 are shown in **Fig. Q21.22**. Does the fact that they cross each other violate the statement in Section 21.6 that electric field lines never cross? Explain.

Figure **Q21.22**



Q21.23 The air temperature and the velocity of the air have different values at different places in the earth's atmosphere. Is the air velocity a vector field? Why or why not? Is the air temperature a vector field? Again, why or why not?

EXERCISES

Section 21.3 Coulomb's Law

21.1 •• Excess electrons are placed on a small lead sphere with mass 8.00 g so that its net charge is -3.20×10^{-9} C. (a) Find the number of excess electrons on the sphere. (b) How many excess electrons are there per lead atom? The atomic number of lead is 82, and its atomic mass is 207 g/mol.

21.2 • Lightning occurs when there is a flow of electric charge (principally electrons) between the ground and a thundercloud. The maximum rate of charge flow in a lightning bolt is about 20,000 C/s; this lasts for 100 μs or less. How much charge flows between the ground and the cloud in this time? How many electrons flow during this time?

21.3 •• If a proton and an electron are released when they are 2.0×10^{-10} m apart (a typical atomic distance), find the initial acceleration of each particle.

21.4 • **Particles in a Gold Ring.** You have a pure (24-karat) gold ring of mass 10.8 g. Gold has an atomic mass of 197 g/mol and an atomic number of 79. (a) How many protons are in the ring, and what is their total positive charge? (b) If the ring carries no net charge, how many electrons are in it?

21.5 • BIO **Signal Propagation in Neurons.** *Neurons* are components of the nervous system of the body that transmit signals as electric impulses travel along their length. These impulses propagate when charge suddenly rushes into and then out of a part of the neuron called an *axon*. Measurements have shown that, during the inflow part of this cycle, approximately 5.6×10^{11} Na^+ (sodium ions) per meter, each with charge $+e$, enter the axon. How many coulombs of charge enter a 1.5-cm length of the axon during this process?

21.6 • Two small spheres spaced 20.0 cm apart have equal charge. How many excess electrons must be present on each sphere if the magnitude of the force of repulsion between them is 3.33×10^{-21} N?

21.7 •• An average human weighs about 650 N. If each of two average humans could carry 1.0 C of excess charge, one positive and one negative, how far apart would they have to be for the electric attraction between them to equal their 650-N weight?

21.8 •• Two small aluminum spheres, each having mass 0.0250 kg, are separated by 80.0 cm. (a) How many electrons does each sphere contain? (The atomic mass of aluminum is 26.982 g/mol, and its atomic number is 13.) (b) How many electrons would have to be removed from one sphere and added to the other to cause an attractive force between the spheres of magnitude 1.00×10^4 N (roughly 1 ton)? Assume that the spheres may be treated as point charges. (c) What fraction of all the electrons in each sphere does this represent?

21.9 •• Two small plastic spheres are given positive electric charges. When they are 15.0 cm apart, the repulsive force between them has magnitude 0.220 N. What is the charge on each sphere (a) if the two charges are equal and (b) if one sphere has four times the charge of the other?

21.10 •• **Just How Strong Is the Electric Force?** Suppose you had two small boxes, each containing 1.0 g of protons. (a) If one were placed on the moon by an astronaut and the other were left on the earth, and if they were connected by a very light (and very long!) string, what would be the tension in the string? Express your answer in newtons and in pounds. Do you need to take into account the gravitational forces of the earth and moon on the protons? Why? (b) What gravitational force would each box of protons exert on the other box?

21.11 • In an experiment in space, one proton is held fixed and another proton is released from rest a distance of 2.50 mm away. (a) What is the initial acceleration of the proton after it is released? (b) Sketch qualitative (no numbers!) acceleration–time and velocity–time graphs of the released proton's motion.

21.12 • A negative charge of $-0.550\ \mu$C exerts an upward 0.600-N force on an unknown charge that is located 0.300 m directly below the first charge. What are (a) the value of the unknown charge (magnitude and sign); (b) the magnitude and direction of the force that the unknown charge exerts on the -0.550-μC charge?

21.13 •• Three point charges are arranged on a line. Charge $q_3 = +5.00$ nC and is at the origin. Charge $q_2 = -3.00$ nC and is at $x = +4.00$ cm. Charge q_1 is at $x = +2.00$ cm. What is q_1 (magnitude and sign) if the net force on q_3 is zero?

21.14 •• In Example 21.4, suppose the point charge on the y-axis at $y = -0.30$ m has negative charge $-2.0\ \mu$C, and the other charges remain the same. Find the magnitude and direction of the net force on Q. How does your answer differ from that in Example 21.4? Explain the differences.

21.15 •• In Example 21.3, calculate the net force on charge q_1.

21.16 •• In Example 21.4, what is the net force (magnitude and direction) on charge q_1 exerted by the other two charges?

21.17 •• Three point charges are arranged along the x-axis. Charge $q_1 = +3.00\ \mu$C is at the origin, and charge $q_2 = -5.00\ \mu$C is at $x = 0.200$ m. Charge $q_3 = -8.00\ \mu$C. Where is q_3 located if the net force on q_1 is 7.00 N in the $-x$-direction?

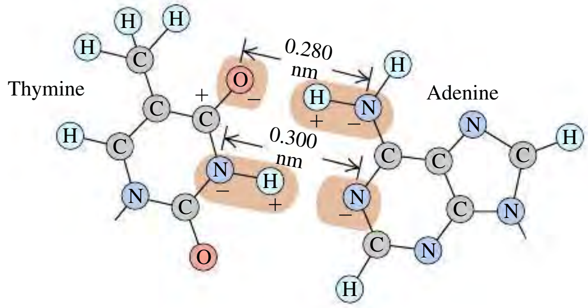
21.18 •• Repeat Exercise 21.17 for $q_3 = +8.00\ \mu$C.

21.19 •• Two point charges are located on the y-axis as follows: charge $q_1 = -1.50$ nC at $y = -0.600$ m, and charge $q_2 = +3.20$ nC at the origin ($y = 0$). What is the total force (magnitude and direction) exerted by these two charges on a third charge $q_3 = +5.00$ nC located at $y = -0.400$ m?

21.20 •• Two point charges are placed on the x-axis as follows: Charge $q_1 = +4.00$ nC is located at $x = 0.200$ m, and charge $q_2 = +5.00$ nC is at $x = -0.300$ m. What are the magnitude and direction of the total force exerted by these two charges on a negative point charge $q_3 = -6.00$ nC that is placed at the origin?

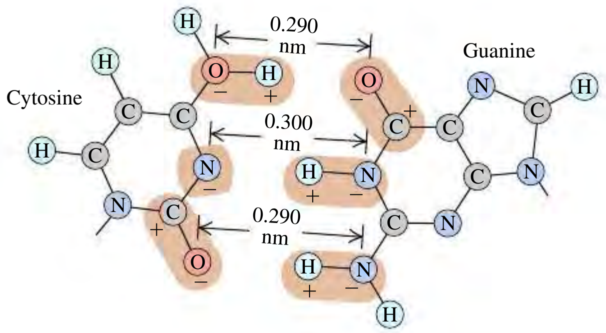
21.21 •• BIO **Base Pairing in DNA, I.** The two sides of the DNA double helix are connected by pairs of bases (adenine, thymine, cytosine, and guanine). Because of the geometric shape of these molecules, adenine bonds with thymine and cytosine bonds with guanine. **Figure E21.21** shows the bonding of thymine and adenine. Each charge shown is $\pm e$, and the H—N distance is 0.110 nm. (a) Calculate the *net* force that thymine exerts on adenine. Is it attractive or repulsive? To keep the calculations fairly simple, yet reasonable, consider only the forces due to the O—H—N and the N—H—N combinations, assuming that these two combinations are parallel to each other. Remember, however, that in the O—H—N set, the O^- exerts a force on both the H^+ and the N^-, and likewise along the N—H—N set. (b) Calculate the force on the electron in the hydrogen atom, which is 0.0529 nm from the proton. Then compare the strength of the bonding force of the electron in hydrogen with the bonding force of the adenine–thymine molecules.

Figure **E21.21**



21.22 •• BIO **Base Pairing in DNA, II.** Refer to Exercise 21.21. **Figure E21.22** shows the bonding of cytosine and guanine. The O—H and H—N distances are each 0.110 nm. In this case, assume that the bonding is due only to the forces along the O—H—O, N—H—N, and O—H—N combinations, and assume also that these three combinations are parallel to each other. Calculate the *net* force that cytosine exerts on guanine due to the preceding three combinations. Is this force attractive or repulsive?

Figure **E21.22**



Section 21.4 Electric Field and Electric Forces

21.23 • CP A proton is placed in a uniform electric field of 2.75×10^3 N/C. Calculate (a) the magnitude of the electric force felt by the proton; (b) the proton's acceleration; (c) the proton's speed after 1.00 μs in the field, assuming it starts from rest.

21.24 • A particle has charge -5.00 nC. (a) Find the magnitude and direction of the electric field due to this particle at a point 0.250 m directly above it. (b) At what distance from this particle does its electric field have a magnitude of 12.0 N/C?

21.25 • **CP** A proton is traveling horizontally to the right at $4.50 \times 10^6\text{ m/s}$. (a) Find the magnitude and direction of the weakest electric field that can bring the proton uniformly to rest over a distance of 3.20 cm. (b) How much time does it take the proton to stop after entering the field? (c) What minimum field (magnitude and direction) would be needed to stop an electron under the conditions of part (a)?

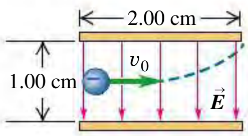
21.26 • **CP** An electron is released from rest in a uniform electric field. The electron accelerates vertically upward, traveling 4.50 m in the first 3.00 μs after it is released. (a) What are the magnitude and direction of the electric field? (b) Are we justified in ignoring the effects of gravity? Justify your answer quantitatively.

21.27 •• (a) What must the charge (sign and magnitude) of a 1.45-g particle be for it to remain stationary when placed in a downward-directed electric field of magnitude 650 N/C? (b) What is the magnitude of an electric field in which the electric force on a proton is equal in magnitude to its weight?

21.28 •• **Electric Field of the Earth.** The earth has a net electric charge that causes a field at points near its surface equal to 150 N/C and directed in toward the center of the earth. (a) What magnitude and sign of charge would a 60-kg human have to acquire to overcome his or her weight by the force exerted by the earth's electric field? (b) What would be the force of repulsion between two people each with the charge calculated in part (a) and separated by a distance of 100 m? Is use of the earth's electric field a feasible means of flight? Why or why not?

21.29 •• **CP** An electron is projected with an initial speed $v_0 = 1.60 \times 10^6\text{ m/s}$ into the uniform field between two parallel plates (**Fig. E21.29**). Assume that the field between the plates is uniform and directed vertically downward and that the field outside the plates is zero. The electron enters the field at a point midway between the plates. (a) If the electron just misses the upper plate as it emerges from the field, find the magnitude of the electric field. (b) Suppose that the electron in Fig. E21.29 is replaced by a proton with the same initial speed v_0. Would the proton hit one of the plates? If not, what would be the magnitude and direction of its vertical displacement as it exits the region between the plates? (c) Compare the paths traveled by the electron and the proton, and explain the differences. (d) Discuss whether it is reasonable to ignore the effects of gravity for each particle.

Figure **E21.29**



21.30 • (a) Calculate the magnitude and direction (relative to the $+x$-axis) of the electric field in Example 21.6. (b) A -2.5-nC point charge is placed at point P in Fig. 21.19. Find the magnitude and direction of (i) the force that the -8.0-nC charge at the origin exerts on this charge and (ii) the force that this charge exerts on the -8.0-nC charge at the origin.

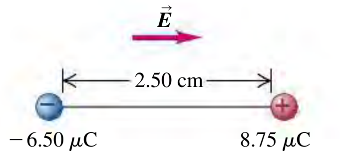
21.31 •• **CP** In Exercise 21.29, what is the speed of the electron as it emerges from the field?

21.32 •• **CP** A uniform electric field exists in the region between two oppositely charged plane parallel plates. A proton is released from rest at the surface of the positively charged plate and strikes the surface of the opposite plate, 1.60 cm distant from the first, in a time interval of $3.20 \times 10^{-6}\text{ s}$. (a) Find the magnitude of the electric field. (b) Find the speed of the proton when it strikes the negatively charged plate.

21.33 • A point charge is at the origin. With this point charge as the source point, what is the unit vector $\hat{\boldsymbol{r}}$ in the direction of the field point (a) at $x = 0$, $y = -1.35\text{ m}$; (b) at $x = 12.0\text{ cm}$, $y = 12.0\text{ cm}$; (c) at $x = -1.10\text{ m}$, $y = 2.60\text{ m}$? Express your results in terms of the unit vectors $\hat{\imath}$ and $\hat{\jmath}$.

21.34 •• A $+8.75\text{-}\mu\text{C}$ point charge is glued down on a horizontal frictionless table. It is tied to a $-6.50\text{-}\mu\text{C}$ point charge by a light, nonconducting 2.50-cm wire. A uniform electric field of magnitude $1.85 \times 10^8\text{ N/C}$ is directed parallel to the wire, as shown in **Fig. E21.34**. (a) Find the tension in the wire. (b) What would the tension be if both charges were negative?

Figure **E21.34**

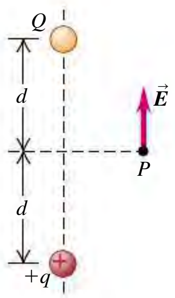


21.35 •• (a) An electron is moving east in a uniform electric field of 1.50 N/C directed to the west. At point A, the velocity of the electron is $4.50 \times 10^5\text{ m/s}$ toward the east. What is the speed of the electron when it reaches point B, 0.375 m east of point A? (b) A proton is moving in the uniform electric field of part (a). At point A, the velocity of the proton is $1.90 \times 10^4\text{ m/s}$, east. What is the speed of the proton at point B?

Section 21.5 Electric-Field Calculations

21.36 • Two point charges Q and $+q$ (where q is positive) produce the net electric field shown at point P in **Fig. E21.36**. The field points parallel to the line connecting the two charges. (a) What can you conclude about the sign and magnitude of Q? Explain your reasoning. (b) If the lower charge were negative instead, would it be possible for the field to have the direction shown in the figure? Explain your reasoning.

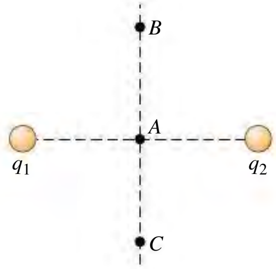
Figure **E21.36**



21.37 •• Two positive point charges q are placed on the x-axis, one at $x = a$ and one at $x = -a$. (a) Find the magnitude and direction of the electric field at $x = 0$. (b) Derive an expression for the electric field at points on the x-axis. Use your result to graph the x-component of the electric field as a function of x, for values of x between $-4a$ and $+4a$.

21.38 • The two charges q_1 and q_2 shown in **Fig. E21.38** have equal magnitudes. What is the direction of the net electric field due to these two charges at points A (midway between the charges), B, and C if (a) both charges are negative, (b) both charges are positive, (c) q_1 is positive and q_2 is negative.

Figure **E21.38**

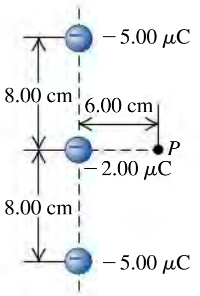


21.39 • A $+2.00\text{-nC}$ point charge is at the origin, and a second -5.00-nC point charge is on the x-axis at $x = 0.800\text{ m}$. (a) Find the electric field (magnitude and direction) at each of the following points on the x-axis: (i) $x = 0.200\text{ m}$; (ii) $x = 1.20\text{ m}$; (iii) $x = -0.200\text{ m}$. (b) Find the net electric force that the two charges would exert on an electron placed at each point in part (a).

21.40 • Repeat Exercise 21.39, but now let the charge at the origin be −4.00 nC.

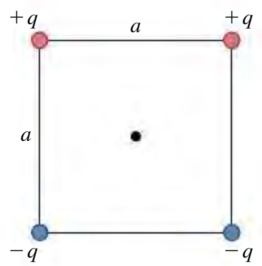
21.41 • Three negative point charges lie along a line as shown in **Fig. E21.41**. Find the magnitude and direction of the electric field this combination of charges produces at point P, which lies 6.00 cm from the −2.00-μC charge measured perpendicular to the line connecting the three charges.

Figure **E21.41**



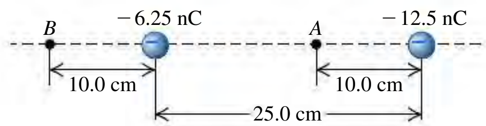
21.42 •• A point charge is placed at each corner of a square with side length a. All charges have magnitude q. Two of the charges are positive and two are negative (**Fig. E21.42**). What is the direction of the net electric field at the center of the square due to the four charges, and what is its magnitude in terms of q and a?

Figure **E21.42**



21.43 • Two point charges are separated by 25.0 cm (**Fig. E21.43**). Find the net electric field these charges produce at (a) point A and (b) point B. (c) What would be the magnitude and direction of the electric force this combination of charges would produce on a proton at A?

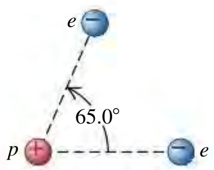
Figure **E21.43**



21.44 •• Point charge $q_1 = -5.00$ nC is at the origin and point charge $q_2 = +3.00$ nC is on the x-axis at $x = 3.00$ cm. Point P is on the y-axis at $y = 4.00$ cm. (a) Calculate the electric fields $\vec{E}_1$ and $\vec{E}_2$ at point P due to the charges q_1 and q_2. Express your results in terms of unit vectors (see Example 21.6). (b) Use the results of part (a) to obtain the resultant field at P, expressed in unit vector form.

21.45 •• If two electrons are each 1.50×10^{-10} m from a proton (**Fig. E21.45**), find the magnitude and direction of the net electric force they will exert on the proton.

Figure **E21.45**



21.46 •• **BIO** **Electric Field of Axons.** A nerve signal is transmitted through a neuron when an excess of Na^+ ions suddenly enters the axon, a long cylindrical part of the neuron. Axons are approximately 10.0 μm in diameter, and measurements show that about 5.6×10^{11} Na^+ ions per meter (each of charge $+e$) enter during this process. Although the axon is a long cylinder, the charge does not all enter everywhere at the same time. A plausible model would be a series of point charges moving along the axon. Consider a 0.10-mm length of the axon and model it as a point charge. (a) If the charge that enters each meter of the axon gets distributed uniformly along it, how many coulombs of charge enter a 0.10-mm length of the axon? (b) What electric field (magnitude and direction) does the sudden influx of charge produce at the surface of the body if the axon is 5.00 cm below the skin? (c) Certain sharks can respond to electric fields as weak as 1.0 μN/C. How far from this segment of axon could a shark be and still detect its electric field?

21.47 • In a rectangular coordinate system a positive point charge $q = 6.00 \times 10^{-9}$ C is placed at the point $x = +0.150$ m, $y = 0$, and an identical point charge is placed at $x = -0.150$ m, $y = 0$. Find the x- and y-components, the magnitude, and the direction of the electric field at the following points: (a) the origin; (b) $x = 0.300$ m, $y = 0$; (c) $x = 0.150$ m, $y = -0.400$ m; (d) $x = 0$, $y = 0.200$ m.

21.48 •• A point charge $q_1 = -4.00$ nC is at the point $x = 0.600$ m, $y = 0.800$ m, and a second point charge $q_2 = +6.00$ nC is at the point $x = 0.600$ m, $y = 0$. Calculate the magnitude and direction of the net electric field at the origin due to these two point charges.

21.49 •• A charge of −6.50 nC is spread uniformly over the surface of one face of a nonconducting disk of radius 1.25 cm. (a) Find the magnitude and direction of the electric field this disk produces at a point P on the axis of the disk a distance of 2.00 cm from its center. (b) Suppose that the charge were all pushed away from the center and distributed uniformly on the outer rim of the disk. Find the magnitude and direction of the electric field at point P. (c) If the charge is all brought to the center of the disk, find the magnitude and direction of the electric field at point P. (d) Why is the field in part (a) stronger than the field in part (b)? Why is the field in part (c) the strongest of the three fields?

21.50 •• A very long, straight wire has charge per unit length 3.20×10^{-10} C/m. At what distance from the wire is the electric-field magnitude equal to 2.50 N/C?

21.51 • A ring-shaped conductor with radius $a = 2.50$ cm has a total positive charge $Q = +0.125$ nC uniformly distributed around it (see Fig. 21.23). The center of the ring is at the origin of coordinates O. (a) What is the electric field (magnitude and direction) at point P, which is on the x-axis at $x = 40.0$ cm? (b) A point charge $q = -2.50$ μC is placed at P. What are the magnitude and direction of the force exerted *by* the charge q *on* the ring?

21.52 •• A straight, nonconducting plastic wire 8.50 cm long carries a charge density of +175 nC/m distributed uniformly along its length. It is lying on a horizontal tabletop. (a) Find the magnitude and direction of the electric field this wire produces at a point 6.00 cm directly above its midpoint. (b) If the wire is now bent into a circle lying flat on the table, find the magnitude and direction of the electric field it produces at a point 6.00 cm directly above its center.

Section 21.7 Electric Dipoles

21.53 • Point charges $q_1 = -4.5$ nC and $q_2 = +4.5$ nC are separated by 3.1 mm, forming an electric dipole. (a) Find the electric dipole moment (magnitude and direction). (b) The charges are in a uniform electric field whose direction makes an angle of 36.9° with the line connecting the charges. What is the magnitude of this field if the torque exerted on the dipole has magnitude 7.2×10^{-9} N·m?

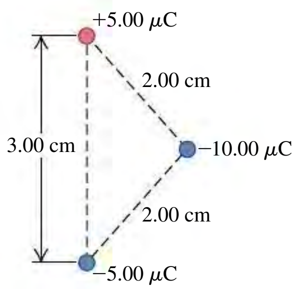
21.54 • The ammonia molecule (NH_3) has a dipole moment of 5.0×10^{-30} C·m. Ammonia molecules in the gas phase are placed in a uniform electric field $\vec{E}$ with magnitude 1.6×10^6 N/C. (a) What is the change in electric potential energy when the dipole moment of a molecule changes its orientation with respect to $\vec{E}$ from parallel to perpendicular? (b) At what absolute temperature T is the average translational kinetic energy $\frac{3}{2}kT$ of a molecule equal to the change in potential energy calculated in part (a)? (*Note:* Above this temperature, thermal agitation prevents the dipoles from aligning with the electric field.)

21.55 • **Torque on a Dipole.** An electric dipole with dipole moment $\vec{p}$ is in a uniform external electric field $\vec{E}$. (a) Find the orientations of the dipole for which the torque on the dipole is zero. (b) Which of the orientations in part (a) is stable, and which is unstable? (*Hint:* Consider a small rotation away from the equilibrium position and see what happens.) (c) Show that for the stable orientation in part (b), the dipole's own electric field tends to oppose the external field.

21.56 • The dipole moment of the water molecule (H_2O) is 6.17×10^{-30} C·m. Consider a water molecule located at the origin whose dipole moment $\vec{p}$ points in the $+x$-direction. A chlorine ion (Cl^-), of charge -1.60×10^{-19} C, is located at $x = 3.00 \times 10^{-9}$ m. Find the magnitude and direction of the electric force that the water molecule exerts on the chlorine ion. Is this force attractive or repulsive? Assume that x is much larger than the separation d between the charges in the dipole, so that the approximate expression for the electric field along the dipole axis derived in Example 21.14 can be used.

21.57 • Three charges are at the corners of an isosceles triangle as shown in **Fig. E21.57**. The ± 5.00-μC charges form a dipole. (a) Find the force (magnitude and direction) the -10.00-μC charge exerts on the dipole. (b) For an axis perpendicular to the line connecting the ± 5.00-μC charges at the midpoint of this line, find the torque (magnitude and direction) exerted on the dipole by the -10.00-μC charge.

Figure **E21.57**



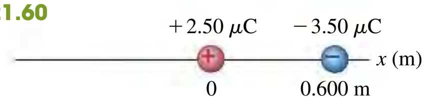
21.58 •• Consider the electric dipole of Example 21.14. (a) Derive an expression for the magnitude of the electric field produced by the dipole at a point on the x-axis in Fig. 21.33. What is the direction of this electric field? (b) How does the electric field at points on the x-axis depend on x when x is very large?

PROBLEMS

21.59 ••• Four identical charges Q are placed at the corners of a square of side L. (a) In a free-body diagram, show all of the forces that act on one of the charges. (b) Find the magnitude and direction of the total force exerted on one charge by the other three charges.

21.60 ••• Two charges are placed on the x-axis: one, of 2.50 μC, at the origin and the other, of -3.50 μC, at $x = 0.600$ m (**Fig. P21.60**). Find the position on the x-axis where the net force on a small charge $+q$ would be zero.

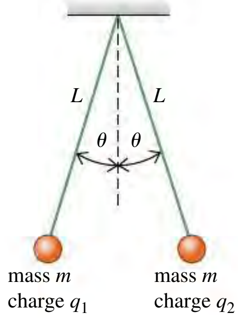
Figure **P21.60**



21.61 •• A charge $q_1 = +5.00$ nC is placed at the origin of an xy-coordinate system, and a charge $q_2 = -2.00$ nC is placed on the positive x-axis at $x = 4.00$ cm. (a) If a third charge $q_3 = +6.00$ nC is now placed at the point $x = 4.00$ cm, $y = 3.00$ cm, find the x- and y-components of the total force exerted on this charge by the other two. (b) Find the magnitude and direction of this force.

21.62 •• **CP** Two identical spheres with mass m are hung from silk threads of length L (**Fig. P21.62**). The spheres have the same charge, so $q_1 = q_2 = q$. The radius of each sphere is very small compared to the distance between the spheres, so they may be treated as point charges. Show that if the angle θ is small, the equilibrium separation d between the spheres is $d = (q^2L/2\pi\epsilon_0 mg)^{1/3}$. (*Hint:* If θ is small, then $\tan\theta \cong \sin\theta$.)

Figure **P21.62**

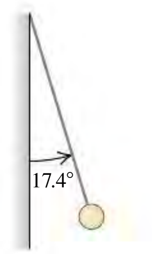


21.63 ••• **CP** Two small spheres with mass $m = 15.0$ g are hung by silk threads of length $L = 1.20$ m from a common point (Fig. P21.62). When the spheres are given equal quantities of negative charge, so that $q_1 = q_2 = q$, each thread hangs at $\theta = 25.0°$ from the vertical. (a) Draw a diagram showing the forces on each sphere. Treat the spheres as point charges. (b) Find the magnitude of q. (c) Both threads are now shortened to length $L = 0.600$ m, while the charges q_1 and q_2 remain unchanged. What new angle will each thread make with the vertical? (*Hint:* This part of the problem can be solved numerically by using trial values for θ and adjusting the values of θ until a self-consistent answer is obtained.)

21.64 •• **CP** Two identical spheres are each attached to silk threads of length $L = 0.500$ m and hung from a common point (Fig. P21.62). Each sphere has mass $m = 8.00$ g. The radius of each sphere is very small compared to the distance between the spheres, so they may be treated as point charges. One sphere is given positive charge q_1, and the other a different positive charge q_2; this causes the spheres to separate so that when the spheres are in equilibrium, each thread makes an angle $\theta = 20.0°$ with the vertical. (a) Draw a free-body diagram for each sphere when in equilibrium, and label all the forces that act on each sphere. (b) Determine the magnitude of the electrostatic force that acts on each sphere, and determine the tension in each thread. (c) Based on the given information, what can you say about the magnitudes of q_1 and q_2? Explain. (d) A small wire is now connected between the spheres, allowing charge to be transferred from one sphere to the other until the two spheres have equal charges; the wire is then removed. Each thread now makes an angle of 30.0° with the vertical. Determine the original charges. (*Hint:* The total charge on the pair of spheres is conserved.)

21.65 •• **CP** A small 12.3-g plastic ball is tied to a very light 28.6-cm string that is attached to the vertical wall of a room (**Fig. P21.65**). A uniform horizontal electric field exists in this room. When the ball has been given an excess charge of -1.11 μC, you observe that it remains suspended, with the string making an angle of 17.4° with the wall. Find the magnitude and direction of the electric field in the room.

Figure **P21.65**



21.66 •• Point charge $q_1 = -6.00 \times 10^{-6}$ C is on the x-axis at $x = -0.200$ m. Point charge q_2 is on the x-axis at $x = +0.400$ m. Point charge $q_3 = +3.00 \times 10^{-6}$ C is at the origin. What is q_2 (magnitude and sign) (a) if the net force on q_3 is 6.00 N in the $+x$-direction; (b) if the net force on q_3 is 6.00 N in the $-x$-direction?

21.67 •• Two particles having charges $q_1 = 0.500$ nC and $q_2 = 8.00$ nC are separated by a distance of 1.20 m. At what point along the line connecting the two charges is the total electric field due to the two charges equal to zero?

21.68 •• A -3.00-nC point charge is on the x-axis at $x = 1.20$ m. A second point charge, Q, is on the x-axis at -0.600 m. What must be the sign and magnitude of Q for the resultant electric field at the origin to be (a) 45.0 N/C in the $+x$-direction, (b) 45.0 N/C in the $-x$-direction?

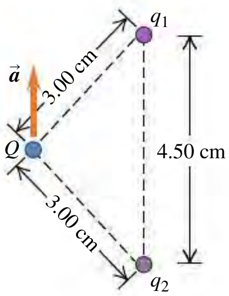
21.69 ••• A charge $+Q$ is located at the origin, and a charge $+4Q$ is at distance d away on the x-axis. Where should a third charge, q, be placed, and what should be its sign and magnitude, so that all three charges will be in equilibrium?

21.70 •• A charge of -3.00 nC is placed at the origin of an xy-coordinate system, and a charge of 2.00 nC is placed on the y-axis at $y = 4.00$ cm. (a) If a third charge, of 5.00 nC, is now placed at the point $x = 3.00$ cm, $y = 4.00$ cm, find the x- and y-components of the total force exerted on this charge by the other two charges. (b) Find the magnitude and direction of this force.

21.71 • Three identical point charges q are placed at each of three corners of a square of side L. Find the magnitude and direction of the net force on a point charge $-3q$ placed (a) at the center of the square and (b) at the vacant corner of the square. In each case, draw a free-body diagram showing the forces exerted on the $-3q$ charge by each of the other three charges.

21.72 ••• Two point charges q_1 and q_2 are held in place 4.50 cm apart. Another point charge $Q = -1.75\ \mu$C, of mass 5.00 g, is initially located 3.00 cm from both of these charges (**Fig. P21.72**) and released from rest. You observe that the initial acceleration of Q is 324 m/s^2 upward, parallel to the line connecting the two point charges. Find q_1 and q_2.

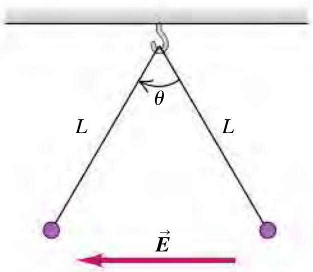
Figure **P21.72**



21.73 •• **CP** **Strength of the Electric Force.** Imagine two 1.0-g bags of protons, one at the earth's north pole and the other at the south pole. (a) How many protons are in each bag? (b) Calculate the gravitational attraction and the electric repulsion that each bag exerts on the other. (c) Are the forces in part (b) large enough for you to feel if you were holding one of the bags?

21.74 ••• **CP** Two tiny spheres of mass 6.80 mg carry charges of equal magnitude, 72.0 nC, but opposite sign. They are tied to the same ceiling hook by light strings of length 0.530 m. When a horizontal uniform electric field E that is directed to the left is turned on, the spheres hang at rest with the angle θ between the strings equal to 58.0° (**Fig. P21.74**). (a) Which ball (the one on the right or the one on the left) has positive charge? (b) What is the magnitude E of the field?

Figure **P21.74**



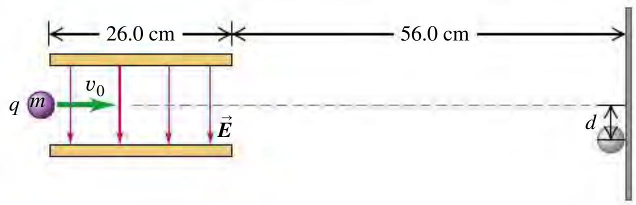
21.75 •• **CP** Consider a model of a hydrogen atom in which an electron is in a circular orbit of radius $r = 5.29 \times 10^{-11}$ m around a stationary proton. What is the speed of the electron in its orbit?

21.76 •• The earth has a downward-directed electric field near its surface of about 150 N/C. If a raindrop with a diameter of 0.020 mm is suspended, motionless, in this field, how many excess electrons must it have on its surface?

21.77 •• **CP** A proton is projected into a uniform electric field that points vertically upward and has magnitude E. The initial velocity of the proton has a magnitude v_0 and is directed at an angle α below the horizontal. (a) Find the maximum distance h_{max} that the proton descends vertically below its initial elevation. Ignore gravitational forces. (b) After what horizontal distance d does the proton return to its original elevation? (c) Sketch the trajectory of the proton. (d) Find the numerical values of h_{max} and d if $E = 500$ N/C, $v_0 = 4.00 \times 10^5$ m/s, and $\alpha = 30.0°$.

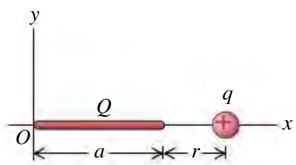
21.78 ••• A small object with mass m, charge q, and initial speed $v_0 = 5.00 \times 10^3$ m/s is projected into a uniform electric field between two parallel metal plates of length 26.0 cm (**Fig. P21.78**). The electric field between the plates is directed downward and has magnitude $E = 800$ N/C. Assume that the field is zero outside the region between the plates. The separation between the plates is large enough for the object to pass between the plates without hitting the lower plate. After passing through the field region, the object is deflected downward a vertical distance $d = 1.25$ cm from its original direction of motion and reaches a collecting plate that is 56.0 cm from the edge of the parallel plates. Ignore gravity and air resistance. Calculate the object's charge-to-mass ratio, q/m.

Figure **P21.78**



21.79 •• **CALC** Positive charge Q is distributed uniformly along the x-axis from $x = 0$ to $x = a$. A positive point charge q is located on the positive x-axis at $x = a + r$, a distance r to the right of the end of Q (**Fig. P21.79**). (a) Calculate the x- and y-components of the electric field produced by the charge distribution Q at points on the positive x-axis where $x > a$. (b) Calculate the force (magnitude and direction) that the charge distribution Q exerts on q. (c) Show that if $r \gg a$, the magnitude of the force in part (b) is approximately $Qq/4\pi\epsilon_0 r^2$. Explain why this result is obtained.

Figure **P21.79**

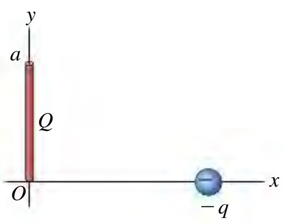


21.80 •• In a region where there is a uniform electric field that is upward and has magnitude 3.60×10^4 N/C, a small object is projected upward with an initial speed of 1.92 m/s. The object travels upward a distance of 6.98 cm in 0.200 s. What is the object's charge-to-mass ratio q/m? Assume $g = 9.80$ m/s^2, and ignore air resistance.

21.81 • A negative point charge $q_1 = -4.00$ nC is on the x-axis at $x = 0.60$ m. A second point charge q_2 is on the x-axis at $x = -1.20$ m. What must the sign and magnitude of q_2 be for the net electric field at the origin to be (a) 50.0 N/C in the $+x$-direction and (b) 50.0 N/C in the $-x$-direction?

21.82 •• CALC Positive charge Q is distributed uniformly along the positive y-axis between $y = 0$ and $y = a$. A negative point charge $-q$ lies on the positive x-axis, a distance x from the origin (**Fig. P21.82**). (a) Calculate the x- and y-components of the electric field produced by the charge distribution Q at points on the positive x-axis. (b) Calculate the x- and y-components of the force that the charge distribution Q exerts on q. (c) Show that if $x \gg a$, $F_x \cong -Qq/4\pi\epsilon_0 x^2$ and $F_y \cong +Qqa/8\pi\epsilon_0 x^3$. Explain why this result is obtained.

Figure **P21.82**



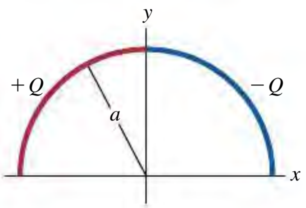
21.83 ••• A uniformly charged disk like the disk in Fig. 21.25 has radius 2.50 cm and carries a total charge of 7.0×10^{-12} C. (a) Find the electric field (magnitude and direction) on the x-axis at $x = 20.0$ cm. (b) Show that for $x \gg R$, Eq. (21.11) becomes $E = Q/4\pi\epsilon_0 x^2$, where Q is the total charge on the disk. (c) Is the magnitude of the electric field you calculated in part (a) larger or smaller than the electric field 20.0 cm from a point charge that has the same total charge as this disk? In terms of the approximation used in part (b) to derive $E = Q/4\pi\epsilon_0 x^2$ for a point charge from Eq. (21.11), explain why this is so. (d) What is the percent difference between the electric fields produced by the finite disk and by a point charge with the same charge at $x = 20.0$ cm and at $x = 10.0$ cm?

21.84 •• CP A small sphere with mass m carries a positive charge q and is attached to one end of a silk fiber of length L. The other end of the fiber is attached to a large vertical insulating sheet that has a positive surface charge density σ. Show that when the sphere is in equilibrium, the fiber makes an angle equal to arctan $(q\sigma/2mg\epsilon_0)$ with the vertical sheet.

21.85 •• CALC Negative charge $-Q$ is distributed uniformly around a quarter-circle of radius a that lies in the first quadrant, with the center of curvature at the origin. Find the x- and y-components of the net electric field at the origin.

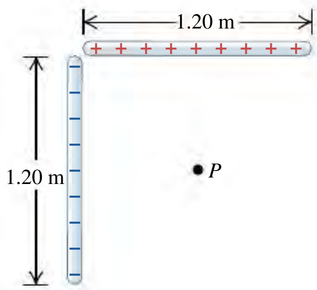
21.86 •• CALC A semicircle of radius a is in the first and second quadrants, with the center of curvature at the origin. Positive charge $+Q$ is distributed uniformly around the left half of the semicircle, and negative charge $-Q$ is distributed uniformly around the right half of the semicircle (**Fig. P21.86**). What are the magnitude and direction of the net electric field at the origin produced by this distribution of charge?

Figure **P21.86**



21.87 •• Two 1.20-m nonconducting rods meet at a right angle. One rod carries $+2.50\ \mu$C of charge distributed uniformly along its length, and the other carries $-2.50\ \mu$C distributed uniformly along it (**Fig. P21.87**). (a) Find the magnitude and direction of the electric field these rods produce at point P, which is 60.0 cm from each rod. (b) If an electron is released at P, what are the magnitude and direction of the net force that these rods exert on it?

Figure **P21.87**



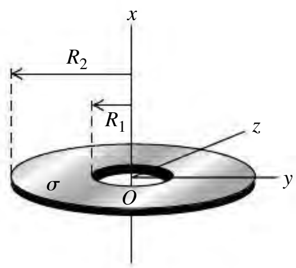
21.88 • Two very large parallel sheets are 5.00 cm apart. Sheet A carries a uniform surface charge density of $-8.80\ \mu\text{C/m}^2$, and sheet B, which is to the right of A, carries a uniform charge density of $-11.6\ \mu\text{C/m}^2$. Assume that the sheets are large enough to be treated as infinite. Find the magnitude and direction of the net electric field these sheets produce at a point (a) 4.00 cm to the right of sheet A; (b) 4.00 cm to the left of sheet A; (c) 4.00 cm to the right of sheet B.

21.89 • Repeat Problem 21.88 for the case where sheet B is positive.

21.90 • Two very large horizontal sheets are 4.25 cm apart and carry equal but opposite uniform surface charge densities of magnitude σ. You want to use these sheets to hold stationary in the region between them an oil droplet of mass 486 μg that carries an excess of five electrons. Assuming that the drop is in vacuum, (a) which way should the electric field between the plates point, and (b) what should σ be?

21.91 •• CP A thin disk with a circular hole at its center, called an *annulus*, has inner radius R_1 and outer radius R_2 (**Fig. P21.91**). The disk has a uniform positive surface charge density σ on its surface. (a) Determine the total electric charge on the annulus. (b) The annulus lies in the yz-plane, with its center at the origin. For an arbitrary point on the x-axis (the axis of the annulus), find the magnitude and direction of the electric field $\vec{E}$. Consider points both above and below the annulus. (c) Show that at points on the x-axis that are sufficiently close to the origin, the magnitude of the electric field is approximately proportional to the distance between the center of the annulus and the point. How close is "sufficiently close"? (d) A point particle with mass m and negative charge $-q$ is free to move along the x-axis (but cannot move off the axis). The particle is originally placed at rest at $x = 0.01R_1$ and released. Find the frequency of oscillation of the particle. (*Hint:* Review Section 14.2. The annulus is held stationary.)

Figure **P21.91**

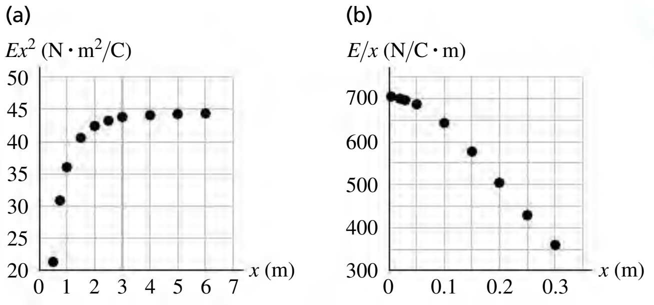


21.92 ••• DATA CP **Design of an Inkjet Printer.** Inkjet printers can be described as either continuous or drop-on-demand. In a continuous inkjet printer, letters are built up by squirting drops of ink at the paper from a rapidly moving nozzle. You are part of an engineering group working on the design of such a printer. Each ink drop will have a mass of 1.4×10^{-8} g. The drops will leave the nozzle and travel toward the paper at 50 m/s, passing through a charging unit that gives each drop a positive charge q by removing some electrons from it. The drops will then pass between parallel deflecting plates, 2.0 cm long, where there is a uniform vertical electric field with magnitude 8.0×10^4 N/C. Your team is working on the design of the charging unit that places the charge on the drops. (a) If a drop is to be deflected 0.30 mm by the time it reaches the end of the deflection plates, what magnitude of charge must be given to the drop? How many electrons must be removed from the drop to give it this charge? (b) If the unit that produces the stream of drops is redesigned so that it produces drops with a speed of 25 m/s, what q value is needed to achieve the same 0.30-mm deflection?

21.93 •• DATA Two small spheres, each carrying a net positive charge, are separated by 0.400 m. You have been asked to perform measurements that will allow you to determine the charge on each sphere. You set up a coordinate system with one sphere (charge q_1) at the origin and the other sphere (charge q_2) at $x = +0.400$ m. Available to you are a third sphere with net charge $q_3 = 4.00 \times 10^{-6}$ C and an apparatus that can accurately measure the location of this sphere and the net force on it. First you place the third sphere on the x-axis at $x = 0.200$ m; you measure the net force on it to be 4.50 N in the $+x$-direction. Then you move the third sphere to $x = +0.600$ m and measure the net force on it now to be 3.50 N in the $+x$-direction. (a) Calculate q_1 and q_2. (b) What is the net force (magnitude and direction) on q_3 if it is placed on the x-axis at $x = -0.200$ m? (c) At what value of x (other than $x = \pm\infty$) could q_3 be placed so that the net force on it is zero?

21.94 ••• DATA Positive charge Q is distributed uniformly around a very thin conducting ring of radius a, as in Fig. 21.23. You measure the electric field E at points on the ring axis, at a distance x from the center of the ring, over a wide range of values of x. (a) Your results for the larger values of x are plotted in **Fig. P21.94a** as Ex^2 versus x. Explain why the quantity Ex^2 approaches a constant value as x increases. Use Fig. P21.94a to calculate the net charge Q on the ring. (b) Your results for smaller values of x are plotted in Fig. P21.94b as E/x versus x. Explain why E/x approaches a constant value as x approaches zero. Use Fig. P21.94b to calculate a.

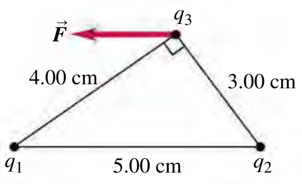
Figure **P21.94**



CHALLENGE PROBLEMS

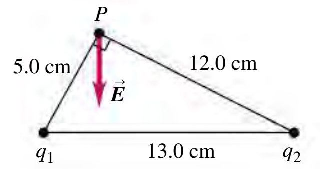
21.95 ••• Three charges are placed as shown in **Fig. P21.95**. The magnitude of q_1 is 2.00 μC, but its sign and the value of the charge q_2 are not known. Charge q_3 is $+4.00$ μC, and the net force $\vec{F}$ on q_3 is entirely in the negative x-direction. (a) Considering the different possible signs of q_1, there are four possible force diagrams representing the forces $\vec{F}_1$ and $\vec{F}_2$ that q_1 and q_2 exert on q_3. Sketch these four possible force configurations. (b) Using the sketches from part (a) and the direction of $\vec{F}$, deduce the signs of the charges q_1 and q_2. (c) Calculate the magnitude of q_2. (d) Determine F, the magnitude of the net force on q_3.

Figure **P21.95**



21.96 ••• Two charges are placed as shown in **Fig. P21.96**. The magnitude of q_1 is 3.00 μC, but its sign and the value of the charge q_2 are not known. The direction of the net electric field $\vec{E}$ at point P is entirely in the negative y-direction. (a) Considering the different possible signs of q_1 and q_2, four possible diagrams could represent the electric fields $\vec{E}_1$ and $\vec{E}_2$ produced by q_1 and q_2. Sketch the four possible electric-field configurations. (b) Using the sketches from part (a) and the direction of $\vec{E}$, deduce the signs of q_1 and q_2. (c) Determine the magnitude of $\vec{E}$.

Figure **P21.96**



21.97 ••• CALC Two thin rods of length L lie along the x-axis, one between $x = \frac{1}{2}a$ and $x = \frac{1}{2}a + L$ and the other between $x = -\frac{1}{2}a$ and $x = -\frac{1}{2}a - L$. Each rod has positive charge Q distributed uniformly along its length. (a) Calculate the electric field produced by the second rod at points along the positive x-axis. (b) Show that the magnitude of the force that one rod exerts on the other is

$$F = \frac{Q^2}{4\pi\epsilon_0 L^2}\ln\left[\frac{(a + L)^2}{a(a + 2L)}\right]$$

(c) Show that if $a \gg L$, the magnitude of this force reduces to $F = Q^2/4\pi\epsilon_0 a^2$. (*Hint:* Use the expansion $\ln(1 + z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \cdots$, valid for $|z| \ll 1$. Carry *all* expansions to at least order L^2/a^2.) Interpret this result.

PASSAGE PROBLEMS

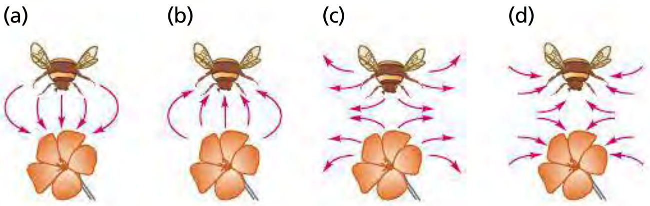
BIO ELECTRIC BEES. Flying insects such as bees may accumulate a small positive electric charge as they fly. In one experiment, the mean electric charge of 50 bees was measured to be $+(30 \pm 5)$ pC per bee. Researchers also observed the electrical properties of a plant consisting of a flower atop a long stem. The charge on the stem was measured as a positively charged bee approached, landed, and flew away. Plants are normally electrically neutral, so the measured net electric charge on the stem was zero when the bee was very far away. As the bee approached the flower, a small net positive charge was detected in the stem, even before the bee landed. Once the bee landed, the whole plant became positively charged, and this positive charge remained on the plant after the bee flew away. By creating artificial flowers with various charge values, experimenters found that bees can distinguish between charged and uncharged flowers and may use the positive electric charge left by a previous bee as a cue indicating whether a plant has already been visited (in which case, little pollen may remain).

21.98 Consider a bee with the mean electric charge found in the experiment. This charge represents roughly how many missing electrons? (a) 1.9×10^8; (b) 3.0×10^8; (c) 1.9×10^{18}; (d) 3.0×10^{18}.

21.99 What is the best explanation for the observation that the electric charge on the stem became positive as the charged bee approached (before it landed)? (a) Because air is a good conductor, the positive charge on the bee's surface flowed through the air from bee to plant. (b) Because the earth is a reservoir of large amounts of charge, positive ions were drawn up the stem from the ground toward the charged bee. (c) The plant became electrically polarized as the charged bee approached. (d) Bees that had visited the plant earlier deposited a positive charge on the stem.

21.100 After one bee left a flower with a positive charge, that bee flew away and another bee with the same amount of positive charge flew close to the plant. Which diagram in **Fig. P21.100** best represents the electric field lines between the bee and the flower?

Figure **P21.100**



21.101 In a follow-up experiment, a charge of +40 pC was placed at the center of an artificial flower at the end of a 30-cm-long stem. Bees were observed to approach no closer than 15 cm from the center of this flower before they flew away. This observation suggests that the smallest external electric field to which bees may be sensitive is closest to which of these values? (a) 2.4 N/C; (b) 16 N/C; (c) 2.7×10^{-10} N/C; (d) 4.8×10^{-10} N/C.

Answers

Chapter Opening Question ?

(ii) Water molecules have a permanent electric dipole moment: One end of the molecule has a positive charge and the other end has a negative charge. These ends attract negative and positive ions, respectively, holding the ions apart in solution. Water is less effective as a solvent for materials whose molecules do not ionize (called *nonionic* substances), such as oils.

Test Your Understanding Questions

21.1 (iv) Two charged objects repel if their charges are of the same sign (either both positive or both negative).

21.2 (a) (i), (b) (ii) Before the two spheres touch, the negatively charged sphere exerts a repulsive force on the electrons in the other sphere, causing zones of positive and negative induced charge (see Fig. 21.7b). The positive zone is closer to the negatively charged sphere than the negative zone, so there is a net force of attraction that pulls the spheres together, like the comb and insulator in Fig. 21.8b. Once the two metal spheres touch, some of the excess electrons on the negatively charged sphere will flow onto the other sphere (because metals are conductors). Then both spheres will have a net negative charge and will repel each other.

21.3 (iv) The force exerted by q_1 on Q is still as in Example 21.4. The magnitude of the force exerted by q_2 on Q is still equal to $F_{1 \text{ on } Q}$, but the direction of the force is now *toward* q_2 at an angle α below the x-axis. Hence the x-components of the two forces cancel while the (negative) y-components add together, and the total electric force is in the negative y-direction.

21.4 (a) (ii), (b) (i) The electric field $\vec{E}$ produced by a positive point charge points directly away from the charge (see Fig. 21.18a) and has a magnitude that depends on the distance r from the charge to the field point. Hence a second, negative point charge $q < 0$ will feel a force $\vec{F} = q\vec{E}$ that points directly toward the positive charge and has a magnitude that depends on the distance r between the two charges. If the negative charge moves directly toward the positive charge, the direction of the force remains the same but the force magnitude increases as the distance r decreases. If the negative charge moves in a circle around the positive charge, the force magnitude stays the same (because the distance r is constant) but the force direction changes.

21.5 (iv) Think of a pair of segments of length dy, one at coordinate $y > 0$ and the other at coordinate $-y < 0$. The upper segment has a positive charge and produces an electric field $d\vec{E}$ at P that points away from the segment, so this $d\vec{E}$ has a positive x-component and a negative y-component, like the vector $d\vec{E}$ in Fig. 21.24. The lower segment has the same amount of negative charge. It produces a $d\vec{E}$ that has the same magnitude but points *toward* the lower segment, so it has a negative x-component and a negative y-component. By symmetry, the two x-components are equal but opposite, so they cancel. Thus the total electric field has only a negative y-component.

21.6 yes If the field lines are straight, $\vec{E}$ must point in the same direction throughout the region. Hence the force $\vec{F} = q\vec{E}$ on a particle of charge q is always in the same direction. A particle released from rest accelerates in a straight line in the direction of $\vec{F}$, and so its trajectory is a straight line along a field line.

21.7 (ii) Equations (21.17) and (21.18) tell us that the potential energy for a dipole in an electric field is $U = -\vec{p} \cdot \vec{E} = -pE\cos\phi$, where ϕ is the angle between the directions of $\vec{p}$ and $\vec{E}$. If $\vec{p}$ and $\vec{E}$ point in opposite directions, so that $\phi = 180°$, we have $\cos\phi = -1$ and $U = +pE$. This is the maximum value that U can have. From our discussion of energy diagrams in Section 7.5, it follows that this is a situation of unstable equilibrium.

Bridging Problem

$E = 2kQ/\pi a^2$ in the $-y$-direction

? This child acquires an electric charge by touching the charged metal shell. The charged hairs on the child's head repel and stand out. What would happen if the child stood *inside* a large, charged metal shell? She would acquire (i) the same sign of charge as on the shell, and her hairs would stand out; (ii) the opposite sign of charge as on the shell, and her hairs would stand out; (iii) no charge, and her hairs would be relaxed; (iv) any of these, depending on the amount of charge on the shell.

22 GAUSS'S LAW

LEARNING GOALS

Looking forward at ...

22.1 How you can determine the amount of charge within a closed surface by examining the electric field on the surface.

22.2 What is meant by electric flux, and how to calculate it.

22.3 How Gauss's law relates the electric flux through a closed surface to the charge enclosed by the surface.

22.4 How to use Gauss's law to calculate the electric field due to a symmetric charge distribution.

22.5 Where the charge is located on a charged conductor.

Looking back at ...

21.4–21.6 Electric fields and their properties.

In physics, an important tool for simplifying problems is the *symmetry properties* of systems. Many physical systems have symmetry; for example, a cylindrical body doesn't look any different after you've rotated it around its axis, and a charged metal sphere looks just the same after you've turned it about any axis through its center.

In this chapter we'll use symmetry ideas along with a new principle, called *Gauss's law,* to simplify electric-field calculations. For example, the field of a straight-line or plane-sheet charge distribution, which we derived in Section 21.5 by using some fairly strenuous integrations, can be obtained in a few steps with the help of Gauss's law. But Gauss's law is more than just a way to make certain calculations easier. Indeed, it is a fundamental statement about the relationship between electric charges and electric fields. Among other things, Gauss's law can help us understand how electric charge distributes itself over conducting bodies.

Here's what Gauss's law is all about. Given any general distribution of charge, we surround it with an imaginary surface that encloses the charge. Then we look at the electric field at various points on this imaginary surface. Gauss's law is a relationship between the field at *all* the points on the surface and the total charge enclosed within the surface. This may sound like a rather indirect way of expressing things, but it turns out to be a tremendously useful relationship. In the next several chapters, we'll make frequent use of the insights that Gauss's law provides into the character of electric fields.

22.1 CHARGE AND ELECTRIC FLUX

The discussion of Gauss's law in this section is based on and inspired by the innovative ideas of Ruth W. Chabay and Bruce A. Sherwood in *Electric and Magnetic Interactions* (John Wiley & Sons, 1994).

In Chapter 21 we asked the question, "Given a charge distribution, what is the electric field produced by that distribution at a point P?" We saw that the answer could be found by representing the distribution as an assembly of point charges, each of which produces an electric field $\vec{E}$ given by Eq. (21.7). The total field at P is then the vector sum of the fields due to all the point charges.

But there is an alternative relationship between charge distributions and electric fields. To discover this relationship, let's stand the question of Chapter 21 on its head and ask, "If the electric-field pattern is known in a given region, what can we determine about the charge distribution in that region?"

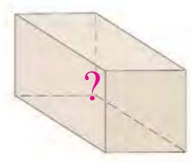
Here's an example. Consider the box shown in **Fig. 22.1a**, which may or may not contain electric charge. We'll imagine that the box is made of a material that has no effect on any electric fields; it's of the same breed as the massless rope and the frictionless incline. Better still, let the box represent an *imaginary* surface that may or may not enclose some charge. We'll refer to the box as a **closed surface** because it completely encloses a volume. How can you determine how much (if any) electric charge lies within the box?

Knowing that a charge distribution produces an electric field and that an electric field exerts a force on a test charge, you move a test charge q_0 around the vicinity of the box. By measuring the force $\vec{F}$ experienced by the test charge at different positions, you make a three-dimensional map of the electric field $\vec{E} = \vec{F}/q_0$ outside the box. In the case shown in Fig. 22.1b, the map turns out to be the same as that of the electric field produced by a positive point charge (Fig. 21.28a). From the details of the map, you can find the exact value of the point charge inside the box.

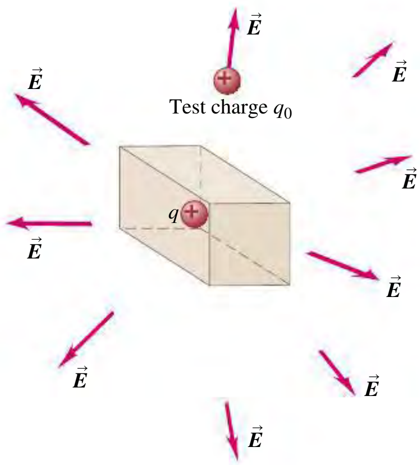
To determine the contents of the box, we actually need to measure $\vec{E}$ on only the *surface* of the box. In **Fig. 22.2a** there is a single *positive* point charge inside the box, and in Fig. 22.2b there are two such charges. The field patterns on the surfaces of the boxes are different in detail, but in each case the electric field points *out* of the box. Figures 22.2c and 22.2d show cases with one and two *negative* point charges, respectively, inside the box. Again, the details of $\vec{E}$ are different for the two cases, but the electric field points *into* each box.

22.1 How can you measure the charge inside a box without opening it?

(a) A box containing an unknown amount of charge



(b) Using a test charge outside the box to probe the amount of charge inside the box



Electric Flux and Enclosed Charge

In Section 21.4 we mentioned the analogy between electric-field vectors and the velocity vectors of a fluid in motion. This analogy can be helpful, even though an electric field does not actually "flow." Using this analogy, in Figs. 22.2a and 22.2b, in which the electric-field vectors point out of the surface, we say that there is an *outward* **electric flux.** (The word "flux" comes from a Latin word meaning "flow.") In Figs. 22.2c and 22.2d the $\vec{E}$ vectors point into the surface, and the electric flux is *inward.*

Figure 22.2 suggests a simple relationship: Positive charge inside the box goes with an outward electric flux through the box's surface, and negative charge inside goes with an inward electric flux. What happens if there is *zero*

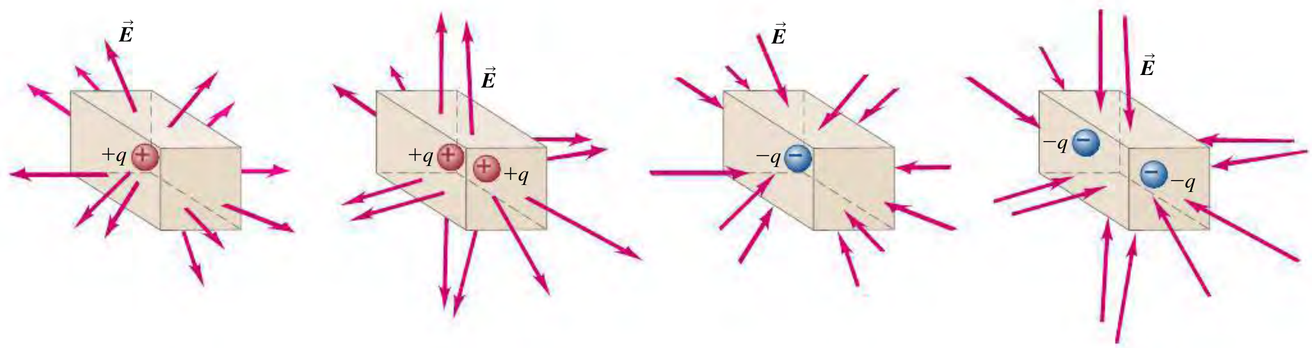
22.2 The electric field on the surface of boxes containing (a) a single positive point charge, (b) two positive point charges, (c) a single negative point charge, or (d) two negative point charges.

(a) Positive charge inside box, outward flux

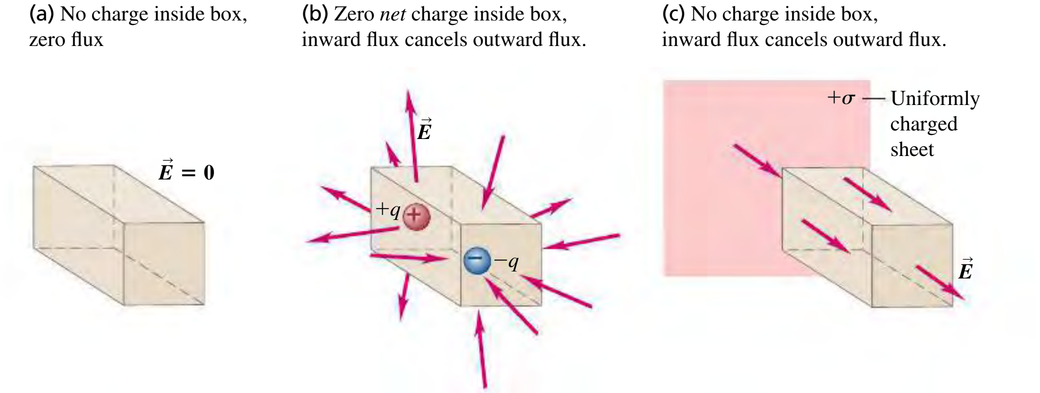
(b) Positive charges inside box, outward flux

(c) Negative charge inside box, inward flux

(d) Negative charges inside box, inward flux



22.3 Three cases in which there is zero *net* charge inside a box and no net electric flux through the surface of the box. (a) An empty box with $\vec{E} = \mathbf{0}$. (b) A box containing one positive and one equal-magnitude negative point charge. (c) An empty box immersed in a uniform electric field.



charge inside the box? In **Fig. 22.3a** the box is empty and $\vec{E} = \mathbf{0}$ everywhere, so there is no electric flux into or out of the box. In Fig. 22.3b, one positive and one negative point charge of equal magnitude are enclosed within the box, so the *net* charge inside the box is zero. There is an electric field, but it "flows into" the box on half of its surface and "flows out of" the box on the other half. Hence there is no *net* electric flux into or out of the box.

The box is again empty in Fig. 22.3c. However, there is charge present *outside* the box; the box has been placed with one end parallel to a uniformly charged infinite sheet, which produces a uniform electric field perpendicular to the sheet (see Example 21.11 of Section 21.5). On one end of the box, $\vec{E}$ points into the box; on the opposite end, $\vec{E}$ points out of the box; and on the sides, $\vec{E}$ is parallel to the surface and so points neither into nor out of the box. As in Fig. 22.3b, the inward electric flux on one part of the box exactly compensates for the outward electric flux on the other part. So in all of the cases shown in Fig. 22.3, there is no *net* electric flux through the surface of the box, and no *net* charge is enclosed in the box.

Figures 22.2 and 22.3 demonstrate a connection between the *sign* (positive, negative, or zero) of the *net* charge enclosed by a closed surface and the sense (outward, inward, or none) of the net electric flux through the surface. There is also a connection between the *magnitude* of the net charge inside the closed surface and the *strength* of the net "flow" of $\vec{E}$ over the surface. In both **Figs. 22.4a** and 22.4b there is a single point charge inside the box, but in Fig. 22.4b the magnitude of the charge is twice as great, and so $\vec{E}$ is everywhere twice as great in magnitude as in Fig. 22.4a. If we keep in mind the fluid-flow analogy, this means that the net outward electric flux is also twice as great in Fig. 22.4b as in Fig. 22.4a. This suggests that the net electric flux through the surface of the box is *directly proportional* to the magnitude of the net charge enclosed by the box.

This conclusion is independent of the size of the box. In Fig. 22.4c the point charge $+q$ is enclosed by a box with twice the linear dimensions of the box in Fig. 22.4a. The magnitude of the electric field of a point charge decreases with distance according to $1/r^2$, so the average magnitude of $\vec{E}$ on each face of the large box in Fig. 22.4c is just $\frac{1}{4}$ of the average magnitude on the corresponding face in Fig. 22.4a. But each face of the large box has exactly four times the area of the corresponding face of the small box. Hence the outward electric flux is the *same* for the two boxes if we *define* electric flux as follows: For each face of the box, take the product of the average perpendicular component of $\vec{E}$ and the area of that face; then add up the results from all faces of the box. With this definition the net electric flux due to a single point charge inside the box is independent of the size of the box and depends only on the net charge inside the box.

To summarize, for the special cases of a closed surface in the shape of a rectangular box and charge distributions made up of point charges or infinite charged sheets, we have found:

1. Whether there is a net outward or inward electric flux through a closed surface depends on the sign of the enclosed charge.
2. Charges *outside* the surface do not give a net electric flux through the surface.
3. The net electric flux is directly proportional to the net amount of charge enclosed within the surface but is otherwise independent of the size of the closed surface.

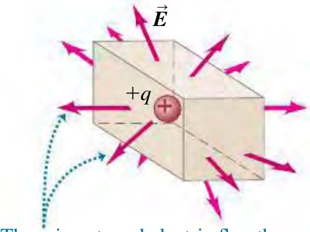
These observations are a qualitative statement of *Gauss's law.*

Do these observations hold true for other kinds of charge distributions and for closed surfaces of arbitrary shape? The answer to these questions will prove to be yes. But to explain why this is so, we need a precise mathematical statement of what we mean by electric flux. We develop this in the next section.

TEST YOUR UNDERSTANDING OF SECTION 22.1 If all of the dimensions of the box in Fig. 22.2a are increased by a factor of 3, how will the electric flux through the box change? (i) The flux will be $3^2 = 9$ times greater; (ii) the flux will be 3 times greater; (iii) the flux will be unchanged; (iv) the flux will be $\frac{1}{3}$ as great; (v) the flux will be $\left(\frac{1}{3}\right)^2 = \frac{1}{9}$ as great; (vi) not enough information is given to decide. ❙

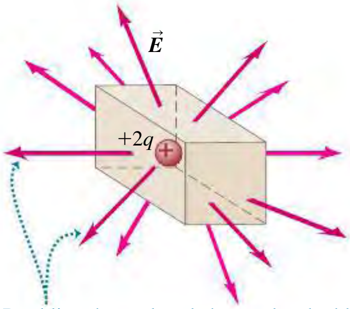
22.4 Three boxes, each of which encloses a positive point charge.

(a) A box containing a positive point charge $+q$



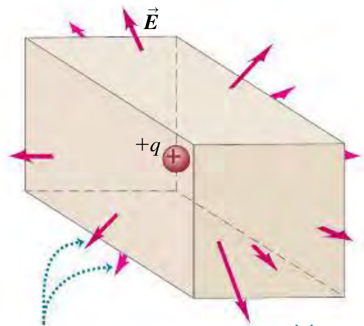
There is outward electric flux through the surface.

(b) The same box as in (a), but containing a positive point charge $+2q$



Doubling the enclosed charge also doubles the magnitude of the electric field on the surface, so the electric flux through the surface is twice as great as in (a).

(c) The same positive point charge $+q$, but enclosed by a box with twice the dimensions



The electric flux is the same as in (a): The magnitude of the electric field on the surface is $\frac{1}{4}$ as great as in (a), but the area through which the field "flows" is 4 times greater.

22.2 CALCULATING ELECTRIC FLUX

In the preceding section we introduced the concept of *electric flux.* We used this to give a rough qualitative statement of Gauss's law: The net electric flux through a closed surface is directly proportional to the net charge inside that surface. To be able to make full use of this law, we need to know how to *calculate* electric flux. To do this, let's again make use of the analogy between an electric field $\vec{E}$ and the field of velocity vectors $\vec{v}$ in a flowing fluid. (Again, keep in mind that this is only an analogy; an electric field is *not* a flow.)

Flux: Fluid-Flow Analogy

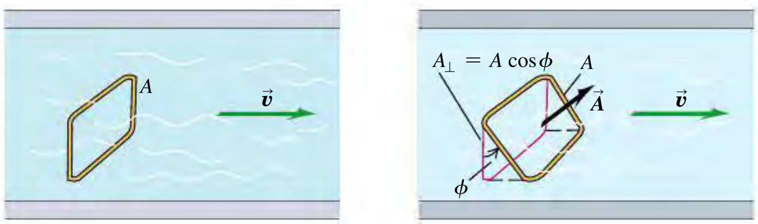
Figure 22.5 shows a fluid flowing steadily from left to right. Let's examine the volume flow rate dV/dt (in, say, cubic meters per second) through the wire rectangle with area A. When the area is perpendicular to the flow velocity $\vec{v}$ (Fig. 22.5a) and the flow velocity is the same at all points in the fluid, the volume flow rate dV/dt is the area A multiplied by the flow speed v:

$$\frac{dV}{dt} = vA$$

When the rectangle is tilted at an angle ϕ (Fig. 22.5b) so that its face is not perpendicular to $\vec{v}$, the area that counts is the silhouette area that we see when we look in the direction of $\vec{v}$. This area, which is outlined in red and labeled $A_\perp$ in Fig. 22.5b, is the *projection* of the area A onto a surface perpendicular to $\vec{v}$. Two sides of the projected rectangle have the same length as the original one, but the

(a) A wire rectangle in a fluid

(b) The wire rectangle tilted by an angle ϕ



22.5 The volume flow rate of fluid through the wire rectangle (a) is vA when the area of the rectangle is perpendicular to $\vec{v}$ and (b) is $vA\cos\phi$ when the rectangle is tilted at an angle ϕ.

other two are foreshortened by a factor of $\cos\phi$, so the projected area $A_\perp$ is equal to $A\cos\phi$. Then the volume flow rate through A is

$$\frac{dV}{dt} = vA\cos\phi$$

If $\phi = 90°$, $dV/dt = 0$; the wire rectangle is edge-on to the flow, and no fluid passes through the rectangle.

Also, $v\cos\phi$ is the component of the vector $\vec{v}$ perpendicular to the plane of the area A. Calling this component $v_\perp$, we can rewrite the volume flow rate as

$$\frac{dV}{dt} = v_\perp A$$

We can express the volume flow rate more compactly by using the concept of *vector area* $\vec{A}$, a vector quantity with magnitude A and a direction perpendicular to the plane of the area we are describing. The vector area $\vec{A}$ describes both the size of an area and its orientation in space. In terms of $\vec{A}$, we can write the volume flow rate of fluid through the rectangle in Fig. 22.5b as a scalar (dot) product:

$$\frac{dV}{dt} = \vec{v}\cdot\vec{A}$$

BIO Application Flux Through a Basking Shark's Mouth Unlike aggressive carnivorous sharks such as great whites, a basking shark feeds passively on plankton in the water that passes through the shark's gills as it swims. To survive on these tiny organisms requires a huge flux of water through a basking shark's immense mouth, which can be up to a meter across. The water flux—the product of the shark's speed through the water and the area of its mouth—can be up to 0.5 m^3/s (500 liters per second, or almost 5×10^5 gallons per hour). In a similar way, the flux of electric field through a surface depends on the magnitude of the field and the area of the surface (as well as the relative orientation of the field and surface).

Flux of a Uniform Electric Field

Using the analogy between electric field and fluid flow, we now define electric flux in the same way as we have just defined the volume flow rate of a fluid; we simply replace the fluid velocity $\vec{v}$ by the electric field $\vec{E}$. The symbol that we use for electric flux is Φ_E (the capital Greek letter phi; the subscript E is a reminder that this is *electric* flux). Consider first a flat area A perpendicular to a uniform electric field $\vec{E}$ (**Fig. 22.6a**). We define the electric flux through this area to be the product of the field magnitude E and the area A:

$$\Phi_E = EA$$

Roughly speaking, we can picture Φ_E in terms of the field lines passing through A. Increasing the area means that more lines of $\vec{E}$ pass through the area, increasing the flux; a stronger field means more closely spaced lines of $\vec{E}$ and therefore more lines per unit area, so again the flux increases.

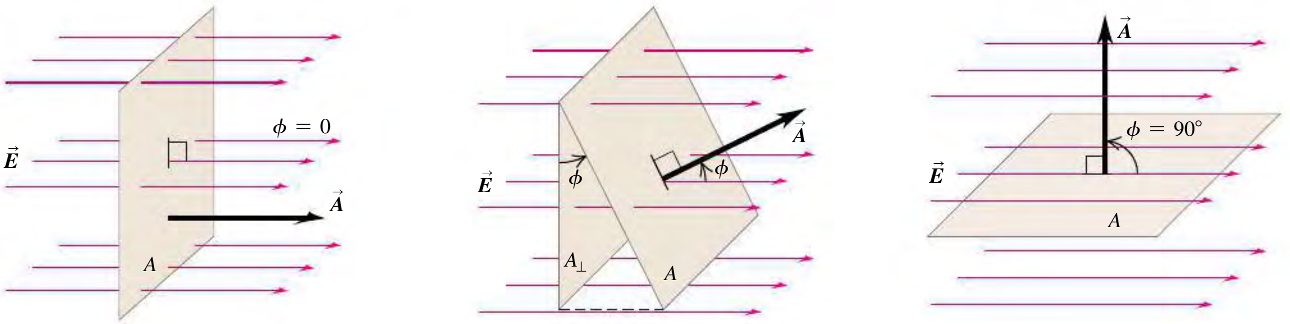
If the area A is flat but not perpendicular to the field $\vec{E}$, then fewer field lines pass through it. In this case the area that counts is the silhouette area that we see

22.6 A flat surface in a uniform electric field. The electric flux Φ_E through the surface equals the scalar product of the electric field $\vec{E}$ and the area vector $\vec{A}$.

(a) Surface is face-on to electric field:
- $\vec{E}$ and $\vec{A}$ are parallel (the angle between $\vec{E}$ and $\vec{A}$ is $\phi = 0$).
- The flux $\Phi_E = \vec{E}\cdot\vec{A} = EA$.

(b) Surface is tilted from a face-on orientation by an angle ϕ:
- The angle between $\vec{E}$ and $\vec{A}$ is ϕ.
- The flux $\Phi_E = \vec{E}\cdot\vec{A} = EA\cos\phi$.

(c) Surface is edge-on to electric field:
- $\vec{E}$ and $\vec{A}$ are perpendicular (the angle between $\vec{E}$ and $\vec{A}$ is $\phi = 90°$).
- The flux $\Phi_E = \vec{E}\cdot\vec{A} = EA\cos 90° = 0$.



when looking in the direction of $\vec{E}$. This is the area $A_{\perp}$ in Fig. 22.6b and is equal to $A\cos\phi$ (compare to Fig. 22.5b). We generalize our definition of electric flux for a uniform electric field to

$$\Phi_E = EA\cos\phi \qquad \text{(electric flux for uniform } \vec{E}\text{, flat surface)} \tag{22.1}$$

Since $E\cos\phi$ is the component of $\vec{E}$ perpendicular to the area, we can rewrite Eq. (22.1) as

$$\Phi_E = E_{\perp}A \qquad \text{(electric flux for uniform } \vec{E}\text{, flat surface)} \tag{22.2}$$

In terms of the vector area $\vec{A}$ perpendicular to the area, we can write the electric flux as the scalar product of $\vec{E}$ and $\vec{A}$:

$$\Phi_E = \vec{E}\cdot\vec{A} \qquad \text{(electric flux for uniform } \vec{E}\text{, flat surface)} \tag{22.3}$$

Equations (22.1), (22.2), and (22.3) express the electric flux for a *flat* surface and a *uniform* electric field in different but equivalent ways. The SI unit for electric flux is $1\ \text{N}\cdot\text{m}^2/\text{C}$. Note that if the area is edge-on to the field, $\vec{E}$ and $\vec{A}$ are perpendicular and the flux is zero (Fig. 22.6c).

We can represent the direction of a vector area $\vec{A}$ by using a *unit vector* $\hat{n}$ perpendicular to the area; $\hat{n}$ stands for "normal." Then

$$\vec{A} = A\hat{n} \tag{22.4}$$

A surface has two sides, so there are two possible directions for $\hat{n}$ and $\vec{A}$. We must always specify which direction we choose. In Section 22.1 we related the charge inside a *closed* surface to the electric flux through the surface. With a closed surface we will always choose the direction of $\hat{n}$ to be *outward,* and we will speak of the flux *out of* a closed surface. Thus what we called "outward electric flux" in Section 22.1 corresponds to a *positive* value of Φ_E, and what we called "inward electric flux" corresponds to a *negative* value of Φ_E.

Flux of a Nonuniform Electric Field

What happens if the electric field $\vec{E}$ isn't uniform but varies from point to point over the area A? Or what if A is part of a curved surface? Then we divide A into many small elements dA, each of which has a unit vector $\hat{n}$ perpendicular to it and a vector area $d\vec{A} = \hat{n}\,dA$. We calculate the electric flux through each element and integrate the results to obtain the total flux:

$$\underbrace{\Phi_E}_{\text{Electric flux through a surface}} = \int \overbrace{E}^{\text{Magnitude of electric field } \vec{E}} \underbrace{\cos\phi}_{\text{Angle between } \vec{E} \text{ and normal to surface}}\, \underbrace{dA}_{\text{Element of surface area}} = \int \overbrace{E_{\perp}}^{\text{Component of } \vec{E} \text{ perpendicular to surface}}\, \underbrace{dA}_{\text{Element of surface area}} = \int \vec{E}\cdot \underbrace{d\vec{A}}_{\text{Vector element of surface area}} \tag{22.5}$$

We call this integral the **surface integral** of the component $E_{\perp}$ over the area, or the surface integral of $\vec{E}\cdot d\vec{A}$. In specific problems, one form of the integral is sometimes more convenient than another. Example 22.3 at the end of this section illustrates the use of Eq. (22.5).

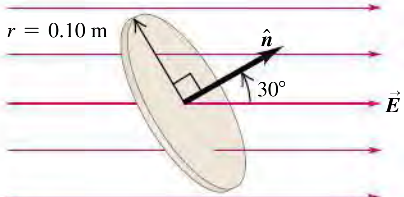
In Eq. (22.5) the electric flux $\int E_{\perp}\,dA$ is equal to the *average* value of the perpendicular component of the electric field, multiplied by the area of the surface. This is the same definition of electric flux that we were led to in Section 22.1, now expressed more mathematically. In the next section we will see the connection between the total electric flux through *any* closed surface, no matter what its shape, and the amount of charge enclosed within that surface.

EXAMPLE 22.1 ELECTRIC FLUX THROUGH A DISK

SOLUTION

A disk of radius 0.10 m is oriented with its normal unit vector $\hat{\boldsymbol{n}}$ at 30° to a uniform electric field $\vec{\boldsymbol{E}}$ of magnitude 2.0×10^3 N/C (**Fig. 22.7**). (Since this isn't a closed surface, it has no "inside" or "outside." That's why we have to specify the direction of $\hat{\boldsymbol{n}}$ in the figure.) (a) What is the electric flux through the disk? (b) What is the flux through the disk if it is turned so that $\hat{\boldsymbol{n}}$ is perpendicular to $\vec{\boldsymbol{E}}$? (c) What is the flux through the disk if $\hat{\boldsymbol{n}}$ is parallel to $\vec{\boldsymbol{E}}$?

22.7 The electric flux Φ_E through a disk depends on the angle between its normal $\hat{\boldsymbol{n}}$ and the electric field $\vec{\boldsymbol{E}}$.



SOLUTION

IDENTIFY and SET UP: This problem is about a flat surface in a uniform electric field, so we can apply the ideas of this section. We calculate the electric flux from Eq. (22.1).

EXECUTE: (a) The area is $A = \pi(0.10\ \text{m})^2 = 0.0314\ \text{m}^2$ and the angle between $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{A}} = A\hat{\boldsymbol{n}}$ is $\phi = 30°$, so from Eq. (22.1),

$$\Phi_E = EA\cos\phi = (2.0 \times 10^3\ \text{N/C})(0.0314\ \text{m}^2)(\cos 30°)$$
$$= 54\ \text{N}\cdot\text{m}^2/\text{C}$$

(b) The normal to the disk is now perpendicular to $\vec{\boldsymbol{E}}$, so $\phi = 90°$, $\cos\phi = 0$, and $\Phi_E = 0$.

(c) The normal to the disk is parallel to $\vec{\boldsymbol{E}}$, so $\phi = 0$ and $\cos\phi = 1$:

$$\Phi_E = EA\cos\phi = (2.0 \times 10^3\ \text{N/C})(0.0314\ \text{m}^2)(1)$$
$$= 63\ \text{N}\cdot\text{m}^2/\text{C}$$

EVALUATE: Our answer to part (b) is smaller than that to part (a), which is in turn smaller than that to part (c). Is all this as it should be?

EXAMPLE 22.2 ELECTRIC FLUX THROUGH A CUBE

SOLUTION

An imaginary cubical surface of side L is in a region of uniform electric field $\vec{\boldsymbol{E}}$. Find the electric flux through each face of the cube and the total flux through the cube when (a) it is oriented with two of its faces perpendicular to $\vec{\boldsymbol{E}}$ (**Fig. 22.8a**) and (b) the cube is turned by an angle θ about a vertical axis (Fig. 22.8b).

SOLUTION

IDENTIFY and SET UP: Since $\vec{\boldsymbol{E}}$ is uniform and each of the six faces of the cube is flat, we find the flux Φ_{Ei} through each face from Eqs. (22.3) and (22.4). The total flux through the cube is the sum of the six individual fluxes.

EXECUTE: (a) Figure 22.8a shows the unit vectors $\hat{\boldsymbol{n}}_1$ through $\hat{\boldsymbol{n}}_6$ for each face; each unit vector points *outward* from the cube's closed surface. The angle between $\vec{\boldsymbol{E}}$ and $\hat{\boldsymbol{n}}_1$ is 180°, the angle between $\vec{\boldsymbol{E}}$ and $\hat{\boldsymbol{n}}_2$ is 0°, and the angle between $\vec{\boldsymbol{E}}$ and each of the other four unit vectors is 90°. Each face of the cube has area L^2, so the fluxes through the faces are

$$\Phi_{E1} = \vec{\boldsymbol{E}}\cdot\hat{\boldsymbol{n}}_1 A = EL^2\cos 180° = -EL^2$$
$$\Phi_{E2} = \vec{\boldsymbol{E}}\cdot\hat{\boldsymbol{n}}_2 A = EL^2\cos 0° = +EL^2$$
$$\Phi_{E3} = \Phi_{E4} = \Phi_{E5} = \Phi_{E6} = EL^2\cos 90° = 0$$

The flux is negative on face 1, where $\vec{\boldsymbol{E}}$ is directed into the cube, and positive on face 2, where $\vec{\boldsymbol{E}}$ is directed out of the cube. The total flux through the cube is

$$\Phi_E = \Phi_{E1} + \Phi_{E2} + \Phi_{E3} + \Phi_{E4} + \Phi_{E5} + \Phi_{E6}$$
$$= -EL^2 + EL^2 + 0 + 0 + 0 + 0 = 0$$

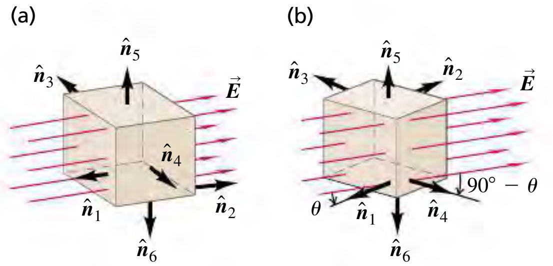
(b) The field $\vec{\boldsymbol{E}}$ is directed into faces 1 and 3, so the fluxes through them are negative; $\vec{\boldsymbol{E}}$ is directed out of faces 2 and 4, so the fluxes through them are positive. We find

$$\Phi_{E1} = \vec{\boldsymbol{E}}\cdot\hat{\boldsymbol{n}}_1 A = EL^2\cos(180° - \theta) = -EL^2\cos\theta$$
$$\Phi_{E2} = \vec{\boldsymbol{E}}\cdot\hat{\boldsymbol{n}}_2 A = +EL^2\cos\theta$$
$$\Phi_{E3} = \vec{\boldsymbol{E}}\cdot\hat{\boldsymbol{n}}_3 A = EL^2\cos(90° + \theta) = -EL^2\sin\theta$$
$$\Phi_{E4} = \vec{\boldsymbol{E}}\cdot\hat{\boldsymbol{n}}_4 A = EL^2\cos(90° - \theta) = +EL^2\sin\theta$$
$$\Phi_{E5} = \Phi_{E6} = EL^2\cos 90° = 0$$

The total flux $\Phi_E = \Phi_{E1} + \Phi_{E2} + \Phi_{E3} + \Phi_{E4} + \Phi_{E5} + \Phi_{E6}$ through the surface of the cube is again zero.

EVALUATE: We came to the same conclusion in our discussion of Fig. 22.3c: There is zero net flux of a uniform electric field through a closed surface that contains no electric charge.

22.8 Electric flux of a uniform field $\vec{\boldsymbol{E}}$ through a cubical box of side L in two orientations.



EXAMPLE 22.3 ELECTRIC FLUX THROUGH A SPHERE

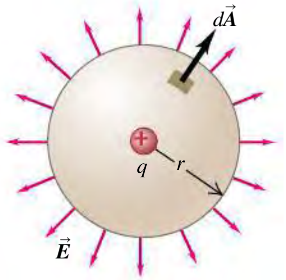
SOLUTION

A point charge $q = +3.0\ \mu C$ is surrounded by an imaginary sphere of radius $r = 0.20$ m centered on the charge (**Fig. 22.9**). Find the resulting electric flux through the sphere.

SOLUTION

IDENTIFY and SET UP: The surface is not flat and the electric field is not uniform, so to calculate the electric flux (our target variable) we must use the general definition, Eq. (22.5). Because the sphere is centered on the point charge, at any point on the spherical surface, $\vec{E}$ is directed out of the sphere perpendicular to the surface. The positive direction for both $\hat{n}$ and $E_\perp$ is outward, so $E_\perp = E$ and the flux through a surface element dA is $\vec{E} \cdot d\vec{A} = E\,dA$. This greatly simplifies the integral in Eq. (22.5).

22.9 Electric flux through a sphere centered on a point charge.



EXECUTE: We must evaluate the integral of Eq. (22.5), $\Phi_E = \int E\,dA$. At any point on the sphere of radius r the electric field has the same magnitude $E = q/4\pi\epsilon_0 r^2$. Hence E can be taken outside the integral, which becomes $\Phi_E = E\int dA = EA$, where A is the area of the spherical surface: $A = 4\pi r^2$. Hence the total flux through the sphere is

$$\Phi_E = EA = \frac{q}{4\pi\epsilon_0 r^2}4\pi r^2 = \frac{q}{\epsilon_0}$$

$$= \frac{3.0 \times 10^{-6}\ \text{C}}{8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2} = 3.4 \times 10^5\ \text{N} \cdot \text{m}^2/\text{C}$$

EVALUATE: The radius r of the sphere cancels out of the result for Φ_E. We would have obtained the same flux with a sphere of radius 2.0 m or 200 m. We came to essentially the same conclusion in our discussion of Fig. 22.4 in Section 22.1, where we considered rectangular closed surfaces of two different sizes enclosing a point charge. There we found that the flux of $\vec{E}$ was independent of the size of the surface; the same result holds true for a spherical surface. Indeed, the flux through *any* surface enclosing a single point charge is independent of the shape or size of the surface, as we'll soon see.

TEST YOUR UNDERSTANDING OF SECTION 22.2 Rank the following surfaces in order from most positive to most negative electric flux: (i) a flat rectangular surface with vector area $\vec{A} = (6.0\ \text{m}^2)\hat{\imath}$ in a uniform electric field $\vec{E} = (4.0\ \text{N/C})\hat{\jmath}$; (ii) a flat circular surface with vector area $\vec{A} = (3.0\ \text{m}^2)\hat{\jmath}$ in a uniform electric field $\vec{E} = (4.0\ \text{N/C})\hat{\imath} + (2.0\ \text{N/C})\hat{\jmath}$; (iii) a flat square surface with vector area $\vec{A} = (3.0\ \text{m}^2)\hat{\imath} + (7.0\ \text{m}^2)\hat{\jmath}$ in a uniform electric field $\vec{E} = (4.0\ \text{N/C})\hat{\imath} - (2.0\ \text{N/C})\hat{\jmath}$; (iv) a flat oval surface with vector area $\vec{A} = (3.0\ \text{m}^2)\hat{\imath} - (7.0\ \text{m}^2)\hat{\jmath}$ in a uniform electric field $\vec{E} = (4.0\ \text{N/C})\hat{\imath} - (2.0\ \text{N/C})\hat{\jmath}$. ❙

22.3 GAUSS'S LAW

Gauss's law is an alternative to Coulomb's law. While completely equivalent to Coulomb's law, Gauss's law provides a different way to express the relationship between electric charge and electric field. It was formulated by Carl Friedrich Gauss (1777–1855), one of the greatest mathematicians of all time (**Fig. 22.10**).

22.10 Carl Friedrich Gauss helped develop several branches of mathematics, including differential geometry, real analysis, and number theory. The "bell curve" of statistics is one of his inventions. Gauss also made state-of-the-art investigations of the earth's magnetism and calculated the orbit of the first asteroid to be discovered.

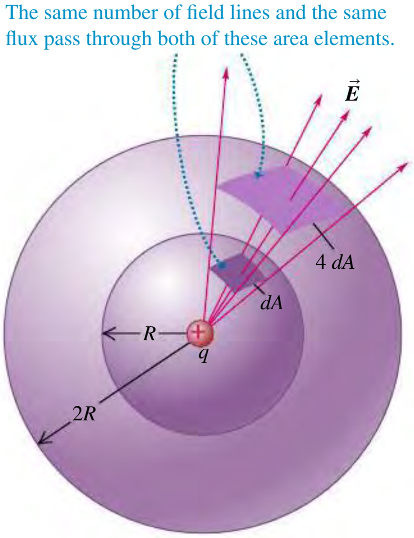
Point Charge Inside a Spherical Surface

Gauss's law states that the total electric flux through any closed surface (a surface enclosing a definite volume) is proportional to the total (net) electric charge inside the surface. In Section 22.1 we observed this relationship qualitatively; now we'll develop it more rigorously. We'll start with the field of a single positive point charge q. The field lines radiate out equally in all directions. We place this charge at the center of an imaginary spherical surface with radius R. The magnitude E of the electric field at every point on the surface is given by

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{R^2}$$

At each point on the surface, $\vec{E}$ is perpendicular to the surface, and its magnitude is the same at every point, as in Example 22.3 (Section 22.2). The total electric

22.11 Projection of an element of area dA of a sphere of radius R onto a concentric sphere of radius $2R$. The projection multiplies each linear dimension by 2, so the area element on the larger sphere is $4\,dA$.



flux is the product of the field magnitude E and the total area $A = 4\pi R^2$ of the sphere:

$$\Phi_E = EA = \frac{1}{4\pi\epsilon_0}\frac{q}{R^2}(4\pi R^2) = \frac{q}{\epsilon_0} \tag{22.6}$$

The flux is independent of the radius R of the sphere. It depends on only the charge q enclosed by the sphere.

We can also interpret this result in terms of field lines. **Figure 22.11** shows two spheres with radii R and $2R$ centered on the point charge q. Every field line that passes through the smaller sphere also passes through the larger sphere, so the total flux through each sphere is the same.

What is true of the entire sphere is also true of any portion of its surface. In Fig. 22.11 an area dA is outlined on the sphere of radius R and projected onto the sphere of radius $2R$ by drawing lines from the center through points on the boundary of dA. The area projected on the larger sphere is clearly $4\,dA$. But since the electric field due to a point charge is inversely proportional to r^2, the field magnitude is $\frac{1}{4}$ as great on the sphere of radius $2R$ as on the sphere of radius R. Hence the electric flux is the same for both areas and is independent of the radius of the sphere.

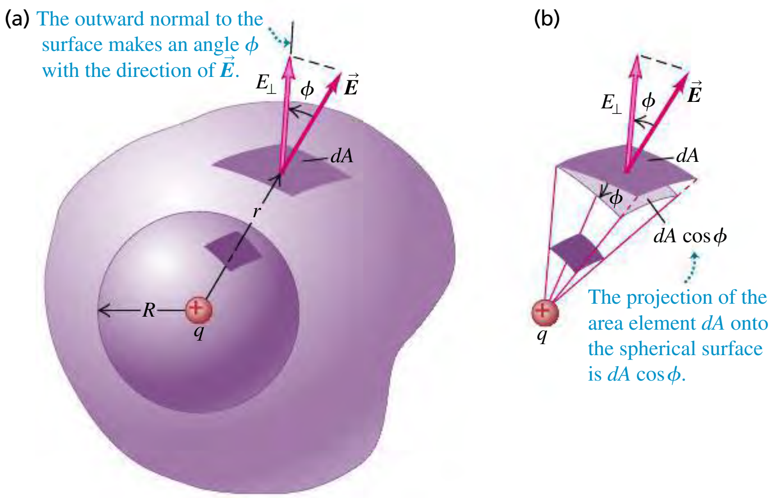
Point Charge Inside a Nonspherical Surface

We can extend this projection technique to nonspherical surfaces. Instead of a second sphere, let us surround the sphere of radius R by a surface of irregular shape, as in **Fig. 22.12a**. Consider a small element of area dA on the irregular surface; we note that this area is *larger* than the corresponding element on a spherical surface at the same distance from q. If a normal to dA makes an angle ϕ with a radial line from q, two sides of the area projected onto the spherical surface are foreshortened by a factor $\cos\phi$ (Fig. 22.12b). The other two sides are unchanged. Thus the electric flux through the spherical surface element is equal to the flux $E\,dA\cos\phi$ through the corresponding irregular surface element.

We can divide the entire irregular surface into elements dA, compute the electric flux $E\,dA\cos\phi$ for each, and sum the results by integrating, as in Eq. (22.5). Each of the area elements projects onto a corresponding spherical surface element. Thus the *total* electric flux through the irregular surface, given by any of the forms of Eq. (22.5), must be the same as the total flux through a sphere, which Eq. (22.6) shows is equal to q/ϵ_0. Thus, for the irregular surface,

$$\Phi_E = \oint \vec{E}\cdot d\vec{A} = \frac{q}{\epsilon_0} \tag{22.7}$$

22.12 Calculating the electric flux through a nonspherical surface.



Equation (22.7) holds for a surface of *any* shape or size, provided only that it is a *closed* surface enclosing the charge q. The circle on the integral sign reminds us that the integral is always taken over a *closed* surface.

The area elements $d\vec{A}$ and the corresponding unit vectors $\hat{n}$ always point *out of* the volume enclosed by the surface. The electric flux is then positive in areas where the electric field points out of the surface and negative where it points inward. Also, $E_\perp$ is positive at points where $\vec{E}$ points out of the surface and negative at points where $\vec{E}$ points into the surface.

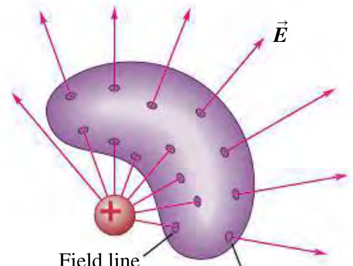
If the point charge in Fig. 22.12 is negative, the $\vec{E}$ field is directed radially *inward;* the angle ϕ is then greater than 90°, its cosine is negative, and the integral in Eq. (22.7) is negative. But since q is also negative, Eq. (22.7) holds.

For a closed surface enclosing *no* charge,

$$\Phi_E = \oint \vec{E} \cdot d\vec{A} = 0$$

This is the mathematical statement that when a region contains no charge, any field lines caused by charges *outside* the region that enter on one side must leave again on the other side. (In Section 22.1 we came to the same conclusion by considering the special case of a rectangular box in a uniform field.) **Figure 22.13** illustrates this point. *Electric field lines can begin or end inside a region of space only when there is charge in that region.*

22.13 A point charge *outside* a closed surface that encloses no charge. If an electric field line from the external charge enters the surface at one point, it must leave at another.



General Form of Gauss's Law

Now comes the final step in obtaining the general form of Gauss's law. Suppose the surface encloses not just one point charge q but several charges q_1, q_2, q_3, ... The total (resultant) electric field $\vec{E}$ at any point is the vector sum of the $\vec{E}$ fields of the individual charges. Let Q_{encl} be the *total* charge enclosed by the surface: $Q_{\text{encl}} = q_1 + q_2 + q_3 + \cdots$. Also let $\vec{E}$ be the *total* field at the position of the surface area element $d\vec{A}$, and let $E_\perp$ be its component perpendicular to the plane of that element (that is, parallel to $d\vec{A}$). Then we can write an equation like Eq. (22.7) for each charge and its corresponding field and add the results. When we do, we obtain the general statement of Gauss's law:

Gauss's law:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A} = \frac{Q_{\text{encl}}}{\epsilon_0} \tag{22.8}$$

Total charge enclosed by surface

Electric flux through a closed surface of area A = surface integral of $\vec{E}$

Electric constant

The total electric flux through a closed surface is equal to the total (net) electric charge inside the surface, divided by ϵ_0.

Using the definition of Q_{encl} and the various ways to express electric flux given in Eq. (22.5), we can express Gauss's law in the following equivalent forms:

Various forms of Gauss's law:

Magnitude of electric field $\vec{E}$ | Component of $\vec{E}$ perpendicular to surface | Total charge enclosed by surface

$$\Phi_E = \oint E\cos\phi \, dA = \oint E_\perp \, dA = \oint \vec{E} \cdot d\vec{A} = \frac{Q_{\text{encl}}}{\epsilon_0} \tag{22.9}$$

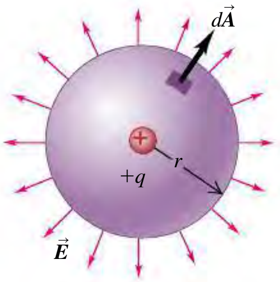
Electric flux through a closed surface | Angle between $\vec{E}$ and normal to surface | Element of surface area | Vector element of surface area | Electric constant

CAUTION Gaussian surfaces are imaginary Remember that the closed surface in Gauss's law is *imaginary;* there need not be any material object at the position of the surface. We often refer to a closed surface used in Gauss's law as a **Gaussian surface.**

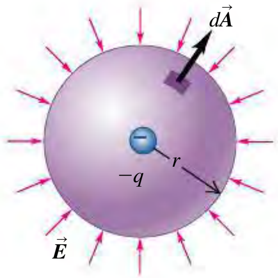
As in Eq. (22.5), the various forms of the integral all express the same thing, the total electric flux through the Gaussian surface, in different terms. One form is sometimes more convenient than another.

22.14 Spherical Gaussian surfaces around (a) a positive point charge and (b) a negative point charge.

(a) Gaussian surface around positive charge: positive (outward) flux



(b) Gaussian surface around negative charge: negative (inward) flux



As an example, **Fig. 22.14a** shows a spherical Gaussian surface of radius r around a positive point charge $+q$. The electric field points out of the Gaussian surface, so at every point on the surface $\vec{E}$ is in the same direction as $d\vec{A}$, $\phi = 0$, and $E_\perp$ is equal to the field magnitude $E = q/4\pi\epsilon_0 r^2$. Since E is the same at all points on the surface, we can take it outside the integral in Eq. (22.9). Then the remaining integral is $\int dA = A = 4\pi r^2$, the area of the sphere. Hence Eq. (22.9) becomes

$$\Phi_E = \oint E_\perp \, dA = \oint \left(\frac{q}{4\pi\epsilon_0 r^2} \right) dA = \frac{q}{4\pi\epsilon_0 r^2} \oint dA = \frac{q}{4\pi\epsilon_0 r^2} 4\pi r^2 = \frac{q}{\epsilon_0}$$

The enclosed charge Q_{encl} is just the charge $+q$, so this agrees with Gauss's law. If the Gaussian surface encloses a *negative* point charge as in Fig. 22.14b, then $\vec{E}$ points *into* the surface at each point in the direction opposite $d\vec{A}$. Then $\phi = 180°$ and $E_\perp$ is equal to the negative of the field magnitude: $E_\perp = -E = -|-q|/4\pi\epsilon_0 r^2 = -q/4\pi\epsilon_0 r^2$. Equation (22.9) then becomes

$$\Phi_E = \oint E_\perp \, dA = \oint \left(\frac{-q}{4\pi\epsilon_0 r^2} \right) dA = \frac{-q}{4\pi\epsilon_0 r^2} \oint dA = \frac{-q}{4\pi\epsilon_0 r^2} 4\pi r^2 = \frac{-q}{\epsilon_0}$$

This again agrees with Gauss's law because the enclosed charge in Fig. 22.14b is $Q_{\text{encl}} = -q$.

In Eqs. (22.8) and (22.9), Q_{encl} is always the algebraic sum of all the positive and negative charges enclosed by the Gaussian surface, and $\vec{E}$ is the *total* field at each point on the surface. Also note that in general, this field is caused partly by charges inside the surface and partly by charges outside. But as Fig. 22.13 shows, the outside charges do *not* contribute to the total (net) flux through the surface. So Eqs. (22.8) and (22.9) are correct even when there are charges outside the surface that contribute to the electric field at the surface. When $Q_{\text{encl}} = 0$, the total flux through the Gaussian surface must be zero, even though some areas may have positive flux and others may have negative flux (see Fig. 22.3b).

Gauss's law is the definitive answer to the question we posed at the beginning of Section 22.1: "If the electric-field pattern is known in a given region, what can we determine about the charge distribution in that region?" It provides a relationship between the electric field on a closed surface and the charge distribution within that surface. But in some cases we can use Gauss's law to answer the reverse question: "If the charge distribution is known, what can we determine about the electric field that the charge distribution produces?" Gauss's law may seem like an unappealing way to address this question, since it may look as though evaluating the integral in Eq. (22.8) is a hopeless task. Sometimes it is, but other times it is surprisingly easy. Here's an example in which *no* integration is involved at all; we'll work out several more examples in the next section.

CONCEPTUAL EXAMPLE 22.4 ELECTRIC FLUX AND ENCLOSED CHARGE

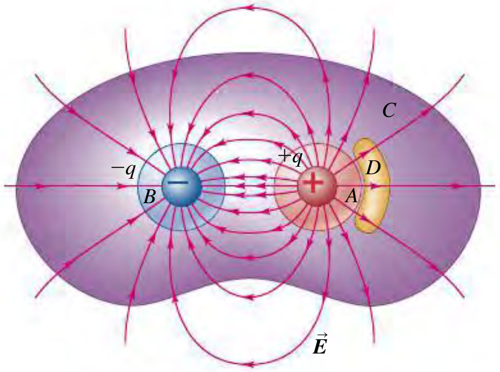
SOLUTION

Figure 22.15 shows the field produced by two point charges $+q$ and $-q$ (an electric dipole). Find the electric flux through each of the closed surfaces A, B, C, and D.

SOLUTION

Gauss's law, Eq. (22.8), says that the total electric flux through a closed surface is equal to the total enclosed charge divided by ϵ_0. In Fig. 22.15, surface A (shown in red) encloses the positive charge, so $Q_{\text{encl}} = +q$; surface B (in blue) encloses the negative charge, so $Q_{\text{encl}} = -q$; surface C (in purple) encloses *both* charges, so $Q_{\text{encl}} = +q + (-q) = 0$; and surface D (in yellow) encloses no charges, so $Q_{\text{encl}} = 0$. Hence, without having to do any integration, we have $\Phi_{EA} = +q/\epsilon_0$, $\Phi_{EB} = -q/\epsilon_0$, and $\Phi_{EC} = \Phi_{ED} = 0$. These results depend only on the charges enclosed within each Gaussian surface, not on the precise shapes of the surfaces.

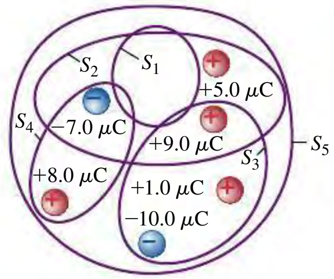
22.15 The net number of field lines leaving a closed surface is proportional to the total charge enclosed by that surface.



We can draw similar conclusions by examining the electric field lines. All the field lines that cross surface A are directed out of the surface, so the flux through A must be positive. Similarly, the flux through B must be negative since all of the field lines that cross that surface point inward. For both surface C and surface D, there are as many field lines pointing into the surface as there are field lines pointing outward, so the flux through each of these surfaces is zero.

TEST YOUR UNDERSTANDING OF SECTION 22.3 **Figure 22.16** shows six point charges that all lie in the same plane. Five Gaussian surfaces—S_1, S_2, S_3, S_4, and S_5—each enclose part of this plane, and Fig. 22.16 shows the intersection of each surface with the plane. Rank these five surfaces in order of the electric flux through them, from most positive to most negative. ❙

22.16 Five Gaussian surfaces and six point charges.



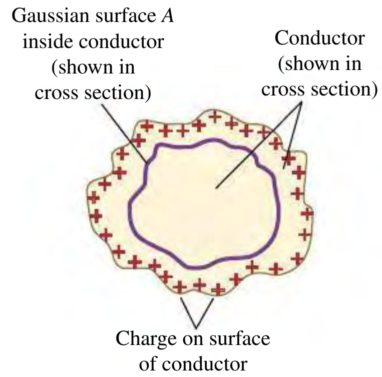
22.4 APPLICATIONS OF GAUSS'S LAW

Gauss's law is valid for *any* distribution of charges and for *any* closed surface. Gauss's law can be used in two ways. If we know the charge distribution, and if it has enough symmetry to let us evaluate the integral in Gauss's law, we can find the field. Or if we know the field, we can use Gauss's law to find the charge distribution, such as charges on conducting surfaces.

In this section we present examples of both kinds of applications. As you study them, watch for the role played by the symmetry properties of each system. We will use Gauss's law to calculate the electric fields caused by several simple charge distributions; the results are collected in a table in the chapter summary.

In practical problems we often encounter situations in which we want to know the electric field caused by a charge distribution on a conductor. These calculations are aided by the following remarkable fact: *When excess charge is placed on a solid conductor and is at rest, it resides entirely on the surface, not in the interior of the material.* (By *excess* we mean charges other than the ions and free electrons that make up the neutral conductor.) Here's the proof. We know from Section 21.4 that in an electrostatic situation (with all charges at rest) the electric field $\vec{E}$ at every point in the interior of a conducting material is zero. If $\vec{E}$ were *not* zero, the excess charges would move. Suppose we construct a Gaussian surface inside the conductor, such as surface A in **Fig. 22.17**. Because $\vec{E} = \mathbf{0}$ everywhere on this surface, Gauss's law requires that the net charge inside the surface is zero. Now imagine shrinking the surface like a collapsing balloon until it encloses a region so small that we may consider it as a point P; then the charge at that point must be zero. We can do this anywhere inside the conductor, so *there can be no excess charge at any point within a solid conductor; any excess charge must reside on the conductor's surface.* (This result is for a *solid* conductor. In the next section we'll discuss what can happen if the conductor has cavities in its interior.) We will make use of this fact frequently in the examples that follow.

22.17 Under electrostatic conditions (charges not in motion), any excess charge on a solid conductor resides entirely on the conductor's surface.



PROBLEM-SOLVING STRATEGY 22.1 GAUSS'S LAW

IDENTIFY *the relevant concepts:* Gauss's law is most useful when the charge distribution has spherical, cylindrical, or planar symmetry. In these cases the symmetry determines the direction of $\vec{E}$. Then Gauss's law yields the magnitude of $\vec{E}$ if we are given the charge distribution, and vice versa. In either case, begin the analysis by asking the question: What is the symmetry?

SET UP *the problem* using the following steps:

1. List the known and unknown quantities and identify the target variable.
2. Select the appropriate closed, imaginary Gaussian surface. For spherical symmetry, use a concentric spherical surface. For cylindrical symmetry, use a coaxial cylindrical surface with flat ends perpendicular to the axis of symmetry (like a soup can). For planar symmetry, use a cylindrical surface (like a tuna can) with its flat ends parallel to the plane.

EXECUTE *the solution* as follows:

1. Determine the appropriate size and placement of your Gaussian surface. To evaluate the field magnitude at a particular point, the surface must include that point. It may help to place one end of a can-shaped surface within a conductor, where $\vec{E}$ and therefore Φ are zero, or to place its ends equidistant from a charged plane.
2. Evaluate the integral $\oint E_\perp \, dA$ in Eq. (22.9). In this equation $E_\perp$ is the perpendicular component of the *total* electric field at each point on the Gaussian surface. A well-chosen Gaussian surface should make integration trivial or unnecessary. If the surface comprises several separate surfaces, such as the sides and ends of a cylinder, the integral $\oint E_\perp \, dA$ over the entire closed surface is the sum of the integrals $\int E_\perp \, dA$ over the separate surfaces. Consider points 3–6 as you work.
3. If $\vec{E}$ is *perpendicular* (normal) at every point to a surface with area A, if it points *outward* from the interior of the surface, and if it has the same *magnitude* at every point on the surface, then $E_\perp = E =$ constant, and $\int E_\perp \, dA$ over that surface is equal to EA. (If $\vec{E}$ is inward, then $E_\perp = -E$ and $\int E_\perp \, dA = -EA$.) This should be the case for part or all of your Gaussian surface. If $\vec{E}$ is tangent to a surface at every point, then $E_\perp = 0$ and the integral over that surface is zero. This may be the case for parts of a cylindrical Gaussian surface. If $\vec{E} = \mathbf{0}$ at every point on a surface, the integral is zero.
4. Even when there is *no* charge within a Gaussian surface, the field at any given point on the surface is not necessarily zero. In that case, however, the total electric flux through the surface is always zero.
5. The flux integral $\oint E_\perp \, dA$ can be approximated as the difference between the numbers of electric lines of force leaving and entering the Gaussian surface. In this sense the flux gives the sign of the enclosed charge, but is only proportional to it; zero flux corresponds to zero enclosed charge.
6. Once you have evaluated $\oint E_\perp \, dA$, use Eq. (22.9) to solve for your target variable.

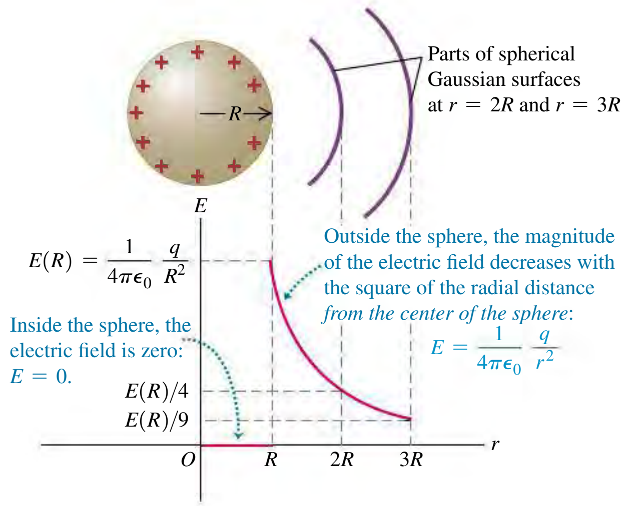
EVALUATE *your answer:* If your result is a *function* that describes how the magnitude of the electric field varies with position, ensure that it makes sense.

EXAMPLE 22.5 FIELD OF A CHARGED CONDUCTING SPHERE

SOLUTION

We place a total positive charge q on a solid conducting sphere with radius R (**Fig. 22.18**). Find $\vec{E}$ at any point inside or outside the sphere.

22.18 Calculating the electric field of a conducting sphere with positive charge q. Outside the sphere, the field is the same as if all of the charge were concentrated at the center of the sphere.



SOLUTION

IDENTIFY and SET UP: As we discussed earlier in this section, all of the charge must be on the surface of the sphere. The charge is free to move on the conductor, and there is no preferred position on the surface; the charge is therefore distributed *uniformly* over the surface, and the system is spherically symmetric. To exploit this symmetry, we take as our Gaussian surface a sphere of radius r centered on the conductor. We can calculate the field inside or outside the conductor by taking $r < R$ or $r > R$, respectively. In either case, the point at which we want to calculate $\vec{E}$ lies on the Gaussian surface.

EXECUTE: The spherical symmetry means that the direction of the electric field must be radial; that's because there is no preferred direction parallel to the surface, so $\vec{E}$ can have no component parallel to the surface. There is also no preferred orientation of the sphere, so the field magnitude E can depend only on the distance r from the center and must have the same value at all points on the Gaussian surface.

For $r > R$ the entire conductor is within the Gaussian surface, so the enclosed charge is q. The area of the Gaussian surface is $4\pi r^2$, and $\vec{E}$ is uniform over the surface and perpendicular to it at each point. The flux integral $\oint E_\perp \, dA$ is then just $E(4\pi r^2)$, and Eq. (22.8) gives

$$E(4\pi r^2) = \frac{q}{\epsilon_0} \quad \text{and}$$

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \qquad \text{(outside a charged conducting sphere)}$$

This expression is the same as that for a point charge; outside the charged sphere, its field is the same as though the entire charge were concentrated at its center. Just outside the surface of the sphere, where $r = R$,

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{R^2} \quad \text{(at the surface of a charged conducting sphere)}$$

CAUTION Flux can be positive or negative Remember that we have chosen the charge q to be *positive*. If the charge is negative, the electric field is radially *inward* instead of radially outward, and the electric flux through the Gaussian surface is negative. The electric-field magnitudes outside and at the surface of the sphere are given by the same expressions as above, except that q denotes the *magnitude* (absolute value) of the charge. ▌

For $r < R$ we again have $E(4\pi r^2) = Q_{\text{encl}}/\epsilon_0$. But now our Gaussian surface (which lies entirely within the conductor) encloses *no* charge, so $Q_{\text{encl}} = 0$. The electric field inside the conductor is therefore zero.

EVALUATE: We already knew that $\vec{\boldsymbol{E}} = \mathbf{0}$ inside a solid conductor (whether spherical or not) when the charges are at rest. Figure 22.18 shows E as a function of the distance r from the center of the sphere. Note that in the limit as $R \rightarrow 0$, the sphere becomes a point charge; there is then only an "outside," and the field is everywhere given by $E = q/4\pi\epsilon_0 r^2$. Thus we have deduced Coulomb's law from Gauss's law. (In Section 22.3 we deduced Gauss's law from Coulomb's law; the two laws are equivalent.)

We can also use this method for a conducting spherical *shell* (a spherical conductor with a concentric spherical hole inside) if there is no charge inside the hole. We use a spherical Gaussian surface with radius r less than the radius of the hole. If there *were* a field inside the hole, it would have to be radial and spherically symmetric as before, so $E = Q_{\text{encl}}/4\pi\epsilon_0 r^2$. But now there is no enclosed charge, so $Q_{\text{encl}} = 0$ and $E = 0$ inside the hole.

Can you use this same technique to find the electric field in the region between a charged sphere and a concentric hollow conducting sphere that surrounds it?

EXAMPLE 22.6 FIELD OF A UNIFORM LINE CHARGE

SOLUTION

Electric charge is distributed uniformly along an infinitely long, thin wire. The charge per unit length is λ (assumed positive). Find the electric field by using Gauss's law.

SOLUTION

IDENTIFY and SET UP: We found in Example 21.10 (Section 21.5) that the field $\vec{\boldsymbol{E}}$ of a uniformly charged, infinite wire is radially outward if λ is positive and radially inward if λ is negative, and that the field magnitude E depends on only the radial distance from the wire. This suggests that we use a *cylindrical* Gaussian surface, of radius r and arbitrary length l, coaxial with the wire and with its ends perpendicular to the wire (**Fig. 22.19**).

EXECUTE: The flux through the flat ends of our Gaussian surface is zero because the radial electric field is parallel to these ends, and so $\vec{\boldsymbol{E}} \cdot \hat{\boldsymbol{n}} = 0$. On the cylindrical part of our surface we have $\vec{\boldsymbol{E}} \cdot \hat{\boldsymbol{n}} = E_\perp = E$ everywhere. (If λ were negative, we would have $\vec{\boldsymbol{E}} \cdot \hat{\boldsymbol{n}} = E_\perp = -E$ everywhere.) The area of the cylindrical surface is $2\pi rl$, so the flux through it—and hence the total flux Φ_E through the Gaussian surface—is $EA = 2\pi rlE$. The total enclosed charge is $Q_{\text{encl}} = \lambda l$, and so from Gauss's law, Eq. (22.8),

$$\Phi_E = 2\pi rlE = \frac{\lambda l}{\epsilon_0} \quad \text{and}$$

$$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r} \qquad \text{(field of an infinite line of charge)}$$

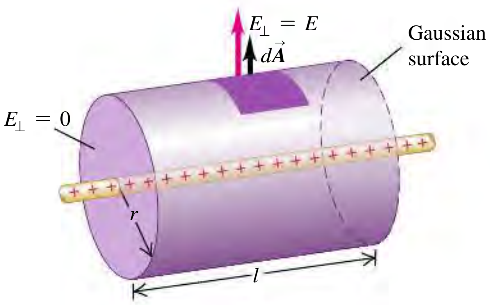
We found this same result in Example 21.10 with *much* more effort.

If λ is *negative*, $\vec{\boldsymbol{E}}$ is directed radially inward, and in the above expression for E we must interpret λ as the absolute value of the charge per unit length.

EVALUATE: We saw in Example 21.10 that the *entire* charge on the wire contributes to the field at any point, and yet we consider only that part of the charge $Q_{\text{encl}} = \lambda l$ within the Gaussian surface when we apply Gauss's law. There's nothing inconsistent here; it takes the entire charge to give the field the properties that allow us to calculate Φ_E so easily, and Gauss's law always applies to the enclosed charge only. If the wire is short, the symmetry of the infinite wire is lost, and E is not uniform over a coaxial, cylindrical Gaussian surface. Gauss's law then *cannot* be used to find Φ; we must solve the problem the hard way, as in Example 21.10.

We can use the Gaussian surface in Fig. 22.19 to show that the field outside a long, uniformly charged cylinder is the same as though all the charge were concentrated on a line along its axis (see Problem 22.41). We can also calculate the electric field in the space between a charged cylinder and a coaxial hollow conducting cylinder surrounding it (see Problem 22.39).

22.19 A coaxial cylindrical Gaussian surface is used to find the electric field outside an infinitely long, charged wire.



EXAMPLE 22.7 FIELD OF AN INFINITE PLANE SHEET OF CHARGE

SOLUTION

Use Gauss's law to find the electric field caused by a thin, flat, infinite sheet with a uniform positive surface charge density σ.

SOLUTION

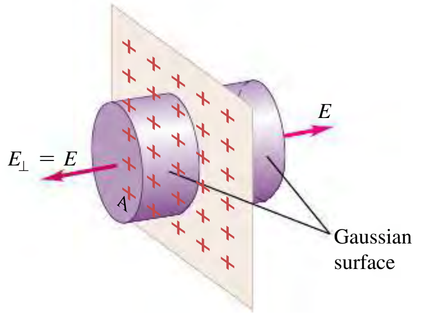
IDENTIFY and SET UP: In Example 21.11 (Section 21.5) we found that the field $\vec{E}$ of a uniformly charged infinite sheet is normal to the sheet, and that its magnitude is independent of the distance from the sheet. To take advantage of these symmetry properties, we use a cylindrical Gaussian surface with ends of area A and with its axis perpendicular to the sheet of charge (**Fig. 22.20**).

EXECUTE: The flux through the cylindrical part of our Gaussian surface is zero because $\vec{E}\cdot\hat{n} = 0$ everywhere. The flux through each flat end of the surface is $+EA$ because $\vec{E}\cdot\hat{n} = E_\perp = E$ everywhere, so the total flux through both ends—and hence the total flux Φ_E through the Gaussian surface—is $+2EA$. The total enclosed charge is $Q_{\text{encl}} = \sigma A$, and so from Gauss's law,

$$2EA = \frac{\sigma A}{\epsilon_0} \quad \text{and}$$

$$E = \frac{\sigma}{2\epsilon_0} \quad \text{(field of an infinite sheet of charge)}$$

22.20 A cylindrical Gaussian surface is used to find the field of an infinite plane sheet of charge.



If σ is negative, $\vec{E}$ is directed *toward* the sheet, the flux through the Gaussian surface in Fig. 22.20 is negative, and σ in the expression $E = \sigma/2\epsilon_0$ denotes the magnitude (absolute value) of the charge density.

EVALUATE: We got the same result for the field of an infinite sheet of charge in Example 21.11 (Section 21.5). That calculation was much more complex and involved a fairly challenging integral. Thanks to the favorable symmetry, Gauss's law makes it much easier to solve this problem.

EXAMPLE 22.8 FIELD BETWEEN OPPOSITELY CHARGED PARALLEL CONDUCTING PLATES

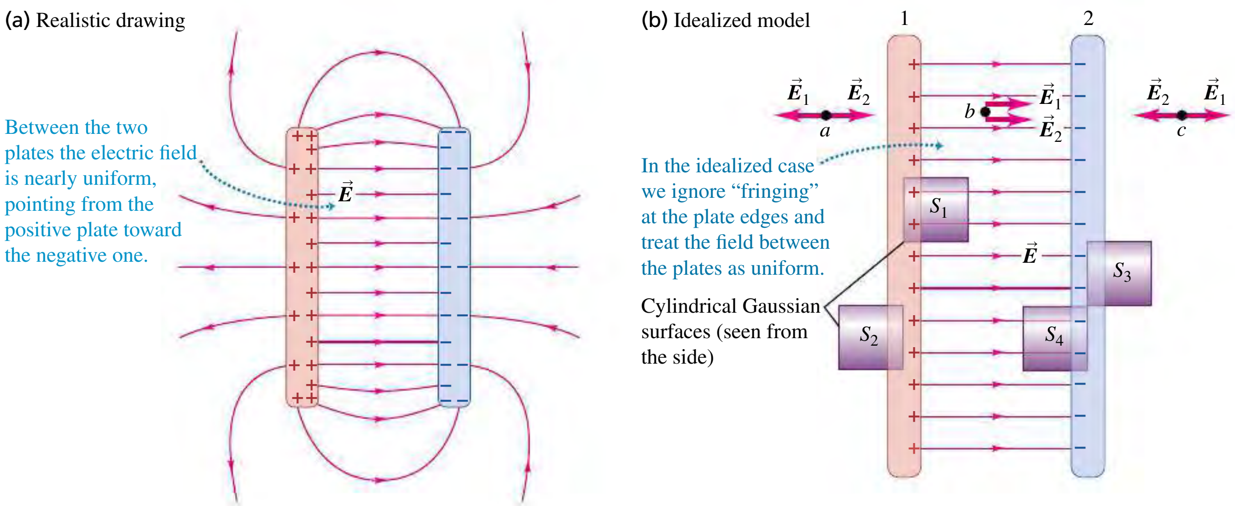
SOLUTION

Two large plane parallel conducting plates are given charges of equal magnitude and opposite sign; the surface charge densities are $+\sigma$ and $-\sigma$. Find the electric field in the region between the plates.

SOLUTION

IDENTIFY and SET UP: Figure 22.21a shows the field. Because opposite charges attract, most of the charge accumulates at the opposing faces of the plates. A small amount of charge resides on the *outer* surfaces of the plates, and there is some spreading or "fringing" of the field at the edges. But if the plates are very large in comparison to the distance between them, the amount of charge on the outer surfaces is negligibly small, and the fringing can be ignored except near the edges. In this case we can assume that the field is uniform in the interior region between the plates, as in Fig. 22.21b, and that the charges are distributed uniformly over the opposing surfaces. To exploit this symmetry, we can use the shaded Gaussian surfaces S_1, S_2, S_3, and S_4. These surfaces are cylinders with flat ends of area A; one end of each surface lies *within* a plate.

22.21 Electric field between oppositely charged parallel plates.



EXECUTE: The left-hand end of surface S_1 is within the positive plate 1. Since the field is zero within the volume of any solid conductor under electrostatic conditions, there is no electric flux through this end. The electric field between the plates is perpendicular to the right-hand end, so on that end, $E_\perp$ is equal to E and the flux is EA; this is positive, since $\vec{E}$ is directed out of the Gaussian surface. There is no flux through the side walls of the cylinder, since these walls are parallel to $\vec{E}$. So the total flux integral in Gauss's law is EA. The net charge enclosed by the cylinder is σA, so Eq. (22.8) yields $EA = \sigma A/\epsilon_0$; we then have

$$E = \frac{\sigma}{\epsilon_0} \quad \text{(field between oppositely charged conducting plates)}$$

The field is uniform and perpendicular to the plates, and its magnitude is independent of the distance from either plate. The Gaussian surface S_4 yields the same result. Surfaces S_2 and S_3 yield $E = 0$ to the left of plate 1 and to the right of plate 2, respectively. We leave these calculations to you (see Exercise 22.27).

EVALUATE: We obtained the same results in Example 21.12 by using the principle of superposition of electric fields. The fields due to the two sheets of charge (one on each plate) are $\vec{E}_1$ and $\vec{E}_2$; from Example 22.7, both of these have magnitude $\sigma/2\epsilon_0$. The total electric field at any point is the vector sum $\vec{E} = \vec{E}_1 + \vec{E}_2$. At points a and c in Fig. 22.21b, $\vec{E}_1$ and $\vec{E}_2$ point in opposite directions, and their sum is zero. At point b, $\vec{E}_1$ and $\vec{E}_2$ are in the same direction; their sum has magnitude $E = \sigma/\epsilon_0$, just as we found by using Gauss's law.

EXAMPLE 22.9 FIELD OF A UNIFORMLY CHARGED SPHERE

Positive electric charge Q is distributed uniformly *throughout the volume* of an *insulating* sphere with radius R. Find the magnitude of the electric field at a point P a distance r from the center of the sphere.

SOLUTION

IDENTIFY and SET UP: As in Example 22.5, the system is spherically symmetric. Hence we can use the conclusions of that example about the direction and magnitude of $\vec{E}$. To make use of the spherical symmetry, we choose as our Gaussian surface a sphere with radius r, concentric with the charge distribution.

EXECUTE: From symmetry, the direction of $\vec{E}$ is radial at every point on the Gaussian surface, so $E_\perp = E$ and the field magnitude E is the same at every point on the surface. Hence the total electric flux through the Gaussian surface is the product of E and the total area of the surface $A = 4\pi r^2$—that is, $\Phi_E = 4\pi r^2 E$.

The amount of charge enclosed within the Gaussian surface depends on r. To find E *inside* the sphere, we choose $r < R$. The volume charge density ρ is the charge Q divided by the volume of the entire charged sphere of radius R:

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

The volume V_{encl} enclosed by the Gaussian surface is $\frac{4}{3}\pi r^3$, so the total charge Q_{encl} enclosed by that surface is

$$Q_{\text{encl}} = \rho V_{\text{encl}} = \left(\frac{Q}{4\pi R^3/3}\right)\left(\tfrac{4}{3}\pi r^3\right) = Q\frac{r^3}{R^3}$$

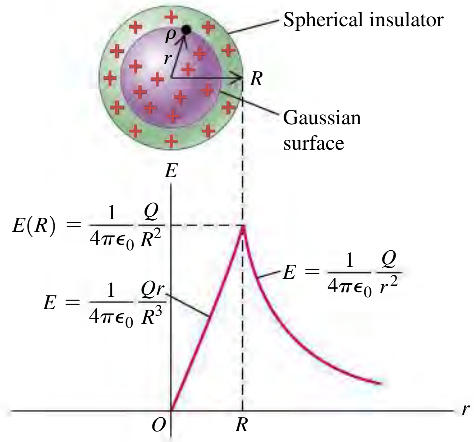
Then Gauss's law, Eq. (22.8), becomes

$$4\pi r^2 E = \frac{Q}{\epsilon_0}\frac{r^3}{R^3} \quad \text{or}$$

$$E = \frac{1}{4\pi\epsilon_0}\frac{Qr}{R^3} \quad \text{(field inside a uniformly charged sphere)}$$

The field magnitude is proportional to the distance r of the field point from the center of the sphere (see the graph of E versus r in **Fig. 22.22**).

22.22 The magnitude of the electric field of a uniformly charged insulating sphere. Compare this with the field for a conducting sphere (see Fig. 22.18).



To find E *outside* the sphere, we take $r > R$. This surface encloses the entire charged sphere, so $Q_{\text{encl}} = Q$, and Gauss's law gives

$$4\pi r^2 E = \frac{Q}{\epsilon_0} \quad \text{or}$$

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2} \quad \text{(field outside a uniformly charged sphere)}$$

The field outside *any* spherically symmetric charged body varies as $1/r^2$, as though the entire charge were concentrated at the center. This is graphed in Fig. 22.22.

If the charge is *negative*, $\vec{E}$ is radially *inward* and in the expressions for E we interpret Q as the absolute value of the charge.

EVALUATE: Notice that if we set $r = R$ in either expression for E, we get the same result $E = Q/4\pi\epsilon_0 R^2$ for the magnitude of the field at the surface of the sphere. This is because the magnitude E is a *continuous* function of r. By contrast, for the charged conducting sphere of Example 22.5 the electric-field magnitude is *discontinuous* at $r = R$ (it jumps from $E = 0$ just inside the

Continued

sphere to $E = Q/4\pi\epsilon_0 R^2$ just outside the sphere). In general, the electric field $\vec{E}$ is discontinuous in magnitude, direction, or both wherever there is a *sheet* of charge, such as at the surface of a charged conducting sphere (Example 22.5), at the surface of an infinite charged sheet (Example 22.7), or at the surface of a charged conducting plate (Example 22.8).

The approach used here can be applied to *any* spherically symmetric distribution of charge, even if it is not radially uniform, as it was here. Such charge distributions occur within many atoms and atomic nuclei, so Gauss's law is useful in atomic and nuclear physics.

EXAMPLE 22.10 CHARGE ON A HOLLOW SPHERE

SOLUTION

A thin-walled, hollow sphere of radius 0.250 m has an unknown charge distributed uniformly over its surface. At a distance of 0.300 m from the center of the sphere, the electric field points radially inward and has magnitude 1.80×10^2 N/C. How much charge is on the sphere?

SOLUTION

IDENTIFY and SET UP: The charge distribution is spherically symmetric. As in Examples 22.5 and 22.9, it follows that the electric field is radial everywhere and its magnitude is a function of only the radial distance r from the center of the sphere. We use a spherical Gaussian surface that is concentric with the charge distribution and has radius $r = 0.300$ m. Our target variable is $Q_{\text{encl}} = q$.

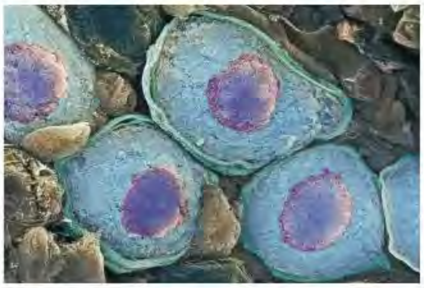
EXECUTE: The charge distribution is the same as if the charge were on the surface of a 0.250-m-radius conducting sphere. Hence we can borrow the results of Example 22.5. We note that the electric field here is directed toward the sphere, so that q must be *negative*. Furthermore, the electric field is directed into the Gaussian surface, so that $E_\perp = -E$ and $\Phi_E = \oint E_\perp \, dA = -E(4\pi r^2)$.

By Gauss's law, the flux is equal to the charge q on the sphere (all of which is enclosed by the Gaussian surface) divided by ϵ_0. Solving for q, we find

$$\begin{aligned} q = -E(4\pi\epsilon_0 r^2) &= -(1.80 \times 10^2 \text{ N/C})(4\pi) \\ &\quad \times (8.854 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)(0.300 \text{ m})^2 \\ &= -1.80 \times 10^{-9} \text{ C} = -1.80 \text{ nC} \end{aligned}$$

EVALUATE: To determine the charge, we had to know the electric field at *all* points on the Gaussian surface so that we could calculate the flux integral. This was possible here because the charge distribution is highly symmetric. If the charge distribution is irregular or lacks symmetry, Gauss's law is not very useful for calculating the charge distribution from the field, or vice versa.

BIO Application Charge Distribution Inside a Nerve Cell The interior of a human nerve cell contains both positive potassium ions (K^+) and negatively charged protein molecules (Pr^-). Potassium ions can flow out of the cell through the cell membrane, but the much larger protein molecules cannot. The result is that the interior of the cell has a net negative charge. (The fluid outside the cell has a positive charge that balances this.) The fluid within the cell is a good conductor, so the Pr^- molecules distribute themselves on the outer surface of the fluid—that is, on the inner surface of the cell membrane, which is an insulator. This is true no matter what the shape of the cell.



DEMO

TEST YOUR UNDERSTANDING OF SECTION 22.4 You place a known amount of charge Q on the irregularly shaped conductor shown in Fig. 22.17. If you know the size and shape of the conductor, can you use Gauss's law to calculate the electric field at an arbitrary position outside the conductor? ❙

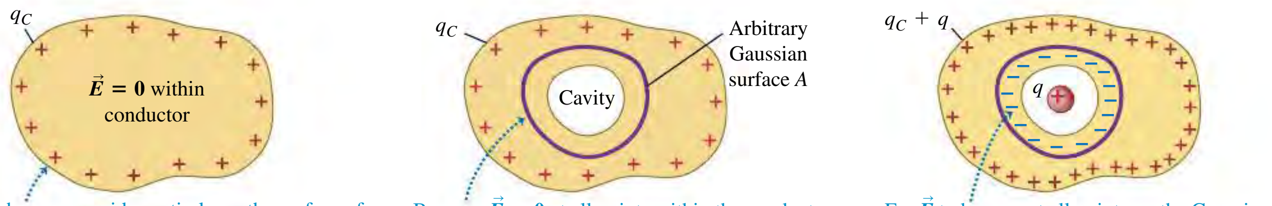
22.5 CHARGES ON CONDUCTORS

We have learned that in an electrostatic situation (in which there is no net motion of charge) the electric field at every point within a conductor is zero and any excess charge on a solid conductor is located entirely on its surface (**Fig. 22.23a**). But what if there is a *cavity* inside the conductor (Fig. 22.23b)? If there is no charge within the cavity, we can use a Gaussian surface such as A (which lies completely within the material of the conductor) to show that the *net* charge on the *surface of the cavity* must be zero, because $\vec{E} = \mathbf{0}$ everywhere on the Gaussian surface. In fact, we can prove in this situation that there can't be any charge *anywhere* on the cavity surface. We will postpone detailed proof of this statement until Chapter 23.

Suppose we place a small body with a charge q inside a cavity within a conductor (Fig. 22.23c). The conductor is uncharged and is insulated from the charge q. Again $\vec{E} = \mathbf{0}$ everywhere on surface A, so according to Gauss's law the *total* charge inside this surface must be zero. Therefore there must be a charge $-q$ distributed on the surface of the cavity, drawn there by the charge q inside the cavity. The *total* charge on the conductor must remain zero, so a charge $+q$ must appear either on its outer surface or inside the material. But we showed that in an electrostatic situation there can't be any excess charge within the material of a conductor. So we conclude that charge $+q$ must appear on the outer surface. By the same reasoning, if the conductor originally had a charge q_C, then the total charge on the outer surface must be $q_C + q$ after charge q is inserted into the cavity.

22.23 Finding the electric field within a charged conductor.

(a) Solid conductor with charge q_C (b) The same conductor with an internal cavity (c) An isolated charge q placed in the cavity



The charge q_C resides entirely on the surface of the conductor. The situation is electrostatic, so $\vec{E} = \mathbf{0}$ within the conductor.

Because $\vec{E} = \mathbf{0}$ at all points within the conductor, the electric field at all points on the Gaussian surface must be zero.

For $\vec{E}$ to be zero at all points on the Gaussian surface, the surface of the cavity must have a total charge $-q$.

CONCEPTUAL EXAMPLE 22.11 A CONDUCTOR WITH A CAVITY

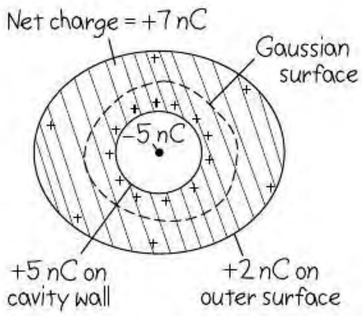
SOLUTION

A conductor with a cavity carries a total charge of +7 nC. Within the cavity, insulated from the conductor, is a point charge of −5 nC. How much charge is on each surface (inner and outer) of the conductor?

SOLUTION

Figure 22.24 shows the situation. If the charge in the cavity is $q = -5$ nC, the charge on the inner cavity surface must be $-q = -(-5\text{ nC}) = +5$ nC. The conductor carries a *total* charge of +7 nC, none of which is in the interior of the material. If +5 nC is on the inner surface of the cavity, then there must be $(+7\text{ nC}) - (+5\text{ nC}) = +2$ nC on the outer surface of the conductor.

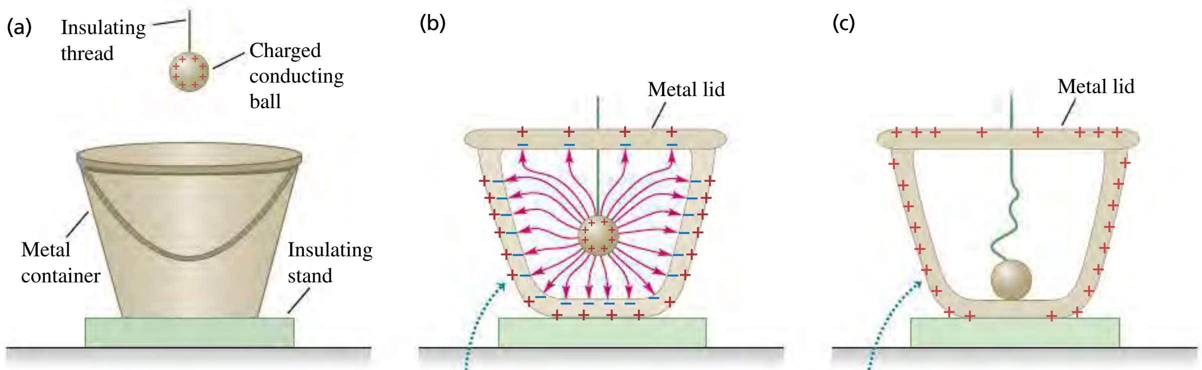
22.24 Our sketch for this problem. There is zero electric field inside the bulk conductor and hence zero flux through the Gaussian surface shown, so the charge on the cavity wall must be the opposite of the point charge.



Testing Gauss's Law Experimentally

We can now consider a historic experiment, shown in **Fig. 22.25**. We mount a conducting container on an insulating stand. The container is initially uncharged. Then we hang a charged metal ball from an insulating thread (Fig. 22.25a), lower it into the container, and put the lid on (Fig. 22.25b). Charges are induced on the walls of the container, as shown. But now we let the ball *touch* the inner wall (Fig. 22.25c).

22.25 (a) A charged conducting ball suspended by an insulating thread outside a conducting container on an insulating stand. (b) The ball is lowered into the container, and the lid is put on. (c) The ball is touched to the inner surface of the container.



Charged ball induces charges on the interior and exterior of the container.

Once the ball touches the container, it is part of the interior surface; all the charge moves to the container's exterior.

DATA *SPEAKS*

Electric Charges on Conductors

When students were given a problem involving electric charges on conductors, more than 41% gave an incorrect response. Common errors:

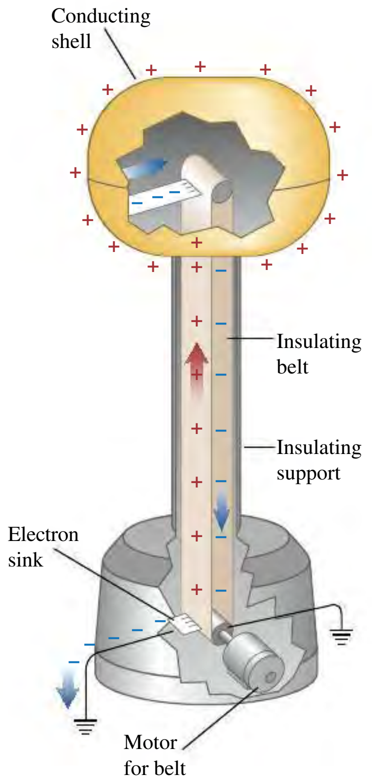
- Not understanding that charges outside a conductor have no effect on the interior of that conductor, even if the conductor has a cavity inside.
- Not understanding that charges inside a cavity within a conductor affect the charge distributions on both the cavity walls and the conductor's outer surface.

The surface of the ball becomes part of the cavity surface. The situation is now the same as Fig. 22.23b; if Gauss's law is correct, the net charge on the cavity surface must be zero. Thus the ball must lose all its charge. Finally, we pull the ball out; we find that it has indeed lost all its charge.

This experiment was performed in the 19th century by the English scientist Michael Faraday, using a metal icepail with a lid, and it is called **Faraday's icepail experiment.** The result confirms the validity of Gauss's law and therefore of Coulomb's law. Faraday's result was significant because Coulomb's experimental method, using a torsion balance and dividing of charges, was not very precise; it is very difficult to confirm the $1/r^2$ dependence of the electrostatic force by direct force measurements. By contrast, experiments like Faraday's test the validity of Gauss's law, and therefore of Coulomb's law, with much greater precision. Modern versions of this experiment have shown that the exponent 2 in the $1/r^2$ of Coulomb's law does not differ from precisely 2 by more than 10^{-16}. So there is no reason to believe it is anything other than exactly 2.

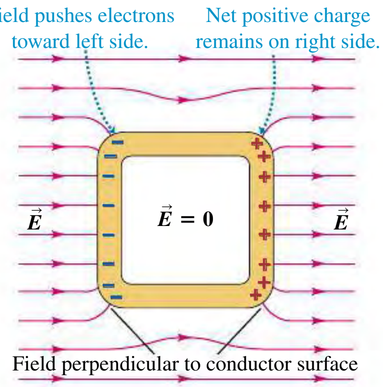
The same principle behind Faraday's icepail experiment is used in a *Van de Graaff electrostatic generator* (**Fig. 22.26**). A charged belt continuously produces a buildup of charge on the inside of a conducting shell. By Gauss's law, there can never be any charge on the inner surface of this shell, so the charge is immediately carried away to the outside surface of the shell. As a result, the charge on the shell and the electric field around it can become very large very rapidly. The Van de Graaff generator is used as an accelerator of charged particles and for physics demonstrations.

This principle also forms the basis for *electrostatic shielding*. Suppose we have a very sensitive electronic instrument that we want to protect from stray electric fields that might cause erroneous measurements. We surround the instrument with a conducting box, or we line the walls, floor, and ceiling of the room with a conducting material such as sheet copper. The external electric field redistributes the free electrons in the conductor, leaving a net positive



22.26 Cutaway view of the essential parts of a Van de Graaff electrostatic generator. The electron sink at the bottom draws electrons from the belt, giving it a positive charge; at the top the belt attracts electrons away from the conducting shell, giving the shell a positive charge.

22.27 (a) A conducting box (a Faraday cage) immersed in a uniform electric field. The field of the induced charges on the box combines with the uniform field to give zero total field inside the box. (b) This person is inside a Faraday cage, and so is protected from the powerful electric discharge.



charge on the outer surface in some regions and a net negative charge in others (**Fig. 22.27**). This charge distribution causes an additional electric field such that the *total* field at every point inside the box is zero, as Gauss's law says it must be. The charge distribution on the box also alters the shapes of the field lines near the box, as the figure shows. Such a setup is often called a *Faraday cage*. The same physics tells you that one of the safest places to be in a lightning storm is inside a car; if the car is struck by lightning, the charge tends to remain on the metal skin of the vehicle, and little or no electric field is produced inside the passenger compartment.

Field at the Surface of a Conductor

Finally, we note that there is a direct relationship between the $\vec{E}$ field at a point just outside any conductor and the surface charge density σ at that point. In general, σ varies from point to point on the surface. We will show in Chapter 23 that at any such point, the direction of $\vec{E}$ is always *perpendicular* to the surface. (You can see this effect in Fig. 22.27a.)

To find a relationship between σ at any point on the surface and the perpendicular component of the electric field at that point, we construct a Gaussian surface in the form of a small cylinder (**Fig. 22.28**). One end face, with area A, lies within the conductor and the other lies just outside. The electric field is zero at all points within the conductor. Outside the conductor the component of $\vec{E}$ perpendicular to the side walls of the cylinder is zero, and over the end face the perpendicular component is equal to $E_\perp$. (If σ is positive, the electric field points out of the conductor and $E_\perp$ is positive; if σ is negative, the field points inward and $E_\perp$ is negative.) Hence the total flux through the surface is $E_\perp A$. The charge enclosed within the Gaussian surface is σA, so from Gauss's law, $E_\perp A = (\sigma A)/\epsilon_0$ and

Electric field at surface of a conductor, $\vec{E}$ perpendicular to surface

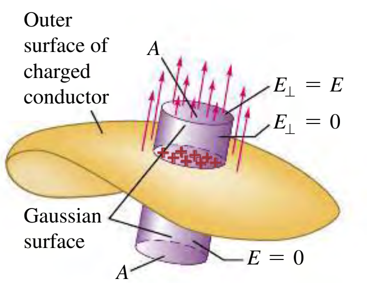
$$E_\perp = \frac{\sigma}{\epsilon_0} \qquad (22.10)$$

σ: Surface charge density; ϵ_0: Electric constant

This agrees with our result for the field at the surface of a charged conducting plate (Example 22.8); soon we'll verify it for a charged conducting sphere.

We showed in Example 22.8 that the field magnitude between two infinite flat oppositely charged conducting plates also equals σ/ϵ_0. In this case the field magnitude E is the same at *all* distances from the plates, but in all other cases E decreases with increasing distance from the surface.

22.28 The field just outside a charged conductor is perpendicular to the surface, and its perpendicular component $E_\perp$ is equal to σ/ϵ_0.



Application Why Lightning Bolts Are Vertical Our planet is a good conductor, and its surface has a negative charge. Hence, the electric field in the atmosphere above the surface points generally downward, toward the negative charge and perpendicular to the surface (see Example 22.13). The negative charge is balanced by positive charges in the atmosphere. In a lightning storm, the vertical electric field becomes great enough to cause charges to flow vertically through the air. The air is excited and ionized by the passage of charge through it, producing a visible lightning bolt.

CONCEPTUAL EXAMPLE 22.12 FIELD AT THE SURFACE OF A CONDUCTING SPHERE

SOLUTION

Verify Eq. (22.10) for a conducting sphere with radius R and total charge q.

SOLUTION

In Example 22.5 (Section 22.4) we showed that the electric field just outside the surface is

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{R^2}$$

The surface charge density is uniform and equal to q divided by the surface area of the sphere:

$$\sigma = \frac{q}{4\pi R^2}$$

Comparing these two expressions, we see that $E = \sigma/\epsilon_0$, which verifies Eq. (22.10).

EXAMPLE 22.13 ELECTRIC FIELD OF THE EARTH

SOLUTION

The earth (a conductor) has a net electric charge. The resulting electric field near the surface has an average value of about 150 N/C, directed toward the center of the earth. (a) What is the corresponding surface charge density? (b) What is the *total* surface charge of the earth?

SOLUTION

IDENTIFY and SET UP: We are given the electric-field magnitude at the surface of the conducting earth. We can calculate the surface charge density σ from Eq. (22.10). The total charge Q on the earth's surface is then the product of σ and the earth's surface area.

EXECUTE: (a) The direction of the field means that σ is negative (corresponding to $\vec{E}$ being directed *into* the surface, so $E_\perp$ is negative). From Eq. (22.10),

$$\sigma = \epsilon_0 E_\perp = (8.85 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2})(-150\ \mathrm{N/C})$$
$$= -1.33 \times 10^{-9}\ \mathrm{C/m^2} = -1.33\ \mathrm{nC/m^2}$$

(b) The earth's surface area is $4\pi R_E^2$, where $R_E = 6.38 \times 10^6$ m is the radius of the earth (see Appendix F). The total charge Q is the product $4\pi R_E^2 \sigma$, or

$$Q = 4\pi(6.38 \times 10^6\ \mathrm{m})^2(-1.33 \times 10^{-9}\ \mathrm{C/m^2})$$
$$= -6.8 \times 10^5\ \mathrm{C} = -680\ \mathrm{kC}$$

EVALUATE: You can check our result in part (b) by using the result of Example 22.5. Solving for Q, we find

$$Q = 4\pi\epsilon_0 R^2 E_\perp$$
$$= \frac{1}{9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2}}(6.38 \times 10^6\ \mathrm{m})^2(-150\ \mathrm{N/C})$$
$$= -6.8 \times 10^5\ \mathrm{C}$$

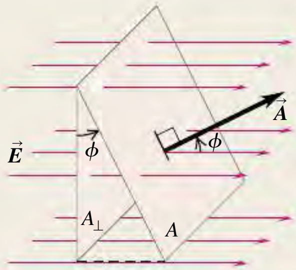
One electron has a charge of -1.60×10^{-19} C. Hence this much excess negative electric charge corresponds to there being $(-6.8 \times 10^5\ \mathrm{C})/(-1.60 \times 10^{-19}\ \mathrm{C}) = 4.2 \times 10^{24}$ excess electrons on the earth, or about 7 moles of excess electrons. This is compensated by an equal *deficiency* of electrons in the earth's upper atmosphere, so the combination of the earth and its atmosphere is electrically neutral.

TEST YOUR UNDERSTANDING OF SECTION 22.5 A hollow conducting sphere has no net charge. There is a positive point charge q at the center of the spherical cavity within the sphere. You connect a conducting wire from the outside of the sphere to ground. Will you measure an electric field outside the sphere?

SOLUTIONS TO ALL EXAMPLES

Electric flux: Electric flux is a measure of the "flow" of electric field through a surface. It is equal to the product of an area element and the perpendicular component of $\vec{E}$, integrated over a surface. (See Examples 22.1–22.3.)

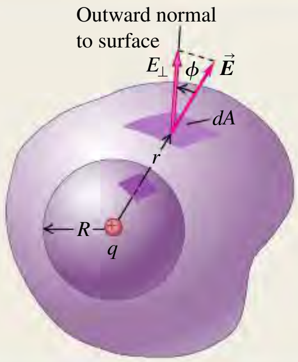
$$\Phi_E = \int E\cos\phi\, dA = \int E_\perp\, dA = \int \vec{E}\cdot d\vec{A} \qquad (22.5)$$



Gauss's law: Gauss's law states that the total electric flux through a closed surface, which can be written as the surface integral of the component of $\vec{E}$ normal to the surface, equals a constant times the total charge Q_{encl} enclosed by the surface. Gauss's law is logically equivalent to Coulomb's law, but its use greatly simplifies problems with a high degree of symmetry. (See Examples 22.4–22.10.)

When excess charge is placed on a solid conductor and is at rest, it resides entirely on the surface, and $\vec{E} = \mathbf{0}$ everywhere in the material of the conductor. (See Examples 22.11–22.13.)

$$\Phi_E = \oint E\cos\phi\, dA = \oint E_\perp\, dA = \oint \vec{E}\cdot d\vec{A} = \frac{Q_{encl}}{\epsilon_0} \qquad (22.8), (22.9)$$



Electric field of various symmetric charge distributions: The following table lists electric fields caused by several symmetric charge distributions. In the table, q, Q, λ, and σ refer to the *magnitudes* of the quantities.

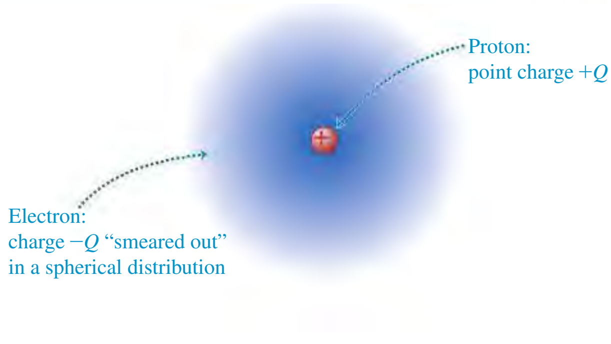
Charge Distribution	Point in Electric Field	Electric Field Magnitude
Single point charge q	Distance r from q	$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}$
Charge q on surface of conducting sphere with radius R	Outside sphere, $r > R$	$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}$
	Inside sphere, $r < R$	$E = 0$
Infinite wire, charge per unit length λ	Distance r from wire	$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r}$
Infinite conducting cylinder with radius R, charge per unit length λ	Outside cylinder, $r > R$	$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r}$
	Inside cylinder, $r < R$	$E = 0$
Solid insulating sphere with radius R, charge Q distributed uniformly throughout volume	Outside sphere, $r > R$	$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$
	Inside sphere, $r < R$	$E = \frac{1}{4\pi\epsilon_0}\frac{Qr}{R^3}$
Infinite sheet of charge with uniform charge per unit area σ	Any point	$E = \frac{\sigma}{2\epsilon_0}$
Two oppositely charged conducting plates with surface charge densities $+\sigma$ and $-\sigma$	Any point between plates	$E = \frac{\sigma}{\epsilon_0}$
Charged conductor	Just outside the conductor	$E = \frac{\sigma}{\epsilon_0}$

BRIDGING PROBLEM ELECTRIC FIELD INSIDE A HYDROGEN ATOM

SOLUTION

A hydrogen atom is made up of a proton of charge $+Q = 1.60 \times 10^{-19}\text{ C}$ and an electron of charge $-Q = -1.60 \times 10^{-19}\text{ C}$. The proton may be regarded as a point charge at $r = 0$, the center of the atom. The motion of the electron causes its charge to be "smeared out" into a spherical distribution around the proton (**Fig. 22.29**), so that the electron is equivalent to a charge per unit volume of $\rho(r) = -(Q/\pi a_0^3)e^{-2r/a_0}$, where $a_0 = 5.29 \times 10^{-11}\text{ m}$ is called the *Bohr radius.* (a) Find the total amount of the hydrogen atom's charge that is enclosed within a sphere with radius r centered on the proton. (b) Find the electric field (magnitude and direction) caused by the charge of the hydrogen atom as a function of r. (c) Make a graph as a function of r of the ratio of the electric-field magnitude E to the magnitude of the field due to the proton alone.

22.29 The charge distribution in a hydrogen atom.



SOLUTION GUIDE

IDENTIFY and SET UP

1. The charge distribution in this problem is spherically symmetric, as in Example 22.9, so you can solve it with Gauss's law.
2. The charge within a sphere of radius r includes the proton charge $+Q$ plus the portion of the electron charge distribution that lies within the sphere. The difference from Example 22.9 is that the electron charge distribution is *not* uniform, so the charge enclosed within a sphere of radius r is *not* simply the charge density multiplied by the volume $4\pi r^3/3$ of the sphere. Instead, you'll have to do an integral.
3. Consider a thin spherical shell centered on the proton, with radius r' and infinitesimal thickness dr'. Since the shell is so thin, every point within the shell is at essentially the same radius from the proton. Hence the amount of electron charge within this shell *is* equal to the electron charge density $\rho(r')$ at this radius multiplied by the volume dV of the shell. What is dV in terms of r'?
4. The total electron charge within a radius r equals the integral of $\rho(r')dV$ from $r' = 0$ to $r' = r$. Set up this integral (but don't solve it yet), and use it to write an expression for the total charge (including the proton) within a sphere of radius r.

EXECUTE

5. Integrate your expression from step 4 to find the charge within radius r. (*Hint:* Integrate by substitution: Change the integration variable from r' to $x = 2r'/a_0$. You can use integration by parts to calculate the integral $\int x^2 e^{-x}\,dx$, or you can look it up in a table of integrals or on the Web.)
6. Use Gauss's law and your results from step 5 to find the electric field at a distance r from the proton.
7. Find the ratio referred to in part (c) and graph it versus r. (You'll actually find it simplest to graph this function versus the quantity r/a_0.)

EVALUATE

8. How do your results for the enclosed charge and the electric-field magnitude behave in the limit $r \to 0$? In the limit $r \to \infty$? Explain your results.

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q22.1 A rubber balloon has a single point charge in its interior. Does the electric flux through the balloon depend on whether or not it is fully inflated? Explain your reasoning.

Q22.2 Suppose that in Fig. 22.15 both charges were positive. What would be the fluxes through each of the four surfaces in the example?

Q22.3 In Fig. 22.15, suppose a third point charge were placed outside the purple Gaussian surface C. Would this affect the electric flux through any of the surfaces A, B, C, or D in the figure? Why or why not?

Q22.4 A certain region of space bounded by an imaginary closed surface contains no charge. Is the electric field always zero everywhere on the surface? If not, under what circumstances is it zero on the surface?

Q22.5 A spherical Gaussian surface encloses a point charge q. If the point charge is moved from the center of the sphere to a point away from the center, does the electric field at a point on the surface change? Does the total flux through the Gaussian surface change? Explain.

Q22.6 You find a sealed box on your doorstep. You suspect that the box contains several charged metal spheres packed in insulating material. How can you determine the total net charge inside the box without opening the box? Or isn't this possible?

Q22.7 A solid copper sphere has a net positive charge. The charge is distributed uniformly over the surface of the sphere, and the electric field inside the sphere is zero. Then a negative point charge outside the sphere is brought close to the surface of the sphere. Is all the net charge on the sphere still on its surface? If so, is this charge still distributed uniformly over the surface? If it is not uniform, how is it distributed? Is the electric field inside the sphere still zero? In each case justify your answers.

Q22.8 If the electric field of a point charge were proportional to $1/r^3$ instead of $1/r^2$, would Gauss's law still be valid? Explain your reasoning. (*Hint:* Consider a spherical Gaussian surface centered on a single point charge.)

Q22.9 In a conductor, one or more electrons from each atom are free to roam throughout the volume of the conductor. Does this contradict the statement that any excess charge on a solid conductor must reside on its surface? Why or why not?

Q22.10 You charge up the Van de Graaff generator shown in Fig. 22.26, and then bring an identical but uncharged hollow conducting sphere near it, without letting the two spheres touch. Sketch the distribution of charges on the second sphere. What is the net flux through the second sphere? What is the electric field inside the second sphere?

Q22.11 A lightning rod is a rounded copper rod mounted on top of a building and welded to a heavy copper cable running down into the ground. Lightning rods are used to protect houses and barns from lightning; the lightning current runs through the copper rather than through the building. Why? Why should the end of the rod be rounded?

Q22.12 A solid conductor has a cavity in its interior. Would the presence of a point charge inside the cavity affect the electric field outside the conductor? Why or why not? Would the presence of a point charge outside the conductor affect the electric field inside the cavity? Again, why or why not?

Q22.13 Explain this statement: "In a static situation, the electric field at the surface of a conductor can have no component parallel to the surface because this would violate the condition that the charges on the surface are at rest." Would this statement be valid for the electric field at the surface of an *insulator*? Explain your answer and the reason for any differences between the cases of a conductor and an insulator.

Q22.14 In a certain region of space, the electric field $\vec{E}$ is uniform. (a) Use Gauss's law to prove that this region of space must be electrically neutral; that is, the volume charge density ρ must be zero. (b) Is the converse true? That is, in a region of space where there is no charge, must $\vec{E}$ be uniform? Explain.

Q22.15 (a) In a certain region of space, the volume charge density ρ has a uniform positive value. Can $\vec{E}$ be uniform in this region? Explain. (b) Suppose that in this region of uniform positive ρ there is a "bubble" within which $\rho = 0$. Can $\vec{E}$ be uniform within this bubble? Explain.

Q22.16 A negative charge $-Q$ is placed inside the cavity of a hollow metal solid. The outside of the solid is grounded by connecting a conducting wire between it and the earth. Is any excess charge induced on the inner surface of the metal? Is there any excess charge on the outside surface of the metal? Why or why not? Would someone outside the solid measure an electric field due to the charge $-Q$? Is it reasonable to say that the grounded conductor has *shielded* the region outside the conductor from the effects of the charge $-Q$? In principle, could the same thing be done for gravity? Why or why not?

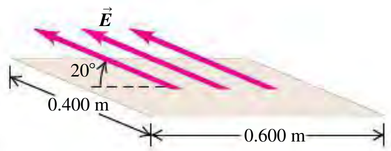
EXERCISES

Section 22.2 Calculating Electric Flux

22.1 • A flat sheet of paper of area $0.250\ \text{m}^2$ is oriented so that the normal to the sheet is at an angle of 60° to a uniform electric field of magnitude 14 N/C. (a) Find the magnitude of the electric flux through the sheet. (b) Does the answer to part (a) depend on the shape of the sheet? Why or why not? (c) For what angle ϕ between the normal to the sheet and the electric field is the magnitude of the flux through the sheet (i) largest and (ii) smallest? Explain your answers.

22.2 •• A flat sheet is in the shape of a rectangle with sides of lengths 0.400 m and 0.600 m. The sheet is immersed in a uniform electric field of magnitude 90.0 N/C that is directed at 20° from the plane of the sheet (**Fig. E22.2**). Find the magnitude of the electric flux through the sheet.

Figure **E22.2**



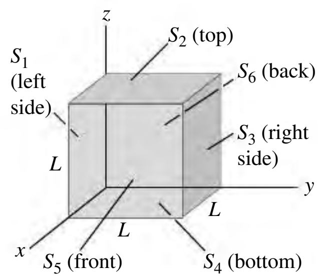
22.3 • You measure an electric field of 1.25×10^6 N/C at a distance of 0.150 m from a point charge. There is no other source of electric field in the region other than this point charge. (a) What is the electric flux through the surface of a sphere that has this charge at its center and that has radius 0.150 m? (b) What is the magnitude of this charge?

22.4 • It was shown in Example 21.10 (Section 21.5) that the electric field due to an infinite line of charge is perpendicular to the line and has magnitude $E = \lambda/2\pi\epsilon_0 r$. Consider an imaginary cylinder with radius $r = 0.250$ m and length $l = 0.400$ m that has an infinite line of positive charge running along its axis. The charge per unit length on the line is $\lambda = 3.00\ \mu\text{C/m}$. (a) What is the electric flux through the cylinder due to this infinite line of charge? (b) What is the flux through the cylinder if its radius is increased to $r = 0.500$ m? (c) What is the flux through the cylinder if its length is increased to $l = 0.800$ m?

22.5 •• A hemispherical surface with radius r in a region of uniform electric field $\vec{E}$ has its axis aligned parallel to the direction of the field. Calculate the flux through the surface.

22.6 • The cube in **Fig. E22.6** has sides of length $L = 10.0$ cm. The electric field is uniform, has magnitude $E = 4.00 \times 10^3$ N/C, and is parallel to the xy-plane at an angle of 53.1° measured from the $+x$-axis toward the $+y$-axis. (a) What is the electric flux through each of the six cube faces S_1, S_2, S_3, S_4, S_5, and S_6? (b) What is the total electric flux through all faces of the cube?

Figure **E22.6**

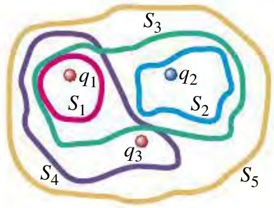


Section 22.3 Gauss's Law

22.7 • **BIO** As discussed in Section 22.5, human nerve cells have a net negative charge and the material in the interior of the cell is a good conductor. If a cell has a net charge of -8.65 pC, what are the magnitude and direction (inward or outward) of the net flux through the cell boundary?

22.8 • The three small spheres shown in **Fig. E22.8** carry charges $q_1 = 4.00$ nC, $q_2 = -7.80$ nC, and $q_3 = 2.40$ nC. Find the net electric flux through each of the following closed surfaces shown in cross section in the figure: (a) S_1; (b) S_2; (c) S_3; (d) S_4; (e) S_5. (f) Do your answers to parts (a)–(e) depend on how the charge is distributed over each small sphere? Why or why not?

Figure **E22.8**



Surface	What it encloses
S_1	q_1
S_2	q_2
S_3	q_1 and q_2
S_4	q_1 and q_3
S_5	q_1 and q_2 and q_3

22.9 •• A charged paint is spread in a very thin uniform layer over the surface of a plastic sphere of diameter 12.0 cm, giving it a charge of $-49.0\ \mu$C. Find the electric field (a) just inside the paint layer; (b) just outside the paint layer; (c) 5.00 cm outside the surface of the paint layer.

22.10 • A point charge $q_1 = 4.00$ nC is located on the x-axis at $x = 2.00$ m, and a second point charge $q_2 = -6.00$ nC is on the y-axis at $y = 1.00$ m. What is the total electric flux due to these two point charges through a spherical surface centered at the origin and with radius (a) 0.500 m, (b) 1.50 m, (c) 2.50 m?

22.11 • A 6.20-μC point charge is at the center of a cube with sides of length 0.500 m. (a) What is the electric flux through one of the six faces of the cube? (b) How would your answer to part (a) change if the sides were 0.250 m long? Explain.

22.12 • **Electric Fields in an Atom.** The nuclei of large atoms, such as uranium, with 92 protons, can be modeled as spherically symmetric spheres of charge. The radius of the uranium nucleus is approximately 7.4×10^{-15} m. (a) What is the electric field this nucleus produces just outside its surface? (b) What magnitude of electric field does it produce at the distance of the electrons, which is about 1.0×10^{-10} m? (c) The electrons can be modeled as forming a uniform shell of negative charge. What net electric field do they produce at the location of the nucleus?

Section 22.4 Applications of Gauss's Law and Section 22.5 Charges on Conductors

22.13 •• Two very long uniform lines of charge are parallel and are separated by 0.300 m. Each line of charge has charge per unit length $+5.20\ \mu$C/m. What magnitude of force does one line of charge exert on a 0.0500-m section of the other line of charge?

22.14 •• A solid metal sphere with radius 0.450 m carries a net charge of 0.250 nC. Find the magnitude of the electric field (a) at a point 0.100 m outside the surface of the sphere and (b) at a point inside the sphere, 0.100 m below the surface.

22.15 •• How many excess electrons must be added to an isolated spherical conductor 26.0 cm in diameter to produce an electric field of magnitude 1150 N/C just outside the surface?

22.16 • Some planetary scientists have suggested that the planet Mars has an electric field somewhat similar to that of the earth, producing a net electric flux of $-3.63 \times 10^{16}\ \text{N} \cdot \text{m}^2/\text{C}$ at the planet's surface. Calculate: (a) the total electric charge on the planet; (b) the electric field at the planet's surface (refer to the astronomical data inside the back cover); (c) the charge density on Mars, assuming all the charge is uniformly distributed over the planet's surface.

22.17 •• A very long uniform line of charge has charge per unit length 4.80 μC/m and lies along the x-axis. A second long uniform line of charge has charge per unit length $-2.40\ \mu$C/m and is parallel to the x-axis at $y = 0.400$ m. What is the net electric field (magnitude and direction) at the following points on the y-axis: (a) $y = 0.200$ m and (b) $y = 0.600$ m?

22.18 •• The electric field 0.400 m from a very long uniform line of charge is 840 N/C. How much charge is contained in a 2.00-cm section of the line?

22.19 •• A hollow, conducting sphere with an outer radius of 0.250 m and an inner radius of 0.200 m has a uniform surface charge density of $+6.37 \times 10^{-6}\ \text{C/m}^2$. A charge of $-0.500\ \mu$C is now introduced at the center of the cavity inside the sphere. (a) What is the new charge density on the outside of the sphere? (b) Calculate the strength of the electric field just outside the sphere. (c) What is the electric flux through a spherical surface just inside the inner surface of the sphere?

22.20 • (a) At a distance of 0.200 cm from the center of a charged conducting sphere with radius 0.100 cm, the electric field is 480 N/C. What is the electric field 0.600 cm from the center of the sphere? (b) At a distance of 0.200 cm from the axis of a very long charged conducting cylinder with radius 0.100 cm, the electric field is 480 N/C. What is the electric field 0.600 cm from the axis of the cylinder? (c) At a distance of 0.200 cm from a large uniform sheet of charge, the electric field is 480 N/C. What is the electric field 1.20 cm from the sheet?

22.21 •• The electric field at a distance of 0.145 m from the surface of a solid insulating sphere with radius 0.355 m is 1750 N/C. (a) Assuming the sphere's charge is uniformly distributed, what is the charge density inside it? (b) Calculate the electric field inside the sphere at a distance of 0.200 m from the center.

22.22 •• A point charge of $-3.00\ \mu$C is located in the center of a spherical cavity of radius 6.50 cm that, in turn, is at the center of an insulating charged solid sphere. The charge density in the solid is $\rho = 7.35 \times 10^{-4}\ \text{C/m}^3$. Calculate the electric field inside the solid at a distance of 9.50 cm from the center of the cavity.

22.23 •• **CP** An electron is released from rest at a distance of 0.300 m from a large insulating sheet of charge that has uniform surface charge density $+2.90 \times 10^{-12}\ \text{C/m}^2$. (a) How much work is done on the electron by the electric field of the sheet as the electron moves from its initial position to a point 0.050 m from the sheet? (b) What is the speed of the electron when it is 0.050 m from the sheet?

22.24 •• Charge Q is distributed uniformly throughout the volume of an insulating sphere of radius $R = 4.00$ cm. At a distance of $r = 8.00$ cm from the center of the sphere, the electric field due to the charge distribution has magnitude $E = 940$ N/C. What are (a) the volume charge density for the sphere and (b) the electric field at a distance of 2.00 cm from the sphere's center?

22.25 • A conductor with an inner cavity, like that shown in Fig. 22.23c, carries a total charge of $+5.00$ nC. The charge within the cavity, insulated from the conductor, is -6.00 nC. How much charge is on (a) the inner surface of the conductor and (b) the outer surface of the conductor?

22.26 •• A very large, horizontal, nonconducting sheet of charge has uniform charge per unit area $\sigma = 5.00 \times 10^{-6}$ C/m^2. (a) A small sphere of mass $m = 8.00 \times 10^{-6}$ kg and charge q is placed 3.00 cm above the sheet of charge and then released from rest. (a) If the sphere is to remain motionless when it is released, what must be the value of q? (b) What is q if the sphere is released 1.50 cm above the sheet?

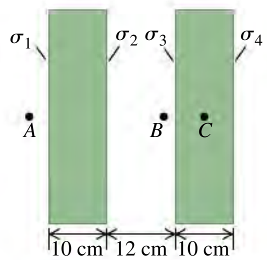
22.27 • Apply Gauss's law to the Gaussian surfaces S_2, S_3, and S_4 in Fig. 22.21b to calculate the electric field between and outside the plates.

22.28 • A square insulating sheet 80.0 cm on a side is held horizontally. The sheet has 4.50 nC of charge spread uniformly over its area. (a) Calculate the electric field at a point 0.100 mm above the center of the sheet. (b) Estimate the electric field at a point 100 m above the center of the sheet. (c) Would the answers to parts (a) and (b) be different if the sheet were made of a conducting material? Why or why not?

22.29 • An infinitely long cylindrical conductor has radius R and uniform surface charge density σ. (a) In terms of σ and R, what is the charge per unit length λ for the cylinder? (b) In terms of σ, what is the magnitude of the electric field produced by the charged cylinder at a distance $r > R$ from its axis? (c) Express the result of part (b) in terms of λ and show that the electric field outside the cylinder is the same as if all the charge were on the axis. Compare your result to the result for a line of charge in Example 22.6 (Section 22.4).

22.30 • Two very large, nonconducting plastic sheets, each 10.0 cm thick, carry uniform charge densities σ_1, σ_2, σ_3, and σ_4 on their surfaces (**Fig. E22.30**). These surface charge densities have the values $\sigma_1 = -6.00\ \mu\text{C/m}^2$, $\sigma_2 = +5.00\ \mu\text{C/m}^2$, $\sigma_3 = +2.00\ \mu\text{C/m}^2$, and $\sigma_4 = +4.00\ \mu\text{C/m}^2$. Use Gauss's law to find the magnitude and direction of the electric field at the following points, far from the edges of these sheets: (a) point A, 5.00 cm from the left face of the left-hand sheet; (b) point B, 1.25 cm from the inner surface of the right-hand sheet; (c) point C, in the middle of the right-hand sheet.

Figure **E22.30**



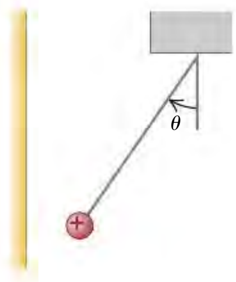
PROBLEMS

22.31 •• **CP** At time $t = 0$ a proton is a distance of 0.360 m from a very large insulating sheet of charge and is moving parallel to the sheet with speed 9.70×10^2 m/s. The sheet has uniform surface charge density 2.34×10^{-9} C/m^2. What is the speed of the proton at $t = 5.00 \times 10^{-8}$ s?

22.32 •• **CP** A very small object with mass 8.20×10^{-9} kg and positive charge 6.50×10^{-9} C is projected directly toward a very large insulating sheet of positive charge that has uniform surface charge density 5.90×10^{-8} C/m^2. The object is initially 0.400 m from the sheet. What initial speed must the object have in order for its closest distance of approach to the sheet to be 0.100 m?

22.33 •• **CP** A small sphere with mass 4.00×10^{-6} kg and charge 5.00×10^{-8} C hangs from a thread near a very large, charged insulating sheet (**Fig. P22.33**). The charge density on the surface of the sheet is uniform and equal to -2.50×10^{-9} C/m^2. Find the angle of the thread.

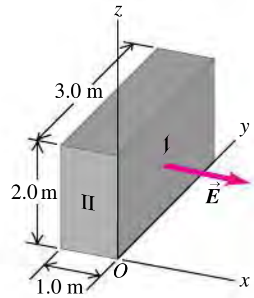
Figure **P22.33**



22.34 •• A cube has sides of length $L = 0.300$ m. One corner is at the origin (Fig. E22.6). The nonuniform electric field is given by $\vec{E} = (-5.00\ \text{N/C}\cdot\text{m})x\hat{\imath} + (3.00\ \text{N/C}\cdot\text{m})z\hat{k}$. (a) Find the electric flux through each of the six cube faces S_1, S_2, S_3, S_4, S_5, and S_6. (b) Find the total electric charge inside the cube.

22.35 • The electric field $\vec{E}$ in **Fig. P22.35** is everywhere parallel to the x-axis, so the components E_y and E_z are zero. The x-component of the field E_x depends on x but not on y or z. At points in the yz-plane (where $x = 0$), $E_x = 125$ N/C. (a) What is the electric flux through surface I in Fig. P22.35? (b) What is the electric flux through surface II? (c) The volume shown is a small section of a very large insulating slab 1.0 m thick. If there is a total charge of -24.0 nC within the volume shown, what are the magnitude and direction of $\vec{E}$ at the face opposite surface I? (d) Is the electric field produced by charges only within the slab, or is the field also due to charges outside the slab? How can you tell?

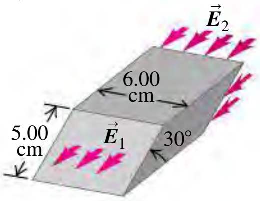
Figure **P22.35**



22.36 •• **CALC** In a region of space there is an electric field $\vec{E}$ that is in the z-direction and that has magnitude $E = [964\ \text{N/(C}\cdot\text{m)}]x$. Find the flux for this field through a square in the xy-plane at $z = 0$ and with side length 0.350 m. One side of the square is along the $+x$-axis and another side is along the $+y$-axis.

22.37 •• The electric field $\vec{E}_1$ at one face of a parallelepiped is uniform over the entire face and is directed out of the face. At the opposite face, the electric field $\vec{E}_2$ is also uniform over the entire face and is directed into that face (**Fig. P22.37**). The two faces in question are inclined at 30.0° from the horizontal, while both $\vec{E}_1$ and $\vec{E}_2$ are horizontal; $\vec{E}_1$ has a magnitude of 2.50×10^4 N/C, and $\vec{E}_2$ has a magnitude of 7.00×10^4 N/C. (a) Assuming that no other electric field lines cross the surfaces of the parallelepiped, determine the net charge contained within. (b) Is the electric field produced by the charges only within the parallelepiped, or is the field also due to charges outside the parallelepiped? How can you tell?

Figure **P22.37**



22.38 • A long line carrying a uniform linear charge density $+50.0\ \mu$C/m runs parallel to and 10.0 cm from the surface of a large, flat plastic sheet that has a uniform surface charge density of $-100\ \mu\text{C/m}^2$ on one side. Find the location of all points where an α particle would feel no force due to this arrangement of charged objects.

22.39 • **The Coaxial Cable.** A long coaxial cable consists of an inner cylindrical conductor with radius a and an outer coaxial cylinder with inner radius b and outer radius c. The outer cylinder is mounted on insulating supports and has no net charge. The inner cylinder has a uniform positive charge per unit length λ. Calculate the electric field (a) at any point between the cylinders a distance r from the axis and (b) at any point outside the outer cylinder. (c) Graph the magnitude of the electric field as a function of the distance r from the axis of the cable, from $r = 0$ to $r = 2c$. (d) Find the charge per unit length on the inner surface and on the outer surface of the outer cylinder.

22.40 • A very long conducting tube (hollow cylinder) has inner radius a and outer radius b. It carries charge per unit length $+\alpha$, where α is a positive constant with units of C/m. A line of charge lies along the axis of the tube. The line of charge has charge per unit length $+\alpha$. (a) Calculate the electric field in terms of α and the distance r from the axis of the tube for (i) $r < a$; (ii) $a < r < b$; (iii) $r > b$. Show your results in a graph of E as a function of r. (b) What is the charge per unit length on (i) the inner surface of the tube and (ii) the outer surface of the tube?

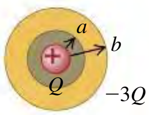
22.41 • A very long, solid cylinder with radius R has positive charge uniformly distributed throughout it, with charge per unit volume ρ. (a) Derive the expression for the electric field inside the volume at a distance r from the axis of the cylinder in terms of the charge density ρ. (b) What is the electric field at a point outside the volume in terms of the charge per unit length λ in the cylinder? (c) Compare the answers to parts (a) and (b) for $r = R$. (d) Graph the electric-field magnitude as a function of r from $r = 0$ to $r = 3R$.

22.42 • **A Sphere in a Sphere.** A solid conducting sphere carrying charge q has radius a. It is inside a concentric hollow conducting sphere with inner radius b and outer radius c. The hollow sphere has no net charge. (a) Derive expressions for the electric-field magnitude in terms of the distance r from the center for the regions $r < a$, $a < r < b$, $b < r < c$, and $r > c$. (b) Graph the magnitude of the electric field as a function of r from $r = 0$ to $r = 2c$. (c) What is the charge on the inner surface of the hollow sphere? (d) On the outer surface? (e) Represent the charge of the small sphere by four plus signs. Sketch the field lines of the system within a spherical volume of radius $2c$.

22.43 • A solid conducting sphere with radius R that carries positive charge Q is concentric with a very thin insulating shell of radius $2R$ that also carries charge Q. The charge Q is distributed uniformly over the insulating shell. (a) Find the electric field (magnitude and direction) in each of the regions $0 < r < R$, $R < r < 2R$, and $r > 2R$. (b) Graph the electric-field magnitude as a function of r.

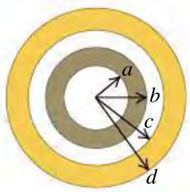
22.44 • A conducting spherical shell with inner radius a and outer radius b has a positive point charge Q located at its center. The total charge on the shell is $-3Q$, and it is insulated from its surroundings (**Fig. P22.44**). (a) Derive expressions for the electric-field magnitude E in terms of the distance r from the center for the regions $r < a$, $a < r < b$, and $r > b$. What is the surface charge density (b) on the inner surface of the conducting shell; (c) on the outer surface of the conducting shell? (d) Sketch the electric field lines and the location of all charges. (e) Graph E as a function of r.

Figure **P22.44**



22.45 • **Concentric Spherical Shells.** A small conducting spherical shell with inner radius a and outer radius b is concentric with a larger conducting spherical shell with inner radius c and outer radius d (**Fig. P22.45**). The inner shell has total charge $+2q$, and the outer shell has charge $+4q$. (a) Calculate the electric field $\vec{E}$ (magnitude and direction) in terms of q and the distance r from the common center of the two shells for (i) $r < a$; (ii) $a < r < b$; (iii) $b < r < c$; (iv) $c < r < d$; (v) $r > d$. Graph the radial component of $\vec{E}$ as a function of r. (b) What is the total charge on the (i) inner surface of the small shell; (ii) outer surface of the small shell; (iii) inner surface of the large shell; (iv) outer surface of the large shell?

Figure **P22.45**



22.46 • Repeat Problem 22.45, but now let the outer shell have charge $-2q$. The inner shell still has charge $+2q$.

22.47 • Negative charge $-Q$ is distributed uniformly over the surface of a thin spherical insulating shell with radius R. Calculate the force (magnitude and direction) that the shell exerts on a positive point charge q located a distance (a) $r > R$ from the center of the shell (outside the shell); (b) $r < R$ from the center of the shell (inside the shell).

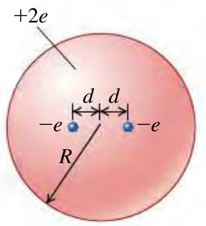
22.48 • A solid conducting sphere with radius R carries a positive total charge Q. The sphere is surrounded by an insulating shell with inner radius R and outer radius $2R$. The insulating shell has a uniform charge density ρ. (a) Find the value of ρ so that the net charge of the entire system is zero. (b) If ρ has the value found in part (a), find the electric field $\vec{E}$ (magnitude and direction) in each of the regions $0 < r < R$, $R < r < 2R$, and $r > 2R$. Graph the radial component of $\vec{E}$ as a function of r. (c) As a general rule, the electric field is discontinuous only at locations where there is a thin sheet of charge. Explain how your results in part (b) agree with this rule.

22.49 ••• **CALC** An insulating hollow sphere has inner radius a and outer radius b. Within the insulating material the volume charge density is given by $\rho(r) = \alpha/r$, where α is a positive constant. (a) In terms of α and a, what is the magnitude of the electric field at a distance r from the center of the shell, where $a < r < b$? (b) A point charge q is placed at the center of the hollow space, at $r = 0$. In terms of α and a, what value must q have (sign and magnitude) in order for the electric field to be constant in the region $a < r < b$, and what then is the value of the constant field in this region?

22.50 •• **CP** **Thomson's Model of the Atom.** Early in the 20th century, a leading model of the structure of the atom was that of English physicist J. J. Thomson (the discoverer of the electron). In Thomson's model, an atom consisted of a sphere of positively charged material in which were embedded negatively charged electrons, like chocolate chips in a ball of cookie dough. Consider such an atom consisting of one electron with mass m and charge $-e$, which may be regarded as a point charge, and a uniformly charged sphere of charge $+e$ and radius R. (a) Explain why the electron's equilibrium position is at the center of the nucleus. (b) In Thomson's model, it was assumed that the positive material provided little or no resistance to the electron's motion. If the electron is displaced from equilibrium by a distance less than R, show that the resulting motion of the electron will be simple harmonic, and calculate the frequency of oscillation. (*Hint:* Review the definition of SHM in Section 14.2. If it can be shown that the net force on the electron is of this form, then it follows that the motion is simple harmonic. Conversely, if the net force on the electron does not follow this form, the motion is not simple harmonic.) (c) By Thomson's time, it was known that excited atoms emit light waves of only certain frequencies. In his model, the frequency of emitted light is the same as the oscillation frequency of the electron(s) in the atom. What radius would a Thomson-model atom need for it to produce red light of frequency 4.57×10^{14} Hz? Compare your answer to the radii of real atoms, which are of the order of 10^{-10} m (see Appendix F). (d) If the electron were displaced from equilibrium by a distance greater than R, would the electron oscillate? Would its motion be simple harmonic? Explain your reasoning. (*Historical note:* In 1910, the atomic nucleus was discovered, proving the Thomson model to be incorrect. An atom's positive charge is not spread over its volume, as Thomson supposed, but is concentrated in the tiny nucleus of radius 10^{-14} to 10^{-15} m.)

22.51 • **Thomson's Model of the Atom, Continued.** Using Thomson's (outdated) model of the atom described in Problem 22.50, consider an atom consisting of two electrons, each of charge $-e$, embedded in a sphere of charge $+2e$ and radius R. In equilibrium, each electron is a distance d from the center of the atom (**Fig. P22.51**). Find the distance d in terms of the other properties of the atom.

Figure **P22.51**



22.52 •• (a) How many excess electrons must be distributed uniformly within the volume of an isolated plastic sphere 30.0 cm in diameter to produce an electric field of magnitude 1390 N/C just outside the surface of the sphere? (b) What is the electric field at a point 10.0 cm outside the surface of the sphere?

22.53 •• **CALC** A nonuniform, but spherically symmetric, distribution of charge has a charge density $\rho(r)$ given as follows:

$$\rho(r) = \rho_0\left(1 - \frac{r}{R}\right) \qquad \text{for } r \leq R$$

$$\rho(r) = 0 \qquad \text{for } r \geq R$$

where $\rho_0 = 3Q/\pi R^3$ is a positive constant. (a) Show that the total charge contained in the charge distribution is Q. (b) Show that the electric field in the region $r \geq R$ is identical to that produced by a point charge Q at $r = 0$. (c) Obtain an expression for the electric field in the region $r \leq R$. (d) Graph the electric-field magnitude E as a function of r. (e) Find the value of r at which the electric field is maximum, and find the value of that maximum field.

22.54 • **A Uniformly Charged Slab.** A slab of insulating material has thickness $2d$ and is oriented so that its faces are parallel to the yz-plane and given by the planes $x = d$ and $x = -d$. The y- and z-dimensions of the slab are very large compared to d; treat them as essentially infinite. The slab has a uniform positive charge density ρ. (a) Explain why the electric field due to the slab is zero at the center of the slab $(x = 0)$. (b) Using Gauss's law, find the electric field due to the slab (magnitude and direction) at all points in space.

22.55 • **CALC** **A Nonuniformly Charged Slab.** Repeat Problem 22.54, but now let the charge density of the slab be given by $\rho(x) = \rho_0(x/d)^2$, where ρ_0 is a positive constant.

22.56 • **CALC** A nonuniform, but spherically symmetric, distribution of charge has a charge density $\rho(r)$ given as follows:

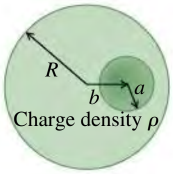
$$\rho(r) = \rho_0\left(1 - \frac{4r}{3R}\right) \qquad \text{for } r \leq R$$

$$\rho(r) = 0 \qquad \text{for } r \geq R$$

where ρ_0 is a positive constant. (a) Find the total charge contained in the charge distribution. Obtain an expression for the electric field in the region (b) $r \geq R$; (c) $r \leq R$. (d) Graph the electric-field magnitude E as a function of r. (e) Find the value of r at which the electric field is maximum, and find the value of that maximum field.

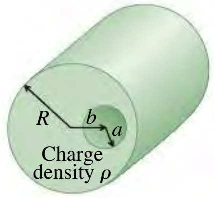
22.57 • (a) An insulating sphere with radius a has a uniform charge density ρ. The sphere is not centered at the origin but at $\vec{r} = \vec{b}$. Show that the electric field inside the sphere is given by $\vec{E} = \rho(\vec{r} - \vec{b})/3\epsilon_0$. (b) An insulating sphere of radius R has a spherical hole of radius a located within its volume and centered a distance b from the center of the sphere, where $a < b < R$ (a cross section of the sphere is shown in **Fig. P22.57**). The solid part of the sphere has a uniform volume charge density ρ. Find the magnitude and direction of the electric field $\vec{E}$ inside the hole, and show that $\vec{E}$ is uniform over the entire hole. [*Hint:* Use the principle of superposition and the result of part (a).]

Figure **P22.57**



22.58 • A very long, solid insulating cylinder has radius R; bored along its entire length is a cylindrical hole with radius a. The axis of the hole is a distance b from the axis of the cylinder, where $a < b < R$ (**Fig. P22.58**). The solid material of the cylinder has a uniform volume charge density ρ. Find the magnitude and direction of the electric field $\vec{E}$ inside the hole, and show that $\vec{E}$ is uniform over the entire hole. (*Hint:* See Problem 22.57.)

Figure **P22.58**



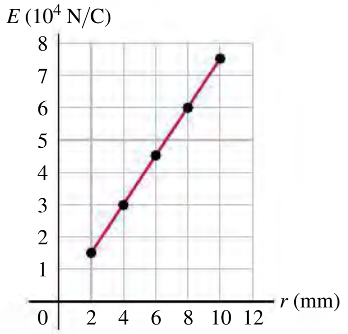
22.59 •• **DATA** In one experiment the electric field is measured for points at distances r from a uniform line of charge that has charge per unit length λ and length l, where $l \gg r$. In a second experiment the electric field is measured for points at distances r from the center of a uniformly charged insulating sphere that has volume charge density ρ and radius $R = 8.00$ mm, where $r > R$. The results of the two measurements are listed in the table, but you aren't told which set of data applies to which experiment:

r (cm)	1.00	1.50	2.00	2.50	3.00	3.50	4.00
Measurement A							
E (10^5 N/C)	2.72	1.79	1.34	1.07	0.902	0.770	0.677
Measurement B							
E (10^5 N/C)	5.45	2.42	1.34	0.861	0.605	0.443	0.335

For each set of data, draw two graphs: one for Er^2 versus r and one for Er versus r. (a) Use these graphs to determine which data set, A or B, is for the uniform line of charge and which set is for the uniformly charged sphere. Explain your reasoning. (b) Use the graphs in part (a) to calculate λ for the uniform line of charge and ρ for the uniformly charged sphere.

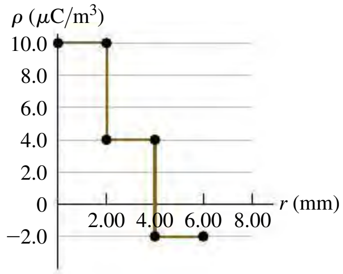
22.60 •• DATA The electric field is measured for points at distances r from the center of a uniformly charged insulating sphere that has volume charge density ρ and radius R, where $r < R$ (**Fig. P22.60**). Calculate ρ.

Figure **P22.60**



22.61 •• DATA The volume charge density ρ for a spherical charge distribution of radius $R = 6.00$ mm is not uniform. **Figure P22.61** shows ρ as a function of the distance r from the center of the distribution. Calculate the electric field at these values of r: (i) 1.00 mm; (ii) 3.00 mm; (iii) 5.00 mm; (iv) 7.00 mm.

Figure **P22.61**



CHALLENGE PROBLEM

22.62 ••• CP CALC A region in space contains a total positive charge Q that is distributed spherically such that the volume charge density $\rho(r)$ is given by

$$\rho(r) = 3\alpha r/2R \qquad \text{for } r \le R/2$$
$$\rho(r) = \alpha[1 - (r/R)^2] \qquad \text{for } R/2 \le r \le R$$
$$\rho(r) = 0 \qquad \text{for } r \ge R$$

Here α is a positive constant having units of C/m^3. (a) Determine α in terms of Q and R. (b) Using Gauss's law, derive an expression for the magnitude of the electric field as a function of r. Do this separately for all three regions. Express your answers in terms of Q. (c) What fraction of the total charge is contained within the region $R/2 \le r \le R$? (d) What is the magnitude of $\vec{E}$ at $r = R/2$? (e) If an electron with charge $q' = -e$ is released from rest at any point in any of the three regions, the resulting motion will be oscillatory but not simple harmonic. Why?

PASSAGE PROBLEMS

SPACE RADIATION SHIELDING. One of the hazards facing humans in space is space radiation: high-energy charged particles emitted by the sun. During a solar flare, the intensity of this radiation can reach lethal levels. One proposed method of protection for astronauts on the surface of the moon or Mars is an array of large, electrically charged spheres placed high above areas where people live and work. The spheres would produce a strong electric field $\vec{E}$ to deflect the charged particles that make up space radiation. The spheres would be similar in construction to a Mylar balloon, with a thin, electrically conducting layer on the outside surface on which a net positive or negative charge would be placed. A typical sphere might be 5 m in diameter.

22.63 Suppose that to repel electrons in the radiation from a solar flare, each sphere must produce an electric field $\vec{E}$ of magnitude 1×10^6 N/C at 25 m from the center of the sphere. What net charge on each sphere is needed? (a) -0.07 C; (b) -8 mC; (c) -80 μC; (d) -1×10^{-20} C.

22.64 What is the magnitude of $\vec{E}$ just outside the surface of such a sphere? (a) 0; (b) 10^6 N/C; (c) 10^7 N/C; (d) 10^8 N/C.

22.65 What is the direction of $\vec{E}$ just outside the surface of such a sphere? (a) Tangent to the surface of the sphere; (b) perpendicular to the surface, pointing toward the sphere; (c) perpendicular to the surface, pointing away from the sphere; (d) there is no electric field just outside the surface.

22.66 Which statement is true about $\vec{E}$ inside a negatively charged sphere as described here? (a) It points from the center of the sphere to the surface and is largest at the center. (b) It points from the surface to the center of the sphere and is largest at the surface. (c) It is zero. (d) It is constant but not zero.

Answers

Chapter Opening Question ?

(iii) The electric field inside a cavity within a conductor is zero, so there would be no electric effect on the child. (See Section 22.5.)

Test Your Understanding Questions

22.1 (iii) Each part of the surface of the box will be three times farther from the charge $+q$, so the electric field will be $\left(\frac{1}{3}\right)^2 = \frac{1}{9}$ as strong. But the area of the box will increase by a factor of $3^2 = 9$. Hence the electric flux will be multiplied by a factor of $\left(\frac{1}{9}\right)(9) = 1$. In other words, the flux will be unchanged.

22.2 (iv), (ii), (i), (iii) In each case the electric field is uniform, so the flux is $\Phi_E = \vec{E} \cdot \vec{A}$. We use the relationships for the scalar products of unit vectors: $\hat{\imath} \cdot \hat{\imath} = \hat{\jmath} \cdot \hat{\jmath} = 1$, $\hat{\imath} \cdot \hat{\jmath} = 0$. In case (i) we have $\Phi_E = (4.0\ \text{N/C})(6.0\ \text{m}^2)\hat{\imath} \cdot \hat{\jmath} = 0$ (the electric field and vector area are perpendicular, so there is zero flux). In case (ii) we have $\Phi_E = [(4.0\ \text{N/C})\hat{\imath} + (2.0\ \text{N/C})\hat{\jmath}] \cdot (3.0\ \text{m}^2)\hat{\jmath} = (2.0\ \text{N/C}) \cdot (3.0\ \text{m}^2) = 6.0\ \text{N} \cdot \text{m}^2/\text{C}$. Similarly, in case (iii) we have $\Phi_E = [(4.0\ \text{N/C})\hat{\imath} - (2.0\ \text{N/C})\hat{\jmath}] \cdot [(3.0\ \text{m}^2)\hat{\imath} + (7.0\ \text{m}^2)\hat{\jmath}] = (4.0\ \text{N/C})(3.0\ \text{m}^2) - (2.0\ \text{N/C})(7.0\ \text{m}^2) = -2\ \text{N} \cdot \text{m}^2/\text{C}$, and in case (iv) we have $\Phi_E = [(4.0\ \text{N/C})\hat{\imath} - (2.0\ \text{N/C})\hat{\jmath}] \cdot [(3.0\ \text{m}^2)\hat{\imath} - (7.0\ \text{m}^2)\hat{\jmath}] = (4.0\ \text{N/C})(3.0\ \text{m}^2) + (2.0\ \text{N/C}) \cdot (7.0\ \text{m}^2) = 26\ \text{N} \cdot \text{m}^2/\text{C}$.

22.3 S_2, S_5, S_4, S_1 and S_3 (tie) Gauss's law tells us that the flux through a closed surface is proportional to the amount of charge enclosed within that surface. So an ordering of these surfaces by their fluxes is the same as an ordering by the amount of enclosed charge. Surface S_1 encloses no charge, surface S_2 encloses $9.0\ \mu\text{C} + 5.0\ \mu\text{C} + (-7.0\ \mu\text{C}) = 7.0\ \mu\text{C}$, surface S_3 encloses $9.0\ \mu\text{C} + 1.0\ \mu\text{C} + (-10.0\ \mu\text{C}) = 0$, surface S_4 encloses $8.0\ \mu\text{C} + (-7.0\ \mu\text{C}) = 1.0\ \mu\text{C}$, and surface S_5 encloses $8.0\ \mu\text{C} + (-7.0\ \mu\text{C}) + (-10.0\ \mu\text{C}) + (1.0\ \mu\text{C}) + (9.0\ \mu\text{C}) + (5.0\ \mu\text{C}) = 6.0\ \mu\text{C}$.

22.4 no You might be tempted to draw a Gaussian surface that is an enlarged version of the conductor, with the same shape and placed so that it completely encloses the conductor. While you know the flux through this Gaussian surface (by Gauss's law, it's $\Phi_E = Q/\epsilon_0$), the direction of the electric field need not be perpendicular to the surface and the magnitude of the field need not be the same at all points on the surface. It's not possible to do the flux integral $\oint E_\perp\, dA$, and we can't calculate the electric field. Gauss's law is useful for calculating the electric field only when the charge distribution is *highly* symmetric.

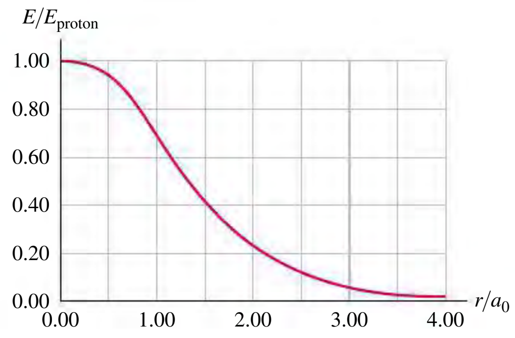
22.5 no Before you connect the wire to the sphere, the presence of the point charge will induce a charge $-q$ on the inner surface of the hollow sphere and a charge q on the outer surface (the net charge on the sphere is zero). There will be an electric field outside the sphere due to the charge on the outer surface. Once you touch the conducting wire to the sphere, however, electrons will flow from ground to the outer surface of the sphere to neutralize the charge there (see Fig. 21.7c). As a result the sphere will have no charge on its outer surface and no electric field outside.

Bridging Problem

(a) $Q(r) = Qe^{-2r/a_0}[2(r/a_0)^2 + 2(r/a_0) + 1]$

(b) $E = \dfrac{kQe^{-2r/a_0}}{r^2}[2(r/a_0)^2 + 2(r/a_0) + 1]$

(c)



? In one type of welding, electric charge flows between the welding tool and the metal pieces that are to be joined. This produces a glowing arc whose high temperature fuses the pieces together. Why must the tool be held close to the metal pieces? (i) To maximize the potential difference between tool and pieces; (ii) to minimize this potential difference; (iii) to maximize the electric field between tool and pieces; (iv) to minimize this electric field; (v) more than one of these.

23 ELECTRIC POTENTIAL

LEARNING GOALS

Looking forward at …

23.1 How to calculate the electric potential energy of a collection of charges.

23.2 The meaning and significance of electric potential.

23.3 How to calculate the electric potential that a collection of charges produces at a point in space.

23.4 How to use equipotential surfaces to visualize how the electric potential varies in space.

23.5 How to use electric potential to calculate the electric field.

Looking back at …

7.1–7.4 Conservative forces and potential energy.

21.1–21.6 Electric force and electric fields.

22.4, 22.5 Applications of Gauss's law.

This chapter is about energy associated with electrical interactions. Every time you turn on a light, use a mobile phone, or make toast in a toaster, you are using electrical energy, an indispensable ingredient of our technological society. In Chapters 6 and 7 we introduced the concepts of *work* and *energy* in the context of mechanics; now we'll combine these concepts with what we've learned about electric charge, electric forces, and electric fields. Just as we found for many problems in mechanics, using energy ideas makes it easier to solve a variety of problems in electricity.

When a charged particle moves in an electric field, the field exerts a force that can do *work* on the particle. This work can be expressed in terms of electric potential energy. Just as gravitational potential energy depends on the height of a mass above the earth's surface, electric potential energy depends on the position of the charged particle in the electric field. We'll use a new concept called *electric potential*, or simply *potential,* to describe electric potential energy. In circuits, a difference in potential from one point to another is often called *voltage*. The concepts of potential and voltage are crucial to understanding how electric circuits work and have equally important applications to electron beams used in cancer radiotherapy, high-energy particle accelerators, and many other devices.

23.1 ELECTRIC POTENTIAL ENERGY

The concepts of work, potential energy, and conservation of energy proved to be extremely useful in our study of mechanics. In this section we'll show that these concepts are just as useful for understanding and analyzing electrical interactions.

Let's begin by reviewing three essential points from Chapters 6 and 7. First, when a force $\vec{F}$ acts on a particle that moves from point a to point b, the work $W_{a\rightarrow b}$ done by the force is given by a *line integral:*

$$W_{a\rightarrow b} = \int_a^b \vec{F}\cdot d\vec{l} = \int_a^b F\cos\phi\, dl \quad \text{(work done by a force)} \qquad (23.1)$$

where $d\vec{l}$ is an infinitesimal displacement along the particle's path and ϕ is the angle between $\vec{F}$ and $d\vec{l}$ at each point along the path.

Second, if the force $\vec{F}$ is *conservative*, as we defined the term in Section 7.3, the work done by $\vec{F}$ can always be expressed in terms of a **potential energy** U. When the particle moves from a point where the potential energy is U_a to a point where it is U_b, the change in potential energy is $\Delta U = U_b - U_a$ and

Work done by a conservative force

Potential energy at initial position

$$W_{a \to b} = U_a - U_b = -(U_b - U_a) = -\Delta U \tag{23.2}$$

Potential energy at final position

Negative of change in potential energy

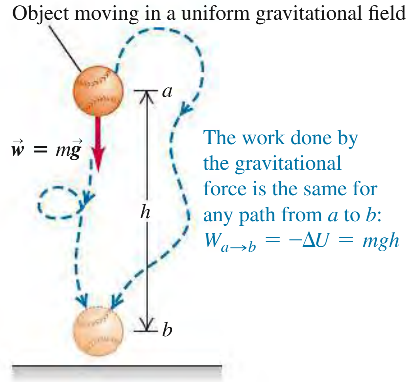
When $W_{a \to b}$ is positive, U_a is greater than U_b, ΔU is negative, and the potential energy *decreases*. That's what happens when a baseball falls from a high point (a) to a lower point (b) under the influence of the earth's gravity; the force of gravity does positive work, and the gravitational potential energy decreases (**Fig. 23.1**). When a tossed ball is moving upward, the gravitational force does negative work during the ascent, and the potential energy increases.

Third, the work–energy theorem says that the change in kinetic energy $\Delta K = K_b - K_a$ during a displacement equals the *total* work done on the particle. If only conservative forces do work, then Eq. (23.2) gives the total work, and $K_b - K_a = -(U_b - U_a)$. We usually write this as

$$K_a + U_a = K_b + U_b \tag{23.3}$$

That is, the total mechanical energy (kinetic plus potential) is *conserved* under these circumstances.

23.1 The work done on a baseball moving in a uniform gravitational field.



Electric Potential Energy in a Uniform Field

Let's look at an electrical example of these concepts. In **Fig. 23.2** a pair of charged parallel metal plates sets up a uniform, downward electric field with magnitude E. The field exerts a downward force with magnitude $F = q_0E$ on a positive test charge q_0. As the charge moves downward a distance d from point a to point b, the force on the test charge is constant and independent of its location. So the work done by the electric field is the product of the force magnitude and the component of displacement in the (downward) direction of the force:

$$W_{a \to b} = Fd = q_0Ed \tag{23.4}$$

This work is positive, since the force is in the same direction as the net displacement of the test charge.

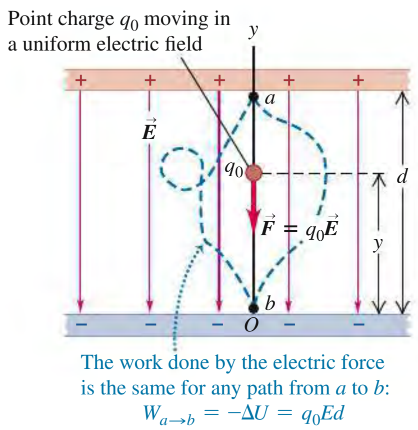
The y-component of the electric force, $F_y = -q_0E$, is constant, and there is no x- or z-component. This is exactly analogous to the gravitational force on a mass m near the earth's surface; for this force, there is a constant y-component $F_y = -mg$ and the x- and z-components are zero. Because of this analogy, we can conclude that the force exerted on q_0 by the uniform electric field in Fig. 23.2 is *conservative*, just as is the gravitational force. This means that the work $W_{a \to b}$ done by the field is independent of the path the particle takes from a to b. We can represent this work with a *potential-energy* function U, just as we did for gravitational potential energy in Section 7.1. The potential energy for the gravitational force $F_y = -mg$ was $U = mgy$; hence the potential energy for the electric force $F_y = -q_0E$ is

$$U = q_0Ey \tag{23.5}$$

When the test charge moves from height y_a to height y_b, the work done on the charge by the field is given by

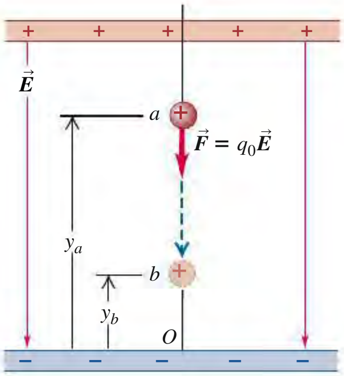
$$W_{a \to b} = -\Delta U = -(U_b - U_a) = -(q_0Ey_b - q_0Ey_a) = q_0E(y_a - y_b) \tag{23.6}$$

23.2 The work done on a point charge moving in a uniform electric field. Compare with Fig. 23.1.

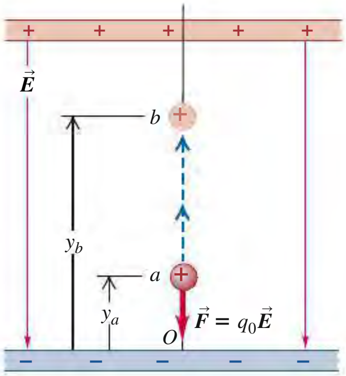


23.3 A positive charge moving (a) in the direction of the electric field $\vec{E}$ and (b) in the direction opposite $\vec{E}$.

(a) Positive charge q_0 moves in the direction of $\vec{E}$:
- Field does *positive* work on charge.
- *U decreases.*

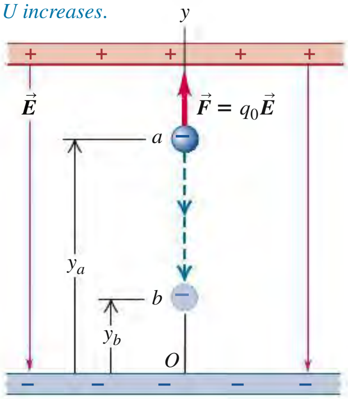


(b) Positive charge q_0 moves opposite $\vec{E}$:
- Field does *negative* work on charge.
- *U increases.*

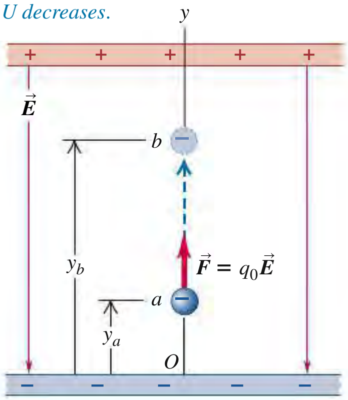


23.4 A negative charge moving (a) in the direction of the electric field $\vec{E}$ and (b) in the direction opposite $\vec{E}$. Compare with Fig. 23.3.

(a) Negative charge q_0 moves in the direction of $\vec{E}$:
- Field does *negative* work on charge.
- *U increases.*



(b) Negative charge q_0 moves opposite $\vec{E}$:
- Field does *positive* work on charge.
- *U decreases.*



When y_a is greater than y_b (**Fig. 23.3a**), the positive test charge q_0 moves downward, in the same direction as $\vec{E}$; the displacement is in the same direction as the force $\vec{F} = q_0\vec{E}$, so the field does positive work and U decreases. [In particular, if $y_a - y_b = d$ as in Fig. 23.2, Eq. (23.6) gives $W_{a \to b} = q_0Ed$, in agreement with Eq. (23.4).] When y_a is less than y_b (Fig. 23.3b), the positive test charge q_0 moves upward, in the opposite direction to $\vec{E}$; the displacement is opposite the force, the field does negative work, and U increases.

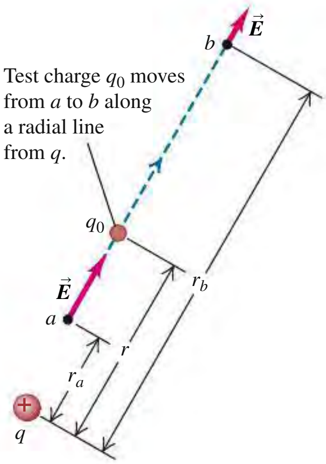
If the test charge q_0 is *negative*, the potential energy increases when it moves with the field and decreases when it moves against the field (**Fig. 23.4**).

Whether the test charge is positive or negative, the following general rules apply: *U increases* if the test charge q_0 moves in the direction *opposite* the electric force $\vec{F} = q_0\vec{E}$ (Figs. 23.3b and Fig. 23.4a); *U decreases* if q_0 moves in the *same* direction as $\vec{F} = q_0\vec{E}$ (Figs. 23.3a and Fig. 23.4b). This is the same behavior as for gravitational potential energy, which increases if a mass m moves upward (opposite the direction of the gravitational force) and decreases if m moves downward (in the same direction as the gravitational force).

CAUTION **Electric potential energy** The relationship between electric potential energy change and motion in an electric field is an important one that we'll use often, but that takes some effort to understand. Studying the preceding paragraph as well as Figs. 23.3 and Fig. 23.4 carefully now will help you tremendously later!

Electric Potential Energy of Two Point Charges

23.5 Test charge q_0 moves along a straight line extending radially from charge q. As it moves from a to b, the distance varies from r_a to r_b.



The idea of electric potential energy isn't restricted to the special case of a uniform electric field. Indeed, we can apply this concept to a point charge in *any* electric field caused by a static charge distribution. Recall from Chapter 21 that we can represent any charge distribution as a collection of point charges. Therefore it's useful to calculate the work done on a test charge q_0 moving in the electric field caused by a single, stationary point charge q.

We'll consider first a displacement along the *radial* line in **Fig. 23.5**. The force on q_0 is given by Coulomb's law, and its radial component is

$$F_r = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r^2} \tag{23.7}$$

If q and q_0 have the same sign $(+$ or $-)$, the force is repulsive and F_r is positive; if the two charges have opposite signs, the force is attractive and F_r is negative. The force is *not* constant during the displacement, and we must integrate to calculate the work $W_{a\to b}$ done on q_0 by this force as q_0 moves from a to b:

$$W_{a\to b} = \int_{r_a}^{r_b} F_r\,dr = \int_{r_a}^{r_b} \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r^2}\,dr = \frac{qq_0}{4\pi\epsilon_0}\left(\frac{1}{r_a} - \frac{1}{r_b}\right) \qquad (23.8)$$

The work done by the electric force for this path depends on only the endpoints.

Now let's consider a more general displacement (**Fig. 23.6**) in which a and b do *not* lie on the same radial line. From Eq. (23.1) the work done on q_0 during this displacement is given by

$$W_{a\to b} = \int_{r_a}^{r_b} F\cos\phi\,dl = \int_{r_a}^{r_b} \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r^2}\cos\phi\,dl$$

But Fig. 23.6 shows that $\cos\phi\,dl = dr$. That is, the work done during a small displacement $d\vec{l}$ depends only on the change dr in the distance r between the charges, which is the *radial component* of the displacement. Thus Eq. (23.8) is valid even for this more general displacement; the work done on q_0 by the electric field $\vec{E}$ produced by q depends only on r_a and r_b, not on the details of the path. Also, if q_0 returns to its starting point a by a different path, the total work done in the round-trip displacement is zero (the integral in Eq. (23.8) is from r_a back to r_a). These are the needed characteristics for a conservative force, as we defined it in Section 7.3. Thus the force on q_0 is a *conservative* force.

We see that Eqs. (23.2) and (23.8) are consistent if we define the potential energy to be $U_a = qq_0/4\pi\epsilon_0 r_a$ when q_0 is a distance r_a from q, and to be $U_b = qq_0/4\pi\epsilon_0 r_b$ when q_0 is a distance r_b from q. Thus

Electric potential energy of two point charges

$$U = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r} \qquad (23.9)$$

Values of two charges
Distance between two charges
Electric constant

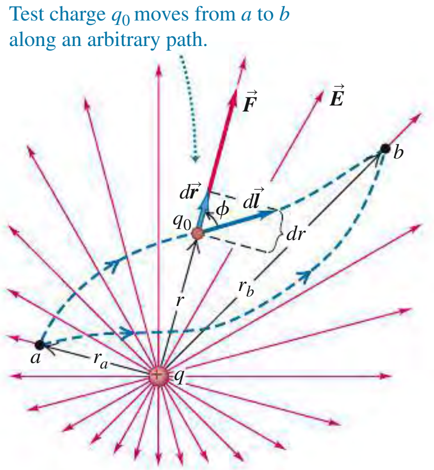
Equation (23.9) is valid no matter what the signs of the charges q and q_0. The potential energy is positive if the charges q and q_0 have the same sign (**Fig. 23.7a**) and negative if they have opposite signs (Fig. 23.7b).

Potential energy is always defined relative to some reference point where $U = 0$. In Eq. (23.9), U is zero when q and q_0 are infinitely far apart and $r = \infty$. Therefore U represents the work that would be done on the test charge q_0 by the field of q if q_0 moved from an initial distance r to infinity. If q and q_0 have the same sign, the interaction is repulsive, this work is positive, and U is positive at any finite separation (Fig. 23.7a). If the charges have opposite signs, the interaction is attractive, the work done is negative, and U is negative (Fig. 23.7b).

We emphasize that the potential energy U given by Eq. (23.9) is a *shared* property of the two charges. If the distance between q and q_0 is changed from r_a to r_b, the change in potential energy is the same whether q is held fixed and q_0 is moved or q_0 is held fixed and q is moved. For this reason, we never use the phrase "the electric potential energy *of* a point charge." (Likewise, if a mass m is at a height h above the earth's surface, the gravitational potential energy is a shared property of the mass m and the earth. We emphasized this in Sections 7.1 and 13.3.)

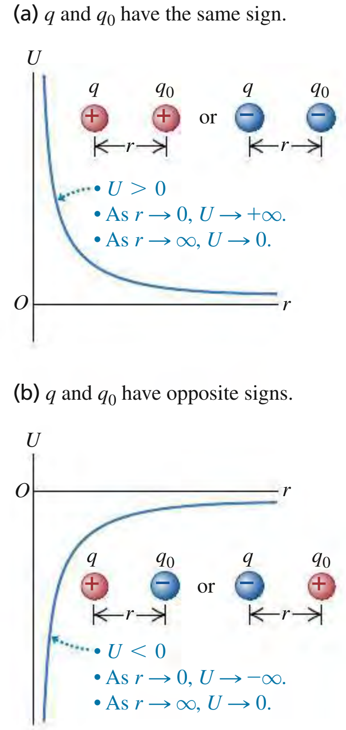
Equation (23.9) also holds if the charge q_0 is outside a spherically symmetric charge *distribution* with total charge q; the distance r is from q_0 to the center of the distribution. That's because Gauss's law tells us that the electric field outside such a distribution is the same as if all of its charge q were concentrated at its center (see Example 22.9 in Section 22.4).

23.6 The work done on charge q_0 by the electric field of charge q does not depend on the path taken, but only on the distances r_a and r_b.



CAUTION Electric potential energy vs. electric force Don't confuse Eq. (23.9) for the potential energy of two point charges with Eq. (23.7) for the radial component of the electric force that one charge exerts on the other. Potential energy U is proportional to $1/r$, while the force component F_r is proportional to $1/r^2$.

23.7 Graphs of the potential energy U of two point charges q and q_0 versus their separation r.



EXAMPLE 23.1 CONSERVATION OF ENERGY WITH ELECTRIC FORCES

A positron (the electron's antiparticle) has mass 9.11×10^{-31} kg and charge $q_0 = +e = +1.60 \times 10^{-19}$ C. Suppose a positron moves in the vicinity of an α (alpha) particle, which has charge $q = +2e = 3.20 \times 10^{-19}$ C and mass 6.64×10^{-27} kg. The α particle's mass is more than 7000 times that of the positron, so we assume that the α particle remains at rest. When the positron is 1.00×10^{-10} m from the α particle, it is moving directly away from the α particle at 3.00×10^6 m/s. (a) What is the positron's speed when the particles are 2.00×10^{-10} m apart? (b) What is the positron's speed when it is very far from the α particle? (c) Suppose the initial conditions are the same but the moving particle is an electron (with the same mass as the positron but charge $q_0 = -e$). Describe the subsequent motion.

SOLUTION

IDENTIFY and SET UP: The electric force between a positron (or an electron) and an α particle is conservative, so mechanical energy (kinetic plus potential) is conserved. Equation (23.9) gives the potential energy U at any separation r: The potential-energy function for parts (a) and (b) looks like that of Fig. 23.7a, and the function for part (c) looks like that of Fig. 23.7b. We are given the positron speed $v_a = 3.00 \times 10^6$ m/s when the separation between the particles is $r_a = 1.00 \times 10^{-10}$ m. In parts (a) and (b) we use Eqs. (23.3) and (23.9) to find the speed for $r = r_b = 2.00 \times 10^{-10}$ m and $r = r_c \to \infty$, respectively. In part (c) we replace the positron with an electron and reconsider the problem.

EXECUTE: (a) Both particles have positive charge, so the positron speeds up as it moves away from the α particle. From the energy-conservation equation, Eq. (23.3), the final kinetic energy is

$$K_b = \tfrac{1}{2}mv_b^2 = K_a + U_a - U_b$$

In this expression,

$$K_a = \tfrac{1}{2}mv_a^2 = \tfrac{1}{2}(9.11 \times 10^{-31}\ \text{kg})(3.00 \times 10^6\ \text{m/s})^2 = 4.10 \times 10^{-18}\ \text{J}$$

$$U_a = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r_a} = (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2) \times \frac{(3.20 \times 10^{-19}\ \text{C})(1.60 \times 10^{-19}\ \text{C})}{1.00 \times 10^{-10}\ \text{m}} = 4.61 \times 10^{-18}\ \text{J}$$

$$U_b = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r_b} = 2.30 \times 10^{-18}\ \text{J}$$

Hence the positron kinetic energy and speed at $r = r_b = 2.00 \times 10^{-10}$ m are

$$K_b = \tfrac{1}{2}mv_b^2 = 4.10 \times 10^{-18}\ \text{J} + 4.61 \times 10^{-18}\ \text{J} - 2.30 \times 10^{-18}\ \text{J} = 6.41 \times 10^{-18}\ \text{J}$$

$$v_b = \sqrt{\frac{2K_b}{m}} = \sqrt{\frac{2(6.41 \times 10^{-18}\ \text{J})}{9.11 \times 10^{-31}\ \text{kg}}} = 3.8 \times 10^6\ \text{m/s}$$

(b) When the positron and α particle are very far apart so that $r = r_c \to \infty$, the final potential energy U_c approaches zero. Again from energy conservation, the final kinetic energy and speed of the positron in this case are

$$K_c = K_a + U_a - U_c = 4.10 \times 10^{-18}\ \text{J} + 4.61 \times 10^{-18}\ \text{J} - 0 = 8.71 \times 10^{-18}\ \text{J}$$

$$v_c = \sqrt{\frac{2K_c}{m}} = \sqrt{\frac{2(8.71 \times 10^{-18}\ \text{J})}{9.11 \times 10^{-31}\ \text{kg}}} = 4.4 \times 10^6\ \text{m/s}$$

(c) The electron and α particle have opposite charges, so the force is attractive and the electron slows down as it moves away. Changing the moving particle's sign from $+e$ to $-e$ means that the initial potential energy is now $U_a = -4.61 \times 10^{-18}$ J, which makes the total mechanical energy *negative*:

$$K_a + U_a = (4.10 \times 10^{-18}\ \text{J}) - (4.61 \times 10^{-18}\ \text{J}) = -0.51 \times 10^{-18}\ \text{J}$$

The total mechanical energy would have to be positive for the electron to move infinitely far away from the α particle. Like a rock thrown upward at low speed from the earth's surface, it will reach a maximum separation $r = r_d$ from the α particle before reversing direction. At this point its speed and its kinetic energy K_d are zero, so at separation r_d we have

$$U_d = K_a + U_a - K_d = (-0.51 \times 10^{-18}\ \text{J}) - 0$$

$$U_d = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r_d} = -0.51 \times 10^{-18}\ \text{J}$$

$$r_d = \frac{1}{U_d}\frac{qq_0}{4\pi\epsilon_0} = \frac{(9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)}{-0.51 \times 10^{-18}\ \text{J}}(3.20 \times 10^{-19}\ \text{C})(-1.60 \times 10^{-19}\ \text{C}) = 9.0 \times 10^{-10}\ \text{m}$$

For $r_b = 2.00 \times 10^{-10}$ m we have $U_b = -2.30 \times 10^{-18}$ J, so the electron kinetic energy and speed at this point are

$$K_b = \tfrac{1}{2}mv_b^2 = 4.10 \times 10^{-18}\ \text{J} + (-4.61 \times 10^{-18}\ \text{J}) - (-2.30 \times 10^{-18}\ \text{J}) = 1.79 \times 10^{-18}\ \text{J}$$

$$v_b = \sqrt{\frac{2K_b}{m}} = \sqrt{\frac{2(1.79 \times 10^{-18}\ \text{J})}{9.11 \times 10^{-31}\ \text{kg}}} = 2.0 \times 10^6\ \text{m/s}$$

EVALUATE: Both particles behave as expected as they move away from the α particle: The positron speeds up, and the electron slows down and eventually turns around. How fast would an electron have to be moving at $r_a = 1.00 \times 10^{-10}$ m to travel infinitely far from the α particle? (*Hint:* See Example 13.5 in Section 13.3.)

Electric Potential Energy with Several Point Charges

Suppose the electric field $\vec{E}$ in which charge q_0 moves is caused by *several* point charges $q_1, q_2, q_3, \ldots$ at distances $r_1, r_2, r_3, \ldots$ from q_0, as in **Fig. 23.8**. For example, q_0 could be a positive ion moving in the presence of other ions (**Fig. 23.9**). The total electric field at each point is the *vector sum* of the fields due to the individual charges, and the total work done on q_0 during any displacement is the sum of the contributions from the individual charges. From Eq. (23.9) we conclude that the potential energy associated with the test charge q_0 at point a in Fig. 23.8 is the *algebraic* sum (*not* a vector sum):

Electric potential energy of point charge q_0 and collection of charges $q_1, q_2, q_3, \ldots$

$$U = \frac{q_0}{4\pi\epsilon_0}\left(\frac{q_1}{r_1} + \frac{q_2}{r_2} + \frac{q_3}{r_3} + \cdots\right) = \frac{q_0}{4\pi\epsilon_0}\sum_i \frac{q_i}{r_i} \tag{23.10}$$

Electric constant   Distances from q_0 to $q_1, q_2, q_3, \ldots$

When q_0 is at a different point b, the potential energy is given by the same expression, but $r_1, r_2, \ldots$ are the distances from $q_1, q_2, \ldots$ to point b. The work done on charge q_0 when it moves from a to b along any path is equal to the difference $U_a - U_b$ between the potential energies when q_0 is at a and at b.

We can represent *any* charge distribution as a collection of point charges, so Eq. (23.10) shows that we can always find a potential-energy function for *any* static electric field. It follows that **for every electric field due to a static charge distribution, the force exerted by that field is conservative.**

Equations (23.9) and (23.10) define U to be zero when distances $r_1, r_2, \ldots$ are infinite—that is, when the test charge q_0 is very far away from all the charges that produce the field. As with any potential-energy function, the point where $U = 0$ is arbitrary; we can always add a constant to make U equal zero at any point we choose. In electrostatics problems it's usually simplest to choose this point to be at infinity. When we analyze electric circuits in Chapters 25 and 26, other choices will be more convenient.

Equation (23.10) gives the potential energy associated with the presence of the test charge q_0 in the $\vec{E}$ field produced by $q_1, q_2, q_3, \ldots$. But there is also potential energy involved in assembling these charges. If we start with charges $q_1, q_2, q_3, \ldots$ all separated from each other by infinite distances and then bring them together so that the distance between q_i and q_j is r_{ij}, the *total* potential energy U is the sum of the potential energies of interaction for each pair of charges. We can write this as

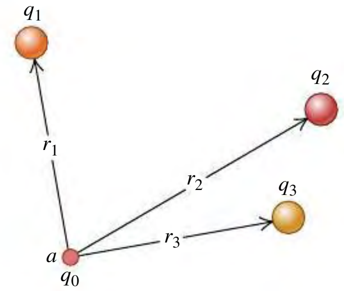
$$U = \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{q_i q_j}{r_{ij}} \tag{23.11}$$

This sum extends over all *pairs* of charges; we don't let $i = j$ (because that would be an interaction of a charge with itself), and we include only terms with $i < j$ to make sure that we count each pair only once. Thus, to account for the interaction between q_3 and q_4, we include a term with $i = 3$ and $j = 4$ but not a term with $i = 4$ and $j = 3$.

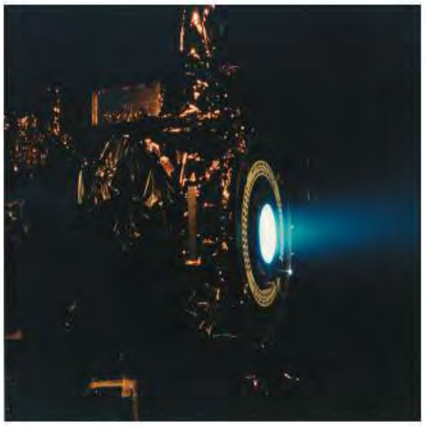
Interpreting Electric Potential Energy

As a final comment, here are two viewpoints on electric potential energy. We have defined it in terms of the work done *by the electric field* on a charged particle moving in the field, just as in Chapter 7 we defined potential energy in terms of the work done by gravity or by a spring. When a particle moves from point a to point b, the work done on it by the electric field is $W_{a\to b} = U_a - U_b$. Thus the potential-energy difference $U_a - U_b$ equals *the work that is done by the electric*

23.8 The potential energy associated with a charge q_0 at point a depends on the other charges q_1, q_2, and q_3 and on their distances r_1, r_2, and r_3 from point a.



23.9 This ion engine for spacecraft uses electric forces to eject a stream of positive xenon ions (Xe^+) at speeds in excess of 30 km/s. The thrust produced is very low (about 0.09 newton) but can be maintained continuously for days, in contrast to chemical rockets, which produce a large thrust for a short time (see Fig. 8.34). Such ion engines have been used for maneuvering interplanetary spacecraft.



force when the particle moves from a to b. When U_a is greater than U_b, the field does positive work on the particle as it "falls" from a point of higher potential energy (a) to a point of lower potential energy (b).

An alternative but equivalent viewpoint is to consider how much work we would have to do to "raise" a particle from a point b where the potential energy is U_b to a point a where it has a greater value U_a (pushing two positive charges closer together, for example). To move the particle slowly (so as not to give it any kinetic energy), we need to exert an additional external force $\vec{F}_{\text{ext}}$ that is equal and opposite to the electric-field force and does positive work. The potential-energy difference $U_a - U_b$ is then defined as *the work that must be done by an external force to move the particle slowly from b to a against the electric force.* Because $\vec{F}_{\text{ext}}$ is the negative of the electric-field force and the displacement is in the opposite direction, this definition of the potential difference $U_a - U_b$ is equivalent to that given above. This alternative viewpoint also works if U_a is less than U_b, corresponding to "lowering" the particle; an example is moving two positive charges away from each other. In this case, $U_a - U_b$ is again equal to the work done by the external force, but now this work is negative.

We will use both of these viewpoints in the next section to interpret what is meant by electric *potential,* or potential energy per unit charge.

SOLUTION

EXAMPLE 23.2 A SYSTEM OF POINT CHARGES

Two point charges are located on the x-axis, $q_1 = -e$ at $x = 0$ and $q_2 = +e$ at $x = a$. (a) Find the work that must be done by an external force to bring a third point charge $q_3 = +e$ from infinity to $x = 2a$. (b) Find the total potential energy of the system of three charges.

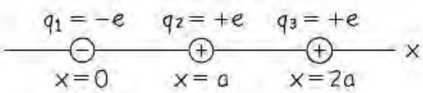
SOLUTION

IDENTIFY and SET UP: **Figure 23.10** shows the final arrangement of the three charges. In part (a) we need to find the work W that must be done on q_3 by an external force $\vec{F}_{\text{ext}}$ to bring q_3 in from infinity to $x = 2a$. We do this by using Eq. (23.10) to find the potential energy associated with q_3 in the presence of q_1 and q_2. In part (b) we use Eq. (23.11), the expression for the potential energy of a collection of point charges, to find the total potential energy of the system.

EXECUTE: (a) The work W equals the difference between (i) the potential energy U associated with q_3 when it is at $x = 2a$ and (ii) the potential energy when it is infinitely far away. The second of these is zero, so the work required is equal to U. The distances between the charges are $r_{13} = 2a$ and $r_{23} = a$, so from Eq. (23.10),

$$W = U = \frac{q_3}{4\pi\epsilon_0}\left(\frac{q_1}{r_{13}} + \frac{q_2}{r_{23}}\right) = \frac{+e}{4\pi\epsilon_0}\left(\frac{-e}{2a} + \frac{+e}{a}\right) = \frac{+e^2}{8\pi\epsilon_0 a}$$

23.10 Our sketch of the situation after the third charge has been brought in from infinity.



This is positive, just as we should expect. If we bring q_3 in from infinity along the $+x$-axis, it is attracted by q_1 but is repelled more strongly by q_2. Hence we must do positive work to push q_3 to the position at $x = 2a$.

(b) From Eq. (23.11), the total potential energy of the three-charge system is

$$U = \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{q_i q_j}{r_{ij}} = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right)$$

$$= \frac{1}{4\pi\epsilon_0}\left[\frac{(-e)(e)}{a} + \frac{(-e)(e)}{2a} + \frac{(e)(e)}{a}\right] = \frac{-e^2}{8\pi\epsilon_0 a}$$

EVALUATE: Our negative result in part (b) means that the system has lower potential energy than it would if the three charges were infinitely far apart. An external force would have to do *negative* work to bring the three charges from infinity to assemble this entire arrangement and would have to do *positive* work to move the three charges back to infinity.

TEST YOUR UNDERSTANDING OF SECTION 23.1 Consider the system of three point charges in Example 21.4 (Section 21.3) and shown in Fig. 21.14. (a) What is the sign of the total potential energy of this system? (i) Positive; (ii) negative; (iii) zero. (b) What is the sign of the total amount of work you would have to do to move these charges infinitely far from each other? (i) Positive; (ii) negative; (iii) zero. ❙

23.2 ELECTRIC POTENTIAL

PhET: Charges and Fields

In Section 23.1 we looked at the potential energy U associated with a test charge q_0 in an electric field. Now we want to describe this potential energy on a "per unit charge" basis, just as electric field describes the force per unit charge on a charged particle in the field. This leads us to the concept of *electric potential,* often called simply *potential*. This concept is very useful in calculations involving energies of charged particles. It also facilitates many electric-field calculations because electric potential is closely related to the electric field $\vec{E}$. When we need to determine an electric field, it is often easier to determine the potential first and then find the field from it.

Potential is *potential energy per unit charge*. We define the potential V at any point in an electric field as the potential energy U *per unit charge* associated with a test charge q_0 at that point:

$$V = \frac{U}{q_0} \quad \text{or} \quad U = q_0 V \tag{23.12}$$

Potential energy and charge are both scalars, so potential is a scalar. From Eq. (23.12) its units are the units of energy divided by those of charge. The SI unit of potential, called one **volt** (1 V) in honor of the Italian electrical experimenter Alessandro Volta (1745–1827), equals 1 joule per coulomb:

$$1\text{ V} = 1\text{ volt} = 1\text{ J/C} = 1\text{ joule/coulomb}$$

Let's put Eq. (23.2), which equates the work done by the electric force during a displacement from a to b to the quantity $-\Delta U = -(U_b - U_a)$, on a "work per unit charge" basis. We divide this equation by q_0, obtaining

$$\frac{W_{a\to b}}{q_0} = -\frac{\Delta U}{q_0} = -\left(\frac{U_b}{q_0} - \frac{U_a}{q_0}\right) = -(V_b - V_a) = V_a - V_b \tag{23.13}$$

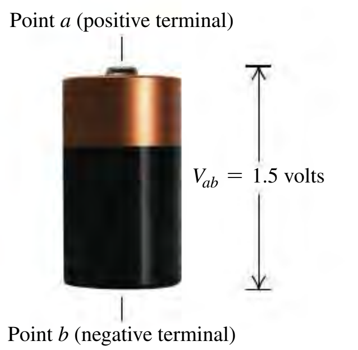
where $V_a = U_a/q_0$ is the potential energy per unit charge at point a and similarly for V_b. We call V_a and V_b the *potential at point a* and *potential at point b*, respectively. Thus the work done per unit charge by the electric force when a charged body moves from a to b is equal to the potential at a minus the potential at b.

The difference $V_a - V_b$ is called the *potential of a with respect to b*; we sometimes abbreviate this difference as $V_{ab} = V_a - V_b$ (note the order of the subscripts). This is often called the potential difference between a and b, but that's ambiguous unless we specify which is the reference point. In electric circuits, which we will analyze in later chapters, the potential difference between two points is often called **voltage** (**Fig. 23.11**). Equation (23.13) then states: **V_{ab}, the potential (in V) of a with respect to b, equals the work (in J) done by the electric force when a UNIT (1-C) charge moves from a to b.**

Another way to interpret the potential difference V_{ab} in Eq. (23.13) is to use the alternative viewpoint mentioned at the end of Section 23.1. In that viewpoint, $U_a - U_b$ is the amount of work that must be done by an *external* force to move a particle of charge q_0 slowly from b to a against the electric force. The work that must be done *per unit charge* by the external force is then $(U_a - U_b)/q_0 = V_a - V_b = V_{ab}$. In other words: **$V_{ab}$, the potential (in V) of a with respect to b, equals the work (in J) that must be done to move a UNIT (1-C) charge slowly from b to a against the electric force.**

An instrument that measures the difference of potential between two points is called a *voltmeter.* (In Chapter 26 we'll discuss how these devices work.) Voltmeters that can measure a potential difference of 1 μV are common, and sensitivities down to 10^{-12} V can be attained.

23.11 The voltage of this battery equals the difference in potential $V_{ab} = V_a - V_b$ between its positive terminal (point a) and its negative terminal (point b).



BIO Application Electrocardiography The electrodes used in an electrocardiogram—EKG or ECG for short—measure the potential differences (typically no greater than 1 mV = 10^{-3} V) between different parts of the patient's skin. These are indicative of the potential differences between regions of the heart, and so provide a sensitive way to detect any abnormalities in the electrical activity that drives cardiac function.

Calculating Electric Potential

To find the potential V due to a single point charge q, we divide Eq. (23.9) by q_0:

Electric potential due to a point charge

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{r} \tag{23.14}$$

Value of point charge (q); Distance from point charge to where potential is measured (r); Electric constant (ϵ_0)

If q is positive, the potential that it produces is positive at all points; if q is negative, it produces a potential that is negative everywhere. In either case, V is equal to zero at $r = \infty$, an infinite distance from the point charge. Note that potential, like electric field, is independent of the test charge q_0 that we use to define it.

Similarly, we divide Eq. (23.10) by q_0 to find the potential due to a collection of point charges:

Electric potential due to a collection of point charges

$$V = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{r_i} \tag{23.15}$$

Value of ith point charge (q_i); Distance from ith point charge to where potential is measured (r_i); Electric constant (ϵ_0)

Just as the electric field due to a collection of point charges is the *vector* sum of the fields produced by each charge, the electric potential due to a collection of point charges is the *scalar* sum of the potentials due to each charge. When we have a continuous distribution of charge along a line, over a surface, or through a volume, we divide the charge into elements dq, and the sum in Eq. (23.15) becomes an integral:

Electric potential due to a continuous distribution of charge

$$V = \frac{1}{4\pi\epsilon_0}\int \frac{dq}{r} \tag{23.16}$$

Integral over charge distribution; Charge element (dq); Distance from charge element to where potential is measured (r); Electric constant (ϵ_0)

We'll work out several examples of such cases. The potential defined by Eqs. (23.15) and (23.16) is zero at points that are infinitely far away from *all* the charges. Later we'll encounter cases in which the charge distribution itself extends to infinity. We'll find that in such cases we cannot set $V = 0$ at infinity, and we'll need to exercise care in using and interpreting Eqs. (23.15) and (23.16).

CAUTION What is electric potential? Before getting too involved in the details of how to calculate electric potential, remind yourself what potential is. The electric *potential* at a certain point is the potential energy per *unit* charge placed at that point. That's why potential is measured in joules per coulomb, or volts. Keep in mind, too, that there doesn't have to be a charge at a given point for a potential V to exist at that point. (In the same way, an electric field can exist at a given point even if there's no charge there to respond to it.)

Finding Electric Potential from Electric Field

When we are given a collection of point charges, Eq. (23.15) is usually the easiest way to calculate the potential V. But in some problems in which the electric field is known or can be found easily, it is easier to determine V from $\vec{E}$. The force $\vec{F}$ on a test charge q_0 can be written as $\vec{F} = q_0\vec{E}$, so from Eq. (23.1) the work done by the electric force as the test charge moves from a to b is given by

$$W_{a\to b} = \int_a^b \vec{F}\cdot d\vec{l} = \int_a^b q_0\vec{E}\cdot d\vec{l}$$

If we divide this by q_0 and compare the result with Eq. (23.13), we find

$$\underbrace{V_a - V_b}_{\text{Electric potential difference}} = \overbrace{\underbrace{\int_a^b \vec{E}\cdot d\vec{l}}_{\text{Scalar product of electric field and displacement vector}} = \int_a^b \underbrace{E}_{\text{Electric-field magnitude}} \underbrace{\cos\phi}_{\text{Angle between } \vec{E} \text{ and } d\vec{l}} \underbrace{dl}_{\text{Displacement}}}^{\text{Integral along path from } a \text{ to } b} \tag{23.17}$$

The value of $V_a - V_b$ is independent of the path taken from a to b, just as the value of $W_{a\rightarrow b}$ is independent of the path. To interpret Eq. (23.17), remember that $\vec{E}$ is the electric force per unit charge on a test charge. If the line integral $\int_a^b \vec{E}\cdot d\vec{l}$ is positive, the electric field does positive work on a positive test charge as it moves from a to b. In this case the electric potential energy decreases as the test charge moves, so the potential energy per unit charge decreases as well; hence V_b is less than V_a and $V_a - V_b$ is positive.

As an illustration, consider a positive point charge (**Fig. 23.12a**). The electric field is directed away from the charge, and $V = q/4\pi\epsilon_0 r$ is positive at any finite distance from the charge. If you move away from the charge, in the direction of $\vec{E}$, you move toward lower values of V; if you move toward the charge, in the direction opposite $\vec{E}$, you move toward greater values of V. For the negative point charge in Fig. 23.12b, $\vec{E}$ is directed toward the charge and $V = q/4\pi\epsilon_0 r$ is negative at any finite distance from the charge. In this case, if you move toward the charge, you are moving in the direction of $\vec{E}$ and in the direction of decreasing (more negative) V. Moving away from the charge, in the direction opposite $\vec{E}$, moves you toward increasing (less negative) values of V. The general rule, valid for *any* electric field, is: Moving *with* the direction of $\vec{E}$ means moving in the direction of *decreasing* V, and moving *against* the direction of $\vec{E}$ means moving in the direction of *increasing* V.

Also, a positive test charge q_0 experiences an electric force in the direction of $\vec{E}$, toward lower values of V; a negative test charge experiences a force opposite $\vec{E}$, toward higher values of V. Thus a positive charge tends to "fall" from a high-potential region to a lower-potential region. The opposite is true for a negative charge.

Notice that Eq. (23.17) can be rewritten as

$$V_a - V_b = -\int_b^a \vec{E}\cdot d\vec{l} \tag{23.18}$$

This has a negative sign compared to the integral in Eq. (23.17), and the limits are reversed; hence Eqs. (23.17) and (23.18) are equivalent. But Eq. (23.18) has a slightly different interpretation. To move a unit charge slowly against the electric force, we must apply an *external* force per unit charge equal to $-\vec{E}$, equal and opposite to the electric force per unit charge $\vec{E}$. Equation (23.18) says that $V_a - V_b = V_{ab}$, the potential of a with respect to b, equals the work done per unit charge by this external force to move a unit charge from b to a. This is the same alternative interpretation we discussed under Eq. (23.13).

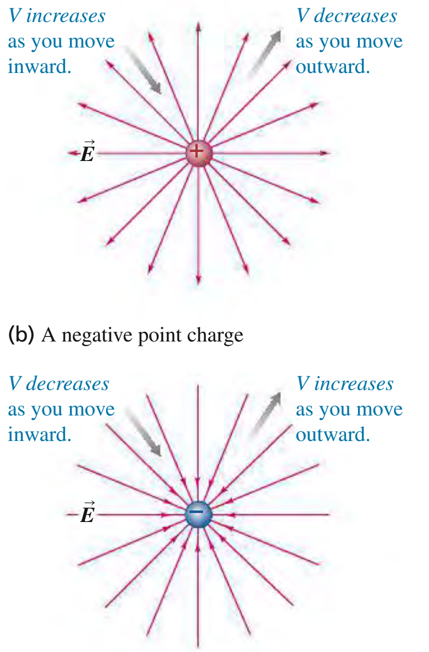
Equations (23.17) and (23.18) show that the unit of potential difference (1 V) is equal to the unit of electric field (1 N/C) multiplied by the unit of distance (1 m). Hence the unit of electric field can be expressed as 1 *volt per meter* (1 V/m), as well as 1 N/C:

$$1\text{ V/m} = 1\text{ volt/meter} = 1\text{ N/C} = 1\text{ newton/coulomb}$$

In practice, the volt per meter is the usual unit of electric-field magnitude.

23.12 If you move in the direction of $\vec{E}$, electric potential V decreases; if you move in the direction opposite $\vec{E}$, V increases.

(a) A positive point charge



DATA *SPEAKS*

Electric Potential and Electric Potential Energy

When students were given a problem involving electric potential and electric potential energy, more than 20% gave an incorrect response. Common errors:

- Confusing potential energy and potential. Electric potential energy is measured in joules; electric potential is potential energy *per unit charge* and is measured in volts.
- Forgetting the relationships among electric potential V, electric field $\vec{E}$, and electric force. $\vec{E}$ always points from regions of high V toward regions of low V; the direction of the electric force on a point charge q is in the direction of $\vec{E}$ if $q > 0$ but opposite $\vec{E}$ if $q < 0$.

BIO Application Electron Volts and Cancer Radiotherapy One way to destroy a cancerous tumor is to aim high-energy electrons directly at it. Each electron has a kinetic energy of 4 to 20 million electron volts, or MeV (1 MeV $= 10^6$ eV), and transfers its energy to the tumor through collisions with the tumor's atoms. Electrons in this energy range can penetrate only a few centimeters into a patient, which makes them useful for treating superficial tumors, such as those on the skin or lips.

Electron Volts

The magnitude e of the electron charge can be used to define a unit of energy that is useful in many calculations with atomic and nuclear systems. When a particle with charge q moves from a point where the potential is V_b to a point where it is V_a, the change in the potential energy U is

$$U_a - U_b = q(V_a - V_b) = qV_{ab}$$

If charge q equals the magnitude e of the electron charge, 1.602×10^{-19} C, and the potential difference is $V_{ab} = 1$ V, the change in energy is

$$U_a - U_b = (1.602 \times 10^{-19}\text{ C})(1\text{ V}) = 1.602 \times 10^{-19}\text{ J}$$

This quantity of energy is defined to be 1 **electron volt** (1 eV):

$$1\text{ eV} = 1.602 \times 10^{-19}\text{ J}$$

The multiples meV, keV, MeV, GeV, and TeV are often used.

CAUTION Electron volts vs. volts Remember that the electron volt is a unit of energy, *not* a unit of potential or potential difference!

When a particle with charge e moves through a potential difference of 1 volt, the change in potential *energy* is 1 eV. If the charge is some multiple of e—say, Ne—the change in potential energy in electron volts is N times the potential difference in volts. For example, when an alpha particle, which has charge $2e$, moves between two points with a potential difference of 1000 V, the change in potential energy is $2(1000\text{ eV}) = 2000$ eV. To confirm this, we write

$$\begin{aligned} U_a - U_b &= qV_{ab} = (2e)(1000\text{ V}) = (2)(1.602 \times 10^{-19}\text{ C})(1000\text{ V}) \\ &= 3.204 \times 10^{-16}\text{ J} = 2000\text{ eV} \end{aligned}$$

Although we defined the electron volt in terms of *potential* energy, we can use it for *any* form of energy, such as the kinetic energy of a moving particle. When we speak of a "one-million-electron-volt proton," we mean a proton with a kinetic energy of one million electron volts (1 MeV), equal to $(10^6)(1.602 \times 10^{-19}\text{ J}) = 1.602 \times 10^{-13}$ J. The Large Hadron Collider near Geneva, Switzerland, is designed to accelerate protons to a kinetic energy of 7 TeV (7×10^{12} eV).

EXAMPLE 23.3 ELECTRIC FORCE AND ELECTRIC POTENTIAL

SOLUTION

A proton (charge $+e = 1.602 \times 10^{-19}$ C) moves a distance $d = 0.50$ m in a straight line between points a and b in a linear accelerator. The electric field is uniform along this line, with magnitude $E = 1.5 \times 10^7$ V/m $= 1.5 \times 10^7$ N/C in the direction from a to b. Determine (a) the force on the proton; (b) the work done on it by the field; (c) the potential difference $V_a - V_b$.

SOLUTION

IDENTIFY and SET UP: This problem uses the relationship between electric field and electric force. It also uses the relationship among force, work, and potential-energy difference. We are given the electric field, so it is straightforward to find the electric force on the proton. Calculating the work is also straightforward because $\vec{E}$ is uniform, so the force on the proton is constant. Once the work is known, we find $V_a - V_b$ from Eq. (23.13).

EXECUTE: (a) The force on the proton is in the same direction as the electric field, and its magnitude is

$$\begin{aligned} F = qE &= (1.602 \times 10^{-19}\text{ C})(1.5 \times 10^7\text{ N/C}) \\ &= 2.4 \times 10^{-12}\text{ N} \end{aligned}$$

(b) The force is constant and in the same direction as the displacement, so the work done on the proton is

$$\begin{aligned} W_{a\to b} = Fd &= (2.4 \times 10^{-12}\text{ N})(0.50\text{ m}) \\ &= 1.2 \times 10^{-12}\text{ J} \\ &= (1.2 \times 10^{-12}\text{ J})\frac{1\text{ eV}}{1.602 \times 10^{-19}\text{ J}} \\ &= 7.5 \times 10^6\text{ eV} = 7.5\text{ MeV} \end{aligned}$$

(c) From Eq. (23.13) the potential difference is the work per unit charge, which is

$$V_a - V_b = \frac{W_{a\to b}}{q} = \frac{1.2 \times 10^{-12}\ \text{J}}{1.602 \times 10^{-19}\ \text{C}}$$
$$= 7.5 \times 10^6\ \text{J/C} = 7.5 \times 10^6\ \text{V} = 7.5\ \text{MV}$$

We can get this same result even more easily by remembering that 1 electron volt equals 1 volt multiplied by the charge e. The work done is 7.5×10^6 eV and the charge is e, so the potential difference is $(7.5 \times 10^6\ \text{eV})/e = 7.5 \times 10^6$ V.

EVALUATE: We can check our result in part (c) by using Eq. (23.17) or Eq. (23.18). The angle ϕ between the constant field $\vec{E}$ and the displacement is zero, so Eq. (23.17) becomes

$$V_a - V_b = \int_a^b E\cos\phi\, dl = \int_a^b E\, dl = E\int_a^b dl$$

The integral of dl from a to b is just the distance d, so we again find

$$V_a - V_b = Ed = (1.5 \times 10^7\ \text{V/m})(0.50\ \text{m}) = 7.5 \times 10^6\ \text{V}$$

EXAMPLE 23.4 POTENTIAL DUE TO TWO POINT CHARGES

SOLUTION

An electric dipole consists of point charges $q_1 = +12$ nC and $q_2 = -12$ nC placed 10.0 cm apart (**Fig. 23.13**). Compute the electric potentials at points a, b, and c.

SOLUTION

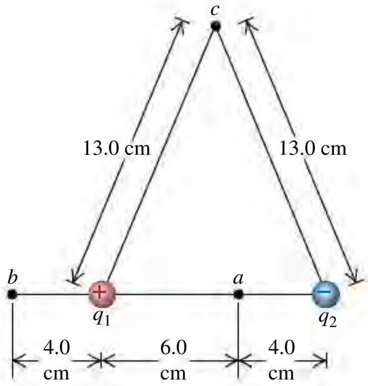
IDENTIFY and SET UP: This is the same arrangement as in Example 21.8, in which we calculated the electric *field* at each point by doing a *vector* sum. Here our target variable is the electric *potential V* at three points, which we find by doing the *algebraic* sum in Eq. (23.15).

EXECUTE: At point a we have $r_1 = 0.060$ m and $r_2 = 0.040$ m, so Eq. (23.15) becomes

$$V_a = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{r_i} = \frac{1}{4\pi\epsilon_0}\frac{q_1}{r_1} + \frac{1}{4\pi\epsilon_0}\frac{q_2}{r_2}$$
$$= (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{12 \times 10^{-9}\ \text{C}}{0.060\ \text{m}}$$
$$+ (9.0 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{(-12 \times 10^{-9}\ \text{C})}{0.040\ \text{m}}$$
$$= 1800\ \text{N}\cdot\text{m/C} + (-2700\ \text{N}\cdot\text{m/C})$$
$$= 1800\ \text{V} + (-2700\ \text{V}) = -900\ \text{V}$$

In a similar way you can show that the potential at point b (where $r_1 = 0.040$ m and $r_2 = 0.140$ m) is $V_b = 1930$ V and that the potential at point c (where $r_1 = r_2 = 0.130$ m) is $V_c = 0$.

23.13 What are the potentials at points a, b, and c due to this electric dipole?



EVALUATE: Let's confirm that these results make sense. Point a is closer to the −12-nC charge than to the +12-nC charge, so the potential at a is negative. The potential is positive at point b, which is closer to the +12-nC charge than the −12-nC charge. Finally, point c is equidistant from the +12-nC charge and the −12-nC charge, so the potential there is zero. (The potential is also equal to zero at a point infinitely far from both charges.)

Comparing this example with Example 21.8 shows that it's much easier to calculate electric potential (a scalar) than electric field (a vector). We'll take advantage of this simplification whenever possible.

EXAMPLE 23.5 POTENTIAL AND POTENTIAL ENERGY

SOLUTION

Compute the potential energy associated with a +4.0-nC point charge if it is placed at points a, b, and c in Fig. 23.13.

SOLUTION

IDENTIFY and SET UP: The potential energy U associated with a point charge q at a location where the electric potential is V is $U = qV$. We use the values of V from Example 23.4.

EXECUTE: At the three points we find

$$U_a = qV_a = (4.0 \times 10^{-9}\ \text{C})(-900\ \text{J/C}) = -3.6 \times 10^{-6}\ \text{J}$$
$$U_b = qV_b = (4.0 \times 10^{-9}\ \text{C})(1930\ \text{J/C}) = 7.7 \times 10^{-6}\ \text{J}$$
$$U_c = qV_c = 0$$

All of these values correspond to U and V being zero at infinity.

EVALUATE: Note that *zero* net work is done on the 4.0-nC charge if it moves from point c to infinity *by any path*. In particular, let the path be along the perpendicular bisector of the line joining the other two charges q_1 and q_2 in Fig. 23.13. As shown in Example 21.8 (Section 21.5), at points on the bisector, the direction of $\vec{E}$ is perpendicular to the bisector. Hence the force on the 4.0-nC charge is perpendicular to the path, and no work is done in any displacement along it.

EXAMPLE 23.6 FINDING POTENTIAL BY INTEGRATION

By integrating the electric field as in Eq. (23.17), find the potential at a distance r from a point charge q.

SOLUTION

IDENTIFY and SET UP: We let point a in Eq. (23.17) be at distance r and let point b be at infinity (**Fig. 23.14**). As usual, we choose the potential to be zero at an infinite distance from the charge q.

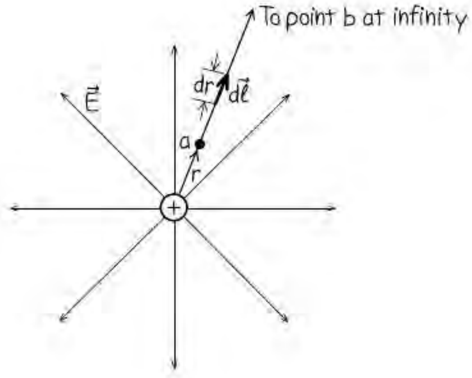
EXECUTE: To carry out the integral, we can choose any path we like between points a and b. The most convenient path is a radial line as shown in Fig. 23.14, so that $d\vec{l}$ is in the radial direction and has magnitude dr. Writing $d\vec{l} = \hat{r}dr$, we have from Eq. (23.17)

$$V - 0 = V = \int_r^\infty \vec{E} \cdot d\vec{l}$$

$$= \int_r^\infty \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r} \cdot \hat{r}dr = \int_r^\infty \frac{q}{4\pi\epsilon_0 r^2}dr$$

$$= -\frac{q}{4\pi\epsilon_0 r}\bigg|_r^\infty = 0 - \left(-\frac{q}{4\pi\epsilon_0 r}\right) = \frac{q}{4\pi\epsilon_0 r}$$

23.14 Calculating the potential by integrating $\vec{E}$ for a single point charge.



EVALUATE: Our result agrees with Eq. (23.14) and is correct for positive or negative q.

EXAMPLE 23.7 MOVING THROUGH A POTENTIAL DIFFERENCE

In **Fig. 23.15** a dust particle with mass $m = 5.0 \times 10^{-9}$ kg $=$ 5.0 μg and charge $q_0 = 2.0$ nC starts from rest and moves in a straight line from point a to point b. What is its speed v at point b?

SOLUTION

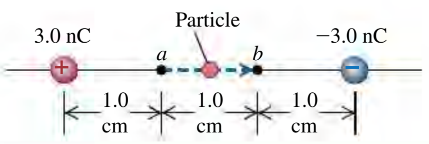
IDENTIFY and SET UP: Only the conservative electric force acts on the particle, so mechanical energy is conserved: $K_a + U_a = K_b + U_b$. We get the potential energies U from the corresponding potentials V from Eq. (23.12): $U_a = q_0V_a$ and $U_b = q_0V_b$.

EXECUTE: We have $K_a = 0$ and $K_b = \frac{1}{2}mv^2$. We substitute these and our expressions for U_a and U_b into the energy-conservation equation, then solve for v. We find

$$0 + q_0V_a = \tfrac{1}{2}mv^2 + q_0V_b$$

$$v = \sqrt{\frac{2q_0(V_a - V_b)}{m}}$$

23.15 The particle moves from point a to point b; its acceleration is not constant.



We calculate the potentials from Eq. (23.15), $V = q/4\pi\epsilon_0 r$:

$$V_a = (9.0 \times 10^9\ \text{N} \cdot \text{m}^2/\text{C}^2) \times \left(\frac{3.0 \times 10^{-9}\ \text{C}}{0.010\ \text{m}} + \frac{(-3.0 \times 10^{-9}\ \text{C})}{0.020\ \text{m}}\right)$$

$$= 1350\ \text{V}$$

$$V_b = (9.0 \times 10^9\ \text{N} \cdot \text{m}^2/\text{C}^2) \times \left(\frac{3.0 \times 10^{-9}\ \text{C}}{0.020\ \text{m}} + \frac{(-3.0 \times 10^{-9}\ \text{C})}{0.010\ \text{m}}\right)$$

$$= -1350\ \text{V}$$

$$V_a - V_b = (1350\ \text{V}) - (-1350\ \text{V}) = 2700\ \text{V}$$

Finally,

$$v = \sqrt{\frac{2(2.0 \times 10^{-9}\ \text{C})(2700\ \text{V})}{5.0 \times 10^{-9}\ \text{kg}}} = 46\ \text{m/s}$$

EVALUATE: Our result makes sense: The positive dust particle speeds up as it moves away from the +3.0-nC charge and toward the −3.0-nC charge. To check unit consistency in the final line of the calculation, note that 1 V = 1 J/C, so the numerator under the radical has units of J or kg $\cdot$ m^2/s^2.

TEST YOUR UNDERSTANDING OF SECTION 23.2 If the electric *potential* at a certain point is zero, does the electric *field* at that point have to be zero? (*Hint:* Consider point c in Example 23.4 and Example 21.8.)

23.3 CALCULATING ELECTRIC POTENTIAL

When calculating the potential due to a charge distribution, we usually follow one of two routes. If we know the charge distribution, we can use Eq. (23.15) or (23.16). Or if we know how the electric field depends on position, we can use Eq. (23.17), defining the potential to be zero at some convenient place. Some problems require a combination of these approaches.

As you read through these examples, compare them with the related examples of calculating electric *field* in Section 21.5. You'll see how much easier it is to calculate scalar electric potentials than vector electric fields. The moral is clear: Whenever possible, solve problems by means of an energy approach (using electric potential and electric potential energy) rather than a dynamics approach (using electric fields and electric forces).

PROBLEM-SOLVING STRATEGY 23.1 CALCULATING ELECTRIC POTENTIAL

IDENTIFY *the relevant concepts:* Remember that electric potential is *potential energy per unit charge*.

SET UP *the problem* using the following steps:

1. Make a drawing showing the locations and values of the charges (which may be point charges or a continuous distribution of charge) and your choice of coordinate axes.
2. Indicate on your drawing the position of the point at which you want to calculate the electric potential V. Sometimes this position will be an arbitrary one (say, a point a distance r from the center of a charged sphere).

EXECUTE *the solution* as follows:

1. To find the potential due to a collection of point charges, use Eq. (23.15). If you are given a continuous charge distribution, devise a way to divide it into infinitesimal elements and use Eq. (23.16). Carry out the integration, using appropriate limits to include the entire charge distribution.
2. If you are given the electric field, or if you can find it from any of the methods presented in Chapters 21 or 22, it may be easier to find the potential difference between points a and b from Eq. (23.17) or (23.18). When appropriate, make use of your freedom to define V to be zero at some convenient place, and choose this place to be point b. (For point charges, this will usually be at infinity. For other distributions of charge—especially those that themselves extend to infinity—it may be necessary to define V_b to be zero at some finite distance from the charge distribution.) Then the potential at any other point, say a, can by found from Eq. (23.17) or (23.18) with $V_b = 0$.
3. Although potential V is a *scalar* quantity, you may have to use components of the vectors $\vec{E}$ and $d\vec{l}$ when you use Eq. (23.17) or (23.18) to calculate V.

EVALUATE *your answer:* Check whether your answer agrees with your intuition. If your result gives V as a function of position, graph the function to see whether it makes sense. If you know the electric field, you can make a rough check of your result for V by verifying that V decreases if you move in the direction of $\vec{E}$.

EXAMPLE 23.8 A CHARGED CONDUCTING SPHERE

SOLUTION

A solid conducting sphere of radius R has a total charge q. Find the electric potential everywhere, both outside and inside the sphere.

SOLUTION

IDENTIFY and SET UP: In Example 22.5 (Section 22.4) we used Gauss's law to find the electric *field* at all points for this charge distribution. We can use that result to determine the potential.

EXECUTE: From Example 22.5, the field *outside* the sphere is the same as if the sphere were removed and replaced by a point charge q. We take $V = 0$ at infinity, as we did for a point charge. Then the potential at a point outside the sphere at a distance r from its center is the same as that due to a point charge q at the center:

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{r}$$

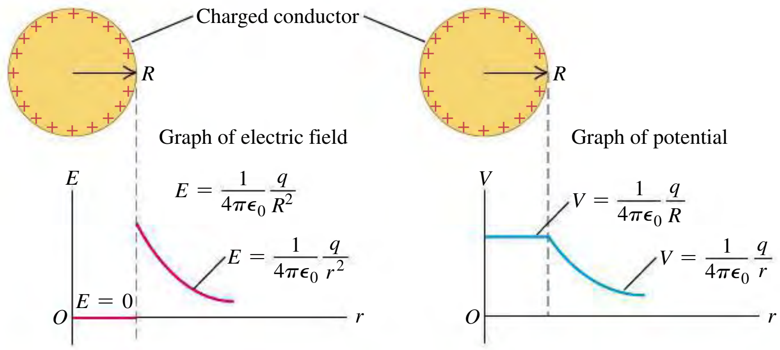
The potential at the surface of the sphere is $V_{\text{surface}} = q/4\pi\epsilon_0 R$.

Continued

Inside the sphere, $\vec{E}$ is zero everywhere. Hence no work is done on a test charge that moves from any point to any other point inside the sphere. Thus the potential is the same at every point inside the sphere and is equal to its value $q/4\pi\epsilon_0 R$ at the surface.

EVALUATE: **Figure 23.16** shows the field and potential for a positive charge q. In this case the electric field points radially away from the sphere. As you move away from the sphere, in the direction of $\vec{E}$, V decreases (as it should).

23.16 Electric-field magnitude E and potential V at points inside and outside a positively charged spherical conductor.



Ionization and Corona Discharge

The results of Example 23.8 have numerous practical consequences. One consequence relates to the maximum potential to which a conductor in air can be raised. This potential is limited because air molecules become *ionized*, and air becomes a conductor, at an electric-field magnitude of about 3×10^6 V/m. Assume for the moment that q is positive. When we compare the expressions in Example 23.8 for the potential V_{surface} and field magnitude E_{surface} at the surface of a charged conducting sphere, we note that $V_{\text{surface}} = E_{\text{surface}}R$. Thus, if E_{m} represents the electric-field magnitude at which air becomes conductive (known as the *dielectric strength* of air), then the maximum potential V_{m} to which a spherical conductor can be raised is

$$V_{\text{m}} = RE_{\text{m}}$$

For a conducting sphere 1 cm in radius in air, $V_{\text{m}} = (10^{-2}\text{ m})(3 \times 10^6\text{ V/m}) =$ 30,000 V. No amount of "charging" could raise the potential of a conducting sphere of this size in air higher than about 30,000 V; attempting to raise the potential further by adding extra charge would cause the surrounding air to become ionized and conductive, and the extra added charge would leak into the air.

To attain even higher potentials, high-voltage machines such as Van de Graaff generators use spherical terminals with very large radii (see Fig. 22.26 and the photograph that opens Chapter 22). For example, a terminal of radius $R = 2$ m has a maximum potential $V_{\text{m}} = (2\text{ m})(3 \times 10^6\text{ V/m}) = 6 \times 10^6\text{ V} = 6$ MV.

Our result in Example 23.8 also explains what happens with a charged conductor with a very *small* radius of curvature, such as a sharp point or thin wire. Because the maximum potential is proportional to the radius, even relatively small potentials applied to sharp points in air produce sufficiently high fields just outside the point to ionize the surrounding air, making it become a conductor. The resulting current and its associated glow (visible in a dark room) are called *corona discharge*. Laser printers and photocopying machines use corona discharge from fine wires to spray charge on the imaging drum (see Fig. 21.2).

A large-radius conductor is used in situations where it's important to *prevent* corona discharge. An example is the blunt end of a metal lightning rod (**Fig. 23.17**). If there is an excess charge in the atmosphere, as happens during thunderstorms, a substantial charge of the opposite sign can build up on this blunt end. As a result,

23.17 The metal mast at the top of the Empire State Building acts as a lightning rod. It is struck by lightning as many as 500 times each year.

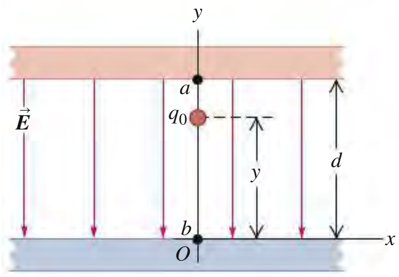
when the atmospheric charge is discharged through a lightning bolt, it tends to be attracted to the charged lightning rod rather than to other structures that could be damaged. (A conducting wire connecting the lightning rod to the ground then allows the acquired charge to dissipate harmlessly.) A lightning rod with a sharp end would allow less charge buildup and hence be less effective.

EXAMPLE 23.9 OPPOSITELY CHARGED PARALLEL PLATES

SOLUTION

Find the potential at any height y between the two oppositely charged parallel plates discussed in Section 23.1 (**Fig. 23.18**).

23.18 The charged parallel plates from Fig. 23.2.



SOLUTION

IDENTIFY and SET UP: We discussed this situation in Section 23.1. From Eq. (23.5), we know the electric *potential energy U* for a test charge q_0 is $U = q_0Ey$. (We set $y = 0$ and $U = 0$ at the bottom plate.) We use Eq. (23.12), $U = q_0V$, to find the electric *potential V* as a function of y.

EXECUTE: The potential $V(y)$ at coordinate y is the potential energy per unit charge:

$$V(y) = \frac{U(y)}{q_0} = \frac{q_0Ey}{q_0} = Ey$$

The potential decreases as we move in the direction of $\vec{E}$ from the upper to the lower plate. At point a, where $y = d$ and $V(y) = V_a$,

$$V_a - V_b = Ed \quad \text{and} \quad E = \frac{V_a - V_b}{d} = \frac{V_{ab}}{d}$$

where V_{ab} is the potential of the positive plate with respect to the negative plate. That is, the electric field equals the potential difference between the plates divided by the distance between them. For a given potential difference V_{ab}, the smaller the distance d between the two plates, the greater the magnitude E of the electric field. (This relationship between E and V_{ab} holds *only* for the planar geometry we have described. It does *not* work for situations such as concentric cylinders or spheres in which the electric field is not uniform.)

EVALUATE: Our result shows that $V = 0$ at the bottom plate (at $y = 0$). This is consistent with our choice that $U = q_0V = 0$ for a test charge placed at the bottom plate.

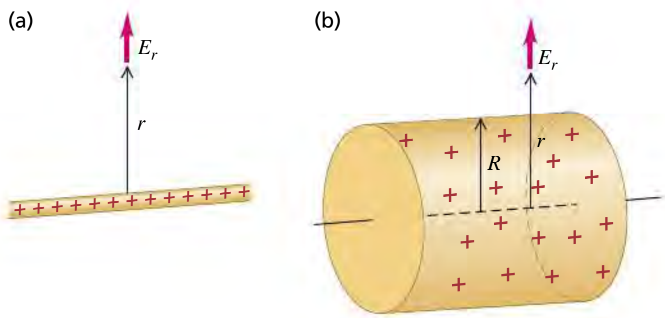
CAUTION "Zero potential" is arbitrary You might think that if a conducting body has zero potential, it must also have zero net charge. But that just isn't so! As an example, the plate at $y = 0$ in Fig. 23.18 has zero potential ($V = 0$) but has a nonzero charge per unit area $-\sigma$. There's nothing special about the place where potential is zero; we *define* this place to be wherever we want it to be.

EXAMPLE 23.10 AN INFINITE LINE CHARGE OR CHARGED CONDUCTING CYLINDER

SOLUTION

Find the potential at a distance r from a very long line of charge with linear charge density (charge per unit length) λ.

23.19 Electric field outside (a) a long, positively charged wire and (b) a long, positively charged cylinder.



SOLUTION

IDENTIFY and SET UP: In both Example 21.10 (Section 21.5) and Example 22.6 (Section 22.4) we found that the electric field at a radial distance r from a long straight-line charge (**Fig. 23.19a**) has only a radial component $E_r = \lambda/2\pi\epsilon_0 r$. We use this expression to find the potential by integrating $\vec{E}$ as in Eq. (23.17).

EXECUTE: Since the field has only a radial component, we have $\vec{E} \cdot d\vec{l} = E_r dr$. Hence from Eq. (23.17) the potential of any point a with respect to any other point b, at radial distances r_a and r_b from the line of charge, is

$$V_a - V_b = \int_a^b \vec{E} \cdot d\vec{l} = \int_a^b E_r dr = \frac{\lambda}{2\pi\epsilon_0}\int_{r_a}^{r_b}\frac{dr}{r} = \frac{\lambda}{2\pi\epsilon_0}\ln\frac{r_b}{r_a}$$

If we take point b at infinity and set $V_b = 0$, we find that V_a is *infinite* for any finite distance r_a from the line charge: $V_a = (\lambda/2\pi\epsilon_0)\ln(\infty/r_a) = \infty$. This is *not* a useful way to define

V for this problem! The difficulty is that the charge distribution itself extends to infinity.

Instead, as recommended in Problem-Solving Strategy 23.1, we set $V_b = 0$ at point b at an arbitrary but *finite* radial distance r_0. Then the potential $V = V_a$ at point a at a radial distance r is given by $V - 0 = (\lambda/2\pi\epsilon_0)\ln(r_0/r)$, or

$$V = \frac{\lambda}{2\pi\epsilon_0}\ln\frac{r_0}{r}$$

EVALUATE: According to our result, if λ is positive, then V decreases as r increases. This is as it should be: V decreases as we move in the direction of $\vec{E}$.

From Example 22.6, the expression for E_r with which we started also applies outside a long, charged conducting cylinder with charge per unit length λ (Fig. 23.19b). Hence our result also gives the potential for such a cylinder, but only for values of r (the distance from the cylinder axis) equal to or greater than the radius R of the cylinder. If we choose r_0 to be the radius R, so that $V = 0$ when $r = R$, then at any point for which $r > R$,

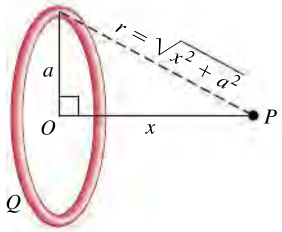
$$V = \frac{\lambda}{2\pi\epsilon_0}\ln\frac{R}{r}$$

Inside the cylinder, $\vec{E} = \mathbf{0}$, and V has the same value (zero) as on the cylinder's surface.

EXAMPLE 23.11 A RING OF CHARGE

SOLUTION

Electric charge Q is distributed uniformly around a thin ring of radius a (**Fig. 23.20**). Find the potential at a point P on the ring axis at a distance x from the center of the ring.



23.20 All the charge in a ring of charge Q is the same distance r from a point P on the ring axis.

SOLUTION

IDENTIFY and SET UP: We divide the ring into infinitesimal segments and use Eq. (23.16) to find V. All parts of the ring (and therefore all elements of the charge distribution) are at the same distance from P.

EXECUTE: Figure 23.20 shows that the distance from each charge element dq to P is $r = \sqrt{x^2 + a^2}$. Hence we can take the factor $1/r$ outside the integral in Eq. (23.16), and

$$V = \frac{1}{4\pi\epsilon_0}\int\frac{dq}{r} = \frac{1}{4\pi\epsilon_0}\frac{1}{\sqrt{x^2 + a^2}}\int dq = \frac{1}{4\pi\epsilon_0}\frac{Q}{\sqrt{x^2 + a^2}}$$

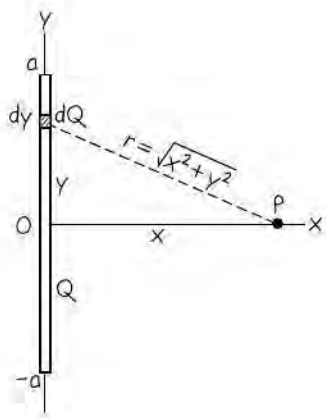
EVALUATE: When x is much larger than a, our expression for V becomes approximately $V = Q/4\pi\epsilon_0 x$, which is the potential at a distance x from a point charge Q. Very far from a charged ring, its electric potential looks like that of a point charge. We drew a similar conclusion about the electric *field* of a ring in Example 21.9 (Section 21.5).

We know the electric field at all points along the x-axis from Example 21.9 (Section 21.5), so we can also find V along this axis by integrating $\vec{E} \cdot d\vec{l}$ as in Eq. (23.17).

EXAMPLE 23.12 POTENTIAL OF A LINE OF CHARGE

SOLUTION

Positive electric charge Q is distributed uniformly along a line of length $2a$ lying along the y-axis between $y = -a$ and $y = +a$ (**Fig. 23.21**). Find the electric potential at a point P on the x-axis at a distance x from the origin.



23.21 Our sketch for this problem.

SOLUTION

IDENTIFY and SET UP: This is the situation of Example 21.10 (Section 21.5), where we found an expression for the electric field $\vec{E}$ at an arbitrary point on the x-axis. We can find V at point P by using Eq. (23.16) to integrate over the charge distribution. Unlike the situation in Example 23.11, each charge element dQ is a *different* distance from point P, so the integration will take a little more effort.

Continued

EXECUTE: As in Example 21.10, the element of charge dQ corresponding to an element of length dy on the rod is $dQ = (Q/2a)dy$. The distance from dQ to P is $\sqrt{x^2 + y^2}$, so the contribution dV that the charge element makes to the potential at P is

$$dV = \frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\frac{dy}{\sqrt{x^2 + y^2}}$$

To find the potential at P due to the entire rod, we integrate dV over the length of the rod from $y = -a$ to $y = a$:

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\int_{-a}^{a}\frac{dy}{\sqrt{x^2 + y^2}}$$

You can look up the integral in a table. The final result is

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{2a}\ln\left(\frac{\sqrt{a^2 + x^2} + a}{\sqrt{a^2 + x^2} - a}\right)$$

EVALUATE: We can check our result by letting x approach infinity. In this limit the point P is infinitely far from all of the charge, so we expect V to approach zero; you can verify that it does.

We know the electric field at all points along the x-axis from Example 21.10. We invite you to use this information to find V along this axis by integrating $\vec{E}$ as in Eq. (23.17).

TEST YOUR UNDERSTANDING OF SECTION 23.3 If the electric *field* at a certain point is zero, does the electric *potential* at that point have to be zero? (*Hint:* Consider the center of the ring in Example 23.11 and Example 21.9.)

23.4 EQUIPOTENTIAL SURFACES

Field lines (see Section 21.6) help us visualize electric fields. In a similar way, the potential at various points in an electric field can be represented graphically by *equipotential surfaces*. These use the same fundamental idea as topographic maps like those used by hikers and mountain climbers (**Fig. 23.22**). On a topographic map, contour lines are drawn through points that are all at the same elevation. Any number of these could be drawn, but typically only a few contour lines are shown at equal spacings of elevation. If a mass m is moved over the terrain along such a contour line, the gravitational potential energy mgy does not change because the elevation y is constant. Thus contour lines on a topographic map are really curves of constant gravitational potential energy. Contour lines are close together where the terrain is steep and there are large changes in elevation over a small horizontal distance; the contour lines are farther apart where the terrain is gently sloping. A ball allowed to roll downhill will experience the greatest downhill gravitational force where contour lines are closest together.

By analogy to contour lines on a topographic map, an **equipotential surface** is a three-dimensional surface on which the *electric potential V* is the same at every point. If a test charge q_0 is moved from point to point on such a surface, the *electric* potential energy q_0V remains constant. In a region where an electric field is present, we can construct an equipotential surface through any point. In diagrams we usually show only a few representative equipotentials, often with equal potential differences between adjacent surfaces. No point can be at two different potentials, so equipotential surfaces for different potentials can never touch or intersect.

23.22 Contour lines on a topographic map are curves of constant elevation and hence of constant gravitational potential energy.

Equipotential Surfaces and Field Lines

Because potential energy does not change as a test charge moves over an equipotential surface, the electric field can do no work on such a charge. It follows that $\vec{E}$ must be perpendicular to the surface at every point so that the electric force $q_0\vec{E}$ is always perpendicular to the displacement of a charge moving on the surface. **Field lines and equipotential surfaces are always mutually perpendicular.** In general, field lines are curves, and equipotentials are curved surfaces. For the special case of a *uniform* field, in which the field lines are straight, parallel, and equally spaced, the equipotentials are parallel *planes* perpendicular to the field lines.

23.23 Cross sections of equipotential surfaces (blue lines) and electric field lines (red lines) for assemblies of point charges. There are equal potential differences between adjacent surfaces. Compare these diagrams to those in Fig. 21.28, which showed only the electric field lines.

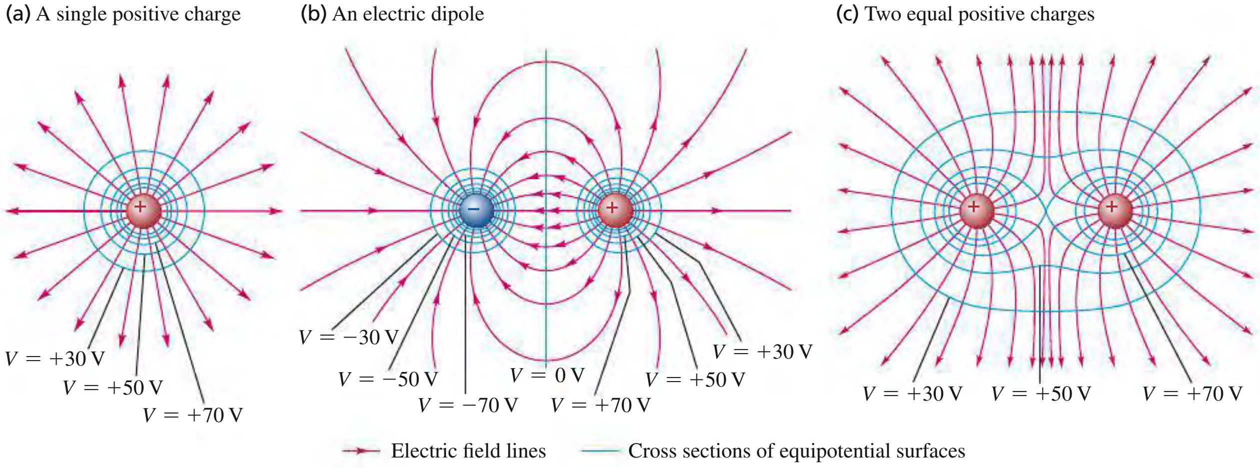
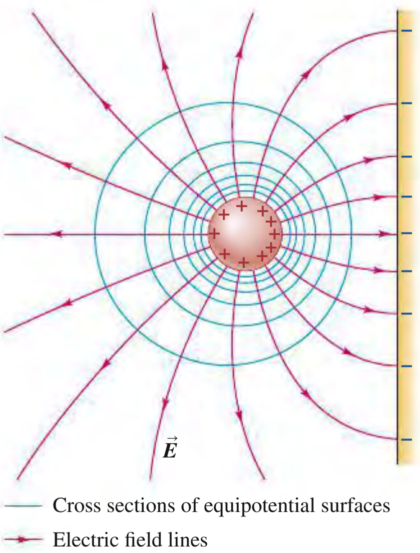


Figure 23.23 shows three arrangements of charges. The field lines in the plane of the charges are represented by red lines, and the intersections of the equipotential surfaces with this plane (that is, cross sections of these surfaces) are shown as blue lines. The actual equipotential surfaces are three-dimensional. At each crossing of an equipotential and a field line, the two are perpendicular.

In Fig. 23.23 we have drawn equipotentials so that there are equal potential differences between adjacent surfaces. In regions where the magnitude of $\vec{E}$ is large, the equipotential surfaces are close together because the field does a relatively large amount of work on a test charge in a relatively small displacement. This is the case near the point charge in Fig. 23.23a or between the two point charges in Fig. 23.23b; note that in these regions the field lines are also closer together. This is directly analogous to the downhill force of gravity being greatest in regions on a topographic map where contour lines are close together. Conversely, in regions where the field is weaker, the equipotential surfaces are farther apart; this happens at larger radii in Fig. 23.23a, to the left of the negative charge or the right of the positive charge in Fig. 23.23b, and at greater distances from both charges in Fig. 23.23c. (It may appear that two equipotential surfaces intersect at the center of Fig. 23.23c, in violation of the rule that this can never happen. In fact this is a single figure-8–shaped equipotential surface.)

CAUTION ***E* need not be constant over an equipotential surface** On a given equipotential surface, the potential V has the same value at every point. In general, however, the electric-field magnitude E is *not* the same at all points on an equipotential surface. For instance, on equipotential surface "$V = -30$ V" in Fig. 23.23b, E is less to the left of the negative charge than it is between the two charges. On the figure-8–shaped equipotential surface in Fig. 23.23c, $E = 0$ at the middle point halfway between the two charges; at any other point on this surface, E is nonzero.

23.24 When charges are at rest, a conducting surface is always an equipotential surface. Field lines are perpendicular to a conducting surface.



Equipotentials and Conductors

Here's an important statement about equipotential surfaces: **When all charges are at rest, the surface of a conductor is always an equipotential surface.** Since the electric field $\vec{E}$ is always perpendicular to an equipotential surface, we can prove this statement by proving that **when all charges are at rest, the electric field just outside a conductor must be perpendicular to the surface at every point** (**Fig. 23.24**). We know that $\vec{E} = \mathbf{0}$ everywhere inside the conductor; otherwise, charges would move. In particular, at any point just inside the surface the component of $\vec{E}$ tangent to the surface is zero. It follows that the tangential component of $\vec{E}$ is also zero just *outside* the surface. If it were not, a charge could move around a rectangular path partly inside and partly outside (**Fig. 23.25**) and return to its starting point with a net amount of work having been done on it. This would violate the conservative nature of electrostatic fields, so the tangential component of $\vec{E}$ just outside the surface must be zero at every point on the surface. Thus $\vec{E}$ is perpendicular to the surface at each point, proving our statement.

It also follows that **when all charges are at rest, the entire solid volume of a conductor is at the same potential.** Equation (23.17) states that the potential difference between two points a and b within the conductor's solid volume, $V_a - V_b$, is equal to the line integral $\int_a^b \vec{E} \cdot d\vec{l}$ of the electric field from a to b. Since $\vec{E} = \mathbf{0}$ everywhere inside the conductor, the integral is guaranteed to be zero for any two such points a and b. Hence the potential is the same for any two points within the solid volume of the conductor. We describe this by saying that the solid volume of the conductor is an *equipotential volume*.

We can now prove a theorem that we quoted without proof in Section 22.5. The theorem is as follows: In an electrostatic situation, if a conductor contains a cavity and if no charge is present inside the cavity, then there can be no net charge *anywhere* on the surface of the cavity. This means that if you're inside a charged conducting box, you can safely touch any point on the inside walls of the box without being shocked. To prove this theorem, we first prove that *every point in the cavity is at the same potential*. In **Fig. 23.26** the conducting surface A of the cavity is an equipotential surface, as we have just proved. Suppose point P in the cavity is at a different potential; then we can construct a different equipotential surface B including point P.

Now consider a Gaussian surface, shown in Fig. 23.26, between the two equipotential surfaces. Because of the relationship between $\vec{E}$ and the equipotentials, we know that the field at every point between the equipotentials is from A toward B, or else at every point it is from B toward A, depending on which equipotential surface is at higher potential. In either case the flux through this Gaussian surface is certainly not zero. But then Gauss's law says that the charge enclosed by the Gaussian surface cannot be zero. This contradicts our initial assumption that there is *no* charge in the cavity. So the potential at P *cannot* be different from that at the cavity wall.

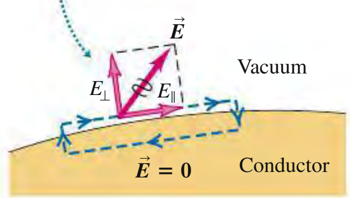
The entire region of the cavity must therefore be at the same potential. But for this to be true, *the electric field inside the cavity must be zero everywhere*. Finally, Gauss's law shows that the electric field at any point on the surface of a conductor is proportional to the surface charge density σ at that point. We conclude that *the surface charge density on the wall of the cavity is zero at every point*. This chain of reasoning may seem tortuous, but it is worth careful study.

CAUTION **Equipotential surfaces vs. Gaussian surfaces** Don't confuse equipotential surfaces with the Gaussian surfaces we encountered in Chapter 22. Gaussian surfaces have relevance only when we are using Gauss's law, and we can choose *any* Gaussian surface that's convenient. We *cannot* choose equipotential surfaces; the shape is determined by the charge distribution.

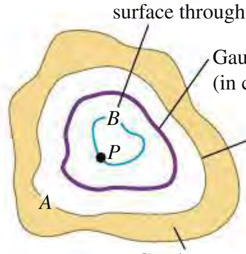
TEST YOUR UNDERSTANDING OF SECTION 23.4 Would the shapes of the equipotential surfaces in Fig. 23.23 change if the sign of each charge were reversed?

DEMO

23.25 At all points on the surface of a conductor, the electric field must be perpendicular to the surface. If $\vec{E}$ had a tangential component, a net amount of work would be done on a test charge by moving it around a loop as shown here—which is impossible because the electric force is conservative.



23.26 A cavity in a conductor. If the cavity contains no charge, every point in the cavity is at the same potential, the electric field is zero everywhere in the cavity, and there is no charge anywhere on the surface of the cavity.



23.5 POTENTIAL GRADIENT

Electric field and potential are closely related. Equation (23.17), restated here, expresses one aspect of that relationship:

$$V_a - V_b = \int_a^b \vec{E} \cdot d\vec{l}$$

If we know $\vec{E}$ at various points, we can use this equation to calculate potential differences. In this section we show how to turn this around; if we know the potential V at various points, we can use it to determine $\vec{E}$. Regarding V as a function of the coordinates (x, y, z) of a point in space, we will show that the components of $\vec{E}$ are related to the *partial derivatives* of V with respect to x, y, and z.

In Eq. (23.17), $V_a - V_b$ is the potential of a with respect to b—that is, the change of potential encountered on a trip from b to a. We can write this as

$$V_a - V_b = \int_b^a dV = -\int_a^b dV$$

where dV is the infinitesimal change of potential accompanying an infinitesimal element $d\vec{l}$ of the path from b to a. Comparing to Eq. (23.17), we have

$$-\int_a^b dV = \int_a^b \vec{E} \cdot d\vec{l}$$

These two integrals must be equal for *any* pair of limits a and b, and for this to be true the *integrands* must be equal. Thus, for *any* infinitesimal displacement $d\vec{l}$,

$$-dV = \vec{E} \cdot d\vec{l}$$

To interpret this expression, we write $\vec{E}$ and $d\vec{l}$ in terms of their components: $\vec{E} = \hat{\imath}\, E_x + \hat{\jmath}\, E_y + \hat{k}\, E_z$ and $d\vec{l} = \hat{\imath}\, dx + \hat{\jmath}\, dy + \hat{k}\, dz$. Then

$$-dV = E_x dx + E_y dy + E_z dz$$

Suppose the displacement is parallel to the x-axis, so $dy = dz = 0$. Then $-dV = E_x dx$ or $E_x = -(dV/dx)_{y,\, z \text{ constant}}$, where the subscript reminds us that only x varies in the derivative; recall that V is in general a function of x, y, and z. But this is just what is meant by the partial derivative $\partial V/\partial x$. The y- and z-components of $\vec{E}$ are related to the corresponding derivatives of V in the same way, so

Electric field components found from potential:

Each electric field component ...

$$E_x = -\frac{\partial V}{\partial x} \qquad E_y = -\frac{\partial V}{\partial y} \qquad E_z = -\frac{\partial V}{\partial z} \tag{23.19}$$

... equals the negative of the corresponding partial derivative of electric potential function V.

This is consistent with the units of electric field being V/m. In terms of unit vectors we can write

Electric field vector found from potential:

Electric field

$$\vec{E} = -\left(\hat{\imath}\frac{\partial V}{\partial x} + \hat{\jmath}\frac{\partial V}{\partial y} + \hat{k}\frac{\partial V}{\partial z}\right) \tag{23.20}$$

Partial derivatives of electric potential function V

The following operation is called the **gradient** of the function f:

$$\vec{\nabla} f = \left(\hat{\imath}\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}\right) f \tag{23.21}$$

The operator denoted by $\vec{\nabla}$ is called "grad" or "del." Thus in vector notation,

$$\vec{E} = -\vec{\nabla} V \tag{23.22}$$

This is read "$\vec{E}$ is the negative of the gradient of V" or "$\vec{E}$ equals negative grad V." The quantity $\vec{\nabla} V$ is called the *potential gradient.*

At each point, the potential gradient $\vec{\nabla}V$ points in the direction in which V *increases* most rapidly with a change in position. Hence at each point the direction of $\vec{E} = -\vec{\nabla}V$ is the direction in which V *decreases* most rapidly and is always perpendicular to the equipotential surface through the point. This agrees with our observation in Section 23.2 that moving in the direction of the electric field means moving in the direction of decreasing potential.

Equation (23.22) doesn't depend on the particular choice of the zero point for V. If we were to change the zero point, the effect would be to change V at every point by the same amount; the derivatives of V would be the same.

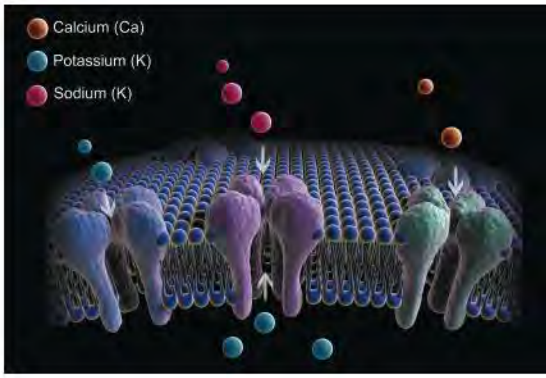
If $\vec{E}$ has a radial component E_r with respect to a point or an axis and r is the distance from the point or axis, the relationship corresponding to Eqs. (23.19) is

$$E_r = -\frac{\partial V}{\partial r} \qquad \text{(radial electric field)} \tag{23.23}$$

Often we can compute the electric field caused by a charge distribution in either of two ways: directly, by adding the $\vec{E}$ fields of point charges, or by first calculating the potential and then taking its gradient to find the field. The second method is often easier because potential is a *scalar* quantity, requiring at worst the integration of a scalar function. Electric field is a *vector* quantity, requiring computation of components for each element of charge and a separate integration for each component. Thus, quite apart from its fundamental significance, potential offers a very useful computational technique in field calculations. In the next two examples, a knowledge of V is used to find the electric field.

We stress once more that if we know $\vec{E}$ as a function of position, we can calculate V from Eq. (23.17) or (23.18), and if we know V as a function of position, we can calculate $\vec{E}$ from Eq. (23.19), (23.20), or (23.23). Deriving V from $\vec{E}$ requires integration, and deriving $\vec{E}$ from V requires differentiation.

BIO Application Potential Gradient Across a Cell Membrane The interior of a human cell is at a lower electric potential V than the exterior. (The potential difference when the cell is inactive is about -70 mV in neurons and about -95 mV in skeletal muscle cells.) Hence there is a potential gradient $\vec{\nabla}V$ that points from the *interior* to the *exterior* of the cell membrane, and an electric field $\vec{E} = -\vec{\nabla}V$ that points from the *exterior* to the *interior*. This field affects how ions flow into or out of the cell through special channels in the membrane.



EXAMPLE 23.13 POTENTIAL AND FIELD OF A POINT CHARGE

From Eq. (23.14) the potential at a radial distance r from a point charge q is $V = q/4\pi\epsilon_0 r$. Find the vector electric field from this expression for V.

SOLUTION

IDENTIFY and SET UP: This problem uses the general relationship between the electric potential as a function of position and the electric-field vector. By symmetry, the electric field here has only a radial component E_r. We use Eq. (23.23) to find this component.

EXECUTE: From Eq. (23.23),

$$E_r = -\frac{\partial V}{\partial r} = -\frac{\partial}{\partial r}\left(\frac{1}{4\pi\epsilon_0}\frac{q}{r}\right) = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}$$

so the vector electric field is

$$\vec{E} = \hat{r}E_r = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$$

EVALUATE: Our result agrees with Eq. (21.7), as it must.

An alternative approach is to ignore the radial symmetry, write the radial distance as $r = \sqrt{x^2 + y^2 + z^2}$, and take the derivatives of V with respect to x, y, and z as in Eq. (23.20). We find

$$\frac{\partial V}{\partial x} = \frac{\partial}{\partial x}\left(\frac{1}{4\pi\epsilon_0}\frac{q}{\sqrt{x^2 + y^2 + z^2}}\right) = -\frac{1}{4\pi\epsilon_0}\frac{qx}{(x^2 + y^2 + z^2)^{3/2}}$$

$$= -\frac{qx}{4\pi\epsilon_0 r^3}$$

and similarly

$$\frac{\partial V}{\partial y} = -\frac{qy}{4\pi\epsilon_0 r^3} \qquad \frac{\partial V}{\partial z} = -\frac{qz}{4\pi\epsilon_0 r^3}$$

Then from Eq. (23.20),

$$\vec{E} = -\left[\hat{\imath}\left(-\frac{qx}{4\pi\epsilon_0 r^3}\right) + \hat{\jmath}\left(-\frac{qy}{4\pi\epsilon_0 r^3}\right) + \hat{k}\left(-\frac{qz}{4\pi\epsilon_0 r^3}\right)\right]$$

$$= \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\left(\frac{x\hat{\imath} + y\hat{\jmath} + z\hat{k}}{r}\right) = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$$

This approach gives us the same answer, but with more effort. Clearly it's best to exploit the symmetry of the charge distribution whenever possible.

EXAMPLE 23.14 POTENTIAL AND FIELD OF A RING OF CHARGE

SOLUTION

In Example 23.11 (Section 23.3) we found that for a ring of charge with radius a and total charge Q, the potential at a point P on the ring's symmetry axis a distance x from the center is

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{\sqrt{x^2 + a^2}}$$

Find the electric field at P.

SOLUTION

IDENTIFY and SET UP: Figure 23.20 shows the situation. We are given V as a function of x along the x-axis, and we wish to find the electric field at a point on this axis. From the symmetry of the charge distribution, the electric field along the symmetry (x-) axis of the ring can have only an x-component. We find it by using the first of Eqs. (23.19).

EXECUTE: The x-component of the electric field is

$$E_x = -\frac{\partial V}{\partial x} = \frac{1}{4\pi\epsilon_0}\frac{Qx}{(x^2 + a^2)^{3/2}}$$

EVALUATE: This agrees with our result in Example 21.9.

CAUTION **Don't use expressions where they don't apply** In this example, V is not a function of y or z on the ring axis, so $\partial V/\partial y = \partial V/\partial z = 0$ and $E_y = E_z = 0$. But that does not mean that it's true *everywhere*; our expressions for V and E_x are valid *on the ring axis only*. If we had an expression for V valid at *all* points in space, we could use it to find the components of $\vec{E}$ at any point by using Eqs. (23.19).

TEST YOUR UNDERSTANDING OF SECTION 23.5 In a certain region of space the potential is given by $V = A + Bx + Cy^3 + Dxy$, where A, B, C, and D are positive constants. Which of these statements about the electric field $\vec{E}$ in this region of space is correct? (There may be more than one correct answer.) (i) Increasing the value of A will increase the value of $\vec{E}$ at all points; (ii) increasing the value of A will decrease the value of $\vec{E}$ at all points; (iii) $\vec{E}$ has no z-component; (iv) the electric field is zero at the origin $(x = 0, y = 0, z = 0)$.

Electric potential energy: The electric force caused by any collection of charges at rest is a conservative force. The work W done by the electric force on a charged particle moving in an electric field can be represented by the change in a potential-energy function U.

The electric potential energy for two point charges q and q_0 depends on their separation r. The electric potential energy for a charge q_0 in the presence of a collection of charges q_1, q_2, q_3 depends on the distance from q_0 to each of these other charges. (See Examples 23.1 and 23.2.)

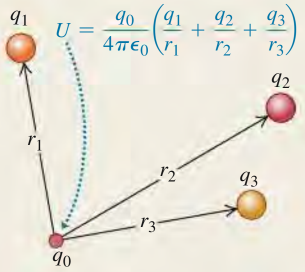
$$W_{a \to b} = U_a - U_b \qquad (23.2)$$

$$U = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r} \qquad (23.9)$$

(two point charges)

$$U = \frac{q_0}{4\pi\epsilon_0}\left(\frac{q_1}{r_1} + \frac{q_2}{r_2} + \frac{q_3}{r_3} + \cdots\right)$$
$$= \frac{q_0}{4\pi\epsilon_0}\sum_i \frac{q_i}{r_i} \qquad (23.10)$$

(q_0 in presence of other point charges)



Electric potential: Potential, denoted by V, is potential energy per unit charge. The potential difference between two points equals the amount of work per charge that would be required to move a positive test charge between those points. The potential V due to a quantity of charge can be calculated by summing (if the charge is a collection of point charges) or by integrating (if the charge is a distribution). (See Examples 23.3, 23.4, 23.5, 23.7, 23.11, and 23.12.)

The potential difference between two points a and b, also called the potential of a with respect to b, is given by the line integral of $\vec{E}$. The potential at a given point can be found by first finding $\vec{E}$ and then carrying out this integral. (See Examples 23.6, 23.8, 23.9, and 23.10.)

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{r} \qquad (23.14)$$

(due to a point charge)

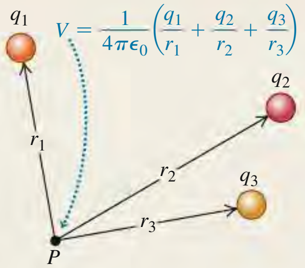
$$V = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{r_i} \qquad (23.15)$$

(due to a collection of point charges)

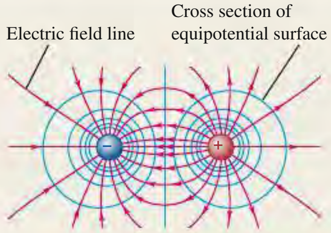
$$V = \frac{1}{4\pi\epsilon_0}\int \frac{dq}{r} \qquad (23.16)$$

(due to a charge distribution)

$$V_a - V_b = \int_a^b \vec{E} \cdot d\vec{l}$$
$$= \int_a^b E\cos\phi\, dl \qquad (23.17)$$



Equipotential surfaces: An equipotential surface is a surface on which the potential has the same value at every point. At a point where a field line crosses an equipotential surface, the two are perpendicular. When all charges are at rest, the surface of a conductor is always an equipotential surface and all points in the interior of a conductor are at the same potential. When a cavity within a conductor contains no charge, the entire cavity is an equipotential region and there is no surface charge anywhere on the surface of the cavity.



Finding electric field from electric potential: If the potential V is known as a function of the coordinates x, y, and z, the components of electric field $\vec{E}$ at any point are given by partial derivatives of V. (See Examples 23.13 and 23.14.)

$$E_x = -\frac{\partial V}{\partial x} \quad E_y = -\frac{\partial V}{\partial y} \quad E_z = -\frac{\partial V}{\partial z} \qquad (23.19)$$

$$\vec{E} = -\left(\hat{\imath}\frac{\partial V}{\partial x} + \hat{\jmath}\frac{\partial V}{\partial y} + \hat{k}\frac{\partial V}{\partial z}\right) \qquad (23.20)$$

(vector form)

BRIDGING PROBLEM A POINT CHARGE AND A LINE OF CHARGE

SOLUTION

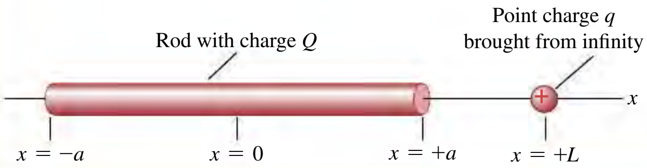
Positive electric charge Q is distributed uniformly along a thin rod of length $2a$. The rod lies along the x-axis between $x = -a$ and $x = +a$ (**Fig. 23.27**). Calculate how much work you must do to bring a positive point charge q from infinity to the point $x = +L$ on the x-axis, where $L > a$.

SOLUTION GUIDE

IDENTIFY and SET UP

1. In this problem you must first calculate the potential V at $x = +L$ due to the charged rod. You can then find the change in potential energy involved in bringing the point charge q from infinity (where $V = 0$) to $x = +L$.

23.27 How much work must you do to bring point charge q in from infinity?



2. To find V, divide the rod into infinitesimal segments of length dx'. How much charge is on such a segment? Consider one such segment located at $x = x'$, where $-a \le x' \le a$. What is the potential dV at $x = +L$ due to this segment?
3. The total potential at $x = +L$ is the integral of dV, including contributions from all of the segments for x' from $-a$ to $+a$. Set up this integral.

EXECUTE

4. Integrate your expression from step 3 to find the potential V at $x = +L$. A simple, standard substitution will do the trick; use a table of integrals only as a last resort.
5. Use your result from step 4 to find the potential energy for a point charge q placed at $x = +L$.
6. Use your result from step 5 to find the work you must do to bring the point charge from infinity to $x = +L$.

EVALUATE

7. What does your result from step 5 become in the limit $a \rightarrow 0$? Does this make sense?
8. Suppose the point charge q were negative rather than positive. How would this affect your result in step 4? In step 5?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q23.1 A student asked, "Since electrical potential is always proportional to potential energy, why bother with the concept of potential at all?" How would you respond?

Q23.2 The potential (relative to a point at infinity) midway between two charges of equal magnitude and opposite sign is zero. Is it possible to bring a test charge from infinity to this midpoint in such a way that no work is done in any part of the displacement? If so, describe how it can be done. If it is not possible, explain why.

Q23.3 Is it possible to have an arrangement of two point charges separated by a finite distance such that the electric potential energy of the arrangement is the same as if the two charges were infinitely far apart? Why or why not? What if there are three charges? Explain.

Q23.4 Since potential can have any value you want depending on the choice of the reference level of zero potential, how does a voltmeter know what to read when you connect it between two points?

Q23.5 If $\vec{E}$ is zero everywhere along a certain path that leads from point A to point B, what is the potential difference between those two points? Does this mean that $\vec{E}$ is zero everywhere along *any* path from A to B? Explain.

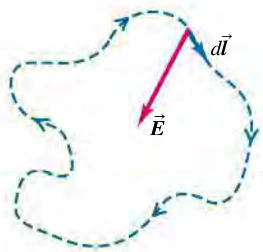
Q23.6 If $\vec{E}$ is zero throughout a certain region of space, is the potential necessarily also zero in this region? Why or why not? If not, what *can* be said about the potential?

Q23.7 Which way do electric field lines point, from high to low potential or from low to high? Explain.

Q23.8 (a) If the potential (relative to infinity) is zero at a point, is the electric field necessarily zero at that point? (b) If the electric field is zero at a point, is the potential (relative to infinity) necessarily zero there? Prove your answers, using simple examples.

Q23.9 If you carry out the integral of the electric field $\int \vec{E} \cdot d\vec{l}$ for a *closed* path like that shown in **Fig. Q23.9**, the integral will *always* be equal to zero, independent of the shape of the path and independent of where charges may be located relative to the path. Explain why.

Figure **Q23.9**



Q23.10 The potential difference between the two terminals of an AA battery (used in flashlights and portable stereos) is 1.5 V. If two AA batteries are placed end to end with the positive terminal of one battery touching the negative terminal of the other, what is the potential difference between the terminals at the exposed ends of the combination? What if the two positive terminals are touching each other? Explain your reasoning.

Q23.11 It is easy to produce a potential difference of several thousand volts between your body and the floor by scuffing your shoes across a nylon carpet. When you touch a metal doorknob, you get a mild shock. Yet contact with a power line of comparable voltage would probably be fatal. Why is there a difference?

Q23.12 If the electric potential at a single point is known, can $\vec{E}$ at that point be determined? If so, how? If not, why not?

Q23.13 Because electric field lines and equipotential surfaces are always perpendicular, two equipotential surfaces can never cross; if they did, the direction of $\vec{E}$ would be ambiguous at the crossing points. Yet two equipotential surfaces appear to cross at the center of Fig. 23.23c. Explain why there is no ambiguity about the direction of $\vec{E}$ in this particular case.

Q23.14 A uniform electric field is directed due east. Point B is 2.00 m west of point A, point C is 2.00 m east of point A, and point D is 2.00 m south of A. For each point, B, C, and D, is the potential at that point larger, smaller, or the same as at point A? Give the reasoning behind your answers.

Q23.15 We often say that if point A is at a higher potential than point B, A is at positive potential and B is at negative potential. Does it necessarily follow that a point at positive potential is positively charged, or that a point at negative potential is negatively charged? Illustrate your answers with clear, simple examples.

Q23.16 A conducting sphere is to be charged by bringing in positive charge a little at a time until the total charge is Q. The total work required for this process is alleged to be proportional to Q^2. Is this correct? Why or why not?

Q23.17 In electronics it is customary to define the potential of ground (thinking of the earth as a large conductor) as zero. Is this consistent with the fact that the earth has a net electric charge that is not zero? (Refer to Exercise 21.28.)

Q23.18 A conducting sphere is placed between two charged parallel plates such as those shown in Fig. 23.2. Does the electric field inside the sphere depend on precisely where between the plates the sphere is placed? What about the electric potential inside the sphere? Do the answers to these questions depend on whether or not there is a net charge on the sphere? Explain your reasoning.

Q23.19 A conductor that carries a net charge Q has a hollow, empty cavity in its interior. Does the potential vary from point to point within the material of the conductor? What about within the cavity? How does the potential inside the cavity compare to the potential within the material of the conductor?

Q23.20 A high-voltage dc power line falls on a car, so the entire metal body of the car is at a potential of 10,000 V with respect to the ground. What happens to the occupants (a) when they are sitting in the car and (b) when they step out of the car? Explain your reasoning.

Q23.21 When a thunderstorm is approaching, sailors at sea sometimes observe a phenomenon called "St. Elmo's fire," a bluish flickering light at the tips of masts. What causes this? Why does it occur at the tips of masts? Why is the effect most pronounced when the masts are wet? (*Hint:* Seawater is a good conductor of electricity.)

Q23.22 A positive point charge is placed near a very large conducting plane. A professor of physics asserted that the field caused by this configuration is the same as would be obtained by removing the plane and placing a negative point charge of equal magnitude in the mirror-image position behind the initial position of the plane. Is this correct? Why or why not? (*Hint:* Inspect Fig. 23.23b.)

EXERCISES

Section 23.1 Electric Potential Energy

23.1 •• A point charge $q_1 = +2.40\ \mu\text{C}$ is held stationary at the origin. A second point charge $q_2 = -4.30\ \mu\text{C}$ moves from the point $x = 0.150$ m, $y = 0$ to the point $x = 0.250$ m, $y = 0.250$ m. How much work is done by the electric force on q_2?

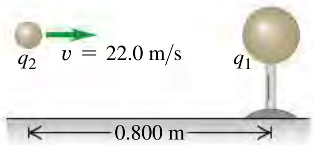
23.2 • A point charge q_1 is held stationary at the origin. A second charge q_2 is placed at point a, and the electric potential energy of the pair of charges is $+5.4 \times 10^{-8}$ J. When the second charge is moved to point b, the electric force on the charge does -1.9×10^{-8} J of work. What is the electric potential energy of the pair of charges when the second charge is at point b?

23.3 •• **Energy of the Nucleus.** How much work is needed to assemble an atomic nucleus containing three protons (such as Li) if we model it as an equilateral triangle of side 2.00×10^{-15} m with a proton at each vertex? Assume the protons started from very far away.

23.4 •• (a) How much work would it take to push two protons very slowly from a separation of 2.00×10^{-10} m (a typical atomic distance) to 3.00×10^{-15} m (a typical nuclear distance)? (b) If the protons are both released from rest at the closer distance in part (a), how fast are they moving when they reach their original separation?

23.5 •• A small metal sphere, carrying a net charge of $q_1 = -2.80\ \mu\text{C}$, is held in a stationary position by insulating supports. A second small metal sphere, with a net charge of $q_2 = -7.80\ \mu\text{C}$ and mass 1.50 g, is projected toward q_1. When the two spheres are 0.800 m apart, q_2, is moving toward q_1 with speed 22.0 m/s (**Fig. E23.5**). Assume that the two spheres can be treated as point charges. You can ignore the force of gravity. (a) What is the speed of q_2 when the spheres are 0.400 m apart? (b) How close does q_2 get to q_1?

Figure **E23.5**



23.6 •• BIO **Energy of DNA Base Pairing.** (See Exercise 21.21.) (a) Calculate the electric potential energy of the adenine–thymine bond, using the same combinations of molecules (O–H–N and N–H–N) as in Exercise 21.21. (b) Compare this energy with the potential energy of the proton–electron pair in the hydrogen atom.

23.7 •• Two protons, starting several meters apart, are aimed directly at each other with speeds of 2.00×10^5 m/s, measured relative to the earth. Find the maximum electric force that these protons will exert on each other.

23.8 •• Three equal 1.20-μC point charges are placed at the corners of an equilateral triangle with sides 0.400 m long. What is the potential energy of the system? (Take as zero the potential energy of the three charges when they are infinitely far apart.)

23.9 •• Two protons are released from rest when they are 0.750 nm apart. (a) What is the maximum speed they will reach? When does this speed occur? (b) What is the maximum acceleration they will achieve? When does this acceleration occur?

23.10 •• Four electrons are located at the corners of a square 10.0 nm on a side, with an alpha particle at its midpoint. How much work is needed to move the alpha particle to the midpoint of one of the sides of the square?

23.11 •• Three point charges, which initially are infinitely far apart, are placed at the corners of an equilateral triangle with sides d. Two of the point charges are identical and have charge q. If zero net work is required to place the three charges at the corners of the triangle, what must the value of the third charge be?

Section 23.2 Electric Potential

23.12 • An object with charge $q = -6.00 \times 10^{-9}$ C is placed in a region of uniform electric field and is released from rest at point A. After the charge has moved to point B, 0.500 m to the right, it has kinetic energy 3.00×10^{-7} J. (a) If the electric potential at point A is $+30.0$ V, what is the electric potential at point B? (b) What are the magnitude and direction of the electric field?

23.13 • A small particle has charge $-5.00\ \mu$C and mass 2.00×10^{-4} kg. It moves from point A, where the electric potential is $V_A = +200$ V, to point B, where the electric potential is $V_B = +800$ V. The electric force is the only force acting on the particle. The particle has speed 5.00 m/s at point A. What is its speed at point B? Is it moving faster or slower at B than at A? Explain.

23.14 • A particle with charge $+4.20$ nC is in a uniform electric field $\vec{E}$ directed to the left. The charge is released from rest and moves to the left; after it has moved 6.00 cm, its kinetic energy is $+2.20 \times 10^{-6}$ J. What are (a) the work done by the electric force, (b) the potential of the starting point with respect to the end point, and (c) the magnitude of $\vec{E}$?

23.15 • A charge of 28.0 nC is placed in a uniform electric field that is directed vertically upward and has a magnitude of 4.00×10^4 V/m. What work is done by the electric force when the charge moves (a) 0.450 m to the right; (b) 0.670 m upward; (c) 2.60 m at an angle of 45.0° downward from the horizontal?

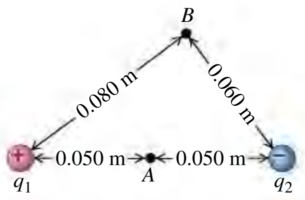
23.16 • Two stationary point charges $+3.00$ nC and $+2.00$ nC are separated by a distance of 50.0 cm. An electron is released from rest at a point midway between the two charges and moves along the line connecting the two charges. What is the speed of the electron when it is 10.0 cm from the $+3.00$-nC charge?

23.17 •• Point charges $q_1 = +2.00\ \mu$C and $q_2 = -2.00\ \mu$C are placed at adjacent corners of a square for which the length of each side is 3.00 cm. Point a is at the center of the square, and point b is at the empty corner closest to q_2. Take the electric potential to be zero at a distance far from both charges. (a) What is the electric potential at point a due to q_1 and q_2? (b) What is the electric potential at point b? (c) A point charge $q_3 = -5.00\ \mu$C moves from point a to point b. How much work is done on q_3 by the electric forces exerted by q_1 and q_2? Is this work positive or negative?

23.18 • Two point charges of equal magnitude Q are held a distance d apart. Consider only points on the line passing through both charges. (a) If the two charges have the same sign, find the location of all points (if there are any) at which (i) the potential (relative to infinity) is zero (is the electric field zero at these points?), and (ii) the electric field is zero (is the potential zero at these points?). (b) Repeat part (a) for two point charges having opposite signs.

23.19 • Two point charges $q_1 = +2.40$ nC and $q_2 = -6.50$ nC are 0.100 m apart. Point A is midway between them; point B is 0.080 m from q_1 and 0.060 m from q_2 (**Fig. E23.19**). Take the electric potential to be zero at infinity. Find (a) the potential at point A; (b) the potential at point B; (c) the work done by the electric field on a charge of 2.50 nC that travels from point B to point A.

Figure **E23.19**



23.20 •• (a) An electron is to be accelerated from 3.00×10^6 m/s to 8.00×10^6 m/s. Through what potential difference must the electron pass to accomplish this? (b) Through what potential difference must the electron pass if it is to be slowed from 8.00×10^6 m/s to a halt?

23.21 •• A positive charge q is fixed at the point $x = 0$, $y = 0$, and a negative charge $-2q$ is fixed at the point $x = a$, $y = 0$. (a) Show the positions of the charges in a diagram. (b) Derive an expression for the potential V at points on the x-axis as a function of the coordinate x. Take V to be zero at an infinite distance from the charges. (c) At which positions on the x-axis is $V = 0$? (d) Graph V at points on the x-axis as a function of x in the range from $x = -2a$ to $x = +2a$. (e) What does the answer to part (b) become when $x \gg a$? Explain why this result is obtained.

23.22 • At a certain distance from a point charge, the potential and electric-field magnitude due to that charge are 4.98 V and 16.2 V/m, respectively. (Take $V = 0$ at infinity.) (a) What is the distance to the point charge? (b) What is the magnitude of the charge? (c) Is the electric field directed toward or away from the point charge?

23.23 • A uniform electric field has magnitude E and is directed in the negative x-direction. The potential difference between point a (at $x = 0.60$ m) and point b (at $x = 0.90$ m) is 240 V. (a) Which point, a or b, is at the higher potential? (b) Calculate the value of E. (c) A negative point charge $q = -0.200\ \mu$C is moved from b to a. Calculate the work done on the point charge by the electric field.

23.24 • For each of the following arrangements of two point charges, find all the points along the line passing through both charges for which the electric potential V is zero (take $V = 0$ infinitely far from the charges) and for which the electric field E is zero: (a) charges $+Q$ and $+2Q$ separated by a distance d, and (b) charges $-Q$ and $+2Q$ separated by a distance d. (c) Are both V and E zero at the same places? Explain.

Section 23.3 Calculating Electric Potential

23.25 •• A thin spherical shell with radius $R_1 = 3.00$ cm is concentric with a larger thin spherical shell with radius $R_2 = 5.00$ cm. Both shells are made of insulating material. The smaller shell has charge $q_1 = +6.00$ nC distributed uniformly over its surface, and the larger shell has charge $q_2 = -9.00$ nC distributed uniformly over its surface. Take the electric potential to be zero at an infinite distance from both shells. (a) What is the electric potential due to the two shells at the following distance from their common center: (i) $r = 0$; (ii) $r = 4.00$ cm; (iii) $r = 6.00$ cm? (b) What is the magnitude of the potential difference between the surfaces of the two shells? Which shell is at higher potential: the inner shell or the outer shell?

23.26 • A total electric charge of 3.50 nC is distributed uniformly over the surface of a metal sphere with a radius of 24.0 cm. If the potential is zero at a point at infinity, find the value of the potential at the following distances from the center of the sphere: (a) 48.0 cm; (b) 24.0 cm; (c) 12.0 cm.

23.27 •• A uniformly charged, thin ring has radius 15.0 cm and total charge $+24.0$ nC. An electron is placed on the ring's axis a distance 30.0 cm from the center of the ring and is constrained to stay on the axis of the ring. The electron is then released from rest. (a) Describe the subsequent motion of the electron. (b) Find the speed of the electron when it reaches the center of the ring.

23.28 • A solid conducting sphere has net positive charge and radius $R = 0.400$ m. At a point 1.20 m from the center of the sphere, the electric potential due to the charge on the sphere is 24.0 V. Assume that $V = 0$ at an infinite distance from the sphere. What is the electric potential at the center of the sphere?

23.29 • Charge $Q = 5.00\ \mu C$ is distributed uniformly over the volume of an insulating sphere that has radius $R = 12.0$ cm. A small sphere with charge $q = +3.00\ \mu C$ and mass 6.00×10^{-5} kg is projected toward the center of the large sphere from an initial large distance. The large sphere is held at a fixed position and the small sphere can be treated as a point charge. What minimum speed must the small sphere have in order to come within 8.00 cm of the surface of the large sphere?

23.30 •• An infinitely long line of charge has linear charge density 5.00×10^{-12} C/m. A proton (mass 1.67×10^{-27} kg, charge $+1.60 \times 10^{-19}$ C) is 18.0 cm from the line and moving directly toward the line at 3.50×10^3 m/s. (a) Calculate the proton's initial kinetic energy. (b) How close does the proton get to the line of charge?

23.31 • A very long wire carries a uniform linear charge density λ. Using a voltmeter to measure potential difference, you find that when one probe of the meter is placed 2.50 cm from the wire and the other probe is 1.00 cm farther from the wire, the meter reads 575 V. (a) What is λ? (b) If you now place one probe at 3.50 cm from the wire and the other probe 1.00 cm farther away, will the voltmeter read 575 V? If not, will it read more or less than 575 V? Why? (c) If you place both probes 3.50 cm from the wire but 17.0 cm from each other, what will the voltmeter read?

23.32 •• A very long insulating cylinder of charge of radius 2.50 cm carries a uniform linear density of 15.0 nC/m. If you put one probe of a voltmeter at the surface, how far from the surface must the other probe be placed so that the voltmeter reads 175 V?

23.33 •• A very long insulating cylindrical shell of radius 6.00 cm carries charge of linear density $8.50\ \mu C/m$ spread uniformly over its outer surface. What would a voltmeter read if it were connected between (a) the surface of the cylinder and a point 4.00 cm above the surface, and (b) the surface and a point 1.00 cm from the central axis of the cylinder?

23.34 • A ring of diameter 8.00 cm is fixed in place and carries a charge of $+5.00\ \mu C$ uniformly spread over its circumference. (a) How much work does it take to move a tiny $+3.00$-μC charged ball of mass 1.50 g from very far away to the center of the ring? (b) Is it necessary to take a path along the axis of the ring? Why? (c) If the ball is slightly displaced from the center of the ring, what will it do and what is the maximum speed it will reach?

23.35 •• A very small sphere with positive charge $q = +8.00\ \mu C$ is released from rest at a point 1.50 cm from a very long line of uniform linear charge density $\lambda = +3.00\ \mu C/m$. What is the kinetic energy of the sphere when it is 4.50 cm from the line of charge if the only force on it is the force exerted by the line of charge?

23.36 • **CP** Two large, parallel conducting plates carrying opposite charges of equal magnitude are separated by 2.20 cm. (a) If the surface charge density for each plate has magnitude $47.0\ nC/m^2$, what is the magnitude of $\vec{E}$ in the region between the plates? (b) What is the potential difference between the two plates? (c) If the separation between the plates is doubled while the surface charge density is kept constant at the value in part (a), what happens to the magnitude of the electric field and to the potential difference?

23.37 • Two large, parallel, metal plates carry opposite charges of equal magnitude. They are separated by 45.0 mm, and the potential difference between them is 360 V. (a) What is the magnitude of the electric field (assumed to be uniform) in the region between the plates? (b) What is the magnitude of the force this field exerts on a particle with charge +2.40 nC? (c) Use the results of part (b) to compute the work done by the field on the particle as it moves from the higher-potential plate to the lower. (d) Compare the result of part (c) to the change of potential energy of the same charge, computed from the electric potential.

23.38 • **BIO** **Electrical Sensitivity of Sharks.** Certain sharks can detect an electric field as weak as $1.0\ \mu V/m$. To grasp how weak this field is, if you wanted to produce it between two parallel metal plates by connecting an ordinary 1.5-V AA battery across these plates, how far apart would the plates have to be?

23.39 • The electric field at the surface of a charged, solid, copper sphere with radius 0.200 m is 3800 N/C, directed toward the center of the sphere. What is the potential at the center of the sphere, if we take the potential to be zero infinitely far from the sphere?

23.40 •• (a) How much excess charge must be placed on a copper sphere 25.0 cm in diameter so that the potential of its center, relative to infinity, is 3.75 kV? (b) What is the potential of the sphere's surface relative to infinity?

Section 23.4 Equipotential Surfaces and Section 23.5 Potential Gradient

23.41 •• **CALC** A metal sphere with radius r_a is supported on an insulating stand at the center of a hollow, metal, spherical shell with radius r_b. There is charge $+q$ on the inner sphere and charge $-q$ on the outer spherical shell. (a) Calculate the potential $V(r)$ for (i) $r < r_a$; (ii) $r_a < r < r_b$; (iii) $r > r_b$. (*Hint:* The net potential is the sum of the potentials due to the individual spheres.) Take V to be zero when r is infinite. (b) Show that the potential of the inner sphere with respect to the outer is

$$V_{ab} = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r_a} - \frac{1}{r_b}\right)$$

(c) Use Eq. (23.23) and the result from part (a) to show that the electric field at any point between the spheres has magnitude

$$E(r) = \frac{V_{ab}}{(1/r_a - 1/r_b)}\frac{1}{r^2}$$

(d) Use Eq. (23.23) and the result from part (a) to find the electric field at a point outside the larger sphere at a distance r from the center, where $r > r_b$. (e) Suppose the charge on the outer sphere is not $-q$ but a negative charge of different magnitude, say $-Q$. Show that the answers for parts (b) and (c) are the same as before but the answer for part (d) is different.

23.42 • A very large plastic sheet carries a uniform charge density of $-6.00\ nC/m^2$ on one face. (a) As you move away from the sheet along a line perpendicular to it, does the potential increase or decrease? How do you know, without doing any calculations? Does your answer depend on where you choose the reference point for potential? (b) Find the spacing between equipotential surfaces that differ from each other by 1.00 V. What type of surfaces are these?

23.43 • **CALC** In a certain region of space, the electric potential is $V(x, y, z) = Axy - Bx^2 + Cy$, where A, B, and C are positive constants. (a) Calculate the x-, y-, and z-components of the electric field. (b) At which points is the electric field equal to zero?

23.44 • **CALC** In a certain region of space the electric potential is given by $V = +Ax^2y - Bxy^2$, where $A = 5.00\ V/m^3$ and $B = 8.00\ V/m^3$. Calculate the magnitude and direction of the electric field at the point in the region that has coordinates $x = 2.00$ m, $y = 0.400$ m, and $z = 0$.

23.45 • A metal sphere with radius $r_a = 1.20$ cm is supported on an insulating stand at the center of a hollow, metal, spherical shell with radius $r_b = 9.60$ cm. Charge $+q$ is put on the inner sphere and charge $-q$ on the outer spherical shell. The magnitude of q is chosen to make the potential difference between the spheres 500 V, with the inner sphere at higher potential. (a) Use the result of Exercise 23.41(b) to calculate q. (b) With the help of the result of Exercise 23.41(a), sketch the equipotential surfaces that correspond to 500, 400, 300, 200, 100, and 0 V. (c) In your sketch, show the electric field lines. Are the electric field lines and equipotential surfaces mutually perpendicular? Are the equipotential surfaces closer together when the magnitude of $\vec{E}$ is largest?

PROBLEMS

23.46 • **CP** A point charge $q_1 = +5.00\ \mu\text{C}$ is held fixed in space. From a horizontal distance of 6.00 cm, a small sphere with mass 4.00×10^{-3} kg and charge $q_2 = +2.00\ \mu\text{C}$ is fired toward the fixed charge with an initial speed of 40.0 m/s. Gravity can be neglected. What is the acceleration of the sphere at the instant when its speed is 25.0 m/s?

23.47 ••• A point charge $q_1 = 4.00$ nC is placed at the origin, and a second point charge $q_2 = -3.00$ nC is placed on the x-axis at $x = +20.0$ cm. A third point charge $q_3 = 2.00$ nC is to be placed on the x-axis between q_1 and q_2. (Take as zero the potential energy of the three charges when they are infinitely far apart.) (a) What is the potential energy of the system of the three charges if q_3 is placed at $x = +10.0$ cm? (b) Where should q_3 be placed to make the potential energy of the system equal to zero?

23.48 •• A positive point charge $q_1 = +5.00 \times 10^{-4}$ C is held at a fixed position. A small object with mass 4.00×10^{-3} kg and charge $q_2 = -3.00 \times 10^{-4}$ C is projected directly at q_1. Ignore gravity. When q_2 is 0.400 m away, its speed is 800 m/s. What is its speed when it is 0.200 m from q_1?

23.49 •• A gold nucleus has a radius of 7.3×10^{-15} m and a charge of $+79e$. Through what voltage must an alpha particle, with charge $+2e$, be accelerated so that it has just enough energy to reach a distance of 2.0×10^{-14} m from the surface of a gold nucleus? (Assume that the gold nucleus remains stationary and can be treated as a point charge.)

23.50 ••• A small sphere with mass 5.00×10^{-7} kg and charge $+7.00\ \mu\text{C}$ is released from rest a distance of 0.400 m above a large horizontal insulating sheet of charge that has uniform surface charge density $\sigma = +8.00\ \text{pC/m}^2$. Using energy methods, calculate the speed of the sphere when it is 0.100 m above the sheet.

23.51 •• **Determining the Size of the Nucleus.** When radium-226 decays radioactively, it emits an alpha particle (the nucleus of helium), and the end product is radon-222. We can model this decay by thinking of the radium-226 as consisting of an alpha particle emitted from the surface of the spherically symmetric radon-222 nucleus, and we can treat the alpha particle as a point charge. The energy of the alpha particle has been measured in the laboratory and has been found to be 4.79 MeV when the alpha particle is essentially infinitely far from the nucleus. Since radon is much heavier than the alpha particle, we can assume that there is no appreciable recoil of the radon after the decay. The radon nucleus contains 86 protons, while the alpha particle has 2 protons and the radium nucleus has 88 protons. (a) What was the electric potential energy of the alpha–radon combination just before the decay, in MeV and in joules? (b) Use your result from part (a) to calculate the radius of the radon nucleus.

23.52 •• **CP** A proton and an alpha particle are released from rest when they are 0.225 nm apart. The alpha particle (a helium nucleus) has essentially four times the mass and two times the charge of a proton. Find the maximum *speed* and maximum *acceleration* of each of these particles. When do these maxima occur: just following the release of the particles or after a very long time?

23.53 • A particle with charge +7.60 nC is in a uniform electric field directed to the left. Another force, in addition to the electric force, acts on the particle so that when it is released from rest, it moves to the right. After it has moved 8.00 cm, the additional force has done 6.50×10^{-5} J of work and the particle has 4.35×10^{-5} J of kinetic energy. (a) What work was done by the electric force? (b) What is the potential of the starting point with respect to the end point? (c) What is the magnitude of the electric field?

23.54 •• Identical charges $q = +5.00\ \mu\text{C}$ are placed at opposite corners of a square that has sides of length 8.00 cm. Point A is at one of the empty corners, and point B is at the center of the square. A charge $q_0 = -3.00\ \mu\text{C}$ is placed at point A and moves along the diagonal of the square to point B. (a) What is the magnitude of the net electric force on q_0 when it is at point A? Sketch the placement of the charges and the direction of the net force. (b) What is the magnitude of the net electric force on q_0 when it is at point B? (c) How much work does the electric force do on q_0 during its motion from A to B? Is this work positive or negative? When it goes from A to B, does q_0 move to higher potential or to lower potential?

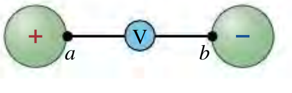
23.55 •• **CALC** A vacuum tube diode consists of concentric cylindrical electrodes, the negative cathode and the positive anode. Because of the accumulation of charge near the cathode, the electric potential between the electrodes is given by

$$V(x) = Cx^{4/3}$$

where x is the distance from the cathode and C is a constant, characteristic of a particular diode and operating conditions. Assume that the distance between the cathode and anode is 13.0 mm and the potential difference between electrodes is 240 V. (a) Determine the value of C. (b) Obtain a formula for the electric field between the electrodes as a function of x. (c) Determine the force on an electron when the electron is halfway between the electrodes.

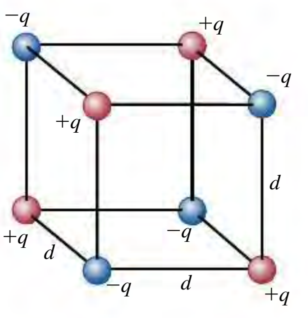
23.56 •• Two oppositely charged, identical insulating spheres, each 50.0 cm in diameter and carrying a uniformly distributed charge of magnitude 250 μC, are placed 1.00 m apart center to center (**Fig. P23.56**). (a) If a voltmeter is connected between the nearest points (a and b) on their surfaces, what will it read? (b) Which point, a or b, is at the higher potential? How can you know this without any calculations?

Figure **P23.56**



23.57 •• **An Ionic Crystal. Figure P23.57** shows eight point charges arranged at the corners of a cube with sides of length d. The values of the charges are $+q$ and $-q$, as shown. This is a model of one cell of a cubic ionic crystal. In sodium chloride (NaCl), for instance, the positive ions are Na^+ and the negative ions are Cl^-. (a) Calculate the potential energy U of this arrangement. (Take as zero the potential energy of the eight charges when they are infinitely far apart.) (b) In part (a), you should have found that $U < 0$. Explain the relationship between this result and the observation that such ionic crystals exist in nature.

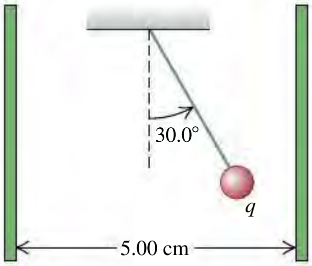
Figure **P23.57**



23.58 • (a) Calculate the potential energy of a system of two small spheres, one carrying a charge of $2.00\,\mu\text{C}$ and the other a charge of $-3.50\,\mu\text{C}$, with their centers separated by a distance of 0.180 m. Assume that $U = 0$ when the charges are infinitely separated. (b) Suppose that one sphere is held in place; the other sphere, with mass 1.50 g, is shot away from it. What minimum initial speed would the moving sphere need to escape completely from the attraction of the fixed sphere? (To escape, the moving sphere would have to reach a velocity of zero when it is infinitely far from the fixed sphere.)

23.59 •• **CP** A small sphere with mass 1.50 g hangs by a thread between two very large parallel vertical plates 5.00 cm apart (**Fig. P23.59**). The plates are insulating and have uniform surface charge densities $+\sigma$ and $-\sigma$. The charge on the sphere is $q = 8.90 \times 10^{-6}\,\text{C}$. What potential difference between the plates will cause the thread to assume an angle of 30.0° with the vertical?

Figure **P23.59**



23.60 •• Two spherical shells have a common center. The inner shell has radius $R_1 = 5.00$ cm and charge $q_1 = +3.00 \times 10^{-6}$ C; the outer shell has radius $R_2 = 15.0$ cm and charge $q_2 = -5.00 \times 10^{-6}$ C. Both charges are spread uniformly over the shell surface. What is the electric potential due to the two shells at the following distances from their common center: (a) $r = 2.50$ cm; (b) $r = 10.0$ cm; (c) $r = 20.0$ cm? Take $V = 0$ at a large distance from the shells.

23.61 • **CALC** **Coaxial Cylinders.** A long metal cylinder with radius a is supported on an insulating stand on the axis of a long, hollow, metal tube with radius b. The positive charge per unit length on the inner cylinder is λ, and there is an equal negative charge per unit length on the outer cylinder. (a) Calculate the potential $V(r)$ for (i) $r < a$; (ii) $a < r < b$; (iii) $r > b$. (*Hint:* The net potential is the sum of the potentials due to the individual conductors.) Take $V = 0$ at $r = b$. (b) Show that the potential of the inner cylinder with respect to the outer is

$$V_{ab} = \frac{\lambda}{2\pi\epsilon_0}\ln\frac{b}{a}$$

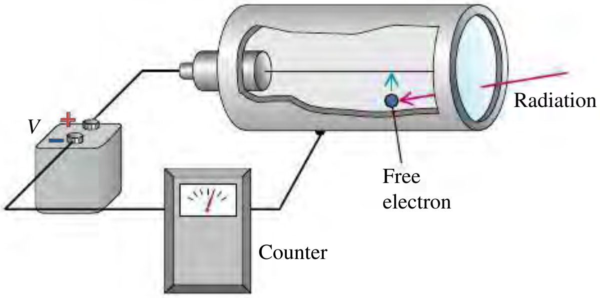
(c) Use Eq. (23.23) and the result from part (a) to show that the electric field at any point between the cylinders has magnitude

$$E(r) = \frac{V_{ab}}{\ln(b/a)}\frac{1}{r}$$

(d) What is the potential difference between the two cylinders if the outer cylinder has no net charge?

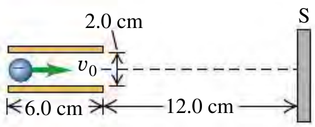
23.62 •• A *Geiger counter* detects radiation such as alpha particles by using the fact that the radiation ionizes the air along its path. A thin wire lies on the axis of a hollow metal cylinder and is insulated from it (**Fig. P23.62**). A large potential difference is established between the wire and the outer cylinder, with the wire at higher potential; this sets up a strong electric field directed radially outward. When ionizing radiation enters the device, it ionizes a few air molecules. The free electrons produced are accelerated by the electric field toward the wire and, on the way there, ionize many more air molecules. Thus a current pulse is produced that can be detected by appropriate electronic circuitry and converted to an audible "click." Suppose the radius of the central wire is 145 μm and the radius of the hollow cylinder is 1.80 cm. What potential difference between the wire and the cylinder produces an electric field of 2.00×10^4 V/m at a distance of 1.20 cm from the axis of the wire? (The wire and cylinder are both very long in comparison to their radii, so the results of Problem 23.61 apply.)

Figure **P23.62**



23.63 • **CP** **Deflection in a CRT.** Cathode-ray tubes (CRTs) were often found in oscilloscopes and computer monitors. In **Fig. P23.63** an electron with an initial speed of 6.50×10^6 m/s is projected along the axis midway between the deflection plates of a cathode-ray tube. The potential difference between the two plates is 22.0 V and the lower plate is the one at higher potential. (a) What is the force (magnitude and direction) on the electron when it is between the plates? (b) What is the acceleration of the electron (magnitude and direction) when acted on by the force in part (a)? (c) How far below the axis has the electron moved when it reaches the end of the plates? (d) At what angle with the axis is it moving as it leaves the plates? (e) How far below the axis will it strike the fluorescent screen S?

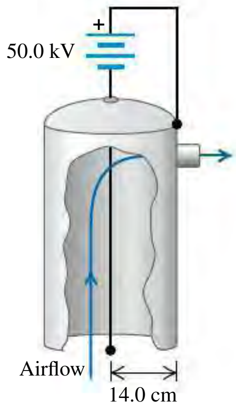
Figure **P23.63**



23.64 •• **CP** **Deflecting Plates of an Oscilloscope.** The vertical deflecting plates of a typical classroom oscilloscope are a pair of parallel square metal plates carrying equal but opposite charges. Typical dimensions are about 3.0 cm on a side, with a separation of about 5.0 mm. The potential difference between the plates is 25.0 V. The plates are close enough that we can ignore fringing at the ends. Under these conditions: (a) how much charge is on each plate, and (b) how strong is the electric field between the plates? (c) If an electron is ejected at rest from the negative plate, how fast is it moving when it reaches the positive plate?

23.65 •• *Electrostatic precipitators* use electric forces to remove pollutant particles from smoke, in particular in the smokestacks of coal-burning power plants. One form of precipitator consists of a vertical, hollow, metal cylinder with a thin wire, insulated from the cylinder, running along its axis (**Fig. P23.65**). A large potential difference is established between the wire and the outer cylinder, with the wire at lower potential. This sets up a strong radial electric field directed inward. The field produces a region of ionized air near the wire. Smoke enters the precipitator at the bottom, ash and dust in it pick up

Figure **P23.65**



electrons, and the charged pollutants are accelerated toward the outer cylinder wall by the electric field. Suppose the radius of the central wire is 90.0 μm, the radius of the cylinder is 14.0 cm, and a potential difference of 50.0 kV is established between the wire and the cylinder. Also assume that the wire and cylinder are both very long in comparison to the cylinder radius, so the results of Problem 23.61 apply. (a) What is the magnitude of the electric field midway between the wire and the cylinder wall? (b) What magnitude of charge must a 30.0-μg ash particle have if the electric field computed in part (a) is to exert a force ten times the weight of the particle?

23.66 •• CALC A disk with radius R has uniform surface charge density σ. (a) By regarding the disk as a series of thin concentric rings, calculate the electric potential V at a point on the disk's axis a distance x from the center of the disk. Assume that the potential is zero at infinity. (*Hint:* Use the result of Example 23.11 in Section 23.3.) (b) Calculate $-\partial V/\partial x$. Show that the result agrees with the expression for E_x calculated in Example 21.11 (Section 21.5).

23.67 ••• CALC **Self-Energy of a Sphere of Charge.** A solid sphere of radius R contains a total charge Q distributed uniformly throughout its volume. Find the energy needed to assemble this charge by bringing infinitesimal charges from far away. This energy is called the "self-energy" of the charge distribution. (*Hint:* After you have assembled a charge q in a sphere of radius r, how much energy would it take to add a spherical shell of thickness dr having charge dq? Then integrate to get the total energy.)

23.68 • CALC A thin insulating rod is bent into a semicircular arc of radius a, and a total electric charge Q is distributed uniformly along the rod. Calculate the potential at the center of curvature of the arc if the potential is assumed to be zero at infinity.

23.69 •• Charge $Q = +4.00\ \mu$C is distributed uniformly over the volume of an insulating sphere that has radius $R = 5.00$ cm. What is the potential difference between the center of the sphere and the surface of the sphere?

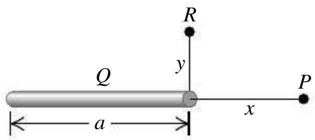
23.70 • An insulating spherical shell with inner radius 25.0 cm and outer radius 60.0 cm carries a charge of $+150.0\ \mu$C uniformly distributed over its outer surface. Point a is at the center of the shell, point b is on the inner surface, and point c is on the outer surface. (a) What will a voltmeter read if it is connected between the following points: (i) a and b; (ii) b and c; (iii) c and infinity; (iv) a and c? (b) Which is at higher potential: (i) a or b; (ii) b or c; (iii) a or c? (c) Which, if any, of the answers would change sign if the charge were $-150\ \mu$C?

23.71 •• CP Two plastic spheres, each carrying charge uniformly distributed throughout its interior, are initially placed in contact and then released. One sphere is 60.0 cm in diameter, has mass 50.0 g, and contains $-10.0\ \mu$C of charge. The other sphere is 40.0 cm in diameter, has mass 150.0 g, and contains $-30.0\ \mu$C of charge. Find the maximum acceleration and the maximum speed achieved by each sphere (relative to the fixed point of their initial location in space), assuming that no other forces are acting on them. (*Hint:* The uniformly distributed charges behave as though they were concentrated at the centers of the two spheres.)

23.72 • (a) If a spherical raindrop of radius 0.650 mm carries a charge of -3.60 pC uniformly distributed over its volume, what is the potential at its surface? (Take the potential to be zero at an infinite distance from the raindrop.) (b) Two identical raindrops, each with radius and charge specified in part (a), collide and merge into one larger raindrop. What is the radius of this larger drop, and what is the potential at its surface, if its charge is uniformly distributed over its volume?

23.73 • CALC Electric charge is distributed uniformly along a thin rod of length a, with total charge Q. Take the potential to be zero at infinity. Find the potential at the following points (**Fig. P23.73**): (a) point P, a distance x to the right of the rod, and (b) point R, a distance y above the right-hand end of the rod. (c) In parts (a) and (b), what does your result reduce to as x or y becomes much larger than a?

Figure **P23.73**



23.74 • An alpha particle with kinetic energy 9.50 MeV (when far away) collides head-on with a lead nucleus at rest. What is the distance of closest approach of the two particles? (Assume that the lead nucleus remains stationary and may be treated as a point charge. The atomic number of lead is 82. The alpha particle is a helium nucleus, with atomic number 2.)

23.75 •• Two metal spheres of different sizes are charged such that the electric potential is the same at the surface of each. Sphere A has a radius three times that of sphere B. Let Q_A and Q_B be the charges on the two spheres, and let E_A and E_B be the electric-field magnitudes at the surfaces of the two spheres. What are (a) the ratio Q_B/Q_A and (b) the ratio E_B/E_A?

23.76 • A metal sphere with radius R_1 has a charge Q_1. Take the electric potential to be zero at an infinite distance from the sphere. (a) What are the electric field and electric potential at the surface of the sphere? This sphere is now connected by a long, thin conducting wire to another sphere of radius R_2 that is several meters from the first sphere. Before the connection is made, this second sphere is uncharged. After electrostatic equilibrium has been reached, what are (b) the total charge on each sphere; (c) the electric potential at the surface of each sphere; (d) the electric field at the surface of each sphere? Assume that the amount of charge on the wire is much less than the charge on each sphere.

23.77 •• CP **Nuclear Fusion in the Sun.** The source of the sun's energy is a sequence of nuclear reactions that occur in its core. The first of these reactions involves the collision of two protons, which fuse together to form a heavier nucleus and release energy. For this process, called *nuclear fusion,* to occur, the two protons must first approach until their surfaces are essentially in contact. (a) Assume both protons are moving with the same speed and they collide head-on. If the radius of the proton is 1.2×10^{-15} m, what is the minimum speed that will allow fusion to occur? The charge distribution within a proton is spherically symmetric, so the electric field and potential outside a proton are the same as if it were a point charge. The mass of the proton is 1.67×10^{-27} kg. (b) Another nuclear fusion reaction that occurs in the sun's core involves a collision between two helium nuclei, each of which has 2.99 times the mass of the proton, charge $+2e$, and radius 1.7×10^{-15} m. Assuming the same collision geometry as in part (a), what minimum speed is required for this fusion reaction to take place if the nuclei must approach a center-to-center distance of about 3.5×10^{-15} m? As for the proton, the charge of the helium nucleus is uniformly distributed throughout its volume. (c) In Section 18.3 it was shown that the average translational kinetic energy of a particle with mass m in a gas at absolute temperature T is $\frac{3}{2}kT$, where k is the Boltzmann constant (given in Appendix F). For two protons with kinetic energy equal to this average value to be able to undergo the process described in part (a), what absolute temperature is required? What absolute temperature is required for two average helium nuclei to be able to undergo the process described in part (b)? (At these temperatures, atoms are completely ionized, so nuclei and electrons move separately.) (d) The temperature in the sun's core is about 1.5×10^7 K. How does this compare to the temperatures calculated in part (c)? How can the reactions described in parts (a) and (b) occur at all in the interior of the sun? (*Hint:* See the discussion of the distribution of molecular speeds in Section 18.5.)

23.78 • **CALC** The electric potential V in a region of space is given by

$$V(x, y, z) = A(x^2 - 3y^2 + z^2)$$

where A is a constant. (a) Derive an expression for the electric field $\vec{E}$ at any point in this region. (b) The work done by the field when a 1.50-μC test charge moves from the point $(x, y, z) = (0, 0, 0.250 \text{ m})$ to the origin is measured to be 6.00×10^{-5} J. Determine A. (c) Determine the electric field at the point (0, 0, 0.250 m). (d) Show that in every plane parallel to the xz-plane the equipotential contours are circles. (e) What is the radius of the equipotential contour corresponding to $V = 1280$ V and $y = 2.00$ m?

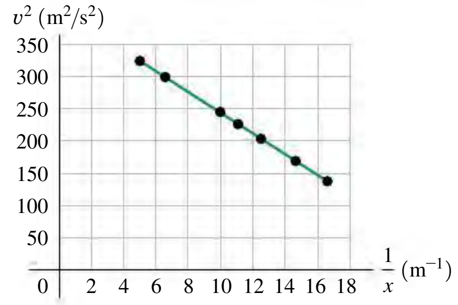
23.79 •• **DATA** The electric potential in a region that is within 2.00 m of the origin of a rectangular coordinate system is given by $V = Ax^l + By^m + Cz^n + D$, where $A, B, C, D, l, m,$ and n are constants. The units of $A, B, C,$ and D are such that if $x, y,$ and z are in meters, then V is in volts. You measure V and each component of the electric field at four points and obtain these results:

Point	(x, y, z) (m)	V (V)	E_x (V/m)	E_y (V/m)	E_z (V/m)
1	(0, 0, 0)	10.0	0	0	0
2	(1.00, 0, 0)	4.0	12.0	0	0
3	(0, 1.00, 0)	6.0	0	12.0	0
4	(0, 0, 1.00)	8.0	0	0	12.0

(a) Use the data in the table to calculate $A, B, C, D, l, m,$ and n. (b) What are V and the magnitude of E at the points $(0, 0, 0)$, $(0.50 \text{ m}, 0.50 \text{ m}, 0.50 \text{ m})$, and $(1.00 \text{ m}, 1.00 \text{ m}, 1.00 \text{ m})$?

23.80 •• **DATA** A small, stationary sphere carries a net charge Q. You perform the following experiment to measure Q: From a large distance you fire a small particle with mass $m = 4.00 \times 10^{-4}$ kg and charge $q = 5.00 \times 10^{-8}$ C directly at the center of the sphere. The apparatus you are using measures the particle's speed v as a function of the distance x from the sphere. The sphere's mass is much greater than the mass of the projectile particle, so you assume that the sphere remains at rest. All of the measured values of x are much larger than the radius of either object, so you treat both objects as point particles. You plot your data on a graph of v^2 versus $(1/x)$ (**Fig. P23.80**). The straight line $v^2 = 400 \text{ m}^2/\text{s}^2 - [(15.75 \text{ m}^3/\text{s}^2)/x]$ gives a good fit to the data points. (a) Explain why the graph is a straight line. (b) What is the initial speed v_0 of the particle when it is very far from the sphere? (c) What is Q? (d) How close does the particle get to the sphere? Assume that this distance is much larger than the radii of the particle and sphere, so continue to treat them as point particles and to assume that the sphere remains at rest.

Figure **P23.80**



23.81 ••• **DATA** **The Millikan Oil-Drop Experiment.** The charge of an electron was first measured by the American physicist Robert Millikan during 1909–1913. In his experiment, oil was sprayed in very fine drops (about 10^{-4} mm in diameter) into the space between two parallel horizontal plates separated by a distance d. A potential difference V_{AB} was maintained between the plates, causing a downward electric field between them. Some of the oil drops acquired a negative charge because of frictional effects or because of ionization of the surrounding air by x rays or radioactivity. The drops were observed through a microscope. (a) Show that an oil drop of radius r at rest between the plates remained at rest if the magnitude of its charge was

$$q = \frac{4\pi}{3}\frac{\rho r^3 g d}{V_{AB}}$$

where ρ is oil's density. (Ignore the buoyant force of the air.) By adjusting V_{AB} to keep a given drop at rest, Millikan determined the charge on that drop, provided its radius r was known. (b) Millikan's oil drops were much too small to measure their radii directly. Instead, Millikan determined r by cutting off the electric field and measuring the *terminal speed* v_t of the drop as it fell. (We discussed terminal speed in Section 5.3.) The viscous force F on a sphere of radius r moving at speed v through a fluid with viscosity η is given by Stokes's law: $F = 6\pi\eta r v$. When a drop fell at v_t, the viscous force just balanced the drop's weight $w = mg$. Show that the magnitude of the charge on the drop was

$$q = 18\pi\frac{d}{V_{AB}}\sqrt{\frac{\eta^3 v_t^3}{2\rho g}}$$

(c) You repeat the Millikan oil-drop experiment. Four of your measured values of V_{AB} and v_t are listed in the table:

Drop	1	2	3	4
V_{AB} (V)	9.16	4.57	12.32	6.28
v_t (10^{-5} m/s)	2.54	0.767	4.39	1.52

In your apparatus, the separation d between the horizontal plates is 1.00 mm. The density of the oil you use is 824 kg/m^3. For the viscosity η of air, use the value 1.81×10^{-5} N·s/m^2. Assume that $g = 9.80$ m/s^2. Calculate the charge q of each drop. (d) If electric charge is *quantized* (that is, exists in multiples of the magnitude of the charge of an electron), then the charge on each drop is $-ne$, where n is the number of excess electrons on each drop. (All four drops in your table have negative charge.) Drop 2 has the smallest magnitude of charge observed in the experiment, for all 300 drops on which measurements were made, so assume that its charge is due to an excess charge of one electron. Determine the number of excess electrons n for each of the other three drops. (e) Use $q = -ne$ to calculate e from the data for each of the four drops, and average these four values to get your best experimental value of e.

CHALLENGE PROBLEMS

23.82 ••• **CALC** A hollow, thin-walled insulating cylinder of radius R and length L (like the cardboard tube in a roll of toilet paper) has charge Q uniformly distributed over its surface. (a) Calculate the electric potential at all points along the axis of the tube. Take the origin to be at the center of the tube, and take the potential to be zero at infinity. (b) Show that if $L \ll R$, the result of part (a) reduces to the potential on the axis of a ring of charge of radius R. (See Example 23.11 in Section 23.3.) (c) Use the result of part (a) to find the electric field at all points along the axis of the tube.

23.83 ••• **CP** In experiments in which atomic nuclei collide, head-on collisions like that described in Problem 23.74 do happen, but "near misses" are more common. Suppose the alpha particle in that problem is not "aimed" at the center of the lead nucleus but has an initial nonzero angular momentum (with respect to the stationary lead nucleus) of magnitude $L = p_0 b$, where p_0 is the magnitude of the particle's initial momentum and $b = 1.00 \times 10^{-12}$ m. What is the distance of closest approach? Repeat for $b = 1.00 \times 10^{-13}$ m and $b = 1.00 \times 10^{-14}$ m.

PASSAGE PROBLEMS

MATERIALS ANALYSIS WITH IONS. *Rutherford backscattering spectrometry* (RBS) is a technique used to determine the structure and composition of materials. A beam of ions (typically helium ions) is accelerated to high energy and aimed at a sample. By analyzing the distribution and energy of the ions that are scattered from (that is, deflected by collisions with) the atoms in the sample, researchers can determine the sample's composition. To accelerate the ions to high energies, a *tandem electrostatic accelerator* may be used. In this device, negative ions (He^-) start at a potential $V = 0$ and are accelerated by a high positive voltage at the midpoint of the accelerator. The high voltage produces a constant electric field in the acceleration tube through which the ions move. When accelerated ions reach the midpoint, the electrons are stripped off, turning the negative ions into doubly positively charged ions (He^{++}). These positive ions are then repelled from the midpoint by the high positive voltage there and continue to accelerate to the far end of the accelerator, where again $V = 0$.

23.84 For a particular experiment, helium ions are to be given a kinetic energy of 3.0 MeV. What should the voltage at the center of the accelerator be, assuming that the ions start essentially at rest? (a) −3.0 MV; (b) +3.0 MV; (c) +1.5 MV; (d) +1.0 MV.

23.85 A helium ion (He^{++}) that comes within about 10 fm of the center of the nucleus of an atom in the sample may induce a nuclear reaction instead of simply scattering. Imagine a helium ion with a kinetic energy of 3.0 MeV heading straight toward an atom at rest in the sample. Assume that the atom stays fixed. What minimum charge can the nucleus of the atom have such that the helium ion gets no closer than 10 fm from the center of the atomic nucleus? ($1 \text{ fm} = 1 \times 10^{-15}$ m, and e is the magnitude of the charge of an electron or a proton.) (a) $2e$; (b) $11e$; (c) $20e$; (d) $22e$.

23.86 The maximum voltage at the center of a typical tandem electrostatic accelerator is 6.0 MV. If the distance from one end of the acceleration tube to the midpoint is 12 m, what is the magnitude of the average electric field in the tube under these conditions? (a) 41,000 V/m; (b) 250,000 V/m; (c) 500,000 V/m; (d) 6,000,000 V/m.

Answers

Chapter Opening Question ?

(iii) A large, constant potential difference V_{ab} is maintained between the welding tool (a) and the metal pieces to be welded (b). For a given potential difference between two conductors a and b, the smaller the distance d separating the conductors, the greater is the magnitude E of the field between them. Hence d must be small in order for E to be large enough to ionize the gas between the conductors (see Section 23.3) and produce an arc through this gas.

Test Your Understanding Questions

23.1 **(a) (i), (b) (ii)** The three charges q_1, q_2, and q_3 are all positive, so all three of the terms in the sum in Eq. (23.11)—q_1q_2/r_{12}, q_1q_3/r_{13}, and q_2q_3/r_{23}—are positive. Hence the total electric potential energy U is positive. This means that it would take positive work to bring the three charges from infinity to the positions shown in Fig. 21.14, and hence *negative* work to move the three charges from these positions back to infinity.

23.2 **no** If $V = 0$ at a certain point, $\vec{E}$ does *not* have to be zero at that point. An example is point c in Figs. 21.23 and 23.13, for which there is an electric field in the $+x$-direction (see Example 21.9 in Section 21.5) even though $V = 0$ (see Example 23.4). This isn't a surprising result because V and $\vec{E}$ are quite different quantities: V is the net amount of work required to bring a unit charge from infinity to the point in question, whereas $\vec{E}$ is the electric force that acts on a unit charge when it arrives at that point.

23.3 **no** If $\vec{E} = \mathbf{0}$ at a certain point, V does *not* have to be zero at that point. An example is point O at the center of the charged ring in Figs. 21.23 and Fig. 23.21. From Example 21.9 (Section 21.5), the electric field is zero at O because the electric-field contributions from different parts of the ring completely cancel. From Example 23.11, however, the potential at O is *not* zero: This point corresponds to $x = 0$, so $V = (1/4\pi\epsilon_0)(Q/a)$. This value of V corresponds to the work that would have to be done to move a unit positive test charge along a path from infinity to point O; it is nonzero because the charged ring repels the test charge, so positive work must be done to move the test charge toward the ring.

23.4 **no** If the positive charges in Fig. 23.23 were replaced by negative charges, and vice versa, the equipotential surfaces would be the same but the sign of the potential would be reversed. For example, the surfaces in Fig. 23.23b with potential $V = +30$ V and $V = -50$ V would have potential $V = -30$ V and $V = +50$ V, respectively.

23.5 **(iii)** From Eqs. (23.19), the components of the electric field are $E_x = -\partial V/\partial x = -(B + Dy)$, $E_y = -\partial V/\partial y = -(3Cy^2 + Dx)$, and $E_z = -\partial V/\partial z = 0$. The value of A has no effect, which means that we can add a constant to the electric potential at all points without changing $\vec{E}$ or the potential difference between two points. The potential does not depend on z, so the z-component of $\vec{E}$ is zero. Note that at the origin the electric field is not zero because it has a nonzero x-component: $E_x = -B$, $E_y = 0$, $E_z = 0$.

Bridging Problem

$$\frac{qQ}{8\pi\epsilon_0 a}\ln\left(\frac{L+a}{L-a}\right)$$

? In flash photography, the energy used to make the flash is stored in a capacitor, which consists of two closely spaced conductors that carry opposite charges. If the amount of charge on the conductors is doubled, by what factor does the stored energy increase? (i) $\sqrt{2}$; (ii) 2; (iii) $2\sqrt{2}$; (iv) 4; (v) 8.

24 CAPACITANCE AND DIELECTRICS

LEARNING GOALS

Looking forward at ...

24.1 The nature of capacitors, and how to calculate a quantity that measures their ability to store charge.

24.2 How to analyze capacitors connected in a network.

24.3 How to calculate the amount of energy stored in a capacitor.

24.4 What dielectrics are, and how they make capacitors more effective.

24.5 How a dielectric inside a charged capacitor becomes polarized.

24.6 How to use Gauss's laws when dielectrics are present.

Looking back at ...

21.2, 21.5, 21.7 Polarization; field of charged conductors; electric dipoles.

22.3–22.5 Gauss's law.

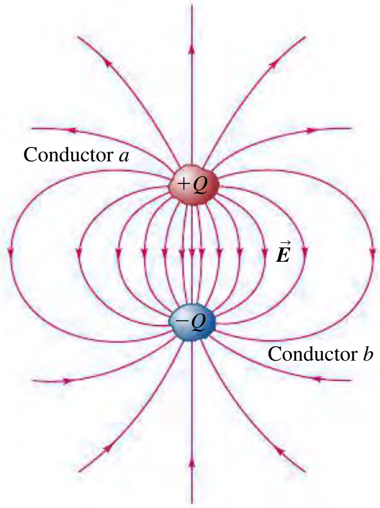
23.3, 23.4 Potential for charged conductors; potential due to a cylindrical charge distribution.

When you stretch the rubber band of a slingshot or pull back the string of an archer's bow, you are storing mechanical energy as elastic potential energy. A capacitor is a device that stores *electric* potential energy and electric charge. To make a capacitor, just insulate two conductors from each other. To store energy in this device, transfer charge from one conductor to the other so that one has a negative charge and the other has an equal amount of positive charge. Work must be done to move the charges through the resulting potential difference between the conductors, and the work done is stored as electric potential energy.

Capacitors have a tremendous number of practical applications in devices such as electronic flash units for photography, pulsed lasers, air bag sensors for cars, and radio and television receivers. We'll encounter many of these applications in later chapters (particularly Chapter 31, in which we'll see the crucial role played by capacitors in the alternating-current circuits that pervade our technological society). In this chapter, however, our emphasis is on the fundamental properties of capacitors. For a particular capacitor, the ratio of the charge on each conductor to the potential difference between the conductors is a constant, called the *capacitance*. The capacitance depends on the sizes and shapes of the conductors and on the insulating material (if any) between them. Compared to the case in which there is only vacuum between the conductors, the capacitance increases when an insulating material (a *dielectric*) is present. This happens because a redistribution of charge, called *polarization,* takes place within the insulating material. Studying polarization will give us added insight into the electrical properties of matter.

Capacitors also give us a new way to think about electric potential energy. The energy stored in a charged capacitor is related to the electric field in the space between the conductors. We will see that electric potential energy can be regarded as being stored *in the field itself.* The idea that the electric field is itself a storehouse of energy is at the heart of the theory of electromagnetic waves and our modern understanding of the nature of light, to be discussed in Chapter 32.

24.1 Any two conductors a and b insulated from each other form a capacitor.



24.1 CAPACITORS AND CAPACITANCE

Any two conductors separated by an insulator (or a vacuum) form a **capacitor** (**Fig. 24.1**). In most practical applications, each conductor initially has zero net charge and electrons are transferred from one conductor to the other; this is called *charging* the capacitor. Then the two conductors have charges with equal magnitude and opposite sign, and the *net* charge on the capacitor as a whole remains zero. We will assume throughout this chapter that this is the case. When we say that a capacitor has charge Q, or that a charge Q is *stored* on the capacitor, we mean that the conductor at higher potential has charge $+Q$ and the conductor at lower potential has charge $-Q$ (assuming that Q is positive). Keep this in mind in the following discussion and examples.

In circuit diagrams a capacitor is represented by either of these symbols:

The vertical lines (straight or curved) represent the conductors, and the horizontal lines represent wires connected to either conductor. One common way to charge a capacitor is to connect these two wires to opposite terminals of a battery. Once the charges Q and $-Q$ are established on the conductors, the battery is disconnected. This gives a fixed *potential difference* V_{ab} between the conductors (that is, the potential of the positively charged conductor a with respect to the negatively charged conductor b) that is just equal to the voltage of the battery.

The electric field at any point in the region between the conductors is proportional to the magnitude Q of charge on each conductor. It follows that the potential difference V_{ab} between the conductors is also proportional to Q. If we double the magnitude of charge on each conductor, the charge density at each point doubles, the electric field at each point doubles, and the potential difference between conductors doubles; however, the *ratio* of charge to potential difference does not change. This ratio is called the **capacitance** C of the capacitor:

Capacitance of a capacitor

$$C = \frac{Q}{V_{ab}} \tag{24.1}$$

Magnitude of charge on each conductor

Potential difference between conductors (a has charge $+Q$, b has charge $-Q$)

The SI unit of capacitance is called one **farad** (1 F), in honor of the 19th-century English physicist Michael Faraday. From Eq. (24.1), one farad is equal to one *coulomb per volt* (1 C/V):

$$1\ \text{F} = 1\ \text{farad} = 1\ \text{C/V} = 1\ \text{coulomb/volt}$$

CAUTION Capacitance vs. coulombs Don't confuse the symbol C for capacitance (which is always in italics) with the abbreviation C for coulombs (which is never italicized).

The greater the capacitance C of a capacitor, the greater the magnitude Q of charge on either conductor for a given potential difference V_{ab} and hence the greater the amount of stored energy. (Remember that potential is potential energy per unit charge.) Thus *capacitance is a measure of the ability of a capacitor to store energy*. We will see that the capacitance depends only on the shapes, sizes, and relative positions of the conductors and on the nature of the insulator between them. (For special types of insulating materials, the capacitance *does* depend on Q and V_{ab}. We won't discuss these materials, however.)

Calculating Capacitance: Capacitors in Vacuum

We can calculate the capacitance C of a given capacitor by finding the potential difference V_{ab} between the conductors for a given magnitude of charge Q and then using Eq. (24.1). For now we'll consider only *capacitors in vacuum;* that is, empty space separates the conductors that make up the capacitor.

The simplest form of capacitor consists of two parallel conducting plates, each with area A, separated by a distance d that is small in comparison with their dimensions (**Fig. 24.2a**). When the plates are charged, the electric field is almost completely localized in the region between the plates (Fig. 24.2b). As we discussed in Example 22.8 (Section 22.4), the field between such plates is essentially *uniform,* and the charges on the plates are uniformly distributed over their opposing surfaces. We call this arrangement a **parallel-plate capacitor.**

We found the electric-field magnitude E for this arrangement in Example 21.12 (Section 21.5) by using the principle of superposition of electric fields and again in Example 22.8 (Section 22.4) by using Gauss's law. It would be a good idea to review those examples. We found that $E = \sigma/\epsilon_0$, where σ is the magnitude (absolute value) of the surface charge density on each plate. This is equal to the magnitude of the total charge Q on each plate divided by the area A of the plate, or $\sigma = Q/A$, so the field magnitude E can be expressed as

$$E = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A}$$

The field is uniform and the distance between the plates is d, so the potential difference (voltage) between the two plates is

$$V_{ab} = Ed = \frac{1}{\epsilon_0}\frac{Qd}{A}$$

Thus

Capacitance of a parallel-plate capacitor in vacuum

$$C = \frac{Q}{V_{ab}} = \epsilon_0 \frac{A}{d} \qquad (24.2)$$

Magnitude of charge on each plate
Area of each plate
Distance between plates
Electric constant
Potential difference between plates

The capacitance depends on only the geometry of the capacitor; it is directly proportional to the area A of each plate and inversely proportional to their separation d. The quantities A and d are constants for a given capacitor, and ϵ_0 is a universal constant. Thus in vacuum the capacitance C is a constant independent of the charge on the capacitor or the potential difference between the plates. If one of the capacitor plates is flexible, C changes as the plate separation d changes. This is the operating principle of a condenser microphone (**Fig. 24.3**).

When matter is present between the plates, its properties affect the capacitance. We will return to this topic in Section 24.4. Meanwhile, we remark that if the space contains air at atmospheric pressure instead of vacuum, the capacitance differs from the prediction of Eq. (24.2) by less than 0.06%.

In Eq. (24.2), if A is in square meters and d in meters, then C is in farads. The units of the electric constant ϵ_0 are $C^2/N \cdot m^2$, so

$$1\ F = 1\ C^2/N \cdot m = 1\ C^2/J$$

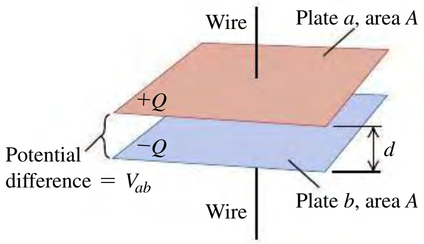
Because $1\ V = 1\ J/C$ (energy per unit charge), this equivalence agrees with our definition $1\ F = 1\ C/V$. Finally, we can express the units of ϵ_0 as $1\ C^2/N \cdot m^2 = 1\ F/m$, so

$$\epsilon_0 = 8.85 \times 10^{-12}\ F/m$$

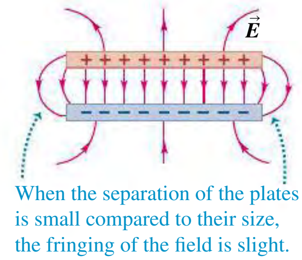
This relationship is useful in capacitance calculations, and it also helps us to verify that Eq. (24.2) is dimensionally consistent.

24.2 A charged parallel-plate capacitor.

(a) Arrangement of the capacitor plates



(b) Side view of the electric field $\vec{E}$



24.3 Inside a condenser microphone is a capacitor with one rigid plate and one flexible plate. The two plates are kept at a constant potential difference V_{ab}. Sound waves cause the flexible plate to move back and forth, varying the capacitance C and causing charge to flow to and from the capacitor in accordance with the relationship $C = Q/V_{ab}$. Thus a sound wave is converted to a charge flow that can be amplified and recorded digitally.

24.4 A commercial capacitor is labeled with the value of its capacitance. For these capacitors, $C = 2200\ \mu\text{F}$, $1000\ \mu\text{F}$, and $470\ \mu\text{F}$.

One farad is a very large capacitance, as the following example shows. In many applications the most convenient units of capacitance are the *microfarad* ($1\ \mu\text{F} = 10^{-6}\ \text{F}$) and the *picofarad* ($1\ \text{pF} = 10^{-12}\ \text{F}$). For example, the flash unit in a point-and-shoot camera uses a capacitor of a few hundred microfarads (**Fig. 24.4**), while capacitances in a radio tuning circuit are typically from 10 to 100 picofarads.

For *any* capacitor in vacuum, the capacitance C depends only on the shapes, dimensions, and separation of the conductors that make up the capacitor. If the conductor shapes are more complex than those of the parallel-plate capacitor, the expression for capacitance is more complicated than in Eq. (24.2). In the following examples we show how to calculate C for two other conductor geometries.

EXAMPLE 24.1 SIZE OF A 1-F CAPACITOR

SOLUTION

The parallel plates of a 1.0-F capacitor are 1.0 mm apart. What is their area?

SOLUTION

IDENTIFY and SET UP: This problem uses the relationship among the capacitance C, plate separation d, and plate area A (our target variable) for a parallel-plate capacitor. We solve Eq. (24.2) for A.

EXECUTE: From Eq. (24.2),

$$A = \frac{Cd}{\epsilon_0} = \frac{(1.0\ \text{F})(1.0 \times 10^{-3}\ \text{m})}{8.85 \times 10^{-12}\ \text{F/m}} = 1.1 \times 10^{8}\ \text{m}^2$$

EVALUATE: This corresponds to a square about 10 km (about 6 miles) on a side! The volume of such a capacitor would be at least $Ad = 1.1 \times 10^5\ \text{m}^3$, equivalent to that of a cube about 50 m on a side. In fact, it's possible to make 1-F capacitors a few *centimeters* on a side. The trick is to have an appropriate substance between the plates rather than a vacuum, so that (among other things) the plate separation d can be greatly reduced. We'll explore this further in Section 24.4.

EXAMPLE 24.2 PROPERTIES OF A PARALLEL-PLATE CAPACITOR

SOLUTION

The plates of a parallel-plate capacitor in vacuum are 5.00 mm apart and $2.00\ \text{m}^2$ in area. A 10.0-kV potential difference is applied across the capacitor. Compute (a) the capacitance; (b) the charge on each plate; and (c) the magnitude of the electric field between the plates.

SOLUTION

IDENTIFY and SET UP: We are given the plate area A, the plate spacing d, and the potential difference $V_{ab} = 1.00 \times 10^4\ \text{V}$ for this parallel-plate capacitor. Our target variables are the capacitance C, the charge Q on each plate, and the electric-field magnitude E. We use Eq. (24.2) to calculate C and then use Eq. (24.1) and V_{ab} to find Q. We use $E = Q/\epsilon_0 A$ to find E.

EXECUTE: (a) From Eq. (24.2),

$$C = \epsilon_0 \frac{A}{d} = (8.85 \times 10^{-12}\ \text{F/m}) \frac{(2.00\ \text{m}^2)}{5.00 \times 10^{-3}\ \text{m}}$$
$$= 3.54 \times 10^{-9}\ \text{F} = 0.00354\ \mu\text{F}$$

(b) The charge on the capacitor is

$$Q = CV_{ab} = (3.54 \times 10^{-9}\ \text{C/V})(1.00 \times 10^4\ \text{V})$$
$$= 3.54 \times 10^{-5}\ \text{C} = 35.4\ \mu\text{C}$$

The plate at higher potential has charge $+35.4\ \mu\text{C}$, and the other plate has charge $-35.4\ \mu\text{C}$.

(c) The electric-field magnitude is

$$E = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A} = \frac{3.54 \times 10^{-5}\ \text{C}}{(8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2)(2.00\ \text{m}^2)}$$
$$= 2.00 \times 10^{6}\ \text{N/C}$$

EVALUATE: We can also find E by recalling that the electric field is equal in magnitude to the potential gradient [Eq. (23.22)]. The field between the plates is uniform, so

$$E = \frac{V_{ab}}{d} = \frac{1.00 \times 10^4\ \text{V}}{5.00 \times 10^{-3}\ \text{m}} = 2.00 \times 10^{6}\ \text{V/m}$$

(Remember that $1\ \text{N/C} = 1\ \text{V/m}$.)

SOLUTION

EXAMPLE 24.3 A SPHERICAL CAPACITOR

Two concentric spherical conducting shells are separated by vacuum (**Fig. 24.5**). The inner shell has total charge $+Q$ and outer radius r_a, and the outer shell has charge $-Q$ and inner radius r_b. Find the capacitance of this spherical capacitor.

SOLUTION

IDENTIFY and SET UP: By definition, the capacitance C is the magnitude Q of the charge on either sphere divided by the potential difference V_{ab} between the spheres. We first find V_{ab}, and then use Eq. (24.1) to find the capacitance $C = Q/V_{ab}$.

EXECUTE: Using a Gaussian surface such as that shown in Fig. 24.5, we found in Example 22.5 (Section 22.4) that the charge on a conducting sphere produces zero field *inside* the sphere, so the outer sphere makes no contribution to the field between the spheres. Therefore the electric field *and* the electric potential between the shells are the same as those outside a charged conducting sphere with charge $+Q$. We considered that problem in Example 23.8 (Section 23.3), so the same result applies here: The potential at any point between the spheres is $V = Q/4\pi\epsilon_0 r$. Hence the potential of the inner (positive) conductor at $r = r_a$ with respect to that of the outer (negative) conductor at $r = r_b$ is

$$V_{ab} = V_a - V_b = \frac{Q}{4\pi\epsilon_0 r_a} - \frac{Q}{4\pi\epsilon_0 r_b}$$

$$= \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_a} - \frac{1}{r_b}\right) = \frac{Q}{4\pi\epsilon_0}\frac{r_b - r_a}{r_a r_b}$$

The capacitance is then

$$C = \frac{Q}{V_{ab}} = 4\pi\epsilon_0 \frac{r_a r_b}{r_b - r_a}$$

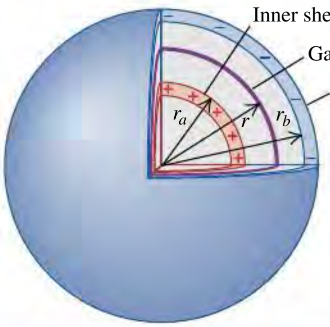
As an example, if $r_a = 9.5$ cm and $r_b = 10.5$ cm,

$$C = 4\pi(8.85 \times 10^{-12}\ \text{F/m})\frac{(0.095\ \text{m})(0.105\ \text{m})}{0.010\ \text{m}}$$

$$= 1.1 \times 10^{-10}\ \text{F} = 110\ \text{pF}$$

EVALUATE: We can relate our expression for C to that for a parallel-plate capacitor. The quantity $4\pi r_a r_b$ is intermediate between the areas $4\pi r_a^2$ and $4\pi r_b^2$ of the two spheres; in fact, it's the *geometric mean* of these two areas, which we can denote by A_{gm}. The distance between spheres is $d = r_b - r_a$, so we can write $C = 4\pi\epsilon_0 r_a r_b/(r_b - r_a) = \epsilon_0 A_{\text{gm}}/d$. This has the same form as for parallel plates: $C = \epsilon_0 A/d$. If the distance between spheres is very small in comparison to their radii, their capacitance is the same as that of parallel plates with the same area and spacing.

24.5 A spherical capacitor.



SOLUTION

EXAMPLE 24.4 A CYLINDRICAL CAPACITOR

Two long, coaxial cylindrical conductors are separated by vacuum (**Fig. 24.6**). The inner cylinder has radius r_a and linear charge density $+\lambda$. The outer cylinder has inner radius r_b and linear charge density $-\lambda$. Find the capacitance per unit length for this capacitor.

SOLUTION

IDENTIFY and SET UP: As in Example 24.3, we use the definition of capacitance, $C = Q/V_{ab}$. We use the result of Example 23.10 (Section 23.3) to find the potential difference V_{ab} between the cylinders, and find the charge Q on a length L of the cylinders from the linear charge density. We then find the corresponding capacitance C from Eq. (24.1). Our target variable is this capacitance divided by L.

EXECUTE: As in Example 24.3, the potential V between the cylinders is not affected by the presence of the charged outer cylinder. Hence our result in Example 23.10 for the potential outside a charged conducting cylinder also holds in this example for potential in the space between the cylinders:

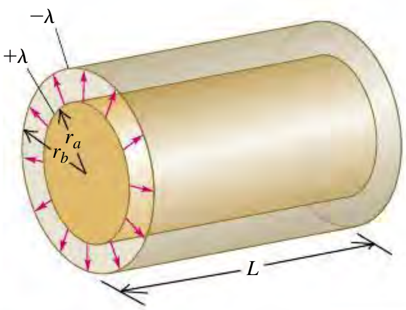
$$V = \frac{\lambda}{2\pi\epsilon_0}\ln\frac{r_0}{r}$$

Here r_0 is the arbitrary, *finite* radius at which $V = 0$. We take $r_0 = r_b$, the radius of the inner surface of the outer cylinder. Then the potential at the outer surface of the inner cylinder (at which $r = r_a$) is just the potential V_{ab} of the inner (positive) cylinder a with respect to the outer (negative) cylinder b:

$$V_{ab} = \frac{\lambda}{2\pi\epsilon_0}\ln\frac{r_b}{r_a}$$

Continued

24.6 A long cylindrical capacitor. The linear charge density λ is assumed to be positive in this figure. The magnitude of charge in a length L of either cylinder is λL.



If λ is positive as in Fig. 24.6, then V_{ab} is positive as well: The inner cylinder is at higher potential than the outer.

The total charge Q in a length L is $Q = \lambda L$, so from Eq. (24.1) the capacitance C of a length L is

$$C = \frac{Q}{V_{ab}} = \frac{\lambda L}{\dfrac{\lambda}{2\pi\epsilon_0}\ln\dfrac{r_b}{r_a}} = \frac{2\pi\epsilon_0 L}{\ln(r_b/r_a)}$$

The capacitance per unit length is

$$\frac{C}{L} = \frac{2\pi\epsilon_0}{\ln(r_b/r_a)}$$

Substituting $\epsilon_0 = 8.85 \times 10^{-12}\ \text{F/m} = 8.85\ \text{pF/m}$, we get

$$\frac{C}{L} = \frac{55.6\ \text{pF/m}}{\ln(r_b/r_a)}$$

EVALUATE: The capacitance of coaxial cylinders is determined entirely by their dimensions, just as for parallel-plate and spherical capacitors. Ordinary coaxial cables are made like this but with an insulating material instead of vacuum between the conductors. A typical cable used for connecting a television to a cable TV feed has a capacitance per unit length of 69 pF/m.

24.7 An assortment of commercially available capacitors.

TEST YOUR UNDERSTANDING OF SECTION 24.1 A capacitor has vacuum in the space between the conductors. If you double the amount of charge on each conductor, what happens to the capacitance? (i) It increases; (ii) it decreases; (iii) it remains the same; (iv) the answer depends on the size or shape of the conductors. ❙

24.2 CAPACITORS IN SERIES AND PARALLEL

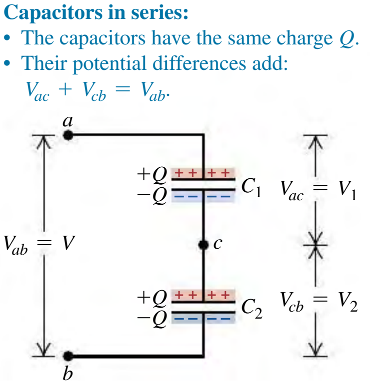
Capacitors are manufactured with certain standard capacitances and working voltages (**Fig. 24.7**). However, these standard values may not be the ones you actually need in a particular application. You can obtain the values you need by combining capacitors; many combinations are possible, but the simplest combinations are a series connection and a parallel connection.

Capacitors in Series

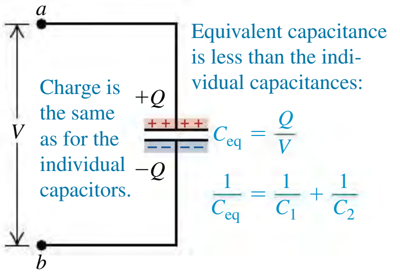
Figure 24.8a is a schematic diagram of a **series connection.** Two capacitors are connected in series (one after the other) by conducting wires between points a and b. Both capacitors are initially uncharged. When a constant positive potential difference V_{ab} is applied between points a and b, the capacitors become charged; the figure shows that the charge on *all* conducting plates has the same magnitude. To see why, note first that the top plate of C_1 acquires a positive charge Q. The electric field of this positive charge pulls negative charge up to the bottom plate of C_1 until all of the field lines that begin on the top plate end on the bottom plate. This requires that the bottom plate have charge $-Q$. These negative charges had to come from the top plate of C_2, which becomes positively charged with charge $+Q$. This positive charge then pulls negative charge $-Q$ from the connection at point b onto the bottom plate of C_2. The total charge on the lower plate of C_1 and the upper plate of C_2 together must always be zero because these plates aren't connected to anything except each other. Thus *in a series connection the magnitude of charge on all plates is the same.*

24.8 A series connection of two capacitors.

(a) Two capacitors in series



(b) The equivalent single capacitor



Referring to Fig. 24.8a, we can write the potential differences between points a and c, c and b, and a and b as

$$V_{ac} = V_1 = \frac{Q}{C_1}, \qquad V_{cb} = V_2 = \frac{Q}{C_2}, \qquad V_{ab} = V = V_1 + V_2 = Q\left(\frac{1}{C_1} + \frac{1}{C_2}\right)$$

and so

$$\frac{V}{Q} = \frac{1}{C_1} + \frac{1}{C_2} \tag{24.3}$$

Following a common convention, we use the symbols V_1, V_2, and V to denote the potential *differences* V_{ac} (across the first capacitor), V_{cb} (across the second capacitor), and V_{ab} (across the entire combination of capacitors), respectively.

The **equivalent capacitance** C_{eq} of the series combination is defined as the capacitance of a *single* capacitor for which the charge Q is the same as for the combination, when the potential difference V is the same. In other words, the combination can be replaced by an *equivalent capacitor* of capacitance C_{eq}. For such a capacitor, shown in Fig. 24.8b,

$$C_{eq} = \frac{Q}{V} \quad \text{or} \quad \frac{1}{C_{eq}} = \frac{V}{Q} \tag{24.4}$$

Combining Eqs. (24.3) and (24.4), we find

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2}$$

We can extend this analysis to any number of capacitors in series. We find the following result for the *reciprocal* of the equivalent capacitance:

Capacitors in series:

$$\underbrace{\frac{1}{C_{eq}}}_{\text{Equivalent capacitance of series combination}} = \underbrace{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots}_{\text{Capacitances of individual capacitors}} \tag{24.5}$$

The reciprocal of the equivalent capacitance of a series combination equals the sum of the reciprocals of the individual capacitances. In a series connection the equivalent capacitance is always *less than* any individual capacitance.

CAUTION Capacitors in series The magnitude of charge is the same on all plates of all the capacitors in a series combination; however, the potential differences of the individual capacitors are *not* the same unless their individual capacitances are the same. The potential differences of the individual capacitors add to give the total potential difference across the series combination: $V_{total} = V_1 + V_2 + V_3 + \cdots$.

Application Touch Screens and Capacitance The touch screen on a mobile phone, an MP3 player, or (as shown here) a medical device uses the physics of capacitors. Behind the screen are two parallel layers, one behind the other, of thin strips of a transparent conductor such as indium tin oxide. A voltage is maintained between the two layers. The strips in one layer are oriented perpendicular to those in the other layer; the points where two strips overlap act as a grid of capacitors. When you bring your finger (a conductor) up to a point on the screen, your finger and the front conducting layer act like a second capacitor in series at that point. The circuitry attached to the conducting layers detects the location of the capacitance change, and so detects where you touched the screen.

Capacitors in Parallel

The arrangement shown in **Fig. 24.9a** is called a **parallel connection.** Two capacitors are connected in parallel between points a and b. In this case the upper plates of the two capacitors are connected by conducting wires to form an equipotential surface, and the lower plates form another. Hence *in a parallel connection the potential difference for all individual capacitors is the same* and is equal to $V_{ab} = V$. The charges Q_1 and Q_2 are not necessarily equal, however, since charges can reach each capacitor independently from the source (such as a battery) of the voltage V_{ab}. The charges are

$$Q_1 = C_1 V \quad \text{and} \quad Q_2 = C_2 V$$

The *total* charge Q of the combination, and thus the total charge on the equivalent capacitor, is

$$Q = Q_1 + Q_2 = (C_1 + C_2)V$$

so

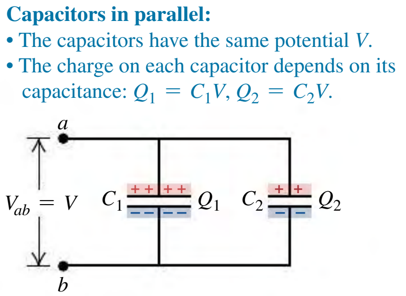
$$\frac{Q}{V} = C_1 + C_2 \tag{24.6}$$

The parallel combination is equivalent to a single capacitor with the same total charge $Q = Q_1 + Q_2$ and potential difference V as the combination (Fig. 24.9b). The equivalent capacitance of the combination, C_{eq}, is the same as the capacitance Q/V of this single equivalent capacitor. So from Eq. (24.6),

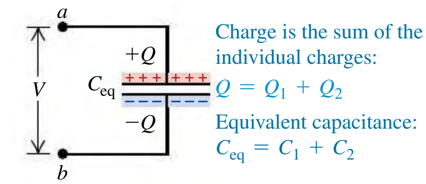
$$C_{eq} = C_1 + C_2$$

24.9 A parallel connection of two capacitors.

(a) Two capacitors in parallel



(b) The equivalent single capacitor



In the same way we can show that for any number of capacitors in parallel,

Capacitors in parallel:

$$\underbrace{C_{eq}}_{\text{Equivalent capacitance of parallel combination}} = \underbrace{C_1 + C_2 + C_3 + \cdots}_{\text{Capacitances of individual capacitors}} \qquad (24.7)$$

The equivalent capacitance of a parallel combination equals the *sum* of the individual capacitances. In a parallel connection the equivalent capacitance is always *greater than* any individual capacitance.

DEMO

CAUTION **Capacitors in parallel** The potential differences are the same for all capacitors in a parallel combination; however, the charges on individual capacitors are *not* the same unless their individual capacitances are the same. The charges on the individual capacitors add to give the total charge on the parallel combination: $Q_{total} = Q_1 + Q_2 + Q_3 + \cdots$. [Compare these statements to those in the "Caution" paragraph following Eq. (24.5).]

PROBLEM-SOLVING STRATEGY 24.1 EQUIVALENT CAPACITANCE

IDENTIFY *the relevant concepts:* The concept of equivalent capacitance is useful whenever two or more capacitors are connected.

SET UP *the problem* using the following steps:
1. Make a drawing of the capacitor arrangement.
2. Identify all groups of capacitors that are connected in series or in parallel.
3. Keep in mind that when we say a capacitor "has charge Q," we mean that the plate at higher potential has charge $+Q$ and the other plate has charge $-Q$.

EXECUTE *the solution* as follows:
1. Use Eq. (24.5) to find the equivalent capacitance of capacitors connected in series, as in Fig. 24.8. Such capacitors each have the *same charge* if they were uncharged before they were connected; that charge is the same as that on the equivalent capacitor. The potential difference across the combination is the sum of the potential differences across the individual capacitors.
2. Use Eq. (24.7) to find the equivalent capacitance of capacitors connected in parallel, as in Fig. 24.9. Such capacitors all have the *same potential difference* across them; that potential difference is the same as that across the equivalent capacitor. The total charge on the combination is the sum of the charges on the individual capacitors.
3. After replacing all the series or parallel groups you initially identified, you may find that more such groups reveal themselves. Replace those groups by using the same procedure as above until no more replacements are possible. If you then need to find the charge or potential difference for an individual original capacitor, you may have to retrace your steps.

EVALUATE *your answer:* Check whether your result makes sense. If the capacitors are connected in series, the equivalent capacitance C_{eq} must be *smaller* than any of the individual capacitances. If the capacitors are connected in parallel, C_{eq} must be *greater* than any of the individual capacitances.

EXAMPLE 24.5 CAPACITORS IN SERIES AND IN PARALLEL

SOLUTION

In Figs. 24.8 and 24.9, let $C_1 = 6.0\ \mu\text{F}$, $C_2 = 3.0\ \mu\text{F}$, and $V_{ab} = 18\ \text{V}$. Find the equivalent capacitance and the charge and potential difference for each capacitor when the capacitors are connected (a) in series (see Fig. 24.8) and (b) in parallel (see Fig. 24.9).

SOLUTION

IDENTIFY and SET UP: In both parts of this example a target variable is the equivalent capacitance C_{eq}, which is given by Eq. (24.5) for the series combination in part (a) and by Eq. (24.7) for the parallel combination in part (b). In each part we find the charge and potential difference from the definition of capacitance, Eq. (24.1), and the rules outlined in Problem-Solving Strategy 24.1.

EXECUTE: (a) From Eq. (24.5) for a series combination,

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} = \frac{1}{6.0\ \mu\text{F}} + \frac{1}{3.0\ \mu\text{F}} \qquad C_{eq} = 2.0\ \mu\text{F}$$

The charge Q on each capacitor in series is the same as that on the equivalent capacitor:

$$Q = C_{eq}V = (2.0\ \mu\text{F})(18\ \text{V}) = 36\ \mu\text{C}$$

The potential difference across each capacitor is inversely proportional to its capacitance:

$$V_{ac} = V_1 = \frac{Q}{C_1} = \frac{36\ \mu\text{C}}{6.0\ \mu\text{F}} = 6.0\ \text{V}$$

$$V_{cb} = V_2 = \frac{Q}{C_2} = \frac{36\ \mu\text{C}}{3.0\ \mu\text{F}} = 12.0\ \text{V}$$

(b) From Eq. (24.7) for a parallel combination,

$$C_{eq} = C_1 + C_2 = 6.0\ \mu\text{F} + 3.0\ \mu\text{F} = 9.0\ \mu\text{F}$$

The potential difference across each of the capacitors is the same as that across the equivalent capacitor, 18 V. The charge on each capacitor is directly proportional to its capacitance:

$$Q_1 = C_1V = (6.0\ \mu\text{F})(18\ \text{V}) = 108\ \mu\text{C}$$

$$Q_2 = C_2V = (3.0\ \mu\text{F})(18\ \text{V}) = 54\ \mu\text{C}$$

EVALUATE: As expected, the equivalent capacitance C_{eq} for the series combination in part (a) is less than either C_1 or C_2, while that for the parallel combination in part (b) is greater than either C_1 or C_2. For two capacitors in series, as in part (a), the charge is the same on either capacitor and the *larger* potential difference appears across the capacitor with the *smaller* capacitance. Furthermore, the sum of the potential differences across the individual capacitors in series equals the potential difference across the equivalent capacitor: $V_{ac} + V_{cb} = V_{ab} = 18$ V. By contrast, for two capacitors in parallel, as in part (b), each capacitor has the same potential difference and the *larger* charge appears on the capacitor with the *larger* capacitance. Can you show that the total charge $Q_1 + Q_2$ on the parallel combination is equal to the charge $Q = C_{eq}V$ on the equivalent capacitor?

EXAMPLE 24.6 A CAPACITOR NETWORK

SOLUTION

Find the equivalent capacitance of the five-capacitor network shown in **Fig. 24.10a**.

SOLUTION

IDENTIFY and SET UP: These capacitors are neither all in series nor all in parallel. We can, however, identify portions of the arrangement that *are* either in series or parallel. We combine these as described in Problem-Solving Strategy 24.1 to find the net equivalent capacitance, using Eq. (24.5) for series connections and Eq. (24.7) for parallel connections.

EXECUTE: The caption of Fig. 24.10 outlines our procedure. We first use Eq. (24.5) to replace the 12-μF and 6-μF series combination by its equivalent capacitance C':

$$\frac{1}{C'} = \frac{1}{12\ \mu\text{F}} + \frac{1}{6\ \mu\text{F}} \qquad C' = 4\ \mu\text{F}$$

This gives us the equivalent combination of Fig. 24.10b. Now we see three capacitors in parallel, and we use Eq. (24.7) to replace them with their equivalent capacitance C'':

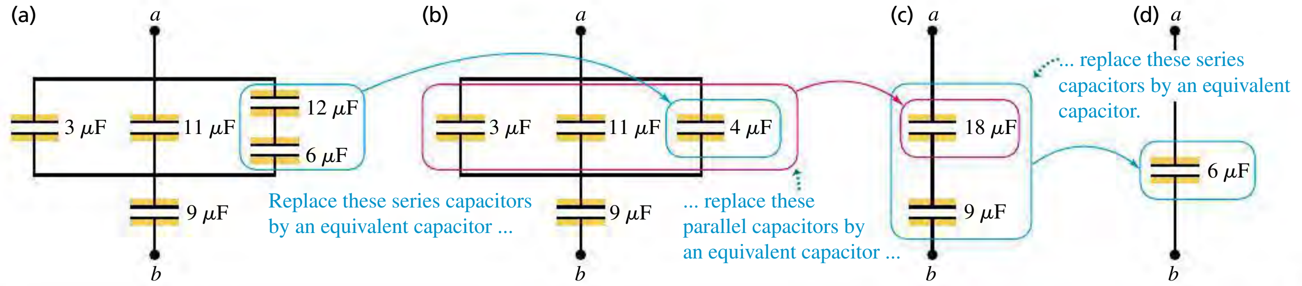
$$C'' = 3\ \mu\text{F} + 11\ \mu\text{F} + 4\ \mu\text{F} = 18\ \mu\text{F}$$

This gives us the equivalent combination of Fig. 24.10c, which has two capacitors in series. We use Eq. (24.5) to replace them with their equivalent capacitance C_{eq}, which is our target variable (Fig. 24.10d):

$$\frac{1}{C_{eq}} = \frac{1}{18\ \mu\text{F}} + \frac{1}{9\ \mu\text{F}} \qquad C_{eq} = 6\ \mu\text{F}$$

EVALUATE: If the potential difference across the entire network in Fig. 24.10a is $V_{ab} = 9.0$ V, the net charge on the network is $Q = C_{eq}V_{ab} = (6\ \mu\text{F})(9.0\ \text{V}) = 54\ \mu\text{C}$. Can you find the charge on, and the voltage across, each of the five individual capacitors?

24.10 (a) A capacitor network between points a and b. (b) The 12-μF and 6-μF capacitors in series in (a) are replaced by an equivalent 4-μF capacitor. (c) The 3-μF, 11-μF, and 4-μF capacitors in parallel in (b) are replaced by an equivalent 18-μF capacitor. (d) Finally, the 18-μF and 9-μF capacitors in series in (c) are replaced by an equivalent 6-μF capacitor.



TEST YOUR UNDERSTANDING OF SECTION 24.2 You want to connect a 4-μF capacitor and an 8-μF capacitor. (a) With which type of connection will the 4-μF capacitor have a greater potential difference across it than the 8-μF capacitor? (i) Series; (ii) parallel; (iii) either series or parallel; (iv) neither series nor parallel. (b) With which type of connection will the 4-μF capacitor have a greater charge than the 8-μF capacitor? (i) Series; (ii) parallel; (iii) either series or parallel; (iv) neither series nor parallel.

24.3 ENERGY STORAGE IN CAPACITORS AND ELECTRIC-FIELD ENERGY

Many of the most important applications of capacitors depend on their ability to store energy. The electric potential energy stored in a charged capacitor is just equal to the amount of work required to charge it—that is, to separate opposite charges and place them on different conductors. When the capacitor is discharged, this stored energy is recovered as work done by electrical forces.

We can calculate the potential energy U of a charged capacitor by calculating the work W required to charge it. Suppose that when we are done charging the capacitor, the final charge is Q and the final potential difference is V. From Eq. (24.1) these quantities are related by

$$V = \frac{Q}{C}$$

Let q and v be the charge and potential difference, respectively, at an intermediate stage during the charging process; then $v = q/C$. At this stage the work dW required to transfer an additional element of charge dq is

$$dW = v\,dq = \frac{q\,dq}{C}$$

The total work W needed to increase the capacitor charge q from zero to Q is

$$W = \int_0^W dW = \frac{1}{C}\int_0^Q q\,dq = \frac{Q^2}{2C} \quad \text{(work to charge a capacitor)} \qquad (24.8)$$

This is also the total work done by the electric field on the charge when the capacitor discharges. Then q *decreases* from an initial value Q to zero as the elements of charge dq "fall" through potential differences v that vary from V down to zero.

If we define the potential energy of an *uncharged* capacitor to be zero, then W in Eq. (24.8) is equal to the potential energy U of the charged capacitor. The final stored charge is $Q = CV$, so we can express U (which is equal to W) as

Potential energy stored in a capacitor

$$U = \frac{Q^2}{2C} = \tfrac{1}{2}CV^2 = \tfrac{1}{2}QV \qquad (24.9)$$

Magnitude of charge on each plate; Capacitance; Potential difference between plates

When Q is in coulombs, C in farads (coulombs per volt), and V in volts (joules per coulomb), U is in joules.

The last form of Eq. (24.9), $U = \frac{1}{2}QV$, shows that the total work W required to charge the capacitor is equal to the total charge Q multiplied by the *average* potential difference $\frac{1}{2}V$ during the charging process.

The expression $U = \frac{1}{2}(Q^2/C)$ in Eq. (24.9) shows that a charged capacitor is the electrical analog of a stretched spring with elastic potential energy $U = \frac{1}{2}kx^2$. The charge Q is analogous to the elongation x, and the *reciprocal* of the capacitance, $1/C$, is analogous to the force constant k. The energy supplied to a capacitor in the charging process is analogous to the work we do on a spring when we stretch it.

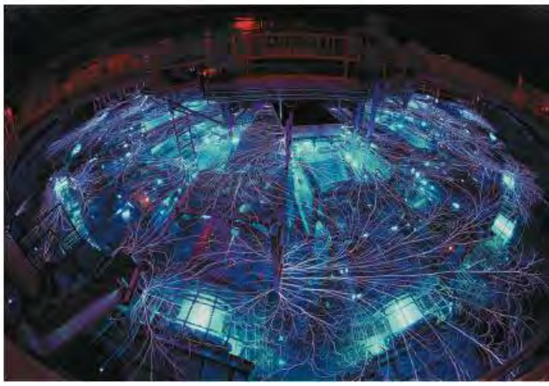
Equations (24.8) and (24.9) tell us that capacitance measures the ability of a capacitor to store both energy and charge. If a capacitor is charged by connecting it to a battery or other source that provides a fixed potential difference V, then increasing the value of C gives a greater charge $Q = CV$ and a greater amount of stored energy $U = \frac{1}{2}CV^2$. If instead the goal is to transfer a given quantity of charge Q from one conductor to another, Eq. (24.8) shows that the work W required is inversely proportional to C; the greater the capacitance, the easier it is to give a capacitor a fixed amount of charge.

Applications of Capacitors: Energy Storage

Most practical applications of capacitors take advantage of their ability to store and release energy. In an electronic flash unit used in photography, the energy stored in a capacitor (see Fig. 24.4) is released when the button is pressed to take a photograph. This provides a conducting path from one capacitor plate to the other through the flash tube. Once this path is established, the stored energy is rapidly converted into a brief but intense flash of light. An extreme example of the same principle is the Z machine at Sandia National Laboratories in New Mexico, which is used in experiments in controlled nuclear fusion (**Fig. 24.11**). A bank of charged capacitors releases more than a million joules of energy in just a few billionths of a second. For that brief space of time, the power output of the Z machine is 2.9×10^{14} W, or about 80 times the power output of all the electric power plants on earth combined!

In other applications, the energy is released more slowly. Springs in the suspension of an automobile help smooth out the ride by absorbing the energy from sudden jolts and releasing that energy gradually; in an analogous way, a capacitor in an electronic circuit can smooth out unwanted variations in voltage due to power surges. We'll discuss these circuits in detail in Chapter 26.

24.11 The Z machine uses a large number of capacitors in parallel to give a tremendous equivalent capacitance C (see Section 24.2). Hence a large amount of energy $U = \frac{1}{2}CV^2$ can be stored with even a modest potential difference V. The arcs shown here are produced when the capacitors discharge their energy into a target, which is no larger than a spool of thread. This heats the target to a temperature higher than 2×10^9 K.



Electric-Field Energy

We can charge a capacitor by moving electrons directly from one plate to another. This requires doing work against the electric field between the plates. Thus we can think of the energy as being stored *in the field* in the region between the plates. To see this, let's find the energy *per unit volume* in the space between the plates of a parallel-plate capacitor with plate area A and separation d. We call this the **energy density,** denoted by u. From Eq. (24.9) the total stored potential energy is $\frac{1}{2}CV^2$ and the volume between the plates is Ad; hence

$$u = \text{Energy density} = \frac{\frac{1}{2}CV^2}{Ad} \tag{24.10}$$

From Eq. (24.2) the capacitance C is given by $C = \epsilon_0 A/d$. The potential difference V is related to the electric-field magnitude E by $V = Ed$. If we use these expressions in Eq. (24.10), the geometric factors A and d cancel, and we find

Electric energy density in a vacuum $$u = \frac{1}{2}\epsilon_0 E^2 \tag{24.11}$$
Magnitude of electric field: E; Electric constant: ϵ_0

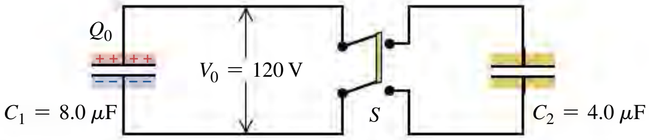
Although we have derived this relationship for a parallel-plate capacitor only, it turns out to be valid for any capacitor in vacuum and indeed *for any electric field configuration in vacuum*. This result has an interesting implication. We think of vacuum as space with no matter in it, but vacuum can nevertheless have electric fields and therefore energy. Thus "empty" space need not be truly empty after all. We will use this idea and Eq. (24.11) in Chapter 32 in connection with the energy transported by electromagnetic waves.

CAUTION **Electric-field energy is electric potential energy** It's a common misconception that electric-field energy is a new kind of energy, different from the electric potential energy described before. This is *not* the case; it is simply a different way of interpreting electric potential energy. We can regard the energy of a given system of charges as being a shared property of all the charges, or we can think of the energy as being a property of the electric field that the charges create. Either interpretation leads to the same value of the potential energy.

EXAMPLE 24.7 TRANSFERRING CHARGE AND ENERGY BETWEEN CAPACITORS

We connect a capacitor $C_1 = 8.0\ \mu\text{F}$ to a power supply, charge it to a potential difference $V_0 = 120\ \text{V}$, and disconnect the power supply (**Fig. 24.12**). Switch S is open. (a) What is the charge Q_0 on C_1? (b) What is the energy stored in C_1? (c) Capacitor $C_2 = 4.0\ \mu\text{F}$ is initially uncharged. We close switch S. After charge no longer flows, what is the potential difference across each capacitor, and what is the charge on each capacitor? (d) What is the final energy of the system?

24.12 When the switch S is closed, the charged capacitor C_1 is connected to an uncharged capacitor C_2. The center part of the switch is an insulating handle; charge can flow only between the two upper terminals and between the two lower terminals.



SOLUTION

IDENTIFY and SET UP: In parts (a) and (b) we find the charge Q_0 and stored energy U_{initial} for the single charged capacitor C_1 from Eqs. (24.1) and (24.9), respectively. After we close switch S, one wire connects the upper plates of the two capacitors and another wire connects the lower plates; the capacitors are now connected in parallel. In part (c) we use the character of the parallel connection to determine how Q_0 is shared between the two capacitors. In part (d) we again use Eq. (24.9) to find the energy stored in capacitors C_1 and C_2; the energy of the system is the sum of these values.

EXECUTE: (a) The initial charge Q_0 on C_1 is

$$Q_0 = C_1 V_0 = (8.0\ \mu\text{F})(120\ \text{V}) = 960\ \mu\text{C}$$

(b) The energy initially stored in C_1 is

$$U_{\text{initial}} = \tfrac{1}{2}Q_0 V_0 = \tfrac{1}{2}(960 \times 10^{-6}\ \text{C})(120\ \text{V}) = 0.058\ \text{J}$$

(c) When we close the switch, the positive charge Q_0 is distributed over the upper plates of both capacitors and the negative charge $-Q_0$ is distributed over the lower plates. Let Q_1 and Q_2 be the magnitudes of the final charges on the capacitors. Conservation of charge requires that $Q_1 + Q_2 = Q_0$. The potential difference V between the plates is the same for both capacitors because they are connected in parallel, so the charges are $Q_1 = C_1 V$ and $Q_2 = C_2 V$. We now have three independent equations relating the three unknowns Q_1, Q_2, and V. Solving these, we find

$$V = \frac{Q_0}{C_1 + C_2} = \frac{960\ \mu\text{C}}{8.0\ \mu\text{F} + 4.0\ \mu\text{F}} = 80\ \text{V}$$

$$Q_1 = 640\ \mu\text{C} \qquad Q_2 = 320\ \mu\text{C}$$

(d) The final energy of the system is

$$U_{\text{final}} = \tfrac{1}{2}Q_1 V + \tfrac{1}{2}Q_2 V = \tfrac{1}{2}Q_0 V$$
$$= \tfrac{1}{2}(960 \times 10^{-6}\ \text{C})(80\ \text{V}) = 0.038\ \text{J}$$

EVALUATE: The final energy is less than the initial energy; the difference was converted to energy of some other form. The conductors become a little warmer because of their resistance, and some energy is radiated as electromagnetic waves. We'll study the behavior of capacitors in more detail in Chapters 26 and Chapter 31.

EXAMPLE 24.8 ELECTRIC-FIELD ENERGY

(a) What is the magnitude of the electric field required to store 1.00 J of electric potential energy in a volume of 1.00 m^3 in vacuum? (b) If the field magnitude is 10 times larger than that, how much energy is stored per cubic meter?

SOLUTION

IDENTIFY and SET UP: We use the relationship between the electric-field magnitude E and the energy density u. In part (a) we use the given information to find u; then we use Eq. (24.11) to find the corresponding value of E. In part (b), Eq. (24.11) tells us how u varies with E.

EXECUTE: (a) The desired energy density is $u = 1.00\ \text{J/m}^3$. Then from Eq. (24.11),

$$E = \sqrt{\frac{2u}{\epsilon_0}} = \sqrt{\frac{2(1.00\ \text{J/m}^3)}{8.85 \times 10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2}}$$
$$= 4.75 \times 10^5\ \text{N/C} = 4.75 \times 10^5\ \text{V/m}$$

(b) Equation (24.11) shows that u is proportional to E^2. If E increases by a factor of 10, u increases by a factor of $10^2 = 100$, so the energy density becomes $u = 100\ \text{J/m}^3$.

EVALUATE: Dry air can sustain an electric field of about $3 \times 10^6\ \text{V/m}$ without experiencing *dielectric breakdown,* which we will discuss in Section 24.4. There we will see that field magnitudes in practical insulators can be even larger than this.

EXAMPLE 24.9 TWO WAYS TO CALCULATE ENERGY STORED IN A CAPACITOR

The spherical capacitor described in Example 24.3 (Section 24.1) has charges $+Q$ and $-Q$ on its inner and outer conductors. Find the electric potential energy stored in the capacitor (a) by using the capacitance C found in Example 24.3 and (b) by integrating the electric-field energy density u.

SOLUTION

IDENTIFY and SET UP: We can determine the energy U stored in a capacitor in two ways: in terms of the work done to put the charges on the two conductors, and in terms of the energy in the electric field between the conductors. The descriptions are equivalent, so they must give us the same result. In Example 24.3 we found the capacitance C and the field magnitude E in the space between the conductors. (The electric field is zero inside the inner sphere and is also zero outside the inner surface of the outer sphere, because a Gaussian surface with radius $r < r_a$ or $r > r_b$ encloses zero net charge. Hence the energy density is nonzero only in the space between the spheres, $r_a < r < r_b$.) In part (a) we use Eq. (24.9) to find U. In part (b) we use Eq. (24.11) to find u, which we integrate over the volume between the spheres to find U.

EXECUTE: (a) From Example 24.3, the spherical capacitor has capacitance

$$C = 4\pi\epsilon_0 \frac{r_a r_b}{r_b - r_a}$$

where r_a and r_b are the radii of the inner and outer conducting spheres, respectively. From Eq. (24.9) the energy stored in this capacitor is

$$U = \frac{Q^2}{2C} = \frac{Q^2}{8\pi\epsilon_0}\frac{r_b - r_a}{r_a r_b}$$

(b) The electric field in the region $r_a < r < r_b$ between the two conducting spheres has magnitude $E = Q/4\pi\epsilon_0 r^2$. The energy density in this region is

$$u = \tfrac{1}{2}\epsilon_0 E^2 = \tfrac{1}{2}\epsilon_0\left(\frac{Q}{4\pi\epsilon_0 r^2}\right)^2 = \frac{Q^2}{32\pi^2\epsilon_0 r^4}$$

The energy density is *not* uniform; it decreases rapidly with increasing distance from the center of the capacitor. To find the total electric-field energy, we integrate u (the energy per unit volume) over the region $r_a < r < r_b$. We divide this region into spherical shells of radius r, surface area $4\pi r^2$, thickness dr, and volume $dV = 4\pi r^2\,dr$. Then

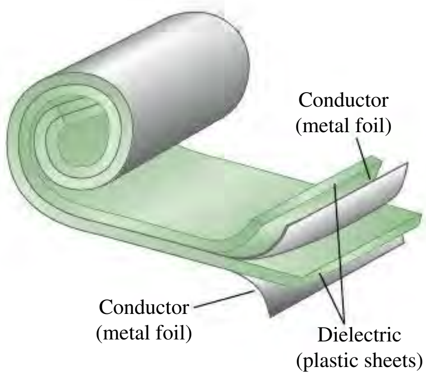
$$\begin{aligned} U &= \int u\,dV = \int_{r_a}^{r_b}\left(\frac{Q^2}{32\pi^2\epsilon_0 r^4}\right)4\pi r^2\,dr \\ &= \frac{Q^2}{8\pi\epsilon_0}\int_{r_a}^{r_b}\frac{dr}{r^2} = \frac{Q^2}{8\pi\epsilon_0}\left(-\frac{1}{r_b} + \frac{1}{r_a}\right) \\ &= \frac{Q^2}{8\pi\epsilon_0}\frac{r_b - r_a}{r_a r_b} \end{aligned}$$

EVALUATE: Electric potential energy can be associated with either the *charges,* as in part (a), or the *field,* as in part (b); the calculated amount of stored energy is the same in either case.

TEST YOUR UNDERSTANDING OF SECTION 24.3 You want to connect a 4-μF capacitor and an 8-μF capacitor. With which type of connection will the 4-μF capacitor have a greater amount of *stored energy* than the 8-μF capacitor? (i) Series; (ii) parallel; (iii) either series or parallel; (iv) neither series nor parallel. ❙

24.4 DIELECTRICS

Most capacitors have a nonconducting material, or **dielectric,** between their conducting plates. A common type of capacitor uses long strips of metal foil for the plates, separated by strips of plastic sheet such as Mylar. A sandwich of these materials is rolled up, forming a unit that can provide a capacitance of several microfarads in a compact package (**Fig. 24.13**).

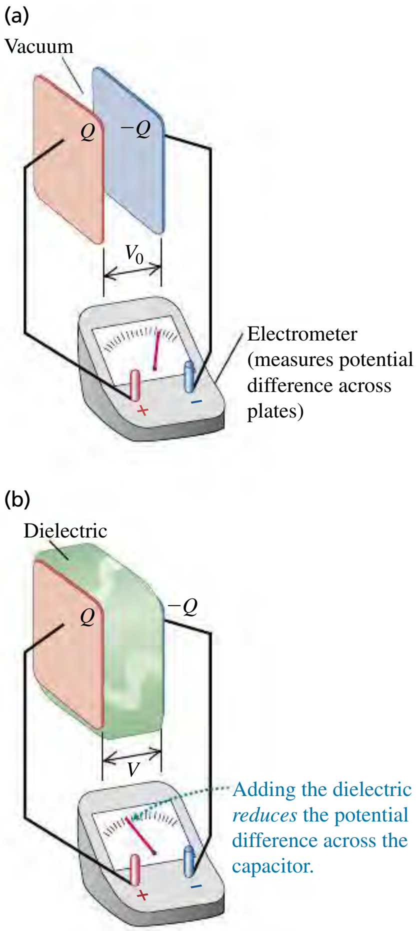


24.13 A common type of capacitor uses dielectric sheets to separate the conductors.

Placing a solid dielectric between the plates of a capacitor serves three functions. First, it solves the mechanical problem of maintaining two large metal sheets at a very small separation without actual contact.

Second, using a dielectric increases the maximum possible potential difference between the capacitor plates. Any insulating material, when subjected to a sufficiently large electric field, experiences a partial ionization that permits conduction through it (Section 23.3). This phenomenon is called **dielectric breakdown.** Many dielectric materials can tolerate stronger electric fields without breakdown

24.14 Effect of a dielectric between the plates of a parallel-plate capacitor. **(a)** With a given charge, the potential difference is V_0. **(b)** With the same charge but with a dielectric between the plates, the potential difference V is smaller than V_0.



than can air. Thus using a dielectric allows a capacitor to sustain a higher potential difference V and so store greater amounts of charge and energy.

Third, the capacitance of a capacitor of given dimensions is *greater* when there is a dielectric material between the plates than when there is vacuum. We can demonstrate this effect with the aid of a sensitive *electrometer,* a device that measures the potential difference between two conductors without letting any appreciable charge flow from one to the other. **Figure 24.14a** shows an electrometer connected across a charged capacitor, with magnitude of charge Q on each plate and potential difference V_0. When we insert an uncharged sheet of dielectric, such as glass, paraffin, or polystyrene, between the plates, experiment shows that the potential difference *decreases* to a smaller value V (Fig. 24.14b). When we remove the dielectric, the potential difference returns to its original value V_0, showing that the original charges on the plates have not changed.

The original capacitance C_0 is given by $C_0 = Q/V_0$, and the capacitance C with the dielectric present is $C = Q/V$. The charge Q is the same in both cases, and V is less than V_0, so we conclude that the capacitance C with the dielectric present is *greater* than C_0. When the space between plates is completely filled by the dielectric, the ratio of C to C_0 (equal to the ratio of V_0 to V) is called the **dielectric constant** of the material, K:

$$K = \frac{C}{C_0} \qquad \text{(definition of dielectric constant)} \tag{24.12}$$

When the charge is constant, $Q = C_0V_0 = CV$ and $C/C_0 = V_0/V$. In this case,

$$V = \frac{V_0}{K} \qquad \text{(when } Q \text{ is constant)} \tag{24.13}$$

With the dielectric present, the potential difference for a given charge Q is *reduced* by a factor K.

The dielectric constant K is a pure number. Because C is always greater than C_0, K is always greater than unity. **Table 24.1** gives some representative values of K. For vacuum, $K = 1$ by definition. For air at ordinary temperatures and pressures, K is about 1.0006; this is so nearly equal to 1 that for most purposes an air capacitor is equivalent to one in vacuum. Note that while water has a very large value of K, it is usually not a very practical dielectric for use in capacitors. The reason is that while pure water is a very poor conductor, it is also an excellent ionic solvent. Any ions that are dissolved in the water will cause charge to flow between the capacitor plates, so the capacitor discharges.

CAUTION **Dielectric constant vs. electric constant** Don't confuse the *dielectric* constant K with the *electric* constant ϵ_0. The value of K is a pure number with no units and is different for different materials (see Table 24.1). By contrast, ϵ_0 is a universal constant with units $C^2/N \cdot m^2$ or F/m.

TABLE 24.1 **Values of Dielectric Constant K at 20°C**

Material	K	Material	K
Vacuum	1	Polyvinyl chloride	3.18
Air (1 atm)	1.00059	Plexiglas®	3.40
Air (100 atm)	1.0548	Glass	5–10
Teflon	2.1	Neoprene	6.70
Polyethylene	2.25	Germanium	16
Benzene	2.28	Glycerin	42.5
Mica	3–6	Water	80.4
Mylar	3.1	Strontium titanate	310

No real dielectric is a perfect insulator. Hence there is always some *leakage current* between the charged plates of a capacitor with a dielectric. We tacitly ignored this effect in Section 24.2 when we derived expressions for the equivalent capacitances of capacitors in series, Eq. (24.5), and in parallel, Eq. (24.7). But if a leakage current flows for a long enough time to substantially change the charges from the values we used to derive Eqs. (24.5) and (24.7), those equations may no longer be accurate.

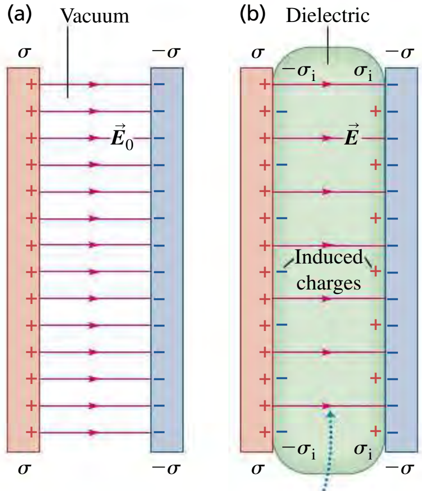
Induced Charge and Polarization

When a dielectric material is inserted between the plates while the charge is kept constant, the potential difference between the plates decreases by a factor K. Therefore the electric field between the plates must decrease by the same factor. If E_0 is the vacuum value and E is the value with the dielectric, then

$$E = \frac{E_0}{K} \qquad \text{(when } Q \text{ is constant)} \tag{24.14}$$

Since the electric-field magnitude is smaller when the dielectric is present, the surface charge density (which causes the field) must be smaller as well. The surface charge on the conducting plates does not change, but an *induced* charge of the opposite sign appears on each surface of the dielectric (**Fig. 24.15**). The dielectric was originally electrically neutral and is still neutral; the induced surface charges arise as a result of *redistribution* of positive and negative charge within the dielectric material, a phenomenon called **polarization.** We first encountered polarization in Section 21.2, and we suggest that you reread the discussion of Fig. 21.8. We will assume that the induced surface charge is *directly proportional* to the electric-field magnitude E in the material; this is indeed the case for many common dielectrics. (This direct proportionality is analogous to Hooke's law for a spring.) In that case, K is a constant for any particular material. When the electric field is very strong or if the dielectric is made of certain crystalline materials, the relationship between induced charge and the electric field can be more complex; we won't consider such cases here.

24.15 Electric field lines with (a) vacuum between the plates and (b) dielectric between the plates.



We can derive a relationship between this induced surface charge and the charge on the plates. Let σ_i be the magnitude of the charge per unit area induced on the surfaces of the dielectric (the induced surface charge density). The magnitude of the surface charge density on the capacitor plates is σ, as usual. Then the *net* surface charge on each side of the capacitor has magnitude $(\sigma - \sigma_i)$; see Fig. 24.15b. As we found in Example 21.12 (Section 21.5) and in Example 22.8 (Section 22.4), the field between the plates is related to the net surface charge density by $E = \sigma_{net}/\epsilon_0$. Without and with the dielectric, respectively,

$$E_0 = \frac{\sigma}{\epsilon_0} \qquad E = \frac{\sigma - \sigma_i}{\epsilon_0} \tag{24.15}$$

Using these expressions in Eq. (24.14) and rearranging the result, we find

$$\sigma_i = \sigma\left(1 - \frac{1}{K}\right) \qquad \text{(induced surface charge density)} \tag{24.16}$$

This equation shows that when K is very large, σ_i is nearly as large as σ. In this case, σ_i nearly cancels σ, and the field and potential difference are much smaller than their values in vacuum.

The product $K\epsilon_0$ is called the **permittivity** of the dielectric, denoted by ϵ:

$$\epsilon = K\epsilon_0 \qquad \text{(definition of permittivity)} \tag{24.17}$$

In terms of ϵ we can express the electric field within the dielectric as

$$E = \frac{\sigma}{\epsilon} \tag{24.18}$$

DATA *SPEAKS*

Capacitors and Capacitance

When students were given a problem involving capacitors and capacitance, more than 25% gave an incorrect response. Common errors:

- Forgetting that the capacitance C of a capacitor depends on only the capacitor geometry (the size, shape, and position of its conductors) and the presence or absence of a dielectric. C does not depend on the amount of charge Q placed on the conductors.
- Not understanding what happens if capacitance changes (for example, by inserting or removing a dielectric). If the capacitor is isolated, Q remains constant but the potential difference V_{ab} changes if C changes. If V_{ab} is held constant, Q changes if C changes.

Application Capacitors in the Toolbox Several practical devices rely on the way a capacitor responds to a change in dielectric constant. One example is an electric stud finder, used to locate metal fasteners hidden behind a wall's surface. It consists of a metal plate with associated circuitry. The plate acts as one half of a capacitor; the wall acts as the other half. If the stud finder moves over a metal fastener, the effective dielectric constant for the capacitor changes, which changes the capacitance and triggers a signal.

Then

Capacitance of a parallel-plate capacitor, dielectric between plates

Dielectric constant; Area of each plate; Permittivity $= K\epsilon_0$

$$C = KC_0 = K\epsilon_0\frac{A}{d} = \epsilon\frac{A}{d} \tag{24.19}$$

Capacitance without dielectric; Electric constant; Distance between plates

We can repeat the derivation of Eq. (24.11) for the energy density u in an electric field for the case in which a dielectric is present. The result is

Electric energy density in a dielectric

Dielectric constant; Permittivity $= K\epsilon_0$

$$u = \tfrac{1}{2}K\epsilon_0 E^2 = \tfrac{1}{2}\epsilon E^2 \tag{24.20}$$

Electric constant; Magnitude of electric field

In empty space, where $K = 1$, $\epsilon = \epsilon_0$ and Eqs. (24.19) and (24.20) reduce to Eqs. (24.2) and (24.11), respectively, for a parallel-plate capacitor in vacuum. For this reason, ϵ_0 is sometimes called the "permittivity of free space" or the "permittivity of vacuum." Because K is a pure number, ϵ and ϵ_0 have the same units, $C^2/N \cdot m^2$ or F/m.

Equation (24.19) shows that extremely high capacitances can be obtained with plates that have a large surface area A and are separated by a small distance d by a dielectric with a large value of K. In an *electrolytic double-layer capacitor,* tiny carbon granules adhere to each plate: A is the combined surface area of the granules, which can be tremendous. The plates with granules attached are separated by a very thin dielectric sheet. Such a capacitor can have a capacitance of 5000 F yet fit in the palm of your hand (compare Example 24.1 in Section 24.1).

PROBLEM-SOLVING STRATEGY 24.2 DIELECTRICS

IDENTIFY *the relevant concepts:* The relationships in this section are useful whenever there is an electric field in a dielectric, such as a dielectric between charged capacitor plates. Typically you must relate the potential difference V_{ab} between the plates, the electric field magnitude E in the capacitor, the charge density σ on the capacitor plates, and the induced charge density σ_i on the surfaces of the capacitor.

SET UP *the problem* using the following steps:

1. Make a drawing of the situation.
2. Identify the target variables, and choose which equations from this section will help you solve for those variables.

EXECUTE *the solution* as follows:

1. In problems such as the next example, it is easy to get lost in a blizzard of formulas. Ask yourself at each step what kind of quantity each symbol represents. For example, distinguish clearly between charges and charge densities, and between electric fields and electric potential differences.
2. Check for consistency of units. Distances must be in meters. A microfarad is 10^{-6} farad, and so on. Don't confuse the numerical value of ϵ_0 with the value of $1/4\pi\epsilon_0$. Electric-field magnitude can be expressed in both N/C and V/m. The units of ϵ_0 are $C^2/N \cdot m^2$ or F/m.

EVALUATE *your answer:* With a dielectric present, (a) the capacitance is greater than without a dielectric; (b) for a given charge on the capacitor, the electric field and potential difference are less than without a dielectric; and (c) the magnitude of the induced surface charge density σ_i on the dielectric is less than that of the charge density σ on the capacitor plates.

EXAMPLE 24.10 A CAPACITOR WITH AND WITHOUT A DIELECTRIC

SOLUTION

Suppose the parallel plates in Fig. 24.15 each have an area of 2000 cm^2 (2.00×10^{-1} m^2) and are 1.00 cm (1.00×10^{-2} m) apart. We connect the capacitor to a power supply, charge it to a potential difference $V_0 = 3.00$ kV, and disconnect the power supply. We then insert a sheet of insulating plastic material between the plates, completely filling the space between them. We find that the potential difference decreases to 1.00 kV while the charge on each capacitor plate remains constant. Find (a) the original capacitance C_0; (b) the magnitude of charge Q on each plate; (c) the capacitance C after the dielectric is inserted; (d) the dielectric constant K of the dielectric; (e) the permittivity ϵ of the dielectric; (f) the magnitude of the induced charge Q_i on each face of the dielectric; (g) the original electric field E_0 between the plates; and (h) the electric field E after the dielectric is inserted.

SOLUTION

IDENTIFY and SET UP: This problem uses most of the relationships we have discussed for capacitors and dielectrics. (Energy relationships are treated in Example 24.11.) Most of the target variables can be obtained in several ways. The methods used below are a sample; we encourage you to think of others and compare your results.

EXECUTE: (a) With vacuum between the plates, we use Eq. (24.19) with $K = 1$:

$$C_0 = \epsilon_0 \frac{A}{d} = (8.85 \times 10^{-12}\ \text{F/m}) \frac{2.00 \times 10^{-1}\ \text{m}^2}{1.00 \times 10^{-2}\ \text{m}}$$
$$= 1.77 \times 10^{-10}\ \text{F} = 177\ \text{pF}$$

(b) From the definition of capacitance, Eq. (24.1),

$$Q = C_0 V_0 = (1.77 \times 10^{-10}\ \text{F})(3.00 \times 10^3\ \text{V})$$
$$= 5.31 \times 10^{-7}\ \text{C} = 0.531\ \mu\text{C}$$

(c) When the dielectric is inserted, Q is unchanged but the potential difference decreases to $V = 1.00\ \text{kV}$. Hence from Eq. (24.1), the new capacitance is

$$C = \frac{Q}{V} = \frac{5.31 \times 10^{-7}\ \text{C}}{1.00 \times 10^3\ \text{V}} = 5.31 \times 10^{-10}\ \text{F}$$
$$= 531\ \text{pF}$$

(d) From Eq. (24.12), the dielectric constant is

$$K = \frac{C}{C_0} = \frac{5.31 \times 10^{-10}\ \text{F}}{1.77 \times 10^{-10}\ \text{F}} = \frac{531\ \text{pF}}{177\ \text{pF}}$$
$$= 3.00$$

Alternatively, from Eq. (24.13),

$$K = \frac{V_0}{V} = \frac{3000\ \text{V}}{1000\ \text{V}} = 3.00$$

(e) With K from part (d) in Eq. (24.17), the permittivity is

$$\epsilon = K\epsilon_0 = (3.00)(8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2)$$
$$= 2.66 \times 10^{-11}\ \text{C}^2/\text{N} \cdot \text{m}^2$$

(f) Multiplying both sides of Eq. (24.16) by the plate area A gives the induced charge $Q_\text{i} = \sigma_\text{i} A$ in terms of the charge $Q = \sigma A$ on each plate:

$$Q_\text{i} = Q\left(1 - \frac{1}{K}\right) = (5.31 \times 10^{-7}\ \text{C})\left(1 - \frac{1}{3.00}\right)$$
$$= 3.54 \times 10^{-7}\ \text{C}$$

(g) Since the electric field between the plates is uniform, its magnitude is the potential difference divided by the plate separation:

$$E_0 = \frac{V_0}{d} = \frac{3000\ \text{V}}{1.00 \times 10^{-2}\ \text{m}} = 3.00 \times 10^5\ \text{V/m}$$

(h) After the dielectric is inserted,

$$E = \frac{V}{d} = \frac{1000\ \text{V}}{1.00 \times 10^{-2}\ \text{m}} = 1.00 \times 10^5\ \text{V/m}$$

or, from Eq. (24.18),

$$E = \frac{\sigma}{\epsilon} = \frac{Q}{\epsilon A} = \frac{5.31 \times 10^{-7}\ \text{C}}{(2.66 \times 10^{-11}\ \text{C}^2/\text{N} \cdot \text{m}^2)(2.00 \times 10^{-1}\ \text{m}^2)}$$
$$= 1.00 \times 10^5\ \text{V/m}$$

or, from Eq. (24.15),

$$E = \frac{\sigma - \sigma_\text{i}}{\epsilon_0} = \frac{Q - Q_\text{i}}{\epsilon_0 A}$$
$$= \frac{(5.31 - 3.54) \times 10^{-7}\ \text{C}}{(8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2)(2.00 \times 10^{-1}\ \text{m}^2)}$$
$$= 1.00 \times 10^5\ \text{V/m}$$

or, from Eq. (24.14),

$$E = \frac{E_0}{K} = \frac{3.00 \times 10^5\ \text{V/m}}{3.00} = 1.00 \times 10^5\ \text{V/m}$$

EVALUATE: Inserting the dielectric increased the capacitance by a factor of $K = 3.00$ and reduced the electric field between the plates by a factor of $1/K = 1/3.00$. It did so by developing induced charges on the faces of the dielectric of magnitude $Q(1 - 1/K) = Q(1 - 1/3.00) = 0.667Q$.

EXAMPLE 24.11 ENERGY STORAGE WITH AND WITHOUT A DIELECTRIC

SOLUTION

Find the energy stored in the electric field of the capacitor in Example 24.10 and the energy density, both before and after the dielectric sheet is inserted.

SOLUTION

IDENTIFY and SET UP: We consider the ideas of energy stored in a capacitor and of electric-field energy density. We use Eq. (24.9) to find the stored energy and Eq. (24.20) to find the energy density.

EXECUTE: From Eq. (24.9), the stored energies U_0 and U without and with the dielectric in place are

$$U_0 = \tfrac{1}{2} C_0 V_0^2 = \tfrac{1}{2}(1.77 \times 10^{-10}\ \text{F})(3000\ \text{V})^2 = 7.97 \times 10^{-4}\ \text{J}$$
$$U = \tfrac{1}{2} C V^2 = \tfrac{1}{2}(5.31 \times 10^{-10}\ \text{F})(1000\ \text{V})^2 = 2.66 \times 10^{-4}\ \text{J}$$

The final energy is one-third of the original energy.

Equation (24.20) gives the energy densities without and with the dielectric:

$$u_0 = \tfrac{1}{2}\epsilon_0 E_0^2 = \tfrac{1}{2}(8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2)(3.00 \times 10^5\ \text{N/C})^2$$
$$= 0.398\ \text{J/m}^3$$
$$u = \tfrac{1}{2}\epsilon E^2 = \tfrac{1}{2}(2.66 \times 10^{-11}\ \text{C}^2/\text{N} \cdot \text{m}^2)(1.00 \times 10^5\ \text{N/C})^2$$
$$= 0.133\ \text{J/m}^3$$

The energy density with the dielectric is one-third of the original energy density.

Continued

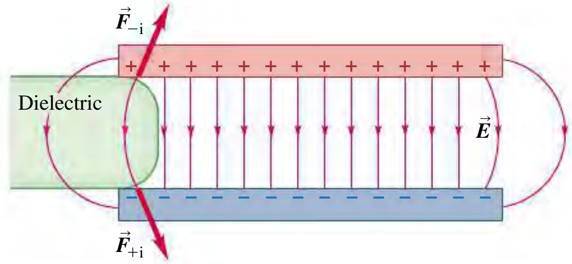
EVALUATE: We can check our answer for u_0 by noting that the volume between the plates is $V = (0.200\ \text{m}^2)(0.0100\ \text{m}) = 0.00200\ \text{m}^3$. Since the electric field between the plates is uniform, u_0 is uniform as well and the energy density is just the stored energy divided by the volume:

$$u_0 = \frac{U_0}{V} = \frac{7.97 \times 10^{-4}\ \text{J}}{0.00200\ \text{m}^3} = 0.398\ \text{J/m}^3$$

This agrees with our earlier answer. You can use the same approach to check our result for u.

In general, when a dielectric is inserted into a capacitor while the charge on each plate remains the same, the permittivity ϵ increases by a factor of K (the dielectric constant), and the electric field E and the energy density $u = \frac{1}{2}\epsilon E^2$ decrease by a factor of $1/K$. Where does the energy go? The answer lies in the fringing field at the edges of a real parallel-plate capacitor. As **Fig. 24.16** shows, that field tends to pull the dielectric into the space between the plates, doing work on it as it does so. We could attach a spring to the left end of the dielectric in Fig. 24.16 and use this force to stretch the spring. Because work is done by the field, the field energy density decreases.

24.16 The fringing field at the edges of the capacitor exerts forces $\vec{F}_{-i}$ and $\vec{F}_{+i}$ on the negative and positive induced surface charges of a dielectric, pulling the dielectric into the capacitor.



Dielectric Breakdown

We mentioned earlier that when a dielectric is subjected to a sufficiently strong electric field, *dielectric breakdown* takes place and the dielectric becomes a conductor. This occurs when the electric field is so strong that electrons are ripped loose from their molecules and crash into other molecules, liberating even more electrons. This avalanche of moving charge forms a spark or arc discharge. Lightning is a dramatic example of dielectric breakdown in air.

Because of dielectric breakdown, capacitors always have maximum voltage ratings. When a capacitor is subjected to excessive voltage, an arc may form through a layer of dielectric, burning or melting a hole in it. This arc creates a conducting path (a short circuit) between the conductors. If a conducting path remains after the arc is extinguished, the device is rendered permanently useless as a capacitor.

The maximum electric-field magnitude that a material can withstand without the occurrence of breakdown is called its **dielectric strength.** This quantity is affected significantly by temperature, trace impurities, small irregularities in the metal electrodes, and other factors that are difficult to control. For this reason we can give only approximate figures for dielectric strengths. The dielectric strength of dry air is about $3 \times 10^6\ \text{V/m}$. **Table 24.2** lists the dielectric strengths of a few common insulating materials. All of the values are substantially greater than the value for air. For example, a layer of polycarbonate 0.01 mm thick (about the smallest practical thickness) has 10 times the dielectric strength of air and can withstand a maximum voltage of about $(3 \times 10^7\ \text{V/m})(1 \times 10^{-5}\ \text{m}) = 300\ \text{V}$.

BIO Application Dielectric Cell Membrane The membrane of a living cell behaves like a dielectric between the plates of a capacitor. The membrane is made of two sheets of lipid molecules, with their water-insoluble ends in the middle and their water-soluble ends (shown in red) on the outer surfaces. Conductive fluids on either side of the membrane (water with negative ions inside the cell, water with positive ions outside) act as charged capacitor plates, and the nonconducting membrane acts as a dielectric with K of about 10. The potential difference V across the membrane is about 0.07 V and the membrane thickness d is about 7×10^{-9} m, so the electric field $E = V/d$ in the membrane is about 10^7 V/m—close to the dielectric strength of the membrane. If the membrane were made of air, V and E would be larger by a factor of $K \approx 10$ and dielectric breakdown would occur.

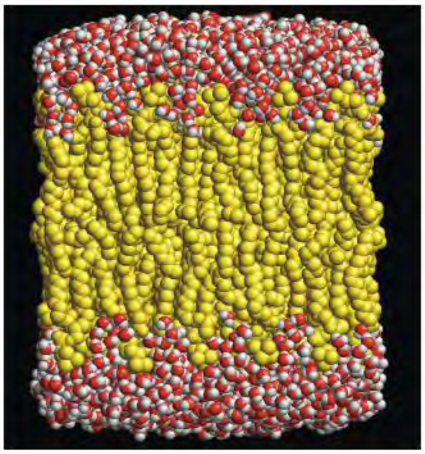


TABLE 24.2 **Dielectric Constant and Dielectric Strength of Some Insulating Materials**

Material	Dielectric Constant, K	Dielectric Strength, E_m (V/m)
Polycarbonate	2.8	3×10^7
Polyester	3.3	6×10^7
Polypropylene	2.2	7×10^7
Polystyrene	2.6	2×10^7
Pyrex glass	4.7	1×10^7

TEST YOUR UNDERSTANDING OF SECTION 24.4 The space between the plates of an isolated parallel-plate capacitor is filled by a slab of dielectric with dielectric constant K. The two plates of the capacitor have charges Q and $-Q$. You pull out the dielectric slab. If the charges do not change, how does the energy in the capacitor change when you remove the slab? (i) It increases; (ii) it decreases; (iii) it remains the same. ❙

24.5 MOLECULAR MODEL OF INDUCED CHARGE

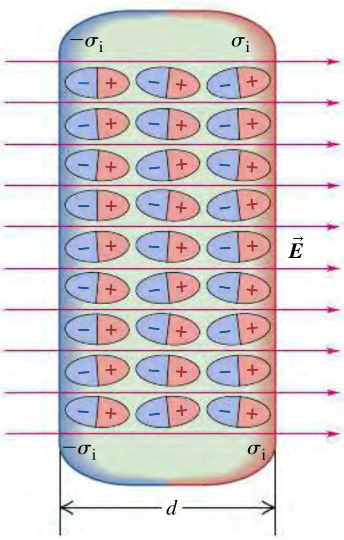
In Section 24.4 we discussed induced surface charges on a dielectric in an electric field. Now let's look at how these surface charges can arise. If the material were a *conductor,* the answer would be simple. Conductors contain charge that is free to move, and when an electric field is present, some of the charge redistributes itself on the surface so that there is no electric field inside the conductor. But an ideal dielectric has *no* charges that are free to move, so how can a surface charge occur?

To understand this, we have to look again at rearrangement of charge at the *molecular* level. Some molecules, such as H_2O and N_2O, have equal amounts of positive and negative charges but a lopsided distribution, with excess positive charge concentrated on one side of the molecule and negative charge on the other. As we described in Section 21.7, such an arrangement is called an *electric dipole,* and the molecule is called a *polar molecule*. When no electric field is present in a gas or liquid with polar molecules, the molecules are oriented randomly (**Fig. 24.17a**). In an electric field, however, they tend to orient themselves as in Fig. 24.17b, as a result of the electric-field torques described in Section 21.7. Because of thermal agitation, the alignment of the molecules with $\vec{E}$ is not perfect.

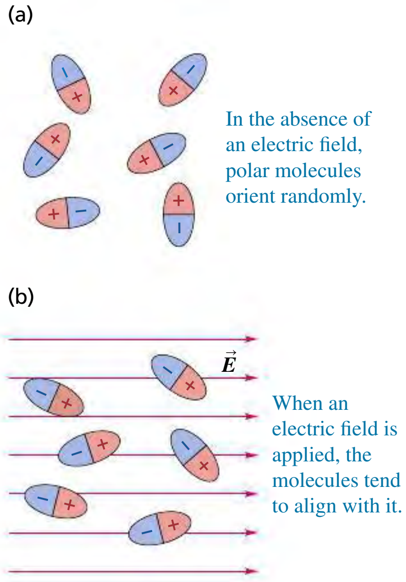
Even a molecule that is *not* ordinarily polar *becomes* a dipole when it is placed in an electric field because the field pushes the positive charges in the molecules in the direction of the field and pushes the negative charges in the opposite direction. This causes a redistribution of charge within the molecule (**Fig. 24.18**). Such dipoles are called *induced* dipoles.

With either polar or nonpolar molecules, the redistribution of charge caused by the field leads to the formation of a layer of charge on each surface of the dielectric material (**Fig. 24.19**). These layers are the surface charges described in Section 24.4; their surface charge density is denoted by σ_i. The charges are *not* free to move indefinitely, as they would be in a conductor, because each charge is bound to a molecule. They are in fact called **bound charges** to distinguish them

24.19 Polarization of a dielectric in an electric field $\vec{E}$ gives rise to thin layers of bound charges on the surfaces, creating surface charge densities σ_i and $-\sigma_i$. The sizes of the molecules are greatly exaggerated for clarity.



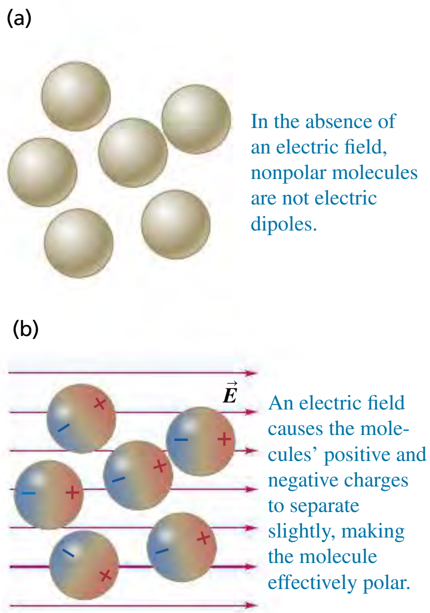
24.17 Polar molecules (a) without and (b) with an applied electric field $\vec{E}$.



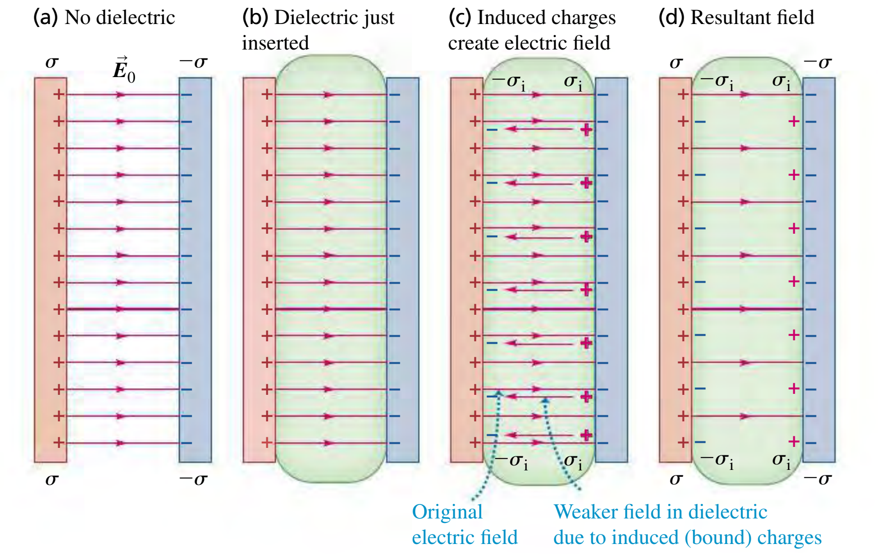
MP

PhET: Molecular Motors
PhET: Optical Tweezers and Applications
PhET: Stretching DNA

24.18 Nonpolar molecules (a) without and (b) with an applied electric field $\vec{E}$.



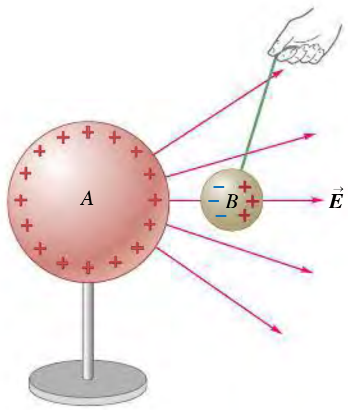
24.20 (a) Electric field of magnitude E_0 between two charged plates. (b) Introduction of a dielectric of dielectric constant K. (c) The induced surface charges and their field. (d) Resultant field of magnitude E_0/K.



from the **free charges** that are added to and removed from the conducting capacitor plates. In the interior of the material the net charge per unit volume remains zero. As we have seen, this redistribution of charge is called *polarization,* and we say that the material is *polarized*.

The four parts of **Fig. 24.20** show the behavior of a slab of dielectric when it is inserted in the field between a pair of oppositely charged capacitor plates. Figure 24.20a shows the original field. Figure 24.20b is the situation after the dielectric has been inserted but before any rearrangement of charges has occurred. Figure 24.20c shows by thinner arrows the additional field set up in the dielectric by its induced surface charges. This field is *opposite* to the original field, but it is not great enough to cancel the original field completely because the charges in the dielectric are not free to move indefinitely. The resultant field in the dielectric, shown in Fig. 24.20d, is therefore decreased in magnitude. In the field-line representation, some of the field lines leaving the positive plate go through the dielectric, while others terminate on the induced charges on the faces of the dielectric.

24.21 A neutral sphere B in the radial electric field of a positively charged sphere A is attracted to the charge because of polarization.



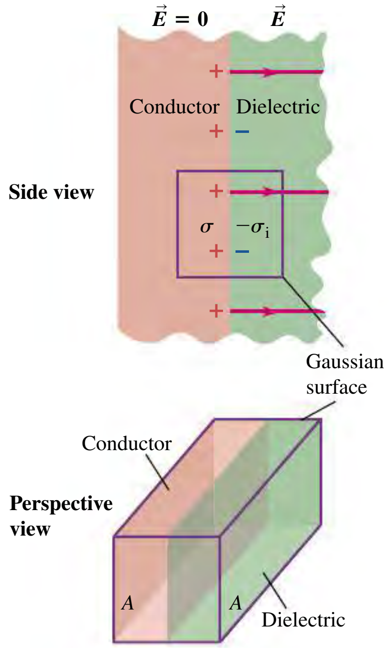
As we discussed in Section 21.2, polarization is also the reason a charged body, such as an electrified plastic rod, can exert a force on an *uncharged* body such as a bit of paper or a pith ball. **Figure 24.21** shows an uncharged dielectric sphere B in the radial field of a positively charged body A. The induced positive charges on B experience a force toward the right, while the force on the induced negative charges is toward the left. The negative charges are closer to A, and thus are in a stronger field, than are the positive charges. The force toward the left is stronger than that toward the right, and B is attracted toward A, even though its net charge is zero. The attraction occurs whether the sign of A's charge is positive or negative (see Fig. 21.8). Furthermore, the effect is not limited to dielectrics; an uncharged conducting body would be attracted in the same way.

TEST YOUR UNDERSTANDING OF SECTION 24.5 A parallel-plate capacitor has charges Q and $-Q$ on its two plates. A dielectric slab with $K = 3$ is then inserted into the space between the plates as shown in Fig. 24.20. Rank the following electric-field magnitudes in order from largest to smallest. (i) The field before the slab is inserted; (ii) the resultant field after the slab is inserted; (iii) the field due to the bound charges. ❙

24.6 GAUSS'S LAW IN DIELECTRICS

We can extend the analysis of Section 24.4 to reformulate Gauss's law in a form that is particularly useful for dielectrics. **Figure 24.22** is a close-up view of the left capacitor plate and left surface of the dielectric in Fig. 24.15b. Let's apply Gauss's law to the rectangular box shown in cross section by the purple line; the surface area of the left and right sides is A. The left side is embedded in the conductor that forms the left capacitor plate, and so the electric field everywhere on that surface is zero. The right side is embedded in the dielectric, where the electric field has magnitude E, and $E_{\perp} = 0$ everywhere on the other four sides. The total charge enclosed, including both the charge on the capacitor plate and the induced charge on the dielectric surface, is $Q_{\text{encl}} = (\sigma - \sigma_{\text{i}})A$, so Gauss's law gives

$$EA = \frac{(\sigma - \sigma_{\text{i}})A}{\epsilon_0} \qquad (24.21)$$



24.22 Gauss's law with a dielectric. This figure shows a close-up of the left-hand capacitor plate in Fig. 24.15b. The Gaussian surface is a rectangular box that lies half in the conductor and half in the dielectric.

This equation is not very illuminating as it stands because it relates two unknown quantities: E inside the dielectric and the induced surface charge density σ_{i}. But now we can use Eq. (24.16), developed for this same situation, to simplify this equation by eliminating σ_{i}. Equation (24.16) is

$$\sigma_{\text{i}} = \sigma\left(1 - \frac{1}{K}\right)$$

or

$$\sigma - \sigma_i = \frac{\sigma}{K}$$

Combining this with Eq. (24.21), we get

$$EA = \frac{\sigma A}{K\epsilon_0}$$

or

$$KEA = \frac{\sigma A}{\epsilon_0} \qquad (24.22)$$

Equation (24.22) says that the flux of $K\vec{E}$, not $\vec{E}$, through the Gaussian surface in Fig. 24.22 is equal to the enclosed *free* charge σA divided by ϵ_0. It turns out that for *any* Gaussian surface, whenever the induced charge is proportional to the electric field in the material, we can rewrite Gauss's law as

Gauss's law in a dielectric:

Dielectric constant

$$\underbrace{\oint K\vec{E}\cdot d\vec{A}}_{\text{Surface integral of } K\vec{E} \text{ over a closed surface}} = \frac{Q_{\text{encl-free}}}{\epsilon_0} \qquad (24.23)$$

Total free charge enclosed by surface

Electric constant

where $Q_{\text{encl-free}}$ is the total *free* charge (not bound charge) enclosed by the Gaussian surface. The significance of these results is that the right sides contain only the *free* charge on the conductor, not the bound (induced) charge. In fact, although we have not proved it, Eq. (24.23) remains valid even when different parts of the Gaussian surface are embedded in dielectrics having different values of K, provided that the value of K in each dielectric is independent of the electric field (usually the case for electric fields that are not too strong) and that we use the appropriate value of K for each point on the Gaussian surface.

EXAMPLE 24.12 A SPHERICAL CAPACITOR WITH DIELECTRIC

SOLUTION

Use Gauss's law to find the capacitance of the spherical capacitor of Example 24.3 (Section 24.1) if the volume between the shells is filled with an insulating oil with dielectric constant K.

SOLUTION

IDENTIFY and SET UP: The spherical symmetry of the problem is not changed by the presence of the dielectric, so as in Example 24.3, we use a concentric spherical Gaussian surface of radius r between the shells. Since a dielectric is present, we use Gauss's law in the form of Eq. (24.23).

EXECUTE: From Eq. (24.23),

$$\oint K\vec{E}\cdot d\vec{A} = \oint KE\,dA = KE\oint dA = (KE)(4\pi r^2) = \frac{Q}{\epsilon_0}$$

$$E = \frac{Q}{4\pi K\epsilon_0 r^2} = \frac{Q}{4\pi\epsilon r^2}$$

where $\epsilon = K\epsilon_0$. Compared to the case in which there is vacuum between the shells, the electric field is reduced by a factor of $1/K$. The potential difference V_{ab} between the shells is reduced by the same factor, and so the capacitance $C = Q/V_{ab}$ is *increased* by a factor of K, just as for a parallel-plate capacitor when a dielectric is inserted. Using the result of Example 24.3, we find that the capacitance with the dielectric is

$$C = \frac{4\pi K\epsilon_0 r_a r_b}{r_b - r_a} = \frac{4\pi\epsilon r_a r_b}{r_b - r_a}$$

EVALUATE: If the dielectric fills the volume between the two conductors, the capacitance is just K times the value with no dielectric. The result is more complicated if the dielectric only partially fills this volume.

TEST YOUR UNDERSTANDING OF SECTION 24.6 A single point charge q is embedded in a very large block of dielectric of dielectric constant K. At a point inside the dielectric a distance r from the point charge, what is the magnitude of the electric field? (i) $q/4\pi\epsilon_0 r^2$; (ii) $Kq/4\pi\epsilon_0 r^2$; (iii) $q/4\pi K\epsilon_0 r^2$; (iv) none of these. ▌

CHAPTER 24 SUMMARY

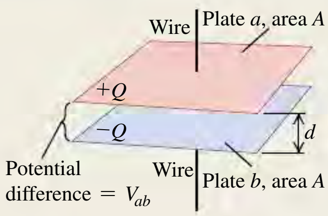
SOLUTIONS TO ALL EXAMPLES

Capacitors and capacitance: A capacitor is any pair of conductors separated by an insulating material. When the capacitor is charged, there are charges of equal magnitude Q and opposite sign on the two conductors, and the potential V_{ab} of the positively charged conductor with respect to the negatively charged conductor is proportional to Q. The capacitance C is defined as the ratio of Q to V_{ab}. The SI unit of capacitance is the farad (F): 1 F = 1 C/V.

A parallel-plate capacitor consists of two parallel conducting plates, each with area A, separated by a distance d. If they are separated by vacuum, the capacitance depends on only A and d. For other geometries, the capacitance can be found by using the definition $C = Q/V_{ab}$. (See Examples 24.1–24.4.)

$$C = \frac{Q}{V_{ab}} \qquad (24.1)$$

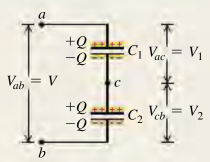
$$C = \frac{Q}{V_{ab}} = \epsilon_0\frac{A}{d} \qquad (24.2)$$

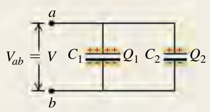


Capacitors in series and parallel: When capacitors with capacitances $C_1, C_2, C_3, \ldots$ are connected in series, the reciprocal of the equivalent capacitance C_{eq} equals the sum of the reciprocals of the individual capacitances. When capacitors are connected in parallel, the equivalent capacitance C_{eq} equals the sum of the individual capacitances. (See Examples 24.5 and 24.6.)

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \qquad (24.5)$$
(capacitors in series)

$$C_{eq} = C_1 + C_2 + C_3 + \cdots \qquad (24.7)$$
(capacitors in parallel)

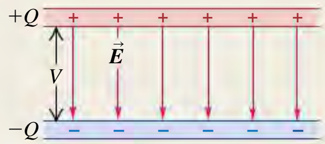




Energy in a capacitor: The energy U required to charge a capacitor C to a potential difference V and a charge Q is equal to the energy stored in the capacitor. This energy can be thought of as residing in the electric field between the conductors; the energy density u (energy per unit volume) is proportional to the square of the electric-field magnitude. (See Examples 24.7–24.9.)

$$U = \frac{Q^2}{2C} = \tfrac{1}{2}CV^2 = \tfrac{1}{2}QV \qquad (24.9)$$

$$u = \tfrac{1}{2}\epsilon_0 E^2 \qquad (24.11)$$



Dielectrics: When the space between the conductors is filled with a dielectric material, the capacitance increases by a factor K, the dielectric constant of the material. The quantity $\epsilon = K\epsilon_0$ is the permittivity of the dielectric. For a fixed amount of charge on the capacitor plates, induced charges on the surface of the dielectric decrease the electric field and potential difference between the plates by the same factor K. The surface charge results from polarization, a microscopic rearrangement of charge in the dielectric. (See Example 24.10.)

Under sufficiently strong electric fields, dielectrics become conductors, a situation called dielectric breakdown. The maximum field that a material can withstand without breakdown is called its dielectric strength.

In a dielectric, the expression for the energy density is the same as in vacuum but with ϵ_0 replaced by $\epsilon = K\epsilon_0$. (See Example 24.11.)

Gauss's law in a dielectric has almost the same form as in vacuum, with two key differences: $\vec{E}$ is replaced by $K\vec{E}$ and Q_{encl} is replaced by $Q_{\text{encl-free}}$, which includes only the free charge (not bound charge) enclosed by the Gaussian surface. (See Example 24.12.)

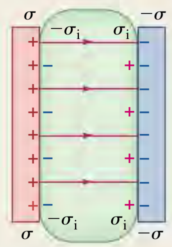
$$C = KC_0 = K\epsilon_0\frac{A}{d} = \epsilon\frac{A}{d} \qquad (24.19)$$

(parallel-plate capacitor filled with dielectric)

$$u = \tfrac{1}{2}K\epsilon_0 E^2 = \tfrac{1}{2}\epsilon E^2 \qquad (24.20)$$

$$\oint K\vec{E}\cdot d\vec{A} = \frac{Q_{\text{encl-free}}}{\epsilon_0} \qquad (24.23)$$

Dielectric between plates



BRIDGING PROBLEM ELECTRIC-FIELD ENERGY AND CAPACITANCE OF A CONDUCTING SPHERE

SOLUTION

A solid conducting sphere of radius R carries a charge Q. Calculate the electric-field energy density at a point a distance r from the center of the sphere for (a) $r < R$ and (b) $r > R$. (c) Calculate the total electric-field energy associated with the charged sphere. (d) How much work is required to assemble the charge Q on the sphere? (e) Use the result of part (c) to find the capacitance of the sphere. (You can think of the second conductor as a hollow conducting shell of infinite radius.)

SOLUTION GUIDE

IDENTIFY and SET UP

1. You know the electric field for this situation at all values of r from Example 22.5 (Section 22.4). You'll use this to find the electric-field energy density u and *total* electric-field energy U. You can then find the capacitance from the relationship $U = Q^2/2C$.
2. To find U, consider a spherical shell of radius r and thickness dr that has volume $dV = 4\pi r^2 dr$. (It will help to make a drawing of such a shell concentric with the conducting sphere.) The energy stored in this volume is $u\,dV$, and the total energy is the integral of $u\,dV$ from $r = 0$ to $r \rightarrow \infty$. Set up this integral.

EXECUTE

3. Find u for $r < R$ and for $r > R$. (*Hint:* What is the field inside a solid conductor?)
4. Substitute your results from step 3 into the expression from step 2. Then calculate the integral to find the total electric-field energy U.
5. Use your understanding of the energy stored in a charge distribution to find the work required to assemble the charge Q.
6. Find the capacitance of the sphere.

EVALUATE

7. Where is the electric-field energy density greatest? Where is it least?
8. How would the results be affected if the solid sphere were replaced by a hollow conducting sphere of the same radius R?
9. You can find the potential difference between the sphere and infinity from $C = Q/V$. Does this agree with the result of Example 23.8 (Section 23.3)?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q24.1 Equation (24.2) shows that the capacitance of a parallel-plate capacitor becomes larger as the plate separation d decreases. However, there is a practical limit to how small d can be made, which places limits on how large C can be. Explain what sets the limit on d. (*Hint:* What happens to the magnitude of the electric field as $d \rightarrow 0$?)

Q24.2 Suppose several different parallel-plate capacitors are charged up by a constant-voltage source. Thinking of the actual movement and position of the charges on an atomic level, why does it make sense that the capacitances are proportional to the surface areas of the plates? Why does it make sense that the capacitances are *inversely* proportional to the distance between the plates?

Q24.3 Suppose the two plates of a capacitor have different areas. When the capacitor is charged by connecting it to a battery, do the charges on the two plates have equal magnitude, or may they be different? Explain your reasoning.

Q24.4 To store the maximum amount of energy in a parallel-plate capacitor with a given battery (voltage source), would it be better to have the plates far apart or close together?

Q24.5 In the parallel-plate capacitor of Fig. 24.2, suppose the plates are pulled apart so that the separation d is much larger than the size of the plates. (a) Is it still accurate to say that the electric field between the plates is uniform? Why or why not? (b) In the situation shown in Fig. 24.2, the potential difference between the plates is $V_{ab} = Qd/\epsilon_0 A$. If the plates are pulled apart as described above, is V_{ab} more or less than this formula would indicate? Explain your reasoning. (c) With the plates pulled apart as described above, is the capacitance more than, less than, or the same as that given by Eq. (24.2)? Explain your reasoning.

Q24.6 A parallel-plate capacitor is charged by being connected to a battery and is kept connected to the battery. The separation between the plates is then doubled. How does the electric field change? The charge on the plates? The total energy? Explain.

Q24.7 A parallel-plate capacitor is charged by being connected to a battery and is then disconnected from the battery. The separation between the plates is then doubled. How does the electric field change? The potential difference? The total energy? Explain.

Q24.8 Two parallel-plate capacitors, identical except that one has twice the plate separation of the other, are charged by the same voltage source. Which capacitor has a stronger electric field between the plates? Which capacitor has a greater charge? Which has greater energy density? Explain your reasoning.

Q24.9 The charged plates of a capacitor attract each other, so to pull the plates farther apart requires work by some external force. What becomes of the energy added by this work? Explain.

Q24.10 You have two capacitors and want to connect them across a voltage source (battery) to store the maximum amount of energy. Should they be connected in series or in parallel?

Q24.11 As shown in Table 24.1, water has a very large dielectric constant $K = 80.4$. Why do you think water is not commonly used as a dielectric in capacitors?

Q24.12 Is dielectric strength the same thing as dielectric constant? Explain any differences between the two quantities. Is there a simple relationship between dielectric strength and dielectric constant (see Table 24.2)?

Q24.13 A capacitor made of aluminum foil strips separated by Mylar film was subjected to excessive voltage, and the resulting dielectric breakdown melted holes in the Mylar. After this, the capacitance was found to be about the same as before, but the breakdown voltage was much less. Why?

Q24.14 Suppose you bring a slab of dielectric close to the gap between the plates of a charged capacitor, preparing to slide it between the plates. What force will you feel? What does this force tell you about the energy stored between the plates once the dielectric is in place, compared to before the dielectric is in place?

Q24.15 The freshness of fish can be measured by placing a fish between the plates of a capacitor and measuring the capacitance. How does this work? (*Hint:* As time passes, the fish dries out. See Table 24.1.)

Q24.16 *Electrolytic* capacitors use as their dielectric an extremely thin layer of nonconducting oxide between a metal plate and a conducting solution. Discuss the advantage of such a capacitor over one constructed using a solid dielectric between the metal plates.

Q24.17 In terms of the dielectric constant K, what happens to the electric flux through the Gaussian surface shown in Fig. 24.22 when the dielectric is inserted into the previously empty space between the plates? Explain.

Q24.18 A parallel-plate capacitor is connected to a power supply that maintains a fixed potential difference between the plates. (a) If a sheet of dielectric is then slid between the plates, what happens to (i) the electric field between the plates, (ii) the magnitude of charge on each plate, and (iii) the energy stored in the capacitor? (b) Now suppose that before the dielectric is inserted, the charged capacitor is disconnected from the power supply. In this case, what happens to (i) the electric field between the plates, (ii) the magnitude of charge on each plate, and (iii) the energy stored in the capacitor? Explain any differences between the two situations.

Q24.19 Liquid dielectrics that have polar molecules (such as water) always have dielectric constants that decrease with increasing temperature. Why?

Q24.20 A conductor is an extreme case of a dielectric, since if an electric field is applied to a conductor, charges are free to move within the conductor to set up "induced charges." What is the dielectric constant of a perfect conductor? Is it $K = 0$, $K \rightarrow \infty$, or something in between? Explain your reasoning.

Q24.21 The two plates of a capacitor are given charges $\pm Q$. The capacitor is then disconnected from the charging device so that the charges on the plates can't change, and the capacitor is immersed in a tank of oil. Does the electric field between the plates increase, decrease, or stay the same? Explain your reasoning. How can this field be measured?

EXERCISES

Section 24.1 Capacitors and Capacitance

24.1 • The plates of a parallel-plate capacitor are 2.50 mm apart, and each carries a charge of magnitude 80.0 nC. The plates are in vacuum. The electric field between the plates has a magnitude of 4.00×10^6 V/m. What is (a) the potential difference between the plates; (b) the area of each plate; (c) the capacitance?

24.2 • The plates of a parallel-plate capacitor are 3.28 mm apart, and each has an area of 9.82 cm^2. Each plate carries a charge of magnitude 4.35×10^{-8} C. The plates are in vacuum. What is (a) the capacitance; (b) the potential difference between the plates; (c) the magnitude of the electric field between the plates?

24.3 • A parallel-plate air capacitor of capacitance 245 pF has a charge of magnitude 0.148 μC on each plate. The plates are 0.328 mm apart. (a) What is the potential difference between the plates? (b) What is the area of each plate? (c) What is the electric-field magnitude between the plates? (d) What is the surface charge density on each plate?

24.4 •• Cathode-ray-tube oscilloscopes have parallel metal plates inside them to deflect the electron beam. These plates are called the *deflecting plates*. Typically, they are squares 3.0 cm on a side and separated by 5.0 mm, with vacuum in between. What is the capacitance of these deflecting plates and hence of the oscilloscope? (*Note:* This capacitance can sometimes have an effect on the circuit you are trying to study and must be taken into consideration in your calculations.)

24.5 • A 10.0-μF parallel-plate capacitor with circular plates is connected to a 12.0-V battery. (a) What is the charge on each plate? (b) How much charge would be on the plates if their separation were doubled while the capacitor remained connected to the battery? (c) How much charge would be on the plates if the capacitor were connected to the 12.0-V battery after the radius of each plate was doubled without changing their separation?

24.6 • A 5.00-μF parallel-plate capacitor is connected to a 12.0-V battery. After the capacitor is fully charged, the battery is disconnected without loss of any of the charge on the plates. (a) A voltmeter is connected across the two plates without discharging them. What does it read? (b) What would the voltmeter read if (i) the plate separation were doubled; (ii) the radius of each plate were doubled but their separation was unchanged?

24.7 • A parallel-plate air capacitor is to store charge of magnitude 240.0 pC on each plate when the potential difference between the plates is 42.0 V. (a) If the area of each plate is 6.80 cm^2, what is the separation between the plates? (b) If the separation between the two plates is double the value calculated in part (a), what potential difference is required for the capacitor to store charge of magnitude 240.0 pC on each plate?

24.8 • A 5.00-pF, parallel-plate, air-filled capacitor with circular plates is to be used in a circuit in which it will be subjected to potentials of up to 1.00×10^2 V. The electric field between the plates is to be no greater than 1.00×10^4 N/C. As a budding electrical engineer for Live-Wire Electronics, your tasks are to (a) design the capacitor by finding what its physical dimensions and separation must be; (b) find the maximum charge these plates can hold.

24.9 • A capacitor is made from two hollow, coaxial, iron cylinders, one inside the other. The inner cylinder is negatively charged and the outer is positively charged; the magnitude of the charge on each is 10.0 pC. The inner cylinder has radius 0.50 mm, the outer one has radius 5.00 mm, and the length of each cylinder is 18.0 cm. (a) What is the capacitance? (b) What applied potential difference is necessary to produce these charges on the cylinders?

24.10 • A cylindrical capacitor consists of a solid inner conducting core with radius 0.250 cm, surrounded by an outer hollow conducting tube. The two conductors are separated by air, and the length of the cylinder is 12.0 cm. The capacitance is 36.7 pF. (a) Calculate the inner radius of the hollow tube. (b) When the capacitor is charged to 125 V, what is the charge per unit length λ on the capacitor?

24.11 •• A spherical capacitor contains a charge of 3.30 nC when connected to a potential difference of 220 V. If its plates are separated by vacuum and the inner radius of the outer shell is 4.00 cm, calculate: (a) the capacitance; (b) the radius of the inner sphere; (c) the electric field just outside the surface of the inner sphere.

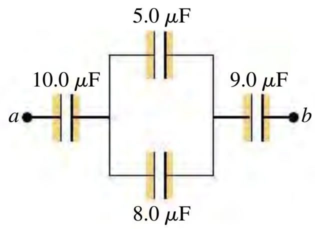
24.12 •• A cylindrical capacitor has an inner conductor of radius 2.2 mm and an outer conductor of radius 3.5 mm. The two conductors are separated by vacuum, and the entire capacitor is 2.8 m long. (a) What is the capacitance per unit length? (b) The potential of the inner conductor is 350 mV higher than that of the outer conductor. Find the charge (magnitude and sign) on both conductors.

24.13 • A spherical capacitor is formed from two concentric, spherical, conducting shells separated by vacuum. The inner sphere has radius 15.0 cm and the capacitance is 116 pF. (a) What is the radius of the outer sphere? (b) If the potential difference between the two spheres is 220 V, what is the magnitude of charge on each sphere?

Section 24.2 Capacitors in Series and Parallel

24.14 • **Figure E24.14** shows a system of four capacitors, where the potential difference across *ab* is 50.0 V. (a) Find the equivalent capacitance of this system between *a* and *b*. (b) How much charge is stored by this combination of capacitors? (c) How much charge is stored in each of the 10.0-μF and the 9.0-μF capacitors?

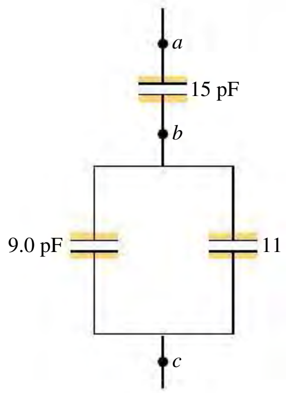
Figure **E24.14**



24.15 • BIO **Electric Eels.** Electric eels and electric fish generate large potential differences that are used to stun enemies and prey. These potentials are produced by cells that each can generate 0.10 V. We can plausibly model such cells as charged capacitors. (a) How should these cells be connected (in series or in parallel) to produce a total potential of more than 0.10 V? (b) Using the connection in part (a), how many cells must be connected together to produce the 500-V surge of the electric eel?

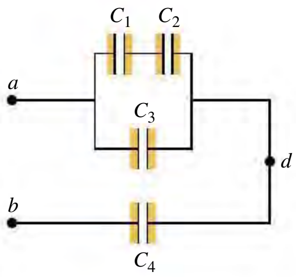
24.16 • For the system of capacitors shown in **Fig. E24.16**, find the equivalent capacitance (a) between *b* and *c*, and (b) between *a* and *c*.

Figure **E24.16**



24.17 • In **Fig. E24.17**, each capacitor has $C = 4.00\ \mu\text{F}$ and $V_{ab} = +28.0\ \text{V}$. Calculate (a) the charge on each capacitor; (b) the potential difference across each capacitor; (c) the potential difference between points a and d.

Figure **E24.17**

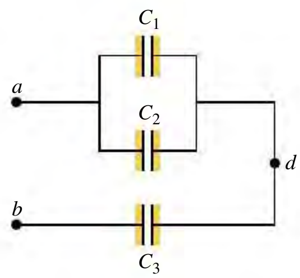


24.18 • In Fig. 24.8a, let $C_1 = 3.00\ \mu\text{F}$, $C_2 = 5.00\ \mu\text{F}$, and $V_{ab} = +64.0\ \text{V}$. Calculate (a) the charge on each capacitor and (b) the potential difference across each capacitor.

24.19 • In Fig. 24.9a, let $C_1 = 3.00\ \mu\text{F}$, $C_2 = 5.00\ \mu\text{F}$, and $V_{ab} = +52.0\ \text{V}$. Calculate (a) the charge on each capacitor and (b) the potential difference across each capacitor.

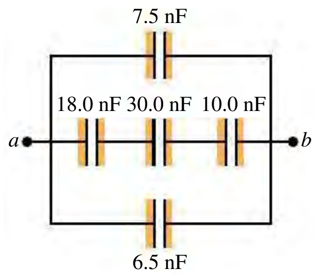
24.20 • In **Fig. E24.20**, $C_1 = 6.00\ \mu\text{F}$, $C_2 = 3.00\ \mu\text{F}$, and $C_3 = 5.00\ \mu\text{F}$. The capacitor network is connected to an applied potential V_{ab}. After the charges on the capacitors have reached their final values, the charge on C_2 is $30.0\ \mu\text{C}$. (a) What are the charges on capacitors C_1 and C_3? (b) What is the applied voltage V_{ab}?

Figure **E24.20**



24.21 •• For the system of capacitors shown in **Fig. E24.21**, a potential difference of 25 V is maintained across ab. (a) What is the equivalent capacitance of this system between a and b? (b) How much charge is stored by this system? (c) How much charge does the 6.5-nF capacitor store? (d) What is the potential difference across the 7.5-nF capacitor?

Figure **E24.21**



24.22 •• Suppose the 3-μF capacitor in Fig. 24.10a were removed and replaced by a different one, and that this changed the equivalent capacitance between points a and b to 8 μF. What would be the capacitance of the replacement capacitor?

Section 24.3 Energy Storage in Capacitors and Electric-Field Energy

24.23 • A 5.80-μF, parallel-plate, air capacitor has a plate separation of 5.00 mm and is charged to a potential difference of 400 V. Calculate the energy density in the region between the plates, in units of J/m^3.

24.24 • A parallel-plate air capacitor has a capacitance of 920 pF. The charge on each plate is $3.90\ \mu\text{C}$. (a) What is the potential difference between the plates? (b) If the charge is kept constant, what will be the potential difference if the plate separation is doubled? (c) How much work is required to double the separation?

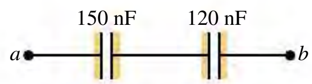
24.25 • An air capacitor is made from two flat parallel plates 1.50 mm apart. The magnitude of charge on each plate is $0.0180\ \mu\text{C}$ when the potential difference is 200 V. (a) What is the capacitance? (b) What is the area of each plate? (c) What maximum voltage can be applied without dielectric breakdown? (Dielectric breakdown for air occurs at an electric-field strength of $3.0 \times 10^6\ \text{V/m}$.) (d) When the charge is $0.0180\ \mu\text{C}$, what total energy is stored?

24.26 •• A parallel-plate vacuum capacitor has 8.38 J of energy stored in it. The separation between the plates is 2.30 mm. If the separation is decreased to 1.15 mm, what is the energy stored (a) if the capacitor is disconnected from the potential source so the charge on the plates remains constant, and (b) if the capacitor remains connected to the potential source so the potential difference between the plates remains constant?

24.27 • You have two identical capacitors and an external potential source. (a) Compare the total energy stored in the capacitors when they are connected to the applied potential in series and in parallel. (b) Compare the maximum amount of charge stored in each case. (c) Energy storage in a capacitor can be limited by the maximum electric field between the plates. What is the ratio of the electric field for the series and parallel combinations?

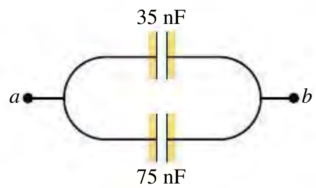
24.28 • For the capacitor network shown in **Fig. E24.28**, the potential difference across ab is 48 V. Find (a) the total charge stored in this network; (b) the charge on each capacitor; (c) the total energy stored in the network; (d) the energy stored in each capacitor; (e) the potential differences across each capacitor.

Figure **E24.28**



24.29 • For the capacitor network shown in **Fig. E24.29**, the potential difference across ab is 220 V. Find (a) the total charge stored in this network; (b) the charge on each capacitor; (c) the total energy stored in the network; (d) the energy stored in each capacitor; (e) the potential difference across each capacitor.

Figure **E24.29**



24.30 • A 0.350-m-long cylindrical capacitor consists of a solid conducting core with a radius of 1.20 mm and an outer hollow conducting tube with an inner radius of 2.00 mm. The two conductors are separated by air and charged to a potential difference of 6.00 V. Calculate (a) the charge per length for the capacitor; (b) the total charge on the capacitor; (c) the capacitance; (d) the energy stored in the capacitor when fully charged.

24.31 • A cylindrical air capacitor of length 15.0 m stores 3.20×10^{-9} J of energy when the potential difference between the two conductors is 4.00 V. (a) Calculate the magnitude of the charge on each conductor. (b) Calculate the ratio of the radii of the inner and outer conductors.

24.32 •• A capacitor is formed from two concentric spherical conducting shells separated by vacuum. The inner sphere has radius 12.5 cm, and the outer sphere has radius 14.8 cm. A potential difference of 120 V is applied to the capacitor. (a) What is the energy density at $r = 12.6$ cm, just outside the inner sphere? (b) What is the energy density at $r = 14.7$ cm, just inside the outer sphere? (c) For a parallel-plate capacitor the energy density is uniform in the region between the plates, except near the edges of the plates. Is this also true for a spherical capacitor?

Section 24.4 Dielectrics

24.33 • A 12.5-μF capacitor is connected to a power supply that keeps a constant potential difference of 24.0 V across the plates. A piece of material having a dielectric constant of 3.75 is placed between the plates, completely filling the space between them. (a) How much energy is stored in the capacitor before and after the dielectric is inserted? (b) By how much did the energy change during the insertion? Did it increase or decrease?

24.34 • A parallel-plate capacitor has capacitance $C_0 = 8.00$ pF when there is air between the plates. The separation between the plates is 1.50 mm. (a) What is the maximum magnitude of charge Q that can be placed on each plate if the electric field in the region between the plates is not to exceed 3.00×10^4 V/m? (b) A dielectric with $K = 2.70$ is inserted between the plates of the capacitor, completely filling the volume between the plates. Now what is the maximum magnitude of charge on each plate if the electric field between the plates is not to exceed 3.00×10^4 V/m?

24.35 • Two parallel plates have equal and opposite charges. When the space between the plates is evacuated, the electric field is $E = 3.20 \times 10^5$ V/m. When the space is filled with dielectric, the electric field is $E = 2.50 \times 10^5$ V/m. (a) What is the charge density on each surface of the dielectric? (b) What is the dielectric constant?

24.36 • A budding electronics hobbyist wants to make a simple 1.0-nF capacitor for tuning her crystal radio, using two sheets of aluminum foil as plates, with a few sheets of paper between them as a dielectric. The paper has a dielectric constant of 3.0, and the thickness of one sheet of it is 0.20 mm. (a) If the sheets of paper measure 22×28 cm and she cuts the aluminum foil to the same dimensions, how many sheets of paper should she use between her plates to get the proper capacitance? (b) Suppose for convenience she wants to use a single sheet of posterboard, with the same dielectric constant but a thickness of 12.0 mm, instead of the paper. What area of aluminum foil will she need for her plates to get her 1.0 nF of capacitance? (c) Suppose she goes high-tech and finds a sheet of Teflon of the same thickness as the posterboard to use as a dielectric. Will she need a larger or smaller area of Teflon than of posterboard? Explain.

24.37 • The dielectric to be used in a parallel-plate capacitor has a dielectric constant of 3.60 and a dielectric strength of 1.60×10^7 V/m. The capacitor is to have a capacitance of 1.25×10^{-9} F and must be able to withstand a maximum potential difference of 5500 V. What is the minimum area the plates of the capacitor may have?

24.38 •• BIO **Potential in Human Cells.** Some cell walls in the human body have a layer of negative charge on the inside surface and a layer of positive charge of equal magnitude on the outside surface. Suppose that the charge density on either surface is $\pm 0.50 \times 10^{-3}$ C/m^2, the cell wall is 5.0 nm thick, and the cell-wall material is air. (a) Find the magnitude of $\vec{E}$ in the wall between the two layers of charge. (b) Find the potential difference between the inside and the outside of the cell. Which is at the higher potential? (c) A typical cell in the human body has a volume of 10^{-16} m^3. Estimate the total electric-field energy stored in the wall of a cell of this size. (*Hint:* Assume that the cell is spherical, and calculate the volume of the cell wall.) (d) In reality, the cell wall is made up, not of air, but of tissue with a dielectric constant of 5.4. Repeat parts (a) and (b) in this case.

24.39 • A constant potential difference of 12 V is maintained between the terminals of a 0.25-μF, parallel-plate, air capacitor. (a) A sheet of Mylar is inserted between the plates of the capacitor, completely filling the space between the plates. When this is done, how much additional charge flows onto the positive plate of the capacitor (see Table 24.1)? (b) What is the total induced charge on either face of the Mylar sheet? (c) What effect does the Mylar sheet have on the electric field between the plates? Explain how you can reconcile this with the increase in charge on the plates, which acts to *increase* the electric field.

24.40 •• Polystyrene has dielectric constant 2.6 and dielectric strength 2.0×10^7 V/m. A piece of polystyrene is used as a dielectric in a parallel-plate capacitor, filling the volume between the plates. (a) When the electric field between the plates is 80% of the dielectric strength, what is the energy density of the stored energy? (b) When the capacitor is connected to a battery with voltage 500.0 V, the electric field between the plates is 80% of the dielectric strength. What is the area of each plate if the capacitor stores 0.200 mJ of energy under these conditions?

24.41 • When a 360-nF air capacitor ($1 \text{ nF} = 10^{-9}$ F) is connected to a power supply, the energy stored in the capacitor is 1.85×10^{-5} J. While the capacitor is kept connected to the power supply, a slab of dielectric is inserted that completely fills the space between the plates. This increases the stored energy by 2.32×10^{-5} J. (a) What is the potential difference between the capacitor plates? (b) What is the dielectric constant of the slab?

24.42 • A parallel-plate capacitor has capacitance $C = 12.5$ pF when the volume between the plates is filled with air. The plates are circular, with radius 3.00 cm. The capacitor is connected to a battery, and a charge of magnitude 25.0 pC goes onto each plate. With the capacitor still connected to the battery, a slab of dielectric is inserted between the plates, completely filling the space between the plates. After the dielectric has been inserted, the charge on each plate has magnitude 45.0 pC. (a) What is the dielectric constant K of the dielectric? (b) What is the potential difference between the plates before and after the dielectric has been inserted? (c) What is the electric field at a point midway between the plates before and after the dielectric has been inserted?

Section 24.6 Gauss's Law in Dielectrics

24.43 • A parallel-plate capacitor has the volume between its plates filled with plastic with dielectric constant K. The magnitude of the charge on each plate is Q. Each plate has area A, and the distance between the plates is d. (a) Use Gauss's law as stated in Eq. (24.23) to calculate the magnitude of the electric field in the dielectric. (b) Use the electric field determined in part (a) to calculate the potential difference between the two plates. (c) Use the result of part (b) to determine the capacitance of the capacitor. Compare your result to Eq. (24.12).

24.44 • A parallel-plate capacitor has plates with area 0.0225 m^2 separated by 1.00 mm of Teflon. (a) Calculate the charge on the plates when they are charged to a potential difference of 12.0 V. (b) Use Gauss's law (Eq. 24.23) to calculate the electric field inside the Teflon. (c) Use Gauss's law to calculate the electric field if the voltage source is disconnected and the Teflon is removed.

PROBLEMS

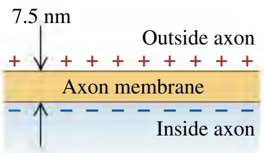
24.45 • Electronic flash units for cameras contain a capacitor for storing the energy used to produce the flash. In one such unit, the flash lasts for $\frac{1}{675}$ s with an average light power output of 2.70×10^5 W. (a) If the conversion of electrical energy to light is 95% efficient (the rest of the energy goes to thermal energy), how much energy must be stored in the capacitor for one flash? (b) The capacitor has a potential difference between its plates of 125 V when the stored energy equals the value calculated in part (a). What is the capacitance?

24.46 • A parallel-plate air capacitor is made by using two plates 12 cm square, spaced 3.7 mm apart. It is connected to a 12-V battery. (a) What is the capacitance? (b) What is the charge on each plate? (c) What is the electric field between the plates? (d) What is the energy stored in the capacitor? (e) If the battery is disconnected and then the plates are pulled apart to a separation of 7.4 mm, what are the answers to parts (a)–(d)?

24.47 ••• In one type of computer keyboard, each key holds a small metal plate that serves as one plate of a parallel-plate, air-filled capacitor. When the key is depressed, the plate separation decreases and the capacitance increases. Electronic circuitry detects the change in capacitance and thus detects that the key has been pressed. In one particular keyboard, the area of each metal plate is 42.0 mm^2, and the separation between the plates is 0.700 mm before the key is depressed. (a) Calculate the capacitance before the key is depressed. (b) If the circuitry can detect a change in capacitance of 0.250 pF, how far must the key be depressed before the circuitry detects its depression?

24.48 ••• BIO **Cell Membranes.** Cell membranes (the walled enclosure around a cell) are typically about 7.5 nm thick. They are partially permeable to allow charged material to pass in and out, as needed. Equal but opposite charge densities build up on the inside and outside faces of such a membrane, and these charges prevent additional charges from passing through the cell wall. We can model a cell membrane as a parallel-plate capacitor, with the membrane itself containing proteins embedded in an organic material to give the membrane a dielectric constant of about 10. (See **Fig. P24.48**.) (a) What is the capacitance per square centimeter of such a cell wall? (b) In its normal resting state, a cell has a potential difference of 85 mV across its membrane. What is the electric field inside this membrane?

Figure **P24.48**

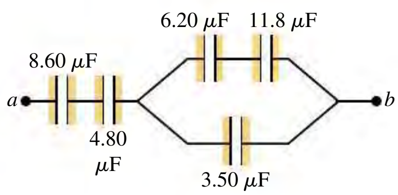


24.49 •• A 20.0-μF capacitor is charged to a potential difference of 800 V. The terminals of the charged capacitor are then connected to those of an uncharged 10.0-μF capacitor. Compute (a) the original charge of the system, (b) the final potential difference across each capacitor, (c) the final energy of the system, and (d) the decrease in energy when the capacitors are connected.

24.50 •• In Fig. 24.9a, let $C_1 = 9.0\ \mu$F, $C_2 = 4.0\ \mu$F, and $V_{ab} = 64$ V. Suppose the charged capacitors are disconnected from the source and from each other, and then reconnected to each other with plates of *opposite* sign together. By how much does the energy of the system decrease?

24.51 • For the capacitor network shown in **Fig. P24.51**, the potential difference across *ab* is 12.0 V. Find (a) the total energy stored in this network and (b) the energy stored in the 4.80-μF capacitor.

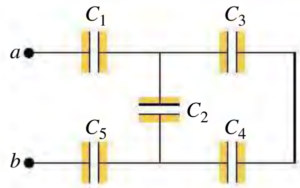
Figure **P24.51**



24.52 •• In Fig. E24.17, $C_1 = 6.00\ \mu$F, $C_2 = 3.00\ \mu$F, $C_3 = 4.00\ \mu$F, and $C_4 = 8.00\ \mu$F. The capacitor network is connected to an applied potential difference V_{ab}. After the charges on the capacitors have reached their final values, the voltage across C_3 is 40.0 V. What are (a) the voltages across C_1 and C_2, (b) the voltage across C_4, and (c) the voltage V_{ab} applied to the network?

24.53 • In **Fig. P24.53**, $C_1 = C_5 = 8.4\ \mu$F and $C_2 = C_3 = C_4 = 4.2\ \mu$F. The applied potential is $V_{ab} = 220$ V. (a) What is the equivalent capacitance of the network between points *a* and *b*? (b) Calculate the charge on each capacitor and the potential difference across each capacitor.

Figure **P24.53**

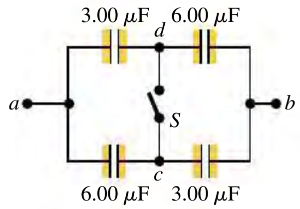


24.54 •• Current materials-science technology allows engineers to construct capacitors with much higher values of *C* than were previously possible. A capacitor has $C = 3000$ F and is rated to withstand a maximum potential difference of 2.7 V. The cylindrical capacitor has diameter 6.0 cm and length 13.5 cm. (a) Find the maximum electric potential energy that can be stored in this capacitor. (b) Does your value in part (a) agree with the 3.0-Wh value printed on the capacitor? (c) What is the maximum attainable energy density in this capacitor? (d) Compare this maximum energy density to the maximum possible energy density for polyester (see Table 24.2).

24.55 •• In Fig. E24.20, $C_1 = 3.00\ \mu$F and $V_{ab} = 150$ V. The charge on capacitor C_1 is 150 μC and the charge on C_3 is 450 μC. What are the values of the capacitances of C_2 and C_3?

24.56 • The capacitors in **Fig. P24.56** are initially uncharged and are connected, as in the diagram, with switch *S* open. The applied potential difference is $V_{ab} = +210$ V. (a) What is the potential difference V_{cd}? (b) What is the potential difference across each capacitor after switch *S* is closed? (c) How much charge flowed through the switch when it was closed?

Figure **P24.56**

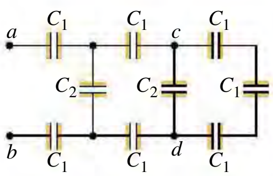


24.57 •• Three capacitors having capacitances of 8.4, 8.4, and 4.2 μF are connected in series across a 36-V potential difference. (a) What is the charge on the 4.2-μF capacitor? (b) What is the total energy stored in all three capacitors? (c) The capacitors are disconnected from the potential difference without allowing them to discharge. They are then reconnected in parallel with each other, with the positively charged plates connected together. What is the voltage across each capacitor in the parallel combination? (d) What is the total energy now stored in the capacitors?

24.58 • **Capacitance of a Thundercloud.** The charge center of a thundercloud, drifting 3.0 km above the earth's surface, contains 20 C of negative charge. Assuming the charge center has a radius of 1.0 km, and modeling the charge center and the earth's surface as parallel plates, calculate: (a) the capacitance of the system; (b) the potential difference between charge center and ground; (c) the average strength of the electric field between cloud and ground; (d) the electrical energy stored in the system.

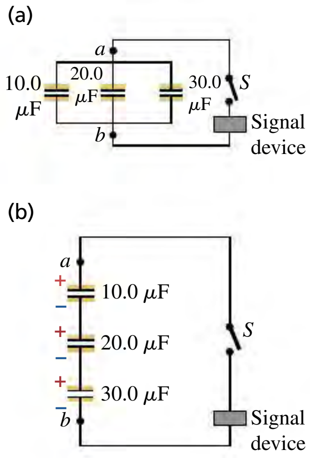
24.59 •• In **Fig. P24.59**, each capacitance C_1 is 6.9 μF, and each capacitance C_2 is 4.6 μF. (a) Compute the equivalent capacitance of the network between points a and b. (b) Compute the charge on each of the three capacitors nearest a and b when $V_{ab} = 420$ V. (c) With 420 V across a and b, compute V_{cd}.

Figure **P24.59**



24.60 • Each combination of capacitors between points a and b in **Fig. P24.60** is first connected across a 120-V battery, charging the combination to 120 V. These combinations are then connected to make the circuits shown. When the switch S is thrown, a surge of charge for the discharging capacitors flows to trigger the signal device. How much charge flows through the signal device in each case?

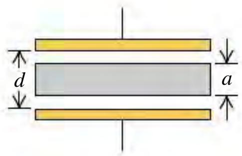
Figure **P24.60**



24.61 • A parallel-plate capacitor with only air between the plates is charged by connecting it to a battery. The capacitor is then disconnected from the battery, without any of the charge leaving the plates. (a) A voltmeter reads 45.0 V when placed across the capacitor. When a dielectric is inserted between the plates, completely filling the space, the voltmeter reads 11.5 V. What is the dielectric constant of this material? (b) What will the voltmeter read if the dielectric is now pulled partway out so it fills only one-third of the space between the plates?

24.62 •• An air capacitor is made by using two flat plates, each with area A, separated by a distance d. Then a metal slab having thickness a (less than d) and the same shape and size as the plates is inserted between them, parallel to the plates and not touching either plate (**Fig. P24.62**). (a) What is the capacitance of this arrangement? (b) Express the capacitance as a multiple of the capacitance C_0 when the metal slab is not present. (c) Discuss what happens to the capacitance in the limits $a \rightarrow 0$ and $a \rightarrow d$.

Figure **P24.62**



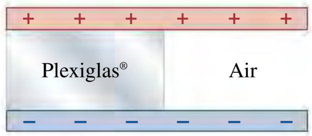
24.63 •• A potential difference $V_{ab} = 48.0$ V is applied across the capacitor network of Fig. E24.17. If $C_1 = C_2 = 4.00\ \mu$F and $C_4 = 8.00\ \mu$F, what must the capacitance C_3 be if the network is to store 2.90×10^{-3} J of electrical energy?

24.64 • **CALC** The inner cylinder of a long, cylindrical capacitor has radius r_a and linear charge density $+\lambda$. It is surrounded by a coaxial cylindrical conducting shell with inner radius r_b and linear charge density $-\lambda$ (see Fig. 24.6). (a) What is the energy density in the region between the conductors at a distance r from the axis? (b) Integrate the energy density calculated in part (a) over the volume between the conductors in a length L of the capacitor to obtain the total electric-field energy per unit length. (c) Use Eq. (24.9) and the capacitance per unit length calculated in Example 24.4 (Section 24.1) to calculate U/L. Does your result agree with that obtained in part (b)?

24.65 •• A parallel-plate capacitor has square plates that are 8.00 cm on each side and 3.80 mm apart. The space between the plates is completely filled with two square slabs of dielectric, each 8.00 cm on a side and 1.90 mm thick. One slab is Pyrex glass and the other is polystyrene. If the potential difference between the plates is 86.0 V, how much electrical energy is stored in the capacitor?

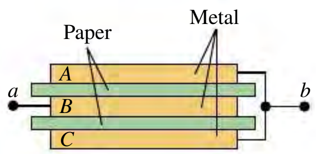
24.66 •• A parallel-plate capacitor is made from two plates 12.0 cm on each side and 4.50 mm apart. Half of the space between these plates contains only air, but the other half is filled with Plexiglas® of dielectric constant 3.40 (**Fig. P24.66**). An 18.0-V battery is connected across the plates. (a) What is the capacitance of this combination? (*Hint:* Can you think of this capacitor as equivalent to two capacitors in parallel?) (b) How much energy is stored in the capacitor? (c) If we remove the Plexiglas® but change nothing else, how much energy will be stored in the capacitor?

Figure **P24.66**



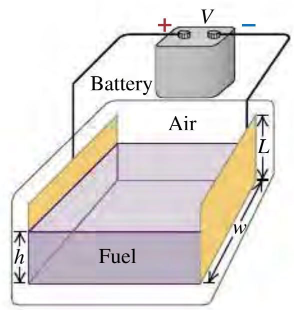
24.67 •• Three square metal plates A, B, and C, each 12.0 cm on a side and 1.50 mm thick, are arranged as in **Fig. P24.67**. The plates are separated by sheets of paper 0.45 mm thick and with dielectric constant 4.2. The outer plates are connected together and connected to point b. The inner plate is connected to point a. (a) Copy the diagram and show by plus and minus signs the charge distribution on the plates when point a is maintained at a positive potential relative to point b. (b) What is the capacitance between points a and b?

Figure **P24.67**



24.68 •• A fuel gauge uses a capacitor to determine the height of the fuel in a tank. The effective dielectric constant K_{eff} changes from a value of 1 when the tank is empty to a value of K, the dielectric constant of the fuel, when the tank is full. The appropriate electronic circuitry can determine the effective dielectric constant of the combined air and fuel between the capacitor plates. Each of the two rectangular plates has a width w and a length L (**Fig. P24.68**). The height of the fuel between the plates is h. You can ignore any fringing effects. (a) Derive an expression for K_{eff} as a function of h. (b) What is the effective dielectric constant for a tank $\frac{1}{4}$ full, $\frac{1}{2}$ full, and $\frac{3}{4}$ full if the fuel is gasoline ($K = 1.95$)? (c) Repeat part (b) for methanol ($K = 33.0$). (d) For which fuel is this fuel gauge more practical?

Figure **P24.68**

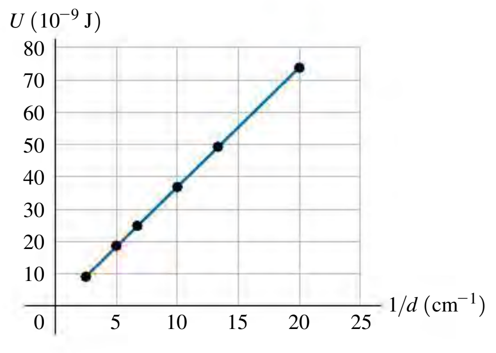


24.69 •• **DATA** Your electronics company has several identical capacitors with capacitance C_1 and several others with capacitance C_2. You must determine the values of C_1 and C_2 but don't have access to C_1 and C_2 individually. Instead, you have a network with C_1 and C_2 connected in series and a network with C_1 and C_2 connected in parallel. You have a 200.0-V battery and instrumentation that measures the total energy supplied by the battery when it is connected to the network. When the parallel combination is connected to the battery, 0.180 J of energy is stored in the network. When the series combination is connected, 0.0400 J of energy is stored. You are told that C_1 is greater than C_2. (a) Calculate C_1 and C_2. (b) For the series combination, does C_1 or C_2 store more charge, or are the values equal? Does C_1 or C_2 store more energy, or are the values equal? (c) Repeat part (b) for the parallel combination.

24.70 •• **DATA** You are designing capacitors for various applications. For one application, you want the maximum possible stored energy. For another, you want the maximum stored charge. For a third application, you want the capacitor to withstand a large applied voltage without dielectric breakdown. You start with an air-filled parallel-plate capacitor that has $C_0 = 6.00$ pF and a plate separation of 2.50 mm. You then consider the use of each of the dielectric materials listed in Table 24.2. In each application, the dielectric will fill the volume between the plates, and the electric field between the plates will be 50% of the dielectric strength given in the table. (a) For each of the five materials given in the table, calculate the energy stored in the capacitor. Which dielectric allows the maximum stored energy? (b) For each material, what is the charge Q stored on each plate of the capacitor? (c) For each material, what is the voltage applied across the capacitor? (d) Is one dielectric material in the table your best choice for all three applications?

24.71 •• **DATA** You are conducting experiments with an air-filled parallel-plate capacitor. You connect the capacitor to a battery with voltage 24.0 V. Initially the separation d between the plates is 0.0500 cm. In one experiment, you leave the battery connected to the capacitor, increase the separation between the plates, and measure the energy stored in the capacitor for each value of d. In a second experiment, you make the same measurements but disconnect the battery before you change the plate separation. One set of your data is given in **Fig. P24.71**, where you have plotted the stored energy U versus $1/d$. (a) For which experiment does this data set apply: the first (battery remains connected) or the second (battery disconnected before d is changed)? Explain. (b) Use the data plotted in Fig. P24.71 to calculate the area A of each plate. (c) For which case, the battery connected or the battery disconnected, is there more energy stored in the capacitor when $d = 0.400$ cm? Explain.

Figure **P24.71**

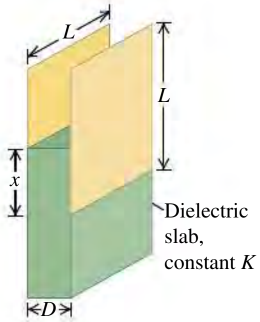


CHALLENGE PROBLEM

24.72 ••• Two square conducting plates with sides of length L are separated by a distance D. A dielectric slab with constant K with dimensions $L \times L \times D$ is inserted a distance x into the space between the plates, as shown in **Fig. P24.72**. (a) Find the capacitance C of this system. (b) Suppose that the capacitor is connected to a battery that maintains a constant potential difference V between the plates. If the dielectric slab is inserted an additional distance dx into the space between the plates, show that the change in stored energy is

$$dU = +\frac{(K-1)\epsilon_0 V^2 L}{2D}dx$$

Figure **P24.72**



(c) Suppose that before the slab is moved by dx, the plates are disconnected from the battery, so that the charges on the plates remain constant. Determine the magnitude of the charge on each plate, and then show that when the slab is moved dx farther into the space between the plates, the stored energy changes by an amount that is the *negative* of the expression for dU given in part (b). (d) If F is the force exerted on the slab by the charges on the plates, then dU should equal the work done *against* this force to move the slab a distance dx. Thus $dU = -F\,dx$. Show that applying this expression to the result of part (b) suggests that the electric force on the slab pushes it *out* of the capacitor, while the result of part (c) suggests that the force pulls the slab *into* the capacitor. (e) Figure 24.16 shows that the force in fact pulls the slab into the capacitor. Explain why the result of part (b) gives an incorrect answer for the direction of this force, and calculate the magnitude of the force. (This method does not require knowledge of the nature of the fringing field.)

PASSAGE PROBLEMS

BIO THE ELECTRIC EGG. Upon fertilization, the eggs of many species undergo a rapid change in potential difference across their outer membrane. This change affects the physiological development of the eggs. The potential difference across the membrane is called the *membrane potential,* V_m, which is the potential inside the membrane minus the potential outside it. The membrane potential arises when enzymes use the energy available in ATP to expel three sodium ions (Na^+) actively and accumulate two potassium ions (K^+) inside the membrane—making the interior less positively charged than the exterior. For a sea urchin egg, V_m is about -70 mV; that is, the potential inside is 70 mV less than that outside. The egg membrane behaves as a capacitor with a capacitance of about 1 $\mu F/cm^2$. The membrane of the unfertilized egg is *selectively permeable* to K^+; that is, K^+ can readily pass through certain channels in the membrane, but other ions cannot. When a sea urchin egg is fertilized, Na^+ channels in the membrane open, Na^+ enters the egg, and V_m rapidly increases to $+30$ mV, where it remains for several minutes. The concentration of Na^+ is about 30 mmol/L in the egg's interior but 450 mmol/L in the surrounding seawater. The K^+ concentration is about 200 mmol/L inside but 10 mmol/L outside. A useful constant that connects electrical and chemical units is the *Faraday number,* which has a value of approximately 10^5 C/mol; that is, Avogadro's number (a mole) of monovalent ions, such as Na^+ or K^+, carries a charge of 10^5 C.

24.73 How many moles of Na^+ must m ove per unit area of membrane to change V_m from -70 mV to $+30$ mV, if we assume that the membrane behaves purely as a capacitor? (a) 10^{-4} mol/cm^2; (b) 10^{-9} mol/cm^2; (c) 10^{-12} mol/cm^2; (d) 10^{-14} mol/cm^2.

24.74 Suppose that the egg has a diameter of 200 μm. What fractional change in the internal Na^+ concentration results from the fertilization-induced change in V_m? Assume that Na^+ ions are distributed throughout the cell volume. The concentration increases by (a) 1 part in 10^4; (b) 1 part in 10^5; (c) 1 part in 10^6; (d) 1 part in 10^7.

24.75 Suppose that the change in V_m was caused by the entry of Ca^{2+} instead of Na^+. How many Ca^{2+} ions would have to enter the cell per unit membrane to produce the change? (a) Half as many as for Na^+; (b) the same as for Na^+; (c) twice as many as for Na^+; (d) cannot say without knowing the inside and outside concentrations of Ca^{2+}.

24.76 What is the minimum amount of work that must be done by the cell to restore V_m to -70 mV? (a) 3 mJ; (b) 3 μJ; (c) 3 nJ; (d) 3 pJ.

Answers

Chapter Opening Question ?

(iv) Equation (24.9) shows that the energy stored in a capacitor with capacitance C and charge Q is $U = Q^2/2C$. If Q is doubled, the stored energy increases by a factor of $2^2 = 4$. Note that if the value of Q is too great, the electric-field magnitude inside the capacitor will exceed the dielectric strength of the material between the plates and dielectric breakdown will occur (see Section 24.4). This puts a practical limit on the amount of energy that can be stored.

Test Your Understanding Questions

24.1 (iii) The capacitance does not depend on the value of the charge Q. Doubling Q causes the potential difference V_{ab} to double, so the capacitance $C = Q/V_{ab}$ remains the same. These statements are true no matter what the geometry of the capacitor.

24.2 (a) (i), (b) (iv) In a series connection the two capacitors carry the same charge Q but have different potential differences $V_{ab} = Q/C$; the capacitor with the smaller capacitance C has the greater potential difference. In a parallel connection the two capacitors have the same potential difference V_{ab} but carry different charges $Q = CV_{ab}$; the capacitor with the larger capacitance C has the greater charge. Hence a 4-μF capacitor will have a greater potential difference than an 8-μF capacitor if the two are connected in series. The 4-μF capacitor cannot carry more charge than the 8-μF capacitor no matter how they are connected: In a series connection they will carry the same charge, and in a parallel connection the 8-μF capacitor will carry more charge.

24.3 (i) Capacitors connected in series carry the same charge Q. To compare the amount of energy stored, we use the expression $U = Q^2/2C$ from Eq. (24.9); it shows that the capacitor with the *smaller* capacitance ($C = 4\ \mu$F) has more stored energy in a series combination. By contrast, capacitors in parallel have the same potential difference V, so to compare them we use $U = \frac{1}{2}CV^2$ from Eq. (24.9). It shows that in a parallel combination, the capacitor with the *larger* capacitance ($C = 8\ \mu$F) has more stored energy. (If we had instead used $U = \frac{1}{2}CV^2$ to analyze the series combination, we would have to account for the different potential differences across the two capacitors. Likewise, using $U = Q^2/2C$ to study the parallel combination would require us to account for the different charges on the capacitors.)

24.4 (i) Here Q remains the same, so we use $U = Q^2/2C$ from Eq. (24.9) for the stored energy. Removing the dielectric lowers the capacitance by a factor of $1/K$; since U is inversely proportional to C, the stored energy *increases* by a factor of K. It takes work to pull the dielectric slab out of the capacitor because the fringing field tries to pull the slab back in (Fig. 24.16). The work that you do goes into the energy stored in the capacitor.

24.5 (i), (iii), (ii) Equation (24.14) says that if E_0 is the initial electric-field magnitude (before the dielectric slab is inserted), then the resultant field magnitude after the slab is inserted is $E_0/K = E_0/3$. The magnitude of the resultant field equals the difference between the initial field magnitude and the magnitude E_i of the field due to the bound charges (see Fig. 24.20). Hence $E_0 - E_i = E_0/3$ and $E_i = 2E_0/3$.

24.6 (iii) Equation (24.23) shows that this situation is the same as an isolated point charge in vacuum but with $\vec{E}$ replaced by $K\vec{E}$. Hence KE at the point of interest is equal to $q/4\pi\epsilon_0 r^2$, and so $E = q/4\pi K\epsilon_0 r^2$. As in Example 24.12, filling the space with a dielectric reduces the electric field by a factor of $1/K$.

Bridging Problem

(a) 0 **(b)** $Q^2/32\pi^2\epsilon_0 r^4$ **(c)** $Q^2/8\pi\epsilon_0 R$
(d) $Q^2/8\pi\epsilon_0 R$ **(e)** $C = 4\pi\epsilon_0 R$

? In a flashlight, how does the amount of current that flows out of the bulb compare to the amount that flows into the bulb? (i) Current out is less than current in; (ii) current out is greater than current in; (iii) current out equals current in; (iv) the answer depends on the brightness of the bulb.

25 CURRENT, RESISTANCE, AND ELECTROMOTIVE FORCE

LEARNING GOALS

Looking forward at ...

25.1 The meaning of electric current, and how charges move in a conductor.

25.2 What is meant by the resistivity and conductivity of a substance.

25.3 How to calculate the resistance of a conductor from its dimensions and its resistivity.

25.4 How an electromotive force (emf) makes it possible for current to flow in a circuit.

25.5 How to do calculations involving energy and power in circuits.

25.6 How to use a simple model to understand the flow of current in metals.

Looking back at ...

17.7 Thermal conductivity.

23.2 Voltmeters, electric field, and electric potential.

24.4 Dielectric breakdown in insulators.

In the past four chapters we studied the interactions of electric charges *at rest;* now we're ready to study charges *in motion*. An *electric current* consists of charges in motion from one region to another. If the charges follow a conducting path that forms a closed loop, the path is called an *electric circuit*.

Fundamentally, electric circuits are a means for conveying *energy* from one place to another. As charged particles move within a circuit, electric potential energy is transferred from a source (such as a battery or generator) to a device in which that energy is either stored or converted to another form: into sound in a stereo system or into heat and light in a toaster or light bulb. Electric circuits are useful because they allow energy to be transported without any moving parts (other than the moving charged particles themselves). They are at the heart of computers, television transmitters and receivers, and household and industrial power distribution systems. Your nervous system is a specialized electric circuit that carries vital signals from one part of your body to another.

In Chapter 26 we will see how to analyze electric circuits and will examine some practical applications of circuits. To prepare you for that, in this chapter we'll examine the basic properties of electric currents. We'll begin by describing the nature of electric conductors and considering how they are affected by temperature. We'll learn why a short, fat, cold copper wire is a better conductor than a long, skinny, hot steel wire. We'll study the properties of batteries and see how they cause current and energy transfer in a circuit. In this analysis we will use the concepts of current, potential difference (or voltage), resistance, and electromotive force. Finally, we'll look at electric current in a material from a microscopic viewpoint.

25.1 CURRENT

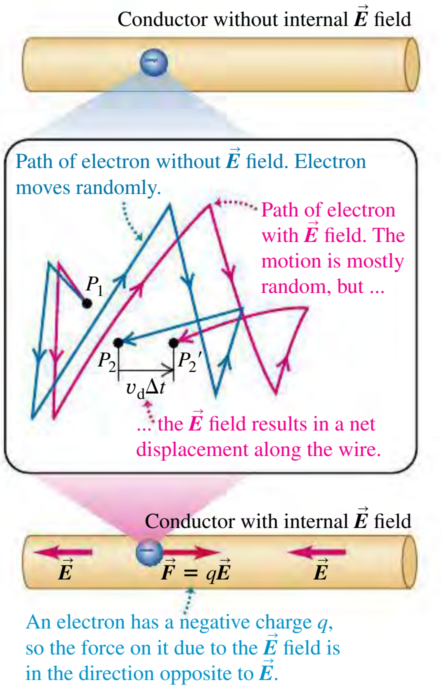
A **current** is any motion of charge from one region to another. In this section we'll discuss currents in conducting materials. The vast majority of technological applications of charges in motion involve currents of this kind.

In electrostatic situations (discussed in Chapters 21 through 24) the electric field is zero everywhere within the conductor, and there is *no* current. However, this does not mean that all charges within the conductor are at rest. In an ordinary metal such as copper or aluminum, some of the electrons are free to move within the conducting material. These free electrons move randomly in all directions, somewhat like the molecules of a gas but with much greater speeds, of the order of 10^6 m/s. The electrons nonetheless do not escape from the conducting material, because they are attracted to the positive ions of the material. The motion of the electrons is random, so there is no *net* flow of charge in any direction and hence no current.

Now consider what happens if a constant, steady electric field $\vec{E}$ is established inside a conductor. (We'll see later how this can be done.) A charged particle (such as a free electron) inside the conducting material is then subjected to a steady force $\vec{F} = q\vec{E}$. If the charged particle were moving in *vacuum,* this steady force would cause a steady acceleration in the direction of $\vec{F}$, and after a time the charged particle would be moving in that direction at high speed. But a charged particle moving in a *conductor* undergoes frequent collisions with the massive, nearly stationary ions of the material. In each such collision the particle's direction of motion undergoes a random change. The net effect of the electric field $\vec{E}$ is that in addition to the random motion of the charged particles within the conductor, there is also a very slow net motion or *drift* of the moving charged particles as a group in the direction of the electric force $\vec{F} = q\vec{E}$ (**Fig. 25.1**). This motion is described in terms of the **drift velocity** $\vec{v}_d$ of the particles. As a result, there is a net current in the conductor.

While the random motion of the electrons has a very fast average speed of about 10^6 m/s, the drift speed is very slow, often on the order of 10^{-4} m/s. Given that the electrons move so slowly, you may wonder why the light comes on immediately when you turn on the switch of a flashlight. The reason is that the electric field is set up in the wire with a speed approaching the speed of light, and electrons start to move all along the wire at very nearly the same time. The time that it takes any individual electron to get from the switch to the light bulb isn't really relevant. A good analogy is a group of soldiers standing at attention when the sergeant orders them to start marching; the order reaches the soldiers' ears at the speed of sound, which is much faster than their marching speed, so all the soldiers start to march essentially in unison.

25.1 If there is no electric field inside a conductor, an electron moves randomly from point P_1 to point P_2 in a time Δt. If an electric field $\vec{E}$ is present, the electric force $\vec{F} = q\vec{E}$ imposes a small drift (greatly exaggerated here) that takes the electron to point P'_2, a distance $v_d \Delta t$ from P_2 in the direction of the force.



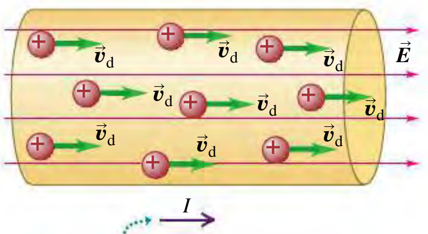
The Direction of Current Flow

The drift of moving charges through a conductor can be interpreted in terms of work and energy. The electric field $\vec{E}$ does work on the moving charges. The resulting kinetic energy is transferred to the material of the conductor by means of collisions with the ions, which vibrate about their equilibrium positions in the crystalline structure of the conductor. This energy transfer increases the average vibrational energy of the ions and therefore the temperature of the material. Thus much of the work done by the electric field goes into heating the conductor, *not* into making the moving charges move ever faster and faster. This heating is sometimes useful, as in an electric toaster, but in many situations is simply an unavoidable by-product of current flow.

In different current-carrying materials, the charges of the moving particles may be positive or negative. In metals the moving charges are always (negative) electrons, while in an ionized gas (plasma) or an ionic solution the moving charges may include both electrons and positively charged ions. In a semiconductor

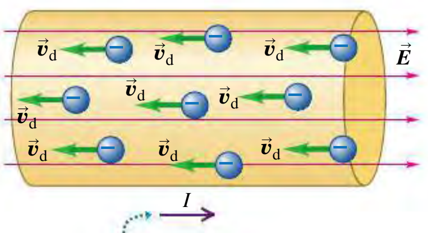
25.2 The same current is produced by (a) positive charges moving in the direction of the electric field $\vec{E}$ or (b) the same number of negative charges moving at the same speed in the direction opposite to $\vec{E}$.

(a)



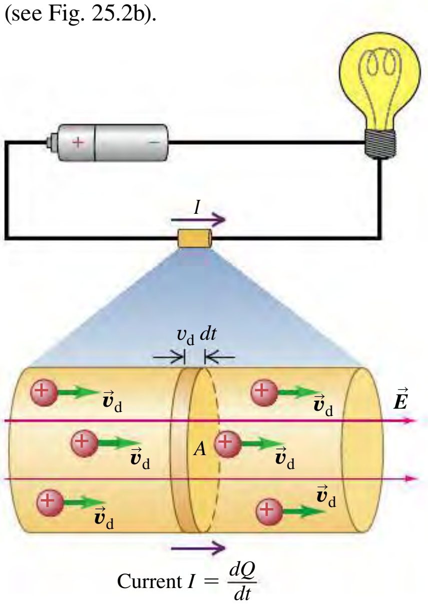
A **conventional current** is treated as a flow of positive charges, regardless of whether the free charges in the conductor are positive, negative, or both.

(b)



In a metallic conductor, the moving charges are electrons — but the *current* still points in the direction positive charges would flow.

25.3 The current I is the time rate of charge transfer through the cross-sectional area A. The random component of each moving charged particle's motion averages to zero, and the current is in the same direction as $\vec{E}$ whether the moving charges are positive (as shown here) or negative (see Fig. 25.2b).



material such as germanium or silicon, conduction is partly by electrons and partly by motion of *vacancies,* also known as *holes;* these are sites of missing electrons and act like positive charges.

Figure 25.2 shows segments of two different current-carrying materials. In Fig. 25.2a the moving charges are positive, the electric force is in the same direction as $\vec{E}$, and the drift velocity $\vec{v}_d$ is from left to right. In Fig. 25.2b the charges are negative, the electric force is opposite to $\vec{E}$, and the drift velocity $\vec{v}_d$ is from right to left. In both cases there is a net flow of positive charge from left to right, and positive charges end up to the right of negative ones. We *define* the current, denoted by I, to be in the direction in which there is a flow of *positive* charge. Thus we describe currents as though they consisted entirely of positive charge flow, even in cases in which we know that the actual current is due to electrons. Hence the current is to the right in both Figs. 25.2a and 25.2b. This choice or convention for the direction of current flow is called **conventional current.** While the direction of the conventional current is *not* necessarily the same as the direction in which charged particles are actually moving, we'll find that the sign of the moving charges is of little importance in analyzing electric circuits.

Figure 25.3 shows a segment of a conductor in which a current is flowing. We consider the moving charges to be *positive,* so they are moving in the same direction as the current. We define the current through the cross-sectional area A to be *the net charge flowing through the area per unit time.* Thus, if a net charge dQ flows through an area in a time dt, the current I through the area is

$$I = \frac{dQ}{dt} \qquad \text{(definition of current)} \tag{25.1}$$

CAUTION Current is not a vector Although we refer to the *direction* of a current, current as defined by Eq. (25.1) is *not* a vector quantity. In a current-carrying wire, the current is always along the length of the wire, regardless of whether the wire is straight or curved. No single vector could describe motion along a curved path. We'll usually describe the direction of current either in words (as in "the current flows clockwise around the circuit") or by choosing a current to be positive if it flows in one direction along a conductor and negative if it flows in the other direction.

The SI unit of current is the **ampere;** one ampere is defined to be *one coulomb per second* $(1\text{ A} = 1\text{ C/s})$. This unit is named in honor of the French scientist André Marie Ampère (1775–1836). When an ordinary flashlight (D-cell size) is turned on, the current in the flashlight is about 0.5 to 1 A; the current in the wires of a car engine's starter motor is around 200 A. Currents in radio and television circuits are usually expressed in *milliamperes* $(1\text{ mA} = 10^{-3}\text{ A})$ or *microamperes* $(1\ \mu\text{A} = 10^{-6}\text{ A})$, and currents in computer circuits are expressed in *nanoamperes* $(1\text{ nA} = 10^{-9}\text{ A})$ or *picoamperes* $(1\text{ pA} = 10^{-12}\text{ A})$.

Current, Drift Velocity, and Current Density

We can express current in terms of the drift velocity of the moving charges. Let's reconsider the situation of Fig. 25.3 of a conductor with cross-sectional area A and an electric field $\vec{E}$ directed from left to right. To begin with, we'll assume that the free charges in the conductor are positive; then the drift velocity is in the same direction as the field.

Suppose there are n moving charged particles per unit volume. We call n the **concentration** of particles; its SI unit is m^{-3}. Assume that all the particles move with the same drift velocity with magnitude v_d. In a time interval dt, each particle moves a distance $v_d\,dt$. The particles that flow out of the right end of the shaded cylinder with length $v_d\,dt$ during dt are the particles that were within this cylinder at the beginning of the interval dt. The volume of the cylinder is $Av_d\,dt$, and the number of particles within it is $nAv_{\text{d}}\,dt$. If each particle has a charge q, the charge dQ that flows out of the end of the cylinder during time dt is

$$dQ = q(nAv_{\text{d}}\,dt) = nqv_{\text{d}}A\,dt$$

and the current is

$$I = \frac{dQ}{dt} = nqv_{\text{d}}A$$

The current *per unit cross-sectional area* is called the **current density** J:

$$J = \frac{I}{A} = nqv_{\text{d}}$$

The units of current density are amperes per square meter (A/m^2).

If the moving charges are negative rather than positive, as in Fig. 25.2b, the drift velocity is opposite to $\vec{E}$. But the *current* is still in the same direction as $\vec{E}$ at each point in the conductor. Hence current I and current density J don't depend on the sign of the charge, and so we replace the charge q by its absolute value $|q|$:

Rate at which charge flows through area

Current through an area

Drift speed

$$I = \frac{dQ}{dt} = n|q|v_{\text{d}}A \qquad (25.2)$$

Cross-sectional area

Concentration of moving charged particles

Charge per particle

$$J = \frac{I}{A} = n|q|v_{\text{d}} \qquad \text{(current density)} \qquad (25.3)$$

The current in a conductor is the product of the concentration of moving charged particles, the magnitude of charge of each such particle, the magnitude of the drift velocity, and the cross-sectional area of the conductor.

We can also define a *vector* current density $\vec{J}$ that includes the direction of the drift velocity:

Vector current density

Drift velocity

$$\vec{J} = nq\vec{v}_{\text{d}} \qquad (25.4)$$

Concentration of moving charged particles

Charge per particle

There are *no* absolute value signs in Eq. (25.4). If q is positive, $\vec{v}_{\text{d}}$ is in the same direction as $\vec{E}$; if q is negative, $\vec{v}_{\text{d}}$ is opposite to $\vec{E}$. In either case, $\vec{J}$ is in the same direction as $\vec{E}$. Equation (25.3) gives the *magnitude* J of the vector current density $\vec{J}$.

CAUTION **Current density vs. current** Current density $\vec{J}$ is a vector, but current I is not. The difference is that the current density $\vec{J}$ describes how charges flow at a certain point, and the vector's direction tells you about the direction of the flow at that point. By contrast, the current I describes how charges flow through an extended object such as a wire. For example, I has the same value at all points in the circuit of Fig. 25.3, but $\vec{J}$ does not: $\vec{J}$ is directed downward in the left-hand side of the loop and upward in the right-hand side. The magnitude of $\vec{J}$ can also vary around a circuit. In Fig. 25.3 $J = I/A$ is less in the battery (which has a large cross-sectional area A) than in the wires (which have a small A).

In general, a conductor may contain several different kinds of moving charged particles having charges $q_1, q_2, \ldots$, concentrations $n_1, n_2, \ldots$, and drift velocities with magnitudes $v_{\text{d}1}, v_{\text{d}2}, \ldots$. An example is current flow in an ionic solution (**Fig. 25.4**). In a sodium chloride solution, current can be carried by both positive sodium ions and negative chlorine ions; the total current I is found by adding up the currents due to each kind of charged particle, from Eq. (25.2). Likewise, the total vector current density $\vec{J}$ is found by using Eq. (25.4) for each kind of charged particle and adding the results.

25.4 Part of the electric circuit that includes this light bulb passes through a beaker with a solution of sodium chloride. The current in the solution is carried by both positive charges (Na^+ ions) and negative charges (Cl^- ions).

We will see in Section 25.4 that it is possible to have a current that is *steady* (constant in time) only if the conducting material forms a closed loop, called a *complete circuit*. In such a steady situation, the total charge in every segment of the conductor is constant. Hence the rate of flow of charge *out* at one end of a segment at any instant equals the rate of flow of charge *in* at the other end of the segment, and *the current is the same at all cross sections of the circuit*. We'll use this observation when we analyze electric circuits later in this chapter.

In many simple circuits, such as flashlights or cordless electric drills, the direction of the current is always the same; this is called *direct current*. But home appliances such as toasters, refrigerators, and televisions use *alternating current,* in which the current continuously changes direction. In this chapter we'll consider direct current only. Alternating current has many special features worthy of detailed study, which we'll examine in Chapter 31.

EXAMPLE 25.1 CURRENT DENSITY AND DRIFT VELOCITY IN A WIRE

SOLUTION

An 18-gauge copper wire (the size usually used for lamp cords), with a diameter of 1.02 mm, carries a constant current of 1.67 A to a 200-W lamp. The free-electron density in the wire is 8.5×10^{28} per cubic meter. Find (a) the current density and (b) the drift speed.

SOLUTION

IDENTIFY and SET UP: This problem uses the relationships among current I, current density J, and drift speed v_d. We are given I and the wire diameter d, so we use Eq. (25.3) to find J. We use Eq. (25.3) again to find v_d from J and the known electron density n.

EXECUTE: (a) The cross-sectional area is

$$A = \frac{\pi d^2}{4} = \frac{\pi(1.02 \times 10^{-3}\ \text{m})^2}{4} = 8.17 \times 10^{-7}\ \text{m}^2$$

The magnitude of the current density is then

$$J = \frac{I}{A} = \frac{1.67\ \text{A}}{8.17 \times 10^{-7}\ \text{m}^2} = 2.04 \times 10^6\ \text{A/m}^2$$

(b) From Eq. (25.3) for the drift velocity magnitude v_d, we find

$$v_d = \frac{J}{n|q|} = \frac{2.04 \times 10^6\ \text{A/m}^2}{(8.5 \times 10^{28}\ \text{m}^{-3})|-1.60 \times 10^{-19}\ \text{C}|}$$
$$= 1.5 \times 10^{-4}\ \text{m/s} = 0.15\ \text{mm/s}$$

EVALUATE: At this speed an electron would require 6700 s (almost 2 h) to travel 1 m along this wire. The speeds of random motion of the electrons are roughly 10^6 m/s, around 10^{10} times the drift speed. Picture the electrons as bouncing around frantically, with a very slow drift!

TEST YOUR UNDERSTANDING OF SECTION 25.1 Suppose we replaced the wire in Example 25.1 with 12-gauge copper wire, which has twice the diameter of 18-gauge wire. If the current remains the same, what effect would this have on the magnitude of the drift velocity v_d? (i) None—v_d would be unchanged; (ii) v_d would be twice as great; (iii) v_d would be four times greater; (iv) v_d would be half as great; (v) v_d would be one-fourth as great. ▌

25.2 RESISTIVITY

The current density $\vec{J}$ in a conductor depends on the electric field $\vec{E}$ and on the properties of the material. In general, this dependence can be quite complex. But for some materials, especially metals, at a given temperature, $\vec{J}$ is nearly *directly proportional* to $\vec{E}$, and the ratio of the magnitudes of E and J is constant. This relationship, called Ohm's law, was discovered in 1826 by the German physicist Georg Simon Ohm (1787–1854). The word "law" should actually be in quotation marks, since **Ohm's law,** like the ideal-gas equation and Hooke's law, is an *idealized model* that describes the behavior of some materials quite well but is not a general description of *all* matter. In the following discussion we'll assume that Ohm's law is valid, even though there are many situations in which it is not.

TABLE 25.1 Resistivities at Room Temperature (20°C)

	Substance	$\rho\ (\Omega \cdot \mathrm{m})$	Substance	$\rho\ (\Omega \cdot \mathrm{m})$
Conductors			**Semiconductors**	
Metals	Silver	1.47×10^{-8}	Pure carbon (graphite)	3.5×10^{-5}
	Copper	1.72×10^{-8}	Pure germanium	0.60
	Gold	2.44×10^{-8}	Pure silicon	2300
	Aluminum	2.75×10^{-8}	**Insulators**	
	Tungsten	5.25×10^{-8}	Amber	5×10^{14}
	Steel	20×10^{-8}	Glass	10^{10}–10^{14}
	Lead	22×10^{-8}	Lucite	$>10^{13}$
	Mercury	95×10^{-8}	Mica	10^{11}–10^{15}
Alloys	Manganin (Cu 84%, Mn 12%, Ni 4%)	44×10^{-8}	Quartz (fused)	75×10^{16}
	Constantan (Cu 60%, Ni 40%)	49×10^{-8}	Sulfur	10^{15}
	Nichrome	100×10^{-8}	Teflon	$>10^{13}$
			Wood	10^{8}–10^{11}

We define the **resistivity** ρ of a material as

$$\underbrace{\rho}_{\text{Resistivity of a material}} = \frac{\overbrace{E}^{\text{Magnitude of electric field in material}}}{\underbrace{J}_{\text{Magnitude of current density caused by electric field}}} \tag{25.5}$$

The greater the resistivity, the greater the field needed to cause a given current density, or the smaller the current density caused by a given field. From Eq. (25.5) the units of ρ are $(\mathrm{V/m})/(\mathrm{A/m^2}) = \mathrm{V \cdot m/A}$. As we will discuss in Section 25.3, 1 V/A is called one *ohm* (1 Ω; the Greek letter Ω, omega, is alliterative with "ohm"). So the SI units for ρ are $\Omega \cdot \mathrm{m}$ (ohm-meters). **Table 25.1** lists some representative values of resistivity. A perfect conductor would have zero resistivity, and a perfect insulator would have an infinite resistivity. Metals and alloys have the smallest resistivities and are the best conductors. The resistivities of insulators are greater than those of the metals by an enormous factor, on the order of 10^{22}.

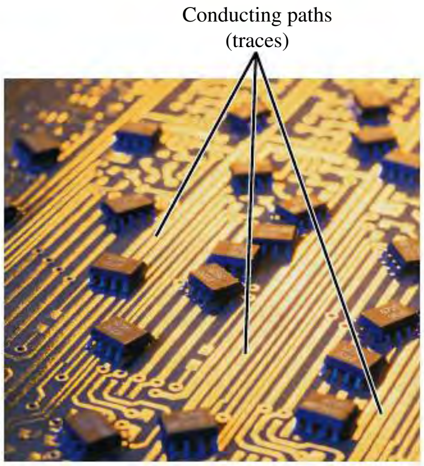
The reciprocal of resistivity is **conductivity.** Its units are $(\Omega \cdot \mathrm{m})^{-1}$. Good conductors of electricity have larger conductivity than insulators. Conductivity is the direct electrical analog of thermal conductivity. Comparing Table 25.1 with Table 17.5 (Thermal Conductivities), we note that good electrical conductors, such as metals, are usually also good conductors of heat. Poor electrical conductors, such as ceramic and plastic materials, are also poor thermal conductors. In a metal the free electrons that carry charge in electrical conduction also provide the principal mechanism for heat conduction, so we should expect a correlation between electrical and thermal conductivity. Because of the enormous difference in conductivity between electrical conductors and insulators, it is easy to confine electric currents to well-defined paths or circuits (**Fig. 25.5**). The variation in *thermal* conductivity is much less, only a factor of 10^3 or so, and it is usually impossible to confine heat currents to that extent.

Semiconductors have resistivities intermediate between those of metals and those of insulators. These materials are important because of the way their resistivities are affected by temperature and by small amounts of impurities.

A material that obeys Ohm's law reasonably well is called an *ohmic* conductor or a *linear* conductor. For such materials, at a given temperature, ρ is a *constant* that does not depend on the value of E. Many materials show substantial departures from Ohm's-law behavior; they are *nonohmic,* or *nonlinear.* In these materials, J depends on E in a more complicated manner.

Analogies with fluid flow can be a big help in developing intuition about electric current and circuits. For example, in the making of wine or maple syrup, the product is sometimes filtered to remove sediments. A pump forces the fluid

25.5 The copper "wires," or traces, on this circuit board are printed directly onto the surface of the dark-colored insulating board. Even though the traces are very close to each other (only about a millimeter apart), the board has such a high resistivity (and low conductivity) that essentially no current can flow between the traces.



BIO Application Resistivity and Nerve Conduction This false-color image from an electron microscope shows a cross section through a nerve fiber about 1 μm (10^{-6} m) in diameter. A layer of an insulating fatty substance called myelin is wrapped around the conductive material of the axon. The resistivity of myelin is much greater than that of the axon, so an electric signal traveling along the nerve fiber remains confined to the axon. This makes it possible for a signal to travel much more rapidly than if the myelin were absent.

through the filter under pressure; if the flow rate (analogous to J) is proportional to the pressure difference between the upstream and downstream sides (analogous to E), the behavior is analogous to Ohm's law.

Resistivity and Temperature

The resistivity of a *metallic* conductor nearly always increases with increasing temperature, as shown in **Fig. 25.6a**. As temperature increases, the ions of the conductor vibrate with greater amplitude, making it more likely that a moving electron will collide with an ion as in Fig. 25.1; this impedes the drift of electrons through the conductor and hence reduces the current. Over a small temperature range (up to 100 C° or so), the resistivity of a metal can be represented approximately by the equation

Temperature dependence of resistivity:

$$\underbrace{\rho(T)}_{\text{Resistivity at temperature } T} = \underbrace{\rho_0}_{\text{Resistivity at reference temperature } T_0}[1 + \underbrace{\alpha}_{\text{Temperature coefficient of resistivity}}(T - T_0)] \quad (25.6)$$

The reference temperature T_0 is often taken as 0°C or 20°C; the temperature T may be higher or lower than T_0. The factor α is called the **temperature coefficient of resistivity.** Some representative values are given in **Table 25.2**. The resistivity of the alloy manganin is practically independent of temperature.

The resistivity of graphite (a nonmetal) *decreases* with increasing temperature, since at higher temperatures, more electrons "shake loose" from the atoms and become mobile; hence the temperature coefficient of resistivity of graphite is negative. The same behavior occurs for semiconductors (Fig. 25.6b). Measuring the resistivity of a small semiconductor crystal is therefore a sensitive measure of temperature; this is the principle of a type of thermometer called a *thermistor*.

Some materials, including several metallic alloys and oxides, show a phenomenon called *superconductivity*. As the temperature decreases, the resistivity at first decreases smoothly, like that of any metal. But then at a certain critical temperature T_c a phase transition occurs and the resistivity suddenly drops to zero, as shown in Fig. 25.6c. Once a current has been established in a superconducting ring, it continues indefinitely without the presence of any driving field.

Superconductivity was discovered in 1911 by the Dutch physicist Heike Kamerlingh Onnes (1853–1926). He discovered that at very low temperatures, below 4.2 K, the resistivity of mercury suddenly dropped to zero. For the next 75 years, the highest T_c attained was about 20 K. This meant that superconductivity occurred only when the material was cooled by using expensive liquid helium, with a boiling-point temperature of 4.2 K, or explosive liquid hydrogen, with a boiling point of 20.3 K. But in 1986 Karl Müller and Johannes Bednorz discovered an oxide of barium, lanthanum, and copper with a T_c of nearly 40 K, and the race was on to develop "high-temperature" superconducting materials.

By 1987 a complex oxide of yttrium, copper, and barium had been found that has a value of T_c well above the 77 K boiling temperature of liquid nitrogen, a refrigerant that is both inexpensive and safe. The current (2014) record for T_c

25.6 Variation of resistivity ρ with absolute temperature T for **(a)** a normal metal, **(b)** a semiconductor, and **(c)** a superconductor. In (a) the linear approximation to ρ as a function of T is shown as a green line; the approximation agrees exactly at $T = T_0$, where $\rho = \rho_0$.

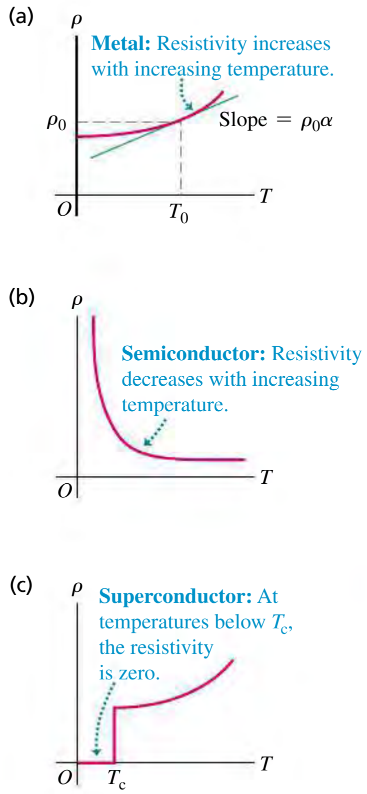


TABLE 25.2 Temperature Coefficients of Resistivity (Approximate Values Near Room Temperature)

Material	α [(°C)$^{-1}$]	Material	α [(°C)$^{-1}$]
Aluminum	0.0039	Lead	0.0043
Brass	0.0020	Manganin	0.00000
Carbon (graphite)	−0.0005	Mercury	0.00088
Constantan	0.00001	Nichrome	0.0004
Copper	0.00393	Silver	0.0038
Iron	0.0050	Tungsten	0.0045

at atmospheric pressure is 138 K, and materials that are superconductors at room temperature may become a reality. The implications of these discoveries for power-distribution systems, computer design, and transportation are enormous. Meanwhile, superconducting electromagnets cooled by liquid helium are used in particle accelerators and some experimental magnetic-levitation railroads. Superconductors have other exotic properties that require an understanding of magnetism to explore; we will discuss these further in Chapter 29.

TEST YOUR UNDERSTANDING OF SECTION 25.2 You maintain a constant electric field inside a piece of semiconductor while lowering the semiconductor's temperature. What happens to the current density in the semiconductor? (i) It increases; (ii) it decreases; (iii) it remains the same.

25.3 RESISTANCE

For a conductor with resistivity ρ, the current density $\vec{J}$ at a point where the electric field is $\vec{E}$ is given by Eq. (25.5), which we can write as

$$\vec{E} = \rho\vec{J} \tag{25.7}$$

When Ohm's law is obeyed, ρ is constant and independent of the magnitude of the electric field, so $\vec{E}$ is directly proportional to $\vec{J}$. Often, however, we are more interested in the total current I in a conductor than in $\vec{J}$ and more interested in the potential difference V between the ends of the conductor than in $\vec{E}$. This is so largely because I and V are much easier to measure than are $\vec{J}$ and $\vec{E}$.

Suppose our conductor is a wire with uniform cross-sectional area A and length L, as shown in **Fig. 25.7**. Let V be the potential difference between the higher-potential and lower-potential ends of the conductor, so that V is positive. (Another name for V is the *voltage across* the conductor.) The *direction* of the current is always from the higher-potential end to the lower-potential end. That's because current in a conductor flows in the direction of $\vec{E}$, no matter what the sign of the moving charges (Fig. 25.2), and because $\vec{E}$ points in the direction of *decreasing* electric potential (see Section 23.2). As the current flows through the potential difference, electric potential energy is lost; this energy is transferred to the ions of the conducting material during collisions.

We can also relate the *value* of the current I to the potential difference between the ends of the conductor. If the magnitudes of the current density $\vec{J}$ and the electric field $\vec{E}$ are uniform throughout the conductor, the total current I is $I = JA$, and the potential difference V between the ends is $V = EL$. We solve these equations for J and E, respectively, and substitute the results into Eq. (25.7):

$$\frac{V}{L} = \frac{\rho I}{A} \quad \text{or} \quad V = \frac{\rho L}{A} I \tag{25.8}$$

This shows that when ρ is constant, the total current I is proportional to the potential difference V.

The ratio of V to I for a particular conductor is called its **resistance** R:

$$R = \frac{V}{I} \tag{25.9}$$

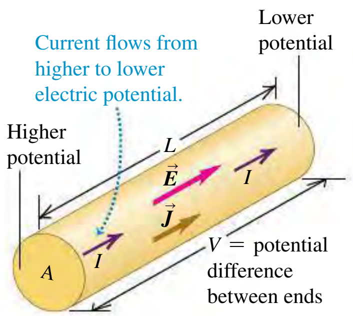
Comparing this definition of R to Eq. (25.8), we see that

Resistance of a conductor
$$R = \frac{\rho L}{A} \tag{25.10}$$
ρ: Resistivity of conductor material; L: Length of conductor; A: Cross-sectional area of conductor

If ρ is constant, as is the case for ohmic materials, then so is R.

PhET: Resistance in a Wire

25.7 A conductor with uniform cross section. The current density is uniform over any cross section, and the electric field is constant along the length.



The following equation is often called Ohm's law:

Relationship among voltage, current, and resistance:

$$V = IR \tag{25.11}$$

Voltage between ends of conductor
Resistance of conductor
Current in conductor

However, it's important to understand that the real content of Ohm's law is the direct proportionality (for some materials) of V to I or of J to E. Equation (25.9) or (25.11) *defines* resistance R for *any* conductor, but only when R is constant can we correctly call this relationship Ohm's law.

Interpreting Resistance

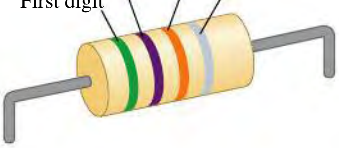
Equation (25.10) shows that the resistance of a wire or other conductor of uniform cross section is directly proportional to its length and inversely proportional to its cross-sectional area. It is also proportional to the resistivity of the material of which the conductor is made.

25.8 A long fire hose offers substantial resistance to water flow. To make water pass through the hose rapidly, the upstream end of the hose must be at much higher pressure than the end where the water emerges. In an analogous way, there must be a large potential difference between the ends of a long wire in order to cause a substantial electric current through the wire.

The flowing-fluid analogy is again useful. In analogy to Eq. (25.10), a narrow water hose offers more resistance to flow than a fat one, and a long hose has more resistance than a short one (**Fig. 25.8**). We can increase the resistance to flow by stuffing the hose with cotton or sand; this corresponds to increasing the resistivity. The flow rate is approximately proportional to the pressure difference between the ends. Flow rate is analogous to current, and pressure difference is analogous to potential difference (voltage). Let's not stretch this analogy too far, though; the water flow rate in a pipe is usually *not* proportional to its cross-sectional area (see Section 14.6).

The SI unit of resistance is the **ohm,** equal to one volt per ampere $(1\ \Omega = 1\ \mathrm{V/A})$. The *kilohm* $(1\ \mathrm{k}\Omega = 10^3\ \Omega)$ and the *megohm* $(1\ \mathrm{M}\Omega = 10^6\ \Omega)$ are also in common use. A 100-m length of 12-gauge copper wire, the size usually used in household wiring, has a resistance at room temperature of about $0.5\ \Omega$. A 100-W, 120-V incandescent light bulb has a resistance (at operating temperature) of $140\ \Omega$. If the same current I flows in both the copper wire and the light bulb, the potential difference $V = IR$ is much greater across the light bulb, and much more potential energy is lost per charge in the light bulb. This lost energy is converted by the light bulb filament into light and heat. You don't want your household wiring to glow white-hot, so its resistance is kept low by using wire of low resistivity and large cross-sectional area.

25.9 This resistor has a resistance of $5.7\ \mathrm{k}\Omega$ with an accuracy (tolerance) of $\pm 10\%$.



Because the resistivity of a material varies with temperature, the resistance of a specific conductor also varies with temperature. For temperature ranges that are not too great, this variation is approximately linear, analogous to Eq. (25.6):

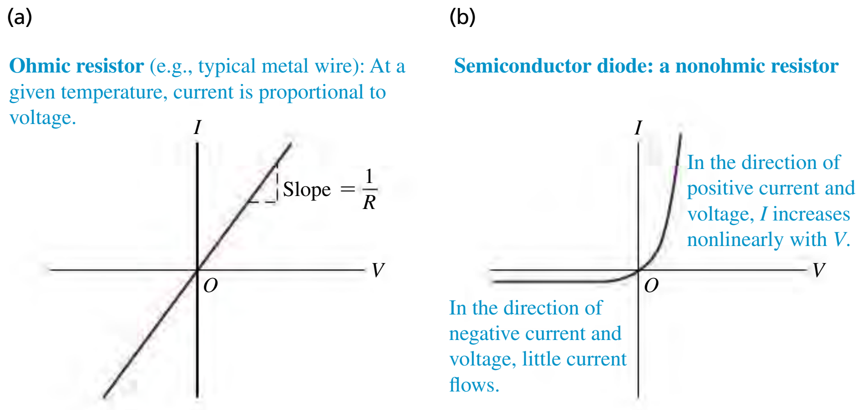
$$R(T) = R_0[1 + \alpha(T - T_0)] \tag{25.12}$$

In this equation, $R(T)$ is the resistance at temperature T and R_0 is the resistance at temperature T_0, often taken to be 0°C or 20°C. The *temperature coefficient of resistance* α is the same constant that appears in Eq. (25.6) if the dimensions L and A in Eq. (25.10) do not change appreciably with temperature; this is indeed the case for most conducting materials. Within the limits of validity of Eq. (25.12), the *change* in resistance resulting from a temperature change $T - T_0$ is given by $R_0\alpha(T - T_0)$.

TABLE 25.3 Color Codes for Resistors

Color	Value as Digit	Value as Multiplier
Black	0	1
Brown	1	10
Red	2	10^2
Orange	3	10^3
Yellow	4	10^4
Green	5	10^5
Blue	6	10^6
Violet	7	10^7
Gray	8	10^8
White	9	10^9

A circuit device made to have a specific value of resistance between its ends is called a **resistor**. Resistors in the range 0.01 to $10^7\ \Omega$ can be bought off the shelf. Individual resistors used in electronic circuitry are often cylindrical, a few millimeters in diameter and length, with wires coming out of the ends. The resistance may be marked with a standard code that uses three or four color bands near one end (**Fig. 25.9**), according to the scheme in **Table 25.3**. The first two bands (starting with the band nearest an end) are digits, and the third is a power-of-10 multiplier. For example, green–violet–red means $57 \times 10^2\ \Omega$, or $5.7\ \mathrm{k}\Omega$. The fourth band, if present, indicates the accuracy (tolerance) of the value;



25.10 Current–voltage relationships for two devices. Only for a resistor that obeys Ohm's law as in (a) is current I proportional to voltage V.

no band means $\pm 20\%$, a silver band $\pm 10\%$, and a gold band $\pm 5\%$. Another important characteristic of a resistor is the maximum *power* it can dissipate without damage. We'll return to this point in Section 25.5.

For a resistor that obeys Ohm's law, a graph of current as a function of potential difference (voltage) is a straight line (**Fig. 25.10a**). The slope of the line is $1/R$. If the sign of the potential difference changes, so does the sign of the current produced; in Fig. 25.7 this corresponds to interchanging the higher- and lower-potential ends of the conductor, so the electric field, current density, and current all reverse direction.

In devices that do not obey Ohm's law, the relationship of voltage to current may not be a direct proportion, and it may be different for the two directions of current. Figure 25.10b shows the behavior of a semiconductor *diode*, a device used to convert alternating current to direct current and to perform a wide variety of logic functions in computer circuitry. For positive potentials V of the anode (one of two terminals of the diode) with respect to the cathode (the other terminal), I increases exponentially with increasing V; for negative potentials the current is extremely small. Thus a positive V causes a current to flow in the positive direction, but a potential difference of the other sign causes little or no current. Hence a diode acts like a one-way valve in a circuit.

EXAMPLE 25.2 ELECTRIC FIELD, POTENTIAL DIFFERENCE, AND RESISTANCE IN A WIRE

The 18-gauge copper wire of Example 25.1 has a cross-sectional area of $8.20 \times 10^{-7}\ \text{m}^2$. It carries a current of 1.67 A. Find (a) the electric-field magnitude in the wire; (b) the potential difference between two points in the wire 50.0 m apart; (c) the resistance of a 50.0-m length of this wire.

SOLUTION

IDENTIFY and SET UP: We are given the cross-sectional area A and current I. Our target variables are the electric-field magnitude E, potential difference V, and resistance R. The current density is $J = I/A$. We find E from Eq. (25.5), $E = \rho J$ (Table 25.1 gives the resistivity ρ for copper). The potential difference is then the product of E and the length of the wire. We can use either Eq. (25.10) or Eq. (25.11) to find R.

EXECUTE: (a) From Table 25.1, $\rho = 1.72 \times 10^{-8}\ \Omega \cdot \text{m}$. Hence, from Eq. (25.5),

$$E = \rho J = \frac{\rho I}{A} = \frac{(1.72 \times 10^{-8}\ \Omega \cdot \text{m})(1.67\ \text{A})}{8.20 \times 10^{-7}\ \text{m}^2} = 0.0350\ \text{V/m}$$

(b) The potential difference is

$$V = EL = (0.0350\ \text{V/m})(50.0\ \text{m}) = 1.75\ \text{V}$$

(c) From Eq. (25.10) the resistance of 50.0 m of this wire is

$$R = \frac{\rho L}{A} = \frac{(1.72 \times 10^{-8}\ \Omega \cdot \text{m})(50.0\ \text{m})}{8.20 \times 10^{-7}\ \text{m}^2} = 1.05\ \Omega$$

Alternatively, we can find R from Eq. (25.11):

$$R = \frac{V}{I} = \frac{1.75\ \text{V}}{1.67\ \text{A}} = 1.05\ \Omega$$

EVALUATE: We emphasize that the resistance of the wire is *defined* to be the ratio of voltage to current. If the wire is made of nonohmic material, then R is different for different values of V but is always given by $R = V/I$. Resistance is also always given by $R = \rho L/A$; if the material is nonohmic, ρ is not constant but depends on E (or, equivalently, on $V = EL$).

EXAMPLE 25.3 TEMPERATURE DEPENDENCE OF RESISTANCE

Suppose the resistance of a copper wire is 1.05 Ω at 20°C. Find the resistance at 0°C and 100°C.

SOLUTION

IDENTIFY and SET UP: We are given the resistance $R_0 = 1.05\ \Omega$ at a reference temperature $T_0 = 20°\text{C}$. We use Eq. (25.12) to find the resistances at $T = 0°\text{C}$ and $T = 100°\text{C}$ (our target variables), taking the temperature coefficient of resistivity from Table 25.2.

EXECUTE: From Table 25.2, $\alpha = 0.00393\ (\text{C}°)^{-1}$ for copper. Then from Eq. (25.12),

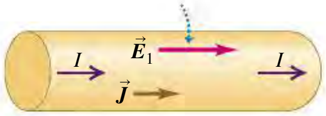
$$
\begin{aligned}
R &= R_0[1 + \alpha(T - T_0)] \\
&= (1.05\ \Omega)\{1 + [0.00393\ (\text{C}°)^{-1}][0°\text{C} - 20°\text{C}]\} \\
&= 0.97\ \Omega \text{ at } T = 0°\text{C} \\
R &= (1.05\ \Omega)\{1 + [0.00393\ (\text{C}°)^{-1}][100°\text{C} - 20°\text{C}]\} \\
&= 1.38\ \Omega \text{ at } T = 100°\text{C}
\end{aligned}
$$

EVALUATE: The resistance at 100°C is greater than that at 0°C by a factor of $(1.38\ \Omega)/(0.97\ \Omega) = 1.42$: Raising the temperature of copper wire from 0°C to 100°C increases its resistance by 42%. From Eq. (25.11), $V = IR$, this means that 42% more voltage is required to produce the same current at 100°C than at 0°C. Designers of electric circuits that must operate over a wide temperature range must take this substantial effect into account.

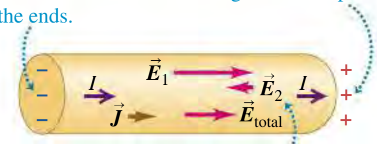
TEST YOUR UNDERSTANDING OF SECTION 25.3 Suppose you increase the voltage across the copper wire in Examples 25.2 and 25.3. The increased voltage causes more current to flow, which makes the temperature of the wire increase. (The same thing happens to the coils of an electric oven or a toaster when a voltage is applied to them. We'll explore this issue in more depth in Section 25.5.) If you double the voltage across the wire, the current in the wire increases. By what factor does it increase? (i) 2; (ii) greater than 2; (iii) less than 2. ❙

25.11 If an electric field is produced inside a conductor that is *not* part of a complete circuit, current flows for only a very short time.

(a) An electric field $\vec{E}_1$ produced inside an isolated conductor causes a current.

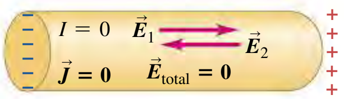


(b) The current causes charge to build up at the ends.



The charge buildup produces an opposing field $\vec{E}_2$, thus reducing the current.

(c) After a very short time $\vec{E}_2$ has the same magnitude as $\vec{E}_1$; then the total field is $\vec{E}_{\text{total}} = \mathbf{0}$ and the current stops completely.



25.4 ELECTROMOTIVE FORCE AND CIRCUITS

For a conductor to have a steady current, it must be part of a path that forms a closed loop or **complete circuit**. Here's why. If you establish an electric field $\vec{E}_1$ inside an isolated conductor with resistivity ρ that is *not* part of a complete circuit, a current begins to flow with current density $\vec{J} = \vec{E}_1/\rho$ (**Fig. 25.11a**). As a result a net positive charge quickly accumulates at one end of the conductor and a net negative charge accumulates at the other end (Fig. 25.11b). These charges themselves produce an electric field $\vec{E}_2$ in the direction opposite to $\vec{E}_1$, causing the total electric field and hence the current to decrease. Within a very small fraction of a second, enough charge builds up on the conductor ends that the total electric field $\vec{E} = \vec{E}_1 + \vec{E}_2 = \mathbf{0}$ inside the conductor. Then $\vec{J} = \mathbf{0}$ as well, and the current stops altogether (Fig. 25.11c). So there can be no steady motion of charge in such an *incomplete* circuit.

To see how to maintain a steady current in a *complete* circuit, we recall a basic fact about electric potential energy: If a charge q goes around a complete circuit and returns to its starting point, the potential energy must be the same at the end of the round trip as at the beginning. As described in Section 25.3, there is always a *decrease* in potential energy when charges move through an ordinary conducting material with resistance. So there must be some part of the circuit in which the potential energy *increases*.

The problem is analogous to an ornamental water fountain that recycles its water. The water pours out of openings at the top, cascades down over the terraces and spouts (moving in the direction of decreasing gravitational potential energy), and collects in a basin in the bottom. A pump then lifts it back to the top (increasing the potential energy) for another trip. Without the pump, the water would just fall to the bottom and stay there.

Electromotive Force

In an electric circuit there must be a device somewhere in the loop that acts like the water pump in a water fountain (**Fig. 25.12**). In this device a charge travels "uphill," from lower to higher potential energy, even though the electrostatic force is trying to push it from higher to lower potential energy. The direction of current in such a device is from lower to higher potential, just the opposite of what happens in an ordinary conductor.

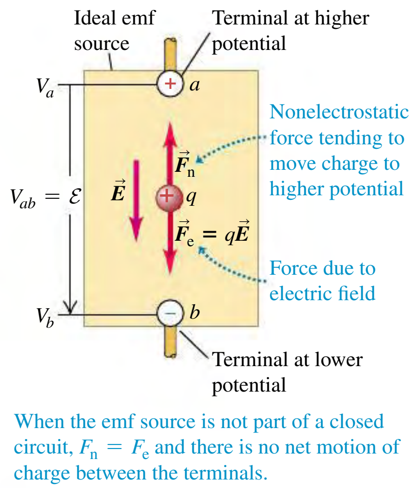
25.12 Just as a water fountain requires a pump, an electric circuit requires a source of electromotive force to sustain a steady current.

The influence that makes current flow from lower to higher potential is called **electromotive force** (abbreviated **emf** and pronounced "ee-em-eff"), and a circuit device that provides emf is called a **source of emf.** Note that "electromotive force" is a poor term because emf is *not* a force but an energy-per-unit-charge quantity, like potential. The SI unit of emf is the same as that for potential, the volt $(1\ \text{V} = 1\ \text{J/C})$. A typical flashlight battery has an emf of 1.5 V; this means that the battery does 1.5 J of work on every coulomb of charge that passes through it. We'll use the symbol $\mathcal{E}$ (a script capital E) for emf.

Every complete circuit with a steady current must include a source of emf. Batteries, electric generators, solar cells, thermocouples, and fuel cells are all examples of sources of emf. All such devices convert energy of some form (mechanical, chemical, thermal, and so on) into electric potential energy and transfer it into the circuit to which the device is connected. An *ideal* source of emf maintains a constant potential difference between its terminals, independent of the current through it. We define electromotive force quantitatively as the magnitude of this potential difference. As we will see, such an ideal source is a mythical beast, like the frictionless plane and the massless rope. We will discuss later how real-life sources of emf differ in their behavior from this idealized model.

Figure 25.13 is a schematic diagram of an ideal source of emf that maintains a potential difference between conductors a and b, called the *terminals* of the device. Terminal a, marked $+$, is maintained at *higher* potential than terminal b, marked $-$. Associated with this potential difference is an electric field $\vec{E}$ in the region around the terminals, both inside and outside the source. The electric field inside the device is directed from a to b, as shown. A charge q within the source experiences an electric force $\vec{F}_e = q\vec{E}$. But the source also provides an additional influence, which we represent as a nonelectrostatic force $\vec{F}_n$. This force, operating inside the device, pushes charge from b to a in an "uphill" direction against the electric force $\vec{F}_e$. Thus $\vec{F}_n$ maintains the potential difference between the terminals. If $\vec{F}_n$ were not present, charge would flow between the terminals until the potential difference was zero. The origin of the additional influence $\vec{F}_n$ depends on the kind of source. In a generator it results from magnetic-field forces on moving charges. In a battery or fuel cell it is associated with diffusion processes and varying electrolyte concentrations resulting from chemical reactions. In an electrostatic machine such as a Van de Graaff generator (see Fig. 22.26), an actual mechanical force is applied by a moving belt or wheel.

25.13 Schematic diagram of a source of emf in an "open-circuit" situation. The electric-field force $\vec{F}_e = q\vec{E}$ and the nonelectrostatic force $\vec{F}_n$ are shown for a positive charge q.



If a positive charge q is moved from b to a inside the source, the nonelectrostatic force $\vec{F}_n$ does a positive amount of work $W_n = q\mathcal{E}$ on the charge. This displacement is *opposite* to the electrostatic force $\vec{F}_e$, so the potential energy associated with the charge *increases* by an amount equal to qV_{ab}, where $V_{ab} = V_a - V_b$ is the (positive) potential of point a with respect to point b. For the ideal source of emf that we've described, $\vec{F}_e$ and $\vec{F}_n$ are equal in magnitude but opposite in direction, so the total work done on the charge q is zero; there is an increase in potential energy but *no* change in the kinetic energy of the charge.

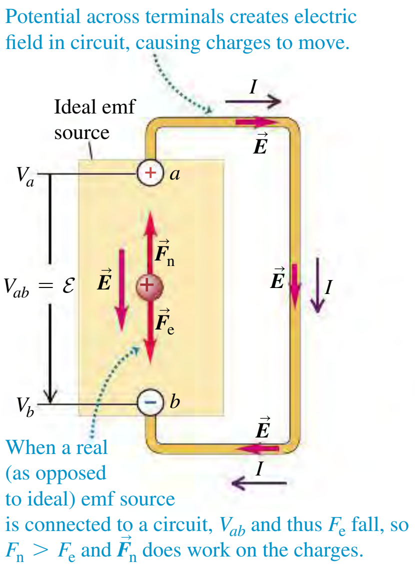
MP

PhET: Battery Voltage
PhET: Signal Circuit

It's like lifting a book from the floor to a high shelf at constant speed. The increase in potential energy is just equal to the nonelectrostatic work W_n, so $q\mathcal{E} = qV_{ab}$, or

$$V_{ab} = \mathcal{E} \qquad \text{(ideal source of emf)} \qquad (25.13)$$

25.14 Schematic diagram of an ideal source of emf in a complete circuit. The electric-field force $\vec{F}_e = q\vec{E}$ and the nonelectrostatic force $\vec{F}_n$ are shown for a positive charge q. The current is in the direction from a to b in the external circuit and from b to a within the source.



Now let's make a complete circuit by connecting a wire with resistance R to the terminals of a source (**Fig. 25.14**). The potential difference between terminals a and b sets up an electric field within the wire; this causes current to flow around the loop from a toward b, from higher to lower potential. Where the wire bends, equal amounts of positive and negative charge persist on the "inside" and "outside" of the bend. These charges exert the forces that cause the current to follow the bends in the wire.

From Eq. (25.11) the potential difference between the ends of the wire in Fig. 25.14 is given by $V_{ab} = IR$. Combining with Eq. (25.13), we have

$$\mathcal{E} = V_{ab} = IR \qquad \text{(ideal source of emf)} \qquad (25.14)$$

That is, when a positive charge q flows around the circuit, the potential *rise* $\mathcal{E}$ as it passes through the ideal source is numerically equal to the potential *drop* $V_{ab} = IR$ as it passes through the remainder of the circuit. Once $\mathcal{E}$ and R are known, this relationship determines the current in the circuit.

CAUTION **Current is not "used up" in a circuit** It's a common misconception that in a closed circuit, current squirts out of the positive terminal of a battery and is consumed or "used up" by the time it reaches the negative terminal. In fact the current is the *same* at every point in a simple loop circuit like that in Fig. 25.14, even if the wire thickness is not constant throughout the circuit. This happens because charge is conserved (it can be neither created nor destroyed) and because charge cannot accumulate in the circuit devices we have described. It's like the flow of water in an ornamental fountain; water flows out of the top at the same rate at which it reaches the bottom, no matter what the dimensions of the fountain. None of the water is "used up" along the way!

BIO Application Danger: Electric Ray! Electric rays deliver electric shocks to stun their prey and to discourage predators. (In ancient Rome, physicians practiced a primitive form of electroconvulsive therapy by placing electric rays on their patients to cure headaches and gout.) The shocks are produced by specialized flattened cells called electroplaques. Such a cell moves ions across membranes to produce an emf of about 0.05 V. Thousands of electroplaques are stacked on top of each other, so their emfs add to a total of as much as 200 V. These stacks make up more than half of an electric ray's body mass. A ray can use these to deliver an impressive current of up to 30 A for a few milliseconds.

Internal Resistance

Real sources of emf in a circuit don't behave in exactly the way we have described; the potential difference across a real source in a circuit is *not* equal to the emf as in Eq. (25.14). The reason is that charge moving through the material of any real source encounters *resistance*. We call this the **internal resistance** of the source, denoted by r. If this resistance behaves according to Ohm's law, r is constant and independent of the current I. As the current moves through r, it experiences an associated drop in potential equal to Ir. Thus, when a current is flowing through a source from the negative terminal b to the positive terminal a, the potential difference V_{ab} between the terminals is

Terminal voltage, source with internal resistance — emf of source — Current through source — Internal resistance of source

$$V_{ab} = \mathcal{E} - Ir \qquad (25.15)$$

The potential V_{ab}, called the **terminal voltage,** is less than the emf $\mathcal{E}$ because of the term Ir representing the potential drop across the internal resistance r. Hence the increase in potential energy qV_{ab} as a charge q moves from b to a within the source is less than the work $q\mathcal{E}$ done by the nonelectrostatic force $\vec{F}_n$, since some potential energy is lost in traversing the internal resistance.

A 1.5-V battery has an emf of 1.5 V, but the terminal voltage V_{ab} of the battery is equal to 1.5 V only if no current is flowing through it so that $I = 0$ in Eq. (25.15). If the battery is part of a complete circuit through which current is flowing, the terminal voltage will be less than 1.5 V. *For a real source of emf, the terminal voltage equals the emf only if no current is flowing through the source* (**Fig. 25.15**). Thus we can describe the behavior of a source in terms of two properties: an emf $\mathcal{E}$, which supplies a constant potential difference independent of current, in series with an internal resistance r.

The current in the external circuit connected to the source terminals a and b is still determined by $V_{ab} = IR$. Combining this with Eq. (25.15), we find

$$\mathcal{E} - Ir = IR \quad \text{or} \quad I = \frac{\mathcal{E}}{R + r} \qquad \begin{matrix}\text{(current, source with} \\ \text{internal resistance)}\end{matrix} \tag{25.16}$$

That is, the current equals the source emf divided by the *total* circuit resistance $(R + r)$.

CAUTION **A battery is not a "current source"** You might have thought that a battery or other source of emf always produces the same current, no matter what circuit it's used in. Equation (25.16) shows that this isn't so! The greater the resistance R of the external circuit, the less current the source will produce.

25.15 The emf of this battery—that is, the terminal voltage when it's not connected to anything—is 12 V. But because the battery has internal resistance, the terminal voltage of the battery is less than 12 V when it is supplying current to a light bulb.

Symbols for Circuit Diagrams

An important part of analyzing any electric circuit is drawing a schematic *circuit diagram*. **Table 25.4** shows the usual symbols used in circuit diagrams. We will use these symbols extensively in this chapter and the next. We usually assume that the wires that connect the various elements of the circuit have negligible resistance; from Eq. (25.11), $V = IR$, the potential difference between the ends of such a wire is zero.

Table 25.4 includes two *meters* that are used to measure the properties of circuits. Idealized meters do not disturb the circuit in which they are connected. A **voltmeter,** introduced in Section 23.2, measures the potential difference between its terminals; an idealized voltmeter has infinitely large resistance and measures potential difference without having any current diverted through it. An ammeter measures the current passing through it; an idealized **ammeter** has zero resistance and has no potential difference between its terminals. The following examples illustrate how to analyze circuits that include meters.

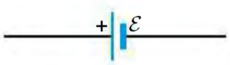
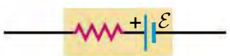
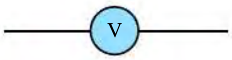
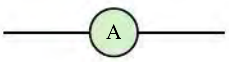
DATA *SPEAKS*

Circuits, emf, and Current

When students were given a problem involving a source of emf, more than 25% gave an incorrect response. Common errors:

- Forgetting that the internal resistance r affects the potential difference V_{ab} between the terminals of the source of emf $\mathcal{E}$. If the current I inside the source is from the negative terminal b to the positive terminal a, then $V_{ab} < \mathcal{E}$ by an amount Ir; if the current flows the other way, $V_{ab} > \mathcal{E}$ by an amount Ir.
- Forgetting that the internal resistance is an intrinsic part of a source of emf. Although we draw the emf and internal resistance as adjacent parts of the circuit, both are parts of the source and cannot be separated.

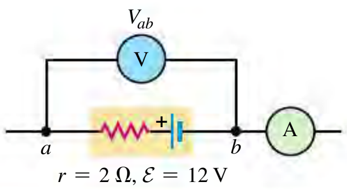
TABLE 25.4 **Symbols for Circuit Diagrams**

	Conductor with negligible resistance
	Resistor
	Source of emf (longer vertical line always represents the positive terminal, usually the terminal with higher potential)
 or 	Source of emf with internal resistance r (r can be placed on either side)
	Voltmeter (measures potential difference between its terminals)
	Ammeter (measures current through it)

CONCEPTUAL EXAMPLE 25.4 A SOURCE IN AN OPEN CIRCUIT

Figure 25.16 shows a source (a battery) with emf $\mathcal{E} = 12\text{ V}$ and internal resistance $r = 2\ \Omega$. (For comparison, the internal resistance of a commercial 12-V lead storage battery is only a few thousandths of an ohm.) The wires to the left of a and to the right of the ammeter A are not connected to anything. Determine the respective readings V_{ab} and I of the idealized voltmeter V and the idealized ammeter A.

25.16 A source of emf in an open circuit.



SOLUTION

There is *zero* current because there is no complete circuit. (Our idealized voltmeter has an infinitely large resistance, so no current flows through it.) Hence the ammeter reads $I = 0$. Because there is no current through the battery, there is no potential difference across its internal resistance. From Eq. (25.15) with $I = 0$, the potential difference V_{ab} across the battery terminals is equal to the emf. So the voltmeter reads $V_{ab} = \mathcal{E} = 12\text{ V}$. The terminal voltage of a real, nonideal source equals the emf *only* if there is no current flowing through the source, as in this example.

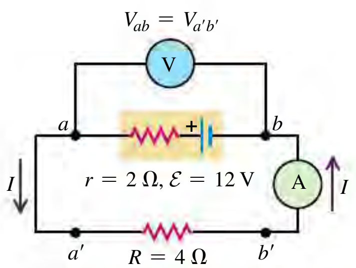
EXAMPLE 25.5 A SOURCE IN A COMPLETE CIRCUIT

We add a 4-Ω resistor to the battery in Conceptual Example 25.4, forming a complete circuit (**Fig. 25.17**). What are the voltmeter and ammeter readings V_{ab} and I now?

SOLUTION

IDENTIFY and SET UP: Our target variables are the current I through the circuit $aa'b'b$ and the potential difference V_{ab}. We first find I from Eq. (25.16). To find V_{ab}, we can use either Eq. (25.11) or Eq. (25.15).

25.17 A source of emf in a complete circuit.



EXECUTE: The ideal ammeter has zero resistance, so the total resistance external to the source is $R = 4\ \Omega$. From Eq. (25.16), the current through the circuit $aa'b'b$ is then

$$I = \frac{\mathcal{E}}{R + r} = \frac{12\text{ V}}{4\ \Omega + 2\ \Omega} = 2\text{ A}$$

Our idealized conducting wires and the idealized ammeter have zero resistance, so there is no potential difference between points a and a' or between points b and b'; that is, $V_{ab} = V_{a'b'}$. We find V_{ab} by considering a and b as the terminals of the resistor: From Ohm's law, Eq. (25.11), we then have

$$V_{a'b'} = IR = (2\text{ A})(4\ \Omega) = 8\text{ V}$$

Alternatively, we can consider a and b as the terminals of the source. Then, from Eq. (25.15),

$$V_{ab} = \mathcal{E} - Ir = 12\text{ V} - (2\text{ A})(2\ \Omega) = 8\text{ V}$$

Either way, we see that the voltmeter reading is 8 V.

EVALUATE: With this current flowing through the source, the terminal voltage V_{ab} is less than the emf $\mathcal{E}$. The smaller the internal resistance r, the less the difference between V_{ab} and $\mathcal{E}$.

CONCEPTUAL EXAMPLE 25.6 USING VOLTMETERS AND AMMETERS

We move the voltmeter and ammeter in Example 25.5 to different positions in the circuit. What are the readings of the ideal voltmeter and ammeter in the situations shown in (a) **Fig. 25.18a** and (b) Fig. 25.18b?

SOLUTION

(a) The voltmeter now measures the potential difference between points a' and b'. As in Example 25.5, $V_{ab} = V_{a'b'}$, so the voltmeter reads the same as in Example 25.5: $V_{a'b'} = 8\text{ V}$.

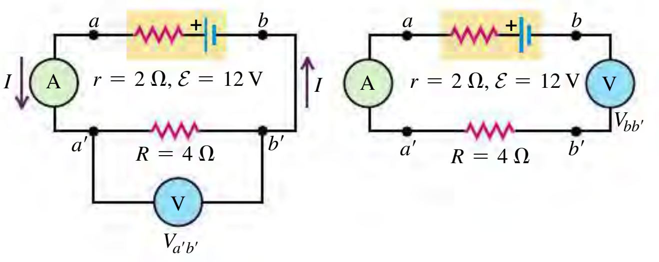
CAUTION **Current in a simple loop** As charges move through a resistor, there is a decrease in electric potential energy, but there is *no* change in the current. *The current in a simple loop is the same at every point;* it is not "used up" as it moves through a resistor. Hence the ammeter in Fig. 25.17 ("downstream" of the 4-Ω resistor) and the ammeter in Fig. 25.18b ("upstream" of the resistor) both read $I = 2\text{ A}$.

(b) There is no current through the ideal voltmeter because it has infinitely large resistance. Since the voltmeter is now part of the circuit, there is no current at all in the circuit, and the ammeter reads $I = 0$.

The voltmeter measures the potential difference $V_{bb'}$ between points b and b'. Since $I = 0$, the potential difference across the resistor is $V_{a'b'} = IR = 0$, and the potential difference between the ends a and a' of the idealized ammeter is also zero. So $V_{bb'}$ is equal to V_{ab}, the terminal voltage of the source. As in Conceptual Example 25.4, there is no current, so the terminal voltage equals the emf, and the voltmeter reading is $V_{ab} = \mathcal{E} = 12\text{ V}$.

This example shows that ammeters and voltmeters are circuit elements, too. Moving the voltmeter from the position in Fig. 25.18a to that in Fig. 25.18b makes large changes in the current and potential differences in the circuit. If you want to measure the potential difference between two points in a circuit without disturbing the circuit, use a voltmeter as in Fig. 25.17 or 25.18a, *not* as in Fig. 25.18b.

25.18 Different placements of a voltmeter and an ammeter in a complete circuit.

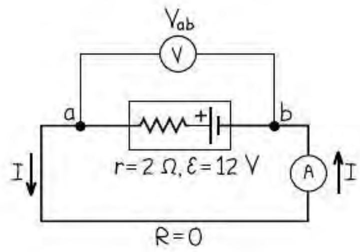


EXAMPLE 25.7 A SOURCE WITH A SHORT CIRCUIT

SOLUTION

In the circuit of Example 25.5 we replace the 4-Ω resistor with a zero-resistance conductor. What are the meter readings now?

25.19 Our sketch for this problem.



SOLUTION

IDENTIFY and SET UP: Figure 25.19 shows the new circuit. Our target variables are again I and V_{ab}. There is now a zero-resistance path between points a and b, through the lower loop, so the potential difference between these points must be zero.

EXECUTE: We must have $V_{ab} = IR = I(0) = 0$, no matter what the current. We can therefore find the current I from Eq. (25.15):

$$V_{ab} = \mathcal{E} - Ir = 0$$

$$I = \frac{\mathcal{E}}{r} = \frac{12\text{ V}}{2\ \Omega} = 6\text{ A}$$

EVALUATE: The current in this circuit has a different value than in Example 25.5, even though the same battery is used; the current depends on both the internal resistance r and the resistance of the external circuit.

The situation here is called a *short circuit*. The external-circuit resistance is zero, because terminals of the battery are connected directly to each other. The short-circuit current is equal to the emf $\mathcal{E}$ divided by the internal resistance r. *Warning:* Short circuits can be dangerous! An automobile battery or a household power line has very small internal resistance (much less than in these examples), and the short-circuit current can be great enough to melt a small wire or cause a storage battery to explode.

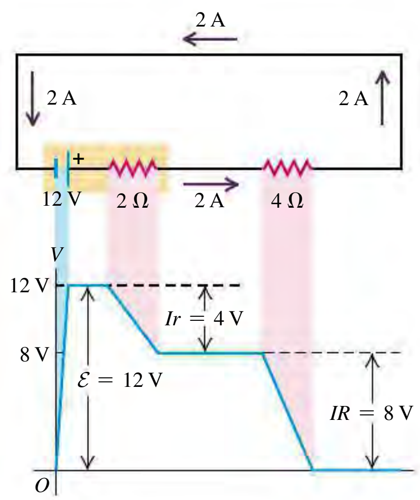
Potential Changes around a Circuit

The net change in potential energy for a charge q making a round trip around a complete circuit must be zero. Hence the net change in *potential* around the circuit must also be zero; in other words, the algebraic sum of the potential differences and emfs around the loop is zero. We can see this by rewriting Eq. (25.16) in the form

$$\mathcal{E} - Ir - IR = 0$$

A potential gain of $\mathcal{E}$ is associated with the emf, and potential drops of Ir and IR are associated with the internal resistance of the source and the external circuit, respectively. **Figure 25.20** shows how the potential varies as we go around the complete circuit of Fig. 25.17. The horizontal axis doesn't necessarily represent actual distances, but rather various points in the loop. If we take the potential to be zero at the negative terminal of the battery, then we have a rise $\mathcal{E}$ and a drop Ir in the battery and an additional drop IR in the external resistor, and as we finish our trip around the loop, the potential is back where it started.

25.20 Potential rises and drops in a circuit.



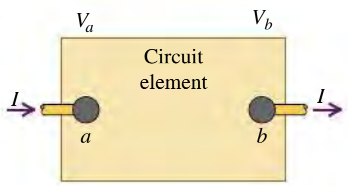
In this section we have considered only situations in which the resistances are ohmic. If the circuit includes a nonlinear device such as a diode (see Fig. 25.10b), Eq. (25.16) is still valid but cannot be solved algebraically because R is not a constant. In such a situation, I can be found by using numerical techniques.

Finally, we remark that Eq. (25.15) is not always an adequate representation of the behavior of a source. The emf may not be constant, and what we have described as an internal resistance may actually be a more complex voltage–current relationship that doesn't obey Ohm's law. Nevertheless, the concept of internal resistance frequently provides an adequate description of batteries, generators, and other energy converters. The principal difference between a fresh flashlight battery and an old one is not in the emf, which decreases only slightly with use, but in the internal resistance, which may increase from less than an ohm when the battery is fresh to as much as 1000 Ω or more after long use. Similarly, a car battery can deliver less current to the starter motor on a cold morning than when the battery is warm, not because the emf is appreciably less but because the internal resistance increases with decreasing temperature.

TEST YOUR UNDERSTANDING OF SECTION 25.4 Rank the following circuits in order from highest to lowest current: (i) A 1.4-Ω resistor connected to a 1.5-V battery that has an internal resistance of 0.10 Ω; (ii) a 1.8-Ω resistor connected to a 4.0-V battery that has a terminal voltage of 3.6 V but an unknown internal resistance; (iii) an unknown resistor connected to a 12.0-V battery that has an internal resistance of 0.20 Ω and a terminal voltage of 11.0 V. ▌

25.5 ENERGY AND POWER IN ELECTRIC CIRCUITS

25.21 The power input to the circuit element between a and b is $P = (V_a - V_b)I = V_{ab}I$.



MP

PhET: Battery-Resistor Circuit
PhET: Circuit Construction Kit (AC+DC)
PhET: Circuit Construction Kit (DC Only)
PhET: Ohm's Law

Let's now look at some energy and power relationships in electric circuits. The box in **Fig. 25.21** represents a circuit element with potential difference $V_a - V_b = V_{ab}$ between its terminals and current I passing through it in the direction from a toward b. This element might be a resistor, a battery, or something else; the details don't matter. As charge passes through the circuit element, the electric field does work on the charge. In a source of emf, additional work is done by the force $\vec{F}_n$ that we mentioned in Section 25.4.

As an amount of charge q passes through the circuit element, there is a change in potential energy equal to qV_{ab}. For example, if $q > 0$ and $V_{ab} = V_a - V_b$ is positive, potential energy decreases as the charge "falls" from potential V_a to lower potential V_b. The moving charges don't gain *kinetic* energy, because the current (the rate of charge flow) out of the circuit element must be the same as the current into the element. Instead, the quantity qV_{ab} represents energy transferred into the circuit element. This situation occurs in the coils of a toaster or electric oven, in which electrical energy is converted to thermal energy.

If the potential at a is lower than at b, then V_{ab} is negative and there is a net transfer of energy *out* of the circuit element. The element then acts as a source, delivering electrical energy into the circuit to which it is attached. This is the usual situation for a battery, which converts chemical energy into electrical energy and delivers it to the external circuit. Thus qV_{ab} can denote a quantity of energy that is either delivered to a circuit element or extracted from that element.

In electric circuits we are most often interested in the *rate* at which energy is either delivered to or extracted from a circuit element. If the current through the element is I, then in a time interval dt an amount of charge $dQ = I\,dt$ passes through the element. The potential energy change for this amount of charge is $V_{ab}\,dQ = V_{ab}I\,dt$. Dividing this expression by dt, we obtain the *rate* at which energy is transferred either into or out of the circuit element. The time rate of energy transfer is *power,* denoted by P, so we write

Power delivered to or extracted from a circuit element

$$P = V_{ab} I \qquad (25.17)$$

Voltage across circuit element (V_{ab})

Current in circuit element (I)

The unit of V_{ab} is one volt, or one joule per coulomb, and the unit of I is one ampere, or one coulomb per second. Hence the unit of $P = V_{ab}I$ is one watt:

$$(1\ \text{J/C})(1\ \text{C/s}) = 1\ \text{J/s} = 1\ \text{W}$$

Let's consider a few special cases.

Power Input to a Pure Resistance

If the circuit element in Fig. 25.21 is a resistor, the potential difference is $V_{ab} = IR$. From Eq. (25.17) the power delivered to the resistor by the circuit is

Voltage across resistor

Power delivered to a resistor

$$P = V_{ab} I = I^2 R = \frac{V_{ab}{}^2}{R} \qquad (25.18)$$

Current in resistor Resistance of resistor

DEMO

In this case the potential at a (where the current enters the resistor) is always higher than that at b (where the current exits). Current enters the higher-potential terminal of the device, and Eq. (25.18) represents the rate of transfer of electric potential energy *into* the circuit element.

What becomes of this energy? The moving charges collide with atoms in the resistor and transfer some of their energy to these atoms, increasing the internal energy of the material. Either the temperature of the resistor increases or there is a flow of heat out of it, or both. In any of these cases we say that energy is *dissipated* in the resistor at a rate I^2R. Every resistor has a *power rating*, the maximum power the device can dissipate without becoming overheated and damaged. Some devices, such as electric heaters, are designed to get hot and transfer heat to their surroundings. But if the power rating is exceeded, even such a device may melt or even explode.

Power Output of a Source

The upper rectangle in **Fig. 25.22a** represents a source with emf $\mathcal{E}$ and internal resistance r, connected by ideal (resistanceless) conductors to an external circuit represented by the lower box. This could describe a car battery connected to one of the car's headlights (Fig. 25.22b). Point a is at higher potential than point b, so $V_a > V_b$ and V_{ab} is positive. Note that the current I is *leaving* the source at the higher-potential terminal (rather than entering there). Energy is being delivered to the external circuit, at a rate given by Eq. (25.17):

$$P = V_{ab}I$$

For a source that can be described by an emf $\mathcal{E}$ and an internal resistance r, we may use Eq. (25.15):

$$V_{ab} = \mathcal{E} - Ir$$

Multiplying this equation by I, we find

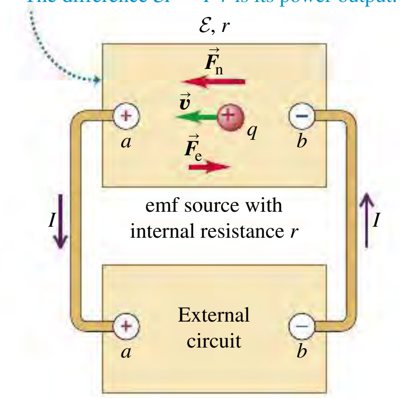
$$P = V_{ab}I = \mathcal{E}I - I^2r \qquad (25.19)$$

What do the terms $\mathcal{E}I$ and I^2r mean? In Section 25.4 we defined the emf $\mathcal{E}$ as the work per unit charge performed on the charges by the nonelectrostatic force as the charges are pushed "uphill" from b to a in the source. In a time dt, a charge $dQ = I\,dt$ flows through the source; the work done on it by this nonelectrostatic force is $\mathcal{E}\,dQ = \mathcal{E}I\,dt$. Thus $\mathcal{E}I$ is the *rate* at which work is done on the circulating charges by whatever agency causes the nonelectrostatic force in the source. This term represents the rate of conversion of nonelectrical energy to electrical energy within the source. The term I^2r is the rate at which electrical energy is

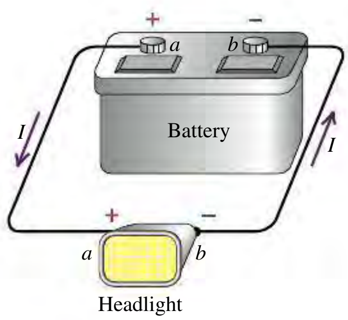
25.22 Energy conversion in a simple circuit.

(a) Diagrammatic circuit

- The emf source converts nonelectrical to electrical energy at a rate $\mathcal{E}I$.
- Its internal resistance *dissipates* energy at a rate I^2r.
- The difference $\mathcal{E}I - I^2r$ is its power output.

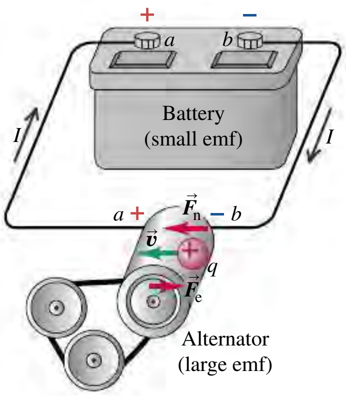


(b) A real circuit of the type shown in (a)



dissipated in the internal resistance of the source. The difference $\mathcal{E}I - I^2r$ is the *net* electrical power output of the source—that is, the rate at which the source delivers electrical energy to the remainder of the circuit.

25.23 When two sources are connected in a simple loop circuit, the source with the larger emf delivers energy to the other source.



Power Input to a Source

Suppose that the lower rectangle in Fig. 25.22a is itself a source, with an emf *larger* than that of the upper source and opposite to that of the upper source. **Figure 25.23** shows a practical example, an automobile battery (the upper circuit element) being charged by the car's alternator (the lower element). The current I in the circuit is then *opposite* to that shown in Fig. 25.22; the lower source is pushing current backward through the upper source. Because of this reversal of current, instead of Eq. (25.15), we have for the upper source

$$V_{ab} = \mathcal{E} + Ir$$

and instead of Eq. (25.19), we have

$$P = V_{ab}I = \mathcal{E}I + I^2r \tag{25.20}$$

Work is being done *on,* rather than *by*, the agent that causes the nonelectrostatic force in the upper source. There is a conversion of electrical energy into nonelectrical energy in the upper source at a rate $\mathcal{E}I$. The term I^2r in Eq. (25.20) is again the rate of dissipation of energy in the internal resistance of the upper source, and the sum $\mathcal{E}I + I^2r$ is the total electrical power *input* to the upper source. This is what happens when a rechargeable battery (a storage battery) is connected to a charger. The charger supplies electrical energy to the battery; part of it is converted to chemical energy, to be reconverted later, and the remainder is dissipated (wasted) in the battery's internal resistance, warming the battery and causing a heat flow out of it. If you have a power tool or laptop computer with a rechargeable battery, you may have noticed that it gets warm while it is charging.

PROBLEM-SOLVING STRATEGY 25.1 POWER AND ENERGY IN CIRCUITS

IDENTIFY *the relevant concepts:* The ideas of electrical power input and output can be applied to any electric circuit. Many problems will ask you to explicitly consider power or energy.

SET UP *the problem* using the following steps:
1. Make a drawing of the circuit.
2. Identify the circuit elements, including sources of emf and resistors. We will introduce other circuit elements later, including capacitors (Chapter 26) and inductors (Chapter 30).
3. Identify the target variables. Typically they will be the power input or output for each circuit element, or the total amount of energy put into or taken out of a circuit element in a given time.

EXECUTE *the solution* as follows:
1. A source of emf $\mathcal{E}$ delivers power $\mathcal{E}I$ into a circuit when current I flows through the source in the direction from $-$ to $+$. (For example, energy is converted from chemical energy in a battery, or from mechanical energy in a generator.) In this case there is a *positive* power output to the circuit or, equivalently, a *negative* power input to the source.
2. A source of emf takes power $\mathcal{E}I$ from a circuit when current passes through the source from $+$ to $-$. (This occurs in charging a storage battery, when electrical energy is converted to chemical energy.) In this case there is a *negative* power output to the circuit or, equivalently, a *positive* power input to the source.
3. There is always a *positive* power input to a resistor through which current flows, irrespective of the direction of current flow. This process removes energy from the circuit, converting it to heat at the rate $VI = I^2R = V^2/R$, where V is the potential difference across the resistor.
4. Just as in item 3, there always is a positive power input to the internal resistance r of a source through which current flows, irrespective of the direction of current flow. This process likewise removes energy from the circuit, converting it into heat at the rate I^2r.
5. If the power into or out of a circuit element is constant, the energy delivered to or extracted from that element is the product of power and elapsed time. (In Chapter 26 we will encounter situations in which the power is not constant. In such cases, calculating the total energy requires an integral over the relevant time interval.)

EVALUATE *your answer:* Check your results; in particular, check that energy is conserved. This conservation can be expressed in either of two forms: "net power input = net power output" or "the algebraic sum of the power inputs to the circuit elements is zero."

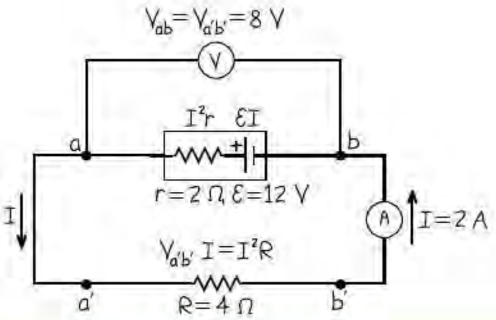
EXAMPLE 25.8 POWER INPUT AND OUTPUT IN A COMPLETE CIRCUIT

For the circuit that we analyzed in Example 25.5, find the rates of energy conversion (chemical to electrical) and energy dissipation in the battery, the rate of energy dissipation in the 4-Ω resistor, and the battery's net power output.

SOLUTION

IDENTIFY and SET UP: Figure 25.24 shows the circuit, gives values of quantities known from Example 25.5, and indicates how we find the target variables. We use Eq. (25.19) to find the battery's net power output, the rate of chemical-to-electrical energy conversion, and the rate of energy dissipation in the battery's internal resistance. We use Eq. (25.18) to find the power delivered to (and dissipated in) the 4-Ω resistor.

25.24 Our sketch for this problem.



EXECUTE: From the first term in Eq. (25.19), the rate of energy conversion in the battery is

$$\mathcal{E}I = (12\text{ V})(2\text{ A}) = 24\text{ W}$$

From the second term in Eq. (25.19), the rate of dissipation of energy in the battery is

$$I^2r = (2\text{ A})^2(2\ \Omega) = 8\text{ W}$$

The *net* electrical power output of the battery is the difference between these: $\mathcal{E}I - I^2r = 16$ W. From Eq. (25.18), the electrical power input to, and the equal rate of dissipation of electrical energy in, the 4-Ω resistor are

$$V_{a'b'}I = (8\text{ V})(2\text{ A}) = 16\text{ W} \quad \text{and}$$

$$I^2R = (2\text{ A})^2(4\ \Omega) = 16\text{ W}$$

EVALUATE: The rate $V_{a'b'}I$ at which energy is supplied to the 4-Ω resistor equals the rate I^2R at which energy is dissipated there. This is also equal to the battery's net power output: $P = V_{ab}I = (8\text{ V})(2\text{ A}) = 16$ W. In summary, the rate at which the source of emf supplies energy is $\mathcal{E}I = 24$ W, of which $I^2r = 8$ W is dissipated in the battery's internal resistor and $I^2R = 16$ W is dissipated in the external resistor.

EXAMPLE 25.9 INCREASING THE RESISTANCE

Suppose we replace the external 4-Ω resistor in Fig. 25.24 with an 8-Ω resistor. How does this affect the electrical power dissipated in this resistor?

SOLUTION

IDENTIFY and SET UP: Our target variable is the power dissipated in the resistor to which the battery is connected. The situation is the same as in Example 25.8, but with a higher external resistance R.

EXECUTE: According to Eq. (25.18), the power dissipated in the resistor is $P = I^2R$. You might conclude that making the resistance R twice as great as in Example 25.8 should make the power twice as great, or $2(16\text{ W}) = 32$ W. If instead you used the formula $P = V_{ab}^2/R$, you might conclude that the power should be one-half as great as in the preceding example, or $(16\text{ W})/2 = 8$ W. Which answer is correct?

In fact, *both* of these answers are *incorrect*. The first is wrong because changing the resistance R also changes the current in the circuit (remember, a source of emf does *not* generate the same current in all situations). The second answer is wrong because the potential difference V_{ab} across the resistor changes when the current changes. To get the correct answer, we first find the current just as we did in Example 25.5:

$$I = \frac{\mathcal{E}}{R + r} = \frac{12\text{ V}}{8\ \Omega + 2\ \Omega} = 1.2\text{ A}$$

The greater resistance causes the current to decrease. The potential difference across the resistor is

$$V_{ab} = IR = (1.2\text{ A})(8\ \Omega) = 9.6\text{ V}$$

which is greater than that with the 4-Ω resistor. We can then find the power dissipated in the resistor in either of two ways:

$$P = I^2R = (1.2\text{ A})^2(8\ \Omega) = 12\text{ W} \quad \text{or}$$

$$P = \frac{V_{ab}^2}{R} = \frac{(9.6\text{ V})^2}{8\ \Omega} = 12\text{ W}$$

EVALUATE: Increasing the resistance R causes a *reduction* in the power input to the resistor. In the expression $P = I^2R$ the decrease in current is more important than the increase in resistance; in the expression $P = V_{ab}^2/R$ the increase in resistance is more important than the increase in V_{ab}. This same principle applies to ordinary light bulbs; a 50-W light bulb has a greater resistance than does a 100-W light bulb.

Can you show that replacing the 4-Ω resistor with an 8-Ω resistor decreases both the rate of energy conversion (chemical to electrical) in the battery and the rate of energy dissipation in the battery?

EXAMPLE 25.10 POWER IN A SHORT CIRCUIT

For the short-circuit situation of Example 25.7, find the rates of energy conversion and energy dissipation in the battery and the net power output of the battery.

SOLUTION

IDENTIFY and SET UP: Our target variables are again the power inputs and outputs associated with the battery. **Figure 25.25** shows the circuit. This is the same situation as in Example 25.8, but now the external resistance R is zero.

EXECUTE: We found in Example 25.7 that the current in this situation is $I = 6$ A. From Eq. (25.19), the rate of energy conversion (chemical to electrical) in the battery is then

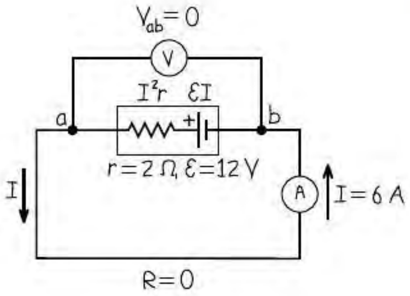
$$\mathcal{E}I = (12\text{ V})(6\text{ A}) = 72\text{ W}$$

and the rate of dissipation of energy in the battery is

$$I^2r = (6\text{ A})^2(2\ \Omega) = 72\text{ W}$$

The net power output of the source is $\mathcal{E}I - I^2r = 0$. We get this same result from the expression $P = V_{ab}I$, because the terminal voltage V_{ab} of the source is zero.

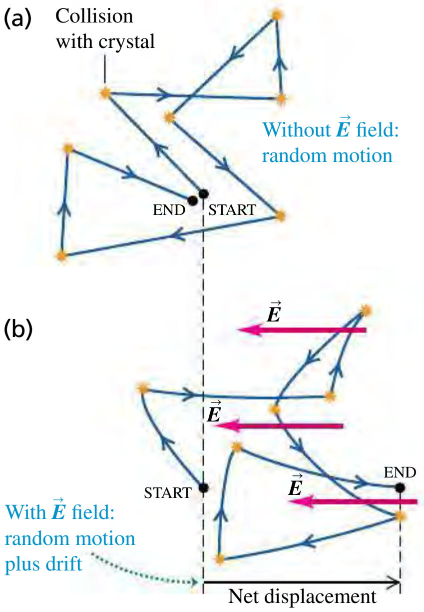
25.25 Our sketch for this problem.



EVALUATE: With ideal wires and an ideal ammeter, so that $R = 0$, *all* of the converted energy from the source is dissipated within the source. This is why a short-circuited battery is quickly ruined and may explode.

TEST YOUR UNDERSTANDING OF SECTION 25.5 Rank the following circuits in order from highest to lowest values of the net power output of the battery. (i) A 1.4-Ω resistor connected to a 1.5-V battery that has an internal resistance of 0.10 Ω; (ii) a 1.8-Ω resistor connected to a 4.0-V battery that has a terminal voltage of 3.6 V but an unknown internal resistance; (iii) an unknown resistor connected to a 12.0-V battery that has an internal resistance of 0.20 Ω and a terminal voltage of 11.0 V. ❙

25.26 Random motions of an electron in a metallic crystal (a) with zero electric field and (b) with an electric field that causes drift. The curvatures of the paths are greatly exaggerated.



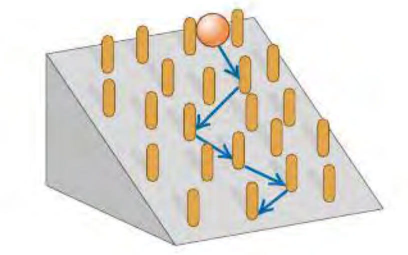
25.6 THEORY OF METALLIC CONDUCTION

We can gain additional insight into electrical conduction by looking at the microscopic origin of conductivity. We'll consider a very simple model that treats the electrons as classical particles and ignores their quantum-mechanical behavior in solids. Using this model, we'll derive an expression for the resistivity of a metal. Even though this model is not entirely correct, it will still help you to develop an intuitive idea of the microscopic basis of conduction.

In the simplest microscopic model of conduction in a metal, each atom in the metallic crystal gives up one or more of its outer electrons. These electrons are then free to move through the crystal, colliding at intervals with the stationary positive ions. The motion of the electrons is analogous to the motion of molecules of a gas moving through a porous bed of sand.

If there is no electric field, the electrons move in straight lines between collisions, the directions of their velocities are random, and on average they never get anywhere (**Fig. 25.26a**). But if an electric field is present, the paths curve slightly because of the acceleration caused by electric-field forces. Figure 25.26b shows a few paths of an electron in an electric field directed from right to left. As we mentioned in Section 25.1, the average speed of random motion is of the order of 10^6 m/s, while the average drift speed is *much* slower, of the order of 10^{-4} m/s.

The average time between collisions is called the **mean free time,** denoted by τ. **Figure 25.27** shows a mechanical analog of this electron motion.



25.27 The motion of a ball rolling down an inclined plane and bouncing off pegs in its path is analogous to the motion of an electron in a metallic conductor with an electric field present.

MP

PhET: Conductivity

We would like to derive from this model an expression for the resistivity ρ of a material, defined by Eq. (25.5):

$$\rho = \frac{E}{J} \tag{25.21}$$

where E and J are the magnitudes of electric field and current density, respectively. The current density $\vec{J}$ is in turn given by Eq. (25.4):

$$\vec{J} = nq\vec{v}_{\mathrm{d}} \tag{25.22}$$

where n is the number of free electrons per unit volume (the electron concentration), $q = -e$ is the charge of each, and $\vec{v}_{\mathrm{d}}$ is their average drift velocity.

We need to relate the drift velocity $\vec{v}_{\mathrm{d}}$ to the electric field $\vec{E}$. The value of $\vec{v}_{\mathrm{d}}$ is determined by a steady-state condition in which, on average, the velocity *gains* of the charges due to the force of the $\vec{E}$ field are just balanced by the velocity *losses* due to collisions. To clarify this process, let's imagine turning on the two effects one at a time. Suppose that before time $t = 0$ there is no field. The electron motion is then completely random. A typical electron has velocity $\vec{v}_0$ at time $t = 0$, and the value of $\vec{v}_0$ averaged over many electrons (that is, the initial velocity of an average electron) is zero: $(\vec{v}_0)_{\mathrm{av}} = \mathbf{0}$. Then at time $t = 0$ we turn on a constant electric field $\vec{E}$. The field exerts a force $\vec{F} = q\vec{E}$ on each charge, and this causes an acceleration $\vec{a}$ in the direction of the force, given by

$$\vec{a} = \frac{\vec{F}}{m} = \frac{q\vec{E}}{m}$$

where m is the electron mass. *Every* electron has this acceleration.

After a time τ, the average time between collisions, we "turn on" the collisions. At time $t = \tau$ an electron that has velocity $\vec{v}_0$ at time $t = 0$ has a velocity

$$\vec{v} = \vec{v}_0 + \vec{a}\tau$$

The velocity $\vec{v}_{\mathrm{av}}$ of an *average* electron at this time is the sum of the averages of the two terms on the right. As we have pointed out, the initial velocity $\vec{v}_0$ is zero for an average electron, so

$$\vec{v}_{\mathrm{av}} = \vec{a}\tau = \frac{q\tau}{m}\vec{E} \tag{25.23}$$

After time $t = \tau$, the tendency of the collisions to decrease the velocity of an average electron (by means of randomizing collisions) just balances the tendency of the $\vec{E}$ field to increase this velocity. Thus the velocity of an average electron, given by Eq. (25.23), is maintained over time and is equal to the drift velocity $\vec{v}_{\mathrm{d}}$:

$$\vec{v}_{\mathrm{d}} = \frac{q\tau}{m}\vec{E}$$

Now we substitute this equation for the drift velocity $\vec{v}_{\mathrm{d}}$ into Eq. (25.22):

$$\vec{J} = nq\vec{v}_{\mathrm{d}} = \frac{nq^2\tau}{m}\vec{E}$$

Comparing this with Eq. (25.21), which we can rewrite as $\vec{J} = \vec{E}/\rho$, and substituting $q = -e$ for an electron, we see that

Resistivity of a metal

$$\rho = \frac{m}{ne^2\tau} \tag{25.24}$$

Electron mass

Average time between collisions

Number of free electrons per unit volume

Magnitude of electron charge

If n and τ are independent of $\vec{E}$, then the resistivity is independent of $\vec{E}$ and the conducting material obeys Ohm's law.

Turning the interactions on one at a time may seem artificial. But the derivation would come out the same if each electron had its own clock and the $t = 0$ times were different for different electrons. If τ is the average time between collisions, then $\vec{v}_d$ is still the average electron drift velocity, even though the motions of the various electrons aren't actually correlated in the way we postulated.

What about the temperature dependence of resistivity? In a perfect crystal with no atoms out of place, a correct quantum-mechanical analysis would let the free electrons move through the crystal with no collisions at all. But the atoms vibrate about their equilibrium positions. As the temperature increases, the amplitudes of these vibrations increase, collisions become more frequent, and the mean free time τ decreases. So this theory predicts that the resistivity of a metal increases with temperature. In a superconductor, roughly speaking, there are no inelastic collisions, τ is infinite, and the resistivity ρ is zero.

In a pure semiconductor such as silicon or germanium, the number of charge carriers per unit volume, n, is not constant but increases very rapidly with increasing temperature. This increase in n far outweighs the decrease in the mean free time, and in a semiconductor the resistivity always decreases rapidly with increasing temperature. At low temperatures, n is very small, and the resistivity becomes so large that the material can be considered an insulator.

Electrons gain energy between collisions through the work done on them by the electric field. During collisions they transfer some of this energy to the atoms of the material of the conductor. This leads to an increase in the material's internal energy and temperature; that's why wires carrying current get warm. If the electric field in the material is large enough, an electron can gain enough energy between collisions to knock off electrons that are normally bound to atoms in the material. These can then knock off more electrons, and so on, leading to an avalanche of current. This is the basis of dielectric breakdown in insulators (see Section 24.4).

EXAMPLE 25.11 MEAN FREE TIME IN COPPER

SOLUTION

Calculate the mean free time between collisions in copper at room temperature.

SOLUTION

IDENTIFY and SET UP: We can obtain an expression for mean free time τ in terms of n, ρ, e, and m by rearranging Eq. (25.24). From Example 25.1 and Table 25.1, for copper $n = 8.5 \times 10^{28}\ \text{m}^{-3}$ and $\rho = 1.72 \times 10^{-8}\ \Omega \cdot \text{m}$. In addition, $e = 1.60 \times 10^{-19}\ \text{C}$ and $m = 9.11 \times 10^{-31}\ \text{kg}$ for electrons.

EXECUTE: From Eq. (25.24), we get

$$\tau = \frac{m}{ne^2\rho}$$
$$= \frac{9.11 \times 10^{-31}\ \text{kg}}{(8.5 \times 10^{28}\ \text{m}^{-3})(1.60 \times 10^{-19}\ \text{C})^2(1.72 \times 10^{-8}\ \Omega \cdot \text{m})}$$
$$= 2.4 \times 10^{-14}\ \text{s}$$

EVALUATE: The mean free time is the average time between collisions for a given electron. Taking the reciprocal, we find that each electron averages $1/\tau = 4.2 \times 10^{13}$ collisions per second!

TEST YOUR UNDERSTANDING OF SECTION 25.6 Which of the following factors will, if increased, make it more difficult to produce a certain amount of current in a conductor? (There may be more than one correct answer.) (i) The mass of the moving charged particles in the conductor; (ii) the number of moving charged particles per cubic meter; (iii) the amount of charge on each moving particle; (iv) the average time between collisions for a typical moving charged particle. ❙

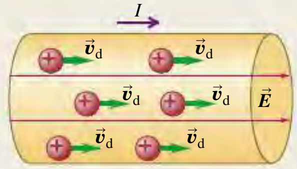
CHAPTER 25 SUMMARY

SOLUTIONS TO ALL EXAMPLES

Current and current density: Current is the amount of charge flowing through a specified area, per unit time. The SI unit of current is the ampere $(1\text{ A} = 1\text{ C/s})$. The current I through an area A depends on the concentration n and charge q of the charge carriers, as well as on the magnitude of their drift velocity $\vec{v}_\text{d}$. The current density is current per unit cross-sectional area. Current is usually described in terms of a flow of positive charge, even when the charges are actually negative or of both signs. (See Example 25.1.)

$$I = \frac{dQ}{dt} = n|q|v_\text{d}A \tag{25.2}$$

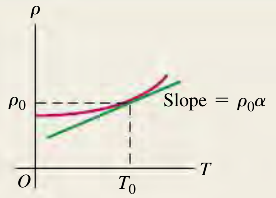
$$\vec{J} = nq\vec{v}_\text{d} \tag{25.4}$$



Resistivity: The resistivity ρ of a material is the ratio of the magnitudes of electric field and current density. Good conductors have small resistivity; good insulators have large resistivity. Ohm's law, obeyed approximately by many materials, states that ρ is a constant independent of the value of E. Resistivity usually increases with temperature; for small temperature changes this variation is represented approximately by Eq. (25.6), where α is the temperature coefficient of resistivity.

$$\rho = \frac{E}{J} \tag{25.5}$$

$$\rho(T) = \rho_0[1 + \alpha(T - T_0)] \tag{25.6}$$

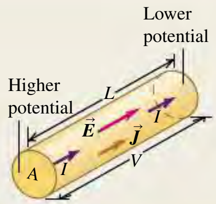


Metal: ρ increases with increasing T.

Resistors: The potential difference V across a sample of material that obeys Ohm's law is proportional to the current I through the sample. The ratio $V/I = R$ is the resistance of the sample. The SI unit of resistance is the ohm $(1\ \Omega = 1\text{ V/A})$. The resistance of a cylindrical conductor is related to its resistivity ρ, length L, and cross-sectional area A. (See Examples 25.2 and 25.3.)

$$V = IR \tag{25.11}$$

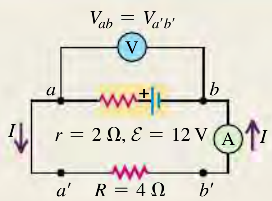
$$R = \frac{\rho L}{A} \tag{25.10}$$



Circuits and emf: A complete circuit has a continuous current-carrying path. A complete circuit carrying a steady current must contain a source of electromotive force (emf) $\mathcal{E}$. The SI unit of electromotive force is the volt (V). Every real source of emf has some internal resistance r, so its terminal potential difference V_{ab} depends on current. (See Examples 25.4–25.7.)

$$V_{ab} = \mathcal{E} - Ir \tag{25.15}$$

(source with internal resistance)



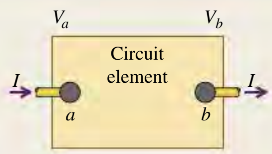
Energy and power in circuits: A circuit element puts energy into a circuit if the current direction is from lower to higher potential in the device, and it takes energy out of the circuit if the current is opposite. The power P equals the product of the potential difference $V_a - V_b = V_{ab}$ and the current I. A resistor always takes electrical energy out of a circuit. (See Examples 25.8–25.10.)

$$P = V_{ab}I \tag{25.17}$$

(general circuit element)

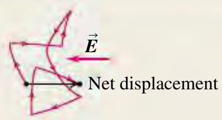
$$P = V_{ab}I = I^2R = \frac{V_{ab}{}^2}{R} \tag{25.18}$$

(power delivered to a resistor)



Conduction in metals: In a metal, current is due to the motion of electrons. They move freely through the metallic crystal but collide with positive ions. In a crude classical model of this motion, the resistivity of the material can be related to the electron mass, charge, speed of random motion, density, and mean free time between collisions. (See Example 25.11.)

$$\rho = \frac{m}{ne^2\tau} \tag{25.24}$$



BRIDGING PROBLEM RESISTIVITY, TEMPERATURE, AND POWER

SOLUTION

A toaster using a Nichrome heating element operates on 120 V. When it is switched on at 20°C, the heating element carries an initial current of 1.35 A. A few seconds later the current reaches the steady value of 1.23 A. (a) What is the final temperature of the element? The average value of the temperature coefficient of resistivity for Nichrome over the relevant temperature range is $4.5 \times 10^{-4}\,(\mathrm{C}°)^{-1}$. (b) What is the power dissipated in the heating element initially and when the current reaches 1.23 A?

SOLUTION GUIDE

IDENTIFY and SET UP

1. A heating element acts as a resistor that converts electrical energy into thermal energy. The resistivity ρ of Nichrome depends on temperature, and hence so does the resistance $R = \rho L/A$ of the heating element and the current $I = V/R$ that it carries.
2. We are given $V = 120$ V and the initial and final values of I. Select an equation that will allow you to find the initial and final values of resistance, and an equation that relates resistance to temperature [the target variable in part (a)].
3. The power P dissipated in the heating element depends on I and V. Select an equation that will allow you to calculate the initial and final values of P.

EXECUTE

4. Combine your equations from step 2 to give a relationship between the initial and final values of I and the initial and final temperatures (20°C and T_{final}).
5. Solve your expression from step 4 for T_{final}.
6. Use your equation from step 3 to find the initial and final powers.

EVALUATE

7. Is the final temperature greater than or less than 20°C? Does this make sense?
8. Is the final resistance greater than or less than the initial resistance? Again, does this make sense?
9. Is the final power greater than or less than the initial power? Does this agree with your observations in step 8?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q25.1 The definition of resistivity ($\rho = E/J$) implies that an electric field exists inside a conductor. Yet we saw in Chapter 21 that there can be no electrostatic electric field inside a conductor. Is there a contradiction here? Explain.

Q25.2 A cylindrical rod has resistance R. If we triple its length and diameter, what is its resistance, in terms of R?

Q25.3 A cylindrical rod has resistivity ρ. If we triple its length and diameter, what is its resistivity, in terms of ρ?

Q25.4 Two copper wires with different diameters are joined end to end. If a current flows in the wire combination, what happens to electrons when they move from the larger-diameter wire into the smaller-diameter wire? Does their drift speed increase, decrease, or stay the same? If the drift speed changes, what is the force that causes the change? Explain your reasoning.

Q25.5 When is a 1.5-V AAA battery *not* actually a 1.5-V battery? That is, when do its terminals provide a potential difference of less than 1.5 V?

Q25.6 Can the potential difference between the terminals of a battery ever be opposite in direction to the emf ? If it can, give an example. If it cannot, explain why not.

Q25.7 A rule of thumb used to determine the internal resistance of a source is that it is the open-circuit voltage divided by the short-circuit current. Is this correct? Why or why not?

Q25.8 Batteries are always labeled with their emf; for instance, an AA flashlight battery is labeled "1.5 volts." Would it also be appropriate to put a label on batteries stating how much current they provide? Why or why not?

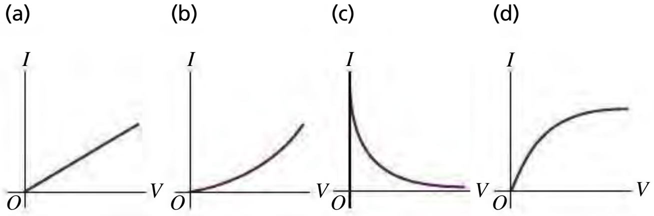
Q25.9 We have seen that a coulomb is an enormous amount of charge; it is virtually impossible to place a charge of 1 C on an object. Yet, a current of 10 A, 10 C/s, is quite reasonable. Explain this apparent discrepancy.

Q25.10 Electrons in an electric circuit pass through a resistor. The wire on either side of the resistor has the same diameter. (a) How does the drift speed of the electrons before entering the resistor compare to the speed after leaving the resistor? Explain your reasoning. (b) How does the potential energy for an electron before entering the resistor compare to the potential energy after leaving the resistor? Explain your reasoning.

Q25.11 Temperature coefficients of resistivity are given in Table 25.2. (a) If a copper heating element is connected to a source of constant voltage, does the electrical power consumed by the heating element increase or decrease as its temperature increases? Explain. (b) A resistor in the form of a carbon cylinder is connected to the voltage source. As the temperature of the cylinder increases, does the electrical power it consumes increase or decrease? Explain.

Q25.12 Which of the graphs in **Fig. Q25.12** best illustrates the current I in a real resistor as a function of the potential difference V across it? Explain.

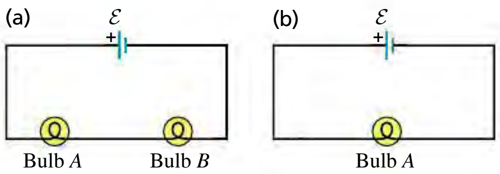
Figure **Q25.12**



Q25.13 Why does an electric light bulb nearly always burn out just as you turn on the light, almost never while the light is shining?

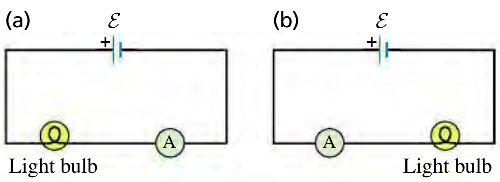
Q25.14 A light bulb glows because it has resistance. The brightness of a light bulb increases with the electrical power dissipated in the bulb. (a) In the circuit shown in **Fig. Q25.14a**, the two bulbs A and B are identical. Compared to bulb A, does bulb B glow more brightly, just as brightly, or less brightly? Explain your reasoning. (b) Bulb B is removed from the circuit and the circuit is completed as shown in Fig. Q25.14b. Compared to the brightness of bulb A in Fig. Q25.14a, does bulb A now glow more brightly, just as brightly, or less brightly? Explain your reasoning.

Figure **Q25.14**



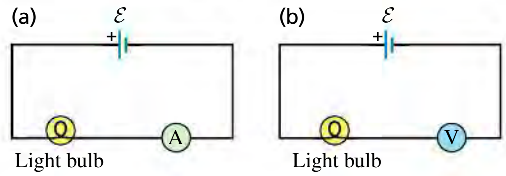
Q25.15 (See Discussion Question Q25.14.) An ideal ammeter A is placed in a circuit with a battery and a light bulb as shown in **Fig. Q25.15a**, and the ammeter reading is noted. The circuit is then reconnected as in Fig. Q25.15b, so that the positions of the ammeter and light bulb are reversed. (a) How does the ammeter reading in the situation shown in Fig. Q25.15a compare to the reading in the situation shown in Fig. Q25.15b? Explain your reasoning. (b) In which situation does the light bulb glow more brightly? Explain your reasoning.

Figure **Q25.15**



Q25.16 (See Discussion Question Q25.14.) Will a light bulb glow more brightly when it is connected to a battery as shown in **Fig. Q25.16a**, in which an ideal ammeter A is placed in the circuit, or when it is connected as shown in Fig. 25.16b, in which an ideal voltmeter V is placed in the circuit? Explain your reasoning.

Figure **Q25.16**



Q25.17 The energy that can be extracted from a storage battery is always less than the energy that goes into it while it is being charged. Why?

Q25.18 Eight flashlight batteries in series have an emf of about 12 V, similar to that of a car battery. Could they be used to start a car with a dead battery? Why or why not?

Q25.19 Small aircraft often have 24-V electrical systems rather than the 12-V systems in automobiles, even though the electrical power requirements are roughly the same in both applications. The explanation given by aircraft designers is that a 24-V system weighs less than a 12-V system because thinner wires can be used. Explain why this is so.

Q25.20 Long-distance, electric-power, transmission lines always operate at very high voltage, sometimes as much as 750 kV. What are the advantages of such high voltages? What are the disadvantages?

Q25.21 Ordinary household electric lines in North America usually operate at 120 V. Why is this a desirable voltage, rather than a value considerably larger or smaller? On the other hand, automobiles usually have 12-V electrical systems. Why is this a desirable voltage?

Q25.22 A fuse is a device designed to break a circuit, usually by melting when the current exceeds a certain value. What characteristics should the material of the fuse have?

Q25.23 High-voltage power supplies are sometimes designed intentionally to have rather large internal resistance as a safety precaution. Why is such a power supply with a large internal resistance safer than a supply with the same voltage but lower internal resistance?

Q25.24 The text states that good thermal conductors are also good electrical conductors. If so, why don't the cords used to connect toasters, irons, and similar heat-producing appliances get hot by conduction of heat from the heating element?

EXERCISES

Section 25.1 Current

25.1 • **Lightning Strikes.** During lightning strikes from a cloud to the ground, currents as high as 25,000 A can occur and last for about 40 μs. How much charge is transferred from the cloud to the earth during such a strike?

25.2 • A silver wire 2.6 mm in diameter transfers a charge of 420 C in 80 min. Silver contains 5.8×10^{28} free electrons per cubic meter. (a) What is the current in the wire? (b) What is the magnitude of the drift velocity of the electrons in the wire?

25.3 • A 5.00-A current runs through a 12-gauge copper wire (diameter 2.05 mm) and through a light bulb. Copper has 8.5×10^{28} free electrons per cubic meter. (a) How many electrons pass through the light bulb each second? (b) What is the current density in the wire? (c) At what speed does a typical electron pass by any given point in the wire? (d) If you were to use wire of twice the diameter, which of the above answers would change? Would they increase or decrease?

25.4 • An 18-gauge copper wire (diameter 1.02 mm) carries a current with a current density of $3.20 \times 10^6\ \text{A/m}^2$. The density of free electrons for copper is 8.5×10^{28} electrons per cubic meter. Calculate (a) the current in the wire and (b) the drift velocity of electrons in the wire.

25.5 •• Copper has 8.5×10^{28} free electrons per cubic meter. A 71.0-cm length of 12-gauge copper wire that is 2.05 mm in diameter carries 4.85 A of current. (a) How much time does it take for an electron to travel the length of the wire? (b) Repeat part (a) for 6-gauge copper wire (diameter 4.12 mm) of the same length that carries the same current. (c) Generally speaking, how does changing the diameter of a wire that carries a given amount of current affect the drift velocity of the electrons in the wire?

25.6 •• You want to produce three 1.00-mm-diameter cylindrical wires, each with a resistance of 1.00 Ω at room temperature. One wire is gold, one is copper, and one is aluminum. Refer to Table 25.1 for the resistivity values. (a) What will be the length of each wire? (b) Gold has a density of $1.93 \times 10^4\ \text{kg/m}^3$. What will be the mass of the gold wire? If you consider the current price of gold, is this wire very expensive?

25.7 • **CALC** The current in a wire varies with time according to the relationship $I = 55\ \text{A} - (0.65\ \text{A/s}^2)t^2$. (a) How many coulombs of charge pass a cross section of the wire in the time interval between $t = 0$ and $t = 8.0$ s? (b) What constant current would transport the same charge in the same time interval?

25.8 • Current passes through a solution of sodium chloride. In 1.00 s, $2.68 \times 10^{16}\ \text{Na}^+$ ions arrive at the negative electrode and $3.92 \times 10^{16}\ \text{Cl}^-$ ions arrive at the positive electrode. (a) What is the current passing between the electrodes? (b) What is the direction of the current?

25.9 • **BIO** **Transmission of Nerve Impulses.** Nerve cells transmit electric signals through their long tubular axons. These signals propagate due to a sudden rush of Na^+ ions, each with charge $+e$, into the axon. Measurements have revealed that typically about $5.6 \times 10^{11}\ \text{Na}^+$ ions enter each meter of the axon during a time of 10 ms. What is the current during this inflow of charge in a meter of axon?

Section 25.2 Resistivity and Section 25.3 Resistance

25.10 • (a) At room temperature, what is the strength of the electric field in a 12-gauge copper wire (diameter 2.05 mm) that is needed to cause a 4.50-A current to flow? (b) What field would be needed if the wire were made of silver instead?

25.11 •• A 1.50-m cylindrical rod of diameter 0.500 cm is connected to a power supply that maintains a constant potential difference of 15.0 V across its ends, while an ammeter measures the current through it. You observe that at room temperature (20.0°C) the ammeter reads 18.5 A, while at 92.0°C it reads 17.2 A. You can ignore any thermal expansion of the rod. Find (a) the resistivity at 20.0°C and (b) the temperature coefficient of resistivity at 20°C for the material of the rod.

25.12 • A copper wire has a square cross section 2.3 mm on a side. The wire is 4.0 m long and carries a current of 3.6 A. The density of free electrons is $8.5 \times 10^{28}/\text{m}^3$. Find the magnitudes of (a) the current density in the wire and (b) the electric field in the wire. (c) How much time is required for an electron to travel the length of the wire?

25.13 • A 14-gauge copper wire of diameter 1.628 mm carries a current of 12.5 mA. (a) What is the potential difference across a 2.00-m length of the wire? (b) What would the potential difference in part (a) be if the wire were silver instead of copper, but all else were the same?

25.14 •• A wire 6.50 m long with diameter of 2.05 mm has a resistance of 0.0290 Ω. What material is the wire most likely made of?

25.15 •• A cylindrical tungsten filament 15.0 cm long with a diameter of 1.00 mm is to be used in a machine for which the temperature will range from room temperature (20°C) up to 120°C. It will carry a current of 12.5 A at all temperatures (consult Tables 25.1 and 25.2). (a) What will be the maximum electric field in this filament, and (b) what will be its resistance with that field? (c) What will be the maximum potential drop over the full length of the filament?

25.16 •• A ductile metal wire has resistance R. What will be the resistance of this wire in terms of R if it is stretched to three times its original length, assuming that the density and resistivity of the material do not change when the wire is stretched? (*Hint:* The amount of metal does not change, so stretching out the wire will affect its cross-sectional area.)

25.17 • In household wiring, copper wire 2.05 mm in diameter is often used. Find the resistance of a 24.0-m length of this wire.

25.18 •• What diameter must a copper wire have if its resistance is to be the same as that of an equal length of aluminum wire with diameter 2.14 mm?

25.19 • A strand of wire has resistance 5.60 $\mu\Omega$. Find the net resistance of 120 such strands if they are (a) placed side by side to form a cable of the same length as a single strand, and (b) connected end to end to form a wire 120 times as long as a single strand.

25.20 • You apply a potential difference of 4.50 V between the ends of a wire that is 2.50 m in length and 0.654 mm in radius. The resulting current through the wire is 17.6 A. What is the resistivity of the wire?

25.21 • A current-carrying gold wire has diameter 0.84 mm. The electric field in the wire is 0.49 V/m. What are (a) the current carried by the wire; (b) the potential difference between two points in the wire 6.4 m apart; (c) the resistance of a 6.4-m length of this wire?

25.22 • A hollow aluminum cylinder is 2.50 m long and has an inner radius of 2.75 cm and an outer radius of 4.60 cm. Treat each surface (inner, outer, and the two end faces) as an equipotential surface. At room temperature, what will an ohmmeter read if it is connected between (a) the opposite faces and (b) the inner and outer surfaces?

25.23 • (a) What is the resistance of a Nichrome wire at 0.0°C if its resistance is 100.00 Ω at 11.5°C? (b) What is the resistance of a carbon rod at 25.8°C if its resistance is 0.0160 Ω at 0.0°C?

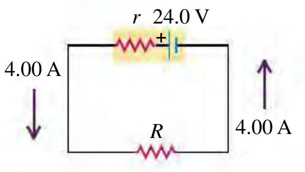
25.24 • A carbon resistor is to be used as a thermometer. On a winter day when the temperature is 4.0°C, the resistance of the carbon resistor is 217.3 Ω. What is the temperature on a spring day when the resistance is 215.8 Ω? (Take the reference temperature T_0 to be 4.0°C.)

Section 25.4 Electromotive Force and Circuits

25.25 • A copper transmission cable 100 km long and 10.0 cm in diameter carries a current of 125 A. (a) What is the potential drop across the cable? (b) How much electrical energy is dissipated as thermal energy every hour?

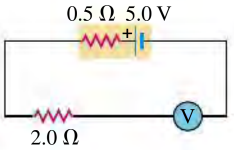
25.26 • Consider the circuit shown in **Fig. E25.26**. The terminal voltage of the 24.0-V battery is 21.2 V. What are (a) the internal resistance r of the battery and (b) the resistance R of the circuit resistor?

Figure **E25.26**



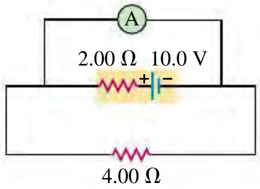
25.27 • An ideal voltmeter V is connected to a 2.0-Ω resistor and a battery with emf 5.0 V and internal resistance 0.5 Ω as shown in **Fig. E25.27**. (a) What is the current in the 2.0-Ω resistor? (b) What is the terminal voltage of the battery? (c) What is the reading on the voltmeter? Explain your answers.

Figure **E25.27**



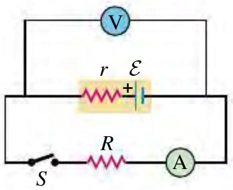
25.28 • An idealized ammeter is connected to a battery as shown in **Fig. E25.28**. Find (a) the reading of the ammeter, (b) the current through the 4.00-Ω resistor, (c) the terminal voltage of the battery.

Figure **E25.28**



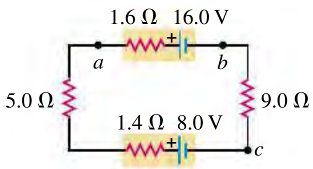
25.29 • When switch S in **Fig. E25.29** is open, the voltmeter V reads 3.08 V. When the switch is closed, the voltmeter reading drops to 2.97 V, and the ammeter A reads 1.65 A. Find the emf, the internal resistance of the battery, and the circuit resistance R. Assume that the two meters are ideal, so they don't affect the circuit.

Figure **E25.29**



25.30 • The circuit shown in **Fig. E25.30** contains two batteries, each with an emf and an internal resistance, and two resistors. Find (a) the current in the circuit (magnitude *and* direction); (b) the terminal voltage V_{ab} of the 16.0-V battery; (c) the potential difference V_{ac} of point a with respect to point c. (d) Using Fig. 25.20 as a model, graph the potential rises and drops in this circuit.

Figure **E25.30**

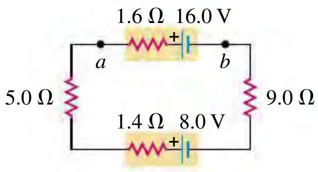


25.31 • In the circuit shown in Fig. E25.30, the 16.0-V battery is removed and reinserted with the opposite polarity, so that its negative terminal is now next to point a. Find (a) the current in the circuit (magnitude *and* direction); (b) the terminal voltage V_{ba} of the 16.0-V battery; (c) the potential difference V_{ac} of point a with respect to point c. (d) Graph the potential rises and drops in this circuit (see Fig. 25.20).

25.32 • In the circuit of Fig. E25.30, the 5.0-Ω resistor is removed and replaced by a resistor of unknown resistance R. When this is done, an ideal voltmeter connected across the points b and c reads 1.9 V. Find (a) the current in the circuit and (b) the resistance R. (c) Graph the potential rises and drops in this circuit (see Fig. 25.20).

25.33 •• The circuit shown in **Fig. E25.33** contains two batteries, each with an emf and an internal resistance, and two resistors. Find (a) the current in the circuit (magnitude and direction) and (b) the terminal voltage V_{ab} of the 16.0-V battery.

Figure **E25.33**



Section 25.5 Energy and Power in Electric Circuits

25.34 •• When a resistor with resistance R is connected to a 1.50-V flashlight battery, the resistor consumes 0.0625 W of electrical power. (Throughout, assume that each battery has negligible internal resistance.) (a) What power does the resistor consume if it is connected to a 12.6-V car battery? Assume that R remains constant when the power consumption changes. (b) The resistor is connected to a battery and consumes 5.00 W. What is the voltage of this battery?

25.35 • **Light Bulbs.** The power rating of a light bulb (such as a 100-W bulb) is the power it dissipates when connected across a 120-V potential difference. What is the resistance of (a) a 100-W bulb and (b) a 60-W bulb? (c) How much current does each bulb draw in normal use?

25.36 • If a "75-W" bulb (see Problem 25.35) is connected across a 220-V potential difference (as is used in Europe), how much power does it dissipate? Ignore the temperature dependence of the bulb's resistance.

25.37 • **European Light Bulb.** In Europe the standard voltage in homes is 220 V instead of the 120 V used in the United States. Therefore a "100-W" European bulb would be intended for use with a 220-V potential difference (see Problem 25.36). (a) If you bring a "100-W" European bulb home to the United States, what should be its U.S. power rating? (b) How much current will the 100-W European bulb draw in normal use in the United States?

25.38 • A battery-powered global positioning system (GPS) receiver operating on 9.0 V draws a current of 0.13 A. How much electrical energy does it consume during 30 minutes?

25.39 • Consider the circuit of Fig. E25.30. (a) What is the total rate at which electrical energy is dissipated in the 5.0-Ω and 9.0-Ω resistors? (b) What is the power output of the 16.0-V battery? (c) At what rate is electrical energy being converted to other forms in the 8.0-V battery? (d) Show that the power output of the 16.0-V battery equals the overall rate of consumption of electrical energy in the rest of the circuit.

25.40 • BIO **Electric Eels.** Electric eels generate electric pulses along their skin that can be used to stun an enemy when they come into contact with it. Tests have shown that these pulses can be up to 500 V and produce currents of 80 mA (or even larger). A typical pulse lasts for 10 ms. What power and how much energy are delivered to the unfortunate enemy with a single pulse, assuming a steady current?

25.41 • BIO **Treatment of Heart Failure.** A heart defibrillator is used to enable the heart to start beating if it has stopped. This is done by passing a large current of 12 A through the body at 25 V for a very short time, usually about 3.0 ms. (a) What power does the defibrillator deliver to the body, and (b) how much energy is transferred?

25.42 •• The battery for a certain cell phone is rated at 3.70 V. According to the manufacturer it can produce 3.15×10^4 J of electrical energy, enough for 5.25 h of operation, before needing to be recharged. Find the average current that this cell phone draws when turned on.

25.43 •• The capacity of a storage battery, such as those used in automobile electrical systems, is rated in ampere-hours $(A \cdot h)$. A 50-A · h battery can supply a current of 50 A for 1.0 h, or 25 A for 2.0 h, and so on. (a) What total energy can be supplied by a 12-V, 60-A · h battery if its internal resistance is negligible? (b) What volume (in liters) of gasoline has a total heat of combustion equal to the energy obtained in part (a)? (See Section 17.6; the density of gasoline is 900 kg/m^3.) (c) If a generator with an average electrical power output of 0.45 kW is connected to the battery, how much time will be required for it to charge the battery fully?

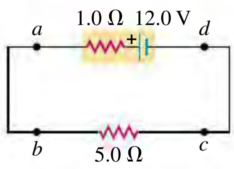
25.44 • An idealized voltmeter is connected across the terminals of a 15.0-V battery, and a 75.0-Ω appliance is also connected across its terminals. If the voltmeter reads 11.9 V, (a) how much power is being dissipated by the appliance, and (b) what is the internal resistance of the battery?

25.45 •• A 25.0-Ω bulb is connected across the terminals of a 12.0-V battery having 3.50 Ω of internal resistance. What percentage of the power of the battery is dissipated across the internal resistance and hence is not available to the bulb?

25.46 •• A typical small flashlight contains two batteries, each having an emf of 1.5 V, connected in series with a bulb having resistance 17 Ω. (a) If the internal resistance of the batteries is negligible, what power is delivered to the bulb? (b) If the batteries last for 5.0 h, what is the total energy delivered to the bulb? (c) The resistance of real batteries increases as they run down. If the initial internal resistance is negligible, what is the combined internal resistance of both batteries when the power to the bulb has decreased to half its initial value? (Assume that the resistance of the bulb is constant. Actually, it will change somewhat when the current through the filament changes, because this changes the temperature of the filament and hence the resistivity of the filament wire.)

25.47 • In the circuit in **Fig. E25.47**, find (a) the rate of conversion of internal (chemical) energy to electrical energy within the battery; (b) the rate of dissipation of electrical energy in the battery; (c) the rate of dissipation of electrical energy in the external resistor.

Figure **E25.47**



25.48 • A "540-W" electric heater is designed to operate from 120-V lines. (a) What is its operating resistance? (b) What current does it draw? (c) If the line voltage drops to 110 V, what power does the heater take? (Assume that the resistance is constant. Actually, it will change because of the change in temperature.) (d) The heater coils are metallic, so that the resistance of the heater decreases with decreasing temperature. If the change of resistance with temperature is taken into account, will the electrical power consumed by the heater be larger or smaller than what you calculated in part (c)? Explain.

Section 25.6 Theory of Metallic Conduction

25.49 •• Pure silicon at room temperature contains approximately 1.0×10^{16} free electrons per cubic meter. (a) Referring to Table 25.1, calculate the mean free time τ for silicon at room temperature. (b) Your answer in part (a) is much greater than the mean free time for copper given in Example 25.11. Why, then, does pure silicon have such a high resistivity compared to copper?

PROBLEMS

25.50 •• In an ionic solution, a current consists of Ca^{2+} ions (of charge $+2e$) and Cl^- ions (of charge $-e$) traveling in opposite directions. If 5.11×10^{18} Cl^- ions go from A to B every 0.50 min, while 3.24×10^{18} Ca^{2+} ions move from B to A, what is the current (in mA) through this solution, and in which direction (from A to B or from B to A) is it going?

25.51 • An electrical conductor designed to carry large currents has a circular cross section 2.50 mm in diameter and is 14.0 m long. The resistance between its ends is 0.104 Ω. (a) What is the resistivity of the material? (b) If the electric-field magnitude in the conductor is 1.28 V/m, what is the total current? (c) If the material has 8.5×10^{28} free electrons per cubic meter, find the average drift speed under the conditions of part (b).

25.52 •• An overhead transmission cable for electrical power is 2000 m long and consists of two parallel copper wires, each encased in insulating material. A short circuit has developed somewhere along the length of the cable where the insulation has worn thin and the two wires are in contact. As a power-company employee, you must locate the short so that repair crews can be sent to that location. Both ends of the cable have been disconnected from the power grid. At one end of the cable (point A), you connect the ends of the two wires to a 9.00-V battery that has negligible internal resistance and measure that 2.86 A of current flows through the battery. At the other end of the cable (point B), you attach those two wires to the battery and measure that 1.65 A of current flows through the battery. How far is the short from point A?

25.53 •• On your first day at work as an electrical technician, you are asked to determine the resistance per meter of a long piece of wire. The company you work for is poorly equipped. You find a battery, a voltmeter, and an ammeter, but no meter for directly measuring resistance (an ohmmeter). You put the leads from the voltmeter across the terminals of the battery, and the meter reads 12.6 V. You cut off a 20.0-m length of wire and connect it to the battery, with an ammeter in series with it to measure the current in the wire. The ammeter reads 7.00 A. You then cut off a 40.0-m length of wire and connect it to the battery, again with the ammeter in series to measure the current. The ammeter reads 4.20 A. Even though the equipment you have available to you is limited, your boss assures you of its high quality: The ammeter has very small resistance, and the voltmeter has very large resistance. What is the resistance of 1 meter of wire?

25.54 • A 2.0-m length of wire is made by welding the end of a 120-cm-long silver wire to the end of an 80-cm-long copper wire. Each piece of wire is 0.60 mm in diameter. The wire is at room temperature, so the resistivities are as given in Table 25.1. A potential difference of 9.0 V is maintained between the ends of the 2.0-m composite wire. What is (a) the current in the copper section; (b) the current in the silver section; (c) the magnitude of $\vec{E}$ in the copper; (d) the magnitude of $\vec{E}$ in the silver; (e) the potential difference between the ends of the silver section of wire?

25.55 • A 3.00-m length of copper wire at 20° C has a 1.20-m-long section with diameter 1.60 mm and a 1.80-m-long section with diameter 0.80 mm. There is a current of 2.5 mA in the 1.60-mm-diameter section. (a) What is the current in the 0.80-mm-diameter section? (b) What is the magnitude of $\vec{E}$ in the 1.60-mm-diameter section? (c) What is the magnitude of $\vec{E}$ in the 0.80-mm-diameter section? (d) What is the potential difference between the ends of the 3.00-m length of wire?

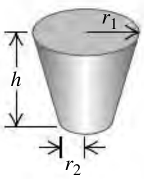
25.56 •• A heating element made of tungsten wire is connected to a large battery that has negligible internal resistance. When the heating element reaches 80.0° C, it consumes electrical energy at a rate of 480 W. What is its power consumption when its temperature is 150.0° C? Assume that the temperature coefficient of resistivity has the value given in Table 25.2 and that it is constant over the temperature range in this problem. In Eq. (25.12) take T_0 to be 20.0° C.

25.57 •• CP BIO Struck by Lightning. Lightning strikes can involve currents as high as 25,000 A that last for about 40 μs. If a person is struck by a bolt of lightning with these properties, the current will pass through his body. We shall assume that his mass is 75 kg, that he is wet (after all, he is in a rainstorm) and therefore has a resistance of 1.0 kΩ, and that his body is all water (which is reasonable for a rough, but plausible, approximation). (a) By how many degrees Celsius would this lightning bolt increase the temperature of 75 kg of water? (b) Given that the internal body temperature is about 37°C, would the person's temperature actually increase that much? Why not? What would happen first?

25.58 •• A resistor with resistance R is connected to a battery that has emf 12.0 V and internal resistance $r = 0.40\ \Omega$. For what two values of R will the power dissipated in the resistor be 80.0 W?

25.59 • CALC A material of resistivity ρ is formed into a solid, truncated cone of height h and radii r_1 and r_2 at either end (**Fig. P25.59**). (a) Calculate the resistance of the cone between the two flat end faces. (*Hint:* Imagine slicing the cone into very many thin disks, and calculate the resistance of one such disk.) (b) Show that your result agrees with Eq. (25.10) when $r_1 = r_2$.

Figure **P25.59**



25.60 • CALC The region between two concentric conducting spheres with radii a and b is filled with a conducting material with resistivity ρ. (a) Show that the resistance between the spheres is given by

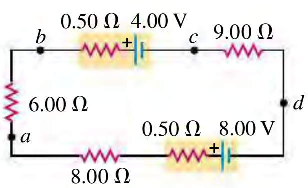
$$R = \frac{\rho}{4\pi}\left(\frac{1}{a} - \frac{1}{b}\right)$$

(b) Derive an expression for the current density as a function of radius, in terms of the potential difference V_{ab} between the spheres. (c) Show that the result in part (a) reduces to Eq. (25.10) when the separation $L = b - a$ between the spheres is small.

25.61 • The potential difference across the terminals of a battery is 8.40 V when there is a current of 1.50 A in the battery from the negative to the positive terminal. When the current is 3.50 A in the reverse direction, the potential difference becomes 10.20 V. (a) What is the internal resistance of the battery? (b) What is the emf of the battery?

25.62 • (a) What is the potential difference V_{ad} in the circuit of **Fig. P25.62**? (b) What is the terminal voltage of the 4.00-V battery? (c) A battery with emf 10.30 V and internal resistance 0.50 Ω is inserted in the circuit at d, with its negative terminal connected to the negative terminal of the 8.00-V battery. What is the difference of potential V_{bc} between the terminals of the 4.00-V battery now?

Figure **P25.62**



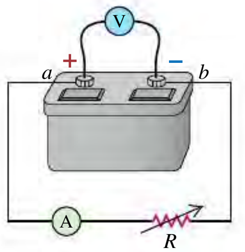
25.63 • BIO The average bulk resistivity of the human body (apart from surface resistance of the skin) is about 5.0 $\Omega \cdot$m. The conducting path between the hands can be represented approximately as a cylinder 1.6 m long and 0.10 m in diameter. The skin resistance can be made negligible by soaking the hands in salt water. (a) What is the resistance between the hands if the skin resistance is negligible? (b) What potential difference between the hands is needed for a lethal shock current of 100 mA? (Note that your result shows that small potential differences produce dangerous currents when the skin is damp.) (c) With the current in part (b), what power is dissipated in the body?

25.64 •• BIO A person with body resistance between his hands of 10 kΩ accidentally grasps the terminals of a 14-kV power supply. (a) If the internal resistance of the power supply is 2000 Ω, what is the current through the person's body? (b) What is the power dissipated in his body? (c) If the power supply is to be made safe by increasing its internal resistance, what should the internal resistance be for the maximum current in the above situation to be 1.00 mA or less?

25.65 • A typical cost for electrical power is \$0.120 per kilowatt-hour. (a) Some people leave their porch light on all the time. What is the yearly cost to keep a 75-W bulb burning day and night? (b) Suppose your refrigerator uses 400 W of power when it's running, and it runs 8 hours a day. What is the yearly cost of operating your refrigerator?

25.66 •• In the circuit shown in **Fig. P25.66**, R is a *variable resistor* whose value ranges from 0 to ∞, and a and b are the terminals of a battery that has an emf $\mathcal{E} = 15.0$ V and an internal resistance of 4.00 Ω. The ammeter and voltmeter are idealized meters. As R varies over its full range of values, what will be the largest and smallest readings of (a) the voltmeter and (b) the ammeter? (c) Sketch qualitative graphs of the readings of both meters as functions of R.

Figure **P25.66**



25.67 •• A Nonideal Ammeter. Unlike the idealized ammeter described in Section 25.4, any real ammeter has a nonzero resistance. (a) An ammeter with resistance R_A is connected in series with a resistor R and a battery of emf $\mathcal{E}$ and internal resistance r. The current measured by the ammeter is I_A. Find the current through the circuit if the ammeter is removed so that the battery and the resistor form a complete circuit. Express your answer in terms of I_A, r, R_A, and R. The more "ideal" the ammeter, the smaller the difference between this current and the current I_A. (b) If $R = 3.80\ \Omega$, $\mathcal{E} = 7.50$ V, and $r = 0.45\ \Omega$, find the maximum value of the ammeter resistance R_A so that I_A is within 1.0% of the current in the circuit when the ammeter is absent. (c) Explain why your answer in part (b) represents a *maximum* value.

25.68 •• A cylindrical copper cable 1.50 km long is connected across a 220.0-V potential difference. (a) What should be its diameter so that it produces heat at a rate of 90.0 W? (b) What is the electric field inside the cable under these conditions?

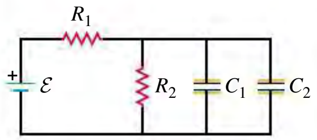
25.69 • CALC A 1.50-m cylinder of radius 1.10 cm is made of a complicated mixture of materials. Its resistivity depends on the distance x from the left end and obeys the formula $\rho(x) = a + bx^2$, where a and b are constants. At the left end, the resistivity is $2.25 \times 10^{-8}\ \Omega \cdot$m, while at the right end it is $8.50 \times 10^{-8}\ \Omega \cdot$m. (a) What is the resistance of this rod? (b) What is the electric field at its midpoint if it carries a 1.75-A current? (c) If we cut the rod into two 75.0-cm halves, what is the resistance of each half?

25.70 •• Compact Fluorescent Bulbs. Compact fluorescent bulbs are much more efficient at producing light than are ordinary incandescent bulbs. They initially cost much more, but they last far longer and use much less electricity. According to one study of these bulbs, a compact bulb that produces as much light as a 100-W incandescent bulb uses only 23 W of power. The compact bulb lasts 10,000 hours, on the average, and costs \$11.00, whereas the incandescent bulb costs only \$0.75, but lasts just 750 hours. The study assumed that electricity costs \$0.080 per kilowatt-hour and that the bulbs are on for 4.0 h per day. (a) What is the total cost (including the price of the bulbs) to run each bulb for 3.0 years? (b) How much do you save over 3.0 years if you use a compact fluorescent bulb instead of an incandescent bulb? (c) What is the resistance of a "100-W" fluorescent bulb? (Remember, it actually uses only 23 W of power and operates across 120 V.)

25.71 • A lightning bolt strikes one end of a steel lightning rod, producing a 15,000-A current burst that lasts for 65 μs. The rod is 2.0 m long and 1.8 cm in diameter, and its other end is connected to the ground by 35 m of 8.0-mm-diameter copper wire. (a) Find the potential difference between the top of the steel rod and the lower end of the copper wire during the current burst. (b) Find the total energy deposited in the rod and wire by the current burst.

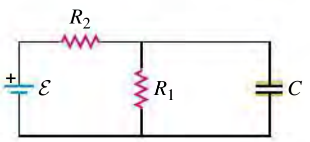
25.72 •• **CP** Consider the circuit shown in **Fig. P25.72**. The battery has emf 72.0 V and negligible internal resistance. $R_2 = 2.00\ \Omega$, $C_1 = 3.00\ \mu$F, and $C_2 = 6.00\ \mu$F. After the capacitors have attained their final charges, the charge on C_1 is $Q_1 = 18.0\ \mu$C. What is (a) the final charge on C_2; (b) the resistance R_1?

Figure **P25.72**



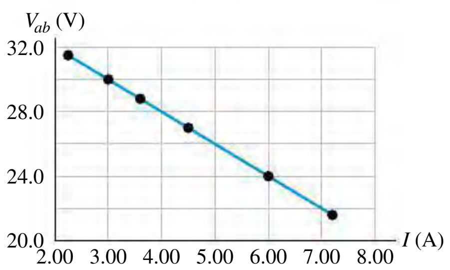
25.73 •• **CP** Consider the circuit shown in **Fig. P25.73**. The emf source has negligible internal resistance. The resistors have resistances $R_1 = 6.00\ \Omega$ and $R_2 = 4.00\ \Omega$. The capacitor has capacitance $C = 9.00\ \mu$F. When the capacitor is fully charged, the magnitude of the charge on its plates is $Q = 36.0\ \mu$C. Calculate the emf $\mathcal{E}$.

Figure **P25.73**



25.74 •• **DATA** An external resistor R is connected between the terminals of a battery. The value of R varies. For each R value, the current I in the circuit and the terminal voltage V_{ab} of the battery are measured. The results are plotted in **Fig. P25.74**, a graph of V_{ab} versus I that shows the best straight-line fit to the data. (a) Use the graph in Fig. P25.74 to calculate the battery's emf and internal resistance. (b) For what value of R is V_{ab} equal to 80.0% of the battery emf?

Figure **P25.74**



25.75 •• **DATA** The voltage drop V_{ab} across each of resistors A and B was measured as a function of the current I in the resistor. The results are shown in the table:

Resistor *A*				
I (A)	0.50	1.00	2.00	4.00
V_{ab} (V)	2.55	3.11	3.77	4.58
Resistor *B*				
I (A)	0.50	1.00	2.00	4.00
V_{ab} (V)	1.94	3.88	7.76	15.52

(a) For each resistor, graph V_{ab} as a function of I and graph the resistance $R = V_{ab}/I$ as a function of I. (b) Does resistor A obey Ohm's law? Explain. (c) Does resistor B obey Ohm's law? Explain. (d) What is the power dissipated in A if it is connected to a 4.00-V battery that has negligible internal resistance? (e) What is the power dissipated in B if it is connected to the battery?

25.76 •• **DATA** According to the U.S. National Electrical Code, copper wire used for interior wiring of houses, hotels, office buildings, and industrial plants is permitted to carry no more than a specified maximum amount of current. The table shows values of the maximum current I_{max} for several common sizes of wire with varnished cambric insulation. The "wire gauge" is a standard used to describe the diameter of wires. Note that the larger the diameter of the wire, the *smaller* the wire gauge.

Wire gauge	Diameter (cm)	I_{max} (A)
14	0.163	18
12	0.205	25
10	0.259	30
8	0.326	40
6	0.412	60
5	0.462	65
4	0.519	85

(a) What considerations determine the maximum current-carrying capacity of household wiring? (b) A total of 4200 W of power is to be supplied through the wires of a house to the household electrical appliances. If the potential difference across the group of appliances is 120 V, determine the gauge of the thinnest permissible wire that can be used. (c) Suppose the wire used in this house is of the gauge found in part (b) and has total length 42.0 m. At what rate is energy dissipated in the wires? (d) The house is built in a community where the consumer cost of electrical energy is \$0.11 per kilowatt-hour. If the house were built with wire of the next larger diameter than that found in part (b), what would be the savings in electricity costs in one year? Assume that the appliances are kept on for an average of 12 hours a day.

CHALLENGE PROBLEMS

25.77 ••• **CALC** The resistivity of a semiconductor can be modified by adding different amounts of impurities. A rod of semiconducting material of length L and cross-sectional area A lies along the x-axis between $x = 0$ and $x = L$. The material obeys Ohm's law, and its resistivity varies along the rod according to $\rho(x) = \rho_0 \exp(-x/L)$. The end of the rod at $x = 0$ is at a potential V_0 greater than the end at $x = L$. (a) Find the total resistance of the rod and the current in the rod. (b) Find the electric-field magnitude $E(x)$ in the rod as a function of x. (c) Find the electric potential $V(x)$ in the rod as a function of x. (d) Graph the functions $\rho(x)$, $E(x)$, and $V(x)$ for values of x between $x = 0$ and $x = L$.

25.78 ••• An external resistor with resistance R is connected to a battery that has emf $\mathcal{E}$ and internal resistance r. Let P be the electrical power output of the source. By conservation of energy, P is equal to the power consumed by R. What is the value of P in the limit that R is (a) very small; (b) very large? (c) Show that the power output of the battery is a maximum when $R = r$. What is this maximum P in terms of $\mathcal{E}$ and r? (d) A battery has $\mathcal{E} = 64.0$ V and $r = 4.00\ \Omega$. What is the power output of this battery when it is connected to a resistor R, for $R = 2.00\ \Omega$, $R = 4.00\ \Omega$, and $R = 6.00\ \Omega$? Are your results consistent with the general result that you derived in part (b)?

PASSAGE PROBLEMS

BIO SPIDERWEB CONDUCTIVITY. Some types of spiders build webs that consist of threads made of dry silk coated with a solution of a variety of compounds. This coating leaves the threads, which are used to capture prey, *hygroscopic*—that is, they attract water from the atmosphere. It has been hypothesized that this aqueous coating makes the threads good electrical conductors. To test the electrical properties of coated thread, researchers placed a 5-mm length of thread between two electrical contacts.* The researchers stretched the thread in 1-mm increments to more than twice its original length, and then allowed it to return to its original length, again in 1-mm increments. Some of the resistance measurements are shown in the table:

Resistance of thread ($10^9\ \Omega$)	9	19	41	63	102	76	50	24
Length of thread (mm)	5	7	9	11	13	9	7	5

*Based on F. Vollrath and D. Edmonds, "Consequences of electrical conductivity in an orb spider's capture web," *Naturwissenschaften* (100:12, December 2013, pp. 1163–69).

25.79 What is the best explanation for the behavior exhibited in the data? (a) Longer threads can carry more current than shorter threads do and so make better electrical conductors. (b) The thread stops being a conductor when it is stretched to 13 mm, due to breaks that occur in the thin coating. (c) As the thread is stretched, the coating thins and its resistance increases; as the thread is relaxed, the coating returns nearly to its original state. (d) The resistance of the thread increases with distance from the end of the thread.

25.80 If the conductivity of the thread results from the aqueous coating only, how does the cross-sectional area A of the coating compare when the thread is 13 mm long versus the starting length of 5 mm? Assume that the resistivity of the coating remains constant and the coating is uniform along the thread. $A_{13\text{ mm}}$ is about (a) $\frac{1}{10}A_{5\text{ mm}}$; (b) $\frac{1}{4}A_{5\text{ mm}}$; (c) $\frac{2}{5}A_{5\text{ mm}}$; (d) the same as $A_{5\text{ mm}}$.

25.81 What is the maximum current that flows in the thread during this experiment if the voltage source is a 9-V battery? (a) about 1 A; (b) about 0.1 A; (c) about 1 μA; (d) about 1 nA.

25.82 In another experiment, a piece of the web is suspended so that it can move freely. When either a positively charged object or a negatively charged object is brought near the web, the thread is observed to move toward the charged object. What is the best interpretation of this observation? The web is (a) a negatively charged conductor; (b) a positively charged conductor; (c) either a positively or negatively charged conductor; (d) an electrically neutral conductor.

Answers

Chapter Opening Question ?

(iii) The current out equals the current in. In other words, charge must enter the bulb at the same rate as it exits the bulb. It is not "used up" or consumed as it flows through the bulb.

Test Your Understanding Questions

25.1 (v) Doubling the diameter increases the cross-sectional area A by a factor of 4. Hence the current-density magnitude $J = I/A$ is reduced to $\frac{1}{4}$ of the value in Example 25.1, and the magnitude of the drift velocity $v_d = J/n|q|$ is reduced by the same factor. The new magnitude is $v_d = (0.15\text{ mm/s})/4 = 0.038\text{ mm/s}$. This behavior is the same as that of an incompressible fluid, which slows down when it moves from a narrow pipe to a broader one (see Section 12.4).

25.2 (ii) Figure 25.6b shows that the resistivity ρ of a semiconductor increases as the temperature decreases. From Eq. (25.5), the magnitude of the current density is $J = E/\rho$, so the current density decreases as the temperature drops and the resistivity increases.

25.3 (iii) Solving Eq. (25.11) for the current shows that $I = V/R$. If the resistance R of the wire remained the same, doubling the voltage V would make the current I double as well. However, we saw in Example 25.3 that the resistance is *not* constant: As the current increases and the temperature increases, R increases as well. Thus doubling the voltage produces a current that is *less* than double the original current. An ohmic conductor is one for which $R = V/I$ has the same value no matter what the voltage, so the wire is *nonohmic*. (In many practical problems the temperature change of the wire is so small that it can be ignored, so we can safely regard the wire as being ohmic. We do so in almost all examples in this book.)

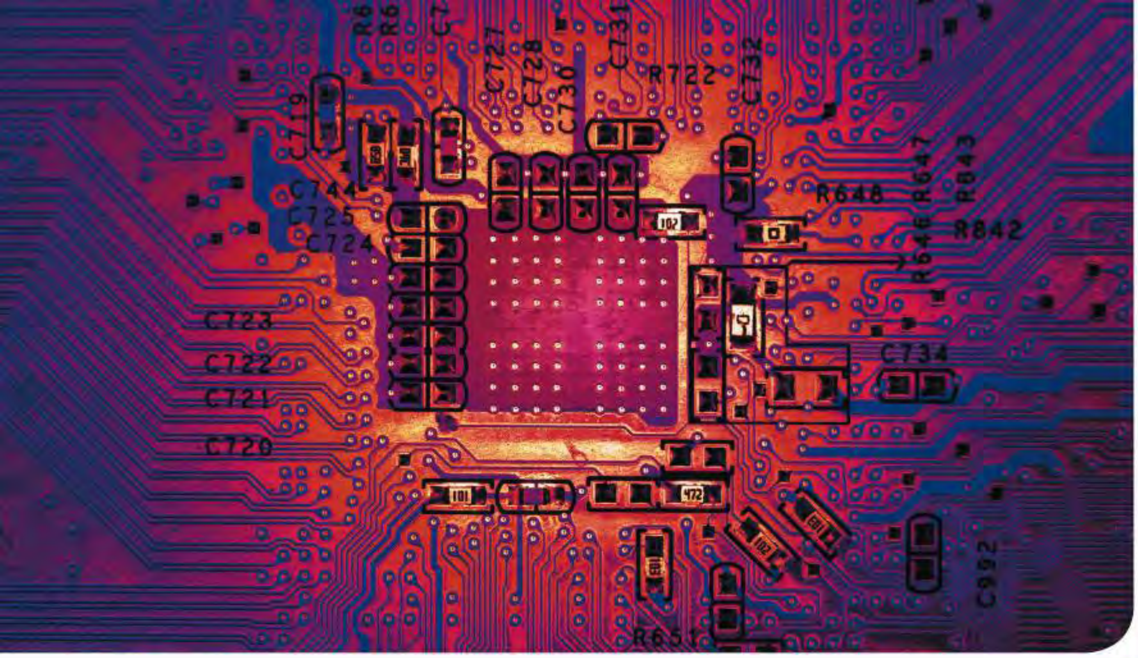
25.4 (iii), (ii), (i) For circuit (i), we find the current from Eq. (25.16): $I = \mathcal{E}/(R + r) = (1.5\text{ V})/(1.4\ \Omega + 0.10\ \Omega) = 1.0\text{ A}$. For circuit (ii), we note that the terminal voltage $v_{ab} = 3.6\text{ V}$ equals the voltage IR across the 1.8-Ω resistor: $V_{ab} = IR$, so $I = V_{ab}/R = (3.6\text{ V})/(1.8\ \Omega) = 2.0\text{ A}$. For circuit (iii), we use Eq. (25.15) for the terminal voltage: $V_{ab} = \mathcal{E} - Ir$, so $I = (\mathcal{E} - V_{ab})/r = (12.0\text{ V} - 11.0\text{ V})/(0.20\ \Omega) = 5.0\text{ A}$.

25.5 (iii), (ii), (i) These are the same circuits that we analyzed in Test Your Understanding of Section 25.4. In each case the net power output of the battery is $P = V_{ab}I$, where V_{ab} is the battery terminal voltage. For circuit (i), we found that $I = 1.0\text{ A}$, so $V_{ab} = \mathcal{E} - Ir = 1.5\text{ V} - (1.0\text{ A})(0.10\ \Omega) = 1.4\text{ V}$, so $P = (1.4\text{ V})(1.0\text{ A}) = 1.4\text{ W}$. For circuit (ii), we have $V_{ab} = 3.6\text{ V}$ and found that $I = 2.0\text{ A}$, so $P = (3.6\text{ V})(2.0\text{ A}) = 7.2\text{ W}$. For circuit (iii), we have $V_{ab} = 11.0\text{ V}$ and found that $I = 5.0\text{ A}$, so $P = (11.0\text{ V})(5.0\text{ A}) = 55\text{ W}$.

25.6 (i) The difficulty of producing a certain amount of current increases as the resistivity ρ increases. From Eq. (25.24), $\rho = m/ne^2\tau$, so increasing the mass m will increase the resistivity. That's because a more massive charged particle will respond more sluggishly to an applied electric field and hence drift more slowly. To produce the same current, a greater electric field would be needed. (Increasing n, e, or τ would decrease the resistivity and make it easier to produce a given current.)

Bridging Problem

(a) 237°C **(b)** 162 W initially, 148 W at 1.23 A



? In a complex circuit like the one on this circuit board, is it possible to connect several resistors with different resistances so that all of them have the same potential difference? (i) Yes, and the current will be the same through all of the resistors; (ii) yes, but the current may be different through different resistors; (iii) no; (iv) the answer depends on the value of the potential difference.

26 DIRECT-CURRENT CIRCUITS

LEARNING GOALS

Looking forward at …

26.1 How to analyze circuits with multiple resistors in series or parallel.

26.2 Rules that you can apply to any circuit with more than one loop.

26.3 How to use an ammeter, voltmeter, ohmmeter, or potentiometer in a circuit.

26.4 How to analyze circuits that include both a resistor and a capacitor.

26.5 How electric power is distributed in the home.

Looking back at …

24.2 Capacitors in series and parallel.

25.4 Current, ammeters, and voltmeters.

25.5 Power in a circuit.

If you look inside a mobile phone, a computer, or under the hood of a car, you will find circuits of much greater complexity than the simple circuits we studied in Chapter 25. Whether connected by wires or integrated in a semiconductor chip, these circuits often include several sources, resistors, and other circuit elements interconnected in a *network*.

In this chapter we study general methods for analyzing such networks, including how to find voltages and currents of circuit elements. We'll learn how to determine the equivalent resistance for several resistors connected in series or in parallel. For more general networks we need two rules called *Kirchhoff's rules*. One is based on the principle of conservation of charge applied to a junction; the other is derived from energy conservation for a charge moving around a closed loop. We'll discuss instruments for measuring various electrical quantities. We'll also look at a circuit containing resistance and capacitance, in which the current varies with time.

Our principal concern in this chapter is with **direct-current** (dc) circuits, in which the direction of the current does not change with time. Flashlights and automobile wiring systems are examples of direct-current circuits. Household electrical power is supplied in the form of **alternating current** (ac), in which the current oscillates back and forth. The same principles for analyzing networks apply to both kinds of circuits, and we conclude this chapter with a look at household wiring systems. We'll discuss alternating-current circuits in detail in Chapter 31.

26.1 RESISTORS IN SERIES AND PARALLEL

Resistors turn up in all kinds of circuits, ranging from hair dryers and space heaters to circuits that limit or divide current or reduce or divide a voltage. Such circuits often contain several resistors, so it's appropriate to consider *combinations* of resistors. A simple example is a string of light bulbs used for holiday decorations; each bulb acts as a resistor, and from a circuit-analysis perspective the string of bulbs is simply a combination of resistors.

Suppose we have three resistors with resistances R_1, R_2, and R_3. **Figure 26.1** shows four different ways in which they might be connected between points a and b. When several circuit elements such as resistors, batteries, and motors are connected in sequence as in Fig. 26.1a, with only a single current path between the points, we say that they are connected in **series**. We studied *capacitors* in series in Section 24.2; we found that, because of conservation of charge, capacitors in series all have the same charge if they are initially uncharged. In circuits we're often more interested in the *current*, which is charge flow per unit time.

The resistors in Fig. 26.1b are said to be connected in **parallel** between points a and b. Each resistor provides an alternative path between the points. For circuit elements that are connected in parallel, the *potential difference* is the same across each element. We studied capacitors in parallel in Section 24.2.

In Fig. 26.1c, resistors R_2 and R_3 are in parallel, and this combination is in series with R_1. In Fig. 26.1d, R_2 and R_3 are in series, and this combination is in parallel with R_1.

For any combination of resistors we can always find a *single* resistor that could replace the combination and result in the same total current and potential difference. For example, a string of holiday light bulbs could be replaced by a single, appropriately chosen light bulb that would draw the same current and have the same potential difference between its terminals as the original string of bulbs. The resistance of this single resistor is called the **equivalent resistance** of the combination. If any one of the networks in Fig. 26.1 were replaced by its equivalent resistance R_{eq}, we could write

$$V_{ab} = IR_{eq} \quad \text{or} \quad R_{eq} = \frac{V_{ab}}{I}$$

where V_{ab} is the potential difference between terminals a and b of the network and I is the current at point a or b. To compute an equivalent resistance, we assume a potential difference V_{ab} across the actual network, compute the corresponding current I, and take the ratio V_{ab}/I.

Resistors in Series

We can derive general equations for the equivalent resistance of a series or parallel combination of resistors. In Fig. 26.1a, the three resistors are in *series,* so the current I is the same in all of them. (Recall from Section 25.4 that current is *not* "used up" as it passes through a circuit.) Applying $V = IR$ to each resistor, we have

$$V_{ax} = IR_1 \qquad V_{xy} = IR_2 \qquad V_{yb} = IR_3$$

The potential differences across each resistor need not be the same (except for the special case in which all three resistances are equal). The potential difference V_{ab} across the entire combination is the sum of these individual potential differences:

$$V_{ab} = V_{ax} + V_{xy} + V_{yb} = I(R_1 + R_2 + R_3)$$

and so

$$\frac{V_{ab}}{I} = R_1 + R_2 + R_3$$

The ratio V_{ab}/I is, by definition, the equivalent resistance R_{eq}. Therefore

$$R_{eq} = R_1 + R_2 + R_3$$

It is easy to generalize this to *any* number of resistors:

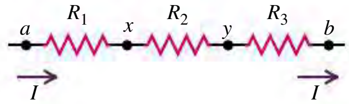
Resistors in series: $R_{eq} = R_1 + R_2 + R_3 + \cdots$ (26.1)

Equivalent resistance of series combination

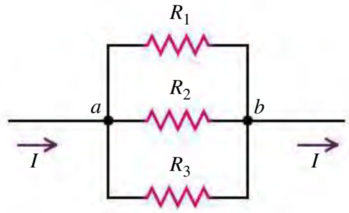
Resistances of individual resistors

26.1 Four different ways of connecting three resistors.

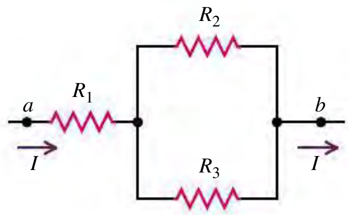
(a) R_1, R_2, and R_3 in series



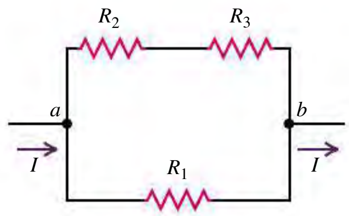
(b) R_1, R_2, and R_3 in parallel



(c) R_1 in series with parallel combination of R_2 and R_3



(d) R_1 in parallel with series combination of R_2 and R_3



The equivalent resistance of a series combination equals the sum of the individual resistances. The equivalent resistance is *greater than* any individual resistance.

CAUTION **Resistors vs. capacitors in series** Don't confuse *resistors* in series with *capacitors* in series. Resistors in series add *directly* [Eq. (26.1)] because the voltage across each is directly proportional to its resistance and to the common current. Capacitors in series [Eq. (24.5)] add *reciprocally;* the voltage across each is directly proportional to the common charge but *inversely* proportional to the individual capacitance.

Resistors in Parallel

If the resistors are in *parallel,* as in Fig. 26.1b, the current through each resistor need not be the same. But the potential difference between the terminals of each resistor must be the same and equal to V_{ab} (**Fig. 26.2**). (Remember that the potential difference between any two points does not depend on the path taken between the points.) Let's call the currents in the three resistors I_1, I_2, and I_3. Then from $I = V/R$,

26.2 A car's headlights and taillights are connected in parallel. Hence each light is exposed to the full potential difference supplied by the car's electrical system, giving maximum brightness. Another advantage is that if one headlight or taillight burns out, the other one keeps shining (see Example 26.2).

$$I_1 = \frac{V_{ab}}{R_1} \qquad I_2 = \frac{V_{ab}}{R_2} \qquad I_3 = \frac{V_{ab}}{R_3}$$

In general, the current is different through each resistor. Because charge neither accumulates at nor drains out of point a, the total current I must equal the sum of the three currents in the resistors:

$$I = I_1 + I_2 + I_3 = V_{ab}\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) \quad \text{or}$$

$$\frac{I}{V_{ab}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

But by the definition of the equivalent resistance R_{eq}, $I/V_{ab} = 1/R_{eq}$, so

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

Again it is easy to generalize to *any* number of resistors in parallel:

Resistors in parallel:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \qquad (26.2)$$

Equivalent resistance of parallel combination

Resistances of individual resistors

DEMO

The reciprocal of the equivalent resistance of a parallel combination equals the sum of the reciprocals of the individual resistances. The equivalent resistance is always *less than* any individual resistance.

CAUTION **Resistors vs. capacitors in parallel** Note the differences between *resistors* in parallel and *capacitors* in parallel. Resistors in parallel add *reciprocally* [Eq. (26.2)] because the current in each is proportional to the common voltage across them and *inversely* proportional to the resistance of each. Capacitors in parallel add *directly* [Eq. (24.7)] because the charge on each is proportional to the common voltage across them and *directly* proportional to the capacitance of each.

For the special case of *two* resistors in parallel,

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_1 + R_2}{R_1 R_2} \quad \text{and}$$

$$R_{eq} = \frac{R_1 R_2}{R_1 + R_2} \quad \text{(two resistors in parallel)} \tag{26.3}$$

Because $V_{ab} = I_1 R_1 = I_2 R_2$, it follows that

$$\frac{I_1}{I_2} = \frac{R_2}{R_1} \quad \text{(two resistors in parallel)} \tag{26.4}$$

Thus the currents carried by two resistors in parallel are *inversely proportional* to their resistances. More current goes through the path of least resistance.

PROBLEM-SOLVING STRATEGY 26.1 RESISTORS IN SERIES AND PARALLEL

IDENTIFY *the relevant concepts:* As in Fig. 26.1, many resistor networks are made up of resistors in series, in parallel, or a combination thereof. Such networks can be replaced by a single equivalent resistor. The logic is similar to that of Problem-Solving Strategy 24.1 for networks of capacitors.

SET UP *the problem* using the following steps:
1. Make a drawing of the resistor network.
2. Identify groups of resistors connected in series or parallel.
3. Identify the target variables. They could include the equivalent resistance of the network, the potential difference across each resistor, or the current through each resistor.

EXECUTE *the solution* as follows:
1. Use Eq. (26.1) or (26.2), respectively, to find the equivalent resistance for series or parallel combinations.
2. If the network is more complex, try reducing it to series and parallel combinations. For example, in Fig. 26.1c we first replace the parallel combination of R_2 and R_3 with its equivalent resistance; this then forms a series combination with R_1. In Fig. 26.1d, the combination of R_2 and R_3 in series forms a parallel combination with R_1.
3. Keep in mind that the total potential difference across resistors connected in series is the sum of the individual potential differences. The potential difference across resistors connected in parallel is the same for every resistor and equals the potential difference across the combination.
4. The current through resistors connected in series is the same through every resistor and equals the current through the combination. The total current through resistors connected in parallel is the sum of the currents through the individual resistors.

EVALUATE *your answer:* Check whether your results are consistent. The equivalent resistance of resistors connected in series should be greater than that of any individual resistor; that of resistors in parallel should be less than that of any individual resistor.

EXAMPLE 26.1 EQUIVALENT RESISTANCE

SOLUTION

Find the equivalent resistance of the network in **Fig. 26.3a** (next page) and the current in each resistor. The source of emf has negligible internal resistance.

SOLUTION

IDENTIFY and SET UP: This network of three resistors is a *combination* of series and parallel resistances, as in Fig. 26.1c. We determine the equivalent resistance of the parallel 6-Ω and 3-Ω resistors, and then that of their series combination with the 4-Ω resistor: This is the equivalent resistance R_{eq} of the network as a whole. We then find the current in the emf, which is the same as that in the 4-Ω resistor. The potential difference is the same across each of the parallel 6-Ω and 3-Ω resistors; we use this to determine how the current is divided between these.

EXECUTE: Figures 26.3b and 26.3c show successive steps in reducing the network to a single equivalent resistance R_{eq}. From Eq. (26.2), the 6-Ω and 3-Ω resistors in parallel in Fig. 26.3a are equivalent to the single 2-Ω resistor in Fig. 26.3b:

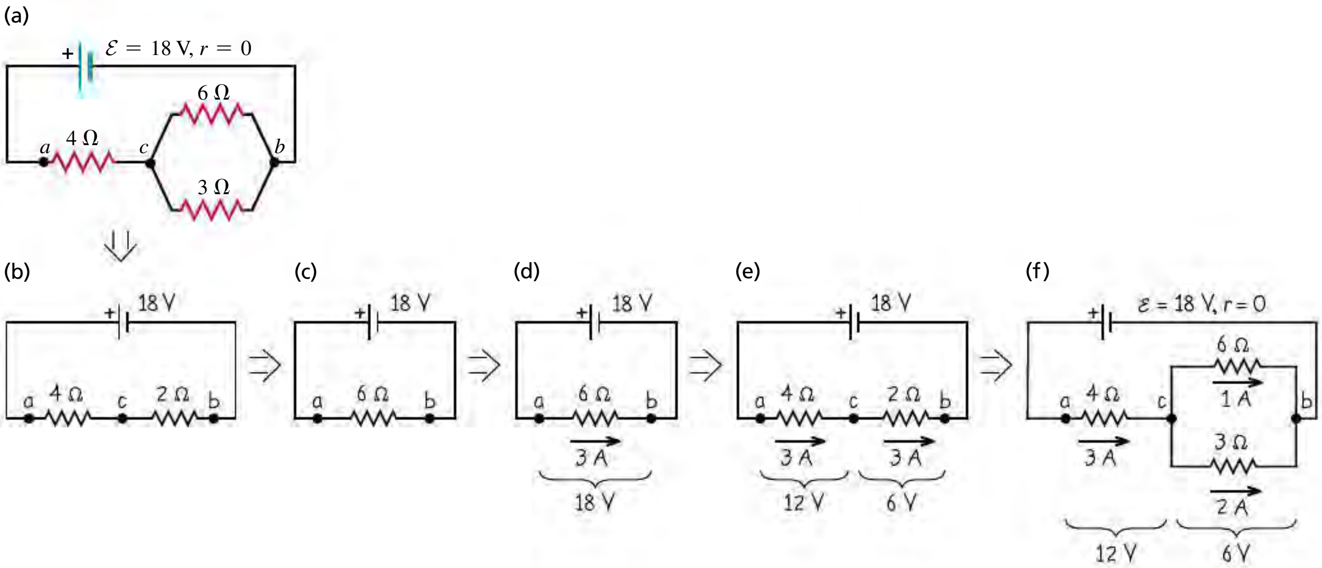
$$\frac{1}{R_{6\,\Omega+3\,\Omega}} = \frac{1}{6\ \Omega} + \frac{1}{3\ \Omega} = \frac{1}{2\ \Omega}$$

[Equation (26.3) gives the same result.] From Eq. (26.1) the series combination of this 2-Ω resistor with the 4-Ω resistor is equivalent to the single 6-Ω resistor in Fig. 26.3c.

We reverse these steps to find the current in each resistor of the original network. In the circuit shown in Fig. 26.3d (identical to Fig. 26.3c), the current is $I = V_{ab}/R = (18\ \text{V})/(6\ \Omega) = 3\ \text{A}$. So the current in the 4-Ω and 2-Ω resistors in Fig. 26.3e (identical

Continued

26.3 Steps in reducing a combination of resistors to a single equivalent resistor and finding the current in each resistor.



to Fig. 26.3b) is also 3 A. The potential difference V_{cb} across the 2-Ω resistor is therefore $V_{cb} = IR = (3\text{ A})(2\ \Omega) = 6\text{ V}$. This potential difference must also be 6 V in Fig. 26.3f (identical to Fig. 26.3a). From $I = V_{cb}/R$, the currents in the 6-Ω and 3-Ω resistors in Fig. 26.3f are, respectively, $(6\text{ V})/(6\ \Omega) = 1\text{ A}$ and $(6\text{ V})/(3\ \Omega) = 2\text{ A}$.

EVALUATE: Note that for the two resistors in parallel between points c and b in Fig. 26.3f, there is twice as much current through the 3-Ω resistor as through the 6-Ω resistor; more current goes through the path of least resistance, in accordance with Eq. (26.4). Note also that the total current through these two resistors is 3 A, the same as it is through the 4-Ω resistor between points a and c.

EXAMPLE 26.2 SERIES VERSUS PARALLEL COMBINATIONS

SOLUTION

Two identical incandescent light bulbs, each with resistance $R = 2\ \Omega$, are connected to a source with $\mathcal{E} = 8\text{ V}$ and negligible internal resistance. Find the current through each bulb, the potential difference across each bulb, and the power delivered to each bulb and to the entire network if the bulbs are connected (a) in series and (b) in parallel. (c) Suppose one of the bulbs burns out; that is, its filament breaks and current can no longer flow through it. What happens to the other bulb in the series case? In the parallel case?

SOLUTION

IDENTIFY and SET UP: The light bulbs are just resistors in simple series and parallel connections (**Figs. 26.4a** and 26.4b). Once we find the current I through each bulb, we can find the power delivered to each bulb by using Eq. (25.18), $P = I^2R = V^2/R$.

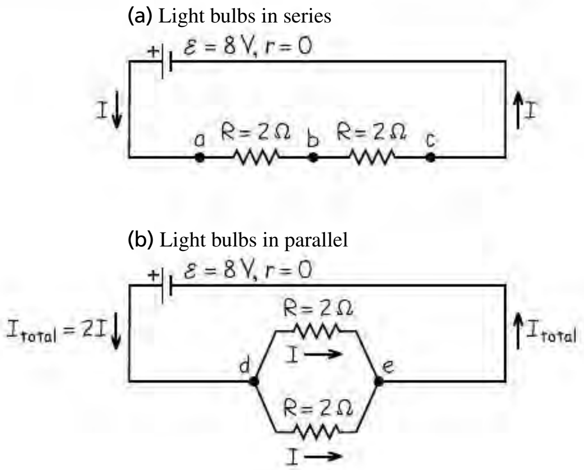
EXECUTE: (a) From Eq. (26.1) the equivalent resistance of the two bulbs between points a and c in Fig. 26.4a is $R_{\text{eq}} = 2R = 2(2\ \Omega) = 4\ \Omega$. In series, the current is the same through each bulb:

$$I = \frac{V_{ac}}{R_{\text{eq}}} = \frac{8\text{ V}}{4\ \Omega} = 2\text{ A}$$

Since the bulbs have the same resistance, the potential difference is the same across each bulb:

$$V_{ab} = V_{bc} = IR = (2\text{ A})(2\ \Omega) = 4\text{ V}$$

26.4 Our sketches for this problem.



From Eq. (25.18), the power delivered to each bulb is

$$P = I^2R = (2\text{ A})^2(2\ \Omega) = 8\text{ W} \qquad \text{or}$$

$$P = \frac{V_{ab}{}^2}{R} = \frac{V_{bc}{}^2}{R} = \frac{(4\text{ V})^2}{2\ \Omega} = 8\text{ W}$$

The total power delivered to both bulbs is $P_{\text{tot}} = 2P = 16\text{ W}$.

(b) If the bulbs are in parallel, as in Fig. 26.4b, the potential difference V_{de} across each bulb is the same and equal to 8 V, the terminal voltage of the source. Hence the current through each light bulb is

$$I = \frac{V_{de}}{R} = \frac{8\text{ V}}{2\ \Omega} = 4\text{ A}$$

and the power delivered to each bulb is

$$P = I^2R = (4\text{ A})^2(2\ \Omega) = 32\text{ W} \qquad \text{or}$$

$$P = \frac{V_{de}{}^2}{R} = \frac{(8\text{ V})^2}{2\ \Omega} = 32\text{ W}$$

Both the potential difference across each bulb and the current through each bulb are twice as great as in the series case. Hence the power delivered to each bulb is *four* times greater, and each bulb is brighter.

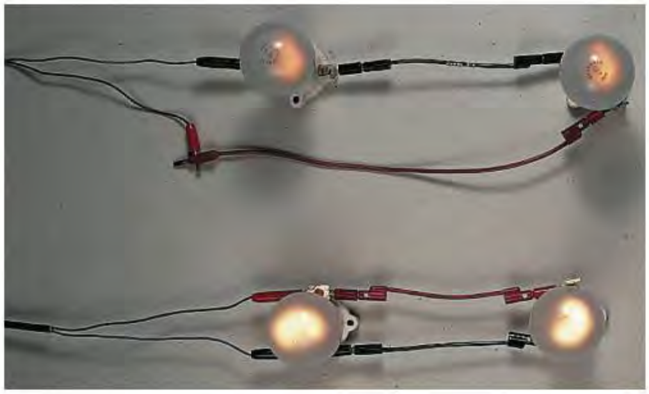
The total power delivered to the parallel network is $P_{\text{total}} = 2P = 64$ W, four times greater than in the series case. The increased power compared to the series case isn't obtained "for free"; energy is extracted from the source four times more rapidly in the parallel case than in the series case. If the source is a battery, it will be used up four times as fast.

(c) In the series case the same current flows through both bulbs. If one bulb burns out, there will be no current in the circuit, and neither bulb will glow.

In the parallel case the potential difference across either bulb is unchanged if a bulb burns out. The current through the functional bulb and the power delivered to it are unchanged.

EVALUATE: Our calculation isn't completely accurate, because the resistance $R = V/I$ of real light bulbs depends on the potential difference V across the bulb. That's because the filament resistance increases with increasing operating temperature and therefore with increasing V. But bulbs connected in series across a source do in fact glow less brightly than when connected in parallel across the same source (**Fig. 26.5**).

26.5 When connected to the same source, two incandescent light bulbs in series (shown at top) draw less power and glow less brightly than when they are in parallel (shown at bottom).



TEST YOUR UNDERSTANDING OF SECTION 26.1 Suppose all three of the resistors shown in Fig. 26.1 have the same resistance, so $R_1 = R_2 = R_3 = R$. Rank the four arrangements shown in parts (a)–(d) of Fig. 26.1 in order of their equivalent resistance, from highest to lowest. ❙

26.2 KIRCHHOFF'S RULES

Many practical resistor networks cannot be reduced to simple series-parallel combinations. **Figure 26.6a** shows a dc power supply with emf $\mathcal{E}_1$ charging a battery with a smaller emf $\mathcal{E}_2$ and feeding current to a light bulb with resistance R. Figure 26.6b is a "bridge" circuit, used in many different types of measurement and control systems. (Problem 26.74 describes one important application of a "bridge" circuit.) To analyze these networks, we'll use the techniques developed by the German physicist Gustav Robert Kirchhoff (1824–1887).

First, here are two terms that we will use often. A **junction** in a circuit is a point where three or more conductors meet. A **loop** is any closed conducting path. In Fig. 26.6a points a and b are junctions, but points c and d are not; in Fig. 26.6b points a, b, c, and d are junctions, but points e and f are not. The blue lines in Figs. 26.6a and 26.6b show some possible loops in these circuits.

Kirchhoff's rules are the following two statements:

Kirchhoff's junction rule (valid at any junction): The sum of the currents into any junction ...

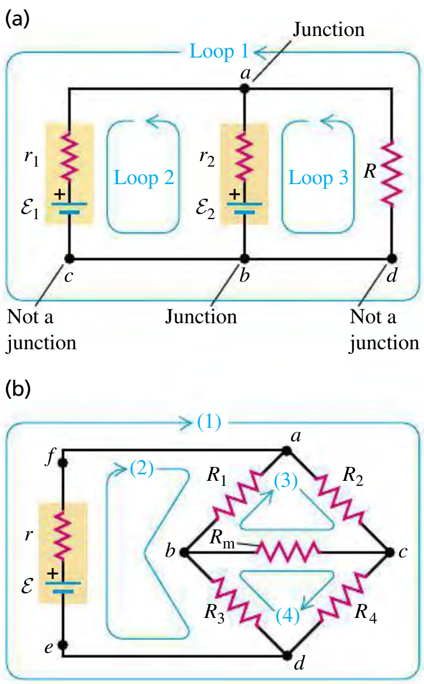
$$\sum I = 0 \quad \text{... equals zero.} \tag{26.5}$$

Kirchhoff's loop rule (valid for any closed loop): The sum of the potential differences around any loop ...

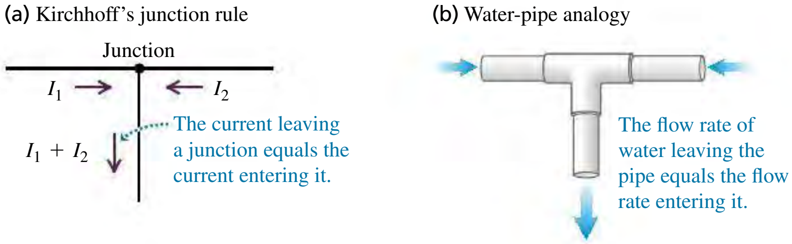
$$\sum V = 0 \quad \text{... equals zero.} \tag{26.6}$$

Note that the potential differences V in Eq. (26.6) include those associated with all circuit elements in the loop, including emfs and resistors.

26.6 Two networks that cannot be reduced to simple series-parallel combinations of resistors.



26.7 Kirchhoff's junction rule states that as much current flows into a junction as flows out of it.



DATA *SPEAKS*

Multiloop Circuits

When students were given a problem involving a circuit with two or more loops, more than 32% gave an incorrect response. Common errors:

- Confusion over whether circuit elements are in series or parallel. If two circuit elements are connected so that the same current passes through both, they are in series; if they are connected so that the potential difference is the same across both, they are in parallel.
- Confusion over what happens at a junction. Current need not split equally among different paths; more current flows along a path with less resistance.

The junction rule is based on *conservation of electric charge*. No charge can accumulate at a junction, so the total charge entering the junction per unit time must equal the total charge leaving per unit time (**Fig. 26.7a**). Charge per unit time is current, so if we consider the currents entering a junction to be positive and those leaving to be negative, the algebraic sum of currents into a junction must be zero. It's like a T branch in a water pipe (Fig. 26.7b); if you have a total of 1 liter per minute coming in the two pipes, you can't have 3 liters per minute going out the third pipe. We used the junction rule (without saying so) in Section 26.1 in the derivation of Eq. (26.2) for resistors in parallel.

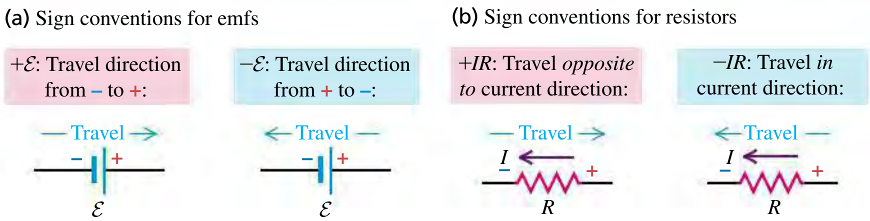
The loop rule is a statement that the electrostatic force is *conservative*. Suppose we go around a loop, measuring potential differences across successive circuit elements as we go. When we return to the starting point, we must find that the *algebraic sum* of these differences is zero; otherwise, we could not say that the potential at this point has a definite value.

Sign Conventions for the Loop Rule

In applying the loop rule, we need some sign conventions. Problem-Solving Strategy 26.2 describes in detail how to use these, but here's a quick overview. We first assume a direction for the current in each branch of the circuit and mark it on a diagram of the circuit. Then, starting at any point in the circuit, we imagine traveling around a loop, adding emfs and IR terms as we come to them. When we travel through a source in the direction from $-$ to $+$, the emf is considered to be *positive*; when we travel from $+$ to $-$, the emf is considered to be *negative* (**Fig. 26.8a**). When we travel through a resistor in the *same* direction as the assumed current, the IR term is *negative* because the current goes in the direction of decreasing potential. When we travel through a resistor in the direction *opposite* to the assumed current, the IR term is *positive* because this represents a rise of potential (Fig. 26.8b).

Kirchhoff's two rules are all we need to solve a wide variety of network problems. Usually, some of the emfs, currents, and resistances are known, and others are unknown. We must always obtain from Kirchhoff's rules a number of independent equations equal to the number of unknowns so that we can solve the equations simultaneously. Often the hardest part of the solution is keeping track of algebraic signs!

26.8 Use these sign conventions when you apply Kirchhoff's loop rule. In each part of the figure "Travel" is the direction that we imagine going around the loop, which is not necessarily the direction of the current.



PROBLEM-SOLVING STRATEGY 26.2 KIRCHHOFF'S RULES

IDENTIFY *the relevant concepts:* Kirchhoff's rules are useful for analyzing any electric circuit.

SET UP *the problem* using the following steps:

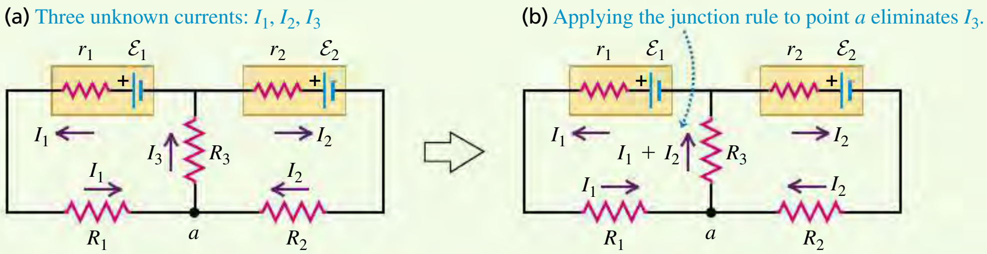
1. Draw a circuit diagram, leaving room to label all quantities, known and unknown. Indicate an assumed direction for each unknown current and emf. (Kirchhoff's rules will yield the magnitudes *and directions* of unknown currents and emfs. If the actual direction of a quantity is opposite to your assumption, the resulting quantity will have a negative sign.)
2. As you label currents, it's helpful to use Kirchhoff's junction rule, as in **Fig. 26.9**, so as to express the currents in terms of as few quantities as possible.
3. Identify the target variables.

EXECUTE *the solution* as follows:

1. Choose any loop in the network and choose a direction (clockwise or counterclockwise) to travel around the loop as you apply Kirchhoff's loop rule. The direction need not be the same as any assumed current direction.
2. Travel around the loop in the chosen direction, adding potential differences algebraically as you cross them. Use the sign conventions of Fig. 26.8.
3. Equate the sum obtained in step 2 to zero in accordance with the loop rule.
4. If you need more independent equations, choose another loop and repeat steps 1–3; continue until you have as many independent equations as unknowns or until every circuit element has been included in at least one loop.
5. Solve the equations simultaneously to determine the unknowns.
6. You can use the loop-rule bookkeeping system to find the potential V_{ab} of any point a with respect to any other point b. Start at b and add the potential changes you encounter in going from b to a; use the same sign rules as in step 2. The algebraic sum of these changes is $V_{ab} = V_a - V_b$.

EVALUATE *your answer:* Check all the steps in your algebra. Apply steps 1 and 2 to a loop you have not yet considered; if the sum of potential drops isn't zero, you've made an error somewhere.

26.9 Applying the junction rule to point a reduces the number of unknown currents from three to two.



EXAMPLE 26.3 A SINGLE-LOOP CIRCUIT

SOLUTION

The circuit shown in **Fig. 26.10a** (next page) contains two batteries, each with an emf and an internal resistance, and two resistors. Find (a) the current in the circuit, (b) the potential difference V_{ab}, and (c) the power output of the emf of each battery.

SOLUTION

IDENTIFY and SET UP: There are no junctions in this single-loop circuit, so we don't need Kirchhoff's junction rule. To apply Kirchhoff's loop rule, we first assume a direction for the current; let's assume a counterclockwise direction as shown in Fig. 26.10a.

EXECUTE: (a) Starting at a and traveling counterclockwise with the current, we add potential increases and decreases and equate the sum to zero as in Eq. (26.6):

$$-I(4\ \Omega) - 4\ \text{V} - I(7\ \Omega) + 12\ \text{V} - I(2\ \Omega) - I(3\ \Omega) = 0$$

Collecting like terms and solving for I, we find

$$8\ \text{V} = I(16\ \Omega) \quad \text{and} \quad I = 0.5\ \text{A}$$

The positive result for I shows that our assumed current direction is correct.

(b) To find V_{ab}, the potential at a with respect to b, we start at b and add potential changes as we go toward a. There are two paths from b to a; taking the lower one, we find

$$V_{ab} = (0.5\ \text{A})(7\ \Omega) + 4\ \text{V} + (0.5\ \text{A})(4\ \Omega) = 9.5\ \text{V}$$

Point a is at 9.5 V higher potential than b. All the terms in this sum, including the IR terms, are positive because each represents an *increase* in potential as we go from b to a. For the upper path,

$$V_{ab} = 12\ \text{V} - (0.5\ \text{A})(2\ \Omega) - (0.5\ \text{A})(3\ \Omega) = 9.5\ \text{V}$$

Here the IR terms are negative because our path goes in the direction of the current, with potential decreases through the resistors. The results for V_{ab} are the same for both paths, as they must be in order for the total potential change around the loop to be zero.

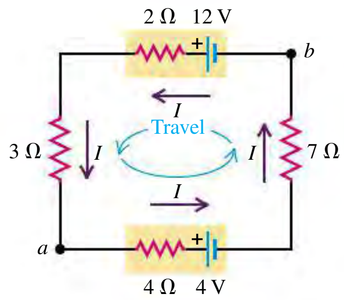
(c) The power outputs of the emf of the 12-V and 4-V batteries are

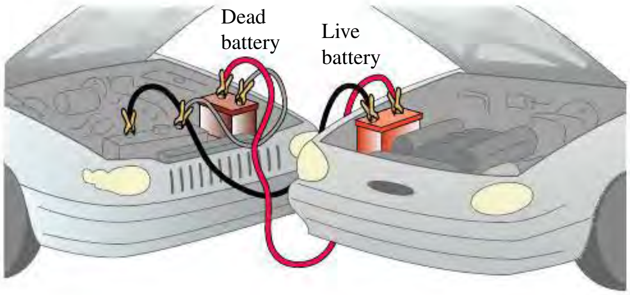
$$P_{12\text{V}} = \mathcal{E}I = (12\ \text{V})(0.5\ \text{A}) = 6\ \text{W}$$

$$P_{4\text{V}} = \mathcal{E}I = (-4\ \text{V})(0.5\ \text{A}) = -2\ \text{W}$$

Continued

26.10 (a) In this example we travel around the loop in the same direction as the assumed current, so all the IR terms are negative. The potential decreases as we travel from + to − through the bottom emf but increases as we travel from − to + through the top emf. (b) A real-life example of a circuit of this kind.





The negative sign in $\mathcal{E}$ for the 4-V battery appears because the current actually runs from the higher-potential side of the battery to the lower-potential side. The negative value of P means that we are *storing* energy in that battery; the 12-V battery is *recharging* it (if it is in fact rechargeable; otherwise, we're destroying it).

EVALUATE: By applying the expression $P = I^2R$ to each of the four resistors in Fig. 26.10a, you can show that the total power dissipated in all four resistors is 4 W. Of the 6 W provided by the emf of the 12-V battery, 2 W goes into storing energy in the 4-V battery and 4 W is dissipated in the resistances.

The circuit shown in Fig. 26.10a is much like that used when a fully charged 12-V storage battery (in a car with its engine running) "jump-starts" a car with a run-down battery (Fig. 26.10b). The run-down battery is slightly recharged in the process. The 3-Ω and 7-Ω resistors in Fig. 26.10a represent the resistances of the jumper cables and of the conducting path through the car with the run-down battery. (The values of the resistances in actual automobiles and jumper cables are considerably lower, and the emf of a run-down car battery isn't much less than 12 V.)

EXAMPLE 26.4 CHARGING A BATTERY

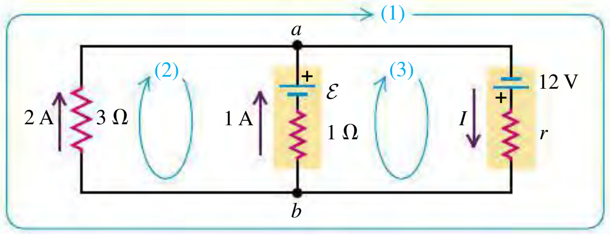
SOLUTION

In the circuit shown in **Fig. 26.11**, a 12-V power supply with unknown internal resistance r is connected to a run-down rechargeable battery with unknown emf $\mathcal{E}$ and internal resistance 1 Ω and to an indicator light bulb of resistance 3 Ω carrying a current of 2 A. The current through the run-down battery is 1 A in the direction shown. Find r, $\mathcal{E}$, and the current I through the power supply.

SOLUTION

IDENTIFY and SET UP: This circuit has more than one loop, so we must apply both the junction and loop rules. We assume the direction of the current through the 12-V power supply, and the polarity of the run-down battery, to be as shown in Fig. 26.11. There are three target variables, so we need three equations.

26.11 In this circuit a power supply charges a run-down battery and lights a bulb. An assumption has been made about the polarity of the emf $\mathcal{E}$ of the battery. Is this assumption correct?



EXECUTE: We apply the junction rule, Eq. (26.5), to point a:

$$-I + 1\ \text{A} + 2\ \text{A} = 0 \qquad \text{so} \qquad I = 3\ \text{A}$$

To determine r, we apply the loop rule, Eq. (26.6), to the large, outer loop (1):

$$12\ \text{V} - (3\ \text{A})r - (2\ \text{A})(3\ \Omega) = 0 \qquad \text{so} \qquad r = 2\ \Omega$$

To determine $\mathcal{E}$, we apply the loop rule to the left-hand loop (2):

$$-\mathcal{E} + (1\ \text{A})(1\ \Omega) - (2\ \text{A})(3\ \Omega) = 0 \qquad \text{so} \qquad \mathcal{E} = -5\ \text{V}$$

The negative value for $\mathcal{E}$ shows that the actual polarity of this emf is opposite to that shown in Fig. 26.11. As in Example 26.3, the battery is being recharged.

EVALUATE: Try applying the junction rule at point b instead of point a, and try applying the loop rule counterclockwise rather than clockwise around loop (1). You'll get the same results for I and r. We can check our result for $\mathcal{E}$ by using loop (3):

$$12\ \text{V} - (3\ \text{A})(2\ \Omega) - (1\ \text{A})(1\ \Omega) + \mathcal{E} = 0$$

which again gives us $\mathcal{E} = -5$ V.

As an additional check, we note that $V_{ba} = V_b - V_a$ equals the voltage across the 3-Ω resistance, which is $(2\ \text{A})(3\ \Omega) = 6$ V. Going from a to b by the right-hand branch, we encounter potential differences $+12\ \text{V} - (3\ \text{A})(2\ \Omega) = +6$ V, and going by the middle branch, we find $-(-5\ \text{V}) + (1\ \text{A})(1\ \Omega) = +6$ V. The three ways of getting V_{ba} give the same results.

EXAMPLE 26.5 POWER IN A BATTERY-CHARGING CIRCUIT

In the circuit of Example 26.4 (shown in Fig. 26.11), find the power delivered by the 12-V power supply and by the battery being recharged, and find the power dissipated in each resistor.

SOLUTION

IDENTIFY and SET UP: We use the results of Section 25.5, in which we found that the power delivered *from* an emf to a circuit is $\mathcal{E}I$ and the power delivered *to* a resistor from a circuit is $V_{ab}I = I^2R$. We know the values of all relevant quantities from Example 26.4.

EXECUTE: The power output from the emf of the power supply is

$$P_{\text{supply}} = \mathcal{E}_{\text{supply}} I_{\text{supply}} = (12\ \text{V})(3\ \text{A}) = 36\ \text{W}$$

The power dissipated in the power supply's internal resistance r is

$$P_{r-\text{supply}} = I_{\text{supply}}{}^2 r_{\text{supply}} = (3\ \text{A})^2(2\ \Omega) = 18\ \text{W}$$

so the power supply's *net* power output is $P_{\text{net}} = 36\ \text{W} - 18\ \text{W} = 18\ \text{W}$. Alternatively, from Example 26.4 the terminal voltage of the battery is $V_{ba} = 6\ \text{V}$, so the net power output is

$$P_{\text{net}} = V_{ba} I_{\text{supply}} = (6\ \text{V})(3\ \text{A}) = 18\ \text{W}$$

The power output of the emf $\mathcal{E}$ of the battery being charged is

$$P_{\text{emf}} = \mathcal{E} I_{\text{battery}} = (-5\ \text{V})(1\ \text{A}) = -5\ \text{W}$$

This is negative because the 1-A current runs through the battery from the higher-potential side to the lower-potential side. (As we mentioned in Example 26.4, the polarity assumed for this battery in Fig. 26.11 was wrong.) We are storing energy in the battery as we charge it. Additional power is dissipated in the battery's internal resistance; this power is

$$P_{r\text{-battery}} = I_{\text{battery}}{}^2 r_{\text{battery}} = (1\ \text{A})^2(1\ \Omega) = 1\ \text{W}$$

The total power input to the battery is thus $1\ \text{W} + |-5\ \text{W}| = 6\ \text{W}$. Of this, 5 W represents useful energy stored in the battery; the remainder is wasted in its internal resistance.

The power dissipated in the light bulb is

$$P_{\text{bulb}} = I_{\text{bulb}}{}^2 R_{\text{bulb}} = (2\ \text{A})^2(3\ \Omega) = 12\ \text{W}$$

EVALUATE: As a check, note that all of the power from the supply is accounted for. Of the 18 W of net power from the power supply, 5 W goes to recharge the battery, 1 W is dissipated in the battery's internal resistance, and 12 W is dissipated in the light bulb.

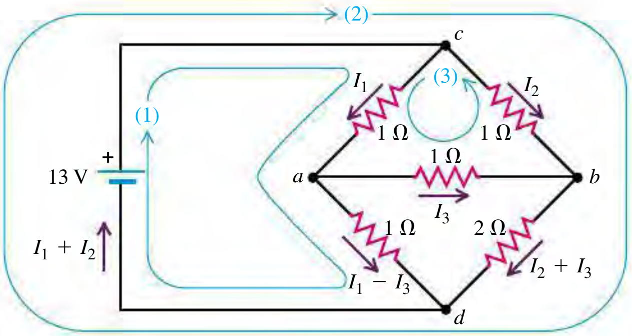
EXAMPLE 26.6 A COMPLEX NETWORK

Figure 26.12 shows a "bridge" circuit of the type described at the beginning of this section (see Fig. 26.6b). Find the current in each resistor and the equivalent resistance of the network of five resistors.

SOLUTION

IDENTIFY and SET UP: This network is neither a series combination nor a parallel combination. Hence we must use Kirchhoff's rules to find the values of the target variables. There are five unknown currents, but by applying the junction rule to junctions a and b, we can represent them in terms of three unknown currents I_1, I_2, and I_3, as shown in Fig. 26.12.

26.12 A network circuit with several resistors.



EXECUTE: We apply the loop rule to the three loops shown:

$$13\ \text{V} - I_1(1\ \Omega) - (I_1 - I_3)(1\ \Omega) = 0 \qquad (1)$$

$$-I_2(1\ \Omega) - (I_2 + I_3)(2\ \Omega) + 13\ \text{V} = 0 \qquad (2)$$

$$-I_1(1\ \Omega) - I_3(1\ \Omega) + I_2(1\ \Omega) = 0 \qquad (3)$$

One way to solve these simultaneous equations is to solve Eq. (3) for I_2, obtaining $I_2 = I_1 + I_3$, and then substitute this expression into Eq. (2) to eliminate I_2. We then have

$$13\ \text{V} = I_1(2\ \Omega) - I_3(1\ \Omega) \qquad (1')$$

$$13\ \text{V} = I_1(3\ \Omega) + I_3(5\ \Omega) \qquad (2')$$

Now we can eliminate I_3 by multiplying Eq. $(1')$ by 5 and adding the two equations. We obtain

$$78\ \text{V} = I_1(13\ \Omega) \qquad I_1 = 6\ \text{A}$$

We substitute this result into Eq. (1′) to obtain $I_3 = -1\ \text{A}$, and from Eq. (3) we find $I_2 = 5\ \text{A}$. The negative value of I_3 tells us that its direction is opposite to the direction we assumed.

The total current through the network is $I_1 + I_2 = 11\ \text{A}$, and the potential drop across it is equal to the battery emf, 13 V. The equivalent resistance of the network is therefore

$$R_{\text{eq}} = \frac{13\ \text{V}}{11\ \text{A}} = 1.2\ \Omega$$

EVALUATE: You can check our results for I_1, I_2, and I_3 by substituting them back into Eqs. (1)–(3). What do you find?

EXAMPLE 26.7 A POTENTIAL DIFFERENCE IN A COMPLEX NETWORK

In the circuit of Example 26.6 (Fig. 26.12), find the potential difference V_{ab}.

SOLUTION

IDENTIFY and SET UP: Our target variable $V_{ab} = V_a - V_b$ is the potential at point a with respect to point b. To find it, we start at point b and follow a path to point a, adding potential rises and drops as we go. We can follow any of several paths from b to a; the result must be the same for all such paths, which gives us a way to check our result.

EXECUTE: The simplest path is through the center 1-Ω resistor. In Example 26.6 we found $I_3 = -1$ A, showing that the actual current direction through this resistor is from right to left. Thus, as we go from b to a, there is a *drop* of potential with magnitude $|I_3|R = (1\text{ A})(1\ \Omega) = 1$ V. Hence $V_{ab} = -1$ V, and the potential at a is 1 V less than at point b.

EVALUATE: To check our result, let's try a path from b to a that goes through the lower two resistors. The currents through these are

$$I_2 + I_3 = 5\text{ A} + (-1\text{ A}) = 4\text{ A} \quad \text{and}$$
$$I_1 - I_3 = 6\text{ A} - (-1\text{ A}) = 7\text{ A}$$

and so

$$V_{ab} = -(4\text{ A})(2\ \Omega) + (7\text{ A})(1\ \Omega) = -1\text{ V}$$

You can confirm this result by using some other paths from b to a.

26.13 Both this ammeter (top) and voltmeter (bottom) are d'Arsonval galvanometers. The difference has to do with their internal connections (see Fig. 26.15).

TEST YOUR UNDERSTANDING OF SECTION 26.2 Subtract Eq. (1) from Eq. (2) in Example 26.6. To which loop in Fig. 26.12 does this equation correspond? Would this equation have simplified the solution of Example 26.6?

26.3 ELECTRICAL MEASURING INSTRUMENTS

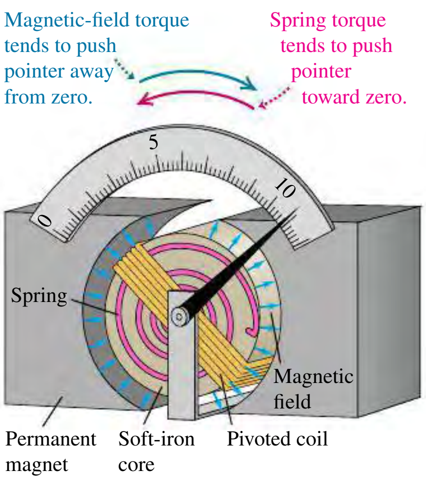
We've been talking about potential difference, current, and resistance for two chapters, so it's about time we said something about how to *measure* these quantities. Many common devices, including car instrument panels, battery chargers, and inexpensive electrical instruments, measure potential difference (voltage), current, or resistance with a **d'Arsonval galvanometer** (**Fig. 26.13**). In the following discussion we'll often call it just a *meter*. A pivoted coil of fine wire is placed in the magnetic field of a permanent magnet (**Fig. 26.14**). Attached to the coil is a spring, similar to the hairspring on the balance wheel of a watch. In the equilibrium position, with no current in the coil, the pointer is at zero. When there is a current in the coil, the magnetic field exerts a torque on the coil that is proportional to the current. (We'll discuss this magnetic interaction in detail in Chapter 27.) As the coil turns, the spring exerts a restoring torque that is proportional to the angular displacement.

Thus the angular deflection of the coil and pointer is directly proportional to the coil current, and the device can be calibrated to measure current. The maximum deflection, typically 90° or so, is called *full-scale deflection*. The essential electrical characteristics of the meter are the current I_{fs} required for full-scale deflection (typically on the order of 10 μA to 10 mA) and the resistance R_c of the coil (typically on the order of 10 Ω to 1000 Ω).

The meter deflection is proportional to the *current* in the coil. If the coil obeys Ohm's law, the current is proportional to the *potential difference* between the terminals of the coil, and the deflection is also proportional to this potential difference. For example, consider a meter whose coil has a resistance $R_c = 20.0\ \Omega$ and that deflects full scale when the current in its coil is $I_{fs} = 1.00$ mA. The corresponding potential difference for full-scale deflection is

$$V = I_{fs}R_c = (1.00 \times 10^{-3}\text{ A})(20.0\ \Omega) = 0.0200\text{ V}$$

26.14 A d'Arsonval galvanometer, showing a pivoted coil with attached pointer, a permanent magnet supplying a magnetic field that is uniform in magnitude, and a spring to provide restoring torque, which opposes magnetic-field torque.



Ammeters

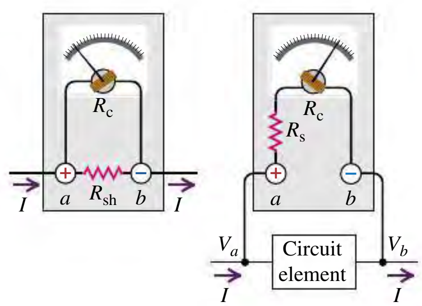
A current-measuring instrument is usually called an **ammeter** (or milliammeter, microammeter, and so forth, depending on the range). *An ammeter always measures the current passing through it.* An *ideal* ammeter, discussed in Section 25.4, would have *zero* resistance, so including it in a branch of a circuit would not affect the current in that branch. Real ammeters always have a finite resistance, but it is always desirable for an ammeter to have as little resistance as possible.

We can adapt any meter to measure currents that are larger than its full-scale reading by connecting a resistor in parallel with it (**Fig. 26.15a**) so that some of the current bypasses the meter coil. The parallel resistor is called a **shunt resistor** or simply a *shunt,* denoted as R_{sh}.

Suppose we want to make a meter with full-scale current I_{fs} and coil resistance R_c into an ammeter with full-scale reading I_a. To determine the shunt resistance R_{sh} needed, note that at full-scale deflection the total current through the parallel combination is I_a, the current through the coil of the meter is I_{fs}, and the current through the shunt is the difference $I_a - I_{fs}$. The potential difference V_{ab} is the same for both paths, so

$$I_{fs}R_c = (I_a - I_{fs})R_{sh} \qquad \text{(for an ammeter)} \tag{26.7}$$

26.15 Using the same meter to measure (a) current and (b) voltage.



EXAMPLE 26.8 DESIGNING AN AMMETER

What shunt resistance is required to make the 1.00-mA, 20.0-Ω meter described above into an ammeter with a range of 0 to 50.0 mA?

SOLUTION

IDENTIFY and SET UP: Since the meter is being used as an ammeter, its internal connections are as shown in Fig. 26.15a. Our target variable is the shunt resistance R_{sh}, which we will find from Eq. (26.7). The ammeter must handle a maximum current $I_a = 50.0 \times 10^{-3}$ A. The coil resistance is $R_c = 20.0\ \Omega$, and the meter shows full-scale deflection when the current through the coil is $I_{fs} = 1.00 \times 10^{-3}$ A.

EXECUTE: Solving Eq. (26.7) for R_{sh}, we find

$$R_{sh} = \frac{I_{fs}R_c}{I_a - I_{fs}} = \frac{(1.00 \times 10^{-3}\ \text{A})(20.0\ \Omega)}{50.0 \times 10^{-3}\ \text{A} - 1.00 \times 10^{-3}\ \text{A}}$$
$$= 0.408\ \Omega$$

EVALUATE: It's useful to consider the equivalent resistance R_{eq} of the ammeter as a whole. From Eq. (26.2),

$$R_{eq} = \left(\frac{1}{R_c} + \frac{1}{R_{sh}}\right)^{-1} = \left(\frac{1}{20.0\ \Omega} + \frac{1}{0.408\ \Omega}\right)^{-1}$$
$$= 0.400\ \Omega$$

The shunt resistance is so small in comparison to the coil resistance that the equivalent resistance is very nearly equal to the shunt resistance. The result is an ammeter with a low equivalent resistance and the desired 0–50.0-mA range. At full-scale deflection, $I = I_a = 50.0$ mA, the current through the galvanometer is 1.00 mA, the current through the shunt resistor is 49.0 mA, and $V_{ab} = 0.0200$ V. If the current I is *less* than 50.0 mA, the coil current and the deflection are proportionally less.

Voltmeters

This same basic meter may also be used to measure potential difference or *voltage*. A voltage-measuring device is called a **voltmeter.** A voltmeter always measures the potential difference between two points, and its terminals must be connected to these points. (Example 25.6 in Section 25.4 described what can happen if a voltmeter is connected incorrectly.) As we discussed in Section 25.4, an ideal voltmeter would have *infinite* resistance, so connecting it between two points in a circuit would not alter any of the currents. Real voltmeters always have finite resistance, but a voltmeter should have large enough resistance that connecting it in a circuit does not change the other currents appreciably.

For the meter of Example 26.8, the voltage across the meter coil at full-scale deflection is only $I_{fs}R_c = (1.00 \times 10^{-3}\ \text{A})(20.0\ \Omega) = 0.0200$ V. We can extend this range by connecting a resistor R_s in *series* with the coil (Fig. 26.15b). Then only a fraction of the total potential difference appears across the coil itself, and the remainder appears across R_s. For a voltmeter with full-scale reading V_V, we need a series resistor R_s in Fig. 26.15b such that

$$V_V = I_{fs}(R_c + R_s) \qquad \text{(for a voltmeter)} \tag{26.8}$$

BIO Application Electromyography A fine needle containing two electrodes is being inserted into a muscle in this patient's hand. By using a sensitive voltmeter to measure the potential difference between these electrodes, a physician can probe the muscle's electrical activity. This is an important technique for diagnosing neurological and neuromuscular diseases.

EXAMPLE 26.9 DESIGNING A VOLTMETER

What series resistance is required to make the 1.00-mA, 20.0-Ω meter described above into a voltmeter with a range of 0 to 10.0 V?

SOLUTION

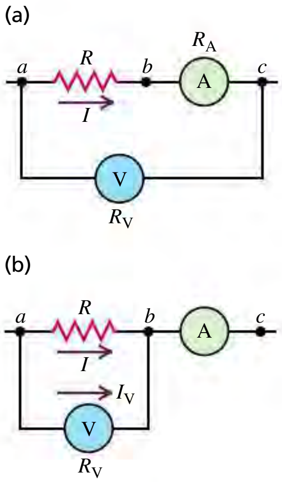
IDENTIFY and SET UP: Since this meter is being used as a voltmeter, its internal connections are as shown in Fig. 26.15b. The maximum allowable voltage across the voltmeter is $V_V = 10.0$ V. We want this to occur when the current through the coil is $I_{fs} = 1.00 \times 10^{-3}$ A. Our target variable is the series resistance R_s, which we find from Eq. (26.8).

EXECUTE: From Eq. (26.8),

$$R_s = \frac{V_V}{I_{fs}} - R_c = \frac{10.0 \text{ V}}{0.00100 \text{ A}} - 20.0\ \Omega = 9980\ \Omega$$

EVALUATE: At full-scale deflection, $V_{ab} = 10.0$ V, the voltage across the meter is 0.0200 V, the voltage across R_s is 9.98 V, and the current through the voltmeter is 0.00100 A. Most of the voltage appears across the series resistor. The meter's equivalent resistance is a desirably high $R_{eq} = 20.0\ \Omega + 9980\ \Omega = 10{,}000\ \Omega$. Such a meter is called a "1000 ohms-per-volt" meter, referring to the ratio of resistance to full-scale deflection. In normal operation the current through the circuit element being measured (I in Fig. 26.15b) is much greater than 0.00100 A, and the resistance between points a and b in the circuit is much less than 10,000 Ω. The voltmeter draws off only a small fraction of the current and thus disturbs the circuit being measured only slightly.

26.16 Ammeter–voltmeter method for measuring resistance.



Ammeters and Voltmeters in Combination

A voltmeter and an ammeter can be used together to measure *resistance* and *power*. The resistance R of a resistor equals the potential difference V_{ab} between its terminals divided by the current I; that is, $R = V_{ab}/I$. The power input P to any circuit element is the product of the potential difference across it and the current through it: $P = V_{ab}I$. In principle, the most straightforward way to measure R or P is to measure V_{ab} and I simultaneously.

With practical ammeters and voltmeters this isn't quite as simple as it seems. In **Fig. 26.16a**, ammeter A reads the current I in the resistor R. Voltmeter V, however, reads the *sum* of the potential difference V_{ab} across the resistor and the potential difference V_{bc} across the ammeter. If we transfer the voltmeter terminal from c to b, as in Fig. 26.16b, then the voltmeter reads the potential difference V_{ab} correctly, but the ammeter now reads the *sum* of the current I in the resistor and the current I_V in the voltmeter. Either way, we have to correct the reading of one instrument or the other unless the corrections are small enough to be negligible.

EXAMPLE 26.10 MEASURING RESISTANCE I

The voltmeter in the circuit of Fig. 26.16a reads 12.0 V and the ammeter reads 0.100 A. The meter resistances are $R_V = 10{,}000\ \Omega$ (for the voltmeter) and $R_A = 2.00\ \Omega$ (for the ammeter). What are the resistance R and the power dissipated in the resistor?

SOLUTION

IDENTIFY and SET UP: The ammeter reads the current $I = 0.100$ A through the resistor, and the voltmeter reads the potential difference between a and c. If the ammeter were *ideal* (that is, if $R_A = 0$), there would be zero potential difference between b and c, the voltmeter reading $V = 12.0$ V would be equal to the potential difference V_{ab} across the resistor, and the resistance would be equal to $R = V/I = (12.0\text{ V})/(0.100\text{ A}) = 120\ \Omega$. The ammeter is *not* ideal, however (its resistance is $R_A = 2.00\ \Omega$), so the voltmeter reading V is actually the sum of the potential differences V_{bc} (across the ammeter) and V_{ab} (across the resistor). We use Ohm's law to find the voltage V_{bc} from the known current and ammeter resistance. Then we solve for V_{ab} and R. Given these, we are able to calculate the power P into the resistor.

EXECUTE: From Ohm's law, $V_{bc} = IR_A = (0.100\text{ A})(2.00\ \Omega) = 0.200$ V and $V_{ab} = IR$. The sum of these is $V = 12.0$ V, so the potential difference across the resistor is $V_{ab} = V - V_{bc} = (12.0\text{ V}) - (0.200\text{ V}) = 11.8$ V. Hence the resistance is

$$R = \frac{V_{ab}}{I} = \frac{11.8\text{ V}}{0.100\text{ A}} = 118\ \Omega$$

The power dissipated in this resistor is

$$P = V_{ab}I = (11.8\text{ V})(0.100\text{ A}) = 1.18\text{ W}$$

EVALUATE: You can confirm this result for the power by using the alternative formula $P = I^2R$. Do you get the same answer?

EXAMPLE 26.11 MEASURING RESISTANCE II

SOLUTION

Suppose the meters of Example 26.10 are connected to a different resistor as shown in Fig. 26.16b, and the readings obtained on the meters are the same as in Example 26.10. What is the value of this new resistance R, and what is the power dissipated in the resistor?

SOLUTION

IDENTIFY and SET UP: In Example 26.10 the ammeter read the actual current through the resistor, but the voltmeter reading was not the same as the potential difference across the resistor. Now the situation is reversed: The voltmeter reading $V = 12.0\text{ V}$ shows the actual potential difference V_{ab} across the resistor, but the ammeter reading $I_A = 0.100\text{ A}$ is *not* equal to the current I through the resistor. Applying the junction rule at b in Fig. 26.16b shows that $I_A = I + I_V$, where I_V is the current through the voltmeter. We find I_V from the given values of V and the voltmeter resistance R_V, and we use this value to find the resistor current I. We then determine the resistance R from I and the voltmeter reading, and calculate the power as in Example 26.10.

EXECUTE: We have $I_V = V/R_V = (12.0\text{ V})/(10{,}000\ \Omega) = 1.20\text{ mA}$. The actual current I in the resistor is $I = I_A - I_V = 0.100\text{ A} - 0.0012\text{ A} = 0.0988\text{ A}$, and the resistance is

$$R = \frac{V_{ab}}{I} = \frac{12.0\text{ V}}{0.0988\text{ A}} = 121\ \Omega$$

The power dissipated in the resistor is

$$P = V_{ab}I = (12.0\text{ V})(0.0988\text{ A}) = 1.19\text{ W}$$

EVALUATE: Had the meters been ideal, our results would have been $R = 12.0\text{ V}/0.100\text{ A} = 120\ \Omega$ and $P = VI = (12.0\text{ V}) \times (0.100\text{ A}) = 1.2\text{ W}$ both here and in Example 26.10. The actual (correct) results are not too different in either case. That's because the ammeter and voltmeter are nearly ideal: Compared with the resistance R under test, the ammeter resistance R_A is very small and the voltmeter resistance R_V is very large. Under these conditions, treating the meters as ideal yields pretty good results; accurate work requires calculations as in these two examples.

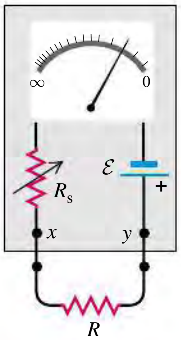
Ohmmeters

An alternative method for measuring resistance is to use a d'Arsonval meter in an arrangement called an **ohmmeter.** It consists of a meter, a resistor, and a source (often a flashlight battery) connected in series (**Fig. 26.17**). The resistance R to be measured is connected between terminals x and y.

The series resistance R_s is variable; it is adjusted so that when terminals x and y are short-circuited (that is, when $R = 0$), the meter deflects full scale. When nothing is connected to terminals x and y, so that the circuit between x and y is *open* (that is, when $R \rightarrow \infty$), there is no current and hence no deflection. For any intermediate value of R the meter deflection depends on the value of R, and the meter scale can be calibrated to read the resistance R directly. Larger currents correspond to smaller resistances, so this scale reads backward compared to the scale showing the current.

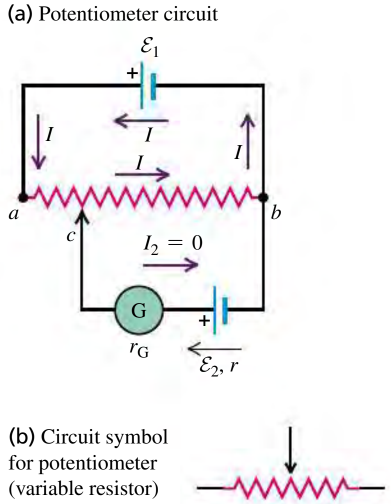
In situations in which high precision is required, instruments containing d'Arsonval meters have been supplanted by electronic instruments with direct digital readouts. Digital voltmeters can be made with extremely high internal resistance, of the order of $100\text{ M}\Omega$. **Figure 26.18** shows a digital *multimeter,* an instrument that can measure voltage, current, or resistance over a wide range.

26.17 Ohmmeter circuit. The resistor R_s has a variable resistance, as is indicated by the arrow through the resistor symbol. To use the ohmmeter, first connect x directly to y and adjust R_s until the meter reads zero. Then connect x and y across the resistor R and read the scale.



26.18 This digital multimeter can be used as a voltmeter (red arc), ammeter (yellow arc), or ohmmeter (green arc).

26.19 A potentiometer.



The Potentiometer

The *potentiometer* is an instrument that can be used to measure the emf of a source without drawing any current from the source; it also has a number of other useful applications. Essentially, it balances an unknown potential difference against an adjustable, measurable potential difference.

The principle of the potentiometer is shown schematically in **Fig. 26.19a**. A resistance wire *ab* of total resistance R_{ab} is permanently connected to the terminals of a source of known emf $\mathcal{E}_1$. A sliding contact *c* is connected through the galvanometer G to a second source whose emf $\mathcal{E}_2$ is to be measured. As contact *c* is moved along the resistance wire, the resistance R_{cb} between points *c* and *b* varies; if the resistance wire is uniform, R_{cb} is proportional to the length of wire between *c* and *b*. To determine the value of $\mathcal{E}_2$, contact *c* is moved until a position is found at which the galvanometer shows no deflection; this corresponds to zero current passing through $\mathcal{E}_2$. With $I_2 = 0$, Kirchhoff's loop rule gives

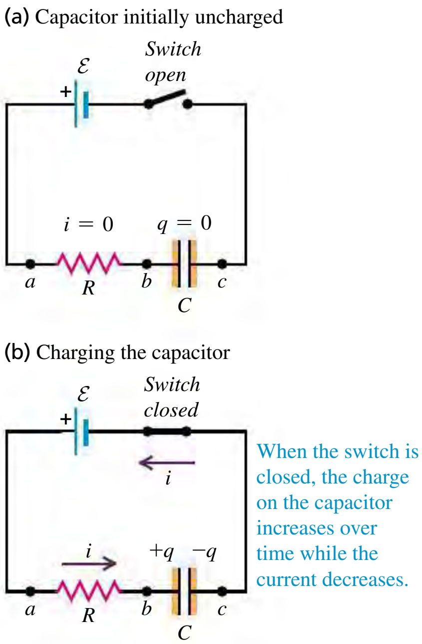
$$\mathcal{E}_2 = IR_{cb}$$

With $I_2 = 0$, the current I produced by the emf $\mathcal{E}_1$ has the same value no matter what the value of the emf $\mathcal{E}_2$. We calibrate the device by replacing $\mathcal{E}_2$ by a source of known emf; then to find any unknown emf $\mathcal{E}_2$, we measure the length of wire *cb* for which $I_2 = 0$. Note: For this to work, V_{ab} must be greater than $\mathcal{E}_2$.

The term *potentiometer* is also used for any variable resistor, usually having a circular resistance element and a sliding contact controlled by a rotating shaft and knob. The circuit symbol for a potentiometer is shown in Fig. 26.19b.

TEST YOUR UNDERSTANDING OF SECTION 26.3 You want to measure the current through and the potential difference across the 2-Ω resistor shown in Fig. 26.12 (Example 26.6 in Section 26.2). (a) How should you connect an ammeter and a voltmeter to do this? (i) Both ammeter and voltmeter in series with the 2-Ω resistor; (ii) ammeter in series with the 2-Ω resistor and voltmeter connected between points *b* and *d*; (iii) ammeter connected between points *b* and *d* and voltmeter in series with the 2-Ω resistor; (iv) both ammeter and voltmeter connected between points *b* and *d*. (b) What resistances should these meters have? (i) Both ammeter and voltmeter resistances should be much greater than 2 Ω; (ii) ammeter resistance should be much greater than 2 Ω and voltmeter resistance should be much less than 2 Ω; (iii) ammeter resistance should be much less than 2 Ω and voltmeter resistance should be much greater than 2 Ω; (iv) both ammeter and voltmeter resistances should be much less than 2 Ω. ▌

26.20 Charging a capacitor. (a) Just before the switch is closed, the charge q is zero. (b) When the switch closes (at $t = 0$), the current jumps from zero to $\mathcal{E}/R$. As time passes, q approaches Q_f and the current i approaches zero.



26.4 *R-C* CIRCUITS

In the circuits we have analyzed up to this point, we have assumed that all the emfs and resistances are *constant* (time independent) so that all the potentials, currents, and powers are also independent of time. But in the simple act of charging or discharging a capacitor we find a situation in which the currents, voltages, and powers *do* change with time.

Many devices incorporate circuits in which a capacitor is alternately charged and discharged. These include flashing traffic lights, automobile turn signals, and electronic flash units. Understanding what happens in such circuits is thus of great practical importance.

Charging a Capacitor

Figure 26.20 shows a simple circuit for charging a capacitor. A circuit such as this that has a resistor and a capacitor in series is called an ***R-C* circuit.** We idealize the battery (or power supply) to have a constant emf $\mathcal{E}$ and zero internal resistance $(r = 0)$, and we ignore the resistance of all the connecting conductors.

We begin with the capacitor initially uncharged (Fig. 26.20a); then at some initial time $t = 0$ we close the switch, completing the circuit and permitting current around the loop to begin charging the capacitor (Fig. 26.20b). For all practical purposes, the current begins at the same instant in every conducting part of the circuit, and at each instant the current is the same in every part.

Because the capacitor in Fig. 26.20 is initially uncharged, the potential difference v_{bc} across it is zero at $t = 0$. At this time, from Kirchhoff's loop law, the voltage v_{ab} across the resistor R is equal to the battery emf $\mathcal{E}$. The initial $(t = 0)$ current through the resistor, which we will call I_0, is given by Ohm's law: $I_0 = v_{ab}/R = \mathcal{E}/R$.

As the capacitor charges, its voltage v_{bc} increases and the potential difference v_{ab} across the resistor decreases, corresponding to a decrease in current. The sum of these two voltages is constant and equal to $\mathcal{E}$. After a long time the capacitor is fully charged, the current decreases to zero, and v_{ab} across the resistor becomes zero. Then the entire battery emf $\mathcal{E}$ appears across the capacitor and $v_{bc} = \mathcal{E}$.

Let q represent the charge on the capacitor and i the current in the circuit at some time t after the switch has been closed. We choose the positive direction for the current to correspond to positive charge flowing onto the left-hand capacitor plate, as in Fig. 26.20b. The instantaneous potential differences v_{ab} and v_{bc} are

$$v_{ab} = iR \qquad v_{bc} = \frac{q}{C}$$

Using these in Kirchhoff's loop rule, we find

$$\mathcal{E} - iR - \frac{q}{C} = 0 \tag{26.9}$$

The potential drops by an amount iR as we travel from a to b and by q/C as we travel from b to c. Solving Eq. (26.9) for i, we find

$$i = \frac{\mathcal{E}}{R} - \frac{q}{RC} \tag{26.10}$$

At time $t = 0$, when the switch is first closed, the capacitor is uncharged, and so $q = 0$. Substituting $q = 0$ into Eq. (26.10), we find that the *initial* current I_0 is given by $I_0 = \mathcal{E}/R$, as we have already noted. If the capacitor were not in the circuit, the last term in Eq. (26.10) would not be present; then the current would be *constant* and equal to $\mathcal{E}/R$.

As the charge q increases, the term q/RC becomes larger and the capacitor charge approaches its final value, which we will call Q_f. The current decreases and eventually becomes zero. When $i = 0$, Eq. (26.10) gives

$$\frac{\mathcal{E}}{R} = \frac{Q_f}{RC} \qquad Q_f = C\mathcal{E} \tag{26.11}$$

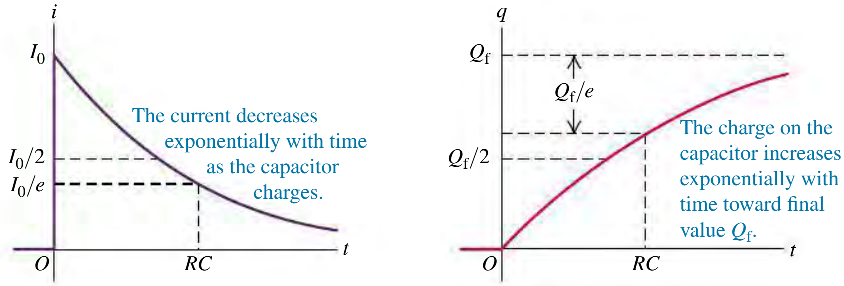
Note that the final charge Q_f does not depend on R.

Figure 26.21 shows the current and capacitor charge as functions of time. At the instant the switch is closed $(t = 0)$, the current jumps from zero to its initial value $I_0 = \mathcal{E}/R$; after that, it gradually approaches zero. The capacitor charge starts at zero and gradually approaches the final value given by Eq. (26.11), $Q_f = C\mathcal{E}$.

CAUTION **Lowercase means time-varying** Up to this point we have been working with constant potential differences (voltages), currents, and charges, and we have used *capital* letters V, I, and Q, respectively, to denote these quantities. To distinguish between quantities that vary with time and those that are constant, we will use *lowercase* letters v, i, and q for time-varying voltages, currents, and charges, respectively. We suggest that you follow this same convention in your own work.

(a) Graph of current versus time for a charging capacitor

(b) Graph of capacitor charge versus time for a charging capacitor



26.21 Current i and capacitor charge q as functions of time for the circuit of Fig. 26.20. The initial current is I_0 and the initial capacitor charge is zero. The current asymptotically approaches zero, and the capacitor charge asymptotically approaches a final value of Q_f.

We can derive general expressions for charge q and current i as functions of time. With our choice of the positive direction for current (Fig. 26.20b), i equals the rate at which positive charge arrives at the left-hand (positive) plate of the capacitor, so $i = dq/dt$. Making this substitution in Eq. (26.10), we have

$$\frac{dq}{dt} = \frac{\mathcal{E}}{R} - \frac{q}{RC} = -\frac{1}{RC}(q - C\mathcal{E})$$

We can rearrange this to

$$\frac{dq}{q - C\mathcal{E}} = -\frac{dt}{RC}$$

and then integrate both sides. We change the integration variables to q' and t' so that we can use q and t for the upper limits. The lower limits are $q' = 0$ and $t' = 0$:

$$\int_0^q \frac{dq'}{q' - C\mathcal{E}} = -\int_0^t \frac{dt'}{RC}$$

When we carry out the integration, we get

$$\ln\left(\frac{q - C\mathcal{E}}{-C\mathcal{E}}\right) = -\frac{t}{RC}$$

Exponentiating both sides (that is, taking the inverse logarithm) and solving for q, we find

$$\frac{q - C\mathcal{E}}{-C\mathcal{E}} = e^{-t/RC}$$

PhET: Circuit Construction Kit (AC+DC)
PhET: Circuit Construction Kit (DC Only)

***R-C* circuit, charging capacitor:**

Capacitor charge · Capacitance · Final capacitor charge $= C\mathcal{E}$

$$q = C\mathcal{E}(1 - e^{-t/RC}) = Q_f(1 - e^{-t/RC}) \quad (26.12)$$

Battery emf · Time since switch closed · Resistance

The instantaneous current i is just the time derivative of Eq. (26.12):

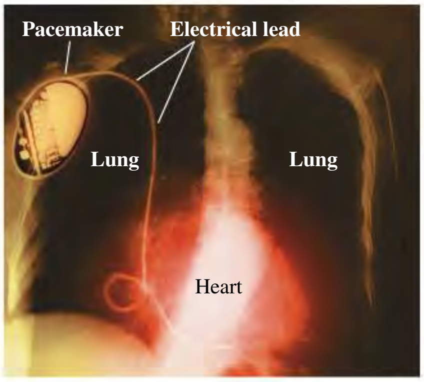
***R-C* circuit, charging capacitor:**

Current · Battery emf · Time since switch closed · Initial current $= \mathcal{E}/R$

$$i = \frac{dq}{dt} = \frac{\mathcal{E}}{R}e^{-t/RC} = I_0 e^{-t/RC} \quad (26.13)$$

Rate of change of capacitor charge · Resistance · Capacitance

BIO Application Pacemakers and Capacitors This x-ray image shows a pacemaker implanted in a patient with a malfunctioning sinoatrial node, the part of the heart that generates the electrical signal to trigger heartbeats. The pacemaker circuit contains a battery, a capacitor, and a computer-controlled switch. To maintain regular beating, once per second the switch discharges the capacitor and sends an electrical pulse along the lead to the heart. The switch then flips to allow the capacitor to recharge for the next pulse.



The charge and current are both *exponential* functions of time. Figure 26.21a is a graph of Eq. (26.13) and Fig. 26.21b is a graph of Eq. (26.12).

Time Constant

After a time equal to RC, the current in the R-C circuit has decreased to $1/e$ (about 0.368) of its initial value. At this time, the capacitor charge has reached $(1 - 1/e) = 0.632$ of its final value $Q_f = C\mathcal{E}$. The product RC is therefore a measure of how quickly the capacitor charges. We call RC the **time constant,** or the **relaxation time**, of the circuit, denoted by τ:

$$\tau = RC \quad \text{(time constant for } R\text{-}C \text{ circuit)} \quad (26.14)$$

When τ is small, the capacitor charges quickly; when it is larger, the charging takes more time. If the resistance is small, it's easier for current to flow, and the capacitor charges more quickly. If R is in ohms and C in farads, τ is in seconds.

In Fig. 26.21a the horizontal axis is an *asymptote* for the curve. Strictly speaking, i never becomes exactly zero. But the longer we wait, the closer it gets. After a time equal to $10RC$, the current has decreased to 0.000045 of its initial value. Similarly, the curve in Fig. 26.21b approaches the horizontal dashed line labeled Q_f as an asymptote. The charge q never attains exactly this value, but after a time equal to $10RC$, the difference between q and Q_f is only 0.000045 of Q_f. We invite you to verify that the product RC has units of time.

Discharging a Capacitor

Now suppose that after the capacitor in Fig. 26.21b has acquired a charge Q_0, we remove the battery from our *R-C* circuit and connect points a and c to an open switch (**Fig. 26.22a**). We then close the switch and at the same instant reset our stopwatch to $t = 0$; at that time, $q = Q_0$. The capacitor then *discharges* through the resistor, and its charge eventually decreases to zero.

Again let i and q represent the time-varying current and charge at some instant after the connection is made. In Fig. 26.22b we make the same choice of the positive direction for current as in Fig. 26.20b. Then Kirchhoff's loop rule gives Eq. (26.10) but with $\mathcal{E} = 0$; that is,

$$i = \frac{dq}{dt} = -\frac{q}{RC} \tag{26.15}$$

The current i is now negative; this is because positive charge q is leaving the left-hand capacitor plate in Fig. 26.22b, so the current is in the direction opposite to that shown. At time $t = 0$, when $q = Q_0$, the initial current is $I_0 = -Q_0/RC$.

To find q as a function of time, we rearrange Eq. (26.15), again change the variables to q' and t', and integrate. This time the limits for q' are Q_0 to q:

$$\int_{Q_0}^{q} \frac{dq'}{q'} = -\frac{1}{RC}\int_0^t dt'$$

$$\ln\frac{q}{Q_0} = -\frac{t}{RC}$$

***R-C* circuit, discharging capacitor:**

$$q = Q_0 e^{-t/RC} \tag{26.16}$$

Capacitor charge (q); Initial capacitor charge (Q_0); Capacitance (C); Time since switch closed (t); Resistance (R)

The instantaneous current i is the derivative of this with respect to time:

***R-C* circuit, discharging capacitor:**

$$i = \frac{dq}{dt} = -\frac{Q_0}{RC}e^{-t/RC} = I_0 e^{-t/RC} \tag{26.17}$$

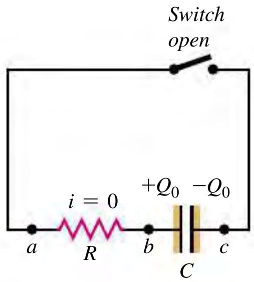
Current (i); Initial capacitor charge (Q_0); Capacitance (C); Time since switch closed (t); Rate of change of capacitor charge (dq/dt); Resistance (R); Initial current $= -Q_0/RC$ (I_0)

We graph the current and the charge in **Fig. 26.23**; both quantities approach zero exponentially with time. Comparing these results with Eqs. (26.12) and (26.13), we note that the expressions for the current are identical, apart from the sign of I_0. The capacitor charge approaches zero asymptotically in Eq. (26.16), while the *difference* between q and Q approaches zero asymptotically in Eq. (26.12).

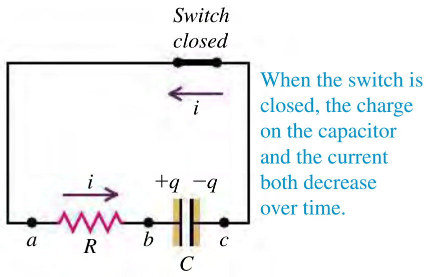
Energy considerations give us additional insight into the behavior of an *R-C* circuit. While the capacitor is charging, the instantaneous rate at which the battery delivers energy to the circuit is $P = \mathcal{E}i$. The instantaneous rate at which

26.22 Discharging a capacitor. (a) Before the switch is closed at time $t = 0$, the capacitor charge is Q_0 and the current is zero. (b) At time t after the switch is closed, the capacitor charge is q and the current is i. The actual current direction is opposite to the direction shown; i is negative. After a long time, q and i both approach zero.

(a) Capacitor initially charged

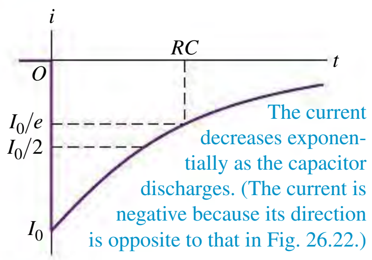


(b) Discharging the capacitor

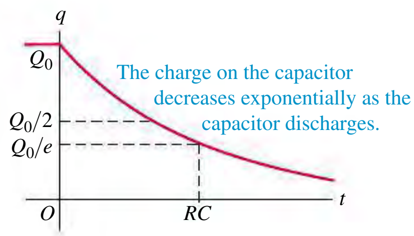


26.23 Current i and capacitor charge q as functions of time for the circuit of Fig. 26.22. The initial current is I_0 and the initial capacitor charge is Q_0. Both i and q asymptotically approach zero.

(a) Graph of current versus time for a discharging capacitor



(b) Graph of capacitor charge versus time for a discharging capacitor



electrical energy is dissipated in the resistor is i^2R, and the rate at which energy is stored in the capacitor is $iv_{bc} = iq/C$. Multiplying Eq. (26.9) by i, we find

$$\mathcal{E}i = i^2R + \frac{iq}{C} \tag{26.18}$$

This means that of the power $\mathcal{E}i$ supplied by the battery, part (i^2R) is dissipated in the resistor and part (iq/C) is stored in the capacitor.

The *total* energy supplied by the battery during charging of the capacitor equals the battery emf $\mathcal{E}$ multiplied by the total charge Q_f, or $\mathcal{E}Q_f$. The total energy stored in the capacitor, from Eq. (24.9), is $Q_f\mathcal{E}/2$. Thus, of the energy supplied by the battery, *exactly half* is stored in the capacitor, and the other half is dissipated in the resistor. This half-and-half division of energy doesn't depend on C, R, or $\mathcal{E}$. You can verify this result by taking the integral over time of each of the power quantities in Eq. (26.18).

EXAMPLE 26.12 CHARGING A CAPACITOR

SOLUTION

A 10-MΩ resistor is connected in series with a 1.0-μF capacitor and a battery with emf 12.0 V. Before the switch is closed at time $t = 0$, the capacitor is uncharged. (a) What is the time constant? (b) What fraction of the final charge Q_f is on the capacitor at $t = 46$ s? (c) What fraction of the initial current I_0 is still flowing at $t = 46$ s?

SOLUTION

IDENTIFY and SET UP: This is the situation shown in Fig. 26.20, with $R = 10\ \text{M}\Omega$, $C = 1.0\ \mu\text{F}$, and $\mathcal{E} = 12.0$ V. The charge q and current i vary with time as shown in Fig. 26.21. Our target variables are (a) the time constant τ, (b) the ratio q/Q_f at $t = 46$ s, and (c) the ratio i/I_0 at $t = 46$ s. Equation (26.14) gives τ. For a capacitor being charged, Eq. (26.12) gives q and Eq. (26.13) gives i.

EXECUTE: (a) From Eq. (26.14),

$$\tau = RC = (10 \times 10^6\ \Omega)(1.0 \times 10^{-6}\ \text{F}) = 10\ \text{s}$$

(b) From Eq. (26.12),

$$\frac{q}{Q_f} = 1 - e^{-t/RC} = 1 - e^{-(46\ \text{s})/(10\ \text{s})} = 0.99$$

(c) From Eq. (26.13),

$$\frac{i}{I_0} = e^{-t/RC} = e^{-(46\ \text{s})/(10\ \text{s})} = 0.010$$

EVALUATE: After 4.6 time constants the capacitor is 99% charged and the charging current has decreased to 1.0% of its initial value. The circuit would charge more rapidly if we reduced the time constant by using a smaller resistance.

EXAMPLE 26.13 DISCHARGING A CAPACITOR

SOLUTION

The resistor and capacitor of Example 26.12 are reconnected as shown in Fig. 26.22. The capacitor has an initial charge of 5.0 μC and is discharged by closing the switch at $t = 0$. (a) At what time will the charge be 0.50 μC? (b) What is the current at this time?

SOLUTION

IDENTIFY and SET UP: Now the capacitor is being discharged, so q and i vary with time as in Fig. 26.23, with $Q_0 = 5.0 \times 10^{-6}$ C. Again we have $RC = \tau = 10$ s. Our target variables are (a) the value of t at which $q = 0.50\ \mu$C and (b) the value of i at this time. We first solve Eq. (26.16) for t, and then solve Eq. (26.17) for i.

EXECUTE: (a) Solving Eq. (26.16) for the time t gives

$$t = -RC \ln\frac{q}{Q_0} = -(10\ \text{s})\ln\frac{0.50\ \mu\text{C}}{5.0\ \mu\text{C}} = 23\ \text{s} = 2.3\tau$$

(b) From Eq. (26.17), with $Q_0 = 5.0\ \mu\text{C} = 5.0 \times 10^{-6}$ C,

$$i = -\frac{Q_0}{RC}e^{-t/RC} = -\frac{5.0 \times 10^{-6}\ \text{C}}{10\ \text{s}}e^{-2.3} = -5.0 \times 10^{-8}\ \text{A}$$

EVALUATE: The current in part (b) is negative because i has the opposite sign when the capacitor is discharging than when it is charging. Note that we could have avoided evaluating $e^{-t/RC}$ by noticing that at the time in question, $q = 0.10Q_0$; from Eq. (26.16) this means that $e^{-t/RC} = 0.10$.

TEST YOUR UNDERSTANDING OF SECTION 26.4 The energy stored in a capacitor is equal to $q^2/2C$. When a capacitor is discharged, what fraction of the initial energy remains after an elapsed time of one time constant? (i) $1/e$; (ii) $1/e^2$; (iii) $1 - 1/e$; (iv) $(1 - 1/e)^2$; (v) answer depends on how much energy was stored initially.

26.5 POWER DISTRIBUTION SYSTEMS

We conclude this chapter with a brief discussion of practical household and automotive electric-power distribution systems. Automobiles use direct-current (dc) systems, while nearly all household, commercial, and industrial systems use alternating current (ac) because of the ease of stepping voltage up and down with transformers. Most of the same basic wiring concepts apply to both. We'll talk about alternating-current circuits in greater detail in Chapter 31.

The various lamps, motors, and other appliances to be operated are always connected in *parallel* to the power source (the wires from the power company for houses, or from the battery and alternator for a car). If appliances were connected in series, shutting one appliance off would shut them all off (see Example 26.2 in Section 26.1). **Figure 26.24** shows the basic idea of house wiring. One side of the "line," as the pair of conductors is called, is called the *neutral* side; it is always connected to "ground" at the entrance panel. For houses, *ground* is an actual electrode driven into the earth (which is usually a good conductor) or sometimes connected to the household water pipes. Electricians speak of the "hot" side and the "neutral" side of the line. Most modern house wiring systems have *two* hot lines with opposite polarity with respect to the neutral. We'll return to this detail later.

Household voltage is nominally 120 V in the United States and Canada, and often 240 V in Europe. (For alternating current, which varies sinusoidally with time, these numbers represent the *root-mean-square* voltage, which is $1/\sqrt{2}$ times the peak voltage. We'll discuss this further in Section 31.1.) The amount of current I drawn by a given device is determined by its power input P, given by Eq. (25.17): $P = VI$. Hence $I = P/V$. For example, the current in a 100-W light bulb is

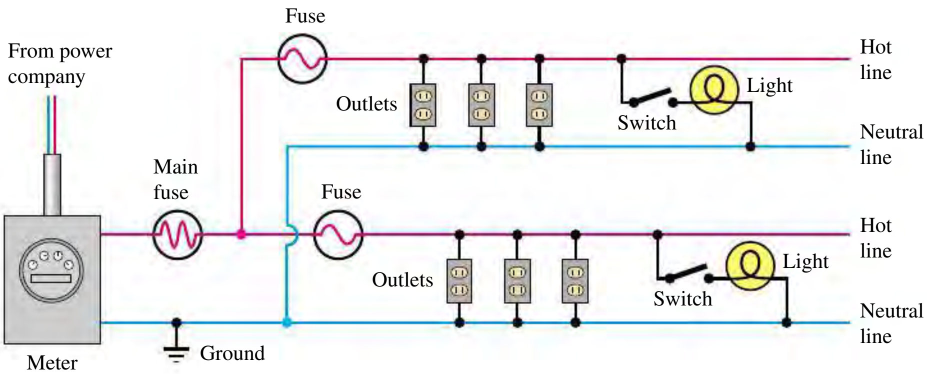
$$I = \frac{P}{V} = \frac{100\text{ W}}{120\text{ V}} = 0.83\text{ A}$$

The power input to this bulb is actually determined by its resistance R. Using Eq. (25.18), which states that $P = VI = I^2R = V^2/R$ for a resistor, we get the resistance of this bulb at operating temperature:

$$R = \frac{V}{I} = \frac{120\text{ V}}{0.83\text{ A}} = 144\ \Omega \qquad \text{or} \qquad R = \frac{V^2}{P} = \frac{(120\text{ V})^2}{100\text{ W}} = 144\ \Omega$$

Similarly, a 1500-W waffle iron draws a current of $(1500\text{ W})/(120\text{ V}) = 12.5\text{ A}$ and has a resistance, at operating temperature, of $9.6\ \Omega$. Because of the temperature dependence of resistivity, the resistances of these devices are considerably less when they are cold. If you measure the resistance of a 100-W light bulb with

26.24 Schematic diagram of part of a house wiring system. Only two branch circuits are shown; an actual system might have four to thirty branch circuits. Lamps and appliances may be plugged into the outlets. The grounding wires, which normally carry no current, are not shown.



an ohmmeter (whose small current causes very little temperature rise), you will probably get a value of about 10 Ω. When a light bulb is turned on, this low resistance causes an initial surge of current until the filament heats up. That's why a light bulb that's ready to burn out nearly always does so just when you turn it on.

Circuit Overloads and Short Circuits

The maximum current available from an individual circuit is limited by the resistance of the wires. As we discussed in Section 25.5, the I^2R power loss in the wires causes them to become hot, and in extreme cases this can cause a fire or melt the wires. Ordinary lighting and outlet wiring in houses usually uses 12-gauge wire. This has a diameter of 2.05 mm and can carry a maximum current of 20 A safely (without overheating). Larger-diameter wires of the same length have lower resistance [see Eq. (25.10)]. Hence 8-gauge (3.26 mm) or 6-gauge (4.11 mm) is used for high-current appliances such as clothes dryers, and 2-gauge (6.54 mm) or larger is used for the main power lines entering a house.

26.25 (a) Excess current will melt the thin wire of lead–tin alloy that runs along the length of a fuse, inside the transparent housing. (b) The switch on this circuit breaker will flip if the maximum allowable current is exceeded.

(a)

(b)

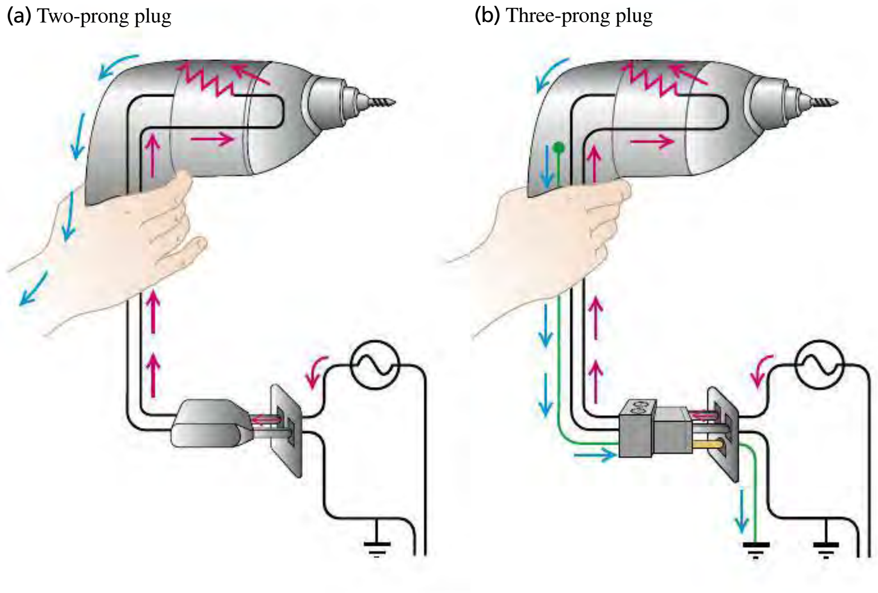
Protection against overloading and overheating of circuits is provided by fuses or circuit breakers. A *fuse* contains a link of lead–tin alloy with a very low melting temperature; the link melts and breaks the circuit when its rated current is exceeded (**Fig. 26.25a**). A *circuit breaker* is an electromechanical device that performs the same function, using an electromagnet or a bimetallic strip to "trip" the breaker and interrupt the circuit when the current exceeds a specified value (Fig. 26.25b). Circuit breakers have the advantage that they can be reset after they are tripped, while a blown fuse must be replaced.

CAUTION **Fuses** If your system has fuses and you plug too many high-current appliances into the same outlet, the fuse blows. Do not replace the fuse with one that has a higher rating; if you do, you risk overheating the wires and starting a fire. The only safe solution is to distribute the appliances among several circuits. Modern kitchens often have three or four separate 20-A circuits.

Contact between the hot and neutral sides of the line causes a *short circuit*. Such a situation, which can be caused by faulty insulation or by a variety of mechanical malfunctions, provides a very low-resistance current path, permitting a very large current that would quickly melt the wires and ignite their insulation if the current were not interrupted by a fuse or circuit breaker (see Example 25.10 in Section 25.5). An equally dangerous situation is a broken wire that interrupts the current path, creating an *open circuit*. This is hazardous because of the sparking that can occur at the point of intermittent contact.

In approved wiring practice, a fuse or breaker is placed *only* in the hot side of the line, never in the neutral side. Otherwise, if a short circuit should develop because of faulty insulation or other malfunction, the ground-side fuse could blow. The hot side would still be live and would pose a shock hazard if you touched the live conductor and a grounded object such as a water pipe. For similar reasons the wall switch for a light fixture is always in the hot side of the line, never the neutral side.

Further protection against shock hazard is provided by a third conductor called the *grounding wire,* included in all present-day wiring. This conductor corresponds to the long round or U-shaped prong of the three-prong connector plug on an appliance or power tool. It is connected to the neutral side of the line at the entrance panel. The grounding wire normally carries no current, but it connects the metal case or frame of the device to ground. If a conductor on the hot side of the line accidentally contacts the frame or case, the grounding conductor provides a current path, and the fuse blows. Without the ground wire, the frame could become "live"—that is, at a potential 120 V above ground. Then if you touched it and a water pipe (or even a damp floor) at the same time, you could



26.26 (a) If a malfunctioning electric drill is connected to a wall socket via a two-prong plug, a person may receive a shock. (b) When the drill malfunctions when connected via a three-prong plug, a person touching it receives no shock, because electric charge flows through the ground wire (shown in green) to the third prong and into the ground rather than into the person's body. If the ground current is appreciable, the fuse blows.

get a dangerous shock (**Fig. 26.26**). In some situations, especially outlets located outdoors or near a sink or other water pipes, a special kind of circuit breaker called a *ground-fault interrupter* (GFI or GFCI) is used. This device senses the difference in current between the hot and neutral conductors (which is normally zero) and trips when some very small value, typically 5 mA, is exceeded.

Household and Automotive Wiring

Most modern household wiring systems actually use a slight elaboration of the system described above. The power company provides *three* conductors. One is neutral; the other two are both at 120 V with respect to the neutral but with opposite polarity, giving a voltage between them of 240 V. The power company calls this a *three-wire line,* in contrast to the 120-V two-wire (plus ground wire) line described above. With a three-wire line, 120-V lamps and appliances can be connected between neutral and either hot conductor, and high-power devices requiring 240 V, such as electric ranges and clothes dryers, are connected between the two hot lines.

All of the above discussion can be applied directly to automobile wiring. The voltage is about 13 V (direct current); the power is supplied by the battery and by the alternator, which charges the battery when the engine is running. The neutral side of each circuit is connected to the body and frame of the vehicle. For this low voltage a separate grounding conductor is not required for safety. The fuse or circuit breaker arrangement is the same in principle as in household wiring. Because of the lower voltage (less energy per charge), more current (a greater number of charges per second) is required for the same power; a 100-W headlight bulb requires a current of about $(100\ \text{W})/(13\ \text{V}) = 8\ \text{A}$.

Although we spoke of *power* in the above discussion, what we buy from the power company is *energy*. Power is energy transferred per unit time, so energy is average power multiplied by time. The usual unit of energy sold by the power company is the kilowatt-hour $(1\ \text{kW}\cdot\text{h})$:

$$1\ \text{kW}\cdot\text{h} = (10^3\ \text{W})(3600\ \text{s}) = 3.6 \times 10^6\ \text{W}\cdot\text{s} = 3.6 \times 10^6\ \text{J}$$

In the United States, one kilowatt-hour typically costs 8 to 27 cents, depending on the location and quantity of energy purchased. To operate a 1500-W (1.5-kW) waffle iron continuously for 1 hour requires 1.5 kW · h of energy; at 10 cents per kilowatt-hour, the energy cost is 15 cents. The cost of operating any lamp or appliance for a specified time can be calculated in the same way if the power rating is known. However, many electric cooking utensils (including waffle irons) cycle on and off to maintain a constant temperature, so the average power may be less than the power rating marked on the device.

EXAMPLE 26.14 A KITCHEN CIRCUIT

SOLUTION

An 1800-W toaster, a 1.3-kW electric frying pan, and a 100-W lamp are plugged into the same 20-A, 120-V circuit. (a) What current is drawn by each device, and what is the resistance of each device? (b) Will this combination trip the circuit breaker?

SOLUTION

IDENTIFY and SET UP: When plugged into the same circuit, the three devices are connected in parallel, so the voltage across each appliance is $V = 120\text{ V}$. We find the current I drawn by each device from the relationship $P = VI$, where P is the power input of the device. To find the resistance R of each device we use the relationship $P = V^2/R$.

EXECUTE: (a) To simplify the calculation of current and resistance, we note that $I = P/V$ and $R = V^2/P$. Hence

$$I_{\text{toaster}} = \frac{1800\text{ W}}{120\text{ V}} = 15\text{ A} \qquad R_{\text{toaster}} = \frac{(120\text{ V})^2}{1800\text{ W}} = 8\ \Omega$$

$$I_{\text{frying pan}} = \frac{1300\text{ W}}{120\text{ V}} = 11\text{ A} \qquad R_{\text{frying pan}} = \frac{(120\text{ V})^2}{1300\text{ W}} = 11\ \Omega$$

$$I_{\text{lamp}} = \frac{100\text{ W}}{120\text{ V}} = 0.83\text{ A} \qquad R_{\text{lamp}} = \frac{(120\text{ V})^2}{100\text{ W}} = 144\ \Omega$$

For constant voltage the device with the *least* resistance (in this case the toaster) draws the most current and receives the most power.

(b) The total current through the line is the sum of the currents drawn by the three devices:

$$\begin{aligned} I &= I_{\text{toaster}} + I_{\text{frying pan}} + I_{\text{lamp}} \\ &= 15\text{ A} + 11\text{ A} + 0.83\text{ A} = 27\text{ A} \end{aligned}$$

This exceeds the 20-A rating of the line, and the circuit breaker will indeed trip.

EVALUATE: We could also find the total current by using $I = P/V$ and the total power P delivered to all three devices:

$$\begin{aligned} I &= \frac{P_{\text{toaster}} + P_{\text{frying pan}} + P_{\text{lamp}}}{V} \\ &= \frac{1800\text{ W} + 1300\text{ W} + 100\text{ W}}{120\text{ V}} = 27\text{ A} \end{aligned}$$

A third way to determine I is to use $I = V/R_{\text{eq}}$, where R_{eq} is the equivalent resistance of the three devices in parallel:

$$I = \frac{V}{R_{\text{eq}}} = (120\text{ V})\left(\frac{1}{8\ \Omega} + \frac{1}{11\ \Omega} + \frac{1}{144\ \Omega}\right) = 27\text{ A}$$

Appliances with such current demands are common, so modern kitchens have more than one 20-A circuit. To keep currents safely below 20 A, the toaster and frying pan should be plugged into different circuits.

TEST YOUR UNDERSTANDING OF SECTION 26.5 To prevent the circuit breaker in Example 26.14 from blowing, a home electrician replaces the circuit breaker with one rated at 40 A. Is this a reasonable thing to do? ❙

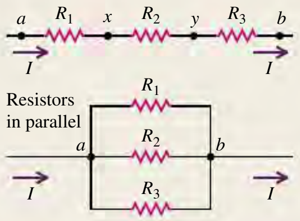
CHAPTER 26 SUMMARY

SOLUTIONS TO ALL EXAMPLES

Resistors in series and parallel: When several resistors $R_1, R_2, R_3, \ldots$ are connected in series, the equivalent resistance R_{eq} is the sum of the individual resistances. The same *current* flows through all the resistors in a series connection. When several resistors are connected in parallel, the reciprocal of equivalent resistance R_{eq} is the sum of the reciprocals of the individual resistances. All resistors in a parallel connection have the same *potential difference* between their terminals. (See Examples 26.1 and 26.2.)

$$R_{eq} = R_1 + R_2 + R_3 + \cdots \quad (26.1)$$
(resistors in series)

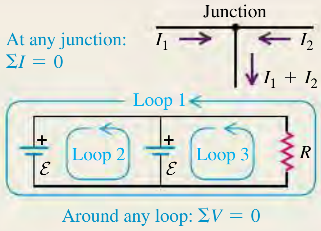
$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \quad (26.2)$$
(resistors in parallel)



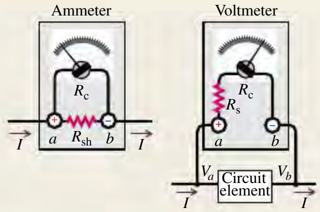
Kirchhoff's rules: Kirchhoff's junction rule is based on conservation of charge. It states that the algebraic sum of the currents into any junction must be zero. Kirchhoff's loop rule is based on conservation of energy and the conservative nature of electrostatic fields. It states that the algebraic sum of potential differences around any loop must be zero. Careful use of consistent sign rules is essential in applying Kirchhoff's rules. (See Examples 26.3–26.7.)

$$\sum I = 0 \quad \text{(junction rule)} \quad (26.5)$$

$$\sum V = 0 \quad \text{(loop rule)} \quad (26.6)$$



Electrical measuring instruments: In a d'Arsonval galvanometer, the deflection is proportional to the current in the coil. For a larger current range, a shunt resistor is added, so some of the current bypasses the meter coil. Such an instrument is called an ammeter. If the coil and any additional series resistance included obey Ohm's law, the meter can also be calibrated to read potential difference or voltage. The instrument is then called a voltmeter. A good ammeter has very low resistance; a good voltmeter has very high resistance. (See Examples 26.8–26.11.)



R-C circuits: When a capacitor is charged by a battery in series with a resistor, the current and capacitor charge are not constant. The charge approaches its final value asymptotically and the current approaches zero asymptotically. The charge and current in the circuit are given by Eqs. (26.12) and (26.13). After a time $\tau = RC$, the charge has approached within $1/e$ of its final value. This time is called the time constant or relaxation time of the circuit. When the capacitor discharges, the charge and current are given as functions of time by Eqs. (26.16) and (26.17). The time constant is the same for charging and discharging. (See Examples 26.12 and 26.13.)

Capacitor charging:

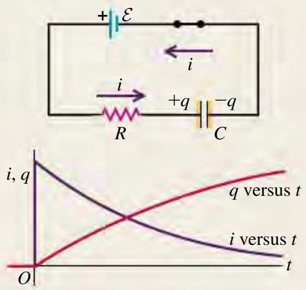
$$q = C\mathcal{E}\left(1 - e^{-t/RC}\right) = Q_f\left(1 - e^{-t/RC}\right) \quad (26.12)$$

$$i = \frac{dq}{dt} = \frac{\mathcal{E}}{R}e^{-t/RC} = I_0 e^{-t/RC} \quad (26.13)$$

Capacitor discharging:

$$q = Q_0 e^{-t/RC} \quad (26.16)$$

$$i = \frac{dq}{dt} = -\frac{Q_0}{RC}e^{-t/RC} = I_0 e^{-t/RC} \quad (26.17)$$



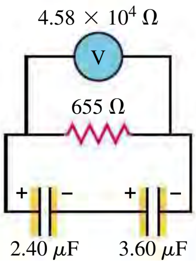
Household wiring: In household wiring systems, the various electrical devices are connected in parallel across the power line, which consists of a pair of conductors, one "hot" and the other "neutral." An additional "ground" wire is included for safety. The maximum permissible current in a circuit is determined by the size of the wires and the maximum temperature they can tolerate. Protection against excessive current and the resulting fire hazard is provided by fuses or circuit breakers. (See Example 26.14.)

BRIDGING PROBLEM TWO CAPACITORS AND TWO RESISTORS

SOLUTION

A 2.40-μF capacitor and a 3.60-μF capacitor are connected in series. (a) A charge of 5.20 mC is placed on each capacitor. What is the energy stored in the capacitors? (b) A 655-Ω resistor is connected to the terminals of the capacitor combination, and a voltmeter with resistance 4.58×10^4 Ω is connected across the resistor (**Fig. 26.27**). What is the rate of change of the energy stored in the capacitors just after the connection is made? (c) How long after the connection is made has the energy stored in the capacitors decreased to $1/e$ of its initial value? (d) At the instant calculated in part (c), what is the rate of change of the energy stored in the capacitors?

26.27 When the connection is made, the charged capacitors discharge.



SOLUTION GUIDE

IDENTIFY AND SET UP

1. The two capacitors act as a single equivalent capacitor (see Section 24.2), and the resistor and voltmeter act as a single equivalent resistor. Select equations that will allow you to calculate the values of these equivalent circuit elements.
2. In part (a) you will need to use Eq. (24.9), which gives the energy stored in a capacitor.
3. For parts (b), (c), and (d), you will need to use Eq. (24.9) as well as Eqs. (26.16) and (26.17), which give the capacitor charge and current as functions of time. (*Hint:* The rate at which energy is lost by the capacitors equals the rate at which energy is dissipated in the resistances.)

EXECUTE

4. Find the stored energy at $t = 0$.
5. Find the rate of change of the stored energy at $t = 0$.
6. Find the value of t at which the stored energy has $1/e$ of the value you found in step 4.
7. Find the rate of change of the stored energy at the time you found in step 6.

EVALUATE

8. Check your results from steps 5 and 7 by calculating the rate of change in a different way. (*Hint:* The rate of change of the stored energy U is dU/dt.)

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

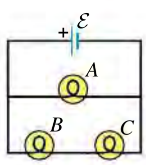
Q26.1 In which 120-V light bulb does the filament have greater resistance: a 60-W bulb or a 120-W bulb? If the two bulbs are connected to a 120-V line in series, through which bulb will there be the greater voltage drop? What if they are connected in parallel? Explain your reasoning.

Q26.2 Two 120-V light bulbs, one 25-W and one 200-W, were connected in series across a 240-V line. It seemed like a good idea at the time, but one bulb burned out almost immediately. Which one burned out, and why?

Q26.3 You connect a number of identical light bulbs to a flashlight battery. (a) What happens to the brightness of each bulb as more and more bulbs are added to the circuit if you connect them (i) in series and (ii) in parallel? (b) Will the battery last longer if the bulbs are in series or in parallel? Explain your reasoning.

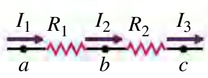
Q26.4 In the circuit shown in **Fig. Q26.4**, three identical light bulbs are connected to a flashlight battery. How do the brightnesses of the bulbs compare? Which light bulb has the greatest current passing through it? Which light bulb has the greatest potential difference between its terminals? What happens if bulb A is unscrewed? Bulb B? Bulb C? Explain your reasoning.

Figure **Q26.4**



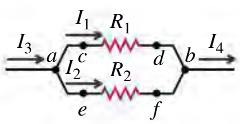
Q26.5 If two resistors R_1 and R_2 ($R_2 > R_1$) are connected in series as shown in **Fig. Q26.5**, which of the following must be true? In each case justify your answer. (a) $I_1 = I_2 = I_3$. (b) The current is greater in R_1 than in R_2. (c) The electrical power consumption is the same for both resistors. (d) The electrical power consumption is greater in R_2 than in R_1. (e) The potential drop is the same across both resistors. (f) The potential at point a is the same as at point c. (g) The potential at point b is lower than at point c. (h) The potential at point c is lower than at point b.

Figure **Q26.5**



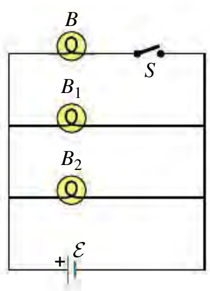
Q26.6 If two resistors R_1 and R_2 ($R_2 > R_1$) are connected in parallel as shown in **Fig. Q26.6**, which of the following must be true? In each case justify your answer. (a) $I_1 = I_2$. (b) $I_3 = I_4$. (c) The current is greater in R_1 than in R_2. (d) The rate of electrical energy consumption is the same for both resistors. (e) The rate of electrical energy consumption is greater in R_2 than in R_1. (f) $V_{cd} = V_{ef} = V_{ab}$. (g) Point c is at higher potential than point d. (h) Point f is at higher potential than point e. (i) Point c is at higher potential than point e.

Figure **Q26.6**



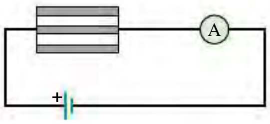
Q26.7 A battery with no internal resistance is connected across identical light bulbs as shown in **Fig. Q26.7**. When you close the switch S, will the brightness of bulbs B_1 and B_2 change? If so, how will it change? Explain.

Figure **Q26.7**



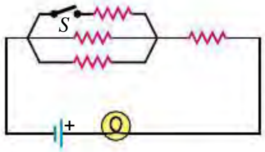
Q26.8 A resistor consists of three identical metal strips connected as shown in **Fig. Q26.8**. If one of the strips is cut out, does the ammeter reading increase, decrease, or stay the same? Why?

Figure **Q26.8**



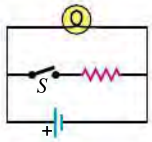
Q26.9 A light bulb is connected in the circuit shown in **Fig. Q26.9**. If we close the switch S, does the bulb's brightness increase, decrease, or remain the same? Explain why.

Figure **Q26.9**



Q26.10 A real battery, having nonnegligible internal resistance, is connected across a light bulb as shown in **Fig. Q26.10**. When the switch S is closed, what happens to the brightness of the bulb? Why?

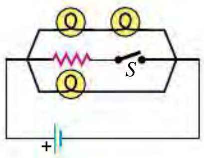
Figure **Q26.10**



Q26.11 If the battery in Discussion Question Q26.10 is ideal with no internal resistance, what will happen to the brightness of the bulb when S is closed? Why?

Q26.12 Consider the circuit shown in **Fig. Q26.12**. What happens to the brightnesses of the bulbs when the switch S is closed if the battery (a) has no internal resistance and (b) has nonnegligible internal resistance? Explain why.

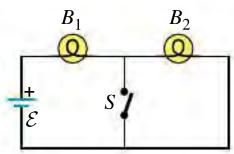
Figure **Q26.12**



Q26.13 Is it possible to connect resistors together in a way that cannot be reduced to some combination of series and parallel combinations? If so, give examples. If not, state why not.

Q26.14 The battery in the circuit shown in **Fig. Q26.14** has no internal resistance. After you close the switch S, will the brightness of bulb B_1 increase, decrease, or stay the same?

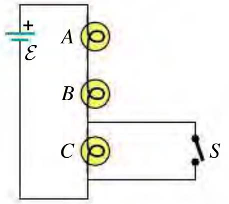
Figure **Q26.14**



Q26.15 In a two-cell flashlight, the batteries are usually connected in series. Why not connect them in parallel? What possible advantage could there be in connecting several identical batteries in parallel?

Q26.16 Identical light bulbs A, B, and C are connected as shown in **Fig. Q26.16**. When the switch S is closed, bulb C goes out. Explain why. What happens to the brightness of bulbs A and B? Explain.

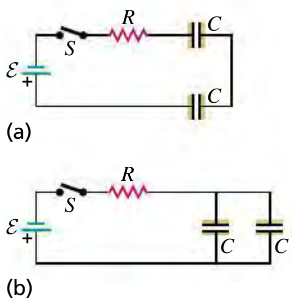
Figure **Q26.16**



Q26.17 The emf of a flashlight battery is roughly constant with time, but its internal resistance increases with age and use. What sort of meter should be used to test the freshness of a battery?

Q26.18 Will the capacitors in the circuits shown in **Fig. Q26.18** charge at the same rate when the switch S is closed? If not, in which circuit will the capacitors charge more rapidly? Explain.

Figure **Q26.18**



Q26.19 Verify that the time constant RC has units of time.

Q26.20 For very large resistances it is easy to construct R-C circuits that have time constants of several seconds or minutes. How might this fact be used to measure very large resistances, those that are too large to measure by more conventional means?

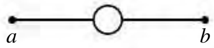
Q26.21 When a capacitor, battery, and resistor are connected in series, does the resistor affect the maximum charge stored on the capacitor? Why or why not? What purpose does the resistor serve?

EXERCISES

Section 26.1 Resistors in Series and Parallel

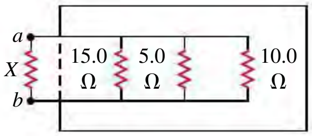
26.1 •• A uniform wire of resistance R is cut into three equal lengths. One of these is formed into a circle and connected between the other two (**Fig. E26.1**). What is the resistance between the opposite ends a and b?

Figure **E26.1**



26.2 •• A machine part has a resistor X protruding from an opening in the side. This resistor is connected to three other resistors, as shown in **Fig. E26.2**. An ohmmeter connected across a and b reads 2.00 Ω. What is the resistance of X?

Figure **E26.2**

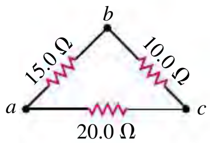


26.3 • A resistor with $R_1 = 25.0\ \Omega$ is connected to a battery that has negligible internal resistance and electrical energy is dissipated by R_1 at a rate of 36.0 W. If a second resistor with $R_2 = 15.0\ \Omega$ is connected in series with R_1, what is the total rate at which electrical energy is dissipated by the two resistors?

26.4 • A 42-Ω resistor and a 20-Ω resistor are connected in parallel, and the combination is connected across a 240-V dc line. (a) What is the resistance of the parallel combination? (b) What is the total current through the parallel combination? (c) What is the current through each resistor?

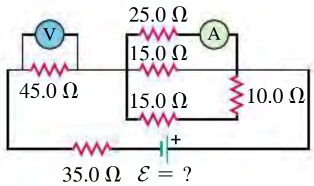
26.5 • A triangular array of resistors is shown in **Fig. E26.5**. What current will this array draw from a 35.0-V battery having negligible internal resistance if we connect it across (a) ab; (b) bc; (c) ac? (d) If the battery has an internal resistance of 3.00 Ω, what current will the array draw if the battery is connected across bc?

Figure **E26.5**



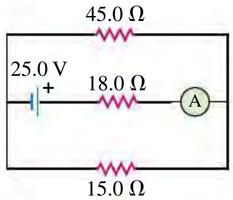
26.6 •• For the circuit shown in **Fig. E26.6** both meters are idealized, the battery has no appreciable internal resistance, and the ammeter reads 1.25 A. (a) What does the voltmeter read? (b) What is the emf $\mathcal{E}$ of the battery?

Figure **E26.6**



26.7 •• For the circuit shown in **Fig. E26.7** find the reading of the idealized ammeter if the battery has an internal resistance of 3.26 Ω.

Figure **E26.7**



26.8 • Three resistors having resistances of 1.60 Ω, 2.40 Ω, and 4.80 Ω are connected in parallel to a 28.0-V battery that has negligible internal resistance. Find (a) the equivalent resistance of the combination; (b) the current in each resistor; (c) the total current through the battery; (d) the voltage across each resistor; (e) the power dissipated in each resistor.

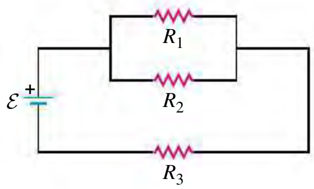
(f) Which resistor dissipates the most power: the one with the greatest resistance or the least resistance? Explain why this should be.

26.9 • Now the three resistors of Exercise 26.8 are connected in series to the same battery. Answer the same questions for this situation.

26.10 •• **Power Rating of a Resistor.** The *power rating* of a resistor is the maximum power the resistor can safely dissipate without too great a rise in temperature and hence damage to the resistor. (a) If the power rating of a 15-kΩ resistor is 5.0 W, what is the maximum allowable potential difference across the terminals of the resistor? (b) A 9.0-kΩ resistor is to be connected across a 120-V potential difference. What power rating is required? (c) A 100.0-Ω and a 150.0-Ω resistor, both rated at 2.00 W, are connected in series across a variable potential difference. What is the greatest this potential difference can be without overheating either resistor, and what is the rate of heat generated in each resistor under these conditions?

26.11 • In **Fig. E26.11**, $R_1 = 3.00\ \Omega$, $R_2 = 6.00\ \Omega$, and $R_3 = 5.00\ \Omega$. The battery has negligible internal resistance. The current I_2 through R_2 is 4.00 A. (a) What are the currents I_1 and I_3? (b) What is the emf of the battery?

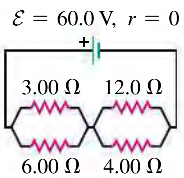
Figure **E26.11**



26.12 •• In Fig. E26.11 the battery has emf 35.0 V and negligible internal resistance. $R_1 = 5.00\ \Omega$. The current through R_1 is 1.50 A, and the current through $R_3 = 4.50$ A. What are the resistances R_2 and R_3?

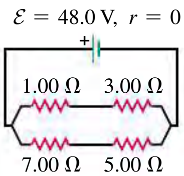
26.13 • Compute the equivalent resistance of the network in **Fig. E26.13**, and find the current in each resistor. The battery has negligible internal resistance.

Figure **E26.13**



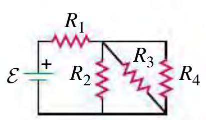
26.14 • Compute the equivalent resistance of the network in **Fig. E26.14**, and find the current in each resistor. The battery has negligible internal resistance.

Figure **E26.14**



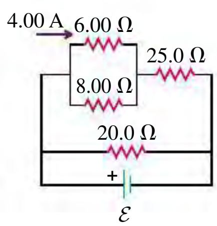
26.15 • In the circuit of **Fig. E26.15**, each resistor represents a light bulb. Let $R_1 = R_2 = R_3 = R_4 = 4.50\ \Omega$ and $\mathcal{E} = 9.00$ V. (a) Find the current in each bulb. (b) Find the power dissipated in each bulb. Which bulb or bulbs glow the brightest? (c) Bulb R_4 is now removed from the circuit, leaving a break in the wire at its position. Now what is the current in each of the remaining bulbs R_1, R_2, and R_3? (d) With bulb R_4 removed, what is the power dissipated in each of the remaining bulbs? (e) Which light bulb(s) glow brighter as a result of removing R_4? Which bulb(s) glow less brightly? Discuss why there are different effects on different bulbs.

Figure **E26.15**



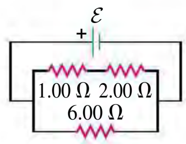
26.16 • Consider the circuit shown in **Fig. E26.16**. The current through the 6.00-Ω resistor is 4.00 A, in the direction shown. What are the currents through the 25.0-Ω and 20.0-Ω resistors?

Figure **E26.16**



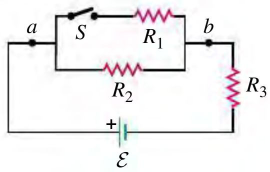
26.17 • In the circuit shown in **Fig. E26.17**, the voltage across the 2.00-Ω resistor is 12.0 V. What are the emf of the battery and the current through the 6.00-Ω resistor?

Figure **E26.17**



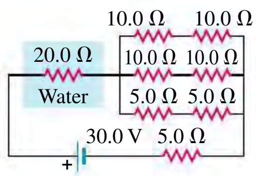
26.18 •• In the circuit shown in **Fig. E26.18**, $\mathcal{E} = 36.0$ V, $R_1 = 4.00\ \Omega$, $R_2 = 6.00\ \Omega$, and $R_3 = 3.00\ \Omega$. (a) What is the potential difference V_{ab} between points a and b when the switch S is open and when S is closed? (b) For each resistor, calculate the current through the resistor with S open and with S closed. For each resistor, does the current increase or decrease when S is closed?

Figure **E26.18**



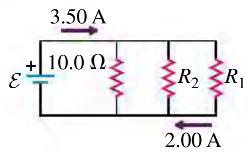
26.19 •• **CP** In the circuit in **Fig. E26.19**, a 20.0-Ω resistor is inside 100 g of pure water that is surrounded by insulating styrofoam. If the water is initially at 10.0°C, how long will it take for its temperature to rise to 58.0°C?

Figure **E26.19**



26.20 • In the circuit shown in **Fig. E26.20**, the rate at which R_1 is dissipating electrical energy is 15.0 W. (a) Find R_1 and R_2. (b) What is the emf of the battery? (c) Find the current through both R_2 and the 10.0-Ω resistor. (d) Calculate the total electrical power consumption in all the resistors and the electrical power delivered by the battery. Show that your results are consistent with conservation of energy.

Figure **E26.20**



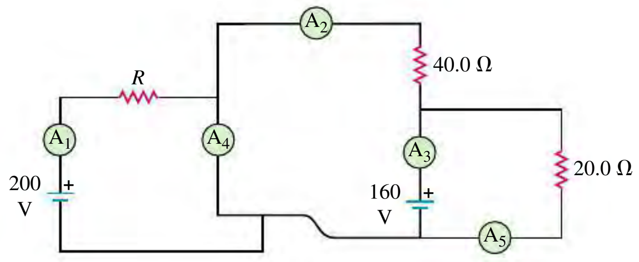
26.21 • **Light Bulbs in Series and in Parallel.** Two light bulbs have constant resistances of 400 Ω and 800 Ω. If the two light bulbs are connected in series across a 120-V line, find (a) the current through each bulb; (b) the power dissipated in each bulb; (c) the total power dissipated in both bulbs. The two light bulbs are now connected in parallel across the 120-V line. Find (d) the current through each bulb; (e) the power dissipated in each bulb; (f) the total power dissipated in both bulbs. (g) In each situation, which of the two bulbs glows the brightest? (h) In which situation is there a greater total light output from both bulbs combined?

26.22 • **Light Bulbs in Series.** A 60-W, 120-V light bulb and a 200-W, 120-V light bulb are connected in series across a 240-V line. Assume that the resistance of each bulb does not vary with current. (*Note:* This description of a light bulb gives the power it dissipates when connected to the stated potential difference; that is, a 25-W, 120-V light bulb dissipates 25 W when connected to a 120-V line.) (a) Find the current through the bulbs. (b) Find the power dissipated in each bulb. (c) One bulb burns out very quickly. Which one? Why?

Section 26.2 Kirchhoff's Rules

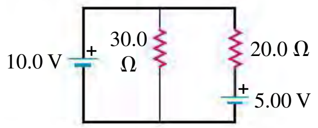
26.23 •• In the circuit shown in **Fig. E26.23**, ammeter A_1 reads 10.0 A and the batteries have no appreciable internal resistance. (a) What is the resistance of R? (b) Find the readings in the other ammeters.

Figure **E26.23**



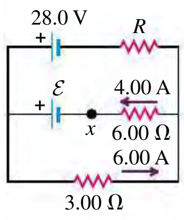
26.24 •• The batteries shown in the circuit in **Fig. E26.24** have negligibly small internal resistances. Find the current through (a) the 30.0-Ω resistor; (b) the 20.0-Ω resistor; (c) the 10.0-V battery.

Figure **E26.24**



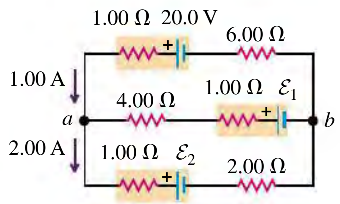
26.25 • In the circuit shown in **Fig. E26.25** find (a) the current in resistor R; (b) the resistance R; (c) the unknown emf $\mathcal{E}$. (d) If the circuit is broken at point x, what is the current in resistor R?

Figure **E26.25**



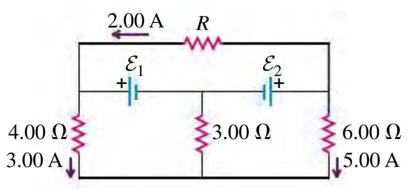
26.26 • Find the emfs $\mathcal{E}_1$ and $\mathcal{E}_2$ in the circuit of **Fig. E26.26**, and find the potential difference of point b relative to point a.

Figure **E26.26**



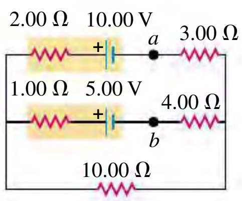
26.27 • In the circuit shown in **Fig. E26.27**, find (a) the current in the 3.00-Ω resistor; (b) the unknown emfs $\mathcal{E}_1$ and $\mathcal{E}_2$; (c) the resistance R. Note that three currents are given.

Figure **E26.27**



26.28 •• In the circuit shown in **Fig. E26.28**, find (a) the current in each branch and (b) the potential difference V_{ab} of point a relative to point b.

Figure **E26.28**

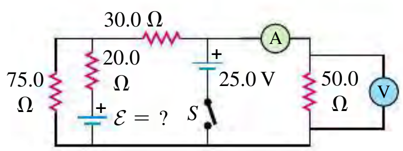


26.29 • The 10.00-V battery in Fig. E26.28 is removed from the circuit and reinserted with the opposite polarity, so that its positive terminal is now next to point a. The rest of the circuit is as shown in the figure. Find (a) the current in each branch and (b) the potential difference V_{ab} of point a relative to point b.

26.30 • The 5.00-V battery in Fig. E26.28 is removed from the circuit and replaced by a 15.00-V battery, with its negative terminal next to point b. The rest of the circuit is as shown in the figure. Find (a) the current in each branch and (b) the potential difference V_{ab} of point a relative to point b.

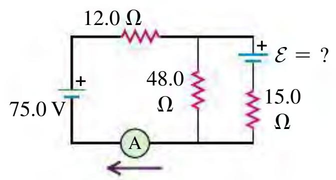
26.31 •• In the circuit shown in **Fig. E26.31** the batteries have negligible internal resistance and the meters are both idealized. With the switch S open, the voltmeter reads 15.0 V. (a) Find the emf $\mathcal{E}$ of the battery. (b) What will the ammeter read when the switch is closed?

Figure **E26.31**



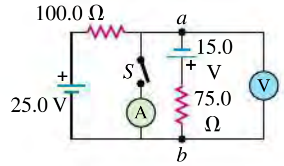
26.32 •• In the circuit shown in **Fig. E26.32** both batteries have insignificant internal resistance and the idealized ammeter reads 1.50 A in the direction shown. Find the emf $\mathcal{E}$ of the battery. Is the polarity shown correct?

Figure **E26.32**



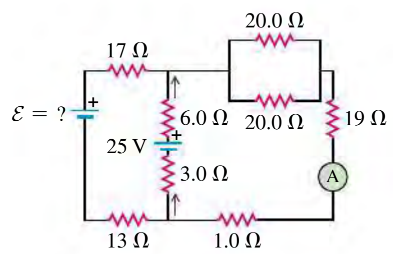
26.33 • In the circuit shown in **Fig. E26.33** all meters are idealized and the batteries have no appreciable internal resistance. (a) Find the reading of the voltmeter with the switch S open. Which point is at a higher potential: a or b? (b) With S closed, find the reading of the voltmeter and the ammeter. Which way (up or down) does the current flow through the switch?

Figure **E26.33**



26.34 •• In the circuit shown in **Fig. E26.34**, the 6.0-Ω resistor is consuming energy at a rate of 24 J/s when the current through it flows as shown. (a) Find the current through the ammeter A. (b) What are the polarity and emf $\mathcal{E}$ of the unknown battery, assuming it has negligible internal resistance?

Figure **E26.34**

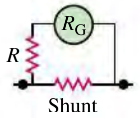


Section 26.3 Electrical Measuring Instruments

26.35 • The resistance of a galvanometer coil is 25.0 Ω, and the current required for full-scale deflection is 500 μA. (a) Show in a diagram how to convert the galvanometer to an ammeter reading 20.0 mA full scale, and compute the shunt resistance. (b) Show how to convert the galvanometer to a voltmeter reading 500 mV full scale, and compute the series resistance.

26.36 • The resistance of the coil of a pivoted-coil galvanometer is 9.36 Ω, and a current of 0.0224 A causes it to deflect full scale. We want to convert this galvanometer to an ammeter reading 20.0 A full scale. The only shunt available has a resistance of 0.0250 Ω. What resistance R must be connected in series with the coil (**Fig. E26.36**)?

Figure **E26.36**



26.37 • A circuit consists of a series combination of 6.00-kΩ and 5.00-kΩ resistors connected across a 50.0-V battery having negligible internal resistance. You want to measure the true potential difference (that is, the potential difference without the meter present) across the 5.00-kΩ resistor using a voltmeter having an internal resistance of 10.0 kΩ. (a) What potential difference does the voltmeter measure across the 5.00-kΩ resistor? (b) What is the *true* potential difference across this resistor when the meter is not present? (c) By what percentage is the voltmeter reading in error from the true potential difference?

26.38 • A galvanometer having a resistance of 25.0 Ω has a 1.00-Ω shunt resistance installed to convert it to an ammeter. It is then used to measure the current in a circuit consisting of a 15.0-Ω resistor connected across the terminals of a 25.0-V battery having no appreciable internal resistance. (a) What current does the ammeter measure? (b) What should be the *true* current in the circuit (that is, the current without the ammeter present)? (c) By what percentage is the ammeter reading in error from the *true* current?

Section 26.4 *R-C* Circuits

26.39 • A capacitor is charged to a potential of 12.0 V and is then connected to a voltmeter having an internal resistance of 3.40 MΩ. After a time of 4.00 s the voltmeter reads 3.0 V. What are (a) the capacitance and (b) the time constant of the circuit?

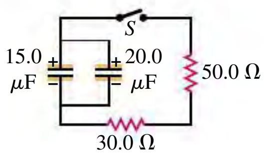
26.40 •• You connect a battery, resistor, and capacitor as in Fig. 26.20a, where $\mathcal{E} = 36.0$ V, $C = 5.00\ \mu$F, and $R = 120\ \Omega$. The switch S is closed at $t = 0$. (a) When the voltage across the capacitor is 8.00 V, what is the magnitude of the current in the circuit? (b) At what time t after the switch is closed is the voltage across the capacitor 8.00 V? (c) When the voltage across the capacitor is 8.00 V, at what rate is energy being stored in the capacitor?

26.41 • A 4.60-μF capacitor that is initially uncharged is connected in series with a 7.50-kΩ resistor and an emf source with $\mathcal{E} = 245$ V and negligible internal resistance. Just after the circuit is completed, what are (a) the voltage drop across the capacitor; (b) the voltage drop across the resistor; (c) the charge on the capacitor; (d) the current through the resistor? (e) A long time after the circuit is completed (after many time constants) what are the values of the quantities in parts (a)–(d)?

26.42 •• You connect a battery, resistor, and capacitor as in Fig. 26.20a, where $R = 12.0\ \Omega$ and $C = 5.00 \times 10^{-6}$ F. The switch S is closed at $t = 0$. When the current in the circuit has magnitude 3.00 A, the charge on the capacitor is 40.0×10^{-6} C. (a) What is the emf of the battery? (b) At what time t after the switch is closed is the charge on the capacitor equal to 40.0×10^{-6} C? (c) When the current has magnitude 3.00 A, at what rate is energy being (i) stored in the capacitor, (ii) supplied by the battery?

26.43 •• CP In the circuit shown in **Fig. E26.43** both capacitors are initially charged to 45.0 V. (a) How long after closing the switch S will the potential across each capacitor be reduced to 10.0 V, and (b) what will be the current at that time?

Figure **E26.43**



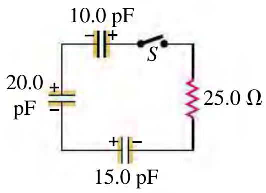
26.44 • A 12.4-μF capacitor is connected through a 0.895-MΩ resistor to a constant potential difference of 60.0 V. (a) Compute the charge on the capacitor at the following times after the connections are made: 0, 5.0 s, 10.0 s, 20.0 s, and 100.0 s. (b) Compute the charging currents at the same instants. (c) Graph the results of parts (a) and (b) for t between 0 and 20 s.

26.45 • An emf source with $\mathcal{E} = 120$ V, a resistor with $R = 80.0\ \Omega$, and a capacitor with $C = 4.00\ \mu$F are connected in series. As the capacitor charges, when the current in the resistor is 0.900 A, what is the magnitude of the charge on each plate of the capacitor?

26.46 • A resistor and a capacitor are connected in series to an emf source. The time constant for the circuit is 0.780 s. (a) A second capacitor, identical to the first, is added in series. What is the time constant for this new circuit? (b) In the original circuit a second capacitor, identical to the first, is connected in parallel with the first capacitor. What is the time constant for this new circuit?

26.47 •• CP In the circuit shown in **Fig. E26.47** each capacitor initially has a charge of magnitude 3.50 nC on its plates. After the switch S is closed, what will be the current in the circuit at the instant that the capacitors have lost 80.0% of their initial stored energy?

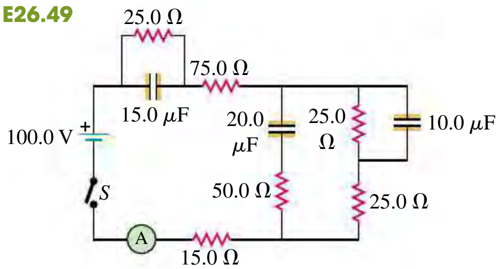
Figure **E26.47**



26.48 • A 1.50-μF capacitor is charging through a 12.0-Ω resistor using a 10.0-V battery. What will be the current when the capacitor has acquired $\frac{1}{4}$ of its maximum charge? Will it be $\frac{1}{4}$ of the maximum current?

26.49 • In the circuit in **Fig. E26.49** the capacitors are initially uncharged, the battery has no internal resistance, and the ammeter is idealized. Find the ammeter reading (a) just after the switch S is closed and (b) after S has been closed for a very long time.

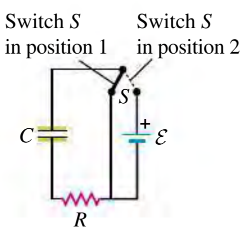
Figure **E26.49**



26.50 • A 12.0-μF capacitor is charged to a potential of 50.0 V and then discharged through a 225-Ω resistor. How long does it take the capacitor to lose (a) half of its charge and (b) half of its stored energy?

26.51 • In the circuit shown in **Fig. E26.51**, $C = 5.90\ \mu$F, $\mathcal{E} = 28.0$ V, and the emf has negligible resistance. Initially the capacitor is uncharged and the switch S is in position 1. The switch is then moved to position 2, so that the capacitor begins to charge. (a) What will be the charge on the capacitor a long time after S is moved to position 2? (b) After S has been in position 2 for 3.00 ms, the charge on the capacitor is measured to be 110 μC. What is the value of the resistance R? (c) How long after S is moved to position 2 will the charge on the capacitor be equal to 99.0% of the final value found in part (a)?

Figure **E26.51**



Section 26.5 Power Distribution Systems

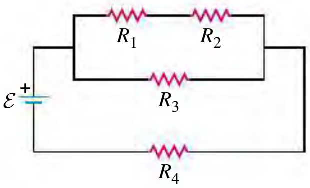
26.52 • The heating element of an electric dryer is rated at 4.1 kW when connected to a 240-V line. (a) What is the current in the heating element? Is 12-gauge wire large enough to supply this current? (b) What is the resistance of the dryer's heating element at its operating temperature? (c) At 11 cents per kWh, how much does it cost per hour to operate the dryer?

26.53 • A 1500-W electric heater is plugged into the outlet of a 120-V circuit that has a 20-A circuit breaker. You plug an electric hair dryer into the same outlet. The hair dryer has power settings of 600 W, 900 W, 1200 W, and 1500 W. You start with the hair dryer on the 600-W setting and increase the power setting until the circuit breaker trips. What power setting caused the breaker to trip?

PROBLEMS

26.54 •• In **Fig. P26.54**, the battery has negligible internal resistance and $\mathcal{E} = 48.0$ V. $R_1 = R_2 = 4.00\ \Omega$ and $R_4 = 3.00\ \Omega$. What must the resistance R_3 be for the resistor network to dissipate electrical energy at a rate of 295 W?

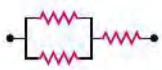
Figure **P26.54**



26.55 • The two identical light bulbs in Example 26.2 (Section 26.1) are connected in parallel to a different source, one with $\mathcal{E} = 8.0$ V and internal resistance 0.8 Ω. Each light bulb has a resistance $R = 2.0\ \Omega$ (assumed independent of the current through the bulb). (a) Find the current through each bulb, the potential difference across each bulb, and the power delivered to each bulb. (b) Suppose one of the bulbs burns out, so that its filament breaks and current no longer flows through it. Find the power delivered to the remaining bulb. Does the remaining bulb glow more or less brightly after the other bulb burns out than before?

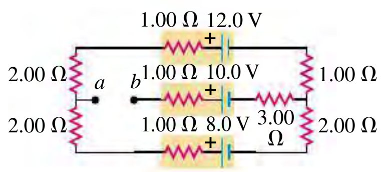
26.56 •• Each of the three resistors in **Fig. P26.56** has a resistance of 2.4 Ω and can dissipate a maximum of 48 W without becoming excessively heated. What is the maximum power the circuit can dissipate?

Figure **P26.56**



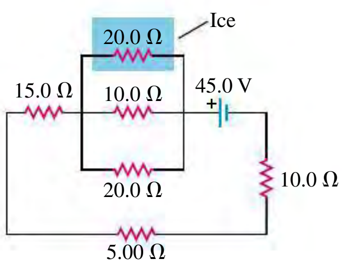
26.57 •• (a) Find the potential of point a with respect to point b in **Fig. P26.57**. (b) If points a and b are connected by a wire with negligible resistance, find the current in the 12.0-V battery.

Figure **P26.57**



26.58 •• CP For the circuit shown in **Fig. P26.58** a 20.0-Ω resistor is embedded in a large block of ice at 0.00°C, and the battery has negligible internal resistance. At what rate (in g/s) is this circuit melting the ice? (The latent heat of fusion for ice is 3.34×10^5 J/kg.)

Figure **P26.58**



26.59 • Calculate the three currents I_1, I_2, and I_3 indicated in the circuit diagram shown in **Fig. P26.59**.

Figure **P26.59**

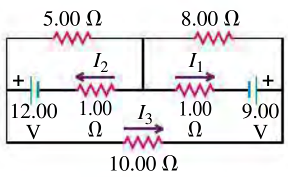
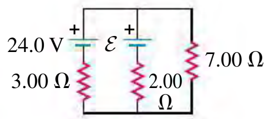


Figure **P26.60**



26.60 ••• What must the emf $\mathcal{E}$ in **Fig. P26.60** be in order for the current through the 7.00-Ω resistor to be 1.80 A? Each emf source has negligible internal resistance.

26.61 • Find the current through each of the three resistors of the circuit shown in **Fig. P26.61**. The emf sources have negligible internal resistance.

Figure **P26.61**

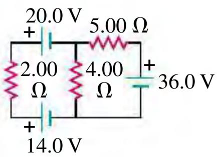
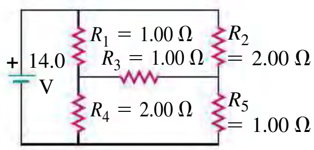


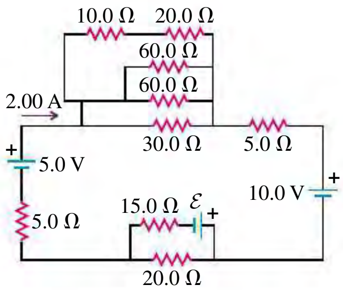
Figure **P26.62**



26.62 • (a) Find the current through the battery and each resistor in the circuit shown in **Fig. P26.62**. (b) What is the equivalent resistance of the resistor network?

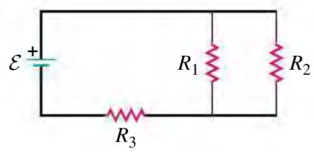
26.63 •• Consider the circuit shown in **Fig. P26.63**. (a) What must the emf $\mathcal{E}$ of the battery be in order for a current of 2.00 A to flow through the 5.00-V battery as shown? Is the polarity of the battery correct as shown? (b) How long does it take for 60.0 J of thermal energy to be produced in the 10.0-Ω resistor?

Figure **P26.63**



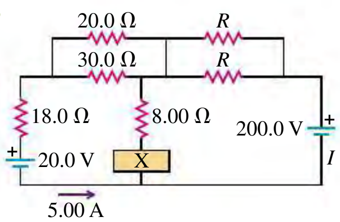
26.64 •• In the circuit shown in **Fig. P26.64**, $\mathcal{E} = 24.0$ V, $R_1 = 6.00\ \Omega$, $R_3 = 12.0\ \Omega$, and R_2 can vary between 3.00 Ω and 24.0 Ω. For what value of R_2 is the power dissipated by heating element R_1 the greatest? Calculate the magnitude of the greatest power.

Figure **P26.64**



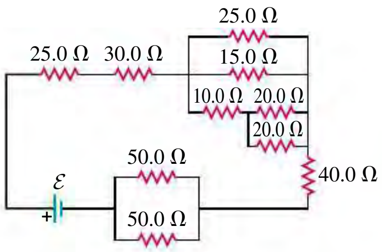
26.65 • In the circuit shown in **Fig. P26.65**, the current in the 20.0-V battery is 5.00 A in the direction shown and the voltage across the 8.00-Ω resistor is 16.0 V, with the lower end of the resistor at higher potential. Find (a) the emf (including its polarity) of the battery X; (b) the current I through the 200.0-V battery (including its direction); (c) the resistance R.

Figure **P26.65**



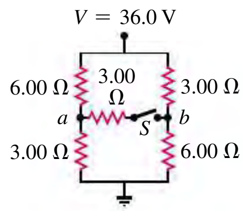
26.66 ••• In the circuit shown in **Fig. P26.66** all the resistors are rated at a maximum power of 2.00 W. What is the maximum emf $\mathcal{E}$ that the battery can have without burning up any of the resistors?

Figure **P26.66**



26.67 • **Figure P26.67** employs a convention often used in circuit diagrams. The battery (or other power supply) is not shown explicitly. It is understood that the point at the top, labeled "36.0 V," is connected to the positive terminal of a 36.0-V battery having negligible internal resistance, and that the *ground* symbol at the bottom is connected to the negative terminal of the battery. The circuit is completed through the battery, even though it is not shown. (a) What is the potential difference V_{ab}, the potential of point a relative to point b, when the switch S is open? (b) What is the current through S when it is closed? (c) What is the equivalent resistance when S is closed?

Figure **P26.67**

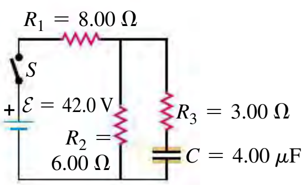


26.68 •• Three identical resistors are connected in series. When a certain potential difference is applied across the combination, the total power dissipated is 45.0 W. What power would be dissipated if the three resistors were connected in parallel across the same potential difference?

26.69 • A resistor R_1 consumes electrical power P_1 when connected to an emf $\mathcal{E}$. When resistor R_2 is connected to the same emf, it consumes electrical power P_2. In terms of P_1 and P_2, what is the total electrical power consumed when they are both connected to this emf source (a) in parallel and (b) in series?

26.70 • The capacitor in **Fig. P26.70** is initially uncharged. The switch S is closed at $t = 0$. (a) Immediately after the switch is closed, what is the current through each resistor? (b) What is the final charge on the capacitor?

Figure **P26.70**

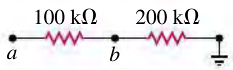


26.71 •• A 2.00-μF capacitor that is initially uncharged is connected in series with a 6.00-kΩ resistor and an emf source with $\mathcal{E} = 90.0$ V and negligible internal resistance. The circuit is completed at $t = 0$. (a) Just after the circuit is completed, what is the rate at which electrical energy is being dissipated in the resistor? (b) At what value of t is the rate at which electrical energy is being dissipated in the resistor equal to the rate at which electrical energy is being stored in the capacitor? (c) At the time calculated in part (b), what is the rate at which electrical energy is being dissipated in the resistor?

26.72 •• A 6.00-μF capacitor that is initially uncharged is connected in series with a 5.00-Ω resistor and an emf source with $\mathcal{E} = 50.0$ V and negligible internal resistance. At the instant when the resistor is dissipating electrical energy at a rate of 300 W, how much energy has been stored in the capacitor?

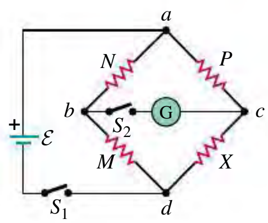
26.73 • Point a in **Fig. P26.73** is maintained at a constant potential of 400 V above ground. (See Problem 26.67.) (a) What is the reading of a voltmeter with the proper range and with resistance 5.00×10^4 Ω when connected between point b and ground? (b) What is the reading of a voltmeter with resistance 5.00×10^6 Ω? (c) What is the reading of a voltmeter with infinite resistance?

Figure **P26.73**



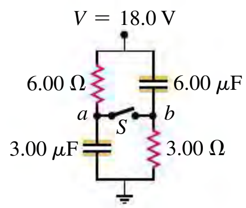
26.74 •• **The Wheatstone Bridge.** The circuit shown in **Fig. P26.74**, called a *Wheatstone bridge,* is used to determine the value of an unknown resistor X by comparison with three resistors M, N, and P whose resistances can be varied. For each setting, the resistance of each resistor is precisely known. With switches S_1 and S_2 closed, these resistors are varied until the current in the galvanometer G is zero; the bridge is then said to be *balanced.* (a) Show that under this condition the unknown resistance is given by $X = MP/N$. (This method permits very high precision in comparing resistors.) (b) If galvanometer G shows zero deflection when $M = 850.0$ Ω, $N = 15.00$ Ω, and $P = 33.48$ Ω, what is the unknown resistance X?

Figure **P26.74**



26.75 • (See Problem 26.67.) (a) What is the potential of point a with respect to point b in **Fig. P26.75** when the switch S is open? (b) Which point, a or b, is at the higher potential? (c) What is the final potential of point b with respect to ground when S is closed? (d) How much does the charge on each capacitor change when S is closed?

Figure **P26.75**



26.76 • A 2.36-μF capacitor that is initially uncharged is connected in series with a 5.86-Ω resistor and an emf source with $\mathcal{E} = 120$ V and negligible internal resistance. (a) Just after the connection is made, what are (i) the rate at which electrical energy is being dissipated in the resistor; (ii) the rate at which the electrical energy stored in the capacitor is increasing; (iii) the electrical power output of the source? How do the answers to parts (i), (ii), and (iii) compare? (b) Answer the same questions as in part (a) at a long time after the connection is made. (c) Answer the same questions as in part (a) at the instant when the charge on the capacitor is one-half its final value.

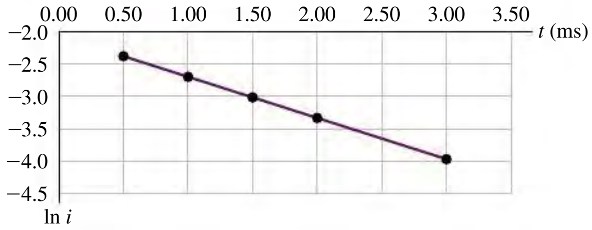
26.77 • A 224-Ω resistor and a 589-Ω resistor are connected in series across a 90.0-V line. (a) What is the voltage across each resistor? (b) A voltmeter connected across the 224-Ω resistor reads 23.8 V. Find the voltmeter resistance. (c) Find the reading of the same voltmeter if it is connected across the 589-Ω resistor. (d) The readings on this voltmeter are lower than the "true" voltages (that is, without the voltmeter present). Would it be possible to design a voltmeter that gave readings *higher* than the "true" voltages? Explain.

26.78 • A resistor with $R = 850$ Ω is connected to the plates of a charged capacitor with capacitance $C = 4.62$ μF. Just before the connection is made, the charge on the capacitor is 6.90 mC. (a) What is the energy initially stored in the capacitor? (b) What is the electrical power dissipated in the resistor just after the connection is made? (c) What is the electrical power dissipated in the resistor at the instant when the energy stored in the capacitor has decreased to half the value calculated in part (a)?

26.79 • A capacitor that is initially uncharged is connected in series with a resistor and an emf source with $\mathcal{E} = 110$ V and negligible internal resistance. Just after the circuit is completed, the current through the resistor is 6.5×10^{-5} A. The time constant for the circuit is 5.2 s. What are the resistance of the resistor and the capacitance of the capacitor?

26.80 •• DATA You set up the circuit shown in Fig. 26.22a, where $R = 196\ \Omega$. You close the switch at time $t = 0$ and measure the magnitude i of the current in the resistor R as a function of time t since the switch was closed. Your results are shown in **Fig. P26.80**, where you have chosen to plot $\ln i$ as a function of t. (a) Explain why your data points lie close to a straight line. (b) Use the graph in Fig. P26.80 to calculate the capacitance C and the initial charge Q_0 on the capacitor. (c) When $i = 0.0500$ A, what is the charge on the capacitor? (d) When $q = 0.500 \times 10^{-4}$ C, what is the current in the resistor?

Figure **P26.80**



26.81 •• DATA You set up the circuit shown in Fig. 26.20, where $C = 5.00 \times 10^{-6}$ F. At time $t = 0$, you close the switch and then measure the charge q on the capacitor as a function of the current i in the resistor. Your results are given in the table:

i (mA)	56.0	48.0	40.0	32.0	24.0
q (μC)	10.1	19.8	30.2	40.0	49.9

(a) Graph q as a function of i. Explain why the data points, when plotted this way, fall close to a straight line. Find the slope and y-intercept of the straight line that gives the best fit to the data. (b) Use your results from part (a) to calculate the resistance R of the resistor and the emf $\mathcal{E}$ of the battery. (c) At what time t after the switch is closed is the voltage across the capacitor equal to 10.0 V? (d) When the voltage across the capacitor is 4.00 V, what is the voltage across the resistor?

26.82 •• DATA The electronics supply company where you work has two different resistors, R_1 and R_2, in its inventory, and you must measure the values of their resistances. Unfortunately, stock is low, and all you have are R_1 and R_2 in parallel and in series—and you can't separate these two resistor combinations. You separately connect each resistor network to a battery with emf 48.0 V and negligible internal resistance and measure the power P supplied by the battery in both cases. For the series combination, $P = 48.0$ W; for the parallel combination, $P = 256$ W. You are told that $R_1 > R_2$. (a) Calculate R_1 and R_2. (b) For the series combination, which resistor consumes more power, or do they consume the same power? Explain. (c) For the parallel combination, which resistor consumes more power, or do they consume the same power?

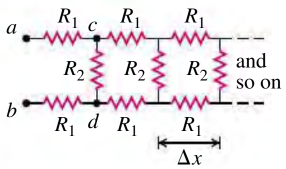
CHALLENGE PROBLEMS

26.83 ••• **An Infinite Network.** As shown in **Fig. P26.83**, a network of resistors of resistances R_1 and R_2 extends to infinity toward the right. Prove that the total resistance R_T of the infinite network is equal to

$$R_T = R_1 + \sqrt{R_1^2 + 2R_1R_2}$$

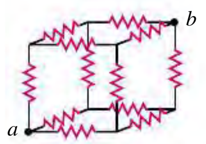
(*Hint:* Since the network is infinite, the resistance of the network to the right of points c and d is also equal to R_T.)

Figure **P26.83**



26.84 ••• Suppose a resistor R lies along each edge of a cube (12 resistors in all) with connections at the corners. Find the equivalent resistance between two diagonally opposite corners of the cube (points a and b in **Fig. P26.84**).

Figure **P26.84**



26.85 ••• BIO **Attenuator Chains and Axons.** The infinite network of resistors shown in Fig. P26.83 is known as an *attenuator chain,* since this chain of resistors causes the potential difference between the upper and lower wires to decrease, or attenuate, along the length of the chain. (a) Show that if the potential difference between the points a and b in Fig. 26.83 is V_{ab}, then the potential difference between points c and d is $V_{cd} = V_{ab}/(1 + \beta)$, where $\beta = 2R_1(R_T + R_2)/R_TR_2$ and R_T, the total resistance of the network, is given in Challenge Problem 26.83. (See the hint given in that problem.) (b) If the potential difference between terminals a and b at the left end of the infinite network is V_0, show that the potential difference between the upper and lower wires n segments from the left end is $V_n = V_0/(1 + \beta)^n$. If $R_1 = R_2$, how many segments are needed to decrease the potential difference V_n to less than 1.0% of V_0? (c) An infinite attenuator chain provides a model of the propagation of a voltage pulse along a nerve fiber, or axon. Each segment of the network in Fig. P26.83 represents a short segment of the axon of length Δx. The resistors R_1 represent the resistance of the fluid inside and outside the membrane wall of the axon. The resistance of the membrane to current flowing through the wall is represented by R_2. For an axon segment of length $\Delta x = 1.0\ \mu$m, $R_1 = 6.4 \times 10^3\ \Omega$ and $R_2 = 8.0 \times 10^8\ \Omega$ (the membrane wall is a good insulator). Calculate the total resistance R_T and β for an infinitely long axon. (This is a good approximation, since the length of an axon is much greater than its width; the largest axons in the human nervous system are longer than 1 m but only about 10^{-7} m in radius.) (d) By what fraction does the potential difference between the inside and outside of the axon decrease over a distance of 2.0 mm? (e) The attenuation of the potential difference calculated in part (d) shows that the axon cannot simply be a passive, current-carrying electrical cable; the potential difference must periodically be reinforced along the axon's length. This reinforcement mechanism is slow, so a signal propagates along the axon at only about 30 m/s. In situations where faster response is required, axons are covered with a segmented sheath of fatty myelin. The segments are about 2 mm long, separated by gaps called the *nodes of Ranvier*. The myelin increases the resistance of a 1.0-μm-long segment of the membrane to $R_2 = 3.3 \times 10^{12}\ \Omega$. For such a myelinated axon, by what fraction does the potential difference between the inside and outside of the axon decrease over the distance from one node of Ranvier to the next? This smaller attenuation means the propagation speed is increased.

PASSAGE PROBLEMS

BIO NERVE CELLS AND *R-C* CIRCUITS. The portion of a nerve cell that conducts signals is called an *axon*. Many of the electrical properties of axons are governed by ion channels, which are protein molecules that span the axon's cell membrane. When open, each ion channel has a pore that is filled with fluid of low resistivity and connects the interior of the cell electrically to the medium outside the cell. In contrast, the lipid-rich cell membrane in which ion channels reside has very high resistivity.

26.86 Assume that a typical open ion channel spanning an axon's membrane has a resistance of $1 \times 10^{11}\ \Omega$. We can model this ion channel, with its pore, as a 12-nm-long cylinder of radius 0.3 nm. What is the resistivity of the fluid in the pore? (a) 10 $\Omega \cdot$ m; (b) 6 $\Omega \cdot$ m; (c) 2 $\Omega \cdot$ m; (d) 1 $\Omega \cdot$ m.

26.87 In a simple model of an axon conducting a nerve signal, ions move across the cell membrane through open ion channels, which act as purely resistive elements. If a typical current density (current per unit cross-sectional area) in the cell membrane is 5 mA/cm^2 when the voltage across the membrane (the *action potential*) is 50 mV, what is the number density of open ion channels in the membrane? (a) 1/cm^2; (b) 10/cm^2; (c) 10/mm^2; (d) 100/μm^2.

26.88 Cell membranes across a wide variety of organisms have a capacitance per unit area of 1 μF/cm^2. For the electrical signal in a nerve to propagate down the axon, the charge on the membrane "capacitor" must change. What time constant is required when the ion channels are open? (a) 1 μs; (b) 10 μs; (c) 100 μs; (d) 1 ms.

Answers

Chapter Opening Question ?

(ii) The potential difference V is the same across resistors connected in parallel. However, there is a different current I through each resistor if the resistances R are different: $I = V/R$.

Test Your Understanding Questions

26.1 (a), (c), (d), (b) Here's why: The three resistors in Fig. 26.1a are in series, so $R_{eq} = R + R + R = 3R$. In Fig. 26.1b the three resistors are in parallel, so $1/R_{eq} = 1/R + 1/R + 1/R = 3/R$ and $R_{eq} = R/3$. In Fig. 26.1c the second and third resistors are in parallel, so their equivalent resistance R_{23} is given by $1/R_{23} = 1/R + 1/R = 2/R$; hence $R_{23} = R/2$. This combination is in series with the first resistor, so the three resistors together have equivalent resistance $R_{eq} = R + R/2 = 3R/2$. In Fig. 26.1d the second and third resistors are in series, so their equivalent resistance is $R_{23} = R + R = 2R$. This combination is in parallel with the first resistor, so the equivalent resistance of the three-resistor combination is given by $1/R_{eq} = 1/R + 1/2R = 3/2R$. Hence $R_{eq} = 2R/3$.

26.2 loop *cbdac*, no Equation (2) minus Eq. (1) gives $-I_2(1\ \Omega) - (I_2 + I_3)(2\ \Omega) + (I_1 - I_3)(1\ \Omega) + I_1(1\ \Omega) = 0$. We can obtain this equation by applying the loop rule around the path from c to b to d to a to c in Fig. 26.12. This isn't an independent equation, so it would not have helped with the solution of Example 26.6.

26.3 (a) (ii), (b) (iii) An ammeter must always be placed in series with the circuit element of interest, and a voltmeter must always be placed in parallel. Ideally the ammeter would have zero resistance and the voltmeter would have infinite resistance so that their presence would have no effect on either the resistor current or the voltage. Neither of these idealizations is possible, but the ammeter resistance should be much less than 2 Ω and the voltmeter resistance should be much greater than 2 Ω.

26.4 (ii) After one time constant, $t = RC$ and the initial charge Q_0 has decreased to $Q_0e^{-t/RC} = Q_0e^{-RC/RC} = Q_0e^{-1} = Q_0/e$. Hence the stored energy has decreased from $Q_0^2/2C$ to $(Q_0/e)^2/2C = Q_0^2/2Ce^2$, a fraction $1/e^2 = 0.135$ of its initial value. This result doesn't depend on the initial value of the energy.

26.5 no This is a very dangerous thing to do. The circuit breaker will allow currents up to 40 A, double the rated value of the wiring. The amount of power $P = I^2R$ dissipated in a section of wire can therefore be up to four times the rated value, so the wires could get very warm and start a fire. (This assumes the resistance R remains unchanged. In fact, R increases with temperature, so the dissipated power can be even greater, and more dangerous, than we have estimated.)

Bridging Problem

(a) 9.39 J **(b)** 2.02×10^4 W **(c)** 4.65×10^{-4} s
(d) 7.43×10^3 W

? The needle of a magnetic compass points north. This alignment is due to (i) a magnetic force on the needle; (ii) a magnetic torque on the needle; (iii) the magnetic field that the needle itself produces; (iv) both (i) and (ii); (v) all of (i), (ii), and (iii).

27 MAGNETIC FIELD AND MAGNETIC FORCES

LEARNING GOALS

Looking forward at ...

27.1 The properties of magnets, and how magnets interact with each other.

27.2 The nature of the force that a moving charged particle experiences in a magnetic field.

27.3 How magnetic field lines are different from electric field lines.

27.4 How to analyze the motion of a charged particle in a magnetic field.

27.5 Some practical applications of magnetic fields in chemistry and physics.

27.6 How to analyze magnetic forces on current-carrying conductors.

27.7 How current loops behave when placed in a magnetic field.

27.8 How direct-current motors work.

27.9 How magnetic forces give rise to the Hall effect.

Looking back at ...

1.10 Vector product of two vectors.

3.4, 5.4 Uniform circular motion.

10.1 Torque.

21.6, 21.7 Electric field lines and electric dipole moment.

22.2, 22.3 Electric flux and Gauss's law.

25.1 Electric current.

26.3 Galvanometers.

Everybody uses magnetic forces. They are at the heart of electric motors, microwave ovens, loudspeakers, computer printers, and disk drives. The most familiar examples of magnetism are permanent magnets, which attract unmagnetized iron objects and can also attract or repel other magnets. A compass needle aligning itself with the earth's magnetism is an example of this interaction. But the *fundamental* nature of magnetism is the interaction of moving electric charges. Unlike electric forces, which act on electric charges whether they are moving or not, magnetic forces act only on *moving* charges.

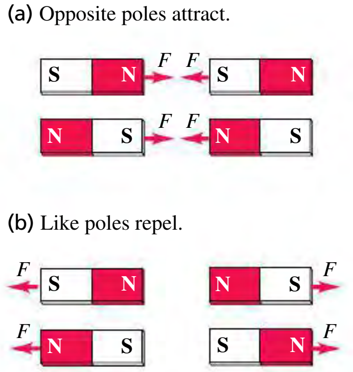
We saw in Chapter 21 that the electric force arises in two stages: (1) a charge produces an electric field in the space around it, and (2) a second charge responds to this field. Magnetic forces also arise in two stages. First, a *moving* charge or a collection of moving charges (that is, an electric current) produces a *magnetic* field. Next, a second current or moving charge responds to this magnetic field, and so experiences a magnetic force.

In this chapter we study the second stage in the magnetic interaction—that is, how moving charges and currents *respond* to magnetic fields. In particular, we will see how to calculate magnetic forces and torques, and we will discover why magnets can pick up iron objects like paper clips. In Chapter 28 we will complete our picture of the magnetic interaction by examining how moving charges and currents *produce* magnetic fields.

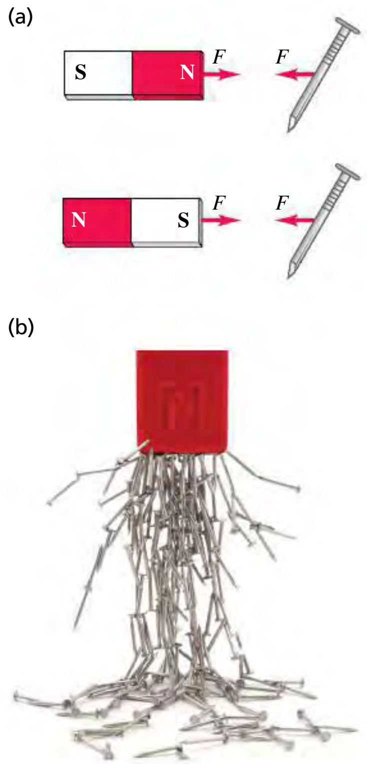
27.1 MAGNETISM

Magnetic phenomena were first observed at least 2500 years ago in fragments of magnetized iron ore found near the ancient city of Magnesia (now Manisa, in western Turkey). These fragments were what are now called **permanent magnets;** you probably have several permanent magnets on your refrigerator door at home. Permanent magnets were found to exert forces on each other as well as on pieces of

27.1 (a) Two bar magnets attract when opposite poles (N and S, or S and N) are next to each other. (b) The bar magnets repel when like poles (N and N, or S and S) are next to each other.



27.2 (a) Either pole of a bar magnet attracts an unmagnetized object that contains iron, such as a nail. (b) A real-life example of this effect.



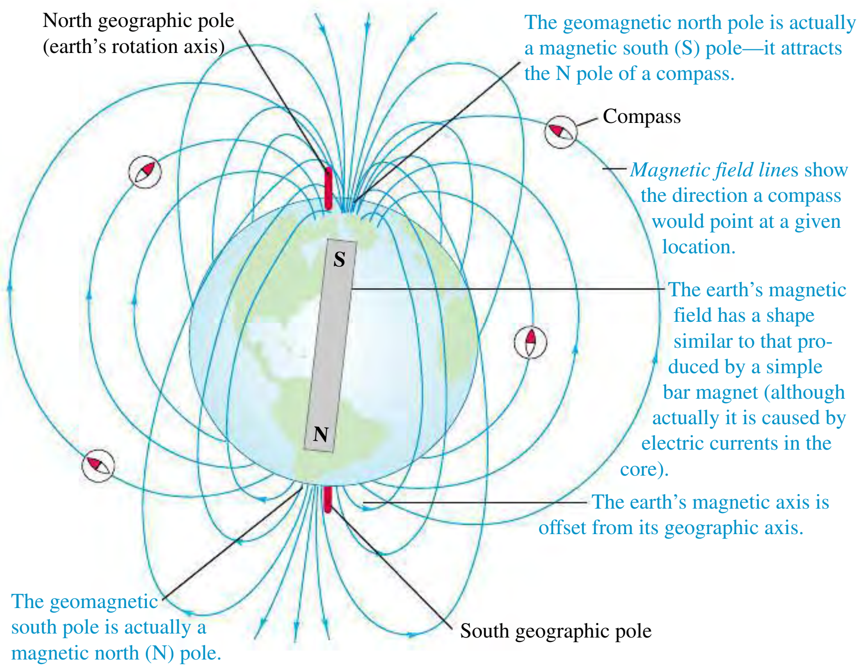
iron that were not magnetized. It was discovered that when an iron rod is brought in contact with a natural magnet, the rod also becomes magnetized. When such a rod is floated on water or suspended by a string from its center, it tends to line itself up in a north-south direction. The needle of an ordinary compass is just such a piece of magnetized iron.

Before the relationship of magnetic interactions to moving charges was understood, the interactions of permanent magnets and compass needles were described in terms of *magnetic poles*. If a bar-shaped permanent magnet, or *bar magnet,* is free to rotate, one end points north. This end is called a *north pole* or *N pole;* the other end is a *south pole* or *S pole*. Opposite poles attract each other, and like poles repel each other (**Fig. 27.1**). An object that contains iron but is not itself magnetized (that is, it shows no tendency to point north or south) is attracted by *either* pole of a permanent magnet (**Fig. 27.2**). This is the attraction that acts between a magnet and the unmagnetized steel door of a refrigerator. By analogy to electric interactions, we describe the interactions in Figs. 27.1 and 27.2 by saying that a bar magnet sets up a *magnetic field* in the space around it and a second body responds to that field. A compass needle tends to align with the magnetic field at the needle's position. ?

The earth itself is a magnet. Its north geographic pole is close to a magnetic *south* pole, which is why the north pole of a compass needle points north. The earth's magnetic axis is not quite parallel to its geographic axis (the axis of rotation), so a compass reading deviates somewhat from geographic north. This deviation, which varies with location, is called *magnetic declination* or *magnetic variation*. Also, the magnetic field is not horizontal at most points on the earth's surface; its angle up or down is called *magnetic inclination*. At the magnetic poles the magnetic field is vertical.

Figure 27.3 is a sketch of the earth's magnetic field. The lines, called *magnetic field lines,* show the direction that a compass would point at each location; they are discussed in detail in Section 27.3. The direction of the field at any point can be defined as the direction of the force that the field would exert on a magnetic north pole. In Section 27.2 we'll describe a more fundamental way to define the direction and magnitude of a magnetic field.

27.3 A sketch of the earth's magnetic field. The field, which is caused by currents in the earth's molten core, changes with time; geologic evidence shows that it reverses direction entirely at irregular intervals of 10^4 to 10^6 years.



Magnetic Poles Versus Electric Charge

The concept of magnetic poles may appear similar to that of electric charge, and north and south poles may seem analogous to positive and negative charges. But the analogy can be misleading. While isolated positive and negative charges exist, there is *no* experimental evidence that one isolated magnetic pole exists; poles always appear in pairs. If a bar magnet is broken in two, each broken end becomes a pole (**Fig. 27.4**). The existence of an isolated magnetic pole, or **magnetic monopole,** would have sweeping implications for theoretical physics. Extensive searches for magnetic monopoles have been carried out, but so far without success.

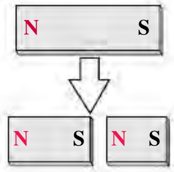
The first evidence of the relationship of magnetism to moving charges was discovered in 1820 by the Danish scientist Hans Christian Oersted. He found that a compass needle was deflected by a current-carrying wire (**Fig. 27.5**). Similar investigations were carried out in France by André Ampère. A few years later, Michael Faraday in England and Joseph Henry in the United States discovered that moving a magnet near a conducting loop can cause a current in the loop. We now know that the magnetic forces between two bodies shown in Figs. 27.1 and 27.2 are fundamentally due to interactions between moving electrons in the atoms of the bodies. (There are also *electric* interactions between the two bodies, but these are far weaker than the magnetic interactions because the bodies are electrically neutral.) Inside a magnetized body such as a permanent magnet, the motion of certain of the atomic electrons is *coordinated;* in an unmagnetized body these motions are not coordinated. (We'll describe these motions further in Section 27.7 and see how the interactions shown in Figs. 27.1 and 27.2 come about.)

Electric and magnetic interactions prove to be intimately connected. Over the next several chapters we will develop the unifying principles of electromagnetism, culminating in the expression of these principles in *Maxwell's equations*. These equations represent the synthesis of electromagnetism, just as Newton's laws of motion are the synthesis of mechanics, and like Newton's laws they represent a towering achievement of the human intellect.

27.4 Breaking a bar magnet. Each piece has a north and south pole, even if the pieces are different sizes. (The smaller the piece, the weaker its magnetism.)

In contrast to electric charges, magnetic poles always come in pairs and can't be isolated.

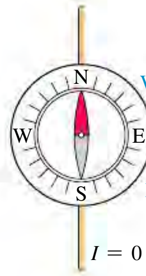
Breaking a magnet in two ...



... yields two magnets, not two isolated poles.

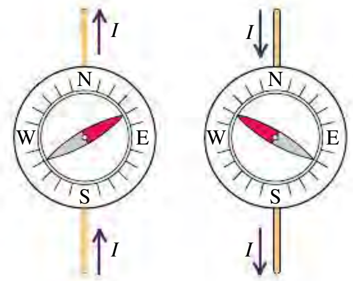
27.5 In Oersted's experiment, a compass is placed directly over a horizontal wire (here viewed from above).

(a)



(b)

When the wire carries a current, the compass needle deflects. The direction of deflection depends on the direction of the current.



TEST YOUR UNDERSTANDING OF SECTION 27.1 Suppose you cut off the part of the compass needle shown in Fig. 27.5a that is painted gray. You discard this part, drill a hole in the remaining red part, and place the red part on the pivot at the center of the compass. Will the red part still swing when a current is applied as in Fig. 27.5b? ❙

27.2 MAGNETIC FIELD

To introduce the concept of magnetic field properly, let's review our formulation of *electric* interactions in Chapter 21, where we introduced the concept of *electric* field. We represented electric interactions in two steps:

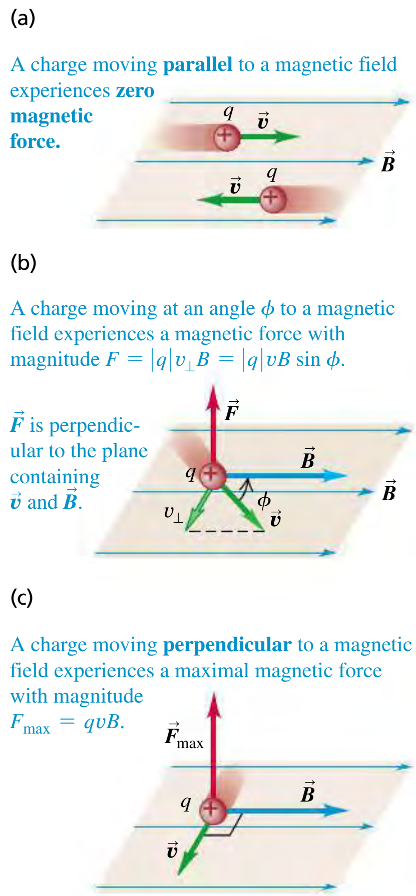
1. A distribution of electric charge creates an electric field $\vec{E}$ in the surrounding space.
2. The electric field exerts a force $\vec{F} = q\vec{E}$ on any other charge q that is present in the field.

We can describe magnetic interactions in a similar way:

1. A moving charge or a current creates a **magnetic field** in the surrounding space (in addition to its *electric* field).
2. The magnetic field exerts a force $\vec{F}$ on any other moving charge or current that is present in the field.

BIO Application Spiny Lobsters and Magnetic Compasses Although the Caribbean spiny lobster (*Panulirus argus*) has a relatively simple nervous system, it is remarkably sensitive to magnetic fields. It has an internal magnetic "compass" that allows it to distinguish north, east, south, and west. This lobster can also sense small differences in the earth's magnetic field from one location to another and may use these differences to help it navigate.

27.6 The magnetic force $\vec{F}$ acting on a positive charge q moving with velocity $\vec{v}$ is perpendicular to both $\vec{v}$ and the magnetic field $\vec{B}$. For given values of speed v and magnetic field strength B, the force is greatest when $\vec{v}$ and $\vec{B}$ are perpendicular.



In this chapter we'll concentrate on the *second* aspect of the interaction: Given the presence of a magnetic field, what force does it exert on a moving charge or a current? In Chapter 28 we will come back to the problem of how magnetic fields are *created* by moving charges and currents.

Like electric field, magnetic field is a *vector field*—that is, a vector quantity associated with each point in space. We will use the symbol $\vec{B}$ for magnetic field. At any position the direction of $\vec{B}$ is defined as the direction in which the north pole of a compass needle tends to point. The arrows in Fig. 27.3 suggest the direction of the earth's magnetic field; for any magnet, $\vec{B}$ points out of its north pole and into its south pole.

Magnetic Forces on Moving Charges

There are four key characteristics of the magnetic force on a moving charge. First, its magnitude is proportional to the magnitude of the charge. If a 1-μC charge and a 2-μC charge move through a given magnetic field with the same velocity, experiments show that the force on the 2-μC charge is twice as great as the force on the 1-μC charge. Second, the magnitude of the force is also proportional to the magnitude, or "strength," of the field; if we double the magnitude of the field (for example, by using two identical bar magnets instead of one) without changing the charge or its velocity, the force doubles.

A third characteristic is that the magnetic force depends on the particle's velocity. This is quite different from the electric-field force, which is the same whether the charge is moving or not. A charged particle at rest experiences *no* magnetic force. And fourth, we find by experiment that the magnetic force $\vec{F}$ *does not* have the same direction as the magnetic field $\vec{B}$ but instead is always *perpendicular* to both $\vec{B}$ and the velocity $\vec{v}$. The magnitude F of the force is proportional to the component of $\vec{v}$ perpendicular to the field; when that component is zero (that is, when $\vec{v}$ and $\vec{B}$ are parallel or antiparallel), the force is zero.

Figure 27.6 shows these relationships. The direction of $\vec{F}$ is always perpendicular to the plane containing $\vec{v}$ and $\vec{B}$. Its magnitude is given by

$$F = |q|v_\perp B = |q|vB\sin\phi \qquad (27.1)$$

where $|q|$ is the magnitude of the charge and ϕ is the angle measured from the direction of $\vec{v}$ to the direction of $\vec{B}$, as shown in the figure.

This description does not specify the direction of $\vec{F}$ completely; there are always two directions, opposite to each other, that are both perpendicular to the plane of $\vec{v}$ and $\vec{B}$. To complete the description, we use the same right-hand rule that we used to define the vector product in Section 1.10. (It would be a good idea to review that section before you go on.) Draw the vectors $\vec{v}$ and $\vec{B}$ with their tails together, as in **Fig. 27.7a**. Imagine turning $\vec{v}$ until it points in the direction of $\vec{B}$ (turning through the smaller of the two possible angles). Wrap the fingers of your right hand around the line perpendicular to the plane of $\vec{v}$ and $\vec{B}$ so that they curl around with the sense of rotation from $\vec{v}$ to $\vec{B}$. Your thumb then points in the direction of the force $\vec{F}$ on a *positive* charge.

This discussion shows that the force on a charge q moving with velocity $\vec{v}$ in a magnetic field $\vec{B}$ is given, both in magnitude and in direction, by

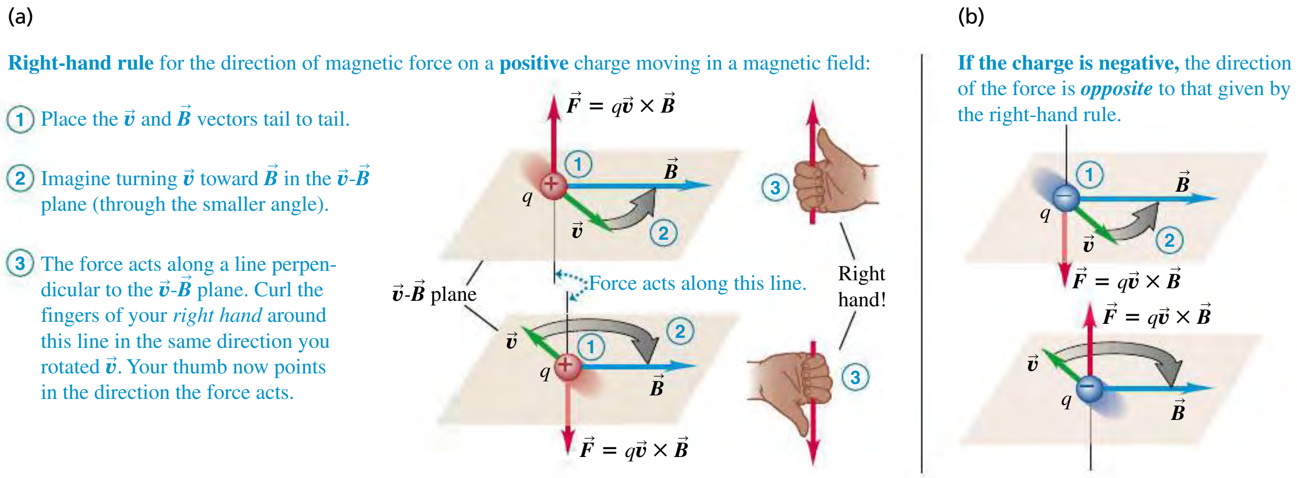
Magnetic force on a moving charged particle

$$\vec{F} = q\vec{v} \times \vec{B} \qquad (27.2)$$

Particle's charge; Particle's velocity; Magnetic field

This is the first of several vector products we will encounter in our study of magnetic-field relationships. It's important to note that Eq. (27.2) was *not* deduced theoretically; it is an observation based on *experiment*.

27.7 Finding the direction of the magnetic force on a moving charged particle.



Equation (27.2) is valid for both positive and negative charges. When q is negative, the direction of the force $\vec{F}$ is opposite to that of $\vec{v} \times \vec{B}$ (Fig. 27.7b). If two charges with equal magnitude and opposite sign move in the same $\vec{B}$ field with the same velocity (**Fig. 27.8**), the forces have equal magnitude and opposite direction. Figures 27.6, 27.7, and 27.8 show several examples of the relationships of the directions of $\vec{F}$, $\vec{v}$, and $\vec{B}$ for both positive and negative charges. Be sure you understand the relationships shown in these figures.

Equation (27.1) gives the magnitude of the magnetic force $\vec{F}$ in Eq. (27.2). Since ϕ is the angle between the directions of vectors $\vec{v}$ and $\vec{B}$, we may interpret $B\sin\phi$ as the component of $\vec{B}$ perpendicular to $\vec{v}$—that is, $B_\perp$. With this notation the force magnitude is

$$F = |q|vB_\perp \tag{27.3}$$

This form may be more convenient, especially in problems involving *currents* rather than individual particles. We'll discuss forces on currents later in this chapter.

From Eq. (27.1) the *units* of B must be the same as the units of F/qv. Therefore the SI unit of B is equivalent to $1\ \mathrm{N \cdot s/C \cdot m}$, or, since one ampere is one coulomb per second $(1\ \mathrm{A} = 1\ \mathrm{C/s})$, $1\ \mathrm{N/A \cdot m}$. This unit is called the **tesla** (abbreviated T), in honor of Nikola Tesla (1856–1943), the prominent Serbian-American scientist and inventor:

$$1\ \text{tesla} = 1\ \mathrm{T} = 1\ \mathrm{N/A \cdot m}$$

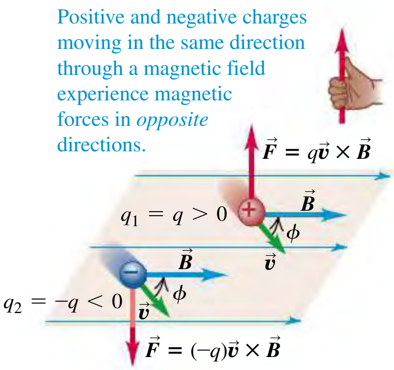
Another unit of B, the **gauss** $(1\ \mathrm{G} = 10^{-4}\ \mathrm{T})$, is also in common use.

The magnetic field of the earth is of the order of 10^{-4} T or 1 G. Magnetic fields of the order of 10 T occur in the interior of atoms and are important in the analysis of atomic spectra. The largest steady magnetic field that can be produced at present in the laboratory is about 45 T. Some pulsed-current electromagnets can produce fields of the order of 120 T for millisecond time intervals.

Measuring Magnetic Fields with Test Charges

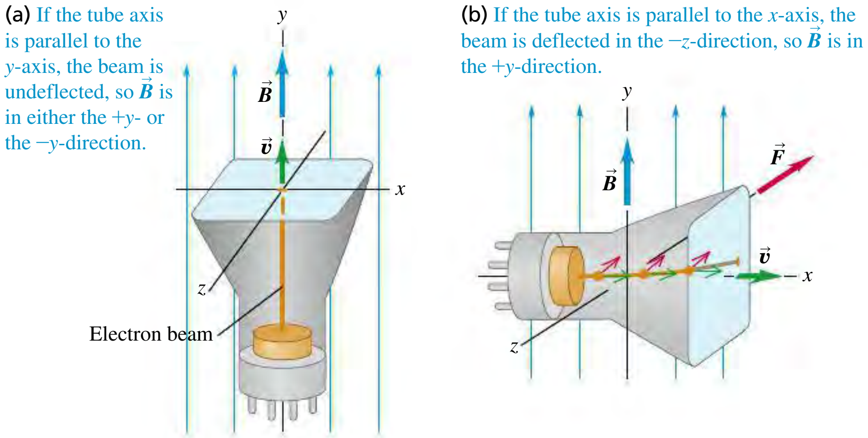
To explore an unknown magnetic field, we can measure the magnitude and direction of the force on a *moving* test charge and then use Eq. (27.2) to determine $\vec{B}$. The electron beam in a cathode-ray tube, such as that in an older television set (not a flat-screen TV), is a convenient device for this. The electron gun shoots out a narrow beam of electrons at a known speed. If there is no force to deflect the beam, it strikes the center of the screen.

27.8 Two charges of the same magnitude but opposite sign moving with the same velocity in the same magnetic field. The magnetic forces on the charges are equal in magnitude but opposite in direction.



BIO Application Magnetic Fields of the Body All living cells are electrically active, and the feeble electric currents within your body produce weak but measurable magnetic fields. The fields produced by skeletal muscles have magnitudes less than 10^{-10} T, about one-millionth as strong as the earth's magnetic field. Your brain produces magnetic fields that are far weaker, only about 10^{-12} T.

27.9 Determining the direction of a magnetic field by using a cathode-ray tube. Because electrons have a negative charge, the magnetic force $\vec{F} = q\vec{v} \times \vec{B}$ in part (b) points opposite to the direction given by the right-hand rule (see Fig. 27.7b).



If a magnetic field is present, in general the electron beam is deflected. But if the beam is parallel or antiparallel to the field, then $\phi = 0$ or π in Eq. (27.1) and $F = 0$; there is no force and hence no deflection. If we find that the electron beam is not deflected when its direction is parallel to a certain axis as in **Fig. 27.9a**, the $\vec{B}$ vector must point either up or down along that axis.

If we then turn the tube 90° (Fig. 27.9b), $\phi = \pi/2$ in Eq. (27.1) and the magnetic force is maximum; the beam is deflected in a direction perpendicular to the plane of $\vec{B}$ and $\vec{v}$. The direction and magnitude of the deflection determine the direction and magnitude of $\vec{B}$. We can perform additional experiments in which the angle between $\vec{B}$ and $\vec{v}$ is between 0° and 90° to confirm Eq. (27.1). We note that the electron has a negative charge; the force in Fig. 27.9b is opposite in direction to the force on a positive charge.

When a charged particle moves through a region of space where *both* electric and magnetic fields are present, both fields exert forces on the particle. The total force $\vec{F}$ is the vector sum of the electric and magnetic forces:

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \tag{27.4}$$

PROBLEM-SOLVING STRATEGY 27.1 MAGNETIC FORCES

IDENTIFY *the relevant concepts:* The equation $\vec{F} = q\vec{v} \times \vec{B}$ allows you to determine the magnetic force on a moving charged particle.

SET UP *the problem* using the following steps:
1. Draw the velocity $\vec{v}$ and magnetic field $\vec{B}$ with their tails together so that you can visualize the plane that contains them.
2. Determine the angle ϕ between $\vec{v}$ and $\vec{B}$.
3. Identify the target variables.

EXECUTE *the solution* as follows:
1. Use Eq. (27.2), $\vec{F} = q\vec{v} \times \vec{B}$, to express the magnetic force. Equation (27.1) gives the magnitude of the force, $F = qvB\sin\phi$.
2. Remember that $\vec{F}$ is perpendicular to the plane containing $\vec{v}$ and $\vec{B}$. The right-hand rule (see Fig. 27.7) gives the direction of $\vec{v} \times \vec{B}$. If q is negative, $\vec{F}$ is *opposite* to $\vec{v} \times \vec{B}$.

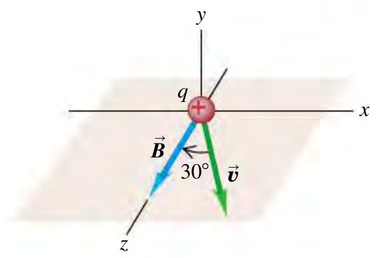
EVALUATE *your answer:* Whenever possible, solve the problem in two ways to confirm that the results agree. Do it directly from the geometric definition of the vector product. Then find the components of the vectors in some convenient coordinate system and calculate the vector product from the components. Verify that the results agree.

EXAMPLE 27.1 MAGNETIC FORCE ON A PROTON

SOLUTION

A beam of protons ($q = 1.6 \times 10^{-19}$ C) moves at 3.0×10^5 m/s through a uniform 2.0-T magnetic field directed along the positive z-axis (**Fig. 27.10**). The velocity of each proton lies in the xz-plane and is directed at 30° to the $+z$-axis. Find the force on a proton.

27.10 Directions of $\vec{v}$ and $\vec{B}$ for a proton in a magnetic field.



SOLUTION

IDENTIFY and SET UP: This problem uses the expression $\vec{F} = q\vec{v} \times \vec{B}$ for the magnetic force $\vec{F}$ on a moving charged particle. The target variable is $\vec{F}$.

EXECUTE: The charge is positive, so the force is in the same direction as the vector product $\vec{v} \times \vec{B}$. From the right-hand rule, this direction is along the negative y-axis. The magnitude of the force, from Eq. (27.1), is

$$\begin{aligned} F &= qvB\sin\phi \\ &= (1.6 \times 10^{-19}\ \text{C})(3.0 \times 10^5\ \text{m/s})(2.0\ \text{T})(\sin 30°) \\ &= 4.8 \times 10^{-14}\ \text{N} \end{aligned}$$

EVALUATE: To check our result, we evaluate the force by using vector language and Eq. (27.2). We have

$$\vec{v} = (3.0 \times 10^5\ \text{m/s})(\sin 30°)\hat{\imath} + (3.0 \times 10^5\ \text{m/s})(\cos 30°)\hat{k}$$

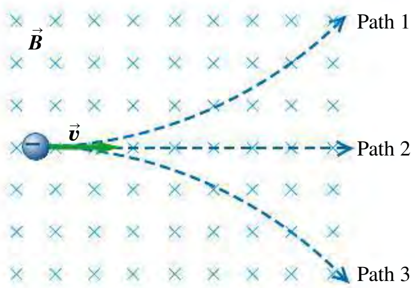
$$\vec{B} = (2.0\ \text{T})\hat{k}$$

$$\begin{aligned} \vec{F} &= q\vec{v} \times \vec{B} \\ &= (1.6 \times 10^{-19}\ \text{C})(3.0 \times 10^5\ \text{m/s})(2.0\ \text{T}) \\ &\quad \times (\sin 30°\hat{\imath} + \cos 30°\hat{k}) \times \hat{k} \\ &= (-4.8 \times 10^{-14}\ \text{N})\hat{\jmath} \end{aligned}$$

(Recall that $\hat{\imath} \times \hat{k} = -\hat{\jmath}$ and $\hat{k} \times \hat{k} = \mathbf{0}$.) We again find that the force is in the negative y-direction with magnitude 4.8×10^{-14} N.

If the beam consists of *electrons* rather than protons, the charge is negative ($q = -1.6 \times 10^{-19}$ C) and the direction of the force is reversed. The force is now directed along the *positive* y-axis, but the magnitude is the same as before, $F = 4.8 \times 10^{-14}$ N.

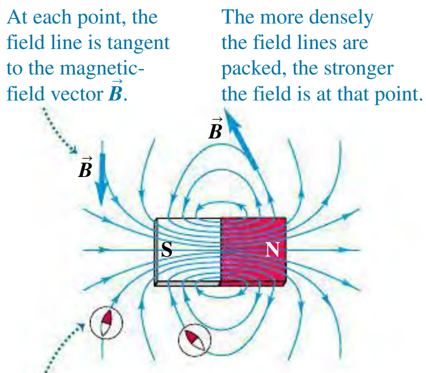
TEST YOUR UNDERSTANDING OF SECTION 27.2 The accompanying figure shows a uniform magnetic field $\vec{B}$ directed into the plane of the paper (shown by the blue ×'s). A particle with a negative charge moves in the plane. Which path—1, 2, or 3—does the particle follow? ❙



27.3 MAGNETIC FIELD LINES AND MAGNETIC FLUX

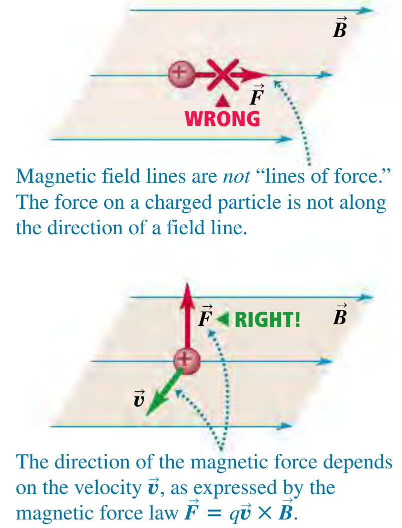
We can represent any magnetic field by **magnetic field lines,** just as we did for the earth's magnetic field in Fig. 27.3. The idea is the same as for the electric field lines we introduced in Section 21.6. We draw the lines so that the line through any point is tangent to the magnetic field vector $\vec{B}$ at that point (**Fig. 27.11**). Just as with electric field lines, we draw only a few representative lines; otherwise, the lines would fill up all of space. Where adjacent field lines are close together, the field magnitude is large; where these field lines are far apart, the field magnitude is small. Also, because the direction of $\vec{B}$ at each point is unique, field lines never intersect.

27.11 Magnetic field lines of a permanent magnet. Note that the field lines pass through the interior of the magnet.



CAUTION Magnetic field lines are not "lines of force" Unlike electric field lines, magnetic field lines *do not* point in the direction of the force on a charge (**Fig. 27.12**, next page). Equation (27.2) shows that the force on a moving charged particle is always perpendicular to the magnetic field and hence to the magnetic field line that passes through the particle's position. The direction of the force depends on the particle's velocity and the sign of its charge, so just looking at magnetic field lines cannot tell you the direction of the force on an arbitrary moving charged particle. Magnetic field lines *do* have the direction that a compass needle would point at each location; this may help you visualize them. ❙

27.12 Magnetic field lines are *not* "lines of force."



Figures 27.11 and 27.13 show magnetic field lines produced by several common sources of magnetic field. In the gap between the poles of the magnet shown in **Fig. 27.13a**, the field lines are approximately straight, parallel, and equally spaced, showing that the magnetic field in this region is approximately *uniform* (that is, constant in magnitude and direction).

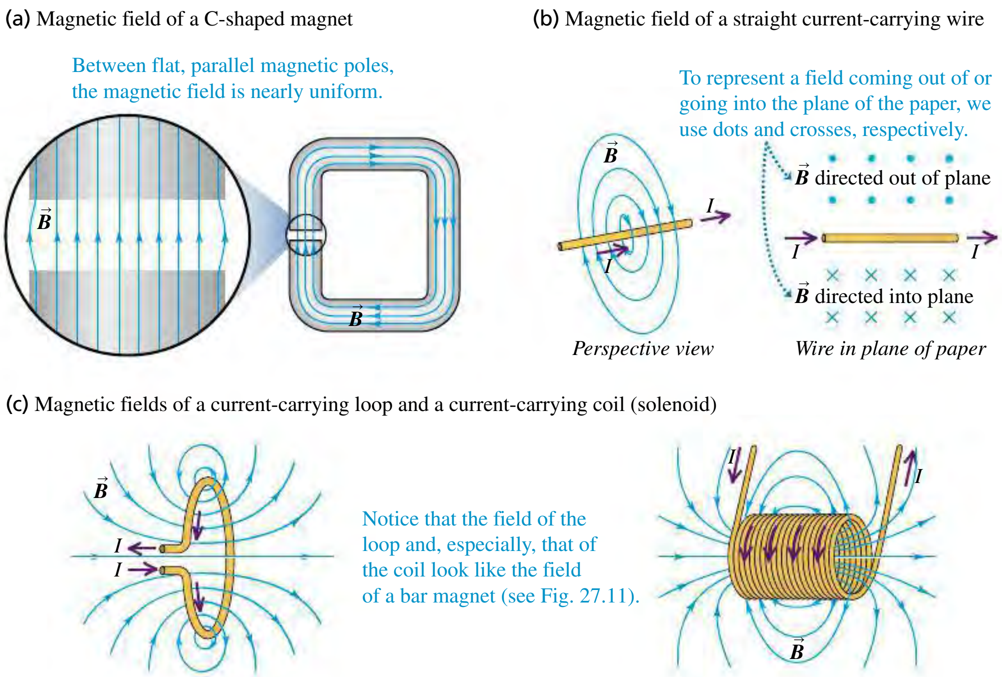
Because magnetic-field patterns are three-dimensional, it's often necessary to draw magnetic field lines that point into or out of the plane of a drawing. To do this we use a dot (•) to represent a vector directed out of the plane and a cross (×) to represent a vector directed into the plane (Fig. 27.13b). To remember these, think of a dot as the head of an arrow coming directly toward you, and think of a cross as the feathers of an arrow flying directly away from you.

Iron filings, like compass needles, tend to align with magnetic field lines. Hence they provide an easy way to visualize field lines (**Fig. 27.14**).

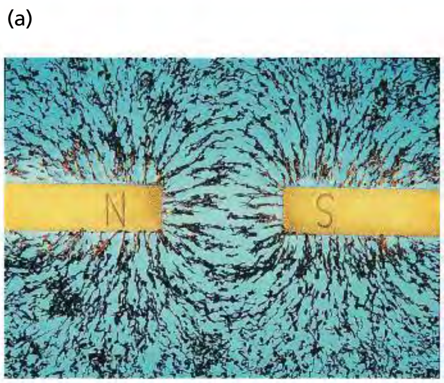
Magnetic Flux and Gauss's Law for Magnetism

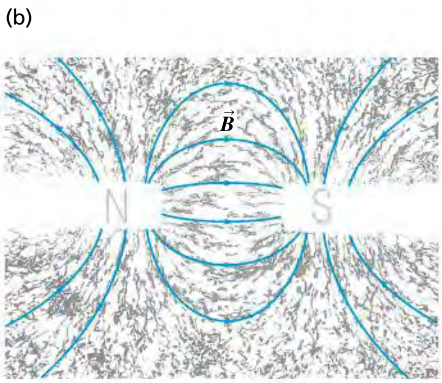
We define the **magnetic flux** Φ_B through a surface just as we defined electric flux in connection with Gauss's law in Section 22.2. We can divide any surface into elements of area dA (**Fig. 27.15**). For each element we determine $B_\perp$, the component of $\vec{B}$ normal to the surface at the position of that element, as shown.

27.13 Magnetic field lines produced by some common sources of magnetic field.



27.14 (a) Like little compass needles, iron filings line up tangent to magnetic field lines. (b) Drawing of field lines for the situation shown in (a).





From the figure, $B_\perp = B\cos\phi$, where ϕ is the angle between the direction of $\vec{B}$ and a line perpendicular to the surface. (Be careful not to confuse ϕ with Φ_B.) In general, this component varies from point to point on the surface. We define the magnetic flux $d\Phi_B$ through this area as

$$d\Phi_B = B_\perp dA = B\cos\phi\, dA = \vec{B}\cdot d\vec{A} \tag{27.5}$$

The *total* magnetic flux through the surface is the sum of the contributions from the individual area elements:

Magnetic flux through a surface; Magnitude of magnetic field $\vec{B}$; Component of $\vec{B}$ perpendicular to surface

$$\Phi_B = \int B\cos\phi\, dA = \int B_\perp\, dA = \int \vec{B}\cdot d\vec{A} \tag{27.6}$$

Angle between $\vec{B}$ and normal to surface; Element of surface area; Vector element of surface area

(Review the concepts of vector area and surface integral in Section 22.2.)

Magnetic flux is a *scalar* quantity. If $\vec{B}$ is uniform over a plane surface with total area A, then $B_\perp$ and ϕ are the same at all points on the surface, and

$$\Phi_B = B_\perp A = BA\cos\phi \tag{27.7}$$

If $\vec{B}$ is also perpendicular to the surface (parallel to the area vector), then $\cos\phi = 1$ and Eq. (27.7) reduces to $\Phi_B = BA$. We will use the concept of magnetic flux extensively during our study of electromagnetic induction in Chapter 29.

The SI unit of magnetic flux is equal to the unit of magnetic field (1 T) times the unit of area $(1\ \text{m}^2)$. This unit is called the **weber** (1 Wb), in honor of the German physicist Wilhelm Weber (1804–1891):

$$1\ \text{Wb} = 1\ \text{T}\cdot\text{m}^2$$

Also, $1\ \text{T} = 1\ \text{N/A}\cdot\text{m}$, so

$$1\ \text{Wb} = 1\ \text{T}\cdot\text{m}^2 = 1\ \text{N}\cdot\text{m/A}$$

In Gauss's law the total *electric* flux through a closed surface is proportional to the total electric charge enclosed by the surface. For example, if the closed surface encloses an electric dipole, the total electric flux is zero because the total charge is zero. (You may want to review Section 22.3 on Gauss's law.) By analogy, if there were such a thing as a single magnetic charge (magnetic monopole), the total *magnetic* flux through a closed surface would be proportional to the total magnetic charge enclosed. But we have mentioned that no magnetic monopole has ever been observed, despite intensive searches. This leads us to **Gauss's law for magnetism:**

Gauss's law for magnetism:

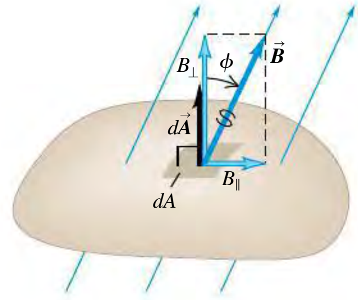
$$\underbrace{\oint \vec{B}\cdot d\vec{A}}_{\text{The total magnetic flux through } any \textbf{ closed surface} \ldots} = 0 \ \ \ldots\text{ equals zero.} \tag{27.8}$$

You can verify Gauss's law for magnetism by examining Figs. 27.11 and 27.13; if you draw a closed surface anywhere in any of the field maps shown in those figures, you will see that every field line that enters the surface also exits from it; the net flux through the surface is zero. It also follows from Eq. (27.8) that magnetic field lines always form closed loops.

CAUTION **Magnetic field lines have no ends** Unlike electric field lines, which begin and end on electric charges, magnetic field lines *never* have endpoints; such a point would indicate the presence of a monopole. You might be tempted to draw magnetic field lines that begin at the north pole of a magnet and end at a south pole. But as Fig. 27.11 shows, a magnet's field lines continue through the interior of the magnet. Like all other magnetic field lines, they form closed loops. ▌

For Gauss's law, which always deals with *closed* surfaces, the vector area element $d\vec{A}$ in Eq. (27.6) always points *out of* the surface. However, some

27.15 The magnetic flux through an area element dA is defined to be $d\Phi_B = B_\perp dA$.



PhET: Magnet and Compass
PhET: Magnets and Electromagnets

DATA *SPEAKS*

Magnetic Forces and Magnetic Field Lines

When students were given a problem involving magnetic forces and field lines, more than 16% gave an incorrect response. Common errors:

- Forgetting that only the component of the magnetic field that is perpendicular to the velocity of a charged particle causes a force on the particle. If there is no perpendicular component, there is no magnetic force.
- Forgetting that the magnetic force on a moving charged particle is not directed along a magnetic field line. That force is perpendicular to the magnetic field as well as to the particle's velocity.

applications of *magnetic* flux involve an *open* surface with a boundary line; there is then an ambiguity of sign in Eq. (27.6) because of the two possible choices of direction for $d\vec{A}$. In these cases we choose one of the two sides of the surface to be the "positive" side and use that choice consistently.

If the element of area dA in Eq. (27.5) is at right angles to the field lines, then $B_\perp = B$; calling the area $dA_\perp$, we have

$$B = \frac{d\Phi_B}{dA_\perp} \tag{27.9}$$

That is, the magnitude of magnetic field is equal to *flux per unit area* across an area at right angles to the magnetic field. For this reason, magnetic field $\vec{B}$ is sometimes called **magnetic flux density.**

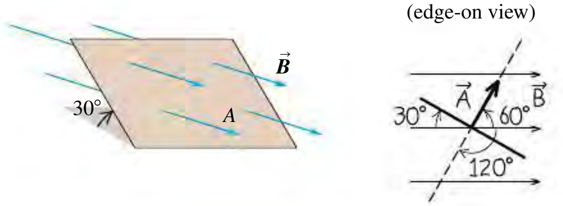
EXAMPLE 27.2 MAGNETIC FLUX CALCULATIONS

Figure 27.16a is a perspective view of a flat surface with area $3.0\ \text{cm}^2$ in a uniform magnetic field $\vec{B}$. The magnetic flux through this surface is $+0.90$ mWb. Find the magnitude of the magnetic field and the direction of the area vector $\vec{A}$.

27.16 (a) A flat area A in a uniform magnetic field $\vec{B}$. (b) The area vector $\vec{A}$ makes a 60° angle with $\vec{B}$. (If we had chosen $\vec{A}$ to point in the opposite direction, ϕ would have been 120° and the magnetic flux Φ_B would have been negative.)

(a) Perspective view

(b) Our sketch of the problem (edge-on view)



SOLUTION

IDENTIFY and SET UP: Our target variables are the field magnitude B and the direction of the area vector. Because $\vec{B}$ is uniform, B and ϕ are the same at all points on the surface. Hence we can use Eq. (27.7), $\Phi_B = BA\cos\phi$.

EXECUTE: The area A is $3.0 \times 10^{-4}\ \text{m}^2$; the direction of $\vec{A}$ is perpendicular to the surface, so ϕ could be either 60° or 120°. But Φ_B, B, and A are all positive, so $\cos\phi$ must also be positive. This rules out 120°, so $\phi = 60°$ (Fig. 27.16b). Hence we find

$$B = \frac{\Phi_B}{A\cos\phi} = \frac{0.90 \times 10^{-3}\ \text{Wb}}{(3.0 \times 10^{-4}\ \text{m}^2)(\cos 60°)} = 6.0\ \text{T}$$

EVALUATE: In many problems we are asked to calculate the flux of a given magnetic field through a given area. This example is somewhat different: It tests your understanding of the definition of magnetic flux.

TEST YOUR UNDERSTANDING OF SECTION 27.3 Imagine moving along the axis of the current-carrying loop in Fig. 27.13c, starting at a point well to the left of the loop and ending at a point well to the right of the loop. (a) How would the magnetic field strength vary as you moved along this path? (i) It would be the same at all points along the path; (ii) it would increase and then decrease; (iii) it would decrease and then increase. (b) Would the magnetic field direction vary as you moved along the path?

27.4 MOTION OF CHARGED PARTICLES IN A MAGNETIC FIELD

When a charged particle moves in a magnetic field, it is acted on by the magnetic force given by Eq. (27.2), and the motion is determined by Newton's laws. **Figure 27.17a** shows a simple example. A particle with positive charge q is at point O, moving with velocity $\vec{v}$ in a uniform magnetic field $\vec{B}$ directed into the plane of the figure. The vectors $\vec{v}$ and $\vec{B}$ are perpendicular, so the magnetic force $\vec{F} = q\vec{v} \times \vec{B}$ has magnitude $F = qvB$ and a direction as shown in the figure. The force is *always* perpendicular to $\vec{v}$, so it cannot change the *magnitude* of the velocity, only its direction. To put it differently, the magnetic force never has a component parallel to the particle's motion, so the magnetic force can never do *work* on the particle. This is true even if the magnetic field is not uniform.

Motion of a charged particle under the action of a magnetic field alone is always motion with constant speed.

Using this principle, we see that in the situation shown in Fig. 27.17a the magnitudes of both $\vec{F}$ and $\vec{v}$ are constant. As the particle of mass m moves from O to P to S, the directions of force and velocity change but their magnitudes stay the same. The particle therefore moves under the influence of a constant-magnitude force that is always at right angles to the velocity of the particle. Comparing the discussion of circular motion in Sections 3.4 and 5.4, we see that the particle's path is a *circle,* traced out with constant speed v. The centripetal acceleration is v^2/R and only the magnetic force acts, so from Newton's second law,

$$F = |q|vB = m\frac{v^2}{R} \tag{27.10}$$

We solve Eq. (27.10) for R:

Radius of a circular orbit in a magnetic field

$$R = \frac{mv}{|q|B} \tag{27.11}$$

Particle's mass; Particle's speed; Magnetic-field magnitude; Particle's charge

If the charge q is negative, the particle moves *clockwise* around the orbit in Fig. 27.17a.

The angular speed ω of the particle can be found from Eq. (9.13), $v = R\omega$. Combining this with Eq. (27.11), we get

$$\omega = \frac{v}{R} = v\frac{|q|B}{mv} = \frac{|q|B}{m} \tag{27.12}$$

The number of revolutions per unit time is $f = \omega/2\pi$. This frequency f is independent of the radius R of the path. It is called the **cyclotron frequency;** in a particle accelerator called a *cyclotron*, particles moving in nearly circular paths are given a boost twice each revolution, increasing their energy and their orbital radii but not their angular speed or frequency. Similarly, one type of *magnetron,* a common source of microwave radiation for microwave ovens and radar systems, emits radiation with a frequency equal to the frequency of circular motion of electrons in a vacuum chamber between the poles of a magnet.

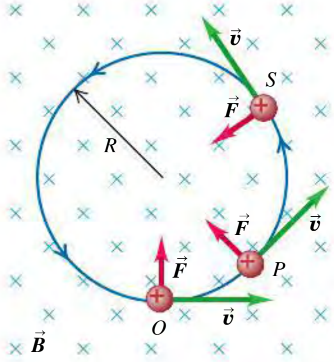
If the direction of the initial velocity is *not* perpendicular to the field, the velocity *component* parallel to the field is constant because there is no force parallel to the field. Then the particle moves in a helix (**Fig. 27.18**). The radius of the helix is given by Eq. (27.11), where v is now the component of velocity perpendicular to the $\vec{B}$ field.

Motion of a charged particle in a nonuniform magnetic field is more complex. **Figure 27.19** shows a field produced by two circular coils separated by some distance. Particles near either coil experience a magnetic force toward the center

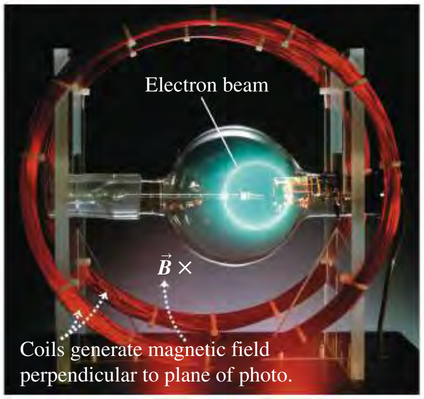
27.17 A charged particle moves in a plane perpendicular to a uniform magnetic field $\vec{B}$.

(a) The orbit of a charged particle in a uniform magnetic field

A charge moving at right angles to a uniform $\vec{B}$ field moves in a circle at constant speed because $\vec{F}$ and $\vec{v}$ are always perpendicular to each other.



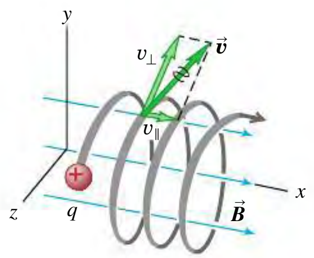
(b) An electron beam (seen as a white arc) curving in a magnetic field



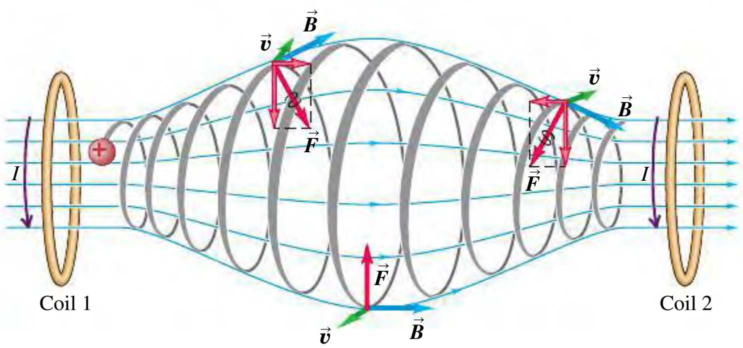
DEMO

27.18 The general case of a charged particle moving in a uniform magnetic field $\vec{B}$. The magnetic field does no work on the particle, so its speed and kinetic energy remain constant.

This particle's motion has components both parallel ($v_\parallel$) and perpendicular ($v_\perp$) to the magnetic field, so it moves in a helical path.

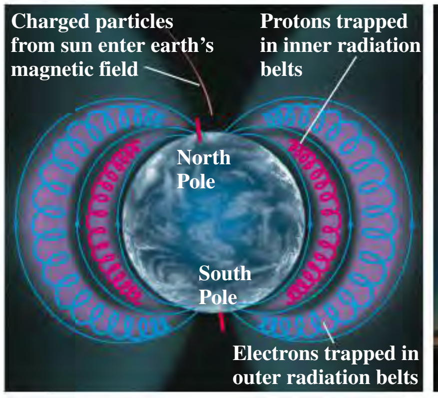


27.19 A magnetic bottle. Particles near either end of the region experience a magnetic force toward the center of the region. This is one way of containing an ionized gas that has a temperature of the order of 10^6 K, which would vaporize any material container.



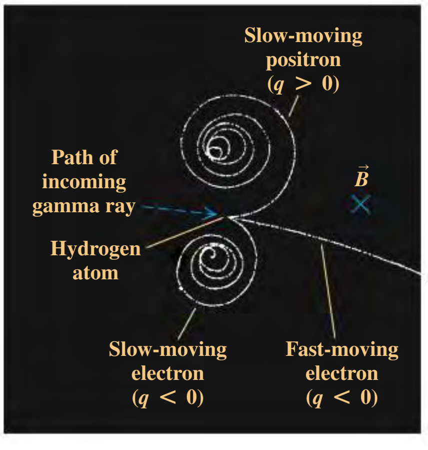
27.20 (a) The Van Allen radiation belts around the earth. Near the poles, charged particles from these belts can enter the atmosphere, producing the aurora borealis ("northern lights") and aurora australis ("southern lights"). (b) A photograph of the aurora borealis.

(a)



(b)

27.21 This bubble chamber image shows the result of a high-energy gamma ray (which does not leave a track) that collides with an electron in a hydrogen atom. This electron flies off to the right at high speed. Some of the energy in the collision is transformed into a second electron and a positron (a positively charged electron). A magnetic field is directed into the plane of the image, which makes the positive and negative particles curve off in different directions.



of the region; particles with appropriate speeds spiral repeatedly from one end of the region to the other and back. Because charged particles can be trapped in such a magnetic field, it is called a *magnetic bottle*. This technique is used to confine very hot plasmas with temperatures of the order of 10^6 K. In a similar way the earth's nonuniform magnetic field traps charged particles coming from the sun in doughnut-shaped regions around the earth, as shown in **Fig. 27.20**. These regions, called the *Van Allen radiation belts,* were discovered in 1958 from data obtained by instruments aboard the Explorer I satellite.

Magnetic forces on charged particles play an important role in studies of elementary particles. **Figure 27.21** shows a chamber filled with liquid hydrogen and with a magnetic field directed into the plane of the photograph. A high-energy gamma ray dislodges an electron from a hydrogen atom, sending it off at high speed and creating a visible track in the liquid hydrogen. The track shows the electron curving downward due to the magnetic force. The energy of the collision also produces another electron and a *positron* (a positively charged electron). Because of their opposite charges, the trajectories of the electron and the positron curve in opposite directions. As these particles plow through the liquid hydrogen, they collide with other charged particles, losing energy and speed. As a result, the radius of curvature decreases as suggested by Eq. (27.11). (The electron's speed is comparable to the speed of light, so Eq. (27.11) isn't directly applicable here.) Similar experiments allow physicists to determine the mass and charge of newly discovered particles.

PROBLEM-SOLVING STRATEGY 27.2 MOTION IN MAGNETIC FIELDS

IDENTIFY *the relevant concepts:* In analyzing the motion of a charged particle in electric and magnetic fields, you will apply Newton's second law of motion, $\sum\vec{F} = m\vec{a}$, with the net force given by $\sum\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$. Often other forces such as gravity can be ignored. Many of the problems are similar to the trajectory and circular-motion problems in Sections 3.3, 3.4, and 5.4; it would be a good idea to review those sections.

SET UP *the problem* using the following steps:

1. Determine the target variable(s).
2. Often the use of components is the most efficient approach. Choose a coordinate system and then express all vector quantities in terms of their components in this system.

EXECUTE *the solution* as follows:

1. If the particle moves perpendicular to a uniform magnetic field, the trajectory is a circle with a radius and angular speed given by Eqs. (27.11) and (27.12), respectively.
2. If your calculation involves a more complex trajectory, use $\sum\vec{F} = m\vec{a}$ in component form: $\sum F_x = ma_x$, and so forth. This approach is particularly useful when both electric and magnetic fields are present.

EVALUATE *your answer:* Check whether your results are reasonable.

EXAMPLE 27.3 ELECTRON MOTION IN A MAGNETRON

SOLUTION

A magnetron in a microwave oven emits electromagnetic waves with frequency $f = 2450\ \text{MHz}$. What magnetic field strength is required for electrons to move in circular paths with this frequency?

SOLUTION

IDENTIFY and SET UP: The problem refers to circular motion as shown in Fig. 27.17a. We use Eq. (27.12) to solve for the field magnitude B.

EXECUTE: The angular speed that corresponds to the frequency f is $\omega = 2\pi f = (2\pi)(2450 \times 10^6\ \text{s}^{-1}) = 1.54 \times 10^{10}\ \text{s}^{-1}$. Then from Eq. (27.12),

$$B = \frac{m\omega}{|q|} = \frac{(9.11 \times 10^{-31}\ \text{kg})(1.54 \times 10^{10}\ \text{s}^{-1})}{1.60 \times 10^{-19}\ \text{C}} = 0.0877\ \text{T}$$

EVALUATE: This is a moderate field strength, easily produced with a permanent magnet. Incidentally, 2450-MHz electromagnetic waves are useful for heating and cooking food because they are strongly absorbed by water molecules.

EXAMPLE 27.4 HELICAL PARTICLE MOTION IN A MAGNETIC FIELD

SOLUTION

In a situation like that shown in Fig. 27.18, the charged particle is a proton ($q = 1.60 \times 10^{-19}\ \text{C}$, $m = 1.67 \times 10^{-27}\ \text{kg}$) and the uniform, 0.500-T magnetic field is directed along the x-axis. At $t = 0$ the proton has velocity components $v_x = 1.50 \times 10^5\ \text{m/s}$, $v_y = 0$, and $v_z = 2.00 \times 10^5\ \text{m/s}$. Only the magnetic force acts on the proton. (a) At $t = 0$, find the force on the proton and its acceleration. (b) Find the radius of the resulting helical path, the angular speed of the proton, and the *pitch* of the helix (the distance traveled along the helix axis per revolution).

SOLUTION

IDENTIFY and SET UP: The magnetic force is $\vec{F} = q\vec{v} \times \vec{B}$; Newton's second law gives the resulting acceleration. Because $\vec{F}$ is perpendicular to $\vec{v}$, the proton's speed does not change. Hence Eq. (27.11) gives the radius of the helical path if we replace v with the velocity component perpendicular to $\vec{B}$. Equation (27.12) gives the angular speed ω, which yields the time T for one revolution (the *period*). Given the velocity component parallel to the magnetic field, we can then determine the pitch.

EXECUTE: (a) With $\vec{B} = B\hat{\imath}$ and $\vec{v} = v_x\hat{\imath} + v_z\hat{k}$, Eq. (27.2) yields

$$\begin{aligned}\vec{F} &= q\vec{v} \times \vec{B} = q(v_x\hat{\imath} + v_z\hat{k}) \times B\hat{\imath} = qv_zB\hat{\jmath}\\ &= (1.60 \times 10^{-19}\ \text{C})(2.00 \times 10^5\ \text{m/s})(0.500\ \text{T})\hat{\jmath}\\ &= (1.60 \times 10^{-14}\ \text{N})\hat{\jmath}\end{aligned}$$

(Recall: $\hat{\imath} \times \hat{\imath} = \mathbf{0}$ and $\hat{k} \times \hat{\imath} = \hat{\jmath}$.) The resulting acceleration is

$$\vec{a} = \frac{\vec{F}}{m} = \frac{1.60 \times 10^{-14}\ \text{N}}{1.67 \times 10^{-27}\ \text{kg}}\hat{\jmath} = (9.58 \times 10^{12}\ \text{m/s}^2)\hat{\jmath}$$

(b) Since $v_y = 0$, the component of velocity perpendicular to $\vec{B}$ is v_z; then from Eq. (27.11),

$$\begin{aligned}R &= \frac{mv_z}{|q|B} = \frac{(1.67 \times 10^{-27}\ \text{kg})(2.00 \times 10^5\ \text{m/s})}{(1.60 \times 10^{-19}\ \text{C})(0.500\ \text{T})}\\ &= 4.18 \times 10^{-3}\ \text{m} = 4.18\ \text{mm}\end{aligned}$$

From Eq. (27.12) the angular speed is

$$\omega = \frac{|q|B}{m} = \frac{(1.60 \times 10^{-19}\ \text{C})(0.500\ \text{T})}{1.67 \times 10^{-27}\ \text{kg}} = 4.79 \times 10^7\ \text{rad/s}$$

The period is $T = 2\pi/\omega = 2\pi/(4.79 \times 10^7\ \text{s}^{-1}) = 1.31 \times 10^{-7}\ \text{s}$. The pitch is the distance traveled along the x-axis in this time, or

$$\begin{aligned}v_xT &= (1.50 \times 10^5\ \text{m/s})(1.31 \times 10^{-7}\ \text{s})\\ &= 0.0197\ \text{m} = 19.7\ \text{mm}\end{aligned}$$

EVALUATE: Although the magnetic force has a tiny magnitude, it produces an immense acceleration because the proton mass is so small. Note that the pitch of the helix is almost five times greater than the radius R, so this helix is much more "stretched out" than that shown in Fig. 27.18.

TEST YOUR UNDERSTANDING OF SECTION 27.4 (a) If you double the speed of the charged particle in Fig. 27.17a while keeping the magnetic field the same (as well as the charge and the mass), how does this affect the radius of the trajectory? (i) The radius is unchanged; (ii) the radius is twice as large; (iii) the radius is four times as large; (iv) the radius is $\frac{1}{2}$ as large; (v) the radius is $\frac{1}{4}$ as large. (b) How does this affect the time required for one complete circular orbit? (i) The time is unchanged; (ii) the time is twice as long; (iii) the time is four times as long; (iv) the time is $\frac{1}{2}$ as long; (v) the time is $\frac{1}{4}$ as long. ▌

27.5 APPLICATIONS OF MOTION OF CHARGED PARTICLES

This section describes several applications of the principles introduced in this chapter. Study them carefully, watching for applications of Problem-Solving Strategy 27.2 (Section 27.4).

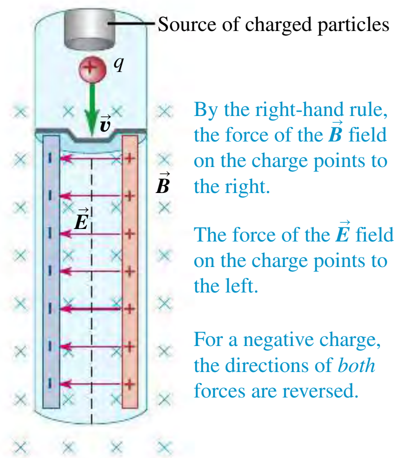
Velocity Selector

In a beam of charged particles produced by a heated cathode or a radioactive material, not all particles move with the same speed. Many applications, however, require a beam in which all the particle speeds are the same. Particles of a specific speed can be selected from the beam by using an arrangement of electric and magnetic fields called a *velocity selector.* In **Fig. 27.22a** a charged particle with mass m, charge q, and speed v enters a region of space where the electric and magnetic fields are perpendicular to the particle's velocity and to each other. The electric field $\vec{E}$ is to the left, and the magnetic field $\vec{B}$ is into the plane of the figure. If q is positive, the electric force is to the left, with magnitude qE, and the magnetic force is to the right, with magnitude qvB. For given field magnitudes E and B, for a particular value of v the electric and magnetic forces will be equal in magnitude; the total force is then zero, and the particle travels in a straight line with constant velocity. This will be the case if $qE = qvB$ (Fig. 27.22b), so the speed v for which there is no deflection is

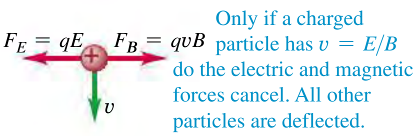
$$v = \frac{E}{B} \tag{27.13}$$

27.22 (a) A velocity selector for charged particles uses perpendicular $\vec{E}$ and $\vec{B}$ fields. Only charged particles with $v = E/B$ move through undeflected. (b) The electric and magnetic forces on a positive charge. The forces are reversed if the charge is negative.

(a) Schematic diagram of velocity selector



(b) Free-body diagram for a positive particle

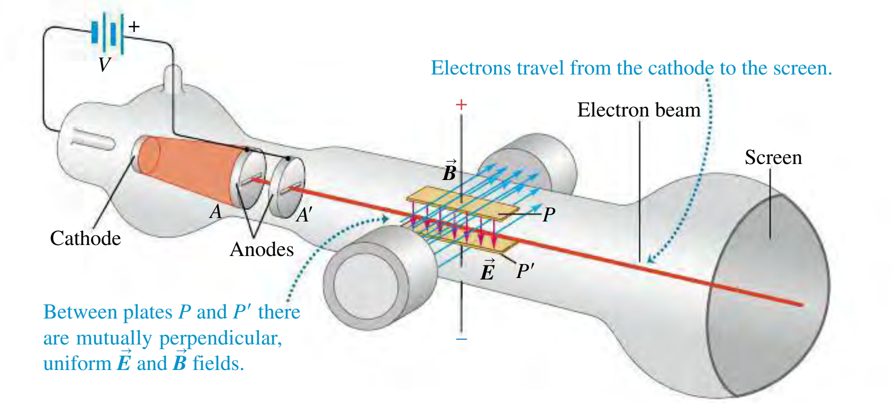


Only particles with speeds equal to E/B can pass through without being deflected by the fields. By adjusting E and B appropriately, we can select particles having a particular speed for use in other experiments. Because q divides out in Eq. (27.13), a velocity selector for positively charged particles also works for electrons or other negatively charged particles.

Thomson's e/m Experiment

In one of the landmark experiments in physics at the end of the 19th century, J. J. Thomson (1856–1940) used the idea just described to measure the ratio of charge to mass for the electron. For this experiment, carried out in 1897 at the Cavendish Laboratory in Cambridge, England, Thomson used the apparatus shown in **Fig. 27.23**. In a highly evacuated glass container, electrons from the hot cathode are accelerated and formed into a beam by a potential difference V between the two anodes A and A'. The speed v of the electrons is determined

27.23 Thomson's apparatus for measuring the ratio e/m for the electron.



by the accelerating potential V. The gained kinetic energy $\frac{1}{2}mv^2$ equals the lost electric potential energy eV, where e is the magnitude of the electron charge:

$$\frac{1}{2}mv^2 = eV \qquad \text{or} \qquad v = \sqrt{\frac{2eV}{m}} \tag{27.14}$$

The electrons pass between the plates P and P' and strike the screen at the end of the tube, which is coated with a material that fluoresces (glows) at the point of impact. The electrons pass straight through the plates when Eq. (27.13) is satisfied; combining this with Eq. (27.14), we get

$$\frac{E}{B} = \sqrt{\frac{2eV}{m}} \qquad \text{so} \qquad \frac{e}{m} = \frac{E^2}{2VB^2} \tag{27.15}$$

All the quantities on the right side can be measured, so the ratio e/m of charge to mass can be determined. It is *not* possible to measure e or m separately by this method, only their ratio.

The most significant aspect of Thomson's e/m measurements was that he found a *single value* for this quantity. It did not depend on the cathode material, the residual gas in the tube, or anything else about the experiment. This independence showed that the particles in the beam, which we now call electrons, are a common constituent of all matter. Thus Thomson is credited with the first discovery of a subatomic particle, the electron.

The most precise value of e/m available as of this writing is

$$e/m = 1.758820088(39) \times 10^{11}\ \text{C/kg}$$

In this expression, (39) indicates the likely uncertainty in the last two digits, 88.

Fifteen years after Thomson's experiments, the American physicist Robert Millikan succeeded in measuring the charge of the electron precisely (see Problem 23.81). This value, together with the value of e/m, enables us to determine the *mass* of the electron. The most precise value available at present is

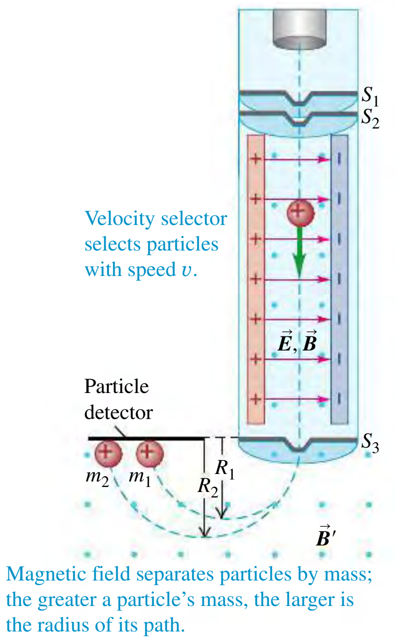
$$m = 9.10938291(40) \times 10^{-31}\ \text{kg}$$

Mass Spectrometers

Techniques similar to Thomson's e/m experiment can be used to measure masses of ions and thus measure atomic and molecular masses. In 1919, Francis Aston (1877–1945), a student of Thomson's, built the first of a family of instruments called **mass spectrometers.** A variation built by Bainbridge is shown in **Fig. 27.24**. Positive ions from a source pass through the slits S_1 and S_2, forming a narrow beam. Then the ions pass through a velocity selector with crossed $\vec{E}$ and $\vec{B}$ fields, as we have described, to block all ions except those with speeds v equal to E/B. Finally, the ions pass into a region with a magnetic field $\vec{B}'$ perpendicular to the figure, where they move in circular arcs with radius R determined by Eq. (27.11): $R = mv/qB'$. Ions with different masses strike the detector at different points, and the values of R can be measured. We assume that each ion has lost one electron, so the net charge of each ion is just $+e$. With everything known in this equation except m, we can compute the mass m of the ion.

One of the earliest results from this work was the discovery that neon has two species of atoms, with atomic masses 20 and 22 g/mol. We now call these species **isotopes** of the element. Later experiments have shown that many elements have several isotopes—atoms with identical chemical behaviors but different masses due to differing numbers of neutrons in their nuclei. This is just one of the many applications of mass spectrometers in chemistry and physics.

27.24 Bainbridge's mass spectrometer utilizes a velocity selector to produce particles with uniform speed v. In the region of magnetic field B', particles with greater mass $(m_2 > m_1)$ travel in paths with larger radius $(R_2 > R_1)$.



SOLUTION

EXAMPLE 27.5 AN e/m DEMONSTRATION EXPERIMENT

You set out to reproduce Thomson's e/m experiment with an accelerating potential of 150 V and a deflecting electric field of magnitude 6.0×10^6 N/C. (a) How fast do the electrons move? (b) What magnetic-field magnitude will yield zero beam deflection? (c) With this magnetic field, how will the electron beam behave if you increase the accelerating potential above 150 V?

SOLUTION

IDENTIFY and SET UP: This is the situation shown in Fig. 27.23. We use Eq. (27.14) to determine the electron speed and Eq. (27.13) to determine the required magnetic field B.

EXECUTE: (a) From Eq. (27.14), the electron speed v is

$$v = \sqrt{2(e/m)V} = \sqrt{2(1.76 \times 10^{11}\ \text{C/kg})(150\ \text{V})}$$
$$= 7.27 \times 10^6\ \text{m/s} = 0.024c$$

(b) From Eq. (27.13), the required field strength is

$$B = \frac{E}{v} = \frac{6.0 \times 10^6\ \text{N/C}}{7.27 \times 10^6\ \text{m/s}} = 0.83\ \text{T}$$

(c) Increasing the accelerating potential V increases the electron speed v. In Fig. 27.23 this doesn't change the upward electric force eE, but it increases the downward magnetic force evB. Therefore the electron beam will turn *downward* and will hit the end of the tube below the undeflected position.

EVALUATE: The required magnetic field is relatively large because the electrons are moving fairly rapidly (2.4% of the speed of light). If the maximum available magnetic field is less than 0.83 T, the electric field strength E would have to be reduced to maintain the desired ratio E/B in Eq. (27.15).

SOLUTION

EXAMPLE 27.6 FINDING LEAKS IN A VACUUM SYSTEM

There is almost no helium in ordinary air, so helium sprayed near a leak in a vacuum system will quickly show up in the output of a vacuum pump connected to such a system. You are designing a leak detector that uses a mass spectrometer to detect He^+ ions (charge $+e = +1.60 \times 10^{-19}$ C, mass 6.65×10^{-27} kg). Ions emerge from the velocity selector with a speed of 1.00×10^5 m/s. They are curved in a semicircular path by a magnetic field B' and are detected at a distance of 10.16 cm from the slit S_3 in Fig. 27.24. Calculate the magnitude of the magnetic field B'.

SOLUTION

IDENTIFY and SET UP: After it passes through the slit, the ion follows a circular path as described in Section 27.4 (see Fig. 27.17). We solve Eq. (27.11) for B'.

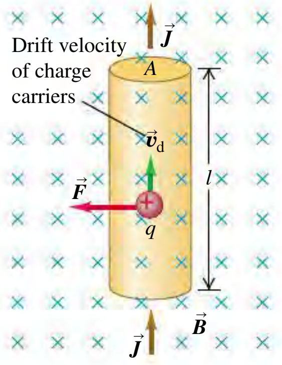
EXECUTE: The distance given is the *diameter* of the semicircular path shown in Fig. 27.24, so the radius is $R = \frac{1}{2}(10.16 \times 10^{-2}\ \text{m})$. From Eq. (27.11), $R = mv/qB'$, we get

$$B' = \frac{mv}{qR} = \frac{(6.65 \times 10^{-27}\ \text{kg})(1.00 \times 10^5\ \text{m/s})}{(1.60 \times 10^{-19}\ \text{C})(5.08 \times 10^{-2}\ \text{m})}$$
$$= 0.0818\ \text{T}$$

EVALUATE: Helium leak detectors are widely used with high-vacuum systems. Our result shows that only a small magnetic field is required, so leak detectors can be relatively compact.

TEST YOUR UNDERSTANDING OF SECTION 27.5 In Example 27.6 He^+ ions with charge $+e$ move at 1.00×10^5 m/s in a straight line through a velocity selector. Suppose the He^+ ions were replaced with He^{2+} ions, in which both electrons have been removed from the helium atom and the ion charge is $+2e$. At what speed must the He^{2+} ions travel through the same velocity selector in order to move in a straight line? (i) 4.00×10^5 m/s; (ii) 2.00×10^5 m/s; (iii) 1.00×10^5 m/s; (iv) 0.50×10^5 m/s; (v) 0.25×10^5 m/s. ❙

27.25 Forces on a moving positive charge in a current-carrying conductor.



27.6 MAGNETIC FORCE ON A CURRENT-CARRYING CONDUCTOR

What makes an electric motor work? Within the motor are conductors that carry currents (that is, whose charges are in motion), as well as magnets that exert forces on the moving charges. Hence there is a magnetic force on each current-carrying conductor, and these forces make the motor turn. The d'Arsonval galvanometer (Section 26.3) also uses magnetic forces on conductors.

We can compute the force on a current-carrying conductor starting with the magnetic force $\vec{F} = q\vec{v} \times \vec{B}$ on a single moving charge. **Figure 27.25** shows a straight segment of a conducting wire, with length l and cross-sectional area A;

the current is from bottom to top. The wire is in a uniform magnetic field $\vec{B}$, perpendicular to the plane of the diagram and directed *into* the plane. Let's assume first that the moving charges are positive. Later we'll see what happens when they are negative.

The drift velocity $\vec{v}_d$ is upward, perpendicular to $\vec{B}$. The average force on each charge is $\vec{F} = q\vec{v}_d \times \vec{B}$, directed to the left as shown in the figure; since $\vec{v}_d$ and $\vec{B}$ are perpendicular, the magnitude of the force is $F = qv_dB$.

We can derive an expression for the *total* force on all the moving charges in a length l of conductor with cross-sectional area A by using the same language we used in Eqs. (25.2) and (25.3) of Section 25.1. The number of charges per unit volume, or charge concentration, is n; a segment of conductor with length l has volume Al and contains a number of charges equal to nAl. The total force $\vec{F}$ on *all* the moving charges in this segment has magnitude

$$F = (nAl)(qv_dB) = (nqv_dA)(lB) \tag{27.16}$$

From Eq. (25.3) the current density is $J = nqv_d$. The product JA is the total current I, so we can rewrite Eq. (27.16) as

$$F = IlB \tag{27.17}$$

If the $\vec{B}$ field is not perpendicular to the wire but makes an angle ϕ with it, as in **Fig. 27.26**, we handle the situation the same way we did in Section 27.2 for a single charge. Only the component of $\vec{B}$ perpendicular to the wire (and to the drift velocities of the charges) exerts a force; this component is $B_\perp = B\sin\phi$. The magnetic force on the wire segment is then

$$F = IlB_\perp = IlB\sin\phi \tag{27.18}$$

The force is always perpendicular to both the conductor and the field, with the direction determined by the same right-hand rule we used for a moving positive charge (Fig. 27.26). Hence this force can be expressed as a vector product, like the force on a single moving charge. We represent the segment of wire with a vector $\vec{l}$ along the wire in the direction of the current; then the force $\vec{F}$ on this segment is

Magnetic force on a straight wire segment

$$\vec{F} = I\vec{l} \times \vec{B} \tag{27.19}$$

Current; Magnetic field; Vector length of segment (points in current direction)

Figure 27.27 illustrates the directions of $\vec{B}$, $\vec{l}$, and $\vec{F}$ for several cases.

If the conductor is not straight, we can divide it into infinitesimal segments $d\vec{l}$. The force $d\vec{F}$ on each segment is

Magnetic force on an infinitesimal wire segment

$$d\vec{F} = I\,d\vec{l} \times \vec{B} \tag{27.20}$$

Current; Magnetic field; Vector length of segment (points in current direction)

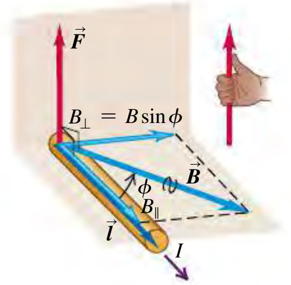
Then we can integrate this expression along the wire to find the total force on a conductor of any shape. The integral is a *line integral*, the same mathematical operation we have used to define work (Section 6.3) and electric potential (Section 23.2).

CAUTION **Current is not a vector** Recall from Section 25.1 that the current I is not a vector. The direction of current flow is described by $d\vec{l}$, not I. If the conductor is curved, I is the same at all points along its length, but $d\vec{l}$ changes direction—it is always tangent to the conductor.

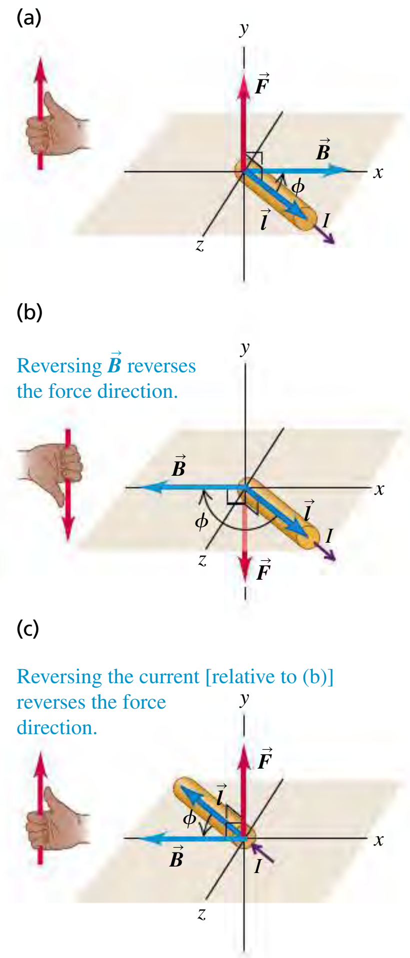
DEMO

27.26 A straight wire segment of length $\vec{l}$ carries a current I in the direction of $\vec{l}$. The magnetic force on this segment is perpendicular to both $\vec{l}$ and the magnetic field $\vec{B}$.

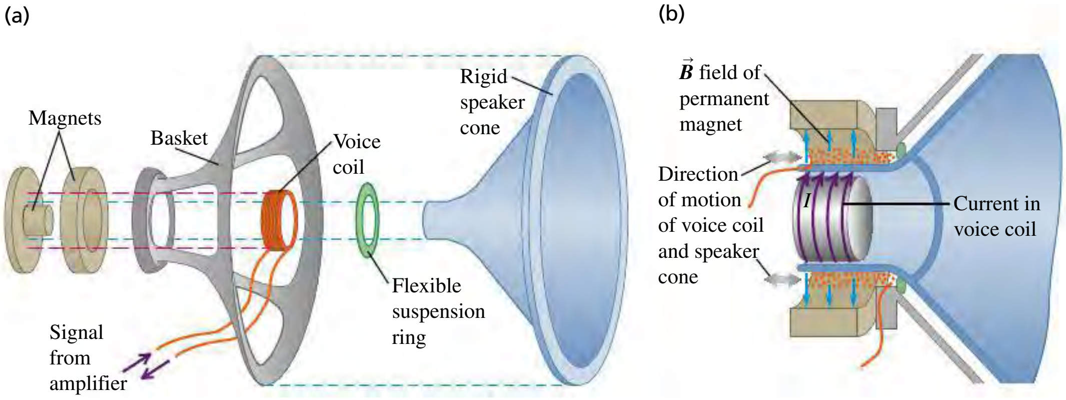
Force $\vec{F}$ on a straight wire carrying a positive current and oriented at an angle ϕ to a magnetic field $\vec{B}$:
- Magnitude is $F = IlB_\perp = IlB\sin\phi$.
- Direction of $\vec{F}$ is given by the right-hand rule.



27.27 Magnetic field $\vec{B}$, length $\vec{l}$, and force $\vec{F}$ vectors for a straight wire carrying a current I.



27.28 (a) Components of a loudspeaker. (b) The permanent magnet creates a magnetic field that exerts forces on the current in the voice coil; for a current I in the direction shown, the force is to the right. If the electric current in the voice coil oscillates, the speaker cone attached to the voice coil oscillates at the same frequency.



Finally, what happens when the moving charges are negative, such as electrons in a metal? Then in Fig. 27.25 an upward current corresponds to a downward drift velocity. But because q is now negative, the direction of the force $\vec{F}$ is the same as before. Thus Eqs. (27.17) through (27.20) are valid for *both* positive and negative charges and even when *both* signs of charge are present at once. This happens in some semiconductor materials and in ionic solutions.

A common application of the magnetic forces on a current-carrying wire is found in loudspeakers (**Fig. 27.28**). The radial magnetic field created by the permanent magnet exerts a force on the voice coil that is proportional to the current in the coil; the direction of the force is either to the left or to the right, depending on the direction of the current. The signal from the amplifier causes the current to oscillate in direction and magnitude. The coil and the speaker cone to which it is attached respond by oscillating with an amplitude proportional to the amplitude of the current in the coil. Turning up the volume knob on the amplifier increases the current amplitude and hence the amplitudes of the cone's oscillation and of the sound wave produced by the moving cone.

SOLUTION

EXAMPLE 27.7 MAGNETIC FORCE ON A STRAIGHT CONDUCTOR

A straight horizontal copper rod carries a current of 50.0 A from west to east in a region between the poles of a large electromagnet. In this region there is a horizontal magnetic field toward the northeast (that is, 45° north of east) with magnitude 1.20 T. (a) Find the magnitude and direction of the force on a 1.00-m section of rod. (b) While keeping the rod horizontal, how should it be oriented to maximize the magnitude of the force? What is the force magnitude in this case?

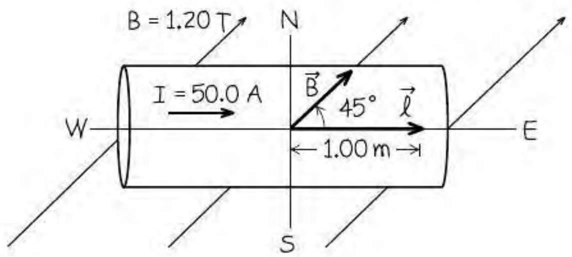
SOLUTION

IDENTIFY and SET UP: Figure 27.29 shows the situation. This is a straight wire segment in a uniform magnetic field, as in Fig. 27.26. Our target variables are the force $\vec{F}$ on the segment and the angle ϕ for which the force magnitude F is greatest. We find the magnitude of the magnetic force from Eq. (27.18) and the direction from the right-hand rule.

EXECUTE: (a) The angle ϕ between the directions of current and field is 45°. From Eq. (27.18) we obtain

$$F = IlB\sin\phi = (50.0\text{ A})(1.00\text{ m})(1.20\text{ T})(\sin 45°) = 42.4\text{ N}$$

27.29 Our sketch of the copper rod as seen from overhead.



The direction of the force is perpendicular to the plane of the current and the field, both of which lie in the horizontal plane. Thus the force must be vertical; the right-hand rule shows that it is vertically *upward* (out of the plane of the figure).

(b) From $F = IlB\sin\phi$, F is maximum for $\phi = 90°$, so that $\vec{l}$ and $\vec{B}$ are perpendicular. To keep $\vec{F} = I\vec{l} \times \vec{B}$ upward, we rotate the rod *clockwise* by 45° from its orientation in Fig. 27.29, so that the current runs toward the southeast. Then $F = IlB = (50.0\text{ A})(1.00\text{ m})(1.20\text{ T}) = 60.0\text{ N}$.

EVALUATE: We check the result in part (a) by using Eq. (27.19) to calculate the force vector. If we use a coordinate system with the x-axis pointing east, the y-axis north, and the z-axis upward, we have $\vec{l} = (1.00\text{ m})\hat{\imath}$, $\vec{B} = (1.20\text{ T})[(\cos 45°)\hat{\imath} + (\sin 45°)\hat{\jmath}]$, and

$$\begin{aligned}\vec{F} &= I\vec{l} \times \vec{B}\\ &= (50.0\text{ A})(1.00\text{ m})\hat{\imath} \times (1.20\text{ T})[(\cos 45°)\hat{\imath} + (\sin 45°)\hat{\jmath}]\\ &= (42.4\text{ N})\hat{k}\end{aligned}$$

Note that the maximum upward force of 60.0 N can hold the conductor in midair against the force of gravity—that is, *magnetically levitate* the conductor—if its weight is 60.0 N and its mass is $m = w/g = (60.0\text{ N})/(9.8\text{ m/s}^2) = 6.12\text{ kg}$. Magnetic levitation is used in some high-speed trains to suspend the train over the tracks. Eliminating rolling friction in this way allows the train to achieve speeds of over 400 km/h.

EXAMPLE 27.8 MAGNETIC FORCE ON A CURVED CONDUCTOR

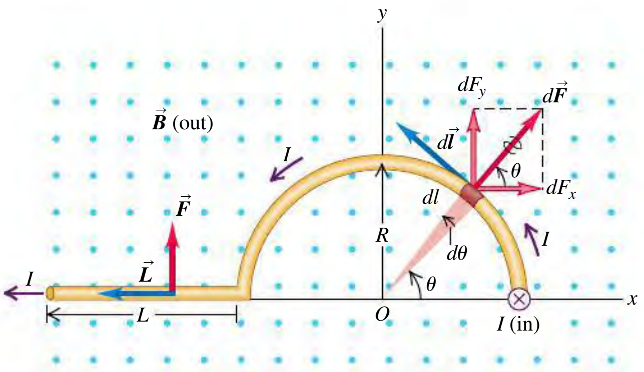
SOLUTION

In **Fig. 27.30** the magnetic field $\vec{B}$ is uniform and perpendicular to the plane of the figure, pointing out of the page. The conductor, carrying current I to the left, has three segments: (1) a straight segment with length L perpendicular to the plane of the figure, (2) a semicircle with radius R, and (3) another straight segment with length L parallel to the x-axis. Find the total magnetic force on this conductor.

SOLUTION

IDENTIFY and SET UP: The magnetic field $\vec{B} = B\hat{k}$ is uniform, so we find the forces $\vec{F}_1$ and $\vec{F}_3$ on the straight segments (1) and (3) from Eq. (27.19). We divide the curved segment (2) into infinitesimal straight segments and find the corresponding force $d\vec{F}_2$ on each straight segment from Eq. (27.20). We then integrate to find $\vec{F}_2$. The total magnetic force on the conductor is then $\vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3$.

27.30 What is the total magnetic force on the conductor?



EXECUTE: For segment (1), $\vec{L} = -L\hat{k}$. Hence from Eq. (27.19), $\vec{F}_1 = I\vec{L} \times \vec{B} = \mathbf{0}$. For segment (3), $\vec{L} = -L\hat{\imath}$, so $\vec{F}_3 = I\vec{L} \times \vec{B} = I(-L\hat{\imath}) \times (B\hat{k}) = ILB\hat{\jmath}$.

For the curved segment (2), Fig. 27.30 shows a segment $d\vec{l}$ with length $dl = R\,d\theta$, at angle θ. The right-hand rule shows that the direction of $d\vec{l} \times \vec{B}$ is radially outward from the center; make sure you can verify this. Because $d\vec{l}$ and $\vec{B}$ are perpendicular, the magnitude dF_2 of the force on the segment $d\vec{l}$ is $dF_2 = I\,dl\,B = I(R\,d\theta)B$. The components of the force on this segment are

$$dF_{2x} = IR\,d\theta\,B\cos\theta \qquad dF_{2y} = IR\,d\theta\,B\sin\theta$$

To find the components of the total force, we integrate these expressions with respect to θ from $\theta = 0$ to $\theta = \pi$ to take in the whole semicircle. The results are

$$F_{2x} = IRB\int_0^{\pi}\cos\theta\,d\theta = 0$$

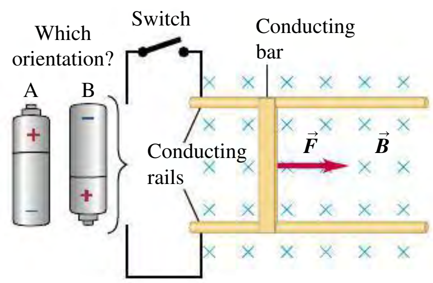
$$F_{2y} = IRB\int_0^{\pi}\sin\theta\,d\theta = 2IRB$$

Hence $\vec{F}_2 = 2IRB\hat{\jmath}$. Finally, adding the forces on all three segments, we find that the total force is in the positive y-direction:

$$\vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 = 0 + 2IRB\hat{\jmath} + ILB\hat{\jmath} = IB(2R + L)\hat{\jmath}$$

EVALUATE: We could have predicted from symmetry that the x-component of $\vec{F}_2$ would be zero: On the right half of the semicircle the x-component of the force is positive (to the right) and on the left half it is negative (to the left); the positive and negative contributions to the integral cancel. The result is that $\vec{F}_2$ is the force that would be exerted if we replaced the semicircle with a *straight* segment of length $2R$ along the x-axis. Do you see why?

TEST YOUR UNDERSTANDING OF SECTION 27.6 The accompanying figure shows a top view of two conducting rails on which a conducting bar can slide. A uniform magnetic field is directed perpendicular to the plane of the figure as shown. A battery is to be connected to the two rails so that when the switch is closed, current will flow through the bar and cause a magnetic force to push the bar to the right. In which orientation, A or B, should the battery be placed in the circuit?



27.7 FORCE AND TORQUE ON A CURRENT LOOP

Current-carrying conductors usually form closed loops, so it is worthwhile to use the results of Section 27.6 to find the *total* magnetic force and torque on a conductor in the form of a loop. Many practical devices make use of the magnetic force or torque on a conducting loop, including loudspeakers (see Fig. 27.28) and galvanometers (see Section 26.3). Hence the results of this section are of substantial practical importance. These results will also help us understand the behavior of bar magnets described in Section 27.1.

As an example, let's look at a rectangular current loop in a uniform magnetic field. We can represent the loop as a series of straight line segments. We will find that the total *force* on the loop is zero but that there can be a net *torque* acting on the loop, with some interesting properties.

Figure 27.31a shows a rectangular loop of wire with side lengths a and b. A line perpendicular to the plane of the loop (i.e., a *normal* to the plane) makes an angle ϕ with the direction of the magnetic field $\vec{B}$, and the loop carries a current I. The wires leading the current into and out of the loop and the source of emf are omitted to keep the diagram simple.

The force $\vec{F}$ on the right side of the loop (length a) is to the right, in the $+x$-direction as shown. On this side, $\vec{B}$ is perpendicular to the current direction, and the force on this side has magnitude

$$F = IaB \tag{27.21}$$

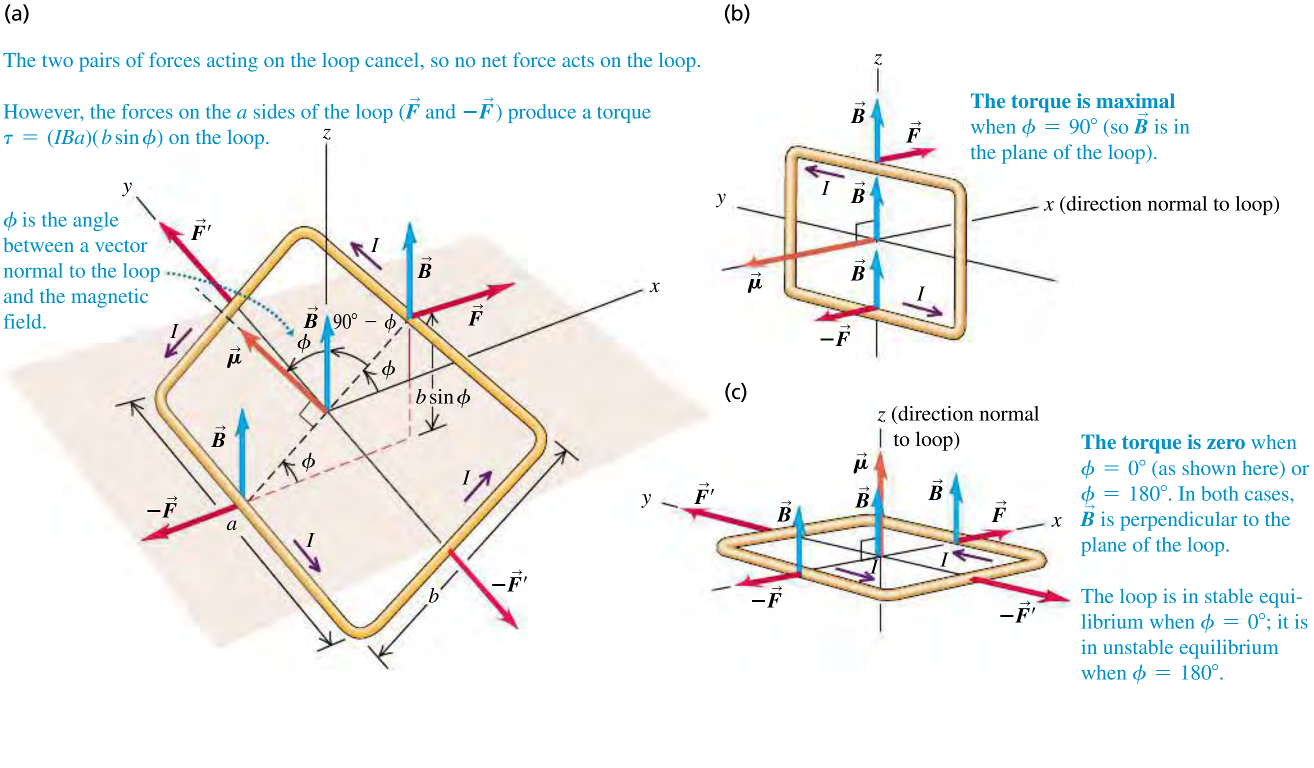
A force $-\vec{F}$ with the same magnitude but opposite direction acts on the opposite side of the loop, as shown in the figure.

The sides with length b make an angle $(90° - \phi)$ with the direction of $\vec{B}$. The forces on these sides are the vectors $\vec{F}'$ and $-\vec{F}'$; their magnitude F' is given by

$$F' = IbB\sin(90° - \phi) = IbB\cos\phi$$

The lines of action of both forces lie along the y-axis.

27.31 Finding the torque on a current-carrying loop in a uniform magnetic field.



The *total* force on the loop is zero because the forces on opposite sides cancel out in pairs.

The net force on a current loop in a uniform magnetic field is zero. However, the net torque is not in general equal to zero.

(You may find it helpful to review the discussion of torque in Section 10.1.) The two forces $\vec{F}'$ and $-\vec{F}'$ in Fig. 27.31a lie along the same line and so give rise to zero net torque with respect to any point. The two forces $\vec{F}$ and $-\vec{F}$ lie along different lines, and each gives rise to a torque about the y-axis. According to the right-hand rule for determining the direction of torques, the vector torques due to $\vec{F}$ and $-\vec{F}$ are both in the $+y$-direction; hence the net vector torque $\vec{\tau}$ is in the $+y$-direction as well. The moment arm for each of these forces (equal to the perpendicular distance from the rotation axis to the line of action of the force) is $(b/2)\sin\phi$, so the torque due to each force has magnitude $F(b/2)\sin\phi$. If we use Eq. (27.21) for F, the magnitude of the net torque is

$$\tau = 2F(b/2)\sin\phi = (IBa)(b\sin\phi) \qquad (27.22)$$

The torque is greatest when $\phi = 90°$, $\vec{B}$ is in the plane of the loop, and the normal to this plane is perpendicular to $\vec{B}$ (Fig. 27.31b). The torque is zero when ϕ is 0° or 180° and the normal to the loop is parallel or antiparallel to the field (Fig. 27.31c). The value $\phi = 0°$ is a stable equilibrium position because the torque is zero there, and when the loop is rotated slightly from this position, the resulting torque tends to rotate it back toward $\phi = 0°$. The position $\phi = 180°$ is an *unstable* equilibrium position; if displaced slightly from this position, the loop tends to move farther away from $\phi = 180°$. Figure 27.31 shows rotation about the y-axis, but because the net force on the loop is zero, Eq. (27.22) for the torque is valid for *any* choice of axis. The torque always tends to rotate the loop in the direction of *decreasing* ϕ—that is, toward the stable equilibrium position $\phi = 0°$.

The area A of the loop is equal to ab, so we can rewrite Eq. (27.22) as

Current · Magnetic-field magnitude

Magnitude of magnetic torque on a current loop

$$\tau = IBA\sin\phi \qquad (27.23)$$

Area of loop · Angle between normal to loop plane and field direction

The product IA is called the **magnetic dipole moment** or **magnetic moment** of the loop, for which we use the symbol μ (the Greek letter mu):

$$\mu = IA \qquad (27.24)$$

It is analogous to the electric dipole moment introduced in Section 21.7. In terms of μ, the magnitude of the torque on a current loop is

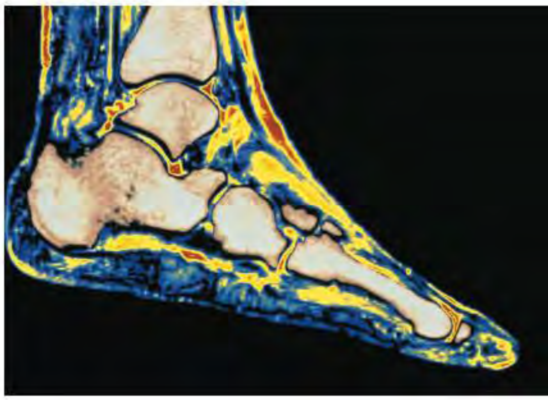
$$\tau = \mu B\sin\phi \qquad (27.25)$$

where ϕ is the angle between the normal to the loop (the direction of the vector area $\vec{A}$) and $\vec{B}$. A current loop, or any other body that experiences a magnetic torque given by Eq. (27.25), is also called a **magnetic dipole.**

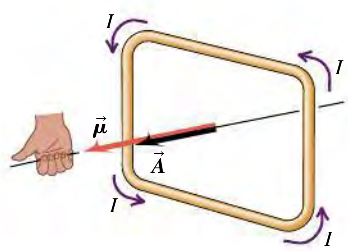
Magnetic Torque: Vector Form

We can also define a vector magnetic moment $\vec{\mu}$ with magnitude IA: This is shown in Fig. 27.31. The direction of $\vec{\mu}$ is defined to be perpendicular to the plane of the loop, with a sense determined by a right-hand rule, as shown in **Fig. 27.32**. Wrap the fingers of your right hand around the perimeter of the loop in the direction of the current. Then extend your thumb so that it is perpendicular to the plane of the loop; its direction is the direction of $\vec{\mu}$ (and of the vector area $\vec{A}$ of the loop). The torque is greatest when $\vec{\mu}$ and $\vec{B}$ are perpendicular and is zero when they are parallel or antiparallel. In the stable equilibrium position, $\vec{\mu}$ and $\vec{B}$ are parallel.

BIO Application Magnetic Resonance Imaging In magnetic resonance imaging (MRI), a patient is placed in a strong magnetic field. Each hydrogen nucleus in the patient acts like a miniature current loop with a magnetic dipole moment that tends to align with the applied field. Radio waves of just the right frequency then flip these magnetic moments out of alignment. The extent to which the radio waves are absorbed is proportional to the amount of hydrogen present. This makes it possible to image details in hydrogen-rich soft tissue that cannot be seen in x-ray images. (X rays are superior to MRI for imaging bone, which is hydrogen deficient.)



27.32 The right-hand rule determines the direction of the magnetic moment of a current-carrying loop. This is also the direction of the loop's area vector $\vec{A}$; $\vec{\mu} = I\vec{A}$ is a vector equation.



Finally, we can express this interaction in terms of the torque vector $\vec{\boldsymbol{\tau}}$, which we used for *electric*-dipole interactions in Section 21.7. From Eq. (27.25) the magnitude of $\vec{\boldsymbol{\tau}}$ is equal to the magnitude of $\vec{\boldsymbol{\mu}} \times \vec{\boldsymbol{B}}$, and reference to Fig. 27.31 shows that the directions are also the same. So we have

Vector magnetic torque on a current loop

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\boldsymbol{B}} \qquad (27.26)$$

Magnetic dipole moment: $\vec{\boldsymbol{\mu}}$; Magnetic field: $\vec{\boldsymbol{B}}$

This result is directly analogous to the result we found in Section 21.7 for the torque exerted by an *electric* field $\vec{\boldsymbol{E}}$ on an *electric* dipole with dipole moment $\vec{\boldsymbol{p}}$.

Potential Energy for a Magnetic Dipole

When a magnetic dipole changes orientation in a magnetic field, the field does work on it. In an infinitesimal angular displacement $d\phi$, the work dW is given by $\tau\, d\phi$, and there is a corresponding change in potential energy. As the above discussion suggests, the potential energy U is least when $\vec{\boldsymbol{\mu}}$ and $\vec{\boldsymbol{B}}$ are parallel and greatest when they are antiparallel. To find an expression for U as a function of orientation, note that the torque on an *electric* dipole in an *electric* field is $\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{p}} \times \vec{\boldsymbol{E}}$; we found in Section 21.7 that the corresponding potential energy is $U = -\vec{\boldsymbol{p}} \cdot \vec{\boldsymbol{E}}$. The torque on a *magnetic* dipole in a *magnetic* field is $\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\boldsymbol{B}}$, so we can conclude immediately that

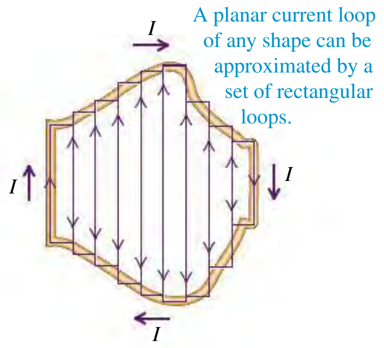
Potential energy for a magnetic dipole in a magnetic field

$$U = -\vec{\boldsymbol{\mu}} \cdot \vec{\boldsymbol{B}} = -\mu B \cos\phi \qquad (27.27)$$

Magnetic dipole moment: $\vec{\boldsymbol{\mu}}$; Magnetic field: $\vec{\boldsymbol{B}}$; Angle between $\vec{\boldsymbol{\mu}}$ and $\vec{\boldsymbol{B}}$: ϕ

With this definition, U is zero when the magnetic dipole moment is perpendicular to the magnetic field ($\phi = 90°$); then $U = -\vec{\boldsymbol{\mu}} \cdot \vec{\mathbf{B}} = -\mu B \cos 90° = 0$.

27.33 The collection of rectangles exactly matches the irregular plane loop in the limit as the number of rectangles approaches infinity and the width of each rectangle approaches zero.



Magnetic Torque: Loops and Coils

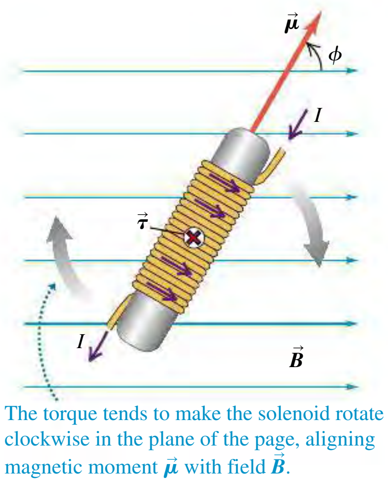
Although we have derived Eqs. (27.21) through (27.27) for a rectangular current loop, all these relationships are valid for a plane loop of any shape at all. Any planar loop may be approximated as closely as we wish by a very large number of rectangular loops, as shown in **Fig. 27.33**. If these loops all carry equal currents in the same clockwise sense, then the forces and torques on the sides of two loops adjacent to each other cancel, and the only forces and torques that do not cancel are due to currents around the boundary. Thus all the above relationships are valid for a plane current loop of any shape, with the magnetic moment $\vec{\boldsymbol{\mu}} = I\vec{\boldsymbol{A}}$.

We can also generalize this whole formulation to a coil consisting of N planar loops close together; the effect is simply to multiply each force, the magnetic moment, the torque, and the potential energy by a factor of N.

An arrangement of particular interest is the **solenoid,** a helical winding of wire, such as a coil wound on a circular cylinder (**Fig. 27.34**). If the windings are closely spaced, the solenoid can be approximated by a number of circular loops lying in planes at right angles to its long axis. The total torque on a solenoid in a magnetic field is simply the sum of the torques on the individual turns. For a solenoid with N turns in a uniform field B, the magnetic moment is $\mu = NIA$ and

$$\tau = NIAB \sin\phi \qquad (27.28)$$

27.34 The torque $\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\boldsymbol{B}}$ on this solenoid in a uniform magnetic field is directed straight into the page. An actual solenoid has many more turns, wrapped closely together.



where ϕ is the angle between the axis of the solenoid and the direction of the field. The magnetic moment vector $\vec{\mu}$ is along the solenoid axis. The torque is greatest when the solenoid axis is perpendicular to the magnetic field and zero when they are parallel. The effect of this torque is to tend to rotate the solenoid into a position where its axis is parallel to the field. Solenoids are also useful as *sources* of magnetic field, as we'll discuss in Chapter 28.

The d'Arsonval galvanometer, described in Section 26.3, makes use of a magnetic torque on a coil carrying a current. As Fig. 26.14 shows, the magnetic field is not uniform but is *radial*, so the side thrusts on the coil are always perpendicular to its plane. Thus the angle ϕ in Eq. (27.28) is always 90°, and the magnetic torque is directly proportional to the current, no matter what the orientation of the coil. A restoring torque proportional to the angular displacement of the coil is provided by two hairsprings, which also serve as current leads to the coil. When current is supplied to the coil, it rotates along with its attached pointer until the restoring spring torque just balances the magnetic torque. Thus the pointer deflection is proportional to the current.

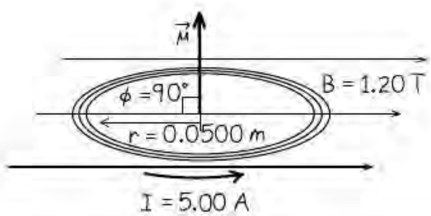
EXAMPLE 27.9 MAGNETIC TORQUE ON A CIRCULAR COIL

A circular coil 0.0500 m in radius, with 30 turns of wire, lies in a horizontal plane. It carries a counterclockwise (as viewed from above) current of 5.00 A. The coil is in a uniform 1.20-T magnetic field directed toward the right. Find the magnitudes of the magnetic moment and the torque on the coil.

SOLUTION

IDENTIFY and SET UP: This problem uses the definition of magnetic moment and the expression for the torque on a magnetic dipole in a magnetic field. **Figure 27.35** shows the situation.

27.35 Our sketch for this problem.



Equation (27.24) gives the magnitude μ of the magnetic moment of a single turn of wire; for N turns, the magnetic moment is N times greater. Equation (27.25) gives the magnitude τ of the torque.

EXECUTE: The area of the coil is $A = \pi r^2$. From Eq. (27.24), the total magnetic moment of all 30 turns is

$$\mu_{\text{total}} = NIA = 30(5.00\text{ A})\pi(0.0500\text{ m})^2 = 1.18\text{ A}\cdot\text{m}^2$$

The angle ϕ between the direction of $\vec{B}$ and the direction of $\vec{\mu}$ (which is along the normal to the plane of the coil) is 90°. From Eq. (27.25), the torque on the coil is

$$\begin{aligned}\tau &= \mu_{\text{total}}B\sin\phi = (1.18\text{ A}\cdot\text{m}^2)(1.20\text{ T})(\sin 90°)\\ &= 1.41\text{ N}\cdot\text{m}\end{aligned}$$

EVALUATE: The torque tends to rotate the right side of the coil down and the left side up, into a position where the normal to its plane is parallel to $\vec{B}$.

EXAMPLE 27.10 POTENTIAL ENERGY FOR A COIL IN A MAGNETIC FIELD

If the coil in Example 27.9 rotates from its initial orientation to one in which its magnetic moment $\vec{\mu}$ is parallel to $\vec{B}$, what is the change in potential energy?

SOLUTION

IDENTIFY and SET UP: Equation (27.27) gives the potential energy for each orientation. The initial position is as shown in Fig. 27.35, with $\phi_1 = 90°$. In the final orientation, the coil has been rotated 90° clockwise so that $\vec{\mu}$ and $\vec{B}$ are parallel, so the angle between these vectors is $\phi_2 = 0$.

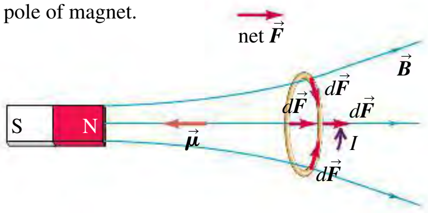
EXECUTE: From Eq. (27.27), the potential energy change is

$$\begin{aligned}\Delta U = U_2 - U_1 &= -\mu B\cos\phi_2 - (-\mu B\cos\phi_1)\\ &= -\mu B(\cos\phi_2 - \cos\phi_1)\\ &= -(1.18\text{ A}\cdot\text{m}^2)(1.20\text{ T})(\cos 0° - \cos 90°) = -1.41\text{ J}\end{aligned}$$

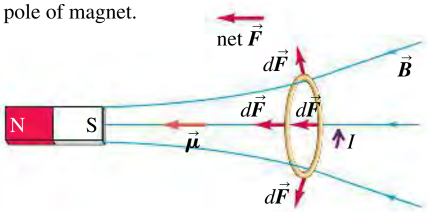
EVALUATE: The potential energy decreases because the rotation is in the direction of the magnetic torque that we found in Example 27.9.

27.36 Forces on current loops in a nonuniform $\vec{B}$ field. In each case the axis of the bar magnet is perpendicular to the plane of the loop and passes through the center of the loop.

(a) Net force on this coil is away from north pole of magnet.



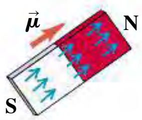
(b) Net force on same coil is toward south pole of magnet.



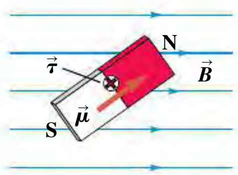
27.37 (a) An unmagnetized piece of iron. (Only a few representative atomic moments are shown.) (b) A magnetized piece of iron (bar magnet). The net magnetic moment of the bar magnet points from its south pole to its north pole. (c) A bar magnet in a magnetic field.

(a) Unmagnetized iron: magnetic moments are oriented randomly.

(b) In a bar magnet, the magnetic moments are aligned.



(c) A magnetic field creates a torque on the bar magnet that tends to align its dipole moment with the $\vec{B}$ field.



Magnetic Dipole in a Nonuniform Magnetic Field

We have seen that a current loop (that is, a magnetic dipole) experiences zero net force in a uniform magnetic field. **Figure 27.36** shows two current loops in the *nonuniform* $\vec{B}$ field of a bar magnet; in both cases the net force on the loop is *not* zero. In Fig. 27.36a the magnetic moment $\vec{\mu}$ is in the direction opposite to the field, and the force $d\vec{F} = I\,d\vec{l} \times \vec{B}$ on a segment of the loop has both a radial component and a component to the right. When these forces are summed to find the net force $\vec{F}$ on the loop, the radial components cancel so that the net force is to the right, away from the magnet. Note that in this case the force is toward the region where the field lines are farther apart and the field magnitude B is less. The polarity of the bar magnet is reversed in Fig. 27.36b, so $\vec{\mu}$ and $\vec{B}$ are parallel; now the net force on the loop is to the left, toward the region of greater field magnitude near the magnet. Later in this section we'll use these observations to explain why bar magnets can pick up unmagnetized iron objects.

Magnetic Dipoles and How Magnets Work

The behavior of a solenoid in a magnetic field (see Fig. 27.34) resembles that of a bar magnet or compass needle; if free to turn, both the solenoid and the magnet orient themselves with their axes parallel to the $\vec{B}$ field. In both cases this is due to the interaction of moving electric charges with a magnetic field; the difference is that in a bar magnet the motion of charge occurs on the microscopic scale of the atom.

Think of an electron as being like a spinning ball of charge. In this analogy the circulation of charge around the spin axis is like a current loop, and so the electron has a net magnetic moment. (This analogy, while helpful, is inexact; an electron isn't really a spinning sphere. A full explanation of the origin of an electron's magnetic moment involves quantum mechanics, which is beyond our scope here.) In an iron atom a substantial fraction of the electron magnetic moments align with each other, and the atom has a nonzero magnetic moment. (By contrast, the atoms of most elements have little or no net magnetic moment.) In an unmagnetized piece of iron there is no overall alignment of the magnetic moments of the atoms; their vector sum is zero, and the net magnetic moment is zero (**Fig. 27.37a**). But in an iron bar magnet the magnetic moments of many of the atoms are parallel, and there is a substantial net magnetic moment $\vec{\mu}$ (Fig. 27.37b). If the magnet is placed in a magnetic field $\vec{B}$, the field exerts a torque given by Eq. (27.26) that tends to align $\vec{\mu}$ with $\vec{B}$ (Fig. 27.37c). A bar magnet tends to align with a $\vec{B}$ field so that a line from the south pole to the north pole of the magnet is in the direction of $\vec{B}$; hence the real significance of a magnet's north and south poles is that they represent the head and tail, respectively, of the magnet's dipole moment $\vec{\mu}$.

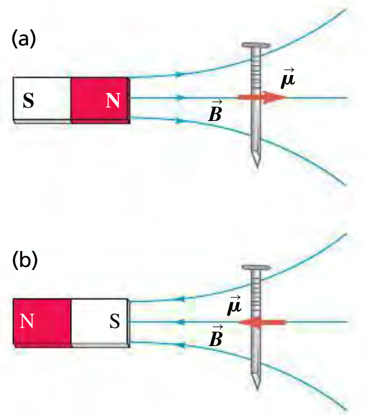
The torque experienced by a current loop in a magnetic field also explains how an unmagnetized iron object like that in Fig. 27.37a becomes magnetized. If an unmagnetized iron paper clip is placed next to a powerful magnet, the magnetic moments of the paper clip's atoms tend to align with the $\vec{B}$ field of the magnet. When the paper clip is removed, its atomic dipoles tend to remain aligned, and the paper clip has a net magnetic moment. The paper clip can be demagnetized by being dropped on the floor or heated; the added internal energy jostles and re-randomizes the atomic dipoles.

The magnetic-dipole picture of a bar magnet explains the attractive and repulsive forces between bar magnets shown in Fig. 27.1. The magnetic moment $\vec{\mu}$ of a bar magnet points from its south pole to its north pole, so the current loops in Figs. 27.36a and 27.36b are both equivalent to a magnet with its north pole on the left. Hence the situation in Fig. 27.36a is equivalent to two bar magnets with their north poles next to each other; the resultant force is repulsive, as in Fig. 27.1b. In Fig. 27.36b we again have the equivalent of two bar magnets end to end, but with the south pole of the left-hand magnet next to the north pole of the right-hand magnet. The resultant force is attractive, as in Fig. 27.1a.

Finally, we can explain how a magnet can attract an unmagnetized iron object (see Fig. 27.2). It's a two-step process. First, the atomic magnetic moments of the iron tend to align with the $\vec{B}$ field of the magnet, so the iron acquires a net magnetic dipole moment $\vec{\mu}$ parallel to the field. Second, the nonuniform field of the magnet attracts the magnetic dipole. **Figure 27.38a** shows an example. The north pole of the magnet is closer to the nail (which contains iron), and the magnetic dipole produced in the nail is equivalent to a loop with a current that circulates in a direction opposite to that shown in Fig. 27.36a. Hence the net magnetic force on the nail is opposite to the force on the loop in Fig. 27.36a, and the nail is attracted toward the magnet. Changing the polarity of the magnet, as in Fig. 27.38b, reverses the directions of both $\vec{B}$ and $\vec{\mu}$. The situation is now equivalent to that shown in Fig. 27.36b; like the loop in that figure, the nail is attracted toward the magnet. Hence a previously unmagnetized object containing iron is attracted to *either* pole of a magnet. By contrast, objects made of brass, aluminum, or wood hardly respond at all to a magnet; the atomic magnetic dipoles of these materials, if present at all, have less tendency to align with an external field.

Our discussion of how magnets and pieces of iron interact has just scratched the surface of a diverse subject known as *magnetic properties of materials*. We'll discuss these properties in more depth in Section 28.8.

27.38 A bar magnet attracts an unmagnetized iron nail in two steps. First, the $\vec{B}$ field of the bar magnet gives rise to a net magnetic moment in the nail. Second, because the field of the bar magnet is not uniform, this magnetic dipole is attracted toward the magnet. The attraction is the same whether the nail is closer to (a) the magnet's north pole or (b) the magnet's south pole.



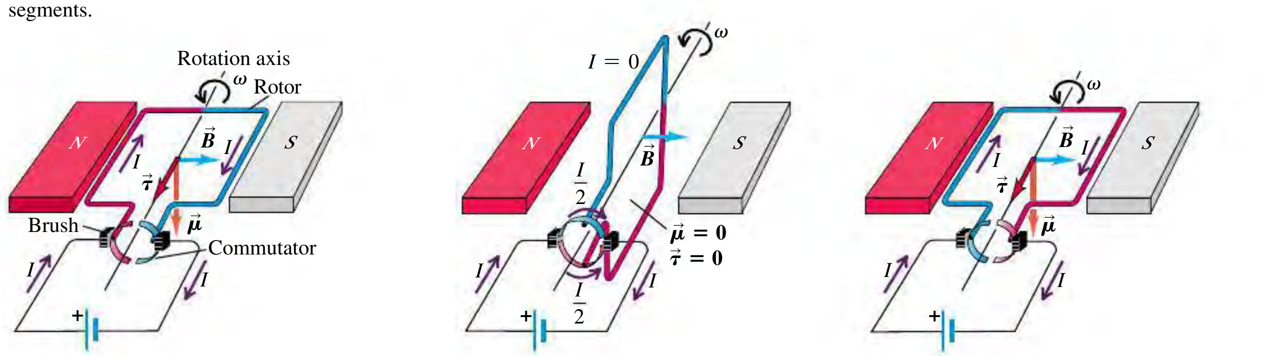
TEST YOUR UNDERSTANDING OF SECTION 27.7 Figure 27.13c depicts the magnetic field lines due to a circular current-carrying loop. (a) What is the direction of the magnetic moment of this loop? (b) Which side of the loop is equivalent to the north pole of a magnet, and which side is equivalent to the south pole? ❙

27.8 THE DIRECT-CURRENT MOTOR

Electric motors play an important role in contemporary society. In a motor a magnetic torque acts on a current-carrying conductor, and electric energy is converted to mechanical energy. As an example, let's look at a simple type of direct-current (dc) motor, shown in **Fig. 27.39**.

The moving part of the motor is the *rotor,* a length of wire formed into an open-ended loop and free to rotate about an axis. The ends of the rotor wires are attached to circular conducting segments that form a *commutator.* In Fig. 27.39a, each of the two commutator segments makes contact with one of the terminals,

27.39 Schematic diagram of a simple dc motor. The rotor is a wire loop that is free to rotate about an axis; the rotor ends are attached to the two curved conductors that form the commutator. (The rotor halves are colored red and blue for clarity.) The commutator segments are insulated from one another.



- Current flows into the red side of the rotor and out of the blue side.
- Therefore the magnetic torque causes the rotor to spin counterclockwise.

- Each brush is in contact with both commutator segments, so the current bypasses the rotor altogether.
- No magnetic torque acts on the rotor.

- The brushes are again aligned with commutator segments. This time the current flows into the blue side of the rotor and out of the red side.
- Therefore the magnetic torque again causes the rotor to spin counterclockwise.

or *brushes,* of an external circuit that includes a source of emf. This causes a current to flow into the rotor on one side, shown in red, and out of the rotor on the other side, shown in blue. Hence the rotor is a current loop with a magnetic moment $\vec{\boldsymbol{\mu}}$. The rotor lies between opposing poles of a permanent magnet, so there is a magnetic field $\vec{\boldsymbol{B}}$ that exerts a torque $\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\boldsymbol{B}}$ on the rotor. For the rotor orientation shown in Fig. 27.39a the torque causes the rotor to turn counterclockwise, in the direction that will align $\vec{\boldsymbol{\mu}}$ with $\vec{\boldsymbol{B}}$.

In Fig. 27.39b the rotor has rotated by 90° from its orientation in Fig. 27.39a. If the current through the rotor were constant, the rotor would now be in its equilibrium orientation; it would simply oscillate around this orientation. But here's where the commutator comes into play; each brush is now in contact with *both* segments of the commutator. There is no potential difference between the commutators, so at this instant no current flows through the rotor, and the magnetic moment is zero. The rotor continues to rotate counterclockwise because of its inertia, and current again flows through the rotor as in Fig. 27.39c. But now current enters on the *blue* side of the rotor and exits on the *red* side, just the opposite of the situation in Fig. 27.39a. While the direction of the current has reversed with respect to the rotor, the rotor itself has rotated 180° and the magnetic moment $\vec{\boldsymbol{\mu}}$ is in the same direction with respect to the magnetic field. Hence the magnetic torque $\vec{\boldsymbol{\tau}}$ is in the same direction in Fig. 27.39c as in Fig. 27.39a. Thanks to the commutator, the current reverses after every 180° of rotation, so the torque is always in the direction to rotate the rotor counterclockwise. When the motor has come "up to speed," the average magnetic torque is just balanced by an opposing torque due to air resistance, friction in the rotor bearings, and friction between the commutator and brushes.

The simple motor shown in Fig. 27.39 has only a single turn of wire in its rotor. In practical motors, the rotor has many turns; this increases the magnetic moment and the torque so that the motor can spin larger loads. The torque can also be increased by using a stronger magnetic field, which is why many motor designs use electromagnets instead of a permanent magnet. Another drawback of the simple design in Fig. 27.39 is that the magnitude of the torque rises and falls as the rotor spins. This can be remedied by having the rotor include several independent coils of wire oriented at different angles (**Fig. 27.40**).

27.40 This motor from a computer disk drive has 12 current-carrying coils. They interact with permanent magnets on the turntable (not shown) to make the turntable rotate. (This design is the reverse of the design in Fig. 27.39, in which the permanent magnets are stationary and the coil rotates.) Because there are multiple coils, the magnetic torque is very nearly constant and the turntable spins at a very constant rate.

Power for Electric Motors

Because a motor converts electric energy to mechanical energy or work, it requires electric energy input. If the potential difference between its terminals is V_{ab} and the current is I, then the power input is $P = V_{ab}I$. Even if the motor coils have negligible resistance, there must be a potential difference between the terminals if P is to be different from zero. This potential difference results principally from magnetic forces exerted on the currents in the conductors of the rotor as they rotate through the magnetic field. The associated electromotive force $\mathcal{E}$ is called an *induced* emf; it is also called a *back* emf because its sense is opposite to that of the current. In Chapter 29 we will study induced emfs resulting from motion of conductors in magnetic fields.

In a *series* motor the rotor is connected in series with the electromagnet that produces the magnetic field; in a *shunt* motor they are connected in parallel. In a series motor with internal resistance r, V_{ab} is greater than $\mathcal{E}$, and the difference is the potential drop Ir across the internal resistance. That is,

$$V_{ab} = \mathcal{E} + Ir \tag{27.29}$$

Because the magnetic force is proportional to velocity, $\mathcal{E}$ is *not* constant but is proportional to the speed of rotation of the rotor.

SOLUTION

EXAMPLE 27.11 A SERIES DC MOTOR

A dc motor with its rotor and field coils connected in series has an internal resistance of 2.00 Ω. When running at full load on a 120-V line, it draws a 4.00-A current. (a) What is the emf in the rotor? (b) What is the power delivered to the motor? (c) What is the rate of dissipation of energy in the internal resistance? (d) What is the mechanical power developed? (e) What is the motor's efficiency? (f) What happens if the machine being driven by the motor jams, so that the rotor suddenly stops turning?

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of power and potential drop in a series dc motor. We are given the internal resistance $r = 2.00\ \Omega$, the voltage $V_{ab} = 120$ V across the motor, and the current $I = 4.00$ A through the motor. We use Eq. (27.29) to determine the emf $\mathcal{E}$ from these quantities. The power delivered to the motor is $V_{ab}I$, the rate of energy dissipation is I^2r, and the power output by the motor is the difference between the power input and the power dissipated. The efficiency e is the ratio of mechanical power output to electric power input.

EXECUTE: (a) From Eq. (27.29), $V_{ab} = \mathcal{E} + Ir$, we have

$$120\text{ V} = \mathcal{E} + (4.00\text{ A})(2.00\ \Omega) \qquad \text{and so} \qquad \mathcal{E} = 112\text{ V}$$

(b) The power delivered to the motor from the source is

$$P_{\text{input}} = V_{ab}I = (120\text{ V})(4.00\text{ A}) = 480\text{ W}$$

(c) The power dissipated in the resistance r is

$$P_{\text{dissipated}} = I^2r = (4.00\text{ A})^2(2.00\ \Omega) = 32\text{ W}$$

(d) The mechanical power output is the electric power input minus the rate of dissipation of energy in the motor's resistance (assuming that there are no other power losses):

$$P_{\text{output}} = P_{\text{input}} - P_{\text{dissipated}} = 480\text{ W} - 32\text{ W} = 448\text{ W}$$

(e) The efficiency e is the ratio of mechanical power output to electric power input:

$$e = \frac{P_{\text{output}}}{P_{\text{input}}} = \frac{448\text{ W}}{480\text{ W}} = 0.93 = 93\%$$

(f) With the rotor stalled, the back emf $\mathcal{E}$ (which is proportional to rotor speed) goes to zero. From Eq. (27.29) the current becomes

$$I = \frac{V_{ab}}{r} = \frac{120\text{ V}}{2.00\ \Omega} = 60\text{ A}$$

and the power dissipated in the resistance r becomes

$$P_{\text{dissipated}} = I^2r = (60\text{ A})^2(2.00\ \Omega) = 7200\text{ W}$$

EVALUATE: If this massive overload doesn't blow a fuse or trip a circuit breaker, the coils will quickly melt. When the motor is first turned on, there's a momentary surge of current until the motor picks up speed. This surge causes greater-than-usual voltage drops ($V = IR$) in the power lines supplying the current. Similar effects are responsible for the momentary dimming of lights in a house when an air conditioner or dishwasher motor starts.

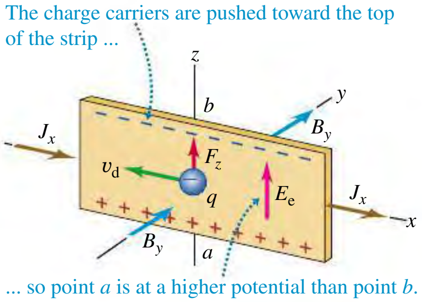
TEST YOUR UNDERSTANDING OF SECTION 27.8 In the circuit shown in Fig. 27.39, you add a switch in series with the source of emf so that the current can be turned on and off. When you close the switch and allow current to flow, will the rotor begin to turn no matter what its original orientation?

27.9 THE HALL EFFECT

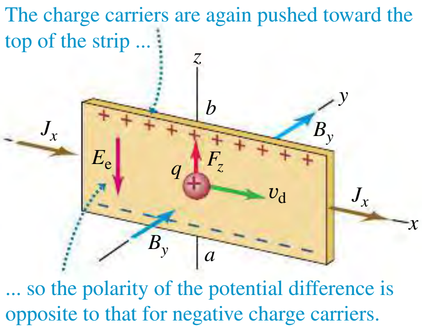
The reality of the forces acting on the moving charges in a conductor in a magnetic field is strikingly demonstrated by the *Hall effect,* an effect analogous to the transverse deflection of an electron beam in a magnetic field in vacuum. (The effect was discovered by the American physicist Edwin Hall in 1879 while he was still a graduate student.) To describe this effect, let's consider a conductor in the form of a flat strip, as shown in **Fig. 27.41** (next page). The current is in the direction of the $+x$-axis and there is a uniform magnetic field $\vec{B}$ perpendicular to the plane of the strip, in the $+y$-direction. The drift velocity of the moving charges (charge magnitude $|q|$) has magnitude v_d. Figure 27.41a shows the case of negative charges, such as electrons in a metal, and Fig. 27.41b shows positive charges. In both cases the magnetic force is upward, just as the magnetic force on a conductor is the same whether the moving charges are positive or negative. In either case a moving charge is driven toward the *upper* edge of the strip by the magnetic force $F_z = |q|v_dB$.

27.41 Forces on charge carriers in a conductor in a magnetic field.

(a) Negative charge carriers (electrons)



(b) Positive charge carriers



If the charge carriers are electrons, as in Fig. 27.41a, an excess negative charge accumulates at the upper edge of the strip, leaving an excess positive charge at its lower edge. This accumulation continues until the resulting transverse electrostatic field $\vec{E}_e$ becomes large enough to cause a force (magnitude $|q|E_e$) that is equal and opposite to the magnetic force (magnitude $|q|v_dB$). After that, there is no longer any net transverse force to deflect the moving charges. This electric field causes a transverse potential difference between opposite edges of the strip, called the *Hall voltage* or the *Hall emf.* The polarity depends on whether the moving charges are positive or negative. Experiments show that for metals the upper edge of the strip in Fig. 27.41a *does* become negatively charged, showing that the charge carriers in a metal are indeed negative electrons.

However, if the charge carriers are *positive,* as in Fig. 27.41b, then *positive* charge accumulates at the upper edge, and the potential difference is *opposite* to the situation with negative charges. Soon after the discovery of the Hall effect in 1879, it was observed that some materials, particularly some *semiconductors,* show a Hall emf opposite to that of the metals, as if their charge carriers were positively charged. We now know that these materials conduct by a process known as *hole conduction*. Within such a material there are locations, called *holes,* that would normally be occupied by an electron but are actually empty. A missing negative charge is equivalent to a positive charge. When an electron moves in one direction to fill a hole, it leaves another hole behind it. The hole migrates in the direction opposite to that of the electron.

In terms of the coordinate axes in Fig. 27.41b, the electrostatic field $\vec{E}_e$ for the positive-q case is in the $-z$-direction; its z-component E_z is negative. The magnetic field is in the $+y$-direction, and we write it as B_y. The magnetic force (in the $+z$-direction) is qv_dB_y. The current density J_x is in the $+x$-direction. In the steady state, when the forces qE_z and qv_dB_y sum to zero,

$$qE_z + qv_dB_y = 0 \qquad \text{or} \qquad E_z = -v_dB_y$$

This confirms that when q is positive, E_z is negative. From Eq. (25.4),

$$J_x = nqv_d$$

Eliminating v_d between these equations, we find

Hall effect:

$$\underbrace{n}_{\text{Concentration of mobile charge carriers}}\underbrace{q}_{\text{Charge per carrier}} = \frac{-\overbrace{J_x}^{\text{Current density}}\overbrace{B_y}^{\text{Magnetic field}}}{\underbrace{E_z}_{\text{Electrostatic field in conductor}}} \tag{27.30}$$

Note that this result (as well as the entire derivation) is valid for both positive and negative q. When q is negative, E_z is positive, and conversely.

We can measure J_x, B_y, and E_z, so we can compute the product nq. In both metals and semiconductors, q is equal in magnitude to the electron charge, so the Hall effect permits a direct measurement of n, the concentration of current-carrying charges in the material. The *sign* of the charges is determined by the polarity of the Hall emf, as we have described.

The Hall effect can also be used for a direct measurement of electron drift speed v_d in metals. As we saw in Chapter 25, these speeds are very small, often of the order of 1 mm/s or less. If we move the entire conductor in the opposite direction to the current with a speed equal to the drift speed, then the electrons are at rest with respect to the magnetic field, and the Hall emf disappears. Thus the conductor speed needed to make the Hall emf vanish is equal to the drift speed.

EXAMPLE 27.12 A HALL-EFFECT MEASUREMENT

SOLUTION

You place a strip of copper, 2.0 mm thick and 1.50 cm wide, in a uniform 0.40-T magnetic field as shown in Fig. 27.41a. When you run a 75-A current in the $+x$-direction, you find that the potential at the bottom of the slab is 0.81 μV higher than at the top. From this measurement, determine the concentration of mobile electrons in copper.

SOLUTION

IDENTIFY and SET UP: This problem describes a Hall-effect experiment. We use Eq. (27.30) to determine the mobile electron concentration n.

EXECUTE: First we find the current density J_x and the electric field E_z:

$$J_x = \frac{I}{A} = \frac{75\ \text{A}}{(2.0 \times 10^{-3}\ \text{m})(1.50 \times 10^{-2}\ \text{m})} = 2.5 \times 10^6\ \text{A/m}^2$$

$$E_z = \frac{V}{d} = \frac{0.81 \times 10^{-6}\ \text{V}}{1.5 \times 10^{-2}\ \text{m}} = 5.4 \times 10^{-5}\ \text{V/m}$$

Then, from Eq. (27.30),

$$n = \frac{-J_x B_y}{qE_z} = \frac{-(2.5 \times 10^6\ \text{A/m}^2)(0.40\ \text{T})}{(-1.60 \times 10^{-19}\ \text{C})(5.4 \times 10^{-5}\ \text{V/m})}$$
$$= 11.6 \times 10^{28}\ \text{m}^{-3}$$

EVALUATE: The actual value of n for copper is $8.5 \times 10^{28}\ \text{m}^{-3}$. The difference shows that our simple model of the Hall effect, which ignores quantum effects and electron interactions with the ions, must be used with caution. This example also shows that with good conductors, the Hall emf is very small even with large current densities. In practice, Hall-effect devices for magnetic-field measurements use semiconductor materials, for which moderate current densities give much larger Hall emfs.

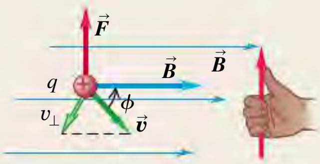
TEST YOUR UNDERSTANDING OF SECTION 27.9 A copper wire of square cross section is oriented vertically. The four sides of the wire face north, south, east, and west. There is a uniform magnetic field directed from east to west, and the wire carries current downward. Which side of the wire is at the highest electric potential? (i) North side; (ii) south side; (iii) east side; (iv) west side. ❙

CHAPTER 27 SUMMARY

SOLUTIONS TO ALL EXAMPLES

Magnetic forces: Magnetic interactions are fundamentally interactions between moving charged particles. These interactions are described by the vector magnetic field, denoted by $\vec{B}$. A particle with charge q moving with velocity $\vec{v}$ in a magnetic field $\vec{B}$ experiences a force $\vec{F}$ that is perpendicular to both $\vec{v}$ and $\vec{B}$. The SI unit of magnetic field is the tesla $(1\ \text{T} = 1\ \text{N/A} \cdot \text{m})$. (See Example 27.1.)

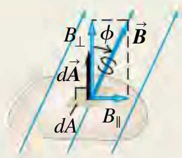
$$\vec{F} = q\vec{v} \times \vec{B} \qquad (27.2)$$



Magnetic field lines and flux: A magnetic field can be represented graphically by magnetic field lines. At each point a magnetic field line is tangent to the direction of $\vec{B}$ at that point. Where field lines are close together, the field magnitude is large, and vice versa. Magnetic flux Φ_B through an area is defined in an analogous way to electric flux. The SI unit of magnetic flux is the weber $(1\ \text{Wb} = 1\ \text{T} \cdot \text{m}^2)$. The net magnetic flux through any closed surface is zero (Gauss's law for magnetism). As a result, magnetic field lines always close on themselves. (See Example 27.2.)

$$\Phi_B = \int B \cos\phi\, dA = \int B_\perp\, dA = \int \vec{B} \cdot d\vec{A} \qquad (27.6)$$

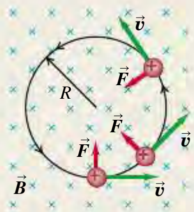
$$\oint \vec{B} \cdot d\vec{A} = 0 \quad \text{(closed surface)} \qquad (27.8)$$



Motion in a magnetic field: The magnetic force is always perpendicular to $\vec{v}$; a particle moving under the action of a magnetic field alone moves with constant speed. In a uniform field, a particle with initial velocity perpendicular to the field moves in a circle with radius R that depends on the magnetic field strength B and the particle mass m, speed v, and charge q. (See Examples 27.3 and 27.4.)

Crossed electric and magnetic fields can be used as a velocity selector. The electric and magnetic forces exactly cancel when $v = E/B$. (See Examples 27.5 and 27.6.)

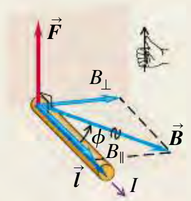
$$R = \frac{mv}{|q|B} \tag{27.11}$$



Magnetic force on a conductor: A straight segment of a conductor carrying current I in a uniform magnetic field $\vec{B}$ experiences a force $\vec{F}$ that is perpendicular to both $\vec{B}$ and the vector $\vec{l}$, which points in the direction of the current and has magnitude equal to the length of the segment. A similar relationship gives the force $d\vec{F}$ on an infinitesimal current-carrying segment $d\vec{l}$. (See Examples 27.7 and 27.8.)

$$\vec{F} = I\vec{l} \times \vec{B} \tag{27.19}$$

$$d\vec{F} = I\,d\vec{l} \times \vec{B} \tag{27.20}$$

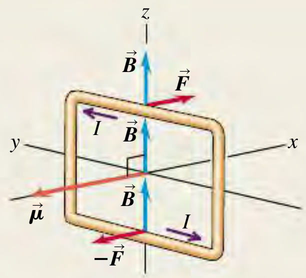


Magnetic torque: A current loop with area A and current I in a uniform magnetic field $\vec{B}$ experiences no net magnetic force, but does experience a magnetic torque of magnitude τ. The vector torque $\vec{\tau}$ can be expressed in terms of the magnetic moment $\vec{\mu} = I\vec{A}$ of the loop, as can the potential energy U of a magnetic moment in a magnetic field $\vec{B}$. The magnetic moment of a loop depends only on the current and the area; it is independent of the shape of the loop. (See Examples 27.9 and 27.10.)

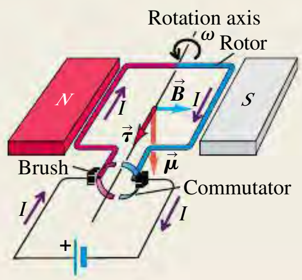
$$\tau = IBA\sin\phi \tag{27.23}$$

$$\vec{\tau} = \vec{\mu} \times \vec{B} \tag{27.26}$$

$$U = -\vec{\mu} \cdot \vec{B} = -\mu B\cos\phi \tag{27.27}$$

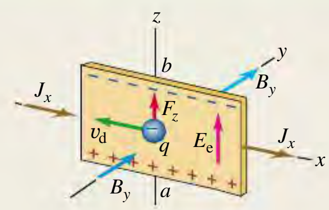


Electric motors: In a dc motor a magnetic field exerts a torque on a current in the rotor. Motion of the rotor through the magnetic field causes an induced emf called a back emf. For a series motor, in which the rotor coil is in series with coils that produce the magnetic field, the terminal voltage is the sum of the back emf and the drop Ir across the internal resistance. (See Example 27.11.)



The Hall effect: The Hall effect is a potential difference perpendicular to the direction of current in a conductor, when the conductor is placed in a magnetic field. The Hall potential is determined by the requirement that the associated electric field must just balance the magnetic force on a moving charge. Hall-effect measurements can be used to determine the sign of charge carriers and their concentration n. (See Example 27.12.)

$$nq = \frac{-J_x B_y}{E_z} \tag{27.30}$$



SOLUTION

BRIDGING PROBLEM MAGNETIC TORQUE ON A CURRENT-CARRYING RING

A circular ring with area 4.45 cm^2 is carrying a current of 12.5 A. The ring, initially at rest, is immersed in a region of uniform magnetic field given by $\vec{B} = (1.15 \times 10^{-2}\ \text{T})(12\hat{\imath} + 3\hat{\jmath} - 4\hat{k})$. The ring is positioned initially such that its magnetic moment is given by $\vec{\mu}_i = \mu(-0.800\hat{\imath} + 0.600\hat{\jmath})$, where μ is the (positive) magnitude of the magnetic moment. (a) Find the initial magnetic torque on the ring. (b) The ring (which is free to rotate around one diameter) is released and turns through an angle of 90.0°, at which point its magnetic moment is given by $\vec{\mu}_f = -\mu\hat{k}$. Determine the decrease in potential energy. (c) If the moment of inertia of the ring about a diameter is $8.50 \times 10^{-7}\ \text{kg} \cdot \text{m}^2$, determine the angular speed of the ring as it passes through the second position.

SOLUTION GUIDE

IDENTIFY and SET UP

1. The current-carrying ring acts as a magnetic dipole, so you can use the equations for a magnetic dipole in a uniform magnetic field.
2. There are no nonconservative forces acting on the ring as it rotates, so the sum of its rotational kinetic energy (discussed in Section 9.4) and the potential energy is conserved.

EXECUTE

3. Use the vector expression for the torque on a magnetic dipole to find the answer to part (a). (*Hint:* Review Section 1.10.)
4. Find the change in potential energy from the first orientation of the ring to the second orientation.
5. Use your result from step 4 to find the rotational kinetic energy of the ring when it is in the second orientation.
6. Use your result from step 5 to find the ring's angular speed when it is in the second orientation.

EVALUATE

7. If the ring were free to rotate around *any* diameter, in what direction would the magnetic moment point when the ring is in a state of stable equilibrium?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q27.1 Can a charged particle move through a magnetic field without experiencing any force? If so, how? If not, why not?

Q27.2 At any point in space, the electric field $\vec{E}$ is defined to be in the direction of the electric force on a positively charged particle at that point. Why don't we similarly define the magnetic field $\vec{B}$ to be in the direction of the magnetic force on a moving, positively charged particle?

Q27.3 Section 27.2 describes a procedure for finding the direction of the magnetic force using your right hand. If you use the same procedure, but with your left hand, will you get the correct direction for the force? Explain.

Q27.4 The magnetic force on a moving charged particle is always perpendicular to the magnetic field $\vec{B}$. Is the trajectory of a moving charged particle always perpendicular to the magnetic field lines? Explain your reasoning.

Q27.5 A charged particle is fired into a cubical region of space where there is a uniform magnetic field. Outside this region, there is no magnetic field. Is it possible that the particle will remain inside the cubical region? Why or why not?

Q27.6 If the magnetic force does no work on a charged particle, how can it have any effect on the particle's motion? Are there other examples of forces that do no work but have a significant effect on a particle's motion?

Q27.7 A charged particle moves through a region of space with constant velocity (magnitude and direction). If the external magnetic field is zero in this region, can you conclude that the external electric field in the region is also zero? Explain. (By "external" we mean fields other than those produced by the charged particle.) If the external electric field is zero in the region, can you conclude that the external magnetic field in the region is also zero?

Q27.8 How might a loop of wire carrying a current be used as a compass? Could such a compass distinguish between north and south? Why or why not?

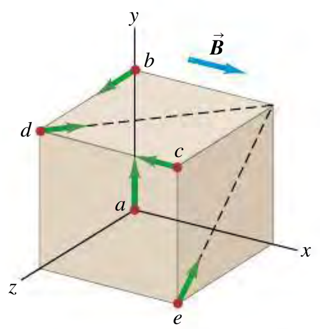
Q27.9 How could the direction of a magnetic field be determined by making only *qualitative* observations of the magnetic force on a straight wire carrying a current?

Q27.10 A loose, floppy loop of wire is carrying current I. The loop of wire is placed on a horizontal table in a uniform magnetic field $\vec{B}$ perpendicular to the plane of the table. This causes the loop of wire to expand into a circular shape while still lying on the table. In a diagram, show all possible orientations of the current I and magnetic field $\vec{B}$ that could cause this to occur. Explain your reasoning.

Q27.11 Several charges enter a uniform magnetic field directed into the page. (a) What path would a positive charge q moving with a velocity of magnitude v follow through the field? (b) What path would a positive charge q moving with a velocity of magnitude $2v$ follow through the field? (c) What path would a negative charge $-q$ moving with a velocity of magnitude v follow through the field? (d) What path would a neutral particle follow through the field?

Q27.12 Each of the lettered points at the corners of the cube in **Fig. Q27.12** represents a positive charge q moving with a velocity of magnitude v in the direction indicated. The region in the figure is in a uniform magnetic field $\vec{B}$, parallel to the x-axis and directed toward the right. Which charges experience a force due to $\vec{B}$? What is the direction of the force on each charge?

Figure **Q27.12**



Q27.13 A student claims that if lightning strikes a metal flagpole, the force exerted by the earth's magnetic field on the current in the pole can be large enough to bend it. Typical lightning currents are of the order of 10^4 to 10^5 A. Is the student's opinion justified? Explain your reasoning.

Q27.14 Could an accelerator be built in which *all* the forces on the particles, for steering and for increasing speed, are magnetic forces? Why or why not?

Q27.15 The magnetic force acting on a charged particle can never do work because at every instant the force is perpendicular to the velocity. The torque exerted by a magnetic field can do work on a current loop when the loop rotates. Explain how these seemingly contradictory statements can be reconciled.

Q27.16 When the polarity of the voltage applied to a dc motor is reversed, the direction of motion does *not* reverse. Why not? How *could* the direction of motion be reversed?

Q27.17 In a Hall-effect experiment, is it possible that *no* transverse potential difference will be observed? Under what circumstances might this happen?

Q27.18 Hall-effect voltages are much greater for relatively poor conductors (such as germanium) than for good conductors (such as copper), for comparable currents, fields, and dimensions. Why?

EXERCISES

Section 27.2 Magnetic Field

27.1 • A particle with a charge of -1.24×10^{-8} C is moving with instantaneous velocity $\vec{v} = (4.19 \times 10^4 \text{ m/s})\hat{\imath} + (-3.85 \times 10^4 \text{ m/s})\hat{\jmath}$. What is the force exerted on this particle by a magnetic field (a) $\vec{B} = (1.40 \text{ T})\hat{\imath}$ and (b) $\vec{B} = (1.40 \text{ T})\hat{k}$?

27.2 • A particle of mass 0.195 g carries a charge of -2.50×10^{-8} C. The particle is given an initial horizontal velocity that is due north and has magnitude 4.00×10^4 m/s. What are the magnitude and direction of the minimum magnetic field that will keep the particle moving in the earth's gravitational field in the same horizontal, northward direction?

27.3 • In a 1.25-T magnetic field directed vertically upward, a particle having a charge of magnitude 8.50 μC and initially moving northward at 4.75 km/s is deflected toward the east. (a) What is the sign of the charge of this particle? Make a sketch to illustrate how you found your answer. (b) Find the magnetic force on the particle.

27.4 • A particle with mass 1.81×10^{-3} kg and a charge of 1.22×10^{-8} C has, at a given instant, a velocity $\vec{v} = (3.00 \times 10^4 \text{ m/s})\hat{\jmath}$. What are the magnitude and direction of the particle's acceleration produced by a uniform magnetic field $\vec{B} = (1.63 \text{ T})\hat{\imath} + (0.980 \text{ T})\hat{\jmath}$?

27.5 • An electron experiences a magnetic force of magnitude 4.60×10^{-15} N when moving at an angle of 60.0° with respect to a magnetic field of magnitude 3.50×10^{-3} T. Find the speed of the electron.

27.6 • An electron moves at 1.40×10^6 m/s through a region in which there is a magnetic field of unspecified direction and magnitude 7.40×10^{-2} T. (a) What are the largest and smallest possible magnitudes of the acceleration of the electron due to the magnetic field? (b) If the actual acceleration of the electron is one-fourth of the largest magnitude in part (a), what is the angle between the electron velocity and the magnetic field?

27.7 •• **CP** A particle with charge 7.80 μC is moving with velocity $\vec{v} = -(3.80 \times 10^3 \text{m/s})\hat{\jmath}$. The magnetic force on the particle is measured to be $\vec{F} = +(7.60 \times 10^{-3} \text{ N})\hat{\imath} - (5.20 \times 10^{-3} \text{ N})\hat{k}$. (a) Calculate all the components of the magnetic field you can from this information. (b) Are there components of the magnetic field that are not determined by the measurement of the force? Explain. (c) Calculate the scalar product $\vec{B} \cdot \vec{F}$. What is the angle between $\vec{B}$ and $\vec{F}$?

27.8 •• **CP** A particle with charge -5.60 nC is moving in a uniform magnetic field $\vec{B} = -(1.25 \text{ T})\hat{k}$. The magnetic force on the particle is measured to be $\vec{F} = -(3.40 \times 10^{-7}\text{N})\hat{\imath} + (7.40 \times 10^{-7} \text{ N})\hat{\jmath}$. (a) Calculate all the components of the velocity of the particle that you can from this information. (b) Are there components of the velocity that are not determined by the measurement of the force? Explain. (c) Calculate the scalar product $\vec{v} \cdot \vec{F}$. What is the angle between $\vec{v}$ and $\vec{F}$?

27.9 •• A group of particles is traveling in a magnetic field of unknown magnitude and direction. You observe that a proton moving at 1.50 km/s in the $+x$-direction experiences a force of 2.25×10^{-16} N in the $+y$-direction, and an electron moving at 4.75 km/s in the $-z$-direction experiences a force of 8.50×10^{-16} N in the $+y$-direction. (a) What are the magnitude and direction of the magnetic field? (b) What are the magnitude and direction of the magnetic force on an electron moving in the $-y$-direction at 3.20 km/s?

Section 27.3 Magnetic Field Lines and Magnetic Flux

27.10 • A flat, square surface with side length 3.40 cm is in the xy-plane at $z = 0$. Calculate the magnitude of the flux through this surface produced by a magnetic field $\vec{B} = (0.200 \text{ T})\hat{\imath} + (0.300 \text{ T})\hat{\jmath} - (0.500 \text{ T})\hat{k}$.

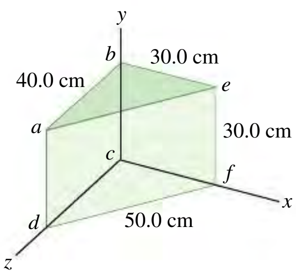
27.11 • A circular area with a radius of 6.50 cm lies in the xy-plane. What is the magnitude of the magnetic flux through this circle due to a uniform magnetic field $B = 0.230$ T (a) in the $+z$-direction; (b) at an angle of 53.1° from the $+z$-direction; (c) in the $+y$-direction?

27.12 • A horizontal rectangular surface has dimensions 2.80 cm by 3.20 cm and is in a uniform magnetic field that is directed at an angle of 30.0° above the horizontal. What must the magnitude of the magnetic field be to produce a flux of 3.10×10^{-4} Wb through the surface?

27.13 •• An open plastic soda bottle with an opening diameter of 2.5 cm is placed on a table. A uniform 1.75-T magnetic field directed upward and oriented 25° from vertical encompasses the bottle. What is the total magnetic flux through the plastic of the soda bottle?

27.14 •• The magnetic field $\vec{B}$ in a certain region is 0.128 T, and its direction is that of the $+z$-axis in **Fig. E27.14**. (a) What is the

Figure **E27.14**

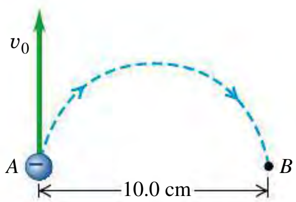


magnetic flux across the surface *abcd* in the figure? (b) What is the magnetic flux across the surface *befc*? (c) What is the magnetic flux across the surface *aefd*? (d) What is the net flux through all five surfaces that enclose the shaded volume?

Section 27.4 Motion of Charged Particles in a Magnetic Field

27.15 •• An electron at point A in **Fig. E27.15** has a speed v_0 of 1.41×10^6 m/s. Find (a) the magnitude and direction of the magnetic field that will cause the electron to follow the semicircular path from A to B, and (b) the time required for the electron to move from A to B.

Figure **E27.15**



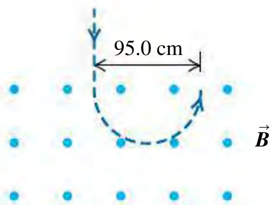
27.16 •• Repeat Exercise 27.15 for the case in which the particle is a proton rather than an electron.

27.17 • CP A 150-g ball containing 4.00×10^8 excess electrons is dropped into a 125-m vertical shaft. At the bottom of the shaft, the ball suddenly enters a uniform horizontal magnetic field that has magnitude 0.250 T and direction from east to west. If air resistance is negligibly small, find the magnitude and direction of the force that this magnetic field exerts on the ball just as it enters the field.

27.18 • An alpha particle (a He nucleus, containing two protons and two neutrons and having a mass of 6.64×10^{-27} kg) traveling horizontally at 35.6 km/s enters a uniform, vertical, 1.80-T magnetic field. (a) What is the diameter of the path followed by this alpha particle? (b) What effect does the magnetic field have on the speed of the particle? (c) What are the magnitude and direction of the acceleration of the alpha particle while it is in the magnetic field? (d) Explain why the speed of the particle does not change even though an unbalanced external force acts on it.

27.19 • In an experiment with cosmic rays, a vertical beam of particles that have charge of magnitude $3e$ and mass 12 times the proton mass enters a uniform horizontal magnetic field of 0.250 T and is bent in a semicircle of diameter 95.0 cm, as shown in **Fig. E27.19**. (a) Find the speed of the particles and the sign of their charge. (b) Is it reasonable to ignore the gravity force on the particles? (c) How does the speed of the particles as they enter the field compare to their speed as they exit the field?

Figure **E27.19**



27.20 • BIO Cyclotrons are widely used in nuclear medicine for producing short-lived radioactive isotopes. These cyclotrons typically accelerate H^- (the *hydride* ion, which has one proton and two electrons) to an energy of 5 MeV to 20 MeV. This ion has a mass very close to that of a proton because the electron mass is negligible—about $\frac{1}{2000}$ of the proton's mass. A typical magnetic field in such cyclotrons is 1.9 T. (a) What is the speed of a 5.0-MeV H^-? (b) If the H^- has energy 5.0 MeV and $B = 1.9$ T, what is the radius of this ion's circular orbit?

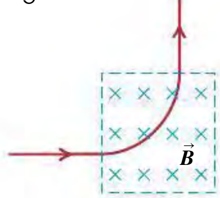
27.21 • A deuteron (the nucleus of an isotope of hydrogen) has a mass of 3.34×10^{-27} kg and a charge of $+e$. The deuteron travels in a circular path with a radius of 6.96 mm in a magnetic field with magnitude 2.50 T. (a) Find the speed of the deuteron. (b) Find the time required for it to make half a revolution. (c) Through what potential difference would the deuteron have to be accelerated to acquire this speed?

27.22 •• In a cyclotron, the orbital radius of protons with energy 300 keV is 16.0 cm. You are redesigning the cyclotron to be used instead for alpha particles with energy 300 keV. An alpha particle has charge $q = +2e$ and mass $m = 6.64 \times 10^{-27}$ kg. If the magnetic field isn't changed, what will be the orbital radius of the alpha particles?

27.23 • An electron in the beam of a cathode-ray tube is accelerated by a potential difference of 2.00 kV. Then it passes through a region of transverse magnetic field, where it moves in a circular arc with radius 0.180 m. What is the magnitude of the field?

27.24 •• A beam of protons traveling at 1.20 km/s enters a uniform magnetic field, traveling perpendicular to the field. The beam exits the magnetic field, leaving the field in a direction perpendicular to its original direction (**Fig. E27.24**). The beam travels a distance of 1.18 cm *while in the field*. What is the magnitude of the magnetic field?

Figure **E27.24**



27.25 •• A proton ($q = 1.60 \times 10^{-19}$ C, $m = 1.67 \times 10^{-27}$ kg) moves in a uniform magnetic field $\vec{B} = (0.500\ \text{T})\hat{\imath}$. At $t = 0$ the proton has velocity components $v_x = 1.50 \times 10^5$ m/s, $v_y = 0$, and $v_z = 2.00 \times 10^5$ m/s (see Example 27.4). (a) What are the magnitude and direction of the magnetic force acting on the proton? In addition to the magnetic field there is a uniform electric field in the $+x$-direction, $\vec{E} = (+2.00 \times 10^4\ \text{V/m})\hat{\imath}$. (b) Will the proton have a component of acceleration in the direction of the electric field? (c) Describe the path of the proton. Does the electric field affect the radius of the helix? Explain. (d) At $t = T/2$, where T is the period of the circular motion of the proton, what is the x-component of the displacement of the proton from its position at $t = 0$?

27.26 • A singly charged ion of ^{7}Li (an isotope of lithium) has a mass of 1.16×10^{-26} kg. It is accelerated through a potential difference of 220 V and then enters a magnetic field with magnitude 0.874 T perpendicular to the path of the ion. What is the radius of the ion's path in the magnetic field?

Section 27.5 Applications of Motion of Charged Particles

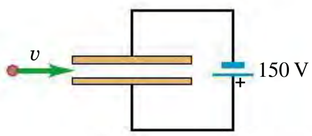
27.27 • **Crossed $\vec{E}$ and $\vec{B}$ Fields.** A particle with initial velocity $\vec{v}_0 = (5.85 \times 10^3\text{m/s})\hat{\jmath}$ enters a region of uniform electric and magnetic fields. The magnetic field in the region is $\vec{B} = -(1.35\ \text{T})\hat{k}$. Calculate the magnitude and direction of the electric field in the region if the particle is to pass through undeflected, for a particle of charge (a) +0.640 nC and (b) −0.320 nC. You can ignore the weight of the particle.

27.28 • (a) What is the speed of a beam of electrons when the simultaneous influence of an electric field of 1.56×10^4 V/m and a magnetic field of 4.62×10^{-3} T, with both fields normal to the beam and to each other, produces no deflection of the electrons? (b) In a diagram, show the relative orientation of the vectors $\vec{v}$, $\vec{E}$, and $\vec{B}$. (c) When the electric field is removed, what is the radius of the electron orbit? What is the period of the orbit?

27.29 •• A 150-V battery is connected across two parallel metal plates of area 28.5 cm^2 and separation 8.20 mm. A beam of alpha particles (charge $+2e$, mass 6.64×10^{-27} kg) is accelerated from rest through a potential difference of 1.75 kV and enters the region between the plates perpendicular to the electric field, as shown in

Fig. E27.29. What magnitude and direction of magnetic field are needed so that the alpha particles emerge undeflected from between the plates?

Figure **E27.29**



27.30 • A singly ionized (one electron removed) ^{40}K atom passes through a velocity selector consisting of uniform perpendicular electric and magnetic fields. The selector is adjusted to allow ions having a speed of 4.50 km/s to pass through undeflected when the magnetic field is 0.0250 T. The ions next enter a second uniform magnetic field (B') oriented at right angles to their velocity. ^{40}K contains 19 protons and 21 neutrons and has a mass of 6.64×10^{-26} kg. (a) What is the magnitude of the electric field in the velocity selector? (b) What must be the magnitude of B' so that the ions will be bent into a semicircle of radius 12.5 cm?

27.31 • Singly ionized (one electron removed) atoms are accelerated and then passed through a velocity selector consisting of perpendicular electric and magnetic fields. The electric field is 155 V/m and the magnetic field is 0.0315 T. The ions next enter a uniform magnetic field of magnitude 0.0175 T that is oriented perpendicular to their velocity. (a) How fast are the ions moving when they emerge from the velocity selector? (b) If the radius of the path of the ions in the second magnetic field is 17.5 cm, what is their mass?

27.32 • In the Bainbridge mass spectrometer (see Fig. 27.24), the magnetic-field magnitude in the velocity selector is 0.510 T, and ions having a speed of 1.82×10^6 m/s pass through undeflected. (a) What is the electric-field magnitude in the velocity selector? (b) If the separation of the plates is 5.20 mm, what is the potential difference between the plates?

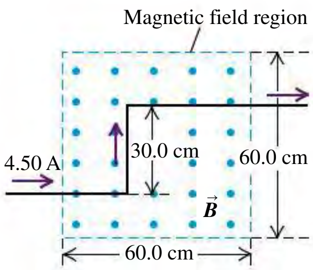
27.33 •• **BIO** **Ancient Meat Eating.** The amount of meat in prehistoric diets can be determined by measuring the ratio of the isotopes ^{15}N to ^{14}N in bone from human remains. Carnivores concentrate ^{15}N, so this ratio tells archaeologists how much meat was consumed. For a mass spectrometer that has a path radius of 12.5 cm for ^{12}C ions (mass 1.99×10^{-26} kg), find the separation of the ^{14}N (mass 2.32×10^{-26} kg) and ^{15}N (mass 2.49×10^{-26} kg) isotopes at the detector.

Section 27.6 Magnetic Force on a Current-Carrying Conductor

27.34 • A straight, 2.5-m wire carries a typical household current of 1.5 A (in one direction) at a location where the earth's magnetic field is 0.55 gauss from south to north. Find the magnitude and direction of the force that our planet's magnetic field exerts on this wire if it is oriented so that the current in it is running (a) from west to east, (b) vertically upward, (c) from north to south. (d) Is the magnetic force ever large enough to cause significant effects under normal household conditions?

27.35 •• A long wire carrying 4.50 A of current makes two 90° bends, as shown in **Fig. E27.35**. The bent part of the wire passes through a uniform 0.240-T magnetic field directed as shown in the figure and confined to a limited region of space. Find the magnitude and direction of the force that the magnetic field exerts on the wire.

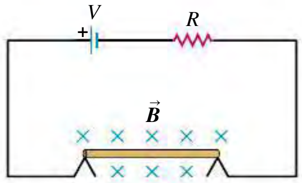
Figure **E27.35**



27.36 •• An electromagnet produces a magnetic field of 0.550 T in a cylindrical region of radius 2.50 cm between its poles. A straight wire carrying a current of 10.8 A passes through the center of this region and is perpendicular to both the axis of the cylindrical region and the magnetic field. What magnitude of force does this field exert on the wire?

27.37 • A thin, 50.0-cm-long metal bar with mass 750 g rests on, but is not attached to, two metallic supports in a uniform 0.450-T magnetic field, as shown in **Fig. E27.37**. A battery and a 25.0-Ω resistor in series are connected to the supports. (a) What is the highest voltage the battery can have without breaking the circuit at the supports? (b) The battery voltage has the maximum value calculated in part (a). If the resistor suddenly gets partially short-circuited, decreasing its resistance to 2.0 Ω, find the initial acceleration of the bar.

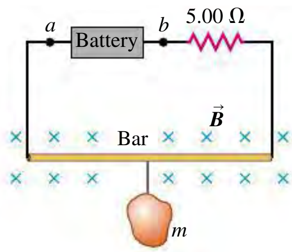
Figure **E27.37**



27.38 • A straight, vertical wire carries a current of 2.60 A downward in a region between the poles of a large superconducting electromagnet, where the magnetic field has magnitude $B = 0.588$ T and is horizontal. What are the magnitude and direction of the magnetic force on a 1.00-cm section of the wire that is in this uniform magnetic field, if the magnetic field direction is (a) east; (b) south; (c) 30.0° south of west?

27.39 • **Magnetic Balance.** The circuit shown in **Fig. E27.39** is used to make a magnetic balance to weigh objects. The mass m to be measured is hung from the center of the bar that is in a uniform magnetic field of 1.50 T, directed into the plane of the figure. The battery voltage can be adjusted to vary the current in the circuit. The horizontal bar is 60.0 cm long and is made of extremely light-weight material. It is connected to the battery by thin vertical wires that can support no appreciable tension; all the weight of the suspended mass m is supported by the magnetic force on the bar. A resistor with $R = 5.00\ \Omega$ is in series with the bar; the resistance of the rest of the circuit is much less than this. (a) Which point, a or b, should be the positive terminal of the battery? (b) If the maximum terminal voltage of the battery is 175 V, what is the greatest mass m that this instrument can measure?

Figure **E27.39**

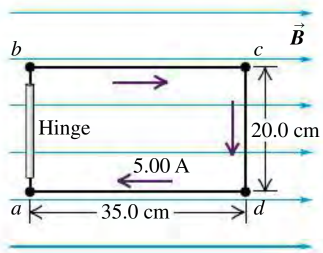


Section 27.7 Force and Torque on a Current Loop

27.40 •• The plane of a 5.0 cm × 8.0 cm rectangular loop of wire is parallel to a 0.19-T magnetic field. The loop carries a current of 6.2 A. (a) What torque acts on the loop? (b) What is the magnetic moment of the loop? (c) What is the maximum torque that can be obtained with the same total length of wire carrying the same current in this magnetic field?

27.41 • The 20.0 cm × 35.0 cm rectangular circuit shown in **Fig. E27.41** is hinged along side ab. It carries a clockwise

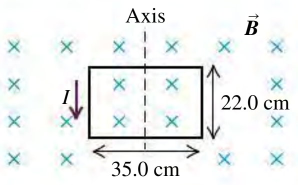
Figure **E27.41**



5.00-A current and is located in a uniform 1.20-T magnetic field oriented perpendicular to two of its sides, as shown. (a) Draw a clear diagram showing the direction of the force that the magnetic field exerts on each segment of the circuit (*ab*, *bc*, etc.). (b) Of the four forces you drew in part (a), decide which ones exert a torque about the hinge *ab*. Then calculate only those forces that exert this torque. (c) Use your results from part (b) to calculate the torque that the magnetic field exerts on the circuit about the hinge axis *ab*.

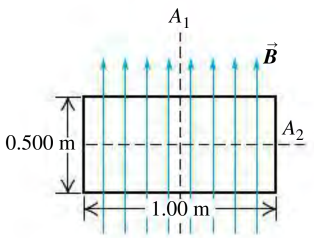
27.42 • A rectangular coil of wire, 22.0 cm by 35.0 cm and carrying a current of 1.95 A, is oriented with the plane of its loop perpendicular to a uniform 1.50-T magnetic field (**Fig. E27.42**). (a) Calculate the net force and torque that the magnetic field exerts on the coil. (b) The coil is rotated through a 30.0° angle about the axis shown, with the left side coming out of the plane of the figure and the right side going into the plane. Calculate the net force and torque that the magnetic field now exerts on the coil. (*Hint:* To visualize this three-dimensional problem, make a careful drawing of the coil as viewed along the rotation axis.)

Figure **E27.42**



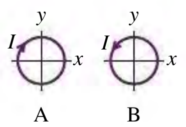
27.43 • **CP** A uniform rectangular coil of total mass 212 g and dimensions 0.500 m × 1.00 m is oriented with its plane parallel to a uniform 3.00-T magnetic field (**Fig. E27.43**). A current of 2.00 A is suddenly started in the coil. (a) About which axis (A_1 or A_2) will the coil begin to rotate? Why? (b) Find the initial angular acceleration of the coil just after the current is started.

Figure **E27.43**



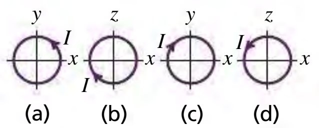
27.44 • Both circular coils A and B (**Fig. E27.44**) have area A and N turns. They are free to rotate about a diameter that coincides with the x-axis. Current I circulates in each coil in the direction shown. There is a uniform magnetic field $\vec{B}$ in the $+z$-direction. (a) What is the direction of the magnetic moment $\vec{\mu}$ for each coil? (b) Explain why the torque on both coils due to the magnetic field is zero, so the coil is in rotational equilibrium. (c) Use Eq. (27.27) to calculate the potential energy for each coil. (d) For each coil, is the equilibrium stable or unstable? Explain.

Figure **E27.44**



27.45 • A circular coil with area A and N turns is free to rotate about a diameter that coincides with the x-axis. Current I is circulating in the coil. There is a uniform magnetic field $\vec{B}$ in the positive y-direction. Calculate the magnitude and direction of the torque $\vec{\tau}$ and the value of the potential energy U, as given in Eq. (27.27), when the coil is oriented as shown in parts (a) through (d) of **Fig. E27.45**.

Figure **E27.45**

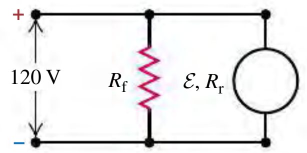


27.46 •• A coil with magnetic moment $1.45\ \text{A}\cdot\text{m}^2$ is oriented initially with its magnetic moment antiparallel to a uniform 0.835-T magnetic field. What is the change in potential energy of the coil when it is rotated 180° so that its magnetic moment is parallel to the field?

Section 27.8 The Direct-Current Motor

27.47 •• In a shunt-wound dc motor with the field coils and rotor connected in parallel (**Fig. E27.47**), the resistance R_f of the field coils is 106 Ω, and the resistance R_r of the rotor is 5.9 Ω. When a potential difference of 120 V is applied to the brushes and the motor is running at full speed delivering mechanical power, the current supplied to it is 4.82 A. (a) What is the current in the field coils? (b) What is the current in the rotor? (c) What is the induced emf developed by the motor? (d) How much mechanical power is developed by this motor?

Figure **E27.47**

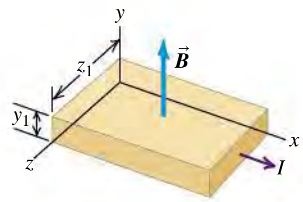


27.48 • A dc motor with its rotor and field coils connected in series has an internal resistance of 3.2 Ω. When the motor is running at full load on a 120-V line, the emf in the rotor is 105 V. (a) What is the current drawn by the motor from the line? (b) What is the power delivered to the motor? (c) What is the mechanical power developed by the motor?

Section 27.9 The Hall Effect

27.49 • **Figure E27.49** shows a portion of a silver ribbon with $z_1 = 11.8$ mm and $y_1 = 0.23$ mm, carrying a current of 120 A in the $+x$-direction. The ribbon lies in a uniform magnetic field, in the y-direction, with magnitude 0.95 T. Apply the simplified model of the Hall effect presented in Section 27.9. If there are 5.85×10^{28} free electrons per cubic meter, find (a) the magnitude of the drift velocity of the electrons in the x-direction; (b) the magnitude and direction of the electric field in the z-direction due to the Hall effect; (c) the Hall emf.

Figure **E27.49**

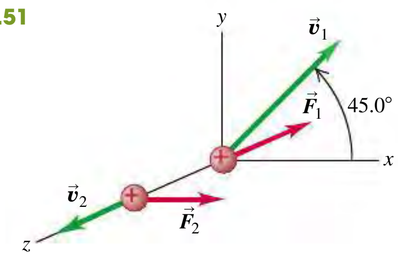


27.50 • Let Fig. E27.49 represent a strip of an unknown metal of the same dimensions as those of the silver ribbon in Exercise 27.49. When the magnetic field is 2.29 T and the current is 78.0 A, the Hall emf is found to be 131 μV. What does the simplified model of the Hall effect presented in Section 27.9 give for the density of free electrons in the unknown metal?

PROBLEMS

27.51 • When a particle of charge $q > 0$ moves with a velocity of $\vec{v}_1$ at 45.0° from the $+x$-axis in the xy-plane, a uniform magnetic field exerts a force $\vec{F}_1$ along the $-z$-axis (**Fig. P27.51**). When the same particle moves with a velocity $\vec{v}_2$ with the same magnitude as $\vec{v}_1$ but along the $+z$-axis, a force $\vec{F}_2$ of magnitude F_2 is exerted on it along the $+x$-axis. (a) What are the magnitude (in terms of q, v_1, and F_2) and direction of the magnetic field? (b) What is the magnitude of $\vec{F}_1$ in terms of F_2?

Figure **P27.51**



27.52 • A particle with charge 7.26×10^{-8} C is moving in a region where there is a uniform 0.650-T magnetic field in the $+x$-direction. At a particular instant, the velocity of the particle has components $v_x = -1.68 \times 10^4$ m/s, $v_y = -3.11 \times 10^4$ m/s, and $v_z = 5.85 \times 10^4$ m/s. What are the components of the force on the particle at this time?

27.53 ••• CP **Fusion Reactor.** If two deuterium nuclei (charge $+e$, mass 3.34×10^{-27} kg) get close enough together, the attraction of the strong nuclear force will fuse them to make an isotope of helium, releasing vast amounts of energy. The range of this force is about 10^{-15} m. This is the principle behind the fusion reactor. The deuterium nuclei are moving much too fast to be contained by physical walls, so they are confined magnetically. (a) How fast would two nuclei have to move so that in a head-on collision they would get close enough to fuse? (Assume their speeds are equal. Treat the nuclei as point charges, and assume that a separation of 1.0×10^{-15} is required for fusion.) (b) What strength magnetic field is needed to make deuterium nuclei with this speed travel in a circle of diameter 2.50 m?

27.54 •• **Magnetic Moment of the Hydrogen Atom.** In the Bohr model of the hydrogen atom (see Section 39.3), in the lowest energy state the electron orbits the proton at a speed of 2.2×10^6 m/s in a circular orbit of radius 5.3×10^{-11} m. (a) What is the orbital period of the electron? (b) If the orbiting electron is considered to be a current loop, what is the current I? (c) What is the magnetic moment of the atom due to the motion of the electron?

27.55 •• You wish to hit a target from several meters away with a charged coin having a mass of 4.25 g and a charge of $+2500\ \mu$C. The coin is given an initial velocity of 12.8 m/s, and a downward, uniform electric field with field strength 27.5 N/C exists throughout the region. If you aim directly at the target and fire the coin horizontally, what magnitude and direction of uniform magnetic field are needed in the region for the coin to hit the target?

27.56 • The magnetic poles of a small cyclotron produce a magnetic field with magnitude 0.85 T. The poles have a radius of 0.40 m, which is the maximum radius of the orbits of the accelerated particles. (a) What is the maximum energy to which protons ($q = 1.60 \times 10^{-19}$C, $m = 1.67 \times 10^{-27}$ kg) can be accelerated by this cyclotron? Give your answer in electron volts and in joules. (b) What is the time for one revolution of a proton orbiting at this maximum radius? (c) What would the magnetic-field magnitude have to be for the maximum energy to which a proton can be accelerated to be twice that calculated in part (a)? (d) For $B = 0.85$ T, what is the maximum energy to which alpha particles ($q = 3.20 \times 10^{-19}$ C, $m = 6.64 \times 10^{-27}$ kg) can be accelerated by this cyclotron? How does this compare to the maximum energy for protons?

27.57 •• A particle with negative charge q and mass $m = 2.58 \times 10^{-15}$ kg is traveling through a region containing a uniform magnetic field $\vec{B} = -(0.120\ \text{T})\hat{k}$. At a particular instant of time the velocity of the particle is $\vec{v} = (1.05 \times 10^6\ \text{m/s})(-3\hat{\imath} + 4\hat{\jmath} + 12\hat{k})$ and the force $\vec{F}$ on the particle has a magnitude of 2.45 N. (a) Determine the charge q. (b) Determine the acceleration $\vec{a}$ of the particle. (c) Explain why the path of the particle is a helix, and determine the radius of curvature R of the circular component of the helical path. (d) Determine the cyclotron frequency of the particle. (e) Although helical motion is not periodic in the full sense of the word, the x- and y-coordinates do vary in a periodic way. If the coordinates of the particle at $t = 0$ are $(x, y, z) = (R, 0, 0)$, determine its coordinates at a time $t = 2T$, where T is the period of the motion in the xy-plane.

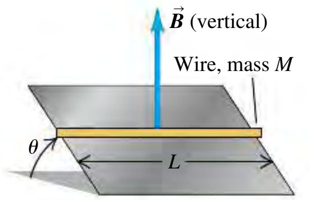
27.58 •• A particle of charge $q > 0$ is moving at speed v in the $+z$-direction through a region of uniform magnetic field $\vec{B}$. The magnetic force on the particle is $\vec{F} = F_0(3\hat{\imath} + 4\hat{\jmath})$, where F_0 is a positive constant. (a) Determine the components B_x, B_y, and B_z, or at least as many of the three components as is possible from the information given. (b) If it is given in addition that the magnetic field has magnitude $6F_0/qv$, determine as much as you can about the remaining components of $\vec{B}$.

27.59 •• Suppose the electric field between the plates in Fig. 27.24 is 1.88×10^4 V/m and the magnetic field in both regions is 0.682 T. If the source contains the three isotopes of krypton, ^{82}Kr, ^{84}Kr, and ^{86}Kr, and the ions are singly charged, find the distance between the lines formed by the three isotopes on the particle detector. Assume the atomic masses of the isotopes (in atomic mass units) are equal to their mass numbers, 82, 84, and 86. (One atomic mass unit = 1 u = 1.66×10^{-27} kg.)

27.60 •• **Mass Spectrograph.** A mass spectrograph is used to measure the masses of ions, or to separate ions of different masses (see Section 27.5). In one design for such an instrument, ions with mass m and charge q are accelerated through a potential difference V. They then enter a uniform magnetic field that is perpendicular to their velocity, and they are deflected in a semicircular path of radius R. A detector measures where the ions complete the semicircle and from this it is easy to calculate R. (a) Derive the equation for calculating the mass of the ion from measurements of B, V, R, and q. (b) What potential difference V is needed so that singly ionized ^{12}C atoms will have $R = 50.0$ cm in a 0.150-T magnetic field? (c) Suppose the beam consists of a mixture of ^{12}C and ^{14}C ions. If v and B have the same values as in part (b), calculate the separation of these two isotopes at the detector. Do you think that this beam separation is sufficient for the two ions to be distinguished? (Make the assumption described in Problem 27.59 for the masses of the ions.)

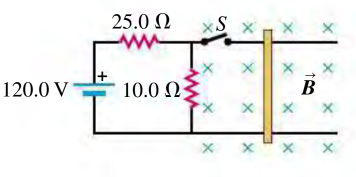
27.61 •• A straight piece of conducting wire with mass M and length L is placed on a frictionless incline tilted at an angle θ from the horizontal (**Fig. P27.61**). There is a uniform, vertical magnetic field $\vec{B}$ at all points (produced by an arrangement of magnets not shown in the figure). To keep the wire from sliding down the incline, a voltage source is attached to the ends of the wire. When just the right amount of current flows through the wire, the wire remains at rest. Determine the magnitude and direction of the current in the wire that will cause the wire to remain at rest. Copy the figure and draw the direction of the current on your copy. In addition, show in a free-body diagram all the forces that act on the wire.

Figure **P27.61**



27.62 •• CP A 2.60-N metal bar, 0.850 m long and having a resistance of 10.0 Ω, rests horizontally on conducting wires connecting it to the circuit shown in **Fig. P27.62**. The bar is in a uniform, horizontal, 1.60-T magnetic field and is not attached to the wires in the circuit. What is the acceleration of the bar just after the switch S is closed?

Figure **P27.62**

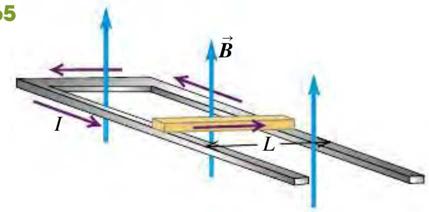


27.63 •• BIO **Determining Diet.** One method for determining the amount of corn in early Native American diets is the *stable* *isotope ratio analysis* (SIRA) technique. As corn photosynthesizes, it concentrates the isotope carbon-13, whereas most other plants concentrate carbon-12. Overreliance on corn consumption can then be correlated with certain diseases, because corn lacks the essential amino acid lysine. Archaeologists use a mass spectrometer to separate the ^{12}C and ^{13}C isotopes in samples of human remains. Suppose you use a velocity selector to obtain singly ionized (missing one electron) atoms of speed 8.50 km/s, and you want to bend them within a uniform magnetic field in a semicircle of diameter 25.0 cm for the ^{12}C. The measured masses of these isotopes are 1.99×10^{-26} kg (^{12}C) and 2.16×10^{-26} kg (^{13}C). (a) What strength of magnetic field is required? (b) What is the diameter of the ^{13}C semicircle? (c) What is the separation of the ^{12}C and ^{13}C ions at the detector at the end of the semicircle? Is this distance large enough to be easily observed?

27.64 •• **CP** A plastic circular loop has radius R, and a positive charge q is distributed uniformly around the circumference of the loop. The loop is then rotated around its central axis, perpendicular to the plane of the loop, with angular speed ω. If the loop is in a region where there is a uniform magnetic field $\vec{B}$ directed parallel to the plane of the loop, calculate the magnitude of the magnetic torque on the loop.

27.65 •• **CP An Electromagnetic Rail Gun.** A conducting bar with mass m and length L slides over horizontal rails that are connected to a voltage source. The voltage source maintains a constant current I in the rails and bar, and a constant, uniform, vertical magnetic field $\vec{B}$ fills the region between the rails (**Fig. P27.65**). (a) Find the magnitude and direction of the net force on the conducting bar. Ignore friction, air resistance, and electrical resistance. (b) If the bar has mass m, find the distance d that the bar must move along the rails from rest to attain speed v. (c) It has been suggested that rail guns based on this principle could accelerate payloads into earth orbit or beyond. Find the distance the bar must travel along the rails if it is to reach the escape speed for the earth (11.2 km/s). Let $B = 0.80$ T, $I = 2.0 \times 10^3$ A, $m = 25$ kg, and $L = 50$ cm. For simplicity assume the net force on the object is equal to the magnetic force, as in parts (a) and (b), even though gravity plays an important role in an actual launch in space.

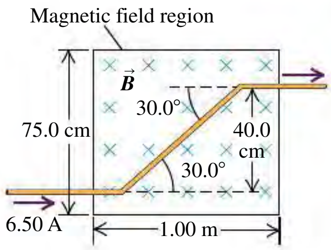
Figure **P27.65**



27.66 • A wire 25.0 cm long lies along the z-axis and carries a current of 7.40 A in the $+z$-direction. The magnetic field is uniform and has components $B_x = -0.242$ T, $B_y = -0.985$ T, and $B_z = -0.336$ T. (a) Find the components of the magnetic force on the wire. (b) What is the magnitude of the net magnetic force on the wire?

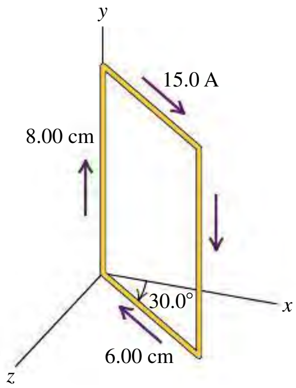
27.67 •• A long wire carrying 6.50 A of current makes two bends, as shown in **Fig. P27.67**. The bent part of the wire passes through a uniform 0.280-T magnetic field directed as shown and confined to a limited region of space. Find the magnitude and direction of the force that the magnetic field exerts on the wire.

Figure **P27.67**



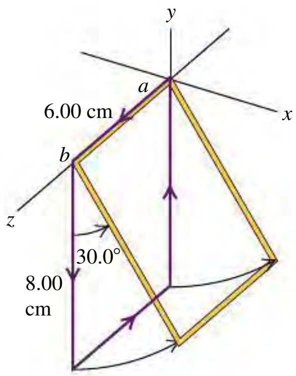
27.68 •• The rectangular loop shown in **Fig. P27.68** is pivoted about the y-axis and carries a current of 15.0 A in the direction indicated. (a) If the loop is in a uniform magnetic field with magnitude 0.48 T in the $+x$-direction, find the magnitude and direction of the torque required to hold the loop in the position shown. (b) Repeat part (a) for the case in which the field is in the $-z$-direction. (c) For each of the above magnetic fields, what torque would be required if the loop were pivoted about an axis through its center, parallel to the y-axis?

Figure **P27.68**



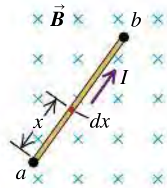
27.69 •• **CP** The rectangular loop of wire shown in **Fig. P27.69** has a mass of 0.15 g per centimeter of length and is pivoted about side ab on a frictionless axis. The current in the wire is 8.2 A in the direction shown. Find the magnitude and direction of the magnetic field parallel to the y-axis that will cause the loop to swing up until its plane makes an angle of 30.0° with the yz-plane.

Figure **P27.69**



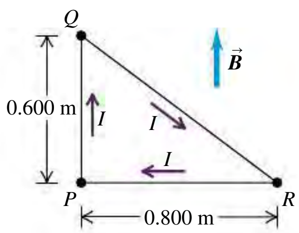
27.70 •• **CALC** A uniform bar of length L carries a current I in the direction from point a to point b (**Fig. P27.70**). The bar is in a uniform magnetic field that is directed into the page. Consider the torque about an axis perpendicular to the bar at point a that is due to the force that the magnetic field exerts on the bar. (a) Suppose that an infinitesimal section of the bar has length dx and is located a distance x from point a. Calculate the torque $d\tau$ about point a due to the magnetic force on this infinitesimal section. (b) Use $\tau = \int_a^b d\tau$ to calculate the total torque τ on the bar. (c) Show that τ is the same as though all of the magnetic force acted at the midpoint of the bar.

Figure **P27.70**



27.71 •• The loop of wire shown in **Fig. P27.71** forms a right triangle and carries a current $I = 5.00$ A in the direction shown. The loop is in a uniform magnetic field that has magnitude $B = 3.00$ T and the same direction as the current in side PQ of the loop. (a) Find the force exerted by the magnetic field on each side of the triangle. If the force is not zero, specify its direction. (b) What is the net force on the loop? (c) The loop is pivoted about an axis that lies along side PR. Use the forces calculated in part (a) to calculate the torque on each side of the loop

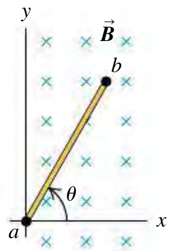
Figure **P27.71**



(see Problem 27.70). (d) What is the magnitude of the net torque on the loop? Calculate the net torque from the torques calculated in part (c) and also from Eq. (27.28). Do these two results agree? (e) Is the net torque directed to rotate point Q into the plane of the figure or out of the plane of the figure?

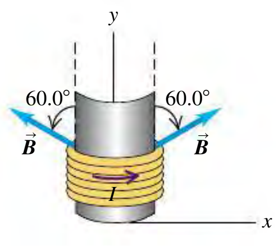
27.72 •• CP A uniform bar has mass 0.0120 kg and is 30.0 cm long. It pivots without friction about an axis perpendicular to the bar at point a (**Fig. P27.72**). The gravitational force on the bar acts in the $-y$-direction. The bar is in a uniform magnetic field that is directed into the page and has magnitude $B = 0.150$ T. (a) What must be the current I in the bar for the bar to be in rotational equilibrium when it is at an angle $\theta = 30.0°$ above the horizontal? Use your result from Problem 27.70. (b) For the bar to be in rotational equilibrium, should I be in the direction from a to b or b to a?

Figure **P27.72**



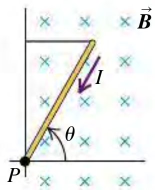
27.73 •• CALC A Voice Coil. It was shown in Section 27.7 that the net force on a current loop in a *uniform* magnetic field is zero. The magnetic force on the voice coil of a loudspeaker (see Fig. 27.28) is nonzero because the magnetic field at the coil is not uniform. A voice coil in a loudspeaker has 50 turns of wire and a diameter of 1.56 cm, and the current in the coil is 0.950 A. Assume that the magnetic field at each point of the coil has a constant magnitude of 0.220 T and is directed at an angle of 60.0° outward from the normal to the plane of the coil (**Fig. P27.73**). Let the axis of the coil be in the y-direction. The current in the coil is in the direction shown (counterclockwise as viewed from a point above the coil on the y-axis). Calculate the magnitude and direction of the net magnetic force on the coil.

Figure **P27.73**



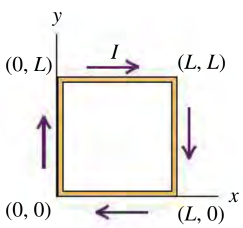
27.74 •• CP The lower end of the thin uniform rod in **Fig. P27.74** is attached to the floor by a frictionless hinge at point P. The rod has mass 0.0840 kg and length 18.0 cm and is in a uniform magnetic field $B = 0.120$ T that is directed into the page. The rod is held at an angle $\theta = 53.0°$ above the horizontal by a horizontal string that connects the top of the rod to the wall. The rod carries a current $I = 12.0$ A in the direction toward P. Calculate the tension in the string. Use your result from Problem 27.70 to calculate the torque due to the magnetic-field force.

Figure **P27.74**



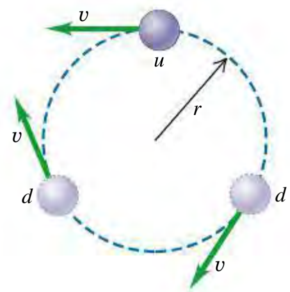
27.75 •• CALC Force on a Current Loop in a Nonuniform Magnetic Field. It was shown in Section 27.7 that the net force on a current loop in a *uniform* magnetic field is zero. But what if $\vec{B}$ is *not* uniform? **Figure P27.75** shows a square loop of wire that lies in the xy-plane. The loop has corners at $(0, 0)$, $(0, L)$, $(L, 0)$, and (L, L) and carries a constant current I in the clockwise direction. The magnetic field has no x-component but has both y- and z-components: $\vec{B} = (B_0 z/L)\hat{\jmath} + (B_0 y/L)\hat{k}$, where B_0 is a positive constant. (a) Sketch the magnetic field lines in the yz-plane. (b) Find the magnitude and direction of the magnetic force exerted on each of the sides of the loop by integrating Eq. (27.20). (c) Find the magnitude and direction of the net magnetic force on the loop.

Figure **P27.75**



27.76 •• Quark Model of the Neutron. The neutron is a particle with zero charge. Nonetheless, it has a nonzero magnetic moment with z-component $9.66 \times 10^{-27}\ \text{A} \cdot \text{m}^2$. This can be explained by the internal structure of the neutron. A substantial body of evidence indicates that a neutron is composed of three fundamental particles called *quarks:* an "up" (u) quark, of charge $+2e/3$, and two "down" (d) quarks, each of charge $-e/3$. The combination of the three quarks produces a net charge of $\frac{2}{3}e - \frac{1}{3}e - \frac{1}{3}e = 0$. If the quarks are in motion, they can produce a nonzero magnetic moment. As a very simple model, suppose the u quark moves in a counterclockwise circular path and the d quarks move in a clockwise circular path, all of radius r and all with the same speed v (**Fig. P27.76**). (a) Determine the current due to the circulation of the u quark. (b) Determine the magnitude of the magnetic moment due to the circulating u quark. (c) Determine the magnitude of the magnetic moment of the three-quark system. (Be careful to use the correct magnetic moment directions.) (d) With what speed v must the quarks move if this model is to reproduce the magnetic moment of the neutron? Use $r = 1.20 \times 10^{-15}$ m (the radius of the neutron) for the radius of the orbits.

Figure **P27.76**



27.77 •• A circular loop of wire with area A lies in the xy-plane. As viewed along the z-axis looking in the $-z$-direction toward the origin, a current I is circulating clockwise around the loop. The torque produced by an external magnetic field $\vec{B}$ is given by $\vec{\tau} = D(4\hat{\imath} - 3\hat{\jmath})$, where D is a positive constant, and for this orientation of the loop the magnetic potential energy $U = -\vec{\mu} \cdot \vec{B}$ is negative. The magnitude of the magnetic field is $B_0 = 13D/IA$. (a) Determine the vector magnetic moment of the current loop. (b) Determine the components B_x, B_y, and B_z of $\vec{B}$.

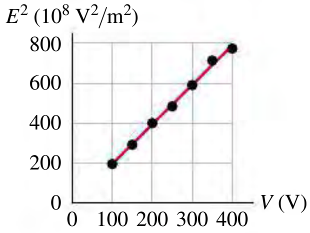
27.78 •• DATA You are using a type of mass spectrometer to measure charge-to-mass ratios of atomic ions. In the device, atoms are ionized with a beam of electrons to produce positive ions, which are then accelerated through a potential difference V. (The final speed of the ions is great enough that you can ignore their initial speed.) The ions then enter a region in which a uniform magnetic field $\vec{B}$ is perpendicular to the velocity of the ions and has magnitude $B = 0.250$ T. In this $\vec{B}$ region, the ions move in a semicircular path of radius R. You measure R as a function of the accelerating voltage V for one particular atomic ion:

V (kV)	10.0	12.0	14.0	16.0	18.0
R (cm)	19.9	21.8	23.6	25.2	26.8

(a) How can you plot the data points so that they will fall close to a straight line? Explain. (b) Construct the graph described in part (a). Use the slope of the best-fit straight line to calculate the charge-to-mass ratio (q/m) for the ion. (c) For $V = 20.0$ kV, what is the speed of the ions as they enter the $\vec{B}$ region? (d) If ions that have $R = 21.2$ cm for $V = 12.0$ kV are singly ionized, what is R when $V = 12.0$ kV for ions that are doubly ionized?

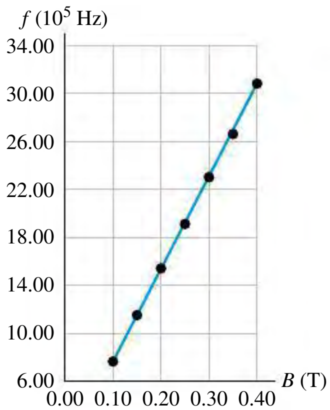
27.79 •• DATA You are a research scientist working on a high-energy particle accelerator. Using a modern version of the Thomson e/m apparatus, you want to measure the mass of a muon (a fundamental particle that has the same charge as an electron but greater mass). The magnetic field between the two charged plates is 0.340 T. You measure the electric field for zero particle deflection as a function of the accelerating potential V. This potential is large enough that you can assume the initial speed of the muons to be zero. **Figure P27.79** is an E^2-versus-V graph of your data. (a) Explain why the data points fall close to a straight line. (b) Use the graph in Fig. P27.79 to calculate the mass m of a muon. (c) If the two charged plates are separated by 6.00 mm, what must be the voltage between the plates in order for the electric field between the plates to be 2.00×10^5 V/m? Assume that the dimensions of the plates are much larger than the separation between them. (d) When $V = 400$ V, what is the speed of the muons as they enter the region between the plates?

Figure **P27.79**



27.80 •• **DATA** You are a technician testing the operation of a cyclotron. An alpha particle in the device moves in a circular path in a magnetic field $\vec{B}$ that is directed perpendicular to the path of the alpha particle. You measure the number of revolutions per second (the frequency f) of the alpha particle as a function of the magnetic field strength B. **Figure P27.80** shows your results and the best straight-line fit to your data. (a) Use the graph in Fig. P27.80 to calculate the charge-to-mass ratio of the alpha particle, which has charge $+2e$. On the basis of your data, what is the mass of an alpha particle? (b) With $B = 0.300$ T, what are the cyclotron frequencies f of a proton and of an electron? How do these f values compare to the frequency of an alpha particle? (c) With $B = 0.300$ T, what speed and kinetic energy does an alpha particle have if the radius of its path is 12.0 cm?

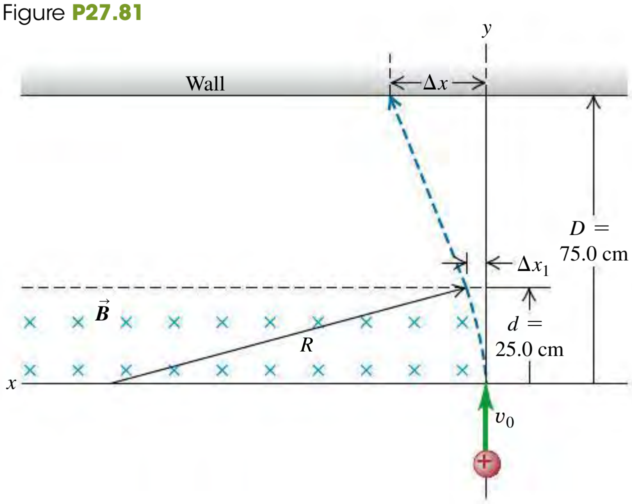
Figure **P27.80**



CHALLENGE PROBLEMS

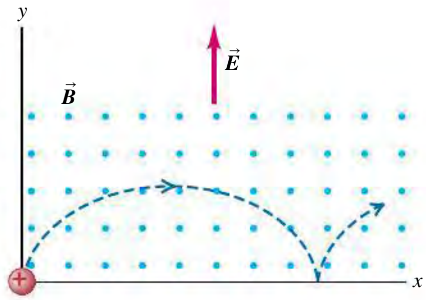
27.81 ••• A particle with charge 2.15 μC and mass 3.20×10^{-11} kg is initially traveling in the $+y$-direction with a speed $v_0 = 1.45 \times 10^5$ m/s. It then enters a region containing a uniform magnetic field that is directed into, and perpendicular to, the page in **Fig. P27.81**. The magnitude of the field is 0.420 T. The region extends a distance of 25.0 cm along the initial direction of travel; 75.0 cm from the point of entry into the magnetic field region is a wall. The length of the field-free region is thus 50.0 cm. When the charged particle enters the magnetic field, it follows a curved path whose radius of curvature is R. It then leaves the magnetic field after a time t_1, having been deflected a distance Δx_1. The particle then travels in the field-free region and strikes the wall after undergoing a total deflection Δx. (a) Determine the radius R of the curved part of the path. (b) Determine t_1, the time the particle spends in the magnetic field. (c) Determine Δx_1, the horizontal deflection at the point of exit from the field. (d) Determine Δx, the total horizontal deflection.

Figure **P27.81**



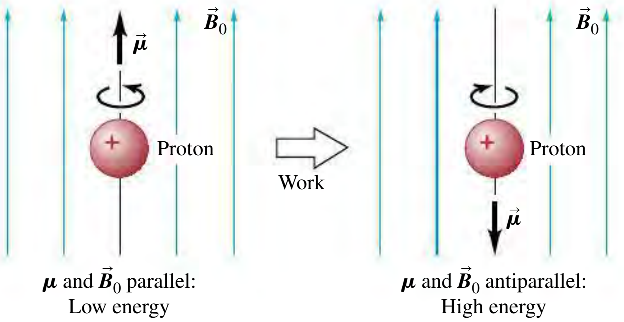
27.82 ••• **CP** **A Cycloidal Path.** A particle with mass m and positive charge q starts from rest at the origin shown in **Fig. P27.82**. There is a uniform electric field $\vec{E}$ in the $+y$-direction and a uniform magnetic field $\vec{B}$ directed out of the page. It is shown in more advanced books that the path is a *cycloid* whose radius of curvature at the top points is twice the y-coordinate at that level. (a) Explain why the path has this general shape and why it is repetitive. (b) Prove that the speed at any point is equal to $\sqrt{2qEy/m}$. (*Hint:* Use energy conservation.) (c) Applying Newton's second law at the top point and taking as given that the radius of curvature here equals $2y$, prove that the speed at this point is $2E/B$.

Figure **P27.82**



PASSAGE PROBLEMS

BIO MAGNETIC FIELDS AND MRI. *Magnetic resonance imaging* (MRI) is a powerful imaging method that, unlike x-ray imaging, allows sharp images of soft tissue to be made without exposing the patient to potentially damaging radiation. A rudimentary understanding of this method can be achieved by the relatively simple application of the classical (that is, non-quantum) physics of magnetism. The starting point for MRI is *nuclear magnetic resonance* (NMR), a technique that depends on the fact that protons in the atomic nucleus have a magnetic field $\vec{B}$. The origin of the proton's magnetic field is the spin of the proton. Being charged, the spinning proton constitutes an electric current analogous to a wire loop through which current flows. Like the wire loop, the proton has a magnetic moment $\vec{\mu}$; thus it will experience a torque when it is subjected to an external magnetic field $\vec{B}_0$. The magnitude of $\vec{\mu}$ is about 1.4×10^{-26} J/T. The proton can be thought of as being in one of two states, with $\vec{\mu}$ oriented parallel or antiparallel to the applied magnetic field, and work must be done to flip the proton from the low-energy state to the high-energy state, as the accompanying figure (next page) shows.



An important consideration is that the net magnetic field of any nucleus, except for that of hydrogen (which has a proton only), consists of contributions from both protons and neutrons. If a nucleus has an even number of protons and neutrons, they will pair in such a way that half of the protons have spins in one orientation and half have spins in the other orientation. Thus the net magnetic moment of the nucleus is zero. Only nuclei with a net magnetic moment are candidates for MRI. Hydrogen is the atom that is most commonly imaged.

27.83 If a proton is exposed to an external magnetic field of 2 T that has a direction perpendicular to the axis of the proton's spin, what will be the torque on the proton? (a) 0; (b) 1.4×10^{-26} N·m; (c) 2.8×10^{-26} N·m; (d) 0.7×10^{-26} N·m.

27.84 Which of following elements is a candidate for MRI? (a) $^{12}C_6$; (b) $^{16}O_8$; (c) $^{40}Ca_{20}$; (d) $^{31}P_{15}$.

27.85 The large magnetic fields used in MRI can produce forces on electric currents within the human body. This effect has been proposed as a possible method for imaging "biocurrents" flowing in the body, such as the current that flows in individual nerves. For a magnetic field strength of 2 T, estimate the magnitude of the maximum force on a 1-mm-long segment of a single cylindrical nerve that has a diameter of 1.5 mm. Assume that the entire nerve carries a current due to an applied voltage of 100 mV (that of a typical action potential). The resistivity of the nerve is 0.6 Ω·m. (a) 6×10^{-7} N; (b) 1×10^{-6} N; (c) 3×10^{-4} N; (d) 0.3 N.

Answers

Chapter Opening Question ?

(ii) A magnetized compass needle has a magnetic dipole moment along its length, and the earth's magnetic field (which points generally northward) exerts a torque that tends to align that dipole moment with the field. See Section 27.7 for details.

Test Your Understanding Questions

27.1 yes When a magnet is cut apart, each part has a north and south pole (see Fig. 27.4). Hence the small red part behaves much like the original, full-sized compass needle.

27.2 path 3 Applying the right-hand rule to the vectors $\vec{v}$ (which points to the right) and $\vec{B}$ (which points into the plane of the figure) says that the force $\vec{F} = q\vec{v} \times \vec{B}$ on a *positive* charge would point *upward*. Since the charge is *negative*, the force points *downward* and the particle follows a trajectory that curves downward.

27.3 (a) (ii), (b) no The magnitude of $\vec{B}$ would increase as you moved to the right, reaching a maximum as you passed through the plane of the loop. As you moved beyond the plane of the loop, the field magnitude would decrease. You can tell this from the spacing of the field lines: The closer the field lines, the stronger the field. The direction of the field would be to the right at all points along the path, since the path is along a field line and the direction of $\vec{B}$ at any point is tangent to the field line through that point.

27.4 (a) (ii), (b) (i) The radius of the orbit as given by Eq. (27.11) is directly proportional to the speed, so doubling the particle speed causes the radius to double as well. The particle has twice as far to travel to complete one orbit but is traveling at double the speed, so the time for one orbit is unchanged. This result also follows from Eq. (27.12), which states that the angular speed ω is independent of the linear speed v. Hence the time per orbit, $T = 2\pi/\omega$, likewise does not depend on v.

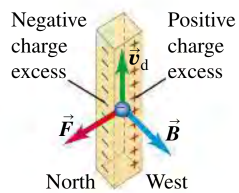
27.5 (iii) From Eq. (27.13), the speed $v = E/B$ at which particles travel straight through the velocity selector does not depend on the magnitude or sign of the charge or the mass of the particle. All that is required is that the particles (in this case, ions) have a nonzero charge.

27.6 A This orientation will cause current to flow clockwise around the circuit and hence through the conducting bar in the direction from the top to the bottom of the figure. From the right-hand rule, the magnetic force $\vec{F} = I\vec{l} \times \vec{B}$ on the bar will then point to the right.

27.7 (a) to the right; (b) north pole on the right, south pole on the left If you wrap the fingers of your right hand around the coil in the direction of the current, your right thumb points to the right (perpendicular to the plane of the coil). This is the direction of the magnetic moment $\vec{\mu}$. The magnetic moment points from the south pole to the north pole, so the right side of the loop is equivalent to a north pole and the left side is equivalent to a south pole.

27.8 no The rotor will not begin to turn when the switch is closed if the rotor is initially oriented as shown in Fig. 27.39b. In this case there is no current through the rotor and hence no magnetic torque. This situation can be remedied by using multiple rotor coils oriented at different angles around the rotation axis. With this arrangement, there is always a magnetic torque no matter what the orientation.

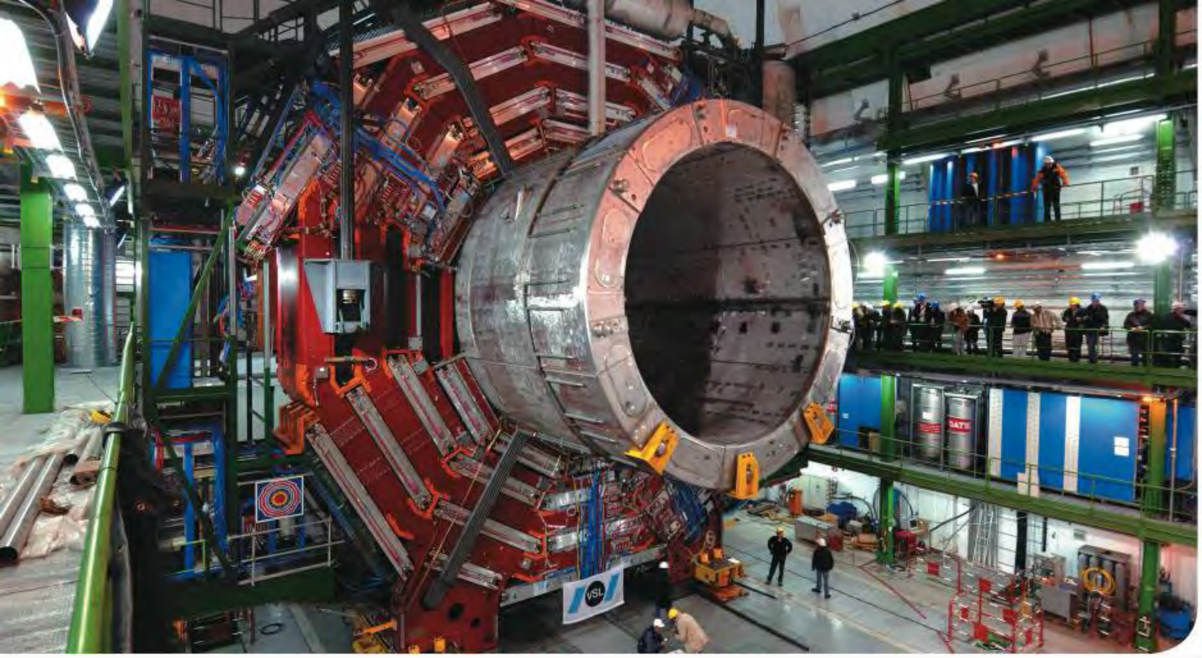
27.9 (ii) The mobile charge carriers in copper are negatively charged electrons, which move upward through the wire to give a downward current. From the right-hand rule, the force on a positively charged particle moving upward in a westward-pointing magnetic field would be to the south; hence the force on a negatively charged particle is to the north. The result is an excess of negative charge on the north side of the wire, leaving an excess of positive charge—and hence a higher electric potential—on the south side.



Bridging Problem

(a) $\tau_x = -1.54 \times 10^{-4}$ N·m,
$\tau_y = -2.05 \times 10^{-4}$ N·m,
$\tau_z = -6.14 \times 10^{-4}$ N·m

(b) -7.55×10^{-4} J **(c)** 42.1 rad/s



? The immense cylinder in this photograph is a current-carrying coil, or solenoid, that generates a uniform magnetic field in its interior as part of an experiment at CERN, the European Organization for Nuclear Research. If two such solenoids were joined end to end, the magnetic field along their common axis would (i) become four times stronger; (ii) double in strength; (iii) become $\sqrt{2}$ times stronger; (iv) not change; (v) weaken.

28 SOURCES OF MAGNETIC FIELD

LEARNING GOALS

Looking forward at …

28.1 The nature of the magnetic field produced by a single moving charged particle.

28.2 How to describe the magnetic field produced by an element of a current-carrying conductor.

28.3 How to calculate the magnetic field produced by a long, straight, current-carrying wire.

28.4 Why wires carrying current in the same direction attract, while wires carrying opposing currents repel.

28.5 How to calculate the magnetic field produced by a current-carrying wire bent into a circle.

28.6 What Ampere's law is, and what it tells us about magnetic fields.

28.7 How to use Ampere's law to calculate the magnetic field of symmetric current distributions.

28.8 How microscopic currents within materials give them their magnetic properties.

Looking back at …

10.5 Angular momentum of a particle.

21.3–21.5 Coulomb's law and electric-field calculations.

22.4 Solving problems with Gauss's law.

27.2–27.9 Magnetic field and magnetic force.

In Chapter 27 we studied the forces exerted on moving charges and on current-carrying conductors in a magnetic field. We didn't worry about how the magnetic field got there; we simply took its existence as a given fact. But how are magnetic fields *created*? We know that both permanent magnets and electric currents in electromagnets create magnetic fields. In this chapter we will study these sources of magnetic field in detail.

We've learned that a charge creates an electric field and that an electric field exerts a force on a charge. But a *magnetic* field exerts a force on only a *moving* charge. Similarly, we'll see that only *moving* charges *create* magnetic fields. We'll begin our analysis with the magnetic field created by a single moving point charge. We can use this analysis to determine the field created by a small segment of a current-carrying conductor. Once we can do that, we can in principle find the magnetic field produced by *any* shape of conductor.

Then we will introduce Ampere's law, which plays a role in magnetism analogous to the role of Gauss's law in electrostatics. Ampere's law lets us exploit symmetry properties in relating magnetic fields to their sources.

Moving charged particles within atoms respond to magnetic fields and can also act as sources of magnetic field. We'll use these ideas to understand how certain magnetic materials can be used to intensify magnetic fields as well as why some materials such as iron act as permanent magnets.

28.1 MAGNETIC FIELD OF A MOVING CHARGE

Let's start with the basics, the magnetic field of a single point charge q moving with a constant velocity $\vec{v}$. In practical applications, such as the solenoid shown in the photo that opens this chapter, magnetic fields are produced by tremendous numbers of charged particles moving together in a current. But once we understand how to calculate the magnetic field due to a single point charge, it's a small leap to calculate the field due to a current-carrying wire or collection of wires.

As we did for electric fields, we call the location of the moving charge at a given instant the **source point** and the point P where we want to find the field the **field point.** In Section 21.4 we found that at a field point a distance r from a point charge q, the magnitude of the *electric* field $\vec{E}$ caused by the charge is proportional to the charge magnitude $|q|$ and to $1/r^2$, and the direction of $\vec{E}$ (for positive q) is along the line from source point to field point. The corresponding relationship for the *magnetic* field $\vec{B}$ of a point charge q moving with constant velocity has some similarities and some interesting differences.

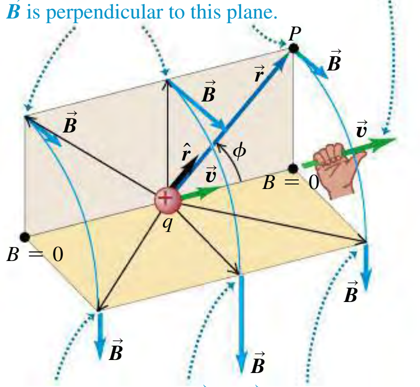
Experiments show that the magnitude of $\vec{B}$ is also proportional to $|q|$ and to $1/r^2$. But the *direction* of $\vec{B}$ is *not* along the line from source point to field point. Instead, $\vec{B}$ is perpendicular to the plane containing this line and the particle's velocity vector $\vec{v}$, as shown in **Fig. 28.1**. Furthermore, the field *magnitude* B is also proportional to the particle's speed v and to the sine of the angle ϕ. Thus the magnetic-field magnitude at point P is

$$B = \frac{\mu_0}{4\pi}\frac{|q|v\sin\phi}{r^2} \tag{28.1}$$

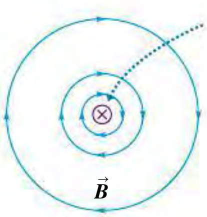
The quantity μ_0 (read as "mu-nought" or "mu-sub-zero") is called the **magnetic constant.** The reason for including the factor of 4π will emerge shortly. We did something similar with Coulomb's law in Section 21.3.

28.1 (a) Magnetic-field vectors due to a moving positive point charge q. At each point, $\vec{B}$ is perpendicular to the plane of $\vec{r}$ and $\vec{v}$, and its magnitude is proportional to the sine of the angle between them. (b) Magnetic field lines in a plane containing a moving positive charge.

(a) Perspective view



(b) View from behind the charge



Moving Charge: Vector Magnetic Field

We can incorporate both the magnitude and direction of $\vec{B}$ into a single vector equation by using the vector product. To avoid having to say "the direction from the source q to the field point P" over and over, we introduce a *unit* vector $\hat{r}$ ("r-hat") that points from the source point to the field point. (We used $\hat{r}$ for the same purpose in Section 21.4.) This unit vector is equal to the vector $\vec{r}$ from the source to the field point divided by its magnitude: $\hat{r} = \vec{r}/r$. Then

Magnetic field due to a point charge with constant velocity

$$\vec{B} = \frac{\mu_0}{4\pi}\frac{q\vec{v}\times\hat{r}}{r^2} \tag{28.2}$$

Magnetic constant: μ_0; Charge: q; Velocity: $\vec{v}$; $\hat{r}$: Unit vector from point charge toward where field is measured; r: Distance from point charge to where field is measured

Figure 28.1 shows the relationship of $\hat{r}$ to P and shows the magnetic field $\vec{B}$ at several points in the vicinity of the charge. At all points along a line through the charge parallel to the velocity $\vec{v}$, the field is zero because $\sin\phi = 0$ at all such points. At any distance r from q, $\vec{B}$ has its greatest magnitude at points lying in the plane perpendicular to $\vec{v}$, because there $\phi = 90°$ and $\sin\phi = 1$. If q is negative, the directions of $\vec{B}$ are opposite to those shown in Fig. 28.1.

Moving Charge: Magnetic Field Lines

A point charge in motion also produces an *electric* field, with field lines that radiate outward from a positive charge. The *magnetic* field lines are completely different. For a point charge moving with velocity $\vec{v}$, the magnetic field lines are *circles* centered on the line of $\vec{v}$ and lying in planes perpendicular to this line. The field-line directions for a positive charge are given by the following *right-hand rule,* one of several that we will encounter in this chapter: Grasp the velocity vector $\vec{v}$ with your right hand so that your right thumb points in the direction of $\vec{v}$; your fingers then curl around the line of $\vec{v}$ in the same sense as the magnetic field lines, assuming q is positive. Figure 28.1a shows parts of a few field lines; Fig. 28.1b shows some field lines in a plane through q, perpendicular to $\vec{v}$. If the point charge is negative, the directions of the field and field lines are the opposite of those shown in Fig. 28.1.

Equations (28.1) and (28.2) describe the $\vec{B}$ field of a point charge moving with *constant* velocity. If the charge *accelerates,* the field can be much more complicated. We won't need these more complicated results for our purposes. (The moving charged particles that make up a current in a wire accelerate at points where the wire bends and the direction of $\vec{v}$ changes. But because the magnitude v_d of the drift velocity in a conductor is typically very small, the centripetal acceleration v_d^2/r is so small that we can ignore its effects.)

As we discussed in Section 27.2, the unit of B is one tesla (1 T):

$$1\ \text{T} = 1\ \text{N} \cdot \text{s/C} \cdot \text{m} = 1\ \text{N/A} \cdot \text{m}$$

Using this with Eq. (28.1) or (28.2), we find that the units of the constant μ_0 are

$$1\ \text{N} \cdot \text{s}^2/\text{C}^2 = 1\ \text{N/A}^2 = 1\ \text{Wb/A} \cdot \text{m} = 1\ \text{T} \cdot \text{m/A}$$

In SI units the numerical value of μ_0 is exactly $4\pi \times 10^{-7}$. Thus

$$\begin{aligned} \mu_0 &= 4\pi \times 10^{-7}\ \text{N} \cdot \text{s}^2/\text{C}^2 = 4\pi \times 10^{-7}\ \text{Wb/A} \cdot \text{m} \\ &= 4\pi \times 10^{-7}\ \text{T} \cdot \text{m/A} \end{aligned} \tag{28.3}$$

It may seem incredible that μ_0 has *exactly* this numerical value! In fact this is a *defined* value that arises from the definition of the ampere, as we'll discuss in Section 28.4.

We mentioned in Section 21.3 that the constant $1/4\pi\epsilon_0$ in Coulomb's law is related to the speed of light c:

$$k = \frac{1}{4\pi\epsilon_0} = (10^{-7}\ \text{N} \cdot \text{s}^2/\text{C}^2)c^2$$

When we study electromagnetic waves in Chapter 32, we will find that their speed of propagation in vacuum, which is equal to the speed of light c, is given by

$$c^2 = \frac{1}{\epsilon_0\mu_0} \tag{28.4}$$

If we solve the equation $k = 1/4\pi\epsilon_0$ for ϵ_0, substitute the resulting expression into Eq. (28.4), and solve for μ_0, we indeed get the value of μ_0 stated above. This discussion is a little premature, but it may give you a hint that electric and magnetic fields are intimately related to the nature of light.

EXAMPLE 28.1 FORCES BETWEEN TWO MOVING PROTONS

SOLUTION

Two protons move parallel to the x-axis in opposite directions (**Fig. 28.2**) at the same speed v (small compared to the speed of light c). At the instant shown, find the electric and magnetic forces on the upper proton and compare their magnitudes.

SOLUTION

IDENTIFY and SET UP: Coulomb's law [Eq. (21.2)] gives the electric force F_E on the upper proton. The magnetic force law [Eq. (27.2)] gives the magnetic force on the upper proton; to use it, we must first use Eq. (28.2) to find the magnetic field that the lower proton produces at the position of the upper proton. The unit vector from the lower proton (the source) to the position of the upper proton is $\hat{r} = \hat{j}$.

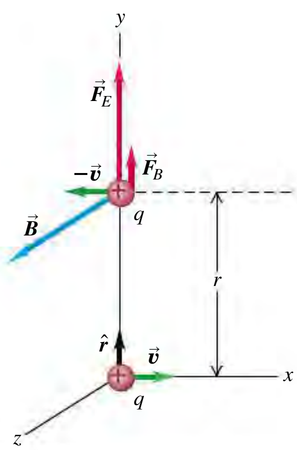
EXECUTE: From Coulomb's law, the magnitude of the electric force on the upper proton is

$$F_E = \frac{1}{4\pi\epsilon_0}\frac{q^2}{r^2}$$

The forces are repulsive, and the force on the upper proton is vertically upward (in the $+y$-direction).

The velocity of the lower proton is $\vec{v} = v\hat{\imath}$. From the right-hand rule for the cross product $\vec{v} \times \hat{r}$ in Eq. (28.2), the $\vec{B}$ field due to the lower proton at the position of the upper proton is in the $+z$-direction (see Fig. 28.2). From Eq. (28.2), the field is

28.2 Electric and magnetic forces between two moving protons.



$$\vec{B} = \frac{\mu_0}{4\pi}\frac{q(v\hat{\imath}) \times \hat{\jmath}}{r^2} = \frac{\mu_0}{4\pi}\frac{qv}{r^2}\hat{k}$$

Continued

The velocity of the upper proton is $-\vec{v} = -v\hat{\imath}$, so the magnetic force on it is

$$\vec{F}_B = q(-\vec{v}) \times \vec{B} = q(-v\hat{\imath}) \times \frac{\mu_0}{4\pi}\frac{qv}{r^2}\hat{k} = \frac{\mu_0}{4\pi}\frac{q^2v^2}{r^2}\hat{\jmath}$$

The magnetic interaction in this situation is also repulsive. The ratio of the force magnitudes is

$$\frac{F_B}{F_E} = \frac{\mu_0 q^2v^2/4\pi r^2}{q^2/4\pi\epsilon_0 r^2} = \frac{\mu_0 v^2}{1/\epsilon_0} = \epsilon_0\mu_0 v^2$$

With the relationship $\epsilon_0\mu_0 = 1/c^2$, Eq. (28.4), this becomes

$$\frac{F_B}{F_E} = \frac{v^2}{c^2}$$

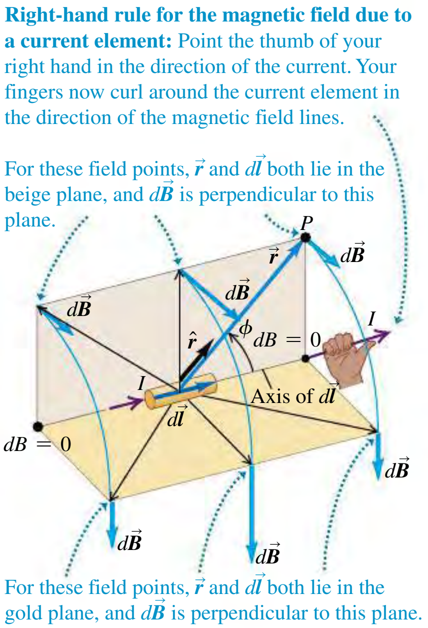
When v is small in comparison to the speed of light, the magnetic force is much smaller than the electric force.

EVALUATE: We have described the velocities, fields, and forces as they are measured by an observer who is stationary in the coordinate system of Fig. 28.2. In a coordinate system that moves with one of the charges, one of the velocities would be zero, so there would be *no* magnetic force. The explanation of this apparent paradox provided one of the paths that led to the special theory of relativity.

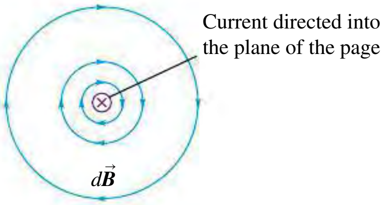
TEST YOUR UNDERSTANDING OF SECTION 28.1 (a) If two protons are traveling parallel to each other in the *same* direction and at the same speed, is the magnetic force between them (i) attractive or (ii) repulsive? (b) Is the net force between them (i) attractive, (ii) repulsive, or (iii) zero? (Assume that the protons' speed is much slower than the speed of light.)

28.3 (a) Magnetic-field vectors due to a current element $d\vec{l}$. (b) Magnetic field lines in a plane containing the current element $d\vec{l}$. Compare this figure to Fig. 28.1 for the field of a moving point charge.

(a) Perspective view



(b) View along the axis of the current element



28.2 MAGNETIC FIELD OF A CURRENT ELEMENT

As for electric fields, there is a **principle of superposition of magnetic fields:**

> **The total magnetic field caused by several moving charges is the vector sum of the fields caused by the individual charges.**

We can use this principle with the results of Section 28.1 to find the magnetic field produced by a current in a conductor.

We begin by calculating the magnetic field caused by a short segment $d\vec{l}$ of a current-carrying conductor, as shown in **Fig. 28.3a**. The volume of the segment is $A\,dl$, where A is the cross-sectional area of the conductor. If there are n moving charged particles per unit volume, each of charge q, the total moving charge dQ in the segment is

$$dQ = nqA\,dl$$

The moving charges in this segment are equivalent to a single charge dQ, traveling with a velocity equal to the *drift* velocity $\vec{v}_d$. (Magnetic fields due to the *random* motions of the charges will, on average, cancel out at every point.) From Eq. (28.1) the magnitude of the resulting field $d\vec{B}$ at any field point P is

$$dB = \frac{\mu_0}{4\pi}\frac{|dQ|v_d \sin\phi}{r^2} = \frac{\mu_0}{4\pi}\frac{n|q|v_d A\,dl \sin\phi}{r^2}$$

But from Eq. (25.2), $n|q|v_d A$ equals the current I in the element. So

$$dB = \frac{\mu_0}{4\pi}\frac{I\,dl \sin\phi}{r^2} \qquad (28.5)$$

Current Element: Vector Magnetic Field

In vector form, using the unit vector $\hat{r}$ as in Section 28.1, we have

$$\underbrace{d\vec{B}}_{\substack{\text{Magnetic field due}\\\text{to an infinitesimal}\\\text{current element}}} = \underbrace{\frac{\mu_0}{4\pi}}_{\substack{\text{Magnetic}\\\text{constant}}}\frac{\overbrace{I}^{\text{Current}}\,\overbrace{d\vec{l}}^{\substack{\text{Vector length of element}\\\text{(points in current direction)}}} \times \overbrace{\hat{r}}^{\substack{\text{Unit vector from}\\\text{element toward where}\\\text{field is measured}}}}{\underbrace{r^2}_{\substack{\text{Distance from element}\\\text{to where field is measured}}}} \qquad (28.6)$$

where $d\vec{l}$ is a vector with length dl, in the same direction as the current.

Equations (28.5) and (28.6) are called the **law of Biot and Savart** (pronounced "Bee-oh" and "Suh-var"). We can use this law to find the total magnetic field $\vec{B}$ at any point in space due to the current in a complete circuit. To do this, we integrate Eq. (28.6) over all segments $d\vec{l}$ that carry current; symbolically,

$$\vec{B} = \frac{\mu_0}{4\pi}\int \frac{I\,d\vec{l} \times \hat{r}}{r^2} \qquad (28.7)$$

In the following sections we will carry out this vector integration for several examples.

Application Currents and Planetary Magnetism The earth's magnetic field is caused by currents circulating within its molten, conducting interior. These currents are stirred by our planet's relatively rapid spin (one rotation per 24 hours). The moon's internal currents are much weaker; it is much smaller than the earth, has a predominantly solid interior, and spins slowly (one rotation per 27.3 days). Hence the moon's magnetic field is only about 10^{-4} as strong as that of the earth.

Current Element: Magnetic Field Lines

As Fig. 28.3 shows, the field vectors $d\vec{B}$ and the magnetic field lines of a current element are exactly like those set up by a positive charge dQ moving in the direction of the drift velocity $\vec{v}_d$. The field lines are circles in planes perpendicular to $d\vec{l}$ and centered on the line of $d\vec{l}$. Their directions are given by the same right-hand rule that we introduced for point charges in Section 28.1.

We can't verify Eq. (28.5) or (28.6) directly because we can never experiment with an isolated segment of a current-carrying circuit. What we measure experimentally is the *total* $\vec{B}$ for a complete circuit. But we can still verify these equations indirectly by calculating $\vec{B}$ for various current configurations with Eq. (28.7) and comparing the results with experimental measurements.

If matter is present in the space around a current-carrying conductor, the field at a field point P in its vicinity will have an additional contribution resulting from the *magnetization* of the material. We'll return to this point in Section 28.8. However, unless the material is iron or some other ferromagnetic material, the additional field is small and is usually negligible. Additional complications arise if time-varying electric or magnetic fields are present or if the material is a superconductor; we'll return to these topics later.

PROBLEM-SOLVING STRATEGY 28.1 MAGNETIC-FIELD CALCULATIONS

IDENTIFY *the relevant concepts:* The Biot–Savart law [Eqs. (28.5) and (28.6)] allows you to calculate the magnetic field at a field point P due to a current-carrying wire of any shape. The idea is to calculate the field element $d\vec{B}$ at P due to a representative current element in the wire and integrate all such field elements to find the field $\vec{B}$ at P.

SET UP *the problem* using the following steps:
1. Make a diagram showing a representative current element and the field point P.
2. Draw the current element $d\vec{l}$, being careful that it points in the direction of the current.
3. Draw the unit vector $\hat{r}$ directed *from* the current element (the source point) to P.
4. Identify the target variable (usually $\vec{B}$).

EXECUTE *the solution* as follows:
1. Use Eq. (28.5) or (28.6) to express the magnetic field $d\vec{B}$ at P from the representative current element.
2. Add up all the $d\vec{B}$'s to find the total field at point P. In some situations the $d\vec{B}$'s at point P have the same direction for all the current elements; then the magnitude of the total $\vec{B}$ field is the sum of the magnitudes of the $d\vec{B}$'s. But often the $d\vec{B}$'s have different directions for different current elements. Then you have to set up a coordinate system and represent each $d\vec{B}$ in terms of its components. The integral for the total $\vec{B}$ is then expressed in terms of an integral for each component.
3. Sometimes you can use the symmetry of the situation to prove that one component of $\vec{B}$ must vanish. Always be alert for ways to use symmetry to simplify the problem.
4. Look for ways to use the principle of superposition of magnetic fields. Later in this chapter we'll determine the fields produced by certain simple conductor shapes; if you encounter a conductor of a complex shape that can be represented as a combination of these simple shapes, you can use superposition to find the field of the complex shape. Examples include a rectangular loop and a semicircle with straight line segments on both sides.

EVALUATE *your answer:* Often your answer will be a mathematical expression for $\vec{B}$ as a function of the position of the field point. Check the answer by examining its behavior in as many limits as you can.

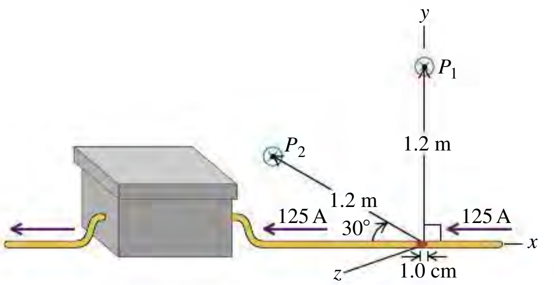
EXAMPLE 28.2 MAGNETIC FIELD OF A CURRENT SEGMENT

A copper wire carries a steady 125-A current to an electroplating tank (**Fig. 28.4**). Find the magnetic field due to a 1.0-cm segment of this wire at a point 1.2 m away from it, if the point is (a) point P_1, straight out to the side of the segment, and (b) point P_2, in the xy-plane and on a line at 30° to the segment.

SOLUTION

IDENTIFY and SET UP: Although Eqs. (28.5) and (28.6) apply only to infinitesimal current elements, we may use either of them here because the segment length is much less than the distance to the field point. The current element is shown in red in Fig. 28.4 and points in the $-x$-direction (the direction of the current), so $d\vec{l} = dl(-\hat{\imath})$. The unit vector $\hat{r}$ for each field point is directed from the current element toward that point: $\hat{r}$ is in the $+y$-direction for point P_1 and at an angle of 30° above the $-x$-direction for point P_2.

28.4 Finding the magnetic field at two points due to a 1.0-cm segment of current-carrying wire (not shown to scale).



EXECUTE: (a) At point P_1, $\hat{r} = \hat{\jmath}$, so

$$\vec{B} = \frac{\mu_0}{4\pi}\frac{I\,d\vec{l} \times \hat{r}}{r^2} = \frac{\mu_0}{4\pi}\frac{I\,dl(-\hat{\imath}) \times \hat{\jmath}}{r^2} = -\frac{\mu_0}{4\pi}\frac{I\,dl}{r^2}\hat{k}$$

$$= -(10^{-7}\,\mathrm{T \cdot m/A})\frac{(125\,\mathrm{A})(1.0 \times 10^{-2}\,\mathrm{m})}{(1.2\,\mathrm{m})^2}\hat{k}$$

$$= -(8.7 \times 10^{-8}\,\mathrm{T})\hat{k}$$

The direction of $\vec{B}$ at P_1 is into the xy-plane of Fig. 28.4.

(b) At P_2, the unit vector is $\hat{r} = (-\cos 30°)\hat{\imath} + (\sin 30°)\hat{\jmath}$. From Eq. (28.6),

$$\vec{B} = \frac{\mu_0}{4\pi}\frac{I\,d\vec{l} \times \hat{r}}{r^2} = \frac{\mu_0}{4\pi}\frac{I\,dl(-\hat{\imath}) \times (-\cos 30°\hat{\imath} + \sin 30°\hat{\jmath})}{r^2}$$

$$= -\frac{\mu_0 I}{4\pi}\frac{dl \sin 30°}{r^2}\hat{k}$$

$$= -(10^{-7}\,\mathrm{T \cdot m/A})\frac{(125\,\mathrm{A})(1.0 \times 10^{-2}\,\mathrm{m})(\sin 30°)}{(1.2\,\mathrm{m})^2}\hat{k}$$

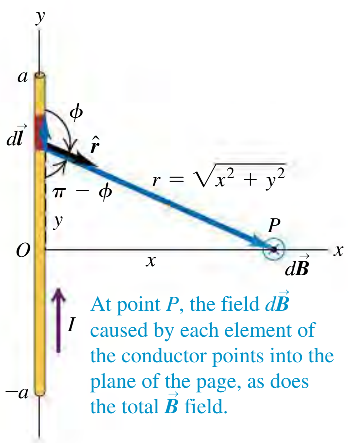
$$= -(4.3 \times 10^{-8}\,\mathrm{T})\hat{k}$$

The direction of $\vec{B}$ at P_2 is also into the xy-plane of Fig. 28.4.

EVALUATE: We can check our results for the direction of $\vec{B}$ by comparing them with Fig. 28.3. The xy-plane in Fig. 28.4 corresponds to the beige plane in Fig. 28.3, but here the direction of the current and hence of $d\vec{l}$ is the reverse of that shown in Fig. 28.3. Hence the direction of the magnetic field is reversed as well. Hence the field at points in the xy-plane in Fig. 28.4 must point *into,* not out of, that plane. This is just what we concluded above.

TEST YOUR UNDERSTANDING OF SECTION 28.2 An infinitesimal current element located at the origin ($x = y = z = 0$) carries current I in the positive y-direction. Rank the following locations in order of the strength of the magnetic field that the current element produces at that location, from largest to smallest value. (i) $x = L$, $y = 0$, $z = 0$; (ii) $x = 0$, $y = L$, $z = 0$; (iii) $x = 0$, $y = 0$, $z = L$; (iv) $x = L/\sqrt{2}$, $y = L/\sqrt{2}$, $z = 0$.

28.5 Magnetic field produced by a straight current-carrying conductor of length $2a$.



28.3 MAGNETIC FIELD OF A STRAIGHT CURRENT-CARRYING CONDUCTOR

Let's use the law of Biot and Savart to find the magnetic field produced by a straight current-carrying conductor. This result is useful because straight conducting wires are found in essentially all electric and electronic devices. **Figure 28.5** shows such a conductor with length $2a$ carrying a current I. We will find $\vec{B}$ at a point a distance x from the conductor on its perpendicular bisector.

We first use the law of Biot and Savart, Eq. (28.5), to find the field $d\vec{B}$ caused by the element of conductor of length $dl = dy$ shown in Fig. 28.5. From the figure, $r = \sqrt{x^2 + y^2}$ and $\sin\phi = \sin(\pi - \phi) = x/\sqrt{x^2 + y^2}$. The right-hand rule for the vector product $d\vec{l} \times \hat{r}$ shows that the *direction* of $d\vec{B}$ is into the plane of the figure, perpendicular to the plane; furthermore, the directions of the $d\vec{B}$'s from *all* elements of the conductor are the same. Thus in integrating Eq. (28.7), we can just add the *magnitudes* of the $d\vec{B}$'s, a significant simplification.

Putting the pieces together, we find that the magnitude of the total $\vec{B}$ field is

$$B = \frac{\mu_0 I}{4\pi} \int_{-a}^{a} \frac{x\,dy}{(x^2 + y^2)^{3/2}}$$

We can integrate this by trigonometric substitution or by using an integral table:

$$B = \frac{\mu_0 I}{4\pi} \frac{2a}{x\sqrt{x^2 + a^2}} \qquad (28.8)$$

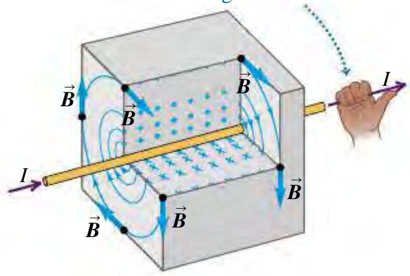
When the length $2a$ of the conductor is much greater than its distance x from point P, we can consider it to be infinitely long. When a is much larger than x, $\sqrt{x^2 + a^2}$ is approximately equal to a; hence in the limit $a \rightarrow \infty$, Eq. (28.8) becomes

$$B = \frac{\mu_0 I}{2\pi x}$$

The physical situation has axial symmetry about the y-axis. Hence $\vec{B}$ must have the same *magnitude* at all points on a circle centered on the conductor and lying in a plane perpendicular to it, and the *direction* of $\vec{B}$ must be everywhere tangent to such a circle (**Fig. 28.6**). Thus, at all points on a circle of radius r around the conductor, the magnitude B is

Magnetic field near a long, straight, current-carrying conductor

$$\underbrace{B}_{\text{Magnetic field near a long, straight, current-carrying conductor}} = \frac{\overbrace{\mu_0}^{\text{Magnetic constant}}\overbrace{I}^{\text{Current}}}{2\pi \underbrace{r}_{\text{Distance from conductor}}} \qquad (28.9)$$



28.6 Magnetic field around a long, straight, current-carrying conductor. The field lines are circles, with directions determined by the right-hand rule.

The geometry in this case is similar to that of Example 21.10 (Section 21.5), in which we solved the problem of the *electric* field caused by an infinite line of charge. The same integral appears in both problems, and the field magnitudes in both problems are proportional to $1/r$. But the lines of $\vec{B}$ in the magnetic problem have completely different shapes than the lines of $\vec{E}$ in the analogous electrical problem. Electric field lines radiate outward from a positive line charge distribution (inward for negative charges). By contrast, magnetic field lines *encircle* the current that acts as their source. Electric field lines due to charges begin and end at those charges, but magnetic field lines always form closed loops and *never* have endpoints, irrespective of the shape of the current-carrying conductor that sets up the field. As we discussed in Section 27.3, this is a consequence of Gauss's law for magnetism, which states that the total magnetic flux through *any* closed surface is always zero:

$$\oint \vec{B} \cdot d\vec{A} = 0 \qquad \text{(magnetic flux through any closed surface)} \qquad (28.10)$$

Any magnetic field line that enters a closed surface must emerge from that surface.

EXAMPLE 28.3 MAGNETIC FIELD OF A SINGLE WIRE

SOLUTION

A long, straight conductor carries a 1.0-A current. At what distance from the axis of the conductor does the resulting magnetic field have magnitude $B = 0.5 \times 10^{-4}$ T (about that of the earth's magnetic field in Pittsburgh)?

SOLUTION

IDENTIFY and SET UP: The length of a "long" conductor is much greater than the distance from the conductor to the field point. Hence we can use the ideas of this section. The geometry is the same as that of Fig. 28.6, so we use Eq. (28.9). All of the quantities in this equation are known except the target variable, the distance r.

EXECUTE: We solve Eq. (28.9) for r:

$$r = \frac{\mu_0 I}{2\pi B} = \frac{(4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A})(1.0\ \text{A})}{(2\pi)(0.5 \times 10^{-4}\ \text{T})}$$
$$= 4 \times 10^{-3}\ \text{m} = 4\ \text{mm}$$

EVALUATE: As we saw in Example 26.14, currents of an ampere or more are typical of those found in the wiring of home appliances. This example shows that the magnetic fields produced by these appliances are very weak even very close to the wire; the fields are proportional to $1/r$, so they become even weaker at greater distances.

EXAMPLE 28.4 MAGNETIC FIELD OF TWO WIRES

Figure 28.7a is an end-on view of two long, straight, parallel wires perpendicular to the xy-plane, each carrying a current I but in opposite directions. (a) Find $\vec{B}$ at points P_1, P_2, and P_3. (b) Find an expression for $\vec{B}$ at any point on the x-axis to the right of wire 2.

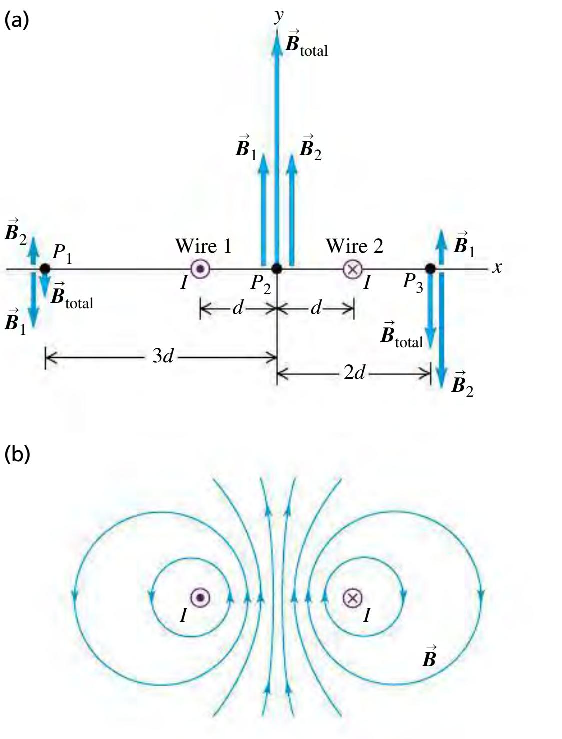
SOLUTION

IDENTIFY and SET UP: We can find the magnetic fields $\vec{B}_1$ and $\vec{B}_2$ due to wires 1 and 2 at each point by using the ideas of this section. By the superposition principle, the magnetic field at each point is then $\vec{B} = \vec{B}_1 + \vec{B}_2$. We use Eq. (28.9) to find the magnitudes B_1 and B_2 of these fields and the right-hand rule to find the corresponding directions. Figure 28.7a shows $\vec{B}_1$, $\vec{B}_2$, and $\vec{B} = \vec{B}_{\text{total}}$ at each point; you should confirm that the directions and relative magnitudes shown are correct. Figure 28.7b shows some of the magnetic field lines due to this two-wire system.

EXECUTE: (a) Since point P_1 is a distance $2d$ from wire 1 and a distance $4d$ from wire 2, $B_1 = \mu_0 I/2\pi(2d) = \mu_0 I/4\pi d$ and $B_2 = \mu_0 I/2\pi(4d) = \mu_0 I/8\pi d$. The right-hand rule shows that $\vec{B}_1$ is in the negative y-direction and $\vec{B}_2$ is in the positive y-direction, so

$$\vec{B}_{\text{total}} = \vec{B}_1 + \vec{B}_2 = -\frac{\mu_0 I}{4\pi d}\hat{\jmath} + \frac{\mu_0 I}{8\pi d}\hat{\jmath} = -\frac{\mu_0 I}{8\pi d}\hat{\jmath} \qquad \text{(point } P_1\text{)}$$

28.7 (a) Two long, straight conductors carrying equal currents in opposite directions. The conductors are seen end-on. (b) Map of the magnetic field produced by the two conductors. The field lines are closest together between the conductors, where the field is strongest.



At point P_2, a distance d from both wires, $\vec{B}_1$ and $\vec{B}_2$ are both in the positive y-direction, and both have the same magnitude $B_1 = B_2 = \mu_0 I/2\pi d$. Hence

$$\vec{B}_{\text{total}} = \vec{B}_1 + \vec{B}_2 = \frac{\mu_0 I}{2\pi d}\hat{\jmath} + \frac{\mu_0 I}{2\pi d}\hat{\jmath} = \frac{\mu_0 I}{\pi d}\hat{\jmath} \qquad \text{(point } P_2\text{)}$$

Finally, at point P_3 the right-hand rule shows that $\vec{B}_1$ is in the positive y-direction and $\vec{B}_2$ is in the negative y-direction. This point is a distance $3d$ from wire 1 and a distance d from wire 2, so $B_1 = \mu_0 I/2\pi(3d) = \mu_0 I/6\pi d$ and $B_2 = \mu_0 I/2\pi d$. The total field at P_3 is

$$\vec{B}_{\text{total}} = \vec{B}_1 + \vec{B}_2 = \frac{\mu_0 I}{6\pi d}\hat{\jmath} - \frac{\mu_0 I}{2\pi d}\hat{\jmath} = -\frac{\mu_0 I}{3\pi d}\hat{\jmath} \qquad \text{(point } P_3\text{)}$$

The same technique can be used to find $\vec{B}_{\text{total}}$ at any point; for points off the x-axis, caution must be taken in vector addition, since $\vec{B}_1$ and $\vec{B}_2$ need no longer be simply parallel or antiparallel.

(b) At any point on the x-axis to the right of wire 2 (that is, for $x > d$), $\vec{B}_1$ and $\vec{B}_2$ are in the same directions as at P_3. Such a point is a distance $x + d$ from wire 1 and a distance $x - d$ from wire 2, so the total field is

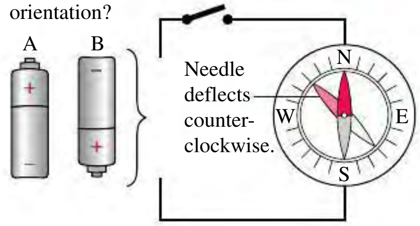
$$\vec{B}_{\text{total}} = \vec{B}_1 + \vec{B}_2 = \frac{\mu_0 I}{2\pi(x+d)}\hat{\jmath} - \frac{\mu_0 I}{2\pi(x-d)}\hat{\jmath}$$
$$= -\frac{\mu_0 I d}{\pi(x^2 - d^2)}\hat{\jmath}$$

where we used a common denominator to combine the two terms.

EVALUATE: Consider our result from part (b) at a point very far from the wires, so that x is much larger than d. Then the d^2 term in the denominator can be ignored, and the magnitude of the total field is approximately $B_{\text{total}} = \mu_0 I d/\pi x^2$. For one wire, Eq. (28.9) shows that the magnetic field decreases with distance in proportion to $1/x$; for two wires carrying opposite currents, $\vec{B}_1$ and $\vec{B}_2$ partially cancel each other, and so B_{total} decreases more rapidly, in proportion to $1/x^2$. This effect is used in communication systems such as telephone or computer networks. The wiring is arranged so that a conductor carrying a signal in one direction and the conductor carrying the return signal are side by side, as in Fig. 28.7a, or twisted around each other (**Fig. 28.8**). As a result, the magnetic field due to these signals *outside* the conductors is very small, making it less likely to exert unwanted forces on other information-carrying currents.

28.8 Computer cables, or cables for audio-video equipment, create little or no magnetic field. This is because within each cable, closely spaced wires carry current in both directions along the length of the cable. The magnetic fields from these opposing currents cancel each other.

TEST YOUR UNDERSTANDING OF SECTION 28.3 The accompanying figure shows a circuit that lies on a horizontal table. A compass is placed on top of the circuit as shown. A battery is to be connected to the circuit so that when the switch is closed, the compass needle deflects counterclockwise. In which orientation, A or B, should the battery be placed in the circuit? ❙



28.4 FORCE BETWEEN PARALLEL CONDUCTORS

Now that we know how to calculate the magnetic field produced by a long, current-carrying conductor, we can find the *magnetic force* that one such conductor exerts on another. This force plays a role in many practical situations in which current-carrying wires are close to each other. **Figure 28.9** shows segments of two long, straight, parallel conductors separated by a distance r and carrying currents I and I' in the same direction. Each conductor lies in the magnetic field set up by the other, so each experiences a force. The figure shows some of the field lines set up by the current in the lower conductor.

From Eq. (28.9) the lower conductor produces a $\vec{B}$ field that, at the position of the upper conductor, has magnitude

$$B = \frac{\mu_0 I}{2\pi r}$$

From Eq. (27.19) the force that this field exerts on a length L of the upper conductor is $\vec{F} = I'\vec{L} \times \vec{B}$, where the vector $\vec{L}$ is in the direction of the current I' and has magnitude L. Since $\vec{B}$ is perpendicular to the length of the conductor and hence to $\vec{L}$, the magnitude of this force is

$$F = I'LB = \frac{\mu_0 I I' L}{2\pi r}$$

and the force *per unit length* F/L is

Magnetic force per unit length between two long, parallel, current-carrying conductors:

$$\frac{F}{L} = \frac{\mu_0 I I'}{2\pi r} \qquad (28.11)$$

Magnetic constant (μ_0); Current in first conductor (I); Current in second conductor (I'); Distance between conductors (r)

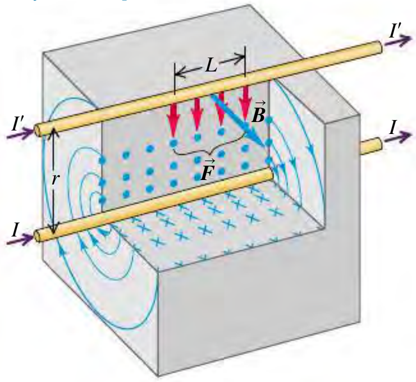
Applying the right-hand rule to $\vec{F} = I'\vec{L} \times \vec{B}$ shows that the force on the upper conductor is directed *downward*.

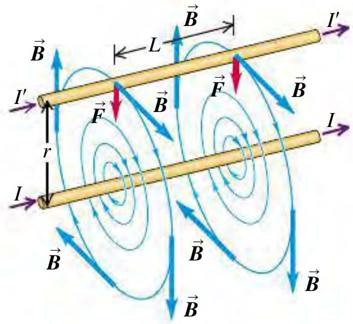
The current in the *upper* conductor also sets up a $\vec{B}$ field at the position of the *lower* conductor. Two successive applications of the right-hand rule for vector products (one to find the direction of the $\vec{B}$ field due to the upper conductor, as in Section 28.2, and one to find the direction of the force that this field exerts on the lower conductor, as in Section 27.6) show that the force on the lower conductor is *upward*. Thus *two parallel conductors carrying current in the same direction attract each other.* If the direction of either current is reversed, the forces also reverse. *Parallel conductors carrying currents in opposite directions repel each other.*

28.9 Parallel conductors carrying currents in the same direction attract each other. The diagrams show how the magnetic field $\vec{B}$ caused by the current in the lower conductor exerts a force $\vec{F}$ on the upper conductor.

The magnetic field of the lower wire exerts an attractive force on the upper wire. By the same token, the upper wire attracts the lower one.

If the wires had currents in *opposite* directions, they would *repel* each other.





Magnetic Forces and Defining the Ampere

The attraction or repulsion between two straight, parallel, current-carrying conductors is the basis of the official SI definition of the **ampere:**

***One ampere* is that unvarying current that, if present in each of two parallel conductors of infinite length and one meter apart in empty space, causes each conductor to experience a force of exactly 2×10^{-7} newtons per meter of length.**

From Eq. (28.11) you can see that this definition of the ampere is what leads us to choose the value of $4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$ for μ_0. The SI definition of the coulomb is the amount of charge transferred in one second by a current of one ampere.

This is an *operational definition;* it gives us an actual experimental procedure for measuring current and defining a unit of current. For high-precision standardization of the ampere, coils of wire are used instead of straight wires, and their separation is only a few centimeters. Even more precise measurements of the standardized ampere are possible with a version of the Hall effect (see Section 27.9).

Mutual forces of attraction exist not only between *wires* carrying currents in the same direction, but also between the elements of a single current-carrying conductor. If the conductor is a liquid or an ionized gas (a plasma), these forces result in a constriction of the conductor called the *pinch effect.* The pinch effect in a plasma has been used in one technique to bring about nuclear fusion.

EXAMPLE 28.5 FORCES BETWEEN PARALLEL WIRES

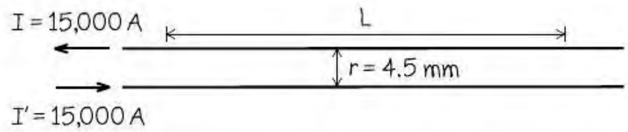
SOLUTION

Two straight, parallel, superconducting wires 4.5 mm apart carry equal currents of 15,000 A in opposite directions. What force, per unit length, does each wire exert on the other?

SOLUTION

IDENTIFY and SET UP: **Figure 28.10** shows the situation. We find F/L, the magnetic force per unit length of wire, from Eq. (28.11).

28.10 Our sketch for this problem.

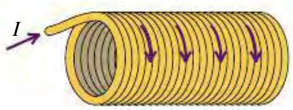


EXECUTE: The conductors *repel* each other because the currents are in opposite directions. From Eq. (28.11) the force per unit length is

$$\frac{F}{L} = \frac{\mu_0 I I'}{2\pi r} = \frac{(4\pi \times 10^{-7}\ \mathrm{T \cdot m/A})(15{,}000\ \mathrm{A})^2}{(2\pi)(4.5 \times 10^{-3}\ \mathrm{m})}$$

$$= 1.0 \times 10^4\ \mathrm{N/m}$$

EVALUATE: This is a large force, more than one ton per meter. Currents and separations of this magnitude are used in superconducting electromagnets in particle accelerators, and mechanical stress analysis is a crucial part of the design process.



TEST YOUR UNDERSTANDING OF SECTION 28.4 A solenoid is a wire wound into a helical coil. The accompanying figure shows a solenoid that carries a current I. (a) Is the *magnetic* force that one turn of the coil exerts on an adjacent turn (i) attractive, (ii) repulsive, or (iii) zero? (b) Is the *electric* force that one turn of the coil exerts on an adjacent turn (i) attractive, (ii) repulsive, or (iii) zero? (c) Is the *magnetic* force between opposite sides of the same turn of the coil (i) attractive, (ii) repulsive, or (iii) zero? (d) Is the *electric* force between opposite sides of the same turn of the coil (i) attractive, (ii) repulsive, or (iii) zero? ❙

28.11 This electromagnet contains a current-carrying coil with numerous turns of wire. The resulting magnetic field can pick up large quantities of steel bars and other iron-bearing items.

28.5 MAGNETIC FIELD OF A CIRCULAR CURRENT LOOP

If you look inside a doorbell, a transformer, an electric motor, or an electromagnet (**Fig. 28.11**), you will find coils of wire with a large number of turns, spaced so closely that each turn is very nearly a planar circular loop. A current in such a coil is used to establish a magnetic field. In Section 27.7 we considered the force and torque on such a current loop placed in an external magnetic field produced by other currents; we are now about to find the magnetic field produced by such a loop or by a collection of closely spaced loops forming a coil.

Figure 28.12 shows a circular conductor with radius a. A current I is led into and out of the loop through two long, straight wires side by side; the currents in these straight wires are in opposite directions, and their magnetic fields very nearly cancel each other (see Example 28.4 in Section 28.3).

We can use the law of Biot and Savart, Eq. (28.5) or (28.6), to find the magnetic field at a point P on the axis of the loop, at a distance x from the center. As the figure shows, $d\vec{l}$ and $\hat{r}$ are perpendicular, and the direction of the field $d\vec{B}$ caused by this particular element $d\vec{l}$ lies in the xy-plane. Since $r^2 = x^2 + a^2$, the magnitude dB of the field due to element $d\vec{l}$ is

$$dB = \frac{\mu_0 I}{4\pi} \frac{dl}{(x^2 + a^2)} \tag{28.12}$$

The components of the vector $d\vec{B}$ are

$$dB_x = dB\cos\theta = \frac{\mu_0 I}{4\pi} \frac{dl}{(x^2 + a^2)} \frac{a}{(x^2 + a^2)^{1/2}} \tag{28.13}$$

$$dB_y = dB\sin\theta = \frac{\mu_0 I}{4\pi} \frac{dl}{(x^2 + a^2)} \frac{x}{(x^2 + a^2)^{1/2}} \tag{28.14}$$

The *total* field $\vec{B}$ at P has only an x-component (it is perpendicular to the plane of the loop). Here's why: For every element $d\vec{l}$ there is a corresponding element on the opposite side of the loop, with opposite direction. These two elements give equal contributions to the x-component of $d\vec{B}$, given by Eq. (28.13), but *opposite* components perpendicular to the x-axis. Thus all the perpendicular components cancel and only the x-components survive.

To obtain the x-component of the total field $\vec{B}$, we integrate Eq. (28.13), including all the $d\vec{l}$'s around the loop. Everything in this expression except dl is constant and can be taken outside the integral, and we have

$$B_x = \int \frac{\mu_0 I}{4\pi} \frac{a\,dl}{(x^2 + a^2)^{3/2}} = \frac{\mu_0 I a}{4\pi(x^2 + a^2)^{3/2}} \int dl$$

The integral of dl is just the circumference of the circle, $\int dl = 2\pi a$, and so

Magnetic field on axis of a circular current-carrying loop

$$B_x = \frac{\mu_0 I a^2}{2(x^2 + a^2)^{3/2}} \tag{28.15}$$

Magnetic constant; Current; Radius of loop; Distance along axis from center of loop to field point

The *direction* of this magnetic field is given by a right-hand rule. If you curl the fingers of your right hand around the loop in the direction of the current, your right thumb points in the direction of the field (**Fig. 28.13**).

Magnetic Field on the Axis of a Coil

Now suppose that instead of the single loop in Fig. 28.12 we have a coil consisting of N loops, all with the same radius. The loops are closely spaced so that the plane of each loop is essentially the same distance x from the field point P. Then the total field is N times the field of a single loop:

$$B_x = \frac{\mu_0 N I a^2}{2(x^2 + a^2)^{3/2}} \quad \text{(on the axis of } N \text{ circular loops)} \tag{28.16}$$

The factor N in Eq. (28.16) is the reason coils of wire, not single loops, are used to produce strong magnetic fields; for a desired field strength, using a single loop might require a current I so great as to exceed the rating of the loop's wire.

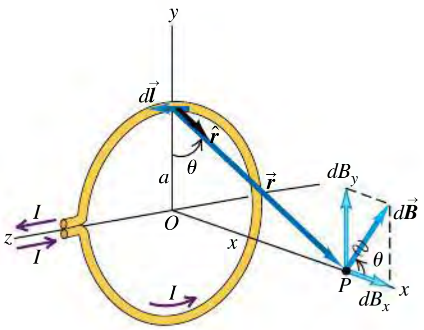
Figure 28.14 shows a graph of B_x as a function of x. The maximum value of the field is at $x = 0$, the center of the loop or coil:

Magnetic field at center of N circular current-carrying loops

$$B_x = \frac{\mu_0 N I}{2a} \tag{28.17}$$

Magnetic constant; Number of loops; Current; Radius of loop

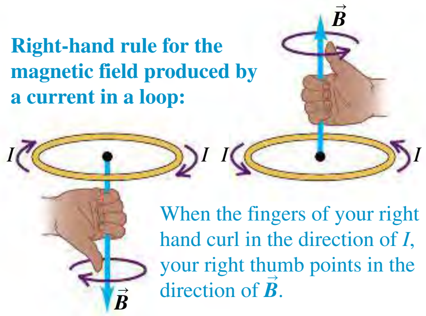
28.12 Magnetic field on the axis of a circular loop. The current in the segment $d\vec{l}$ causes the field $d\vec{B}$, which lies in the xy-plane. The currents in other $d\vec{l}$'s cause $d\vec{B}$'s with different components perpendicular to the x-axis; these components add to zero. The x-components of the $d\vec{B}$'s combine to give the total $\vec{B}$ field at point P.



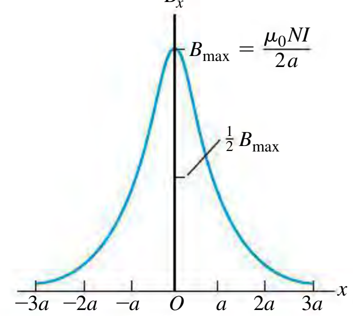
MP

PhET: Faraday's Electromagnetic Lab
PhET: Magnets and Electromagnets

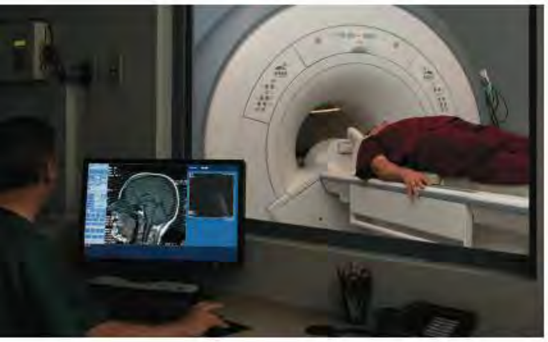
28.13 The right-hand rule for the direction of the magnetic field produced on the axis of a current-carrying coil.



28.14 Graph of the magnetic field along the axis of a circular coil with N turns. When x is much larger than a, the field magnitude decreases approximately as $1/x^3$.



BIO Application Magnetic Fields for MRI Magnetic resonance imaging (see Section 27.7), requires a magnetic field of about 1.5 T. In a typical MRI scan, the patient lies inside a coil that produces the intense field. The currents required are very high, so the coils are bathed in liquid helium at a temperature of 4.2 K to keep them from overheating.



In Section 27.7 we defined the *magnetic dipole moment* μ (or *magnetic moment*) of a current-carrying loop to be equal to IA, where A is the cross-sectional area of the loop. If there are N loops, the total magnetic moment is NIA. The circular loop in Fig. 28.12 has area $A = \pi a^2$, so the magnetic moment of a single loop is $\mu = I\pi a^2$; for N loops, $\mu = NI\pi a^2$. Substituting these results into Eqs. (28.15) and (28.16), we find

$$B_x = \frac{\mu_0 \mu}{2\pi(x^2 + a^2)^{3/2}} \quad \text{(on the axis of any number of circular loops)} \tag{28.18}$$

We described a magnetic dipole in Section 27.7 in terms of its response to a magnetic field produced by currents outside the dipole. But a magnetic dipole is also a *source* of magnetic field; Eq. (28.18) describes the magnetic field *produced* by a magnetic dipole for points along the dipole axis. This field is directly proportional to the magnetic dipole moment μ. Note that the field at all points on the x-axis is in the same direction as the vector magnetic moment $\vec{\mu}$.

CAUTION Magnetic field of a coil Equations (28.15), (28.16), and (28.18) are valid only on the *axis* of a loop or coil. Don't attempt to apply these equations at other points!

28.15 Magnetic field lines produced by the current in a circular loop. At points on the axis the $\vec{B}$ field has the same direction as the magnetic moment of the loop.

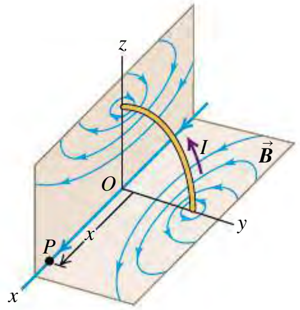


Figure 28.15 shows some of the magnetic field lines surrounding a circular current loop (magnetic dipole) in planes through the axis. The directions of the field lines are given by the same right-hand rule as for a long, straight conductor. Grab the conductor with your right hand, with your thumb in the direction of the current; your fingers curl around in the same direction as the field lines. The field lines for the circular current loop are closed curves that encircle the conductor; they are *not* circles, however.

EXAMPLE 28.6 MAGNETIC FIELD OF A COIL

SOLUTION

A coil consisting of 100 circular loops with radius 0.60 m carries a 5.0-A current. (a) Find the magnetic field at a point along the axis of the coil, 0.80 m from the center. (b) Along the axis, at what distance from the center of the coil is the field magnitude $\frac{1}{8}$ as great as it is at the center?

SOLUTION

IDENTIFY and SET UP: This problem concerns the magnetic field magnitude B along the axis of a current-carrying coil, so we can use Eq. (28.16). We are given $N = 100$, $I = 5.0$ A, and $a = 0.60$ m. In part (a) our target variable is B_x at a given value of x. In part (b) the target variable is the value of x at which the field has $\frac{1}{8}$ of the magnitude that it has at the origin.

EXECUTE: (a) Using $x = 0.80$ m, from Eq. (28.16) we have

$$B_x = \frac{(4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A})(100)(5.0\ \text{A})(0.60\ \text{m})^2}{2[(0.80\ \text{m})^2 + (0.60\ \text{m})^2]^{3/2}}$$
$$= 1.1 \times 10^{-4}\ \text{T}$$

(b) Considering Eq. (28.16), we seek a value of x such that

$$\frac{1}{(x^2 + a^2)^{3/2}} = \frac{1}{8}\frac{1}{(0^2 + a^2)^{3/2}}$$

To solve this for x, we take the reciprocal of the whole thing and then take the 2/3 power of both sides; the result is

$$x = \pm\sqrt{3}a = \pm 1.04\ \text{m}$$

EVALUATE: We check our answer in part (a) by finding the coil's magnetic moment and substituting the result into Eq. (28.18):

$$\mu = NI\pi a^2 = (100)(5.0\ \text{A})\pi(0.60\ \text{m})^2 = 5.7 \times 10^2\ \text{A}\cdot\text{m}^2$$

$$B_x = \frac{(4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A})(5.7 \times 10^2\ \text{A}\cdot\text{m}^2)}{2\pi[(0.80\ \text{m})^2 + (0.60\ \text{m})^2]^{3/2}} = 1.1 \times 10^{-4}\ \text{T}$$

The magnetic moment μ is relatively large, yet it produces a rather small field, comparable to that of the earth. This illustrates how difficult it is to produce strong fields of 1 T or more.

TEST YOUR UNDERSTANDING OF SECTION 28.5 Figure 28.12 shows the magnetic field $d\vec{B}$ produced at point P by a segment $d\vec{l}$ that lies on the positive y-axis (at the top of the loop). This field has components $dB_x > 0$, $dB_y > 0$, $dB_z = 0$. (a) What are the signs of the components of the field $d\vec{B}$ produced at P by a segment $d\vec{l}$ on the negative y-axis (at the bottom of the loop)? (i) $dB_x > 0$, $dB_y > 0$, $dB_z = 0$; (ii) $dB_x > 0$, $dB_y < 0$, $dB_z = 0$; (iii) $dB_x < 0$, $dB_y > 0$, $dB_z = 0$; (iv) $dB_x < 0$, $dB_y < 0$, $dB_z = 0$; (v) none of these. (b) What are the signs of the components of the field $d\vec{B}$ produced at P by a segment $d\vec{l}$ on the negative z-axis (at the right-hand side of the loop)? (i) $dB_x > 0$, $dB_y > 0$, $dB_z = 0$; (ii) $dB_x > 0$, $dB_y < 0$, $dB_z = 0$; (iii) $dB_x < 0$, $dB_y > 0$, $dB_z = 0$; (iv) $dB_x < 0$, $dB_y < 0$, $dB_z = 0$; (v) none of these.

28.6 AMPERE'S LAW

So far our calculations of the magnetic field due to a current have involved finding the infinitesimal field $d\vec{B}$ due to a current element and then summing all the $d\vec{B}$'s to find the total field. This approach is directly analogous to our *electric*-field calculations in Chapter 21.

For the electric-field problem we found that in situations with a highly symmetric charge distribution, it was often easier to use Gauss's law to find $\vec{E}$. There is likewise a law that allows us to more easily find the *magnetic* fields caused by highly symmetric *current* distributions. But the law that allows us to do this, called *Ampere's law,* is rather different in character from Gauss's law.

Gauss's law for electric fields (Chapter 22) involves the flux of $\vec{E}$ through a closed surface; it states that this flux is equal to the total charge enclosed within the surface, divided by the constant ϵ_0. Thus this law relates electric fields and charge distributions. By contrast, Gauss's law for *magnetic* fields, Eq. (28.10), is *not* a relationship between magnetic fields and current distributions; it states that the flux of $\vec{B}$ through *any* closed surface is always zero, whether or not there are currents within the surface. So Gauss's law for $\vec{B}$ can't be used to determine the magnetic field produced by a particular current distribution.

Ampere's law is formulated not in terms of magnetic flux, but rather in terms of the *line integral* of $\vec{B}$ around a closed path, denoted by

$$\oint \vec{B} \cdot d\vec{l}$$

We used line integrals to define work in Chapter 6 and to calculate electric potential in Chapter 23. To evaluate this integral, we divide the path into infinitesimal segments $d\vec{l}$, calculate the scalar product of $\vec{B} \cdot d\vec{l}$ for each segment, and sum these products. In general, $\vec{B}$ varies from point to point, and we must use the value of $\vec{B}$ at the location of each $d\vec{l}$. An alternative notation is $\oint B_{\parallel}\, dl$, where $B_{\parallel}$ is the component of $\vec{B}$ parallel to $d\vec{l}$ at each point. The circle on the integral sign indicates that this integral is always computed for a *closed* path, one whose beginning and end points are the same.

Ampere's Law for a Long, Straight Conductor

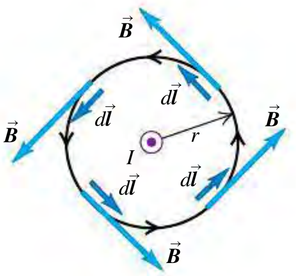
To introduce the basic idea of Ampere's law, let's consider again the magnetic field caused by a long, straight conductor carrying a current I. We found in Section 28.3 that the field at a distance r from the conductor has magnitude

$$B = \frac{\mu_0 I}{2\pi r}$$

28.16 Three integration paths for the line integral of $\vec{B}$ in the vicinity of a long, straight conductor carrying current I *out* of the plane of the page (as indicated by the circle with a dot). The conductor is seen end-on.

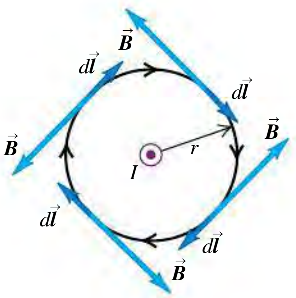
(a) Integration path is a circle centered on the conductor; integration goes around the circle counterclockwise.

Result: $\oint \vec{B} \cdot d\vec{l} = \mu_0 I$



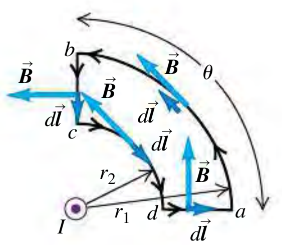
(b) Same integration path as in (a), but integration goes around the circle clockwise.

Result: $\oint \vec{B} \cdot d\vec{l} = -\mu_0 I$



(c) An integration path that does not enclose the conductor

Result: $\oint \vec{B} \cdot d\vec{l} = 0$



The magnetic field lines are circles centered on the conductor. Let's take the line integral of $\vec{B}$ around a circle with radius r, as in **Fig. 28.16a**. At every point on the circle, $\vec{B}$ and $d\vec{l}$ are parallel, and so $\vec{B} \cdot d\vec{l} = B\,dl$; since r is constant around the circle, B is constant as well. Alternatively, we can say that $B_\parallel$ is constant and equal to B at every point on the circle. Hence we can take B outside of the integral. The remaining integral is just the circumference of the circle, so

$$\oint \vec{B} \cdot d\vec{l} = \oint B_\parallel\, dl = B \oint dl = \frac{\mu_0 I}{2\pi r}(2\pi r) = \mu_0 I$$

The line integral is thus independent of the radius of the circle and is equal to μ_0 multiplied by the current passing through the area bounded by the circle.

In Fig. 28.16b the situation is the same, but the integration path now goes around the circle in the opposite direction. Now $\vec{B}$ and $d\vec{l}$ are antiparallel, so $\vec{B} \cdot d\vec{l} = -B\,dl$ and the line integral equals $-\mu_0 I$. We get the same result if the integration path is the same as in Fig. 28.16a, but the direction of the current is reversed. Thus $\oint \vec{B} \cdot d\vec{l}$ equals μ_0 multiplied by the current passing through the area bounded by the integration path, with a positive or negative sign depending on the direction of the current relative to the direction of integration.

There's a simple rule for the sign of the current; you won't be surprised to learn that it uses your right hand. Curl the fingers of your right hand around the integration path so that they curl in the direction of integration (that is, the direction that you use to evaluate $\oint \vec{B} \cdot d\vec{l}$). Then your right thumb indicates the positive current direction. Currents that pass through the integration path in this direction are positive; those in the opposite direction are negative. Using this rule, convince yourself that the current is positive in Fig. 28.16a and negative in Fig. 28.16b. Here's another way to say the same thing: Looking at the surface bounded by the integration path, integrate counterclockwise around the path as in Fig. 28.16a. Currents moving toward you through the surface are positive, and those going away from you are negative.

An integration path that does *not* enclose the conductor is used in Fig. 28.16c. Along the circular arc ab of radius r_1, $\vec{B}$ and $d\vec{l}$ are parallel, and $B_\parallel = B_1 = \mu_0 I/2\pi r_1$; along the circular arc cd of radius r_2, $\vec{B}$ and $d\vec{l}$ are antiparallel and $B_\parallel = -B_2 = -\mu_0 I/2\pi r_2$. The $\vec{B}$ field is perpendicular to $d\vec{l}$ at each point on the straight sections bc and da, so $B_\parallel = 0$ and these sections contribute zero to the line integral. The total line integral is then

$$\oint \vec{B} \cdot d\vec{l} = \oint B_\parallel\, dl = B_1 \int_a^b dl + (0)\int_b^c dl + (-B_2)\int_c^d dl + (0)\int_d^a dl$$

$$= \frac{\mu_0 I}{2\pi r_1}(r_1\theta) + 0 - \frac{\mu_0 I}{2\pi r_2}(r_2\theta) + 0 = 0$$

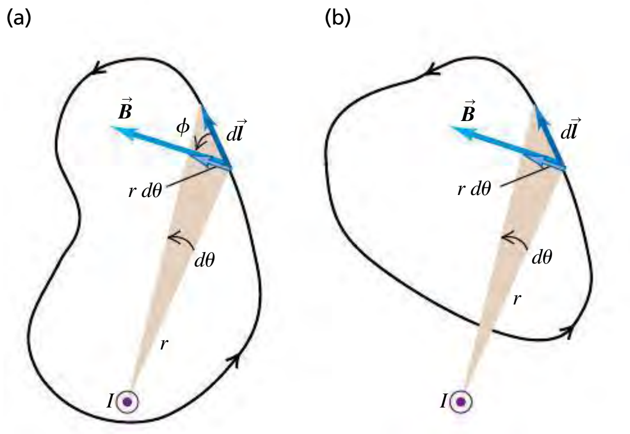
The magnitude of $\vec{B}$ is greater on arc cd than on arc ab, but the arc length is less, so the contributions from the two arcs exactly cancel. Even though there is a magnetic field everywhere along the integration path, the line integral $\oint \vec{B} \cdot d\vec{l}$ is zero if there is no current passing through the area bounded by the path.

We can also derive these results for more general integration paths, such as the one in **Fig. 28.17a**. At the position of the line element $d\vec{l}$, the angle between $d\vec{l}$ and $\vec{B}$ is ϕ, and

$$\vec{B} \cdot d\vec{l} = B\,dl\cos\phi$$

From the figure, $dl\cos\phi = r\,d\theta$, where $d\theta$ is the angle subtended by $d\vec{l}$ at the position of the conductor and r is the distance of $d\vec{l}$ from the conductor. Thus

$$\oint \vec{B} \cdot d\vec{l} = \oint \frac{\mu_0 I}{2\pi r}(r\,d\theta) = \frac{\mu_0 I}{2\pi}\oint d\theta$$



28.17 (a) A more general integration path for the line integral of $\vec{B}$ around a long, straight conductor carrying current I *out* of the plane of the page. The conductor is seen end-on. (b) A more general integration path that does not enclose the conductor.

But $\oint d\theta$ is just equal to 2π, the total angle swept out by the radial line from the conductor to $d\vec{l}$ during a complete trip around the path. So we get

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 I \tag{28.19}$$

This result doesn't depend on the shape of the path or on the position of the wire inside it. If the current in the wire is opposite to that shown, the integral has the opposite sign. But if the path doesn't enclose the wire (Fig. 28.17b), then the net change in θ during the trip around the integration path is zero; $\oint d\theta$ is zero instead of 2π and the line integral is zero.

Ampere's Law: General Statement

We can generalize Ampere's law even further. Suppose *several* long, straight conductors pass through the surface bounded by the integration path. The total magnetic field $\vec{B}$ at any point on the path is the vector sum of the fields produced by the individual conductors. Thus the line integral of the total $\vec{B}$ equals μ_0 times the *algebraic sum* of the currents. In calculating this sum, we use the sign rule for currents described above. If the integration path does not enclose a particular wire, the line integral of the $\vec{B}$ field of that wire is zero, because the angle θ for that wire sweeps through a net change of zero rather than 2π during the integration. Any conductors present that are not enclosed by a particular path may still contribute to the value of $\vec{B}$ at every point, but the *line integrals* of their fields around the path are zero.

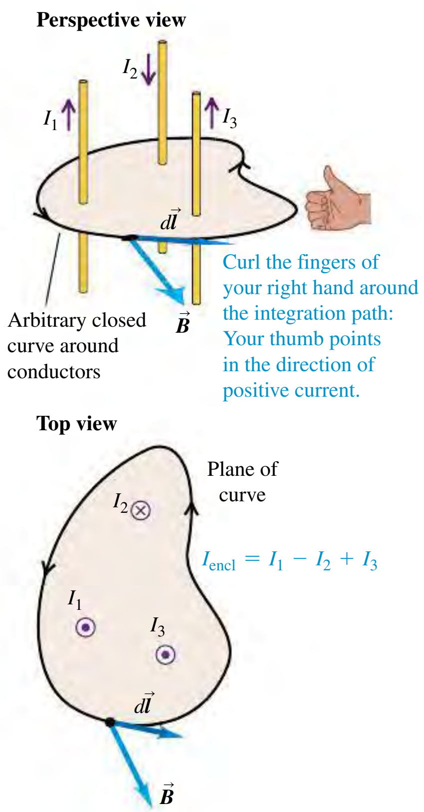
Thus we can replace I in Eq. (28.19) with I_{encl}, the algebraic sum of the currents *enclosed* or *linked* by the integration path, with the sum evaluated by using the sign rule just described (**Fig. 28.18**). Then **Ampere's law** says

Ampere's law:
$$\underbrace{\oint}_{\text{Line integral around a closed path}} \underbrace{\vec{B} \cdot d\vec{l}}_{\text{Scalar product of magnetic field and vector segment of path}} = \underbrace{\mu_0}_{\text{Magnetic constant}} \underbrace{I_{encl}}_{\text{Net current enclosed by path}} \tag{28.20}$$

While we have derived Ampere's law only for the special case of the field of several long, straight, parallel conductors, Eq. (28.20) is in fact valid for conductors and paths of *any* shape. The general derivation is no different in principle from what we have presented, but the geometry is more complicated.

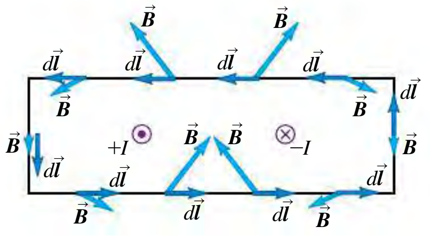
If $\oint \vec{B} \cdot d\vec{l} = 0$, it *does not* necessarily mean that $\vec{B} = \mathbf{0}$ everywhere along the path, only that the total current through an area bounded by the path is zero. In Figs. 28.16c and 28.17b, the integration paths enclose no current at all; in

28.18 Ampere's law.



Ampere's law: If we calculate the line integral of the magnetic field around a closed curve, the result equals μ_0 times the total enclosed current: $\oint \vec{B} \cdot d\vec{l} = \mu_0 I_{encl}$.

28.19 Two long, straight conductors carrying equal currents in opposite directions. The conductors are seen end-on, and the integration path is counterclockwise. The line integral $\oint \vec{B} \cdot d\vec{l}$ gets zero contribution from the upper and lower segments, a positive contribution from the left segment, and a negative contribution from the right segment; the net integral is zero.



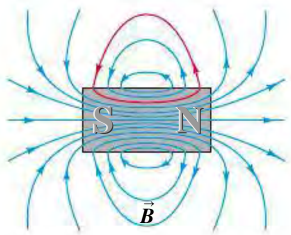


Fig. 28.19 there are positive and negative currents of equal magnitude through the area enclosed by the path. In both cases, $I_{encl} = 0$ and the line integral is zero.

CAUTION **Line integrals of electric and magnetic fields** In Chapter 23 we saw that the line integral of the electrostatic field $\vec{E}$ around any closed path is equal to zero; this is a statement that the electrostatic force $\vec{F} = q\vec{E}$ on a point charge q is conservative, so this force does zero work on a charge that moves around a closed path that returns to the starting point. The value of the line integral $\oint \vec{B} \cdot d\vec{l}$ is not similarly related to the question of whether the *magnetic* force is conservative. Remember that the magnetic force $\vec{F} = q\vec{v} \times \vec{B}$ on a moving charged particle is always *perpendicular* to $\vec{B}$, so $\oint \vec{B} \cdot d\vec{l}$ is *not* related to the work done by the magnetic force; as stated in Ampere's law, this integral is related only to the total current through a surface bounded by the integration path. In fact, the magnetic force on a moving charged particle is *not* conservative. A conservative force depends on only the position of the body on which the force is exerted, but the magnetic force on a moving charged particle also depends on the *velocity* of the particle.

Equation (28.20) turns out to be valid only if the currents are steady and if no magnetic materials or time-varying electric fields are present. In Chapter 29 we will see how to generalize Ampere's law for time-varying fields.

TEST YOUR UNDERSTANDING OF SECTION 28.6 The accompanying figure shows magnetic field lines through the center of a permanent magnet. The magnet is not connected to a source of emf. One of the field lines is colored red. What can you conclude about the currents inside the permanent magnet within the region enclosed by this field line? (i) There are no currents inside the magnet; (ii) there are currents directed out of the plane of the page; (iii) there are currents directed into the plane of the page; (iv) not enough information is given to decide.

28.7 APPLICATIONS OF AMPERE'S LAW

Ampere's law is useful when we can exploit symmetry to evaluate the line integral of $\vec{B}$. Several examples are given below. Problem-Solving Strategy 28.2 is directly analogous to Problem-Solving Strategy 22.1 (Section 22.4) for applications of Gauss's law; we suggest you review that strategy now and compare the two methods.

PROBLEM-SOLVING STRATEGY 28.2 AMPERE'S LAW

IDENTIFY *the relevant concepts:* Like Gauss's law, Ampere's law is most useful when the magnetic field is highly symmetric. In the form $\oint \vec{B} \cdot d\vec{l} = \mu_0 I_{encl}$, it can yield the magnitude of $\vec{B}$ as a function of position if we are given the magnitude and direction of the field-generating electric current.

SET UP *the problem* using the following steps:

1. Determine the target variable(s). Usually one will be the magnitude of the $\vec{B}$ field as a function of position.
2. Select the integration path you will use with Ampere's law. If you want to determine the magnetic field at a certain point, then the path must pass through that point. The integration path doesn't have to be any actual physical boundary. Usually it is a purely geometric curve; it may be in empty space, embedded in a solid body, or some of each. The integration path has to have enough *symmetry* to make evaluation of the integral possible. Ideally the path will be tangent to $\vec{B}$ in regions of interest; elsewhere the path should be perpendicular to $\vec{B}$ or should run through regions in which $\vec{B} = \mathbf{0}$.

EXECUTE *the solution* as follows:

1. Carry out the integral $\oint \vec{B} \cdot d\vec{l}$ along the chosen path. If $\vec{B}$ is tangent to all or some portion of the path and has the same magnitude B at every point, then its line integral is the product of B and the length of that portion of the path. If $\vec{B}$ is perpendicular to some portion of the path, or if $\vec{B} = \mathbf{0}$, that portion makes no contribution to the integral.
2. In the integral $\oint \vec{B} \cdot d\vec{l}$, $\vec{B}$ is the *total* magnetic field at each point on the path; it can be caused by currents enclosed *or not enclosed* by the path. If *no* net current is enclosed by the path, the field at points on the path need not be zero, but the integral $\oint \vec{B} \cdot d\vec{l}$ is always zero.
3. Determine the current I_{encl} enclosed by the integration path. A right-hand rule gives the sign of this current: If you curl the fingers of your right hand so that they follow the path in the direction of integration, then your right thumb points in the direction of positive current. If $\vec{B}$ is tangent to the path everywhere and I_{encl} is positive, the direction of $\vec{B}$ is the same as the direction of integration. If instead I_{encl} is negative, $\vec{B}$ is in the direction opposite to that of the integration.
4. Use Ampere's law $\oint \vec{B} \cdot d\vec{l} = \mu_0 I$ to solve for the target variable.

EVALUATE *your answer:* If your result is an expression for the field magnitude as a function of position, check it by examining how the expression behaves in different limits.

SOLUTION

EXAMPLE 28.7 FIELD OF A LONG, STRAIGHT, CURRENT-CARRYING CONDUCTOR

In Section 28.6 we derived Ampere's law from Eq. (28.9) for the field $\vec{B}$ of a long, straight, current-carrying conductor. Reverse this process, and use Ampere's law to find $\vec{B}$ for this situation.

SOLUTION

IDENTIFY and SET UP: The situation has cylindrical symmetry, so in Ampere's law we take our integration path to be a circle with radius r centered on the conductor and lying in a plane perpendicular to it, as in Fig. 28.16a. The field $\vec{B}$ is everywhere tangent to this circle and has the same magnitude B everywhere on the circle.

EXECUTE: With our choice of the integration path, Ampere's law [Eq. (28.20)] becomes

$$\oint \vec{B} \cdot d\vec{l} = \oint B_{\parallel}\, dl = B(2\pi r) = \mu_0 I$$

Equation (28.9), $B = \mu_0 I/2\pi r$, follows immediately.

Ampere's law determines the direction of $\vec{B}$ as well as its magnitude. Since we chose to go counterclockwise around the integration path, the positive direction for current is out of the plane of Fig. 28.16a; this is the same as the actual current direction in the figure, so I is positive and the integral $\oint \vec{B} \cdot d\vec{l}$ is also positive. Since the $d\vec{l}$'s run counterclockwise, the direction of $\vec{B}$ must be counterclockwise as well, as shown in Fig. 28.16a.

EVALUATE: Our results are consistent with those in Section 28.6.

SOLUTION

EXAMPLE 28.8 FIELD OF A LONG CYLINDRICAL CONDUCTOR

A cylindrical conductor with radius R carries a current I (**Fig. 28.20**). The current is uniformly distributed over the cross-sectional area of the conductor. Find the magnetic field as a function of the distance r from the conductor axis for points both inside $(r < R)$ and outside $(r > R)$ the conductor.

SOLUTION

IDENTIFY and SET UP: As in Example 28.7, the current distribution has cylindrical symmetry, and the magnetic field lines must be circles concentric with the conductor axis. To find the magnetic field inside and outside the conductor, we choose circular integration paths with radii $r < R$ and $r > R$, respectively (see Fig. 28.20).

EXECUTE: In either case the field $\vec{B}$ has the same magnitude at every point on the circular integration path and is tangent to the path. Thus the magnitude of the line integral is simply $B(2\pi r)$. To find the current I_{encl} enclosed by a circular integration path inside the conductor $(r < R)$, note that the current density (current per unit area) is $J = I/\pi R^2$ so $I_{encl} = J(\pi r^2) = Ir^2/R^2$. Hence Ampere's law gives $B(2\pi r) = \mu_0 I r^2/R^2$, or

$$B = \frac{\mu_0 I}{2\pi}\frac{r}{R^2} \quad \text{(inside the conductor, } r < R\text{)} \tag{28.21}$$

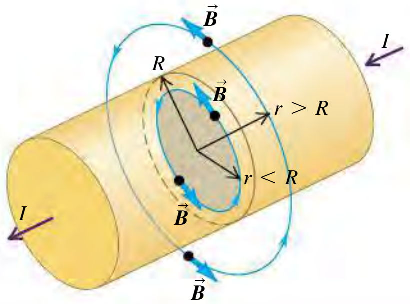
A circular integration path outside the conductor encloses the total current in the conductor, so $I_{encl} = I$. Applying Ampere's law gives the same equation as in Example 28.7, with the same result for B:

$$B = \frac{\mu_0 I}{2\pi r} \quad \text{(outside the conductor, } r > R\text{)} \tag{28.22}$$

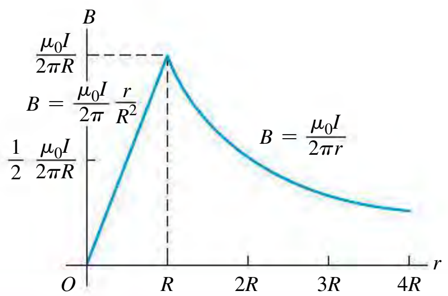
Outside the conductor, the magnetic field is the same as that of a long, straight conductor carrying current I, independent of the radius R over which the current is distributed. Indeed, the magnetic field outside *any* cylindrically symmetric current distribution is the same as if the entire current were concentrated along the axis of the distribution. This is analogous to the results of Examples 22.5 and 22.9 (Section 22.4), in which we found that the *electric* field outside a spherically symmetric *charged* body is the same as though the entire charge were concentrated at the center.

EVALUATE: At the surface of the conductor $(r = R)$, Eqs. (28.21) and (28.22) agree, as they must. **Figure 28.21** shows a graph of B as a function of r.

28.20 To find the magnetic field at radius $r < R$, we apply Ampere's law to the circle enclosing the gray area. The current through the gray area is $(r^2/R^2)I$. To find the magnetic field at radius $r > R$, we apply Ampere's law to the circle enclosing the entire conductor.



28.21 Magnitude of the magnetic field inside and outside a long, straight cylindrical conductor with radius R carrying a current I.



EXAMPLE 28.9 FIELD OF A SOLENOID

A solenoid consists of a helical winding of wire on a cylinder, usually circular in cross section. There can be thousands of closely spaced turns (often in several layers), each of which can be regarded as a circular loop. For simplicity, **Fig. 28.22** shows a solenoid with only a few turns. All turns carry the same current I, and the total $\vec{B}$ field at every point is the vector sum of the fields caused by the individual turns. The figure shows field lines in the xy- and xz-planes. We draw field lines that are uniformly spaced at the center of the solenoid. Exact calculations show that for a long, closely wound solenoid, half of these field lines emerge from the ends and half "leak out" through the windings between the center and the end, as the figure suggests.

If the solenoid is long in comparison with its cross-sectional diameter and the coils are tightly wound, the field inside the solenoid near its midpoint is very nearly uniform over the cross section and parallel to the axis; the *external* field near the midpoint is very small.

Use Ampere's law to find the field at or near the center of such a solenoid if it has n turns per unit length and carries current I.

SOLUTION

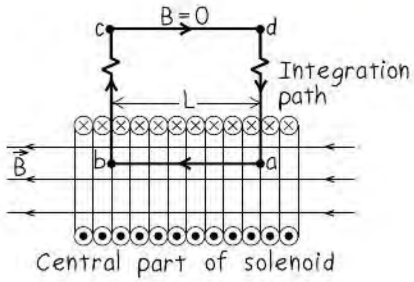
IDENTIFY and SET UP: We assume that $\vec{B}$ is uniform inside the solenoid and zero outside. **Figure 28.23** shows the situation and our chosen integration path, rectangle $abcd$. Side ab, with length L, is parallel to the axis of the solenoid. Sides bc and da are taken to be very long so that side cd is far from the solenoid; then the field at side cd is negligibly small.

EXECUTE: Along side ab, $\vec{B}$ is parallel to the path and is constant. Our Ampere's-law integration takes us along side ab in the same direction as $\vec{B}$, so here $B_{\parallel} = +B$ and

$$\int_a^b \vec{B} \cdot d\vec{l} = BL$$

Along sides bc and da, $\vec{B}$ is perpendicular to the path, and so $B_{\parallel} = 0$; along side cd, $\vec{B} = \mathbf{0}$ and so $B_{\parallel} = 0$. Around the entire closed path, then, we have $\oint \vec{B} \cdot d\vec{l} = BL$.

28.23 Our sketch for this problem.



In a length L there are nL turns, each of which passes once through $abcd$ carrying current I. Hence the total current enclosed by the rectangle is $I_{\text{encl}} = nLI$. The integral $\oint \vec{B} \cdot d\vec{l}$ is positive, so from Ampere's law I_{encl} must be positive as well. This means that the current passing through the surface bounded by the integration path must be in the direction shown in Fig. 28.23. Ampere's law then gives $BL = \mu_0 nLI$, or

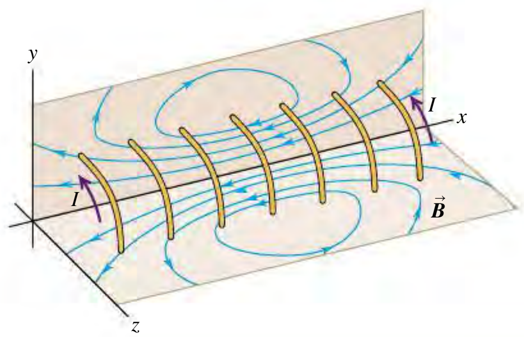
$$B = \mu_0 nI \qquad \text{(solenoid)} \qquad (28.23)$$

Side ab need not lie on the axis of the solenoid, so this result demonstrates that the field is uniform over the entire cross section at the center of the solenoid's length.

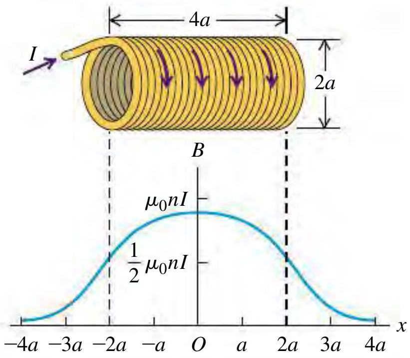
EVALUATE: Note that the *direction* of $\vec{B}$ inside the solenoid is in the same direction as the solenoid's vector magnetic moment $\vec{\mu}$, as we found in Section 28.5 for a single current-carrying loop.

For points along the axis, the field is strongest at the center of the solenoid and drops off near the ends. For a solenoid very long in comparison to its diameter, the field magnitude at each end is exactly half that at the center. This is approximately the case even for a relatively short solenoid, as **Fig. 28.24** shows.

28.22 Magnetic field lines produced by the current in a solenoid. For clarity, only a few turns are shown.



28.24 Magnitude of the magnetic field at points along the axis of a solenoid with length $4a$, equal to four times its radius a. The field magnitude at each end is about half its value at the center. (Compare with Fig. 28.14 for the field of N circular loops.)

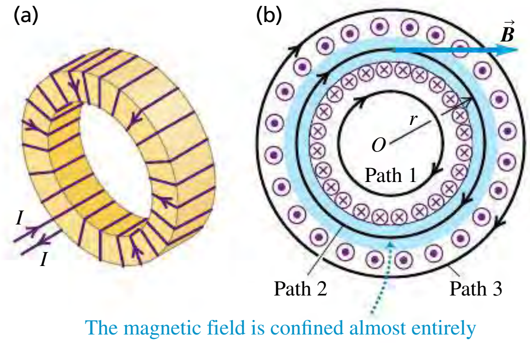


SOLUTION

EXAMPLE 28.10 FIELD OF A TOROIDAL SOLENOID

Figure 28.25a shows a doughnut-shaped **toroidal solenoid,** tightly wound with N turns of wire carrying a current I. (In a practical solenoid the turns would be much more closely spaced than they are in the figure.) Find the magnetic field at all points.

28.25 (a) A toroidal solenoid. For clarity, only a few turns of the winding are shown. (b) Integration paths (black circles) used to compute the magnetic field $\vec{B}$ set up by the current (shown as dots and crosses).



SOLUTION

IDENTIFY and SET UP: Ignoring the slight pitch of the helical windings, we can consider each turn of a tightly wound toroidal solenoid as a loop lying in a plane perpendicular to the large, circular axis of the toroid. The symmetry of the situation then tells us that the magnetic field lines must be circles concentric with the toroid axis. Therefore we choose circular integration paths (of which Fig. 28.25b shows three) for use with Ampere's law, so that the field $\vec{B}$ (if any) is tangent to each path at all points along the path.

EXECUTE: Along each path, $\oint \vec{B} \cdot d\vec{l}$ equals the product of B and the path circumference $l = 2\pi r$. The total current enclosed by path 1 is zero, so from Ampere's law the field $\vec{B} = \mathbf{0}$ everywhere on this path.

Each turn of the winding passes *twice* through the area bounded by path 3, carrying equal currents in opposite directions. The *net* current enclosed is therefore zero, and hence $\vec{B} = \mathbf{0}$ at all points on this path as well. We conclude that *the field of an ideal toroidal solenoid is confined to the space enclosed by the windings.* We can think of such a solenoid as a tightly wound, straight solenoid that has been bent into a circle.

For path 2, we have $\oint \vec{B} \cdot d\vec{l} = 2\pi r B$. Each turn of the winding passes *once* through the area bounded by this path, so $I_{\text{encl}} = NI$. We note that I_{encl} is positive for the clockwise direction of integration in Fig. 28.25b, so $\vec{B}$ is in the direction shown. Ampere's law then says that $2\pi r B = \mu_0 NI$, so

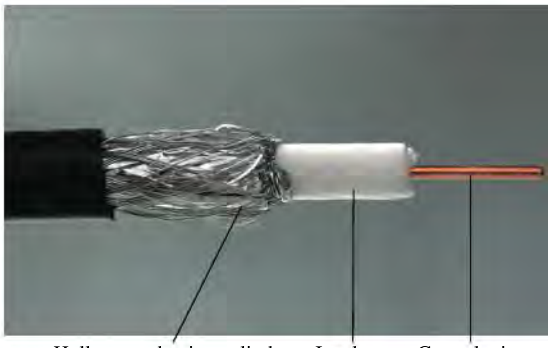
$$B = \frac{\mu_0 NI}{2\pi r} \quad \text{(toroidal solenoid)} \qquad (28.24)$$

EVALUATE: Equation (28.24) indicates that B is *not* uniform over the interior of the core, because different points in the interior are different distances r from the toroid axis. However, if the radial extent of the core is small in comparison to r, the variation is slight. In that case, considering that $2\pi r$ is the circumferential length of the toroid and that $N/2\pi r$ is the number of turns per unit length n, the field may be written as $B = \mu_0 nI$, just as it is at the center of a long, *straight* solenoid.

In a real toroidal solenoid the turns are not precisely circular loops but rather segments of a bent helix. As a result, the external field is not exactly zero. To estimate its magnitude, we imagine Fig. 28.25a as being *very* roughly equivalent, for points outside the torus, to a *single-turn* circular loop with radius r. At the center of such a loop, Eq. (28.17) gives $B = \mu_0 I/2r$; this is smaller than the field inside the solenoid by the factor N/π.

The equations we have derived for the field in a closely wound straight or toroidal solenoid are strictly correct only for windings in *vacuum*. For most practical purposes, however, they can be used for windings in air or on a core of any nonmagnetic, nonsuperconducting material. In the next section we will show how these equations are modified if the core is a magnetic material.

TEST YOUR UNDERSTANDING OF SECTION 28.7 Consider a conducting wire that runs along the central axis of a hollow conducting cylinder. Such an arrangement, called a *coaxial cable*, has many applications in telecommunications. (The cable that connects a television set to a local cable provider is an example of a coaxial cable.) In such a cable a current I runs in one direction along the hollow conducting cylinder and is spread uniformly over the cylinder's cross-sectional area. An equal current runs in the opposite direction along the central wire. How does the magnitude B of the magnetic field outside such a cable depend on the distance r from the central axis of the cable? (i) B is proportional to $1/r$; (ii) B is proportional to $1/r^2$; (iii) B is zero at all points outside the cable. ❙



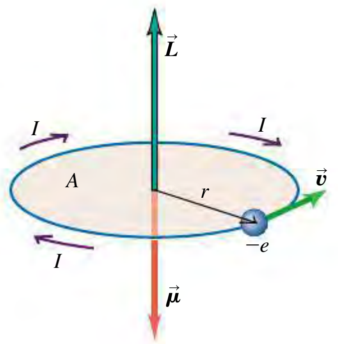
28.8 MAGNETIC MATERIALS

In discussing how currents cause magnetic fields, we have assumed that the conductors are surrounded by vacuum. But the coils in transformers, motors, generators, and electromagnets nearly always have iron cores to increase the magnetic field and confine it to desired regions. Permanent magnets, magnetic recording tapes, and computer disks depend directly on the magnetic properties of materials; when you store information on a computer disk, you are actually setting up an array of microscopic permanent magnets on the disk. So it is worthwhile to examine some aspects of the magnetic properties of materials. After describing the atomic origins of magnetic properties, we will discuss three broad classes of magnetic behavior that occur in materials; these are called *paramagnetism, diamagnetism,* and *ferromagnetism.*

The Bohr Magneton

As we discussed briefly in Section 27.7, the atoms that make up all matter contain moving electrons, and these electrons form microscopic current loops that produce magnetic fields of their own. In many materials these currents are randomly oriented and cause no net magnetic field. But in some materials an external field (a field produced by currents outside the material) can cause these loops to become oriented preferentially with the field, so their magnetic fields *add* to the external field. We then say that the material is *magnetized.*

28.26 An electron moving with speed v in a circular orbit of radius r has an angular momentum $\vec{L}$ and an oppositely directed orbital magnetic dipole moment $\vec{\mu}$. It also has a spin angular momentum and an oppositely directed spin magnetic dipole moment.



Let's look at how these microscopic currents come about. **Figure 28.26** shows a primitive model of an electron in an atom. We picture the electron (mass m, charge $-e$) as moving in a circular orbit with radius r and speed v. This moving charge is equivalent to a current loop. In Section 27.7 we found that a current loop with area A and current I has a magnetic dipole moment μ given by $\mu = IA$; for the orbiting electron the area of the loop is $A = \pi r^2$. To find the current associated with the electron, we note that the orbital period T (the time for the electron to make one complete orbit) is the orbit circumference divided by the electron speed: $T = 2\pi r/v$. The equivalent current I is the total charge passing any point on the orbit per unit time, which is just the magnitude e of the electron charge divided by the orbital period T:

$$I = \frac{e}{T} = \frac{ev}{2\pi r}$$

The magnetic moment $\mu = IA$ is then

$$\mu = \frac{ev}{2\pi r}(\pi r^2) = \frac{evr}{2} \tag{28.25}$$

It is useful to express μ in terms of the *angular momentum* L of the electron. For a particle moving in a circular path, the magnitude of angular momentum equals the magnitude of momentum mv multiplied by the radius r—that is, $L = mvr$ (see Section 10.5). Comparing this with Eq. (28.25), we can write

$$\mu = \frac{e}{2m}L \tag{28.26}$$

Equation (28.26) is useful in this discussion because atomic angular momentum is *quantized;* its component in a particular direction is always an integer multiple of $h/2\pi$, where h is a fundamental physical constant called *Planck's constant.* The numerical value of h is

$$h = 6.626 \times 10^{-34}\ \text{J}\cdot\text{s}$$

The quantity $h/2\pi$ thus represents a fundamental unit of angular momentum in atomic systems, just as e is a fundamental unit of charge. Associated with the quantization of $\vec{L}$ is a fundamental uncertainty in the *direction* of $\vec{L}$ and therefore of $\vec{\mu}$. In the following discussion, when we speak of the magnitude of a magnetic moment, a more precise statement would be "maximum component in a given direction." Thus, to say that a magnetic moment $\vec{\mu}$ is aligned with a magnetic field $\vec{B}$ really means that $\vec{\mu}$ has its maximum possible component in the direction of $\vec{B}$; such components are always quantized.

Equation (28.26) shows that associated with the fundamental unit of angular momentum is a corresponding fundamental unit of magnetic moment. If $L = h/2\pi$, then

$$\mu = \frac{e}{2m}\left(\frac{h}{2\pi}\right) = \frac{eh}{4\pi m} \tag{28.27}$$

DATA *SPEAKS*

Magnetic Fields and Their Sources

When students were given a problem involving currents and the magnetic forces they produce, more than 19% gave an incorrect response. Common errors:

- Forgetting that magnetic fields add according to the laws of vector addition. When two currents or current elements produce a $\vec{B}$ field at a certain point, the net $\vec{B}$ at that point is the *vector* sum of the individual fields.
- Misapplying Ampere's law. You can use Ampere's law to find the magnetic-field magnitude *only* if there is a closed loop over which $\vec{B}$ has a constant magnitude and is directed tangent to the loop.

This quantity is called the **Bohr magneton,** denoted by μ_B. Its numerical value is

$$\mu_B = 9.274 \times 10^{-24}\ \text{A}\cdot\text{m}^2 = 9.274 \times 10^{-24}\ \text{J/T}$$

You should verify that these two sets of units are consistent. The second set is useful when we compute the potential energy $U = -\vec{\boldsymbol{\mu}} \cdot \vec{\boldsymbol{B}}$ for a magnetic moment in a magnetic field.

Electrons also have an intrinsic angular momentum, called *spin,* that is not related to orbital motion but that can be pictured in a classical model as spinning on an axis. This angular momentum also has an associated magnetic moment, and its magnitude turns out to be almost exactly one Bohr magneton. (Effects having to do with quantization of the electromagnetic field cause the spin magnetic moment to be about 1.001 μ_B.)

Paramagnetism

In an atom, most of the various orbital and spin magnetic moments of the electrons add up to zero. However, in some cases the atom has a net magnetic moment that is of the order of μ_B. When such a material is placed in a magnetic field, the field exerts a torque on each magnetic moment, as given by Eq. (27.26): $\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\boldsymbol{B}}$. These torques tend to align the magnetic moments with the field, as we discussed in Section 27.7. In this position, the directions of the current loops are such as to *add* to the externally applied magnetic field.

We saw in Section 28.5 that the $\vec{\boldsymbol{B}}$ field produced by a current loop is proportional to the loop's magnetic dipole moment. In the same way, the additional $\vec{\boldsymbol{B}}$ field produced by microscopic electron current loops is proportional to the total magnetic moment $\vec{\boldsymbol{\mu}}_{\text{total}}$ per unit volume V in the material. We call this vector quantity the **magnetization** of the material, denoted by $\vec{\boldsymbol{M}}$:

$$\vec{\boldsymbol{M}} = \frac{\vec{\boldsymbol{\mu}}_{\text{total}}}{V} \tag{28.28}$$

The additional magnetic field due to magnetization of the material turns out to be equal simply to $\mu_0\vec{\boldsymbol{M}}$, where μ_0 is the same constant that appears in the law of Biot and Savart and Ampere's law. When such a material completely surrounds a current-carrying conductor, the total magnetic field $\vec{\boldsymbol{B}}$ in the material is

$$\vec{\boldsymbol{B}} = \vec{\boldsymbol{B}}_0 + \mu_0\vec{\boldsymbol{M}} \tag{28.29}$$

where $\vec{\boldsymbol{B}}_0$ is the field caused by the current in the conductor.

To check that the units in Eq. (28.29) are consistent, note that magnetization $\vec{\boldsymbol{M}}$ is the magnetic moment per unit volume. The units of magnetic moment are current times area $(\text{A}\cdot\text{m}^2)$, so the units of magnetization are $(\text{A}\cdot\text{m}^2)/\text{m}^3 = \text{A/m}$. From Section 28.1, the units of the constant μ_0 are $\text{T}\cdot\text{m/A}$. So the units of $\mu_0\vec{\boldsymbol{M}}$ are the same as the units of $\vec{\boldsymbol{B}}$: $(\text{T}\cdot\text{m/A})(\text{A/m}) = \text{T}$.

A material showing the behavior just described is said to be **paramagnetic.** The result is that the magnetic field at any point in such a material is greater by a dimensionless factor K_m, called the **relative permeability** of the material, than it would be if the material were replaced by vacuum. The value of K_m is different for different materials; for common paramagnetic solids and liquids at room temperature, K_m typically ranges from 1.00001 to 1.003.

All of the equations in this chapter that relate magnetic fields to their sources can be adapted to the situation in which the current-carrying conductor is embedded in a paramagnetic material. All that need be done is to replace μ_0 by $K_m\mu_0$. This product is usually denoted as μ and is called the **permeability** of the material:

$$\mu = K_m\mu_0 \tag{28.30}$$

CAUTION Two meanings of the symbol μ Equation (28.30) involves dangerous notation because we use μ for magnetic dipole moment as well as for permeability, as is customary. But beware: From now on, every time you see a μ, make sure you know whether it is permeability or magnetic moment. You can usually tell from the context.

TABLE 28.1 Magnetic Susceptibilities of Paramagnetic and Diamagnetic Materials at $T = 20°C$

Material	$\chi_m = K_m - 1$ ($\times 10^{-5}$)
Paramagnetic	
Iron ammonium alum	66
Uranium	40
Platinum	26
Aluminum	2.2
Sodium	0.72
Oxygen gas	0.19
Diamagnetic	
Bismuth	−16.6
Mercury	−2.9
Silver	−2.6
Carbon (diamond)	−2.1
Lead	−1.8
Sodium chloride	−1.4
Copper	−1.0

The amount by which the relative permeability differs from unity is called the **magnetic susceptibility,** denoted by χ_m:

$$\chi_m = K_m - 1 \tag{28.31}$$

Both K_m and χ_m are dimensionless quantities. **Table 28.1** lists values of magnetic susceptibility for several materials. For example, for aluminum, $\chi_m = 2.2 \times 10^{-5}$ and $K_m = 1.000022$. The first group in the table consists of paramagnetic materials; we'll soon discuss the second group, which contains *diamagnetic* materials.

The tendency of atomic magnetic moments to align themselves parallel to the magnetic field (where the potential energy is minimum) is opposed by random thermal motion, which tends to randomize their orientations. For this reason, paramagnetic susceptibility always decreases with increasing temperature. In many cases it is inversely proportional to the absolute temperature T, and the magnetization M can be expressed as

$$M = C\frac{B}{T} \tag{28.32}$$

This relationship is called *Curie's law,* after its discoverer, Pierre Curie (1859–1906). The quantity C is a constant, different for different materials, called the *Curie constant.*

As we described in Section 27.7, a body with atomic magnetic dipoles is attracted to the poles of a magnet. In most paramagnetic substances this attraction is very weak due to thermal randomization of the atomic magnetic moments. But at very low temperatures the thermal effects are reduced, the magnetization increases in accordance with Curie's law, and the attractive forces are greater.

EXAMPLE 28.11 MAGNETIC DIPOLES IN A PARAMAGNETIC MATERIAL

SOLUTION

Nitric oxide (NO) is a paramagnetic compound. The magnetic moment of each NO molecule has a maximum component in any direction of about one Bohr magneton. Compare the interaction energy of such magnetic moments in a 1.5-T magnetic field with the average translational kinetic energy of molecules at 300 K.

SOLUTION

IDENTIFY and SET UP: This problem involves the energy of a magnetic moment in a magnetic field and the average thermal kinetic energy. We have Eqs. (27.27), $U = -\vec{\boldsymbol{\mu}} \cdot \vec{\boldsymbol{B}}$, for the interaction energy of a magnetic moment $\vec{\boldsymbol{\mu}}$ with a $\vec{\boldsymbol{B}}$ field, and (18.16), $K = \frac{3}{2}kT$, for the average translational kinetic energy of a molecule at temperature T.

EXECUTE: We can write $U = -\mu_{\parallel}B$, where $\mu_{\parallel}$ is the component of the magnetic moment $\vec{\boldsymbol{\mu}}$ in the direction of the $\vec{\boldsymbol{B}}$ field. Here the maximum value of $\mu_{\parallel}$ is about μ_B, so

$$|U|_{max} \approx \mu_B B = (9.27 \times 10^{-24}\ \text{J/T})(1.5\ \text{T})$$
$$= 1.4 \times 10^{-23}\ \text{J} = 8.7 \times 10^{-5}\ \text{eV}$$

The average translational kinetic energy K is

$$K = \tfrac{3}{2}kT = \tfrac{3}{2}(1.38 \times 10^{-23}\ \text{J/K})(300\ \text{K})$$
$$= 6.2 \times 10^{-21}\ \text{J} = 0.039\ \text{eV}$$

EVALUATE: At 300 K the magnetic interaction energy is only about 0.2% of the thermal kinetic energy, so we expect only a slight degree of alignment. This is why paramagnetic susceptibilities at ordinary temperature are usually very small.

Diamagnetism

In some materials the total magnetic moment of all the atomic current loops is zero when no magnetic field is present. But even these materials have magnetic effects because an external field alters electron motions within the atoms, causing additional current loops and induced magnetic dipoles comparable to the induced *electric* dipoles we studied in Section 28.5. In this case the additional field caused by these current loops is always *opposite* in direction to that of the external field. (This behavior is explained by Faraday's law of induction, which we will study in Chapter 29. An induced current always tends to cancel the field change that caused it.)

Such materials are said to be **diamagnetic.** They always have negative susceptibility, as shown in Table 28.1, and relative permeability K_m slightly *less* than unity, typically of the order of 0.99990 to 0.99999 for solids and liquids. Diamagnetic susceptibilities are very nearly temperature independent.

Ferromagnetism

There is a third class of materials, called **ferromagnetic** materials, that includes iron, nickel, cobalt, and many alloys containing these elements. In these materials, strong interactions between atomic magnetic moments cause them to line up parallel to each other in regions called **magnetic domains,** even when no external field is present. **Figure 28.27** shows an example of magnetic domain structure. Within each domain, nearly all of the atomic magnetic moments are parallel.

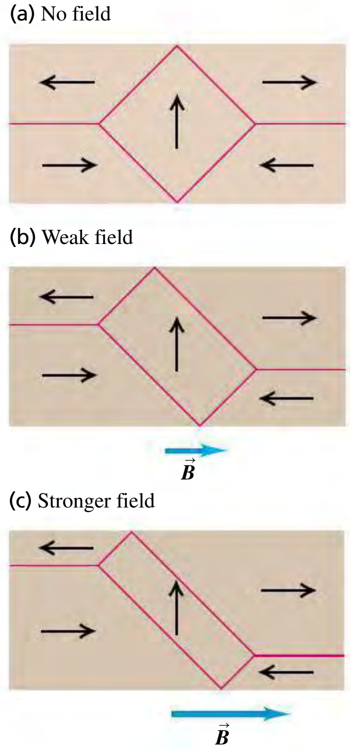
When there is no externally applied field, the domain magnetizations are randomly oriented. But when a field $\vec{B}_0$ (caused by external currents) is present, the domains tend to orient themselves parallel to the field. The domain boundaries also shift; the domains that are magnetized in the field direction grow, and those that are magnetized in other directions shrink. Because the total magnetic moment of a domain may be many thousands of Bohr magnetons, the torques that tend to align the domains with an external field are much stronger than occur with paramagnetic materials. The relative permeability K_m is *much* larger than unity, typically of the order of 1000 to 100,000. As a result, an object made of a ferromagnetic material such as iron is strongly magnetized by the field from a permanent magnet and is attracted to the magnet (see Fig. 27.38). A paramagnetic material such as aluminum is also attracted to a permanent magnet, but K_m for paramagnetic materials is so much smaller for such a material than for ferromagnetic materials that the attraction is very weak. Thus a magnet can pick up iron nails, but not aluminum cans.

As the external field is increased, a point is eventually reached at which nearly *all* the magnetic moments in the ferromagnetic material are aligned parallel to the external field. This condition is called *saturation magnetization;* after it is reached, further increase in the external field causes no increase in magnetization or in the additional field caused by the magnetization.

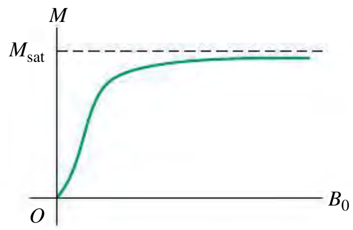
Figure 28.28 shows a "magnetization curve," a graph of magnetization M as a function of external magnetic field B_0, for soft iron. An alternative description of this behavior is that K_m is not constant but decreases as B_0 increases. (Paramagnetic materials also show saturation at sufficiently strong fields. But the magnetic fields required are so large that departures from a linear relationship between M and B_0 in these materials can be observed only at very low temperatures, 1 K or so.)

For many ferromagnetic materials the relationship of magnetization to external magnetic field is different when the external field is increasing from when it is decreasing. **Figure 28.29a** shows this relationship for such a material. When the material is magnetized to saturation and then the external field is reduced to

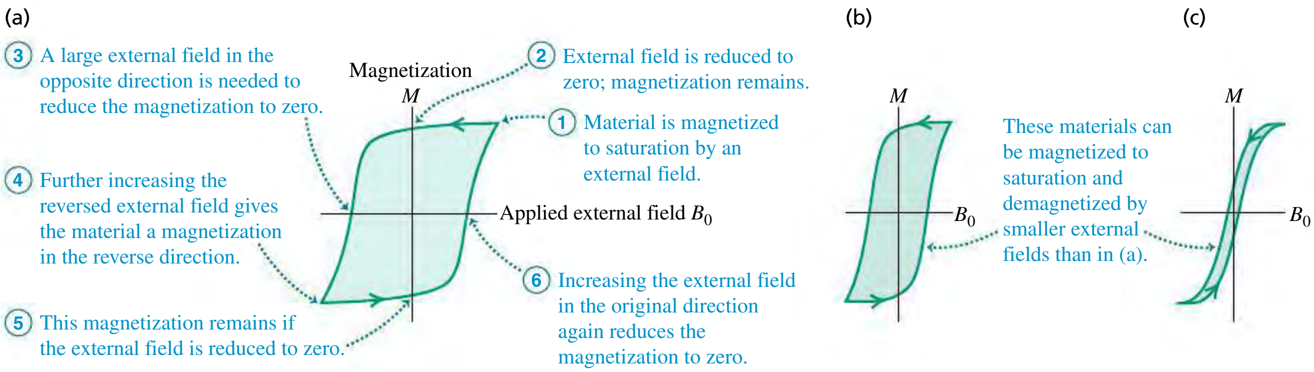
28.27 In this drawing adapted from a magnified photo, the arrows show the directions of magnetization in the domains of a single crystal of nickel. Domains that are magnetized in the direction of an applied magnetic field grow larger.



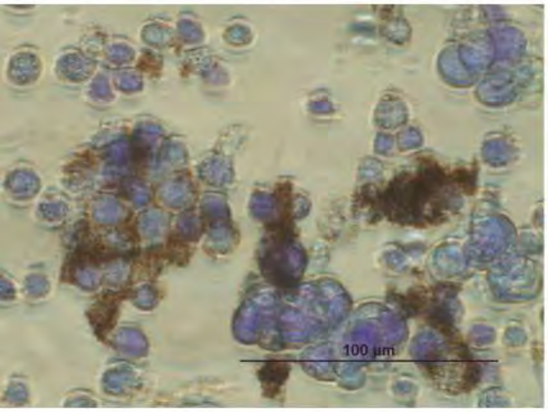
28.28 A magnetization curve for a ferromagnetic material. The magnetization M approaches its saturation value M_{sat} as the magnetic field B_0 (caused by external currents) becomes large.



28.29 Hysteresis loops. The materials of both (a) and (b) remain strongly magnetized when B_0 is reduced to zero. Since (a) is also hard to demagnetize, it would be good for permanent magnets. Since (b) magnetizes and demagnetizes more easily, it could be used as a computer memory material. The material of (c) would be useful for transformers and other alternating-current devices where zero hysteresis would be optimal.



BIO Application Magnetic Nanoparticles for Cancer Therapy The violet blobs in this microscope image are cancer cells that have broken away from a tumor and threaten to spread throughout a patient's body. An experimental technique for fighting these cells uses particles of a magnetic material (shown in brown) injected into the body. These particles are coated with a chemical that preferentially attaches to cancer cells. A magnet outside the patient then "steers" the particles out of the body, taking the cancer cells with them. (Photo courtesy of cancer researcher Dr. Kenneth Scarberry.)



zero, some magnetization remains. This behavior is characteristic of permanent magnets, which retain most of their saturation magnetization when the magnetizing field is removed. To reduce the magnetization to zero requires a magnetic field in the reverse direction.

This behavior is called **hysteresis,** and the curves in Fig. 28.29 are called *hysteresis loops.* Magnetizing and demagnetizing a material that has hysteresis involve the dissipation of energy, and the temperature of the material increases during such a process.

Ferromagnetic materials are widely used in electromagnets, transformer cores, and motors and generators, in which it is desirable to have as large a magnetic field as possible for a given current. Because hysteresis dissipates energy, materials that are used in these applications should usually have as narrow a hysteresis loop as possible. Soft iron is often used; it has high permeability without appreciable hysteresis. For permanent magnets a broad hysteresis loop is usually desirable, with large zero-field magnetization and large reverse field needed to demagnetize. Many kinds of steel and many alloys, such as Alnico, are used for permanent magnets. The remaining magnetic field in such a material, after it has been magnetized to near saturation, is typically of the order of 1 T, corresponding to a remaining magnetization $M = B/\mu_0$ of about 800,000 A/m.

EXAMPLE 28.12 A FERROMAGNETIC MATERIAL

SOLUTION

A cube-shaped permanent magnet is made of a ferromagnetic material with a magnetization M of about 8×10^5 A/m. The side length is 2 cm. (a) Find the magnetic dipole moment of the magnet. (b) Estimate the magnetic field due to the magnet at a point 10 cm from the magnet along its axis.

SOLUTION

IDENTIFY and SET UP: This problem uses the relationship between magnetization M and magnetic dipole moment μ_{total} and the idea that a magnetic dipole produces a magnetic field. We find μ_{total} from Eq. (28.28). To estimate the field, we approximate the magnet as a current loop with this same magnetic moment and use Eq. (28.18).

EXECUTE: (a) From Eq. (28.28),

$$\mu_{\text{total}} = MV = (8 \times 10^5\ \text{A/m})(2 \times 10^{-2}\ \text{m})^3 = 6\ \text{A} \cdot \text{m}^2$$

(b) From Eq. (28.18), the magnetic field on the axis of a current loop with magnetic moment μ_{total} is

$$B = \frac{\mu_0 \mu_{\text{total}}}{2\pi(x^2 + a^2)^{3/2}}$$

where x is the distance from the loop and a is its radius. We can use this expression here if we take a to refer to the size of the permanent magnet. Strictly speaking, there are complications because our magnet does not have the same geometry as a circular current loop. But because $x = 10$ cm is fairly large in comparison to the 2-cm size of the magnet, the term a^2 is negligible in comparison to x^2 and can be ignored. So

$$B \approx \frac{\mu_0 \mu_{\text{total}}}{2\pi x^3} = \frac{(4\pi \times 10^{-7}\ \text{T} \cdot \text{m/A})(6\ \text{A} \cdot \text{m}^2)}{2\pi(0.1\ \text{m})^3}$$

$$= 1 \times 10^{-3}\ \text{T} = 10\ \text{G}$$

which is about ten times stronger than the earth's magnetic field.

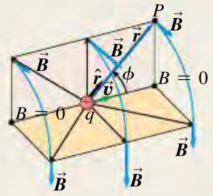
EVALUATE: We calculated B at a point *outside* the magnetic material and therefore used μ_0, not the permeability μ of the magnetic material, in our calculation. You would substitute permeability μ for μ_0 if you were calculating B *inside* a material with relative permeability K_m, for which $\mu = K_m \mu_0$.

TEST YOUR UNDERSTANDING OF SECTION 28.8 Which of the following materials are attracted to a magnet? (i) Sodium; (ii) bismuth; (iii) lead; (iv) uranium. ❙

SOLUTIONS TO ALL EXAMPLES

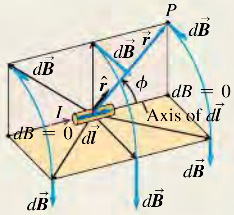
Magnetic field of a moving charge: The magnetic field $\vec{B}$ created by a charge q moving with velocity $\vec{v}$ depends on the distance r from the source point (the location of q) to the field point (where $\vec{B}$ is measured). The $\vec{B}$ field is perpendicular to $\vec{v}$ and to $\hat{r}$, the unit vector directed from the source point to the field point. The principle of superposition of magnetic fields states that the total $\vec{B}$ field produced by several moving charges is the vector sum of the fields produced by the individual charges. (See Example 28.1.)

$$\vec{B} = \frac{\mu_0}{4\pi}\frac{q\vec{v} \times \hat{r}}{r^2} \qquad (28.2)$$



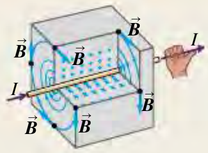
Magnetic field of a current-carrying conductor: The law of Biot and Savart gives the magnetic field $d\vec{B}$ created by an element $d\vec{l}$ of a conductor carrying current I. The field $d\vec{B}$ is perpendicular to both $d\vec{l}$ and $\hat{r}$, the unit vector from the element to the field point. The $\vec{B}$ field created by a finite current-carrying conductor is the integral of $d\vec{B}$ over the length of the conductor. (See Example 28.2.)

$$d\vec{B} = \frac{\mu_0}{4\pi}\frac{I\, d\vec{l} \times \hat{r}}{r^2} \qquad (28.6)$$



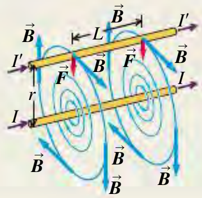
Magnetic field of a long, straight, current-carrying conductor: The magnetic field $\vec{B}$ at a distance r from a long, straight conductor carrying a current I has a magnitude that is inversely proportional to r. The magnetic field lines are circles coaxial with the wire, with directions given by the right-hand rule. (See Examples 28.3 and 28.4.)

$$B = \frac{\mu_0 I}{2\pi r} \qquad (28.9)$$



Magnetic force between current-carrying conductors: Two long, parallel, current-carrying conductors attract if the currents are in the same direction and repel if the currents are in opposite directions. The magnetic force per unit length between the conductors depends on their currents I and I' and separation r. The definition of the ampere is based on this relationship. (See Example 28.5.)

$$\frac{F}{L} = \frac{\mu_0 I I'}{2\pi r} \qquad (28.11)$$



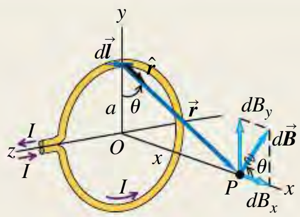
Magnetic field of a current loop: The law of Biot and Savart allows us to calculate the magnetic field produced along the axis of a circular conducting loop of radius a carrying current I. The field depends on the distance x along the axis from the center of the loop to the field point. If there are N loops, the field is multiplied by N. At the center of the loop, $x = 0$. (See Example 28.6.)

$$B_x = \frac{\mu_0 I a^2}{2(x^2 + a^2)^{3/2}} \qquad (28.15)$$

(circular loop)

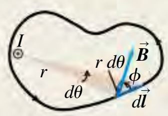
$$B_x = \frac{\mu_0 N I}{2a} \qquad (28.17)$$

(center of N circular loops)



Ampere's law: Ampere's law states that the line integral of $\vec{B}$ around any closed path equals μ_0 times the net current through the area enclosed by the path. The positive sense of current is determined by a right-hand rule. (See Examples 28.7–28.10.)

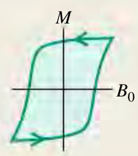
$$\oint \vec{B} \cdot d\vec{l} = \mu_0 I_{\text{encl}} \qquad (28.20)$$



Magnetic fields due to current distributions: The table lists magnetic fields caused by several current distributions. In each case the conductor is carrying current I.

Current Distribution	Point in Magnetic Field	Magnetic-Field Magnitude
Long, straight conductor	Distance r from conductor	$B = \frac{\mu_0 I}{2\pi r}$
Circular loop of radius a	On axis of loop	$B = \frac{\mu_0 I a^2}{2(x^2 + a^2)^{3/2}}$
	At center of loop	$B = \frac{\mu_0 I}{2a}$ (for N loops, multiply these expressions by N)
Long cylindrical conductor of radius R	Inside conductor, $r < R$	$B = \frac{\mu_0 I}{2\pi} \frac{r}{R^2}$
	Outside conductor, $r > R$	$B = \frac{\mu_0 I}{2\pi r}$
Long, closely wound solenoid with n turns per unit length, near its midpoint	Inside solenoid, near center	$B = \mu_0 n I$
	Outside solenoid	$B \approx 0$
Tightly wound toroidal solenoid (toroid) with N turns	Within the space enclosed by the windings, distance r from symmetry axis	$B = \frac{\mu_0 N I}{2\pi r}$
	Outside the space enclosed by the windings	$B \approx 0$

Magnetic materials: When magnetic materials are present, the magnetization of the material causes an additional contribution to $\vec{B}$. For paramagnetic and diamagnetic materials, μ_0 is replaced in magnetic-field expressions by $\mu = K_m \mu_0$, where μ is the permeability of the material and K_m is its relative permeability. The magnetic susceptibility χ_m is defined as $\chi_m = K_m - 1$. Magnetic susceptibilities for paramagnetic materials are small positive quantities; those for diamagnetic materials are small negative quantities. For ferromagnetic materials, K_m is much larger than unity and is not constant. Some ferromagnetic materials are permanent magnets, retaining their magnetization even after the external magnetic field is removed. (See Examples 28.11 and 28.12.)

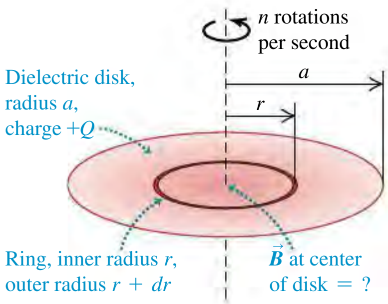


BRIDGING PROBLEM MAGNETIC FIELD OF A CHARGED, ROTATING DIELECTRIC DISK

SOLUTION

A thin dielectric disk with radius a has a total charge $+Q$ distributed uniformly over its surface (**Fig. 28.30**). It rotates n times per second about an axis perpendicular to the surface of the disk and passing through its center. Find the magnetic field at the center of the disk.

28.30 Finding the $\vec{B}$ field at the center of a uniformly charged, rotating disk.



SOLUTION GUIDE

IDENTIFY and SET UP

1. Think of the rotating disk as a series of concentric rotating rings. Each ring acts as a circular current loop that produces a magnetic field at the center of the disk.
2. Use the results of Section 28.5 to find the magnetic field due to a single ring. Then integrate over all rings to find the total field.

EXECUTE

3. Find the charge on a ring with inner radius r and outer radius $r + dr$ (Fig. 28.30).
4. How long does it take the charge found in step 3 to make a complete trip around the rotating ring? Use this to find the current of the rotating ring.
5. Use a result from Section 28.5 to determine the magnetic field that this ring produces at the center of the disk.
6. Integrate your result from step 5 to find the total magnetic field from all rings with radii from $r = 0$ to $r = a$.

EVALUATE

7. Does your answer have the correct units?
8. Suppose all of the charge were concentrated at the rim of the disk (at $r = a$). Would this increase or decrease the field at the center of the disk?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q28.1 A topic of current interest in physics research is the search (thus far unsuccessful) for an isolated magnetic pole, or magnetic *monopole*. If such an entity were found, how could it be recognized? What would its properties be?

Q28.2 Streams of charged particles emitted from the sun during periods of solar activity create a disturbance in the earth's magnetic field. How does this happen?

Q28.3 The text discussed the magnetic field of an infinitely long, straight conductor carrying a current. Of course, there is no such thing as an infinitely long *anything*. How do you decide whether a particular wire is long enough to be considered infinite?

Q28.4 Two parallel conductors carrying current in the same direction attract each other. If they are permitted to move toward each other, the forces of attraction do work. From where does the energy come? Does this contradict the assertion in Chapter 27 that magnetic forces on moving charges do no work? Explain.

Q28.5 Pairs of conductors carrying current into or out of the power-supply components of electronic equipment are sometimes twisted together to reduce magnetic-field effects. Why does this help?

Q28.6 Suppose you have three long, parallel wires arranged so that in cross section they are at the corners of an equilateral triangle. Is there any way to arrange the currents so that all three wires attract each other? So that all three wires repel each other? Explain.

Q28.7 In deriving the force on one of the long, current-carrying conductors in Section 28.4, why did we use the magnetic field due to only one of the conductors? That is, why didn't we use the *total* magnetic field due to *both* conductors?

Q28.8 Two concentric, coplanar, circular loops of wire of different diameter carry currents in the same direction. Describe the nature of the force exerted on the inner loop by the outer loop and on the outer loop by the inner loop.

Q28.9 A current was sent through a helical coil spring. The spring contracted, as though it had been compressed. Why?

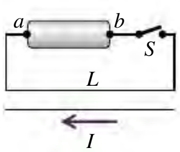
Q28.10 What are the relative advantages and disadvantages of Ampere's law and the law of Biot and Savart for practical calculations of magnetic fields?

Q28.11 Magnetic field lines never have a beginning or an end. Use this to explain why it is reasonable for the field of an ideal toroidal solenoid to be confined entirely to its interior, while a straight solenoid *must* have some field outside.

Q28.12 Two very long, parallel wires carry equal currents in opposite directions. (a) Is there any place that their magnetic fields completely cancel? If so, where? If not, why not? (b) How would the answer to part (a) change if the currents were in the same direction?

Q28.13 In the circuit shown in **Fig. Q28.13**, when switch S is suddenly closed, the wire L is pulled toward the lower wire carrying current I. Which (a or b) is the positive terminal of the battery? How do you know?

Figure **Q28.13**



Q28.14 A metal ring carries a current that causes a magnetic field B_0 at the center of the ring and a field B at point P a distance x from the center along the axis of the ring. If the radius of the ring is doubled, find the magnetic field at the center. Will the field at point P change by the same factor? Why?

Q28.15 Show that the units $A \cdot m^2$ and J/T for the Bohr magneton are equivalent.

Q28.16 Why should the permeability of a paramagnetic material be expected to decrease with increasing temperature?

Q28.17 If a magnet is suspended over a container of liquid air, it attracts droplets to its poles. The droplets contain only liquid oxygen; even though nitrogen is the primary constituent of air, it is not attracted to the magnet. Explain what this tells you about the magnetic susceptibilities of oxygen and nitrogen, and explain why a magnet in ordinary, room-temperature air doesn't attract molecules of oxygen *gas* to its poles.

Q28.18 What features of atomic structure determine whether an element is diamagnetic or paramagnetic? Explain.

Q28.19 The magnetic susceptibility of paramagnetic materials is quite strongly temperature dependent, but that of diamagnetic materials is nearly independent of temperature. Why the difference?

Q28.20 A cylinder of iron is placed so that it is free to rotate around its axis. Initially the cylinder is at rest, and a magnetic field is applied to the cylinder so that it is magnetized in a direction parallel to its axis. If the direction of the *external* field is suddenly reversed, the direction of magnetization will also reverse and the cylinder will begin rotating around its axis. (This is called the *Einstein–de Haas effect*.) Explain why the cylinder begins to rotate.

EXERCISES

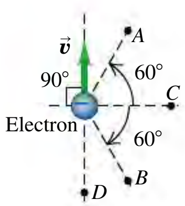
Section 28.1 Magnetic Field of a Moving Charge

28.1 •• A $+6.00\text{-}\mu C$ point charge is moving at a constant 8.00×10^6 m/s in the $+y$-direction, relative to a reference frame. At the instant when the point charge is at the origin of this reference frame, what is the magnetic-field vector $\vec{B}$ it produces at the following points: (a) $x = 0.500$ m, $y = 0$, $z = 0$; (b) $x = 0$, $y = -0.500$ m, $z = 0$; (c) $x = 0$, $y = 0$, $z = +0.500$ m; (d) $x = 0$, $y = -0.500$ m, $z = +0.500$ m?

28.2 • **Fields Within the Atom.** In the Bohr model of the hydrogen atom, the electron moves in a circular orbit of radius 5.3×10^{-11} m with a speed of 2.2×10^6 m/s. If we are viewing the atom in such a way that the electron's orbit is in the plane of the paper with the electron moving clockwise, find the magnitude and direction of the electric and magnetic fields that the electron produces at the location of the nucleus (treated as a point).

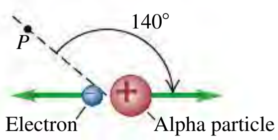
28.3 • An electron moves at $0.100c$ as shown in **Fig. E28.3**. Find the magnitude and direction of the magnetic field this electron produces at the following points, each $2.00\ \mu m$ from the electron: (a) points A and B; (b) point C; (c) point D.

Figure **E28.3**



28.4 •• An alpha particle (charge $+2e$) and an electron move in opposite directions from the same point, each with the speed of 2.50×10^5 m/s (**Fig. E28.4**). Find the magnitude and direction of the total magnetic field these charges produce at point P, which is 8.65 nm from each charge.

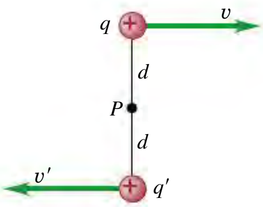
Figure **E28.4**



28.5 • A $-4.80\text{-}\mu C$ charge is moving at a constant speed of 6.80×10^5 m/s in the $+x$-direction relative to a reference frame. At the instant when the point charge is at the origin, what is the magnetic-field vector it produces at the following points: (a) $x = 0.500$ m, $y = 0$, $z = 0$; (b) $x = 0$, $y = 0.500$ m, $z = 0$; (c) $x = 0.500$ m, $y = 0.500$ m, $z = 0$; (d) $x = 0$, $y = 0$, $z = 0.500$ m?

28.6 • Positive point charges $q = +8.00\ \mu C$ and $q' = +3.00\ \mu C$ are moving relative to an observer at point P, as shown in **Fig. E28.6**. The distance d is 0.120 m, $v = 4.50 \times 10^6$ m/s, and $v' = 9.00 \times 10^6$ m/s (a) When the two charges are at the locations shown in the figure, what are the magnitude and direction of the net magnetic field they produce at point P? (b) What are the magnitude and direction of the electric and magnetic forces that each charge exerts on the other, and what is the ratio of the magnitude of the electric force to the magnitude of the magnetic force? (c) If the direction of $\vec{v}\,'$ is reversed, so both charges are moving in the same direction, what are the magnitude and direction of the magnetic forces that the two charges exert on each other?

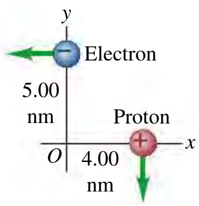
Figure **E28.6**



28.7 • A negative charge $q = -3.60 \times 10^{-6}$ C is located at the origin and has velocity $\vec{v} = (7.50 \times 10^4 \text{m/s})\hat{\imath} + (-4.90 \times 10^4 \text{ m/s})\hat{\jmath}$. At this instant what are the magnitude and direction of the magnetic field produced by this charge at the point $x = 0.200$ m, $y = -0.300$ m, $z = 0$?

28.8 •• An electron and a proton are each moving at 735 km/s in perpendicular paths as shown in **Fig. E28.8**. At the instant when they are at the positions shown, find the magnitude and direction of (a) the total magnetic field they produce at the origin; (b) the magnetic field the electron produces at the location of the proton; (c) the total electric force and the total magnetic force that the electron exerts on the proton.

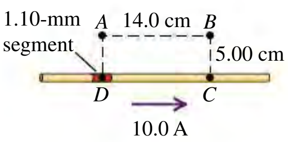
Figure **E28.8**



Section 28.2 Magnetic Field of a Current Element

28.9 • A straight wire carries a 10.0-A current (**Fig. E28.9**). *ABCD* is a rectangle with point D in the middle of a 1.10-mm segment of the wire and point C in the wire. Find the magnitude and direction of the magnetic field due to this segment at (a) point A; (b) point B; (c) point C.

Figure **E28.9**



28.10 • A short current element $d\vec{l} = (0.500 \text{ mm})\hat{\jmath}$ carries a current of 5.40 A in the same direction as $d\vec{l}$. Point P is located at $\vec{r} = (-0.730 \text{ m})\hat{\imath} + (0.390 \text{ m})\hat{k}$. Use unit vectors to express the magnetic field at P produced by this current element.

28.11 •• A long, straight wire lies along the z-axis and carries a 4.00-A current in the $+z$-direction. Find the magnetic field (magnitude and direction) produced at the following points by a 0.500-mm segment of the wire centered at the origin: (a) $x = 2.00$ m, $y = 0$, $z = 0$; (b) $x = 0$, $y = 2.00$ m, $z = 0$; (c) $x = 2.00$ m, $y = 2.00$ m, $z = 0$; (d) $x = 0$, $y = 0$, $z = 2.00$ m,

28.12 •• Two parallel wires are 5.00 cm apart and carry currents in opposite directions, as shown in **Fig. E28.12**. Find the magnitude and direction of the magnetic field at point P due to two 1.50-mm segments of wire that are opposite each other and each 8.00 cm from P.

Figure **E28.12**

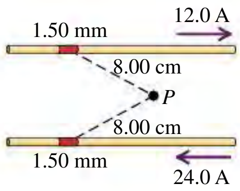
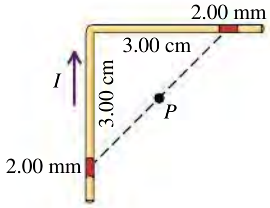


Figure **E28.13**



28.13 • A wire carrying a 28.0-A current bends through a right angle. Consider two 2.00-mm segments of wire, each 3.00 cm from the bend (**Fig. E28.13**). Find the magnitude and direction of the magnetic field these two segments produce at point P, which is midway between them.

28.14 •• A square wire loop 10.0 cm on each side carries a clockwise current of 8.00 A. Find the magnitude and direction of the magnetic field at its center due to the four 1.20-mm wire segments at the midpoint of each side.

Section 28.3 Magnetic Field of a Straight Current-Carrying Conductor

28.15 • **The Magnetic Field from a Lightning Bolt.** Lightning bolts can carry currents up to approximately 20 kA. We can model such a current as the equivalent of a very long, straight wire. (a) If you were unfortunate enough to be 5.0 m away from such a lightning bolt, how large a magnetic field would you experience? (b) How does this field compare to one you would experience by being 5.0 cm from a long, straight household current of 10 A?

28.16 • A very long, straight horizontal wire carries a current such that 8.20×10^{18} electrons per second pass any given point going from west to east. What are the magnitude and direction of the magnetic field this wire produces at a point 4.00 cm directly above it?

28.17 • BIO **Currents in the Heart.** The body contains many small currents caused by the motion of ions in the organs and cells. Measurements of the magnetic field around the chest due to currents in the heart give values of about 10 μG. Although the actual currents are rather complicated, we can gain a rough understanding of their magnitude if we model them as a long, straight wire. If the surface of the chest is 5.0 cm from this current, how large is the current in the heart?

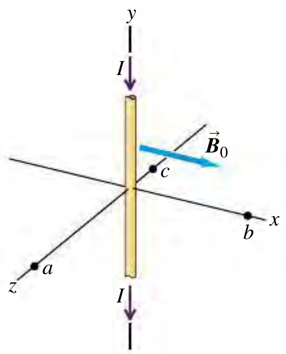
28.18 • BIO **Bacteria Navigation.** Certain bacteria (such as *Aquaspirillum magnetotacticum*) tend to swim toward the earth's geographic north pole because they contain tiny particles, called magnetosomes, that are sensitive to a magnetic field. If a transmission line carrying 100 A is laid underwater, at what range of distances would the magnetic field from this line be great enough to interfere with the migration of these bacteria? (Assume that a field less than 5% of the earth's field would have little effect on the bacteria. Take the earth's field to be 5.0×10^{-5} T, and ignore the effects of the seawater.)

28.19 • (a) How large a current would a very long, straight wire have to carry so that the magnetic field 2.00 cm from the wire is equal to 1.00 G (comparable to the earth's northward-pointing magnetic field)? (b) If the wire is horizontal with the current running from east to west, at what locations would the magnetic field of the wire point in the same direction as the horizontal component of the earth's magnetic field? (c) Repeat part (b) except the wire is vertical with the current going upward.

28.20 • Two long, straight wires, one above the other, are separated by a distance $2a$ and are parallel to the x-axis. Let the $+y$-axis be in the plane of the wires in the direction from the lower wire to the upper wire. Each wire carries current I in the $+x$-direction. What are the magnitude and direction of the net magnetic field of the two wires at a point in the plane of the wires (a) midway between them; (b) at a distance a above the upper wire; (c) at a distance a below the lower wire?

28.21 •• A long, straight wire lies along the y-axis and carries a current $I = 8.00$ A in the $-y$-direction (**Fig. E28.21**). In addition to the magnetic field due to the current in the wire, a uniform magnetic field $\vec{B}_0$ with magnitude 1.50×10^{-6} T is in the $+x$-direction. What is the total field (magnitude and direction) at the following points in the xz-plane: (a) $x = 0$, $z = 1.00$ m; (b) $x = 1.00$ m, $z = 0$; (c) $x = 0$, $z = -0.25$ m?

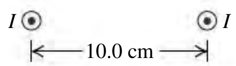
Figure **E28.21**



28.22 •• **BIO** **Transmission Lines and Health.** Currents in dc transmission lines can be 100 A or higher. Some people are concerned that the electromagnetic fields from such lines near their homes could pose health dangers. For a line that has current 150 A and a height of 8.0 m above the ground, what magnetic field does the line produce at ground level? Express your answer in teslas and as a percentage of the earth's magnetic field, which is 0.50 G. Is this value cause for worry?

28.23 • Two long, straight, parallel wires, 10.0 cm apart, carry equal 4.00-A currents in the same direction, as shown in **Fig. E28.23**. Find the magnitude and direction of the magnetic field at (a) point P_1, midway between the wires; (b) point P_2, 25.0 cm to the right of P_1; (c) point P_3, 20.0 cm directly above P_1.

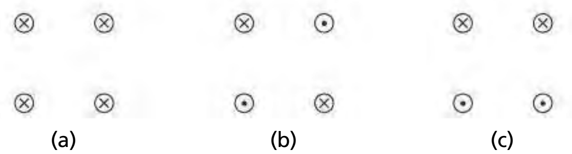
Figure **E28.23**



28.24 •• A rectangular loop with dimensions 4.20 cm by 9.50 cm carries current I. The current in the loop produces a magnetic field at the center of the loop that has magnitude 5.50×10^{-5} T and direction away from you as you view the plane of the loop. What are the magnitude and direction (clockwise or counterclockwise) of the current in the loop?

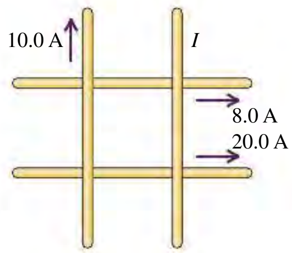
28.25 • Four, long, parallel power lines each carry 100-A currents. A cross-sectional diagram of these lines is a square, 20.0 cm on each side. For each of the three cases shown in **Fig. E28.25**, calculate the magnetic field at the center of the square.

Figure **E28.25**



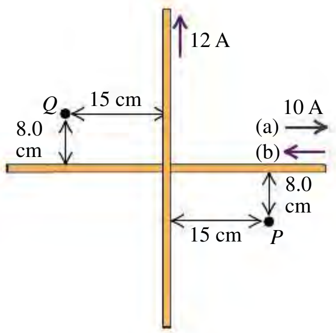
28.26 • Four very long, current-carrying wires in the same plane intersect to form a square 40.0 cm on each side, as shown in **Fig. E28.26**. Find the magnitude and direction of the current I so that the magnetic field at the center of the square is zero.

Figure **E28.26**



28.27 •• Two very long insulated wires perpendicular to each other in the same plane carry currents as shown in **Fig. E28.27**. Find the magnitude of the *net* magnetic field these wires produce at points P and Q if the 10.0-A current is (a) to the right or (b) to the left.

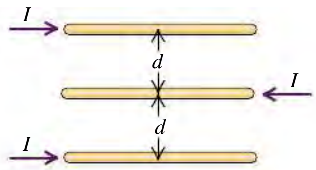
Figure **E28.27**



Section 28.4 Force Between Parallel Conductors

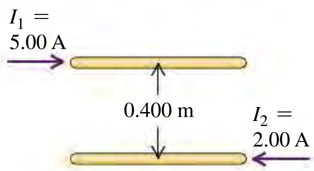
28.28 • Three very long parallel wires each carry current I in the directions shown in **Fig. E28.28**. If the separation between adjacent wires is d, calculate the magnitude and direction of the net magnetic force per unit length on each wire.

Figure **E28.28**



28.29 • Two long, parallel wires are separated by a distance of 0.400 m (**Fig. E28.29**). The currents I_1 and I_2 have the directions shown. (a) Calculate the magnitude of the force exerted by each wire on a 1.20-m length of the other. Is the force attractive or repulsive? (b) Each current is doubled, so that I_1 becomes 10.0 A and I_2 becomes 4.00 A. Now what is the magnitude of the force that each wire exerts on a 1.20-m length of the other?

Figure **E28.29**

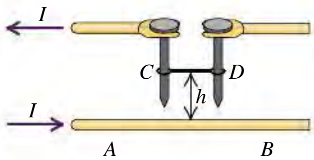


28.30 • Two long, parallel wires are separated by a distance of 2.50 cm. The force per unit length that each wire exerts on the other is 4.00×10^{-5} N/m, and the wires repel each other. The current in one wire is 0.600 A. (a) What is the current in the second wire? (b) Are the two currents in the same direction or in opposite directions?

28.31 • **Lamp Cord Wires.** The wires in a household lamp cord are typically 3.0 mm apart center to center and carry equal currents in opposite directions. If the cord carries direct current to a 100-W light bulb connected across a 120-V potential difference, what force per meter does each wire of the cord exert on the other? Is the force attractive or repulsive? Is this force large enough so it should be considered in the design of the lamp cord? (Model the lamp cord as a very long straight wire.)

28.32 • A long, horizontal wire AB rests on the surface of a table and carries a current I. Horizontal wire CD is vertically above wire AB and is free to slide up and down on the two vertical metal guides C and D (**Fig. E28.32**). Wire CD is connected through the sliding contacts to another wire that also carries a current I, opposite in direction to the current in wire AB. The mass per unit length of the wire CD is λ. To what equilibrium height h will the wire CD rise, assuming that the magnetic force on it is due entirely to the current in the wire AB?

Figure **E28.32**



Section 28.5 Magnetic Field of a Circular Current Loop

28.33 • BIO **Currents in the Brain.** The magnetic field around the head has been measured to be approximately 3.0×10^{-8} G. Although the currents that cause this field are quite complicated, we can get a rough estimate of their size by modeling them as a single circular current loop 16 cm (the width of a typical head) in diameter. What is the current needed to produce such a field at the center of the loop?

28.34 • Calculate the magnitude and direction of the magnetic field at point P due to the current in the semicircular section of wire shown in **Fig. E28.34**. (*Hint:* Does the current in the long, straight section of the wire produce any field at P?)

Figure **E28.34**

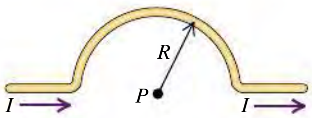
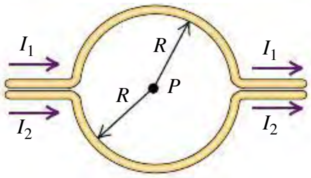


Figure **E28.35**



28.35 •• Calculate the magnitude of the magnetic field at point P of **Fig. E28.35** in terms of R, I_1, and I_2. What does your expression give when $I_1 = I_2$?

28.36 •• A closely wound, circular coil with radius 2.40 cm has 800 turns. (a) What must the current in the coil be if the magnetic field at the center of the coil is 0.0770 T? (b) At what distance x from the center of the coil, on the axis of the coil, is the magnetic field half its value at the center?

28.37 •• A single circular current loop 10.0 cm in diameter carries a 2.00-A current. (a) What is the magnetic field at the center of this loop? (b) Suppose that we now connect 1000 of these loops in series within a 500-cm length to make a solenoid 500 cm long. What is the magnetic field at the center of this solenoid? Is it 1000 times the field at the center of the loop in part (a)? Why or why not?

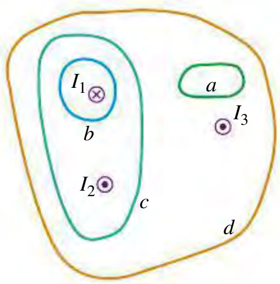
28.38 •• A closely wound coil has a radius of 6.00 cm and carries a current of 2.50 A. How many turns must it have if, at a point on the coil axis 6.00 cm from the center of the coil, the magnetic field is 6.39×10^{-4} T?

28.39 •• Two concentric circular loops of wire lie on a tabletop, one inside the other. The inner wire has a diameter of 20.0 cm and carries a clockwise current of 12.0 A, as viewed from above, and the outer wire has a diameter of 30.0 cm. What must be the magnitude and direction (as viewed from above) of the current in the outer wire so that the net magnetic field due to this combination of wires is zero at the common center of the wires?

Section 28.6 Ampere's Law

28.40 • **Figure E28.40** shows, in cross section, several conductors that carry currents through the plane of the figure. The currents have the magnitudes $I_1 = 4.0$ A, $I_2 = 6.0$ A, and $I_3 = 2.0$ A, and the directions shown. Four paths, labeled a through d, are shown. What is the line integral $\oint \vec{B} \cdot d\vec{l}$ for each path? Each integral involves going around the path in the counterclockwise direction. Explain your answers.

Figure **E28.40**



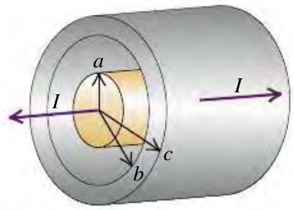
28.41 • A closed curve encircles several conductors. The line integral $\oint \vec{B} \cdot d\vec{l}$ around this curve is 3.83×10^{-4} T·m. (a) What is the net current in the conductors? (b) If you were to integrate around the curve in the opposite direction, what would be the value of the line integral? Explain.

Section 28.7 Applications of Ampere's Law

28.42 •• As a new electrical technician, you are designing a large solenoid to produce a uniform 0.150-T magnetic field near the center of the solenoid. You have enough wire for 4000 circular turns. This solenoid must be 55.0 cm long and 2.80 cm in diameter. What current will you need to produce the necessary field?

28.43 • **Coaxial Cable.** A solid conductor with radius a is supported by insulating disks on the axis of a conducting tube with inner radius b and outer radius c (**Fig. E28.43**). The central conductor and tube carry equal currents I in opposite directions. The currents are distributed uniformly over the cross sections of each conductor. Derive an expression for the magnitude of the magnetic field (a) at points outside the central, solid conductor but inside the tube ($a < r < b$) and (b) at points outside the tube ($r > c$).

Figure **E28.43**



28.44 • Repeat Exercise 28.43 for the case in which the current in the central, solid conductor is I_1, the current in the tube is I_2, and these currents are in the same direction rather than in opposite directions.

28.45 •• A solenoid that is 35 cm long and contains 450 circular coils 2.0 cm in diameter carries a 1.75-A current. (a) What is the magnetic field at the center of the solenoid, 1.0 cm from the coils? (b) Suppose we now stretch out the coils to make a very long wire carrying the same current as before. What is the magnetic field 1.0 cm from the wire's center? Is it the same as that in part (a)? Why or why not?

28.46 •• A 15.0-cm-long solenoid with radius 0.750 cm is closely wound with 600 turns of wire. The current in the windings is 8.00 A. Compute the magnetic field at a point near the center of the solenoid.

28.47 •• A solenoid is designed to produce a magnetic field of 0.0270 T at its center. It has radius 1.40 cm and length 40.0 cm, and the wire can carry a maximum current of 12.0 A. (a) What minimum number of turns per unit length must the solenoid have? (b) What total length of wire is required?

28.48 • A toroidal solenoid has an inner radius of 12.0 cm and an outer radius of 15.0 cm. It carries a current of 1.50 A. How many equally spaced turns must it have so that it will produce a magnetic field of 3.75 mT at points within the coils 14.0 cm from its center?

28.49 • A magnetic field of 37.2 T has been achieved at the MIT Francis Bitter National Magnetic Laboratory. Find the current needed to achieve such a field (a) 2.00 cm from a long, straight wire; (b) at the center of a circular coil of radius 42.0 cm that has 100 turns; (c) near the center of a solenoid with radius 2.40 cm, length 32.0 cm, and 40,000 turns.

28.50 • An ideal toroidal solenoid (see Example 28.10) has inner radius $r_1 = 15.0$ cm and outer radius $r_2 = 18.0$ cm. The solenoid has 250 turns and carries a current of 8.50 A. What is the magnitude of the magnetic field at the following distances from the center of the torus: (a) 12.0 cm; (b) 16.0 cm; (c) 20.0 cm?

28.51 •• A wooden ring whose mean diameter is 14.0 cm is wound with a closely spaced toroidal winding of 600 turns. Compute the magnitude of the magnetic field at the center of the cross section of the windings when the current in the windings is 0.650 A.

Section 28.8 Magnetic Materials

28.52 •• A toroidal solenoid with 400 turns of wire and a mean radius of 6.0 cm carries a current of 0.25 A. The relative permeability of the core is 80. (a) What is the magnetic field in the core? (b) What part of the magnetic field is due to atomic currents?

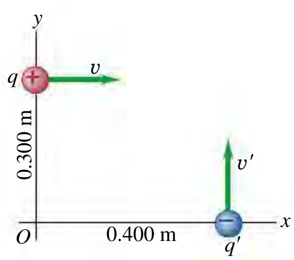
28.53 • A long solenoid with 60 turns of wire per centimeter carries a current of 0.15 A. The wire that makes up the solenoid is wrapped around a solid core of silicon steel ($K_m = 5200$). (The wire of the solenoid is jacketed with an insulator so that none of the current flows into the core.) (a) For a point inside the core, find the magnitudes of (i) the magnetic field $\vec{B}_0$ due to the solenoid current; (ii) the magnetization $\vec{M}$; (iii) the total magnetic field $\vec{B}$. (b) In a sketch of the solenoid and core, show the directions of the vectors $\vec{B}$, $\vec{B}_0$, and $\vec{M}$ inside the core.

28.54 • The current in the windings of a toroidal solenoid is 2.400 A. There are 500 turns, and the mean radius is 25.00 cm. The toroidal solenoid is filled with a magnetic material. The magnetic field inside the windings is found to be 1.940 T. Calculate (a) the relative permeability and (b) the magnetic susceptibility of the material that fills the toroid.

PROBLEMS

28.55 •• A pair of point charges, $q = +8.00\ \mu\text{C}$ and $q' = -5.00\ \mu\text{C}$, are moving as shown in **Fig. P28.55** with speeds $v = 9.00 \times 10^4$ m/s and $v' = 6.50 \times 10^4$ m/s. When the charges are at the locations shown in the figure, what are the magnitude and direction of (a) the magnetic field produced at the origin and (b) the magnetic force that q' exerts on q?

Figure **P28.55**



28.56 •• At a particular instant, charge $q_1 = +4.80 \times 10^{-6}$ C is at the point (0, 0.250 m, 0) and has velocity $\vec{v}_1 = (9.20 \times 10^5\ \text{m/s})\hat{\imath}$. Charge $q_2 = -2.90 \times 10^{-6}$ C is at the point (0.150 m, 0, 0) and has velocity $\vec{v}_2 = (-5.30 \times 10^5\ \text{m/s})\hat{\jmath}$. At this instant, what are the magnitude and direction of the magnetic force that q_1 exerts on q_2?

28.57 ••• Two long, parallel transmission lines, 40.0 cm apart, carry 25.0-A and 75.0-A currents. Find all locations where the net magnetic field of the two wires is zero if these currents are in (a) the same direction and (b) the opposite direction.

28.58 • A long, straight wire carries a current of 8.60 A. An electron is traveling in the vicinity of the wire. At the instant when the electron is 4.50 cm from the wire and traveling at a speed of 6.00×10^4 m/s directly toward the wire, what are the magnitude and direction (relative to the direction of the current) of the force that the magnetic field of the current exerts on the electron?

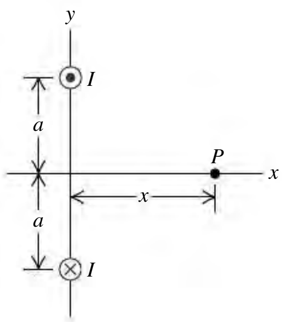
28.59 • **CP** A long, straight wire carries a 13.0-A current. An electron is fired parallel to this wire with a velocity of 250 km/s in the same direction as the current, 2.00 cm from the wire. (a) Find the magnitude and direction of the electron's initial acceleration. (b) What should be the magnitude and direction of a uniform electric field that will allow the electron to continue to travel parallel to the wire? (c) Is it necessary to include the effects of gravity? Justify your answer.

28.60 •• An electron is moving in the vicinity of a long, straight wire that lies along the x-axis. The wire has a constant current of 9.00 A in the $-x$-direction. At an instant when the electron is at point (0, 0.200 m, 0) and the electron's velocity is $\vec{v} = (5.00 \times 10^4\ \text{m/s})\hat{\imath} - (3.00 \times 10^4\ \text{m/s})\hat{\jmath}$, what is the force that the wire exerts on the electron? Express the force in terms of unit vectors, and calculate its magnitude.

28.61 •• An electric bus operates by drawing direct current from two parallel overhead cables, at a potential difference of 600 V, and spaced 55 cm apart. When the power input to the bus's motor is at its maximum power of 65 hp, (a) what current does it draw and (b) what is the attractive force per unit length between the cables?

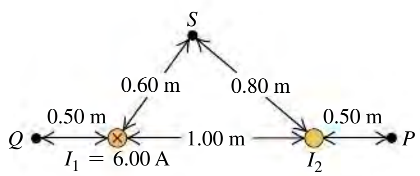
28.62 •• **Figure P28.62** shows an end view of two long, parallel wires perpendicular to the xy-plane, each carrying a current I but in opposite directions. (a) Copy the diagram, and draw vectors to show the $\vec{B}$ field of each wire and the net $\vec{B}$ field at point P. (b) Derive the expression for the magnitude of $\vec{B}$ at any point on the x-axis in terms of the x-coordinate of the point. What is the direction of $\vec{B}$? (c) Graph the magnitude of $\vec{B}$ at points on the x-axis. (d) At what value of x is the magnitude of $\vec{B}$ a maximum? (e) What is the magnitude of $\vec{B}$ when $x \gg a$?

Figure **P28.62**



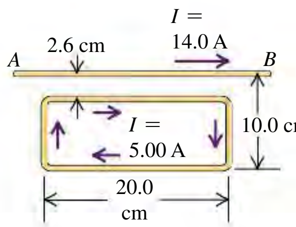
28.63 • Two long, straight, parallel wires are 1.00 m apart (**Fig. P28.63**). The wire on the left carries a current I_1 of 6.00 A into the plane of the paper. (a) What must the magnitude and direction of the current I_2 be for the net field at point P to be zero? (b) Then what are the magnitude and direction of the net field at Q? (c) Then what is the magnitude of the net field at S?

Figure **P28.63**



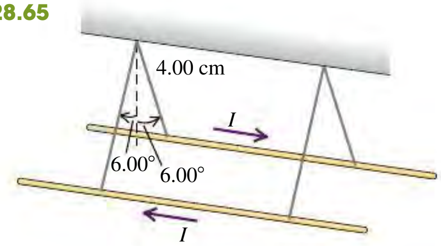
28.64 • The long, straight wire AB shown in **Fig. P28.64** carries a current of 14.0 A. The rectangular loop whose long edges are parallel to the wire carries a current of 5.00 A. Find the magnitude and direction of the net force exerted on the loop by the magnetic field of the wire.

Figure **P28.64**



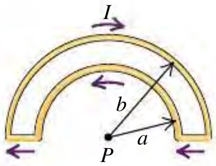
28.65 ••• CP Two long, parallel wires hang by 4.00-cm-long cords from a common axis (**Fig. P28.65**). The wires have a mass per unit length of 0.0125 kg/m and carry the same current in opposite directions. What is the current in each wire if the cords hang at an angle of 6.00° with the vertical?

Figure **P28.65**



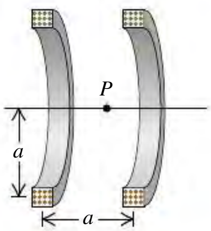
28.66 • The wire semicircles shown in **Fig. P28.66** have radii a and b. Calculate the net magnetic field (magnitude and direction) that the current in the wires produces at point P.

Figure **P28.66**



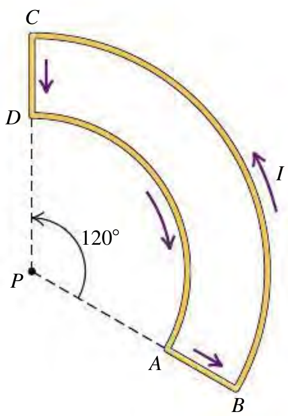
28.67 • CALC **Helmholtz Coils.** **Figure P28.67** is a sectional view of two circular coils with radius a, each wound with N turns of wire carrying a current I, circulating in the same direction in both coils. The coils are separated by a distance a equal to their radii. In this configuration the coils are called Helmholtz coils; they produce a very uniform magnetic field in the region between them. (a) Derive the expression for the magnitude B of the magnetic field at a point on the axis a distance x to the right of point P, which is midway between the coils. (b) Graph B versus x for $x = 0$ to $x = a/2$. Compare this graph to one for the magnetic field due to the right-hand coil alone. (c) From part (a), obtain an expression for the magnitude of the magnetic field at point P. (d) Calculate the magnitude of the magnetic field at P if $N = 300$ turns, $I = 6.00$ A, and $a = 8.00$ cm. (e) Calculate dB/dx and d^2B/dx^2 at $P(x = 0)$. Discuss how your results show that the field is very uniform in the vicinity of P.

Figure **P28.67**



28.68 •• Calculate the magnetic field (magnitude and direction) at a point P due to a current $I = 12.0$ A in the wire shown in **Fig. P28.68**. Segment BC is an arc of a circle with radius 30.0 cm, and point P is at the center of curvature of the arc. Segment DA is an arc of a circle with radius 20.0 cm, and point P is at its center of curvature. Segments CD and AB are straight lines of length 10.0 cm each.

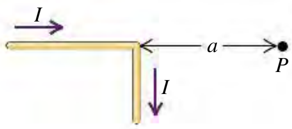
Figure **P28.68**



28.69 • CALC A long, straight wire with a circular cross section of radius R carries a current I. Assume that the current density is not constant across the cross section of the wire, but rather varies as $J = \alpha r$, where α is a constant. (a) By the requirement that J integrated over the cross section of the wire gives the total current I, calculate the constant α in terms of I and R. (b) Use Ampere's law to calculate the magnetic field $B(r)$ for (i) $r \leq R$ and (ii) $r \geq R$. Express your answers in terms of I.

28.70 • CALC The wire shown in **Fig. P28.70** is infinitely long and carries a current I. Calculate the magnitude and direction of the magnetic field that this current produces at point P.

Figure **P28.70**



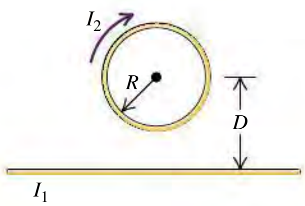
28.71 • CALC A long, straight, solid cylinder, oriented with its axis in the z-direction, carries a current whose current density is $\vec{J}$. The current density, although symmetric about the cylinder axis, is not constant but varies according to the relationship

$$\vec{J} = \frac{2I_0}{\pi a^2}\left[1 - \left(\frac{r}{a}\right)^2\right]\hat{k} \quad \text{for } r \leq a$$
$$= \mathbf{0} \quad \text{for } r \geq a$$

where a is the radius of the cylinder, r is the radial distance from the cylinder axis, and I_0 is a constant having units of amperes. (a) Show that I_0 is the total current passing through the entire cross section of the wire. (b) Using Ampere's law, derive an expression for the magnitude of the magnetic field $\vec{B}$ in the region $r \geq a$. (c) Obtain an expression for the current I contained in a circular cross section of radius $r \leq a$ and centered at the cylinder axis. (d) Using Ampere's law, derive an expression for the magnitude of the magnetic field $\vec{B}$ in the region $r \leq a$. How do your results in parts (b) and (d) compare for $r = a$?

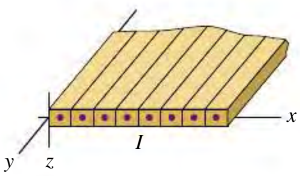
28.72 • A circular loop has radius R and carries current I_2 in a clockwise direction (**Fig. P28.72**). The center of the loop is a distance D above a long, straight wire. What are the magnitude and direction of the current I_1 in the wire if the magnetic field at the center of the loop is zero?

Figure **P28.72**



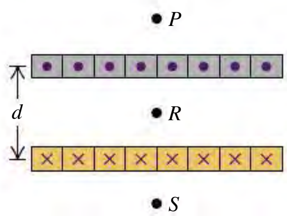
28.73 • **An Infinite Current Sheet.** Long, straight conductors with square cross sections and each carrying current I are laid side by side to form an infinite current sheet (**Fig. P28.73**). The conductors lie in the xy-plane, are parallel to the y-axis, and carry current in the $+y$-direction. There are n conductors per unit length measured along the x-axis. (a) What are the magnitude and direction of the magnetic field a distance a below the current sheet? (b) What are the magnitude and direction of the magnetic field a distance a above the current sheet?

Figure **P28.73**



28.74 • Long, straight conductors with square cross section, each carrying current I, are laid side by side to form an infinite current sheet with current directed out of the plane of the page (**Fig. P28.74**). A second infinite current sheet is a distance d below the first and is parallel to it. The second sheet carries current into the plane of the page. Each sheet has n conductors per unit length. (Refer to Problem 28.73.) Calculate the magnitude and direction of the net magnetic field at (a) point P (above the upper sheet); (b) point R (midway between the two sheets); (c) point S (below the lower sheet).

Figure **P28.74**



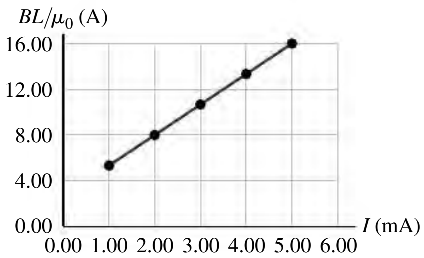
28.75 • A long, straight, solid cylinder, oriented with its axis in the z-direction, carries a current whose current density is $\vec{J}$. The current density, although symmetric about the cylinder axis, is not constant and varies according to the relationship

$$\vec{J} = \left(\frac{b}{r}\right)e^{(r-a)/\delta}\,\hat{k} \quad \text{for } r \le a$$
$$= \mathbf{0} \qquad\qquad\qquad \text{for } r \ge a$$

where the radius of the cylinder is $a = 5.00$ cm, r is the radial distance from the cylinder axis, b is a constant equal to 600 A/m, and δ is a constant equal to 2.50 cm. (a) Let I_0 be the total current passing through the entire cross section of the wire. Obtain an expression for I_0 in terms of b, δ, and a. Evaluate your expression to obtain a numerical value for I_0. (b) Using Ampere's law, derive an expression for the magnetic field $\vec{B}$ in the region $r \ge a$. Express your answer in terms of I_0 rather than b. (c) Obtain an expression for the current I contained in a circular cross section of radius $r \le a$ and centered at the cylinder axis. Express your answer in terms of I_0 rather than b. (d) Using Ampere's law, derive an expression for the magnetic field $\vec{B}$ in the region $r \le a$. (e) Evaluate the magnitude of the magnetic field at $r = \delta$, $r = a$, and $r = 2a$.

28.76 •• DATA As a summer intern at a research lab, you are given a long solenoid that has two separate windings that are wound close together, in the same direction, on the same hollow cylindrical form. You must determine the number of turns in each winding. The solenoid has length $L = 40.0$ cm and diameter 2.80 cm. You let a 2.00-mA current flow in winding 1 and vary the current I in winding 2; both currents flow in the same direction. Then you measure the magnetic-field magnitude B at the center of the solenoid as a function of I. You plot your results as BL/μ_0 versus I. The graph in **Fig. P28.76** shows the best-fit straight line to your data. (a) Explain why the data plotted in this way should fall close to a straight line. (b) Use Fig. P28.76 to calculate N_1 and N_2, the number of turns in windings 1 and 2. (c) If the current in winding 1 remains 2.00 mA in its original direction and winding 2 has $I = 5.00$ mA in the opposite direction, what is B at the center of the solenoid?

Figure **P28.76**



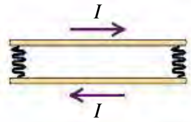
28.77 •• DATA You use a teslameter (a Hall-effect device) to measure the magnitude of the magnetic field at various distances from a long, straight, thick cylindrical copper cable that is carrying a large constant current. To exclude the earth's magnetic field from the measurement, you first set the meter to zero. You then measure the magnetic field B at distances x from the surface of the cable and obtain these data:

x (cm)	2.0	4.0	6.0	8.0	10.0
B (mT)	0.406	0.250	0.181	0.141	0.116

(a) You think you remember from your physics course that the magnetic field of a wire is inversely proportional to the distance from the wire. Therefore, you expect that the quantity Bx from your data will be constant. Calculate Bx for each data point in the table. Is Bx constant for this set of measurements? Explain. (b) Graph the data as x versus $1/B$. Explain why such a plot lies close to a straight line. (c) Use the graph in part (b) to calculate the current I in the cable and the radius R of the cable.

28.78 ••• DATA A pair of long, rigid metal rods, each of length 0.50 m, lie parallel to each other on a frictionless table. Their ends are connected by identical, very lightweight conducting springs with unstretched length l_0 and force constant k (**Fig. P28.78**). When a current I runs through the circuit consisting of the rods and springs, the springs stretch. You measure the distance x each spring stretches for certain values of I. When $I = 8.05$ A, you measure that $x = 0.40$ cm. When $I = 13.1$ A, you find $x = 0.80$ cm. In both cases the rods are much longer than the stretched springs, so it is accurate to use Eq. (28.11) for two infinitely long, parallel conductors. (a) From these two measurements, calculate l_0 and k. (b) If $I = 12.0$ A, what distance x will each spring stretch? (c) What current is required for each spring to stretch 1.00 cm?

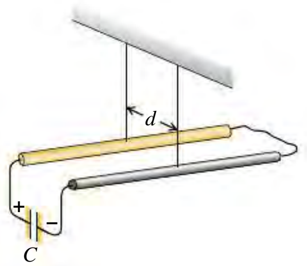
Figure **P28.78**



CHALLENGE PROBLEMS

28.79 ••• CP Two long, straight conducting wires with linear mass density λ are suspended from cords so that they are each horizontal, parallel to each other, and a distance d apart. The back ends of the wires are connected to each other by a slack, low-resistance connecting wire. A charged capacitor (capacitance C) is now added to the system; the positive plate of the capacitor (initial charge $+Q_0$) is connected to the front end of one of the wires, and the negative plate of the capacitor (initial charge $-Q_0$) is connected to the front end of the other wire (**Fig. P28.79**). Both of these connections are also made by slack, low-resistance wires. When the connection is made, the wires are pushed aside by the repulsive force between the wires, and each wire has an initial horizontal velocity of magnitude v_0. Assume that the time constant for the capacitor to discharge is negligible compared to the time it takes for any appreciable displacement in the position of the wires to occur. (a) Show that the initial speed v_0 of either wire is given by

Figure **P28.79**



$$v_0 = \frac{\mu_0 Q_0^2}{4\pi\lambda RCd}$$

where R is the total resistance of the circuit. (b) To what height h will each wire rise as a result of the circuit connection?

28.80 ••• A wide, long, insulating belt has a uniform positive charge per unit area σ on its upper surface. Rollers at each end move the belt to the right at a constant speed v. Calculate the magnitude and direction of the magnetic field produced by the moving belt at a point just above its surface. (*Hint:* At points near the surface and far from its edges or ends, the moving belt can be considered to be an infinite current sheet like that in Problem 28.73.)

PASSAGE PROBLEMS

BIO STUDYING MAGNETIC BACTERIA. Some types of bacteria contain chains of ferromagnetic particles parallel to their long axis. The chains act like small bar magnets that align these *magnetotactic* bacteria with the earth's magnetic field. In one experiment to study the response of such bacteria to magnetic fields, a solenoid is constructed with copper wire 1.0 mm in diameter, evenly wound in a single layer to form a helical coil of length

40 cm and diameter 12 cm. The wire has a very thin layer of insulation, and the coil is wound so that adjacent turns are just touching. The solenoid, which generates a magnetic field, is in an enclosure that shields it from other magnetic fields. A sample of magnetotactic bacteria is placed inside the solenoid. The torque on an individual bacterium in the solenoid's magnetic field is proportional to the magnitude of the magnetic field and to the sine of the angle between the long axis of the bacterium and the magnetic-field direction.

28.81 What current is needed in the wire so that the magnetic field experienced by the bacteria has a magnitude of 150 μT? (a) 0.095 A; (b) 0.12 A; (c) 0.30 A; (d) 14 A.

28.82 To use a larger sample, the experimenters construct a solenoid that has the same length, type of wire, and loop spacing but twice the diameter of the original. How does the maximum possible magnetic torque on a bacterium in this new solenoid compare with the torque the bacterium would have experienced in the original solenoid? Assume that the currents in the solenoids are the same. The maximum torque in the new solenoid is (a) twice that in the original one; (b) half that in the original one; (c) the same as that in the original one; (d) one-quarter that in the original one.

28.83 The solenoid is removed from the enclosure and then used in a location where the earth's magnetic field is 50 μT and points horizontally. A sample of bacteria is placed in the center of the solenoid, and the same current is applied that produced a magnetic field of 150 μT in the lab. Describe the field experienced by the bacteria: The field (a) is still 150 μT; (b) is now 200 μT; (c) is between 100 and 200 μT, depending on how the solenoid is oriented; (d) is between 50 and 150 μT, depending on how the solenoid is oriented.

Answers

Chapter Opening Question ?

(iv) There would be *no* change in the magnetic field strength. From Example 28.9 (Section 28.7), the field inside a solenoid has magnitude $B = \mu_0 nI$, where n is the number of turns of wire per unit length. Joining two solenoids end to end doubles both the number of turns and the length, so the number of turns per unit length is unchanged.

Test Your Understanding Questions

28.1 (a) (i), (b) (ii) The situation is the same as shown in Fig. 28.2 except that the upper proton has velocity $\vec{v}$ rather than $-\vec{v}$. The magnetic field due to the lower proton is the same as shown in Fig. 28.2, but the direction of the magnetic force $\vec{F} = q\vec{v} \times \vec{B}$ on the upper proton is reversed. Hence the magnetic force is attractive. Since the speed v is small compared to c, the magnetic force is much smaller in magnitude than the repulsive electric force and the net force is still repulsive.

28.2 (i) and (iii) (tie), (iv), (ii) From Eq. (28.5), the magnitude of the field dB due to a current element of length dl carrying current I is $dB = (\mu_0/4\pi)(I\,dl\sin\phi/r^2)$. In this expression r is the distance from the element to the field point, and ϕ is the angle between the direction of the current and a vector from the current element to the field point. All four points are the same distance $r = L$ from the current element, so the value of dB is proportional to the value of $\sin\phi$. For the four points the angle is (i) $\phi = 90°$, (ii) $\phi = 0$, (iii) $\phi = 90°$, and (iv) $\phi = 45°$, so the values of $\sin\phi$ are (i) 1, (ii) 0, (iii) 1, and (iv) $1/\sqrt{2}$.

28.3 A This orientation will cause current to flow clockwise around the circuit. Hence current will flow south through the wire that lies under the compass. From the right-hand rule for the magnetic field produced by a long, straight, current-carrying conductor, this will produce a magnetic field that points to the left at the position of the compass (which lies atop the wire). The combination of the northward magnetic field of the earth and the westward field produced by the current gives a net magnetic field to the northwest, so the compass needle will swing counterclockwise to align with this field.

28.4 (a) (i), (b) (iii), (c) (ii), (d) (iii) Current flows in the same direction in adjacent turns of the coil, so the magnetic forces between these turns are attractive. Current flows in opposite directions on opposite sides of the same turn, so the magnetic forces between these sides are repulsive. Thus the magnetic forces on the solenoid turns squeeze them together in the direction along its axis but push them apart radially. The *electric* forces are zero because the wire is electrically neutral, with as much positive charge as there is negative charge.

28.5 (a) (ii), (b) (v) The vector $d\vec{B}$ is in the direction of $d\vec{l} \times \vec{r}$. For a segment on the negative y-axis, $d\vec{l} = -\hat{k}\,dl$ points in the negative z-direction and $\vec{r} = x\hat{\imath} + a\hat{\jmath}$. Hence $d\vec{l} \times \vec{r} = (a\,dl)\hat{\imath} - (x\,dl)\hat{\jmath}$, which has a positive x-component, a negative y-component, and zero z-component. For a segment on the negative z-axis, $d\vec{l} = \hat{\jmath}\,dl$ points in the positive y-direction and $\vec{r} = x\hat{\imath} + a\hat{k}$. Hence $d\vec{l} \times \vec{r} = 1(a\,dl)\hat{\imath} - 1(x\,dl)\hat{k}$, which has a positive x-component, zero y-component, and a negative z-component.

28.6 (ii) Imagine carrying out the integral $\oint \vec{B} \cdot d\vec{l}$ along an integration path that goes counterclockwise around the red magnetic field line. At each point along the path the magnetic field $\vec{B}$ and the infinitesimal segment $d\vec{l}$ are both tangent to the path, so $\vec{B} \cdot d\vec{l}$ is positive at each point and the integral $\oint \vec{B} \cdot d\vec{l}$ is likewise positive. It follows from Ampere's law $\oint \vec{B} \cdot d\vec{l} = \mu_0 I_{encl}$ and the right-hand rule that the integration path encloses a current directed out of the plane of the page. There are no currents in the empty space outside the magnet, so there must be currents inside the magnet (see Section 28.8).

28.7 (iii) By symmetry, any $\vec{B}$ field outside the cable must circulate around the cable, with circular field lines like those surrounding the solid cylindrical conductor in Fig. 28.20. Choose an integration path like the one shown in Fig. 28.20 with radius $r > R$, so that the path completely encloses the cable. As in Example 28.8, the integral $\oint \vec{B} \cdot d\vec{l}$ for this path has magnitude $B(2\pi r)$. From Ampere's law this is equal to $\mu_0 I_{encl}$. The net enclosed current I_{encl} is zero because it includes two currents of equal magnitude but opposite direction: one in the central wire and one in the hollow cylinder. Hence $B(2\pi r) = 0$, and so $B = 0$ for any value of r outside the cable. (The field is nonzero *inside* the cable; see Exercise 28.43.)

28.8 (i), (iv) Sodium and uranium are paramagnetic materials and hence are attracted to a magnet, while bismuth and lead are diamagnetic materials that are repelled by a magnet. (See Table 28.1.)

Bridging Problem

$$B = \frac{\mu_0 n Q}{a}$$

? The card reader at your bank's cash machine scans the information that is coded in a magnetic pattern on the back of your card. Why must you remove the card quickly rather than hold it motionless in the card reader's slot? (i) To maximize the magnetic force on the card; (ii) to maximize the magnetic force on the mobile charges in the card reader; (iii) to generate an electric force on the card; (iv) to generate an electric force on the mobile charges in the card reader.

29 ELECTROMAGNETIC INDUCTION

LEARNING GOALS

Looking forward at …

29.1 The experimental evidence that a changing magnetic field induces an emf.

29.2 How Faraday's law relates the induced emf in a loop to the change in magnetic flux through the loop.

29.3 How to determine the direction of an induced emf.

29.4 How to calculate the emf induced in a conductor moving through a magnetic field.

29.5 How a changing magnetic flux generates a circulating electric field.

29.6 How eddy currents arise in a metal that moves in a magnetic field.

29.7 The four fundamental equations that completely describe both electricity and magnetism.

29.8 The remarkable electric and magnetic properties of superconductors.

Looking back at …

23.1 Conservative electric fields.

25.4 Electromotive force (emf).

27.3, 27.8, 27.9 Magnetic flux; direct-current motors; Hall effect.

28.5–28.7 Magnetic field of a current loop and solenoid; Ampere's law.

Almost every modern device or machine, from a computer to a washing machine to a power drill, has electric circuits at its heart. We learned in Chapter 25 that an electromotive force (emf) is required for a current to flow in a circuit; in Chapters 25 and 26 we almost always took the source of emf to be a battery. But for most devices that you plug into a wall socket, the source of emf is *not* a battery but an electric generating station. Such a station produces electrical energy by converting other forms of energy: gravitational potential energy at a hydroelectric plant, chemical energy in a coal-, gas-, or oil-fired plant, nuclear energy at a nuclear plant. But how is this energy conversion done?

The answer is a phenomenon known as *electromagnetic induction:* If the magnetic flux through a circuit changes, an emf and a current are induced in the circuit. In a power-generating station, magnets move relative to coils of wire to produce a changing magnetic flux in the coils and hence an emf.

The central principle of electromagnetic induction is *Faraday's law*. This law relates induced emf to changing magnetic flux in any loop, including a closed circuit. We also discuss Lenz's law, which helps us to predict the directions of induced emfs and currents. These principles will allow us to understand electrical energy-conversion devices such as motors, generators, and transformers.

Electromagnetic induction tells us that a time-varying magnetic field can act as a source of electric field. We will also see how a time-varying *electric* field can act as a source of *magnetic* field. These remarkable results form part of a neat package of formulas, called *Maxwell's equations,* that describe the behavior of electric and magnetic fields in general. Maxwell's equations pave the way toward an understanding of electromagnetic waves, the topic of Chapter 32.

29.1 INDUCTION EXPERIMENTS

During the 1830s, several pioneering experiments with magnetically induced emf were carried out in England by Michael Faraday and in the United States by Joseph Henry (1797–1878). **Figure 29.1** shows several examples. In Fig. 29.1a, a coil of wire is connected to a galvanometer. When the nearby magnet is stationary, the meter shows no current. This isn't surprising; there is no source of emf in the circuit. But when we *move* the magnet either toward or away from the coil, the meter shows current in the circuit, but *only* while the magnet is moving (Fig. 29.1b). If we keep the magnet stationary and move the coil, we again detect a current during the motion. We call this an **induced current,** and the corresponding emf required to cause this current is called an **induced emf.**

In Fig. 29.1c we replace the magnet with a second coil connected to a battery. When the second coil is stationary, there is no current in the first coil. However, when we move the second coil toward or away from the first or move the first toward or away from the second, there is current in the first coil, but again *only* while one coil is moving relative to the other.

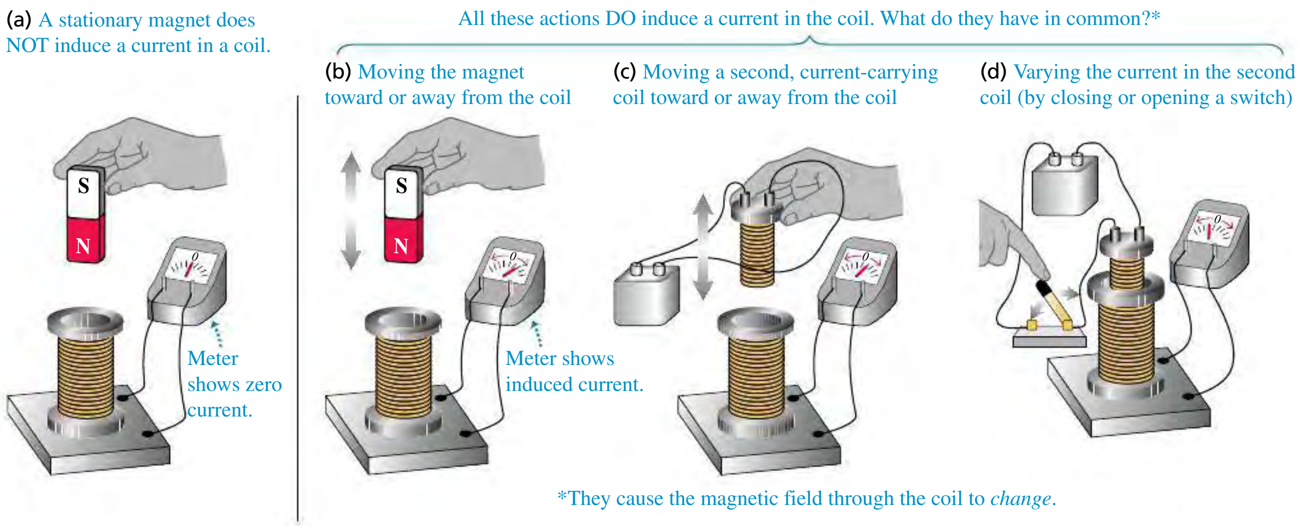
Finally, using the two-coil setup in Fig. 29.1d, we keep both coils stationary and vary the current in the second coil by opening and closing the switch. As we open or close the switch, there is a momentary current pulse in the first coil. The induced current in the first coil is present only while the current in the second coil is changing.

To explore further the common elements in these observations, let's consider a more detailed series of experiments (**Fig. 29.2**). We connect a coil of wire to a galvanometer and then place the coil between the poles of an electromagnet whose magnetic field we can vary. Here's what we observe:

1. When there is no current in the electromagnet, so that $\vec{B} = \mathbf{0}$, the galvanometer shows no current.
2. When the electromagnet is turned on, there is a momentary current through the meter as $\vec{B}$ increases.
3. When $\vec{B}$ levels off at a steady value, the current drops to zero.
4. With the coil in a horizontal plane, we squeeze it so as to decrease the cross-sectional area of the coil. The meter detects current only *during* the deformation, not before or after. When we increase the area to return the coil to its original shape, there is current in the opposite direction, but only while the area of the coil is changing.

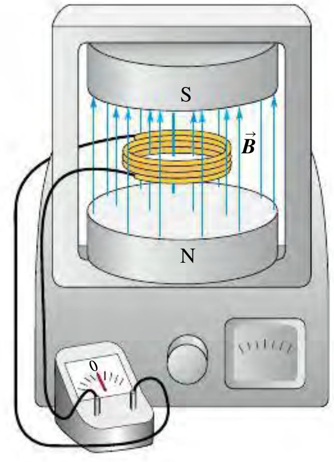
BIO Application Exploring the Brain with Induced emfs Transcranial magnetic stimulation (TMS) is a technique for studying the function of various parts of the brain. A coil held to the subject's head carries a varying electric current, and so produces a varying magnetic field. This field causes an induced emf, and that triggers electric activity in the region of the brain underneath the coil. By observing how the TMS subject responds (for instance, which muscles move as a result of stimulating a certain part of the brain), a physician can test for various neurological conditions.

29.1 Demonstrating the phenomenon of induced current.



5. If we rotate the coil a few degrees about a horizontal axis, the meter detects current during the rotation, in the same direction as when we decreased the area. When we rotate the coil back, there is a current in the opposite direction during this rotation.
6. If we jerk the coil out of the magnetic field, there is a current during the motion, in the same direction as when we decreased the area.
7. If we decrease the number of turns in the coil by unwinding one or more turns, there is a current during the unwinding, in the same direction as when we decreased the area. If we wind more turns onto the coil, there is a current in the opposite direction during the winding.
8. When the magnet is turned off, there is a momentary current in the direction opposite to the current when it was turned on.
9. The faster we carry out any of these changes, the greater the current.
10. If all these experiments are repeated with a coil that has the same shape but different material and different resistance, the current in each case is inversely proportional to the total circuit resistance. This shows that the induced emfs that are causing the current do not depend on the material of the coil but only on its shape and the magnetic field.

29.2 A coil in a magnetic field. When the $\vec{B}$ field is constant and the shape, location, and orientation of the coil do not change, no current is induced in the coil. A current is induced when any of these factors change.



The common element in these experiments is changing *magnetic flux* Φ_B through the coil connected to the galvanometer. In each case the flux changes either because the magnetic field changes with time or because the coil is moving through a nonuniform magnetic field. What's more, in each case the induced emf is proportional to the *rate of change* of magnetic flux Φ_B through the coil. The *direction* of the induced emf depends on whether the flux is increasing or decreasing. If the flux is constant, there is no induced emf.

Induced emfs have a tremendous number of practical applications. If you are reading these words indoors, you are making use of induced emfs right now! At the power plant that supplies your neighborhood, an electric generator produces an emf by varying the magnetic flux through coils of wire. (In the next section we'll see how this is done.) This emf supplies the voltage between the terminals of the wall sockets in your home, and this voltage supplies the power to your reading lamp.

Magnetically induced emfs, just like the emfs discussed in Section 25.4, are the result of *nonelectrostatic* forces. We have to distinguish carefully between the electrostatic electric fields produced by charges (according to Coulomb's law) and the nonelectrostatic electric fields produced by changing magnetic fields. We'll return to this distinction later in this chapter and the next.

29.2 FARADAY'S LAW

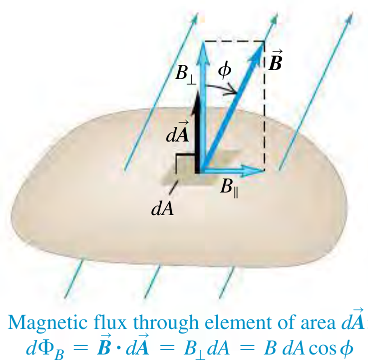
The common element in all induction effects is changing magnetic flux through a circuit. Before stating the simple physical law that summarizes all of the kinds of experiments described in Section 29.1, let's first review the concept of magnetic flux Φ_B (which we introduced in Section 27.3). For an infinitesimal-area element $d\vec{A}$ in a magnetic field $\vec{B}$ (**Fig. 29.3**), the magnetic flux $d\Phi_B$ through the area is

$$d\Phi_B = \vec{B} \cdot d\vec{A} = B_\perp \, dA = B \, dA \cos\phi$$

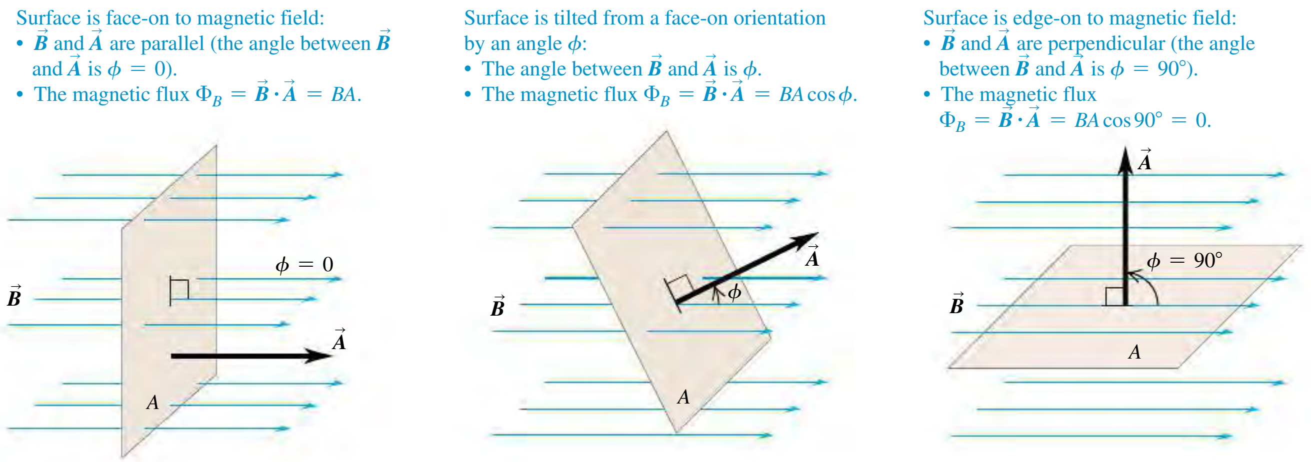
where $B_\perp$ is the component of $\vec{B}$ perpendicular to the surface of the area element and ϕ is the angle between $\vec{B}$ and $d\vec{A}$. (As in Chapter 27, be careful to distinguish between two quantities named "phi," ϕ and Φ_B.) The total magnetic flux Φ_B through a finite area is the integral of this expression over the area:

$$\Phi_B = \int \vec{B} \cdot d\vec{A} = \int B \, dA \cos\phi \qquad (29.1)$$

29.3 Calculating the magnetic flux through an area element.



29.4 Calculating the flux of a uniform magnetic field through a flat area. (Compare to Fig. 22.6, which shows the rules for calculating the flux of a uniform *electric* field.)



If $\vec{B}$ is uniform over a flat area $\vec{A}$, then

$$\Phi_B = \vec{B} \cdot \vec{A} = BA\cos\phi \qquad (29.2)$$

Figure 29.4 reviews the rules for using Eq. (29.2).

CAUTION Choosing the direction of $d\vec{A}$ or $\vec{A}$ In Eqs. (29.1) and (29.2) we must define the direction of the vector area $d\vec{A}$ or $\vec{A}$ unambiguously. There are always two directions perpendicular to any given area, and the sign of the magnetic flux through the area depends on which one we choose. For example, in Fig. 29.3 we chose $d\vec{A}$ to point upward, so ϕ is less than 90° and $\vec{B} \cdot d\vec{A}$ is positive. We could have chosen $d\vec{A}$ to point downward, in which case ϕ would have been greater than 90° and $\vec{B} \cdot d\vec{A}$ would have been negative. Both choices are equally good, but once we make a choice we must stick with it.

Faraday's law of induction states:

Faraday's law: The induced emf in a closed loop ... equals the negative of the time rate of change of magnetic flux through the loop.

$$\mathcal{E} = -\frac{d\Phi_B}{dt} \qquad (29.3)$$

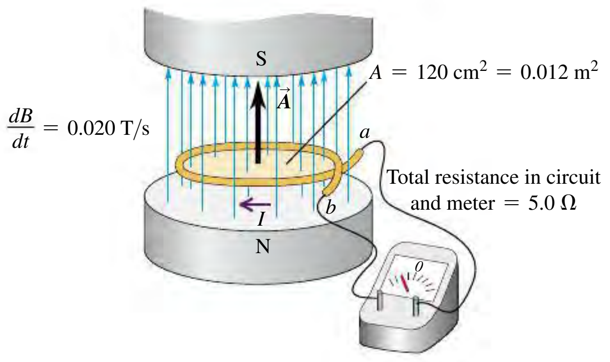
To understand the negative sign, we have to introduce a sign convention for the induced emf $\mathcal{E}$. But first let's look at a simple example of this law in action.

EXAMPLE 29.1 EMF AND CURRENT INDUCED IN A LOOP

SOLUTION

The magnetic field between the poles of the electromagnet in **Fig. 29.5** is uniform at any time, but its magnitude is increasing at the rate of 0.020 T/s. The area of the conducting loop in the field is 120 cm^2, and the total circuit resistance, including the meter, is 5.0 Ω. (a) Find the induced emf and the induced current in the circuit. (b) If the loop is replaced by one made of an insulator, what effect does this have on the induced emf and induced current?

29.5 A stationary conducting loop in an increasing magnetic field.



SOLUTION

IDENTIFY and SET UP: The magnetic flux Φ_B through the loop changes as the magnetic field changes. Hence there will be an induced emf $\mathcal{E}$ and an induced current I in the loop. We calculate

Φ_B from Eq. (29.2), then find $\mathcal{E}$ by using Faraday's law. Finally, we calculate I from $\mathcal{E} = IR$, where R is the total resistance of the circuit that includes the loop.

EXECUTE: (a) The area vector $\vec{A}$ for the loop is perpendicular to the plane of the loop; we take $\vec{A}$ to be vertically upward. Then $\vec{A}$ and $\vec{B}$ are parallel, and because $\vec{B}$ is uniform the magnetic flux through the loop is $\Phi_B = \vec{B} \cdot \vec{A} = BA\cos 0 = BA$. The area $A = 0.012\ \text{m}^2$ is constant, so the rate of change of magnetic flux is

$$\frac{d\Phi_B}{dt} = \frac{d(BA)}{dt} = \frac{dB}{dt}A = (0.020\ \text{T/s})(0.012\ \text{m}^2)$$
$$= 2.4 \times 10^{-4}\ \text{V} = 0.24\ \text{mV}$$

This, apart from a sign that we haven't discussed yet, is the induced emf $\mathcal{E}$. The corresponding induced current is

$$I = \frac{\mathcal{E}}{R} = \frac{2.4 \times 10^{-4}\ \text{V}}{5.0\ \Omega} = 4.8 \times 10^{-5}\ \text{A} = 0.048\ \text{mA}$$

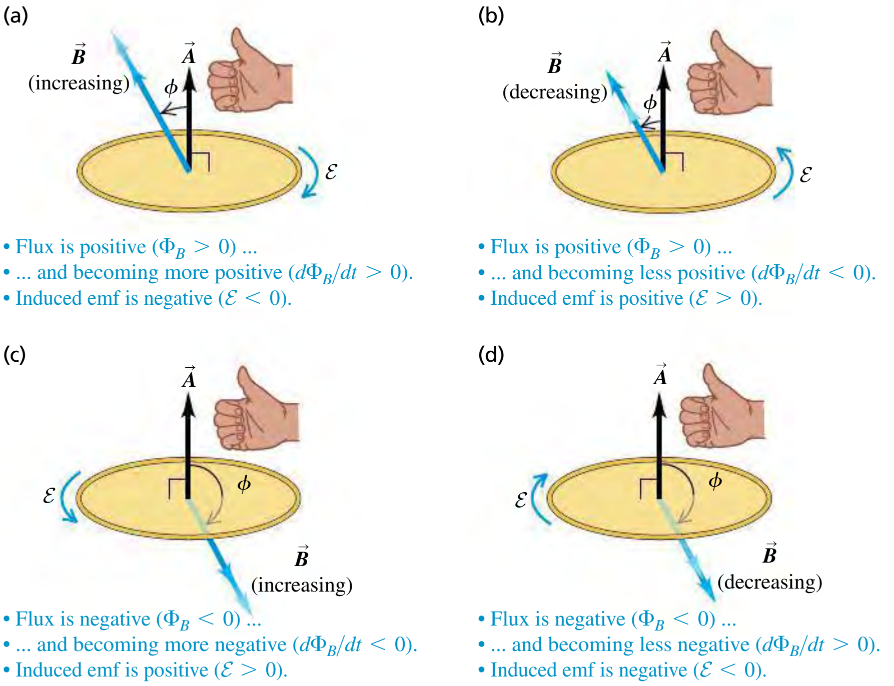
(b) By changing to an insulating loop, we've made the resistance of the loop very high. Faraday's law, Eq. (29.3), does not involve the resistance of the circuit in any way, so the induced *emf* does not change. But the *current* will be smaller, as given by the equation $I = \mathcal{E}/R$. If the loop is made of a perfect insulator with infinite resistance, the induced current is zero. This situation is analogous to an isolated battery whose terminals aren't connected to anything: An emf is present, but no current flows.

EVALUATE: We can verify unit consistency in this calculation by noting that the magnetic-force relationship $\vec{F} = q\vec{v} \times \vec{B}$ implies that the units of $\vec{B}$ are the units of force divided by the units of (charge times velocity): $1\ \text{T} = (1\ \text{N})/(1\ \text{C} \cdot \text{m/s})$. The units of magnetic flux are then $(1\ \text{T})(1\ \text{m}^2) = 1\ \text{N} \cdot \text{s} \cdot \text{m/C}$, and the rate of change of magnetic flux is $1\ \text{N} \cdot \text{m/C} = 1\ \text{J/C} = 1\ \text{V}$. Thus the unit of $d\Phi_B/dt$ is the volt, as required by Eq. (29.3). Also recall that the unit of magnetic flux is the weber (Wb): $1\ \text{T} \cdot \text{m}^2 = 1\ \text{Wb}$, so $1\ \text{V} = 1\ \text{Wb/s}$.

Direction of Induced emf

We can find the direction of an induced emf or current by using Eq. (29.3) together with some simple sign rules. Here's the procedure:

1. Define a positive direction for the vector area $\vec{A}$.
2. From the directions of $\vec{A}$ and the magnetic field $\vec{B}$, determine the sign of the magnetic flux Φ_B and its rate of change $d\Phi_B/dt$. **Figure 29.6** shows several examples.
3. Determine the sign of the induced emf or current. If the flux is increasing, so $d\Phi_B/dt$ is positive, then the induced emf or current is negative; if the flux is decreasing, $d\Phi_B/dt$ is negative and the induced emf or current is positive.
4. Finally, use your right hand to determine the direction of the induced emf or current. Curl the fingers of your right hand around the $\vec{A}$ vector, with your right thumb in the direction of $\vec{A}$. If the induced emf or current in the circuit is *positive,* it is in the same direction as your curled fingers; if the induced emf or current is *negative,* it is in the opposite direction.



29.6 The magnetic flux is becoming **(a)** more positive, **(b)** less positive, **(c)** more negative, and **(d)** less negative. Therefore Φ_B is increasing in (a) and (d) and decreasing in (b) and (c). In (a) and (d) the emfs are negative (they are opposite to the direction of the curled fingers of your right hand when your right thumb points along $\vec{A}$). In (b) and (c) the emfs are positive (in the same direction as the curled fingers).

DATA *SPEAKS*

Magnetic Induction

When students were given a problem involving induced emf and induced currents, more than 68% gave an incorrect response. Common errors:

- Forgetting that *any* change in the magnetic flux through a loop induces an emf in the loop. This can include rotating the loop in a magnetic field or changing the loop's shape.
- Being careless with the sign of magnetic flux. Once you choose the direction of the area vector $\vec{A}$ for a loop, you must use it consistently in flux calculations.

In Example 29.1, in which $\vec{A}$ is upward, a positive $\mathcal{E}$ would be directed counterclockwise around the loop, as seen from above. Both $\vec{A}$ and $\vec{B}$ are upward in this example, so Φ_B is positive; the magnitude B is increasing, so $d\Phi_B/dt$ is positive. Hence by Eq. (29.3), $\mathcal{E}$ in Example 29.1 is *negative*. Its actual direction is thus *clockwise* around the loop, as seen from above.

If the loop in Fig. 29.5 is a conductor, the clockwise induced emf causes a clockwise induced current. This induced current produces an additional magnetic field through the loop, and the right-hand rule described in Section 28.5 shows that this field is *opposite* in direction to the increasing field produced by the electromagnet. This is an example of a general rule called *Lenz's law,* which says that any induction effect tends to oppose the change that caused it; in this case the change is the increase in the flux of the electromagnet's field through the loop. (We'll study Lenz's law in detail in the next section.)

You should check out the signs of the induced emfs and currents for the list of experiments in Section 29.1. For example, when the loop in Fig. 29.2 is in a constant field and we tilt it or squeeze it to *decrease* the flux through it, the induced emf and current are counterclockwise, as seen from above.

CAUTION Induced emfs are caused by changes in flux Since magnetic flux plays a central role in Faraday's law, it's tempting to think that *flux* is the cause of induced emf and that an induced emf will appear in a circuit whenever there is a magnetic field in the region bordered by the circuit. But Eq. (29.3) shows that only a *change* in flux through a circuit, not flux itself, can induce an emf in a circuit. If the flux through a circuit has a constant value, whether positive, negative, or zero, there is no induced emf.

If a coil has N identical turns and if the flux varies at the same rate through each turn, the *total* rate of change through all turns is N times that for a single turn. If Φ_B is the flux through each turn, the total emf in a coil with N turns is

$$\mathcal{E} = -N\frac{d\Phi_B}{dt} \tag{29.4}$$

MP

PhET: Faraday's Electromagnetic Lab
PhET: Faraday's Law
PhET: Generator

As we discussed in this chapter's introduction, induced emfs play an essential role in the generation of electrical power for commercial use. Several of the following examples explore different methods of producing emfs by changing the flux through a circuit.

PROBLEM-SOLVING STRATEGY 29.1 FARADAY'S LAW

IDENTIFY *the relevant concepts:* Faraday's law applies when a magnetic flux is changing. To use the law, identify an area through which there is a flux of magnetic field. This will usually be the area enclosed by a loop made of a conducting material (though not always—see part (b) of Example 29.1). Identify the target variables.

SET UP *the problem* using the following steps:
1. Faraday's law relates the induced emf to the rate of change of magnetic flux. To calculate this rate of change, you first have to understand what is making the flux change. Is the conductor moving or changing orientation? Is the magnetic field changing?
2. The area vector $\vec{A}$ (or $d\vec{A}$) must be perpendicular to the plane of the area. You always have two choices of its direction; for example, if the area is in a horizontal plane, $\vec{A}$ could point up or down. Choose a direction and use it throughout the problem.

EXECUTE *the solution* as follows:
1. Calculate the magnetic flux from Eq. (29.2) if $\vec{B}$ is uniform over the area of the loop or Eq. (29.1) if it isn't uniform. Remember the direction you chose for the area vector.
2. Calculate the induced emf from Eq. (29.3) or (if your conductor has N turns in a coil) Eq. (29.4). Apply the sign rule (described just after Example 29.1) to determine the positive direction of emf.
3. If the circuit resistance is known, you can calculate the magnitude of the induced current I by using $\mathcal{E} = IR$.

EVALUATE *your answer:* Check your results for the proper units, and double-check that you have properly implemented the sign rules for magnetic flux and induced emf.

EXAMPLE 29.2 MAGNITUDE AND DIRECTION OF AN INDUCED EMF

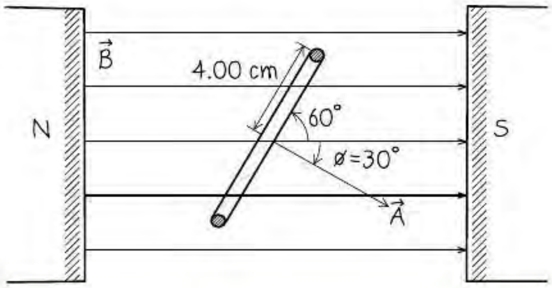
SOLUTION

A 500-loop circular wire coil with radius 4.00 cm is placed between the poles of a large electromagnet. The magnetic field is uniform and makes an angle of 60° with the plane of the coil; it decreases at 0.200 T/s. What are the magnitude and direction of the induced emf?

SOLUTION

IDENTIFY and SET UP: Our target variable is the emf induced by a varying magnetic flux through the coil. The flux varies because the magnetic field decreases in amplitude. We choose the area vector $\vec{A}$ to be in the direction shown in **Fig. 29.7**. With this choice, the geometry is similar to that of Fig. 29.6b. That figure will help us determine the direction of the induced emf.

29.7 Our sketch for this problem.



EXECUTE: The magnetic field is uniform over the loop, so we can calculate the flux from Eq. (29.2): $\Phi_B = BA\cos\phi$, where $\phi = 30°$. In this expression, the only quantity that changes with time is the magnitude B of the field, so $d\Phi_B/dt = (dB/dt)A\cos\phi$.

CAUTION Remember how ϕ is defined You may have been tempted to say that $\phi = 60°$ in this problem. If so, remember that ϕ is the angle between $\vec{A}$ and $\vec{B}$, *not* the angle between $\vec{B}$ and the plane of the loop.

From Eq. (29.4), the induced emf in the coil of $N = 500$ turns is

$$\mathcal{E} = -N\frac{d\Phi_B}{dt} = -N\frac{dB}{dt}A\cos\phi$$

$$= -500(-0.200\ \text{T/s})\pi(0.0400\ \text{m})^2(\cos 30°) = 0.435\ \text{V}$$

The positive answer means that when you point your right thumb in the direction of area vector $\vec{A}$ (30° below field $\vec{B}$ in Fig. 29.7), the positive direction for $\mathcal{E}$ is in the direction of the curled fingers of your right hand. If you viewed the coil from the left in Fig. 29.7 and looked in the direction of $\vec{A}$, the emf would be clockwise.

EVALUATE: If the ends of the wire are connected, the direction of current in the coil is in the same direction as the emf—that is, clockwise as seen from the left side of the coil. A clockwise current increases the magnetic flux through the coil, and therefore tends to oppose the decrease in total flux. This is an example of Lenz's law, which we'll discuss in Section 29.3.

EXAMPLE 29.3 GENERATOR I: A SIMPLE ALTERNATOR

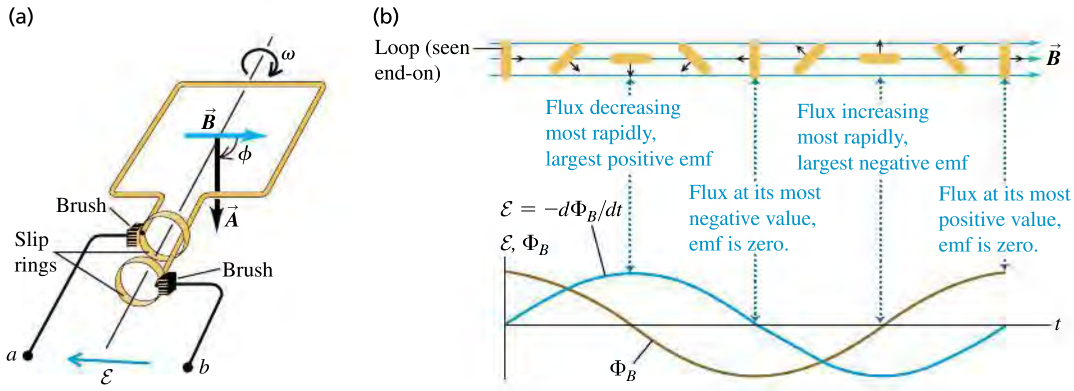
SOLUTION

Figure 29.8a shows a simple *alternator,* a device that generates an emf. A rectangular loop is rotated with constant angular speed ω about the axis shown. The magnetic field $\vec{B}$ is uniform and constant. At time $t = 0$, $\phi = 0$. Determine the induced emf.

SOLUTION

IDENTIFY and SET UP: The magnetic field $\vec{B}$ and the loop area A are constant, but the flux through the loop varies because the loop rotates and so the angle ϕ between $\vec{B}$ and the area vector $\vec{A}$

29.8 (a) Schematic diagram of an alternator. A conducting loop rotates in a magnetic field, producing an emf. Connections from each end of the loop to the external circuit are made by means of that end's slip ring. The system is shown at the time when the angle $\phi = \omega t = 90°$. (b) Graph of the flux through the loop and the resulting emf between terminals a and b, along with the corresponding positions of the loop during one complete rotation.



Continued

changes (Fig. 29.8a). Because the angular speed is constant and $\phi = 0$ at $t = 0$, the angle as a function of time is $\phi = \omega t$.

EXECUTE: The magnetic field is uniform over the loop, so the magnetic flux is $\Phi_B = BA\cos\phi = BA\cos\omega t$. Hence, by Faraday's law [Eq. (29.3)] the induced emf is

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -\frac{d}{dt}(BA\cos\omega t) = \omega BA\sin\omega t$$

EVALUATE: The induced emf $\mathcal{E}$ varies sinusoidally with time (see Fig. 29.8b). When the plane of the loop is perpendicular to $\vec{B}$ ($\phi = 0$ or $180°$), Φ_B reaches its maximum and minimum values. At these times, its instantaneous rate of change is zero and $\mathcal{E}$ is zero. Conversely, $\mathcal{E}$ reaches its maximum and minimum values when the plane of the loop is parallel to $\vec{B}$ ($\phi = 90°$ or $270°$) and Φ_B is changing most rapidly. We note that the induced emf does not depend on the *shape* of the loop, but only on its area.

We can use the alternator as a source of emf in an external circuit by using two *slip rings* that rotate with the loop, as shown in Fig. 29.8a. The rings slide against stationary contacts called *brushes,* which are connected to the output terminals a and b. Since the emf varies sinusoidally, the current that results in the circuit is an *alternating* current that also varies sinusoidally in magnitude and direction. The amplitude of the emf can be increased by increasing the rotation speed, the field magnitude, or the loop area or by using N loops instead of one, as in Eq. (29.4).

Alternators are used in automobiles to generate the currents in the ignition, the lights, and the entertainment system. The arrangement is a little different than in this example; rather than having a rotating loop in a magnetic field, the loop stays fixed and an electromagnet rotates. (The rotation is provided by a mechanical connection between the alternator and the engine.) But the result is the same; the flux through the loop varies sinusoidally, producing a sinusoidally varying emf. Larger alternators of this same type are used in electric power plants (**Fig. 29.9**).

29.9 A commercial alternator uses many loops of wire wound around a barrel-like structure called an armature. The armature and wire remain stationary while electromagnets rotate on a shaft (not shown) through the center of the armature. The resulting induced emf is far larger than would be possible with a single loop of wire.

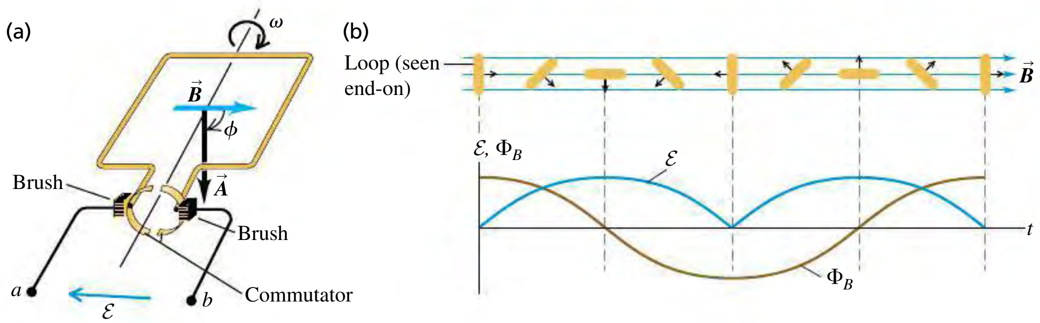
EXAMPLE 29.4 GENERATOR II: A DC GENERATOR AND BACK EMF IN A MOTOR

The alternator in Example 29.3 produces a sinusoidally varying emf and hence an alternating current. **Figure 29.10a** shows a *direct-current* (dc) *generator* that produces an emf that always has the same sign. The arrangement of split rings, called a *commutator,* reverses the connections to the external circuit at angular positions at which the emf reverses. Figure 29.10b shows the resulting emf. Commercial dc generators have a large number of coils and commutator segments, smoothing out the bumps in the emf so that the terminal voltage is not only one-directional but also practically constant. This brush-and-commutator arrangement is the same as that in the direct-current motor discussed in Section 27.8. The motor's *back emf* is just the emf induced by the changing magnetic flux through its rotating coil. Consider a motor with a square, 500-turn coil 10.0 cm on a side. If the magnetic field has magnitude 0.200 T, at what rotation speed is the *average* back emf of the motor equal to 112 V?

SOLUTION

IDENTIFY and SET UP: As far as the rotating loop is concerned, the situation is the same as in Example 29.3 except that we now have N turns of wire. Without the commutator, the emf would alternate between positive and negative values and have an average value of zero (see Fig. 29.8b). With the commutator, the emf is never negative and its average value is positive (Fig. 29.10b). We'll use our result from Example 29.3 to obtain an expression for this average value and solve it for the rotational speed ω.

29.10 (a) Schematic diagram of a dc generator, using a split-ring commutator. The ring halves are attached to the loop and rotate with it. (b) Graph of the resulting induced emf between terminals a and b. Compare to Fig. 29.8b.



EXECUTE: Comparison of Figs. 29.8b and 29.10b shows that the back emf of the motor is just N times the absolute value of the emf found for an alternator in Example 29.3, as in Eq. (29.4): $|\mathcal{E}| = N\omega BA|\sin\omega t|$. To find the *average* back emf, we must replace $|\sin\omega t|$ by its average value. We find this by integrating $|\sin\omega t|$ over half a cycle, from $t = 0$ to $t = T/2 = \pi/\omega$, and dividing by the elapsed time π/ω. During this half cycle, the sine function is positive, so $|\sin\omega t| = \sin\omega t$, and we find

$$(|\sin\omega t|)_{\text{av}} = \frac{\int_0^{\pi/\omega} \sin\omega t\, dt}{\pi/\omega} = \frac{2}{\pi}$$

The average back emf is then

$$\mathcal{E}_{\text{av}} = \frac{2N\omega BA}{\pi}$$

Solving for ω, we obtain

$$\omega = \frac{\pi\mathcal{E}_{\text{av}}}{2NBA}$$
$$= \frac{\pi(112\text{ V})}{2(500)(0.200\text{ T})(0.100\text{ m})^2} = 176\text{ rad/s}$$

(Recall from Example 29.1 that $1\text{ V} = 1\text{ Wb/s} = 1\text{ T}\cdot\text{m}^2/\text{s}$.)

EVALUATE: The average back emf is directly proportional to ω. Hence the slower the rotation speed, the less the back emf and the greater the possibility of burning out the motor, as we described in Example 27.11 (Section 27.8).

EXAMPLE 29.5 GENERATOR III: THE SLIDEWIRE GENERATOR

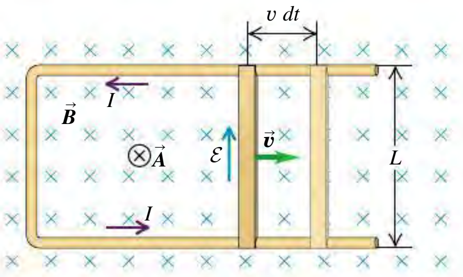
SOLUTION

Figure 29.11 shows a U-shaped conductor in a uniform magnetic field $\vec{B}$ perpendicular to the plane of the figure and directed *into* the page. We lay a metal rod (the "slidewire") with length L across the two arms of the conductor, forming a circuit, and move it to the right with constant velocity $\vec{v}$. This induces an emf and a current, which is why this device is called a *slidewire generator.* Find the magnitude and direction of the resulting induced emf.

SOLUTION

IDENTIFY and SET UP: The magnetic flux changes because the area of the loop—bounded on the right by the moving rod—is increasing. Our target variable is the emf $\mathcal{E}$ induced in this expanding loop. The magnetic field is uniform over the area of the loop, so we can find the flux from $\Phi_B = BA\cos\phi$. We choose the area vector $\vec{A}$ to point straight into the page, in the same direction as $\vec{B}$. With this choice a positive emf will be one that is directed clockwise around the loop. (You can check this with the right-hand rule: Using your right hand, point your thumb into the page and curl your fingers as in Fig. 29.6.)

29.11 A slidewire generator. The magnetic field $\vec{B}$ and the vector area $\vec{A}$ are both directed into the figure. The increase in magnetic flux (caused by an increase in area) induces the emf and current.



EXECUTE: Since $\vec{B}$ and $\vec{A}$ point in the same direction, the angle $\phi = 0$ and $\Phi_B = BA$. The magnetic field magnitude B is constant, so the induced emf is

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -B\frac{dA}{dt}$$

To calculate dA/dt, note that in a time dt the sliding rod moves a distance $v\,dt$ (Fig. 29.11) and the loop area increases by an amount $dA = Lv\,dt$. Hence the induced emf is

$$\mathcal{E} = -B\frac{Lv\,dt}{dt} = -BLv$$

The minus sign tells us that the emf is directed *counterclockwise* around the loop. The induced current is also counterclockwise, as shown in the figure.

EVALUATE: The emf of a slidewire generator is constant if $\vec{v}$ is constant. Hence the slidewire generator is a *direct-current* generator. It's not a very practical device because the rod eventually moves beyond the U-shaped conductor and loses contact, after which the current stops.

EXAMPLE 29.6 WORK AND POWER IN THE SLIDEWIRE GENERATOR

SOLUTION

In the slidewire generator of Example 29.5, energy is dissipated in the circuit owing to its resistance. Let the resistance of the circuit (made up of the moving slidewire and the U-shaped conductor that connects the ends of the slidewire) at a given point in the slidewire's motion be R. Find the rate at which energy is dissipated in the circuit and the rate at which work must be done to move the rod through the magnetic field.

SOLUTION

IDENTIFY and SET UP: Our target variables are the *rates* at which energy is dissipated and at which work is done. Energy is dissipated in the circuit at the rate $P_{\text{dissipated}} = I^2R$. The current I in

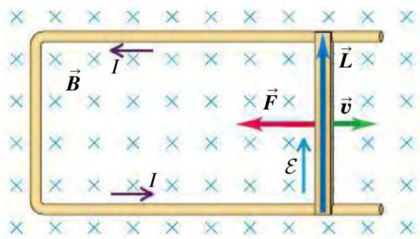
Continued

the circuit equals $|\mathcal{E}|/R$; we found an expression for the induced emf $\mathcal{E}$ in this circuit in Example 29.5. There is a magnetic force $\vec{F} = I\vec{L} \times \vec{B}$ on the rod, where $\vec{L}$ points along the rod in the direction of the current. **Figure 29.12** shows that this force is opposite to the rod velocity $\vec{v}$; to maintain the motion, whoever is pushing the rod must apply a force of equal magnitude in the direction of $\vec{v}$. This force does work at the rate $P_{applied} = Fv$.

EXECUTE: First we'll calculate $P_{dissipated}$. From Example 29.5, $\mathcal{E} = -BLv$, so the current in the rod is $I = |\mathcal{E}|/R = Blv/R$. Hence

$$P_{dissipated} = I^2R = \left(\frac{BLv}{R}\right)^2 R = \frac{B^2L^2v^2}{R}$$

29.12 The magnetic force $\vec{F} = I\vec{L} \times \vec{B}$ that acts on the rod due to the induced current is to the left, opposite to $\vec{v}$.



To calculate $P_{applied}$, we first calculate the magnitude of $\vec{F} = I\vec{L} \times \vec{B}$. Since $\vec{L}$ and $\vec{B}$ are perpendicular, this magnitude is

$$F = ILB = \frac{BLv}{R}LB = \frac{B^2L^2v}{R}$$

The applied force has the same magnitude and does work at the rate

$$P_{applied} = Fv = \frac{B^2L^2v^2}{R}$$

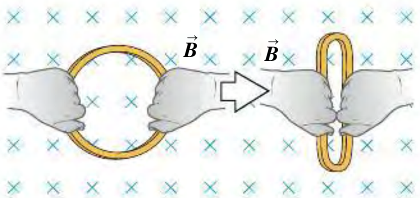
EVALUATE: The rate at which work is done is exactly *equal* to the rate at which energy is dissipated in the resistance.

CAUTION **You can't violate energy conservation** You might think that reversing the direction of $\vec{B}$ or of $\vec{v}$ would allow the magnetic force $\vec{F} = I\vec{L} \times \vec{B}$ to be in the *same* direction as $\vec{v}$. This would be a neat trick. Once the rod was moving, the changing magnetic flux would induce an emf and a current, and the magnetic force on the rod would make it move even faster, increasing the emf and current until the rod was moving at tremendous speed and producing electrical power at a prodigious rate. If this seems too good to be true and a violation of energy conservation, that's because it is. Reversing $\vec{B}$ also reverses the sign of the induced emf and current and hence the direction of $\vec{L}$, so the magnetic force still opposes the motion of the rod; a similar result holds true if we reverse $\vec{v}$.

Generators as Energy Converters

Example 29.6 shows that the slidewire generator doesn't produce electrical energy out of nowhere; the energy is supplied by whatever body exerts the force that keeps the rod moving. All that the generator does is *convert* that energy into a different form. The equality between the rate at which *mechanical* energy is supplied to a generator and the rate at which *electrical* energy is generated holds for all types of generators, including the alternator described in Example 29.3. (We are ignoring the effects of friction in the bearings of an alternator or between the rod and the U-shaped conductor of a slidewire generator. The energy lost to friction is not available for conversion to electrical energy, so in real generators the friction is kept to a minimum.)

In Chapter 27 we stated that the magnetic force on moving charges can never do work. You might think, however, that the magnetic force $\vec{F} = I\vec{L} \times \vec{B}$ in Example 29.6 *is* doing (negative) work on the current-carrying rod as it moves, contradicting our earlier statement. In fact, the work done by the magnetic force is zero. The moving charges that make up the current in the rod in Fig. 29.12 have a vertical component of velocity, causing a horizontal component of force on these charges. As a result, there is a horizontal displacement of charge within the rod, the left side acquiring a net positive charge and the right side a net negative charge. The result is a horizontal component of electric field, perpendicular to the length of the rod (analogous to the Hall effect, described in Section 27.9). It is this field, in the direction of motion of the rod, that does work on the mobile charges in the rod and hence indirectly on the atoms making up the rod.



TEST YOUR UNDERSTANDING OF SECTION 29.2 The accompanying figure shows a wire coil being squeezed in a uniform magnetic field. (a) While the coil is being squeezed, is the induced emf in the coil (i) clockwise, (ii) counterclockwise, or (iii) zero? (b) Once the coil has reached its final squeezed shape, is the induced emf in the coil (i) clockwise, (ii) counterclockwise, or (iii) zero?

29.3 LENZ'S LAW

Lenz's law is a convenient alternative method for determining the direction of an induced current or emf. Lenz's law, named for the Russian physicist H. F. E. Lenz (1804–1865), is not an independent principle; it can be derived from Faraday's law. It always gives the same results as the sign rules we introduced in connection with Faraday's law, but it is often easier to use. Lenz's law also helps us gain intuitive understanding of various induction effects and of the role of energy conservation. **Lenz's law** states:

The direction of any magnetic induction effect is such as to oppose the cause of the effect.

The "cause" may be changing flux through a stationary circuit due to a varying magnetic field, changing flux due to motion of the conductors that make up the circuit, or any combination. If the flux in a stationary circuit changes, as in Examples 29.1 and 29.2, the induced current sets up a magnetic field of its own. Within the area bounded by the circuit, this field is *opposite* to the original field if the original field is *increasing* but is in the *same* direction as the original field if the latter is *decreasing*. That is, the induced current opposes the *change in flux* through the circuit (*not* the flux itself).

If the flux change is due to motion of the conductors, as in Examples 29.3 through 29.6, the direction of the induced current in the moving conductor is such that the direction of the magnetic-field force on the conductor is opposite in direction to its motion. Thus the motion of the conductor, which caused the induced current, is opposed. We saw this explicitly for the slidewire generator in Example 29.6. In all these cases the induced current tries to preserve the *status quo* by opposing motion or a change of flux.

Lenz's law is also directly related to energy conservation. If the induced current in Example 29.6 were in the direction opposite to that given by Lenz's law, the magnetic force on the rod would accelerate it to ever-increasing speed with no external energy source, even though electrical energy is being dissipated in the circuit. This would be a clear violation of energy conservation and doesn't happen in nature.

CONCEPTUAL EXAMPLE 29.7 LENZ'S LAW AND THE SLIDEWIRE GENERATOR

SOLUTION

In Fig. 29.11, the induced current in the loop causes an additional magnetic field in the area bounded by the loop. The direction of the induced current is counterclockwise, so from the discussion of Section 28.5, this additional magnetic field is directed *out of* the plane of the figure. That direction is opposite that of the original magnetic field, so it tends to cancel the effect of that field. This is just what Lenz's law predicts.

CONCEPTUAL EXAMPLE 29.8 LENZ'S LAW AND THE DIRECTION OF INDUCED CURRENT

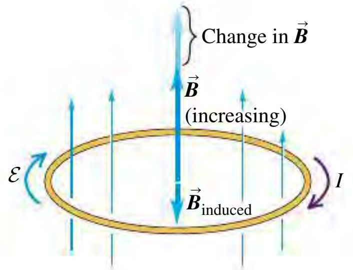
SOLUTION

In **Fig. 29.13** there is a uniform magnetic field $\vec{B}$ through the coil. The magnitude of the field is increasing, so there is an induced emf. Use Lenz's law to determine the direction of the resulting induced current.

SOLUTION

This situation is the same as in Example 29.1 (Section 29.2). By Lenz's law the induced current must produce a magnetic field $\vec{B}_{\text{induced}}$ inside the coil that is downward, opposing the change in flux. From the right-hand rule we described in Section 28.5 for the direction of the magnetic field produced by a circular loop,

29.13 The induced current due to the change in $\vec{B}$ is clockwise, as seen from above the loop. The added field $\vec{B}_{\text{induced}}$ that it causes is downward, opposing the change in the upward field $\vec{B}$.

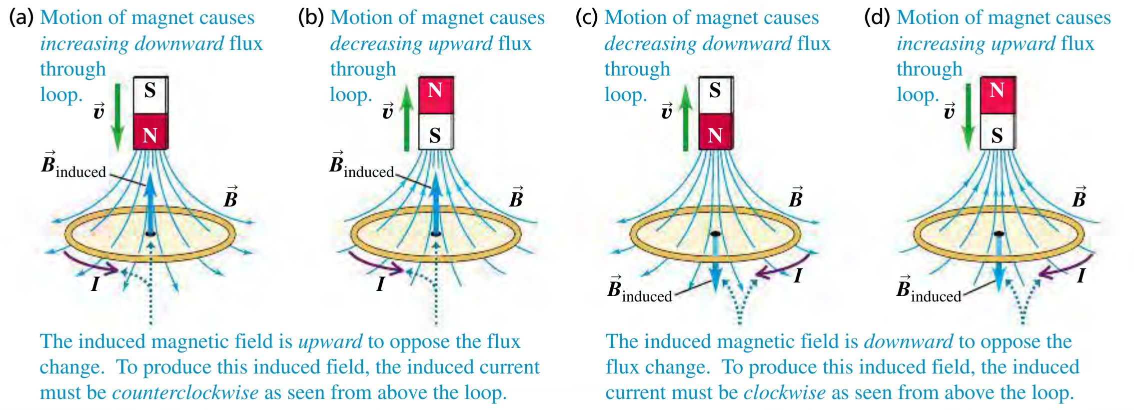


Continued

$\vec{B}_{\text{induced}}$ will be in the desired direction if the induced current flows as shown in Fig. 29.13.

Figure 29.14 shows several applications of Lenz's law to the similar situation of a magnet moving near a conducting loop. In each case, the induced current produces a magnetic field whose direction opposes the change in flux through the loop due to the magnet's motion.

29.14 Directions of induced currents as a bar magnet moves along the axis of a conducting loop. If the bar magnet is stationary, there is no induced current.



Lenz's Law and the Response to Flux Changes

Since an induced current always opposes any change in magnetic flux through a circuit, how is it possible for the flux to change at all? The answer is that Lenz's law gives only the *direction* of an induced current; the *magnitude* of the current depends on the resistance of the circuit. The greater the circuit resistance, the less the induced current that appears to oppose any change in flux and the easier it is for a flux change to take effect. If the loop in Fig. 29.14 were made out of wood (an insulator), there would be almost no induced current in response to changes in the flux through the loop.

Conversely, the less the circuit resistance, the greater the induced current and the more difficult it is to change the flux through the circuit. If the loop in Fig. 29.14 is a good conductor, an induced current flows as long as the magnet moves relative to the loop. Once the magnet and loop are no longer in relative motion, the induced current very quickly decreases to zero because of the nonzero resistance in the loop.

The extreme case occurs when the resistance of the circuit is *zero*. Then the induced current in Fig. 29.14 will continue to flow even after the induced emf has disappeared—that is, even after the magnet has stopped moving relative to the loop. Thanks to this *persistent current,* it turns out that the flux through the loop is exactly the same as it was before the magnet started to move, so the flux through a loop of zero resistance *never* changes. Exotic materials called *superconductors* do indeed have zero resistance; we discuss these further in Section 29.8.

TEST YOUR UNDERSTANDING OF SECTION 29.3 (a) Suppose the magnet in Fig. 29.14a were stationary and the loop of wire moved upward. Would the induced current in the loop be (i) in the same direction as shown in Fig. 29.14a, (ii) in the direction opposite to that shown in Fig. 29.14a, or (iii) zero? (b) Suppose the magnet and loop of wire in Fig. 29.14a both moved downward at the same velocity. Would the induced current in the loop be (i) in the same direction as shown in Fig. 29.14a, (ii) in the direction opposite to that shown in Fig. 29.14a, or (iii) zero? ▌

29.4 MOTIONAL ELECTROMOTIVE FORCE

We've seen several situations in which a conductor moves in a magnetic field, as in the generators discussed in Examples 29.3 through 29.6. We can gain additional insight into the origin of the induced emf in these situations by considering the magnetic forces on mobile charges in the conductor. **Figure 29.15a** shows the same moving rod that we discussed in Example 29.5, separated for the moment from the U-shaped conductor. The magnetic field $\vec{B}$ is uniform and directed into the page, and we move the rod to the right at a constant velocity $\vec{v}$. A charged particle q in the rod then experiences a magnetic force $\vec{F} = q\vec{v} \times \vec{B}$ with magnitude $F = |q|vB$. We'll assume in the following discussion that q is positive; in that case the direction of this force is upward along the rod, from b toward a.

This magnetic force causes the free charges in the rod to move, creating an excess of positive charge at the upper end a and negative charge at the lower end b. This in turn creates an electric field $\vec{E}$ within the rod, in the direction from a toward b (opposite to the magnetic force). Charge continues to accumulate at the ends of the rod until $\vec{E}$ becomes large enough for the downward electric force (with magnitude qE) to cancel exactly the *upward* magnetic force (with magnitude qvB). Then $qE = qvB$ and the charges are in equilibrium.

The magnitude of the potential difference $V_{ab} = V_a - V_b$ is equal to the electric-field magnitude E multiplied by the length L of the rod. From the above discussion, $E = vB$, so

$$V_{ab} = EL = vBL \qquad (29.5)$$

with point a at higher potential than point b.

Now suppose the moving rod slides along a stationary U-shaped conductor, forming a complete circuit (Fig. 29.15b). No *magnetic* force acts on the charges in the stationary U-shaped conductor, but the charge that was near points a and b redistributes itself along the stationary conductor, creating an *electric* field within it. This field establishes a current in the direction shown. The moving rod has become a source of electromotive force; within it, charge moves from lower to higher potential, and in the remainder of the circuit, charge moves from higher to lower potential. We call this emf a **motional electromotive force,** denoted by $\mathcal{E}$. From the above discussion, the magnitude of this emf is

Motional emf, conductor length and velocity perpendicular to uniform $\vec{B}$

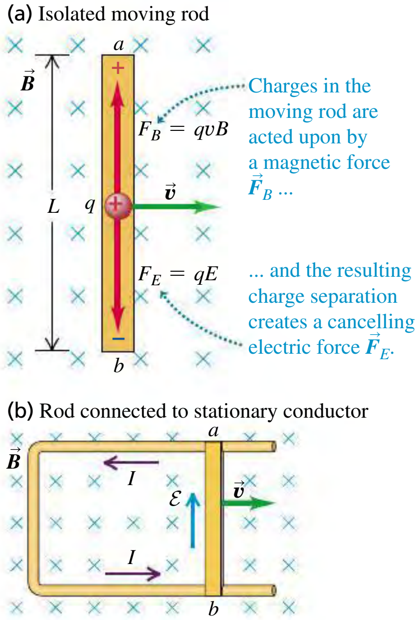
$$\mathcal{E} = vBL \qquad (29.6)$$

(v: Conductor speed; B: Magnitude of uniform magnetic field; L: Conductor length)

This corresponds to a force per unit charge of magnitude vB acting for a distance L along the moving rod. If the total circuit resistance of the U-shaped conductor and the sliding rod is R, the induced current I in the circuit is given by $vBL = IR$. This is the same result we obtained in Section 29.2 by using Faraday's law, and indeed motional emf is a particular case of Faraday's law. Verify that if we express v in meters per second, B in teslas, and L in meters, then $\mathcal{E}$ is in volts. (Recall that $1\ \text{V} = 1\ \text{J/C} = 1\ \text{T} \cdot \text{m}^2/\text{s}$.)

The emf associated with the moving rod in Fig. 29.15b is analogous to that of a battery with its positive terminal at a and its negative terminal at b, although the origins of the two emfs are quite different. In each case a nonelectrostatic force acts on the charges in the device, in the direction from b to a, and the emf is the work per unit charge done by this force when a charge moves from b to a in the device. When the device is connected to an external circuit, the direction of current is from b to a in the device and from a to b in the external circuit. Note that a motional emf is also present in the isolated moving rod in Fig. 29.15a, in the same way that a battery has an emf even when it's not part of a circuit.

29.15 A conducting rod moving in a uniform magnetic field. (a) The rod, the velocity, and the field are mutually perpendicular. (b) Direction of induced current in the circuit.

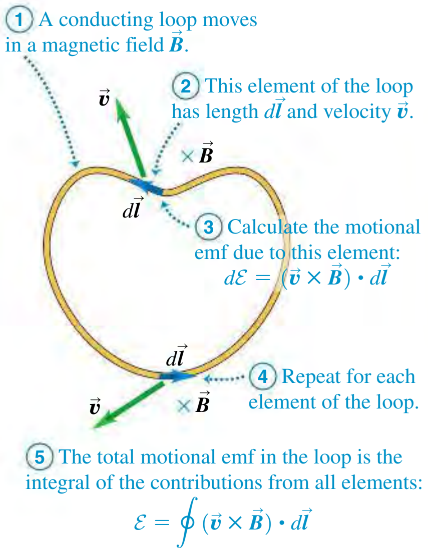


The motional emf $\mathcal{E}$ in the moving rod creates an electric field in the stationary conductor.

You can determine the direction of the induced emf in Fig. 29.15 by using Lenz's law, even if (as in Fig. 29.15a) the conductor does not form a complete circuit. In this case we can mentally complete the circuit between the ends of the conductor and use Lenz's law to determine the direction of the current. From this we can deduce the polarity of the ends of the open-circuit conductor. The direction from the − end to the + end within the conductor is the direction the current would have if the circuit were complete.

Motional emf: General Form

29.16 Calculating the motional emf for a moving current loop. The velocity $\vec{v}$ can be different for different elements if the loop is rotating or changing shape. The magnetic field $\vec{B}$ can also have different values at different points around the loop.



We can generalize the concept of motional emf for a conductor with *any* shape, moving in any magnetic field, uniform or not, if we assume that the magnetic field at each point does not vary with time (**Fig. 29.16**). For an element $d\vec{l}$ of the conductor, the contribution $d\mathcal{E}$ to the emf is the magnitude dl multiplied by the component of $\vec{v}\times\vec{B}$ (the magnetic force per unit charge) parallel to $d\vec{l}$; that is,

$$d\mathcal{E} = (\vec{v}\times\vec{B})\cdot d\vec{l}$$

For any closed conducting loop, the total emf is

Line integral over all elements of **closed conducting loop**

Motional emf, general case

$$\mathcal{E} = \oint(\vec{v}\times\vec{B})\cdot d\vec{l} \qquad (29.7)$$

Length vector of conductor element

Velocity of conductor element

Magnetic field at position of element

This expression looks very different from our original statement of Faraday's law, $\mathcal{E} = -d\Phi_B/dt$ [Eq. (29.3)]. In fact, though, the two statements are equivalent. It can be shown that the rate of change of magnetic flux through a moving conducting loop is always given by the negative of the expression in Eq. (29.7). Thus this equation gives us an alternative formulation of Faraday's law that is often convenient in problems with *moving* conductors. But when we have *stationary* conductors in changing magnetic fields, Eq. (29.7) *cannot* be used; in this case, $\mathcal{E} = -d\Phi_B/dt$ is the only correct way to express Faraday's law.

EXAMPLE 29.9 MOTIONAL EMF IN THE SLIDEWIRE GENERATOR

SOLUTION

Suppose the moving rod in Fig. 29.15b is 0.10 m long, the velocity v is 2.5 m/s, the total resistance of the loop is 0.030 Ω, and B is 0.60 T. Find the motional emf, the induced current, and the force acting on the rod.

SOLUTION

IDENTIFY and SET UP: We'll find the motional emf $\mathcal{E}$ from Eq. (29.6) and the current from the values of $\mathcal{E}$ and the resistance R. The force on the rod is a *magnetic* force, exerted by $\vec{B}$ on the current in the rod; we'll find this force by using $\vec{F} = I\vec{L}\times\vec{B}$.

EXECUTE: From Eq. (29.6) the motional emf is

$$\mathcal{E} = vBL = (2.5\text{ m/s})(0.60\text{ T})(0.10\text{ m}) = 0.15\text{ V}$$

The induced current in the loop is

$$I = \frac{\mathcal{E}}{R} = \frac{0.15\text{ V}}{0.030\ \Omega} = 5.0\text{ A}$$

In the expression for the magnetic force, $\vec{F} = I\vec{L}\times\vec{B}$, the vector $\vec{L}$ points in the same direction as the induced current in the rod (from b to a in Fig. 29.15). The right-hand rule for vector products shows that this force is directed *opposite* to the rod's motion. Since $\vec{L}$ and $\vec{B}$ are perpendicular, the force has magnitude

$$F = ILB = (5.0\text{ A})(0.10\text{ m})(0.60\text{ T}) = 0.30\text{ N}$$

EVALUATE: We can check our answer for the direction of $\vec{F}$ by using Lenz's law. If we take the area vector $\vec{A}$ to point into the plane of the loop, the magnetic flux is positive and increasing as the rod moves to the right and increases the area of the loop. Lenz's law tells us that a force appears to oppose this increase in flux. Hence the force on the rod is to the left, opposite its motion.

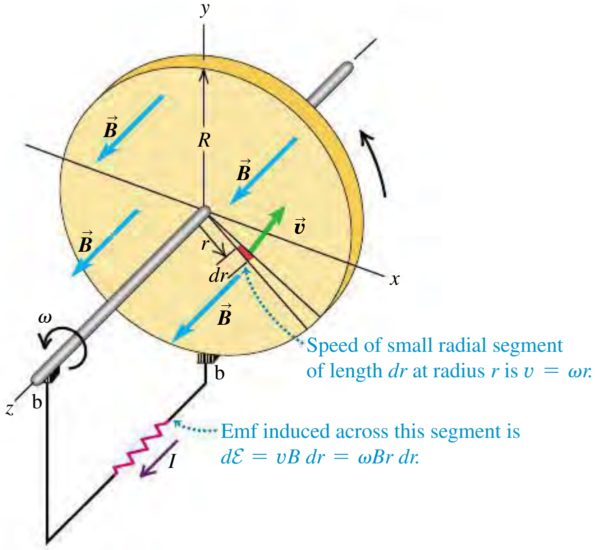
EXAMPLE 29.10 THE FARADAY DISK DYNAMO

Figure 29.17 shows a conducting disk with radius R that lies in the xy-plane and rotates with constant angular velocity ω about the z-axis. The disk is in a uniform, constant $\vec{B}$ field in the z-direction. Find the induced emf between the center and the rim of the disk.

SOLUTION

IDENTIFY and SET UP: A motional emf arises because the conducting disk moves relative to $\vec{B}$. The complication is that different parts of the disk move at different speeds v, depending on their distance from the rotation axis. We'll address this by considering small segments of the disk and integrating their contributions to determine our target variable, the emf between the center and the rim. Consider the small segment of the disk shown in red in Fig. 29.17 and labeled by its velocity vector $\vec{v}$. The magnetic force per unit charge on this segment is $\vec{v} \times \vec{B}$, which points radially outward from the center of the disk. Hence the induced emf tends to make a current flow radially outward, which tells us that the moving conducting path to think about here is a straight line from the center to the rim. We can find the emf from each small disk segment along this line by using $d\mathcal{E} = (\vec{v} \times \vec{B}) \cdot d\vec{l}$ and then integrate to find the total emf.

29.17 A conducting disk with radius R rotating at an angular speed ω in a magnetic field $\vec{B}$. The emf is induced along radial lines of the disk and is applied to an external circuit through the two sliding contacts labeled b.



EXECUTE: The length vector $d\vec{l}$ (of length dr) associated with the segment points radially outward, in the same direction as $\vec{v} \times \vec{B}$. The vectors $\vec{v}$ and $\vec{B}$ are perpendicular, and the magnitude of $\vec{v}$ is $v = \omega r$. The emf from the segment is then $d\mathcal{E} = \omega Br\,dr$. The total emf is the integral of $d\mathcal{E}$ from the center $(r = 0)$ to the rim $(r = R)$:

$$\mathcal{E} = \int_0^R \omega Br\,dr = \tfrac{1}{2}\omega BR^2$$

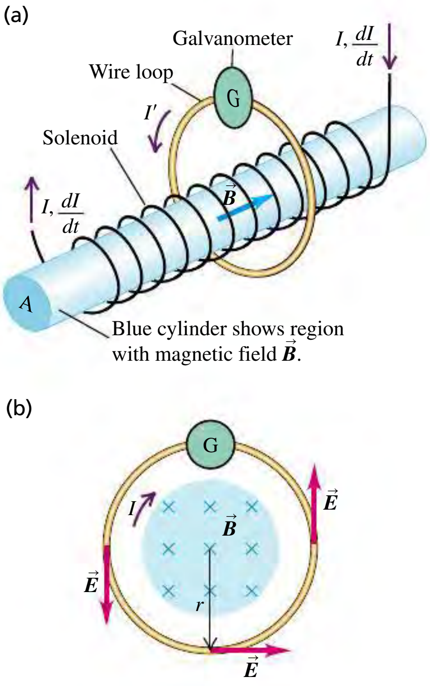
EVALUATE: We can use this device as a source of emf in a circuit by completing the circuit through two stationary brushes (labeled b in the figure) that contact the disk and its conducting shaft as shown. Such a disk is called a *Faraday disk dynamo* or a *homopolar generator*. Unlike the alternator in Example 29.3, the Faraday disk dynamo is a direct-current generator; it produces an emf that is constant in time. Can you use Lenz's law to show that for the direction of rotation in Fig. 29.17, the current in the external circuit must be in the direction shown?

TEST YOUR UNDERSTANDING OF SECTION 29.4 The earth's magnetic field points toward (magnetic) north. For simplicity, assume that the field has no vertical component (as is the case near the earth's equator). (a) If you hold a metal rod in your hand and walk toward the east, how should you orient the rod to get the maximum motional emf between its ends? (i) East-west; (ii) north-south; (iii) up-down; (iv) you get the same motional emf with all of these orientations. (b) How should you hold it to get *zero* emf as you walk toward the east? (i) East-west; (ii) north-south; (iii) up-down; (iv) none of these. (c) In which direction should you travel so that the motional emf across the rod is zero no matter how the rod is oriented? (i) West; (ii) north; (iii) south; (iv) straight up; (v) straight down.

29.5 INDUCED ELECTRIC FIELDS

When a conductor moves in a magnetic field, we can understand the induced emf on the basis of magnetic forces on charges in the conductor, as described in Section 29.4. But an induced emf also occurs when there is a changing flux through a stationary conductor. What is it that pushes the charges around the circuit in this type of situation?

29.18 (a) The windings of a long solenoid carry a current I that is increasing at a rate dI/dt. The magnetic flux in the solenoid is increasing at a rate $d\Phi_B/dt$, and this changing flux passes through a wire loop. An emf $\mathcal{E} = -d\Phi_B/dt$ is induced in the loop, inducing a current I' that is measured by the galvanometer G. (b) Cross-sectional view.



Let's consider the situation shown in **Fig. 29.18**. A long, thin solenoid with cross-sectional area A and n turns per unit length is encircled at its center by a circular conducting loop. The galvanometer G measures the current in the loop. A current I in the winding of the solenoid sets up a magnetic field $\vec{B}$ along the solenoid axis, as shown, with magnitude B as calculated in Example 28.9 (Section 28.7): $B = \mu_0 nI$, where n is the number of turns per unit length. If we ignore the small field outside the solenoid and take the area vector $\vec{A}$ to point in the same direction as $\vec{B}$, then the magnetic flux Φ_B through the loop is

$$\Phi_B = BA = \mu_0 nIA$$

When the solenoid current I changes with time, the magnetic flux Φ_B also changes, and according to Faraday's law the induced emf in the loop is given by

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -\mu_0 nA\frac{dI}{dt} \tag{29.8}$$

If the total resistance of the loop is R, the induced current in the loop, which we may call I', is $I' = \mathcal{E}/R$.

But what *force* makes the charges move around the wire loop? It can't be a magnetic force because the loop isn't even *in* a magnetic field. We are forced to conclude that there has to be an **induced electric field** in the conductor *caused by the changing magnetic flux*. Induced electric fields are *very* different from the electric fields caused by charges, which we discussed in Chapter 23. To see this, note that when a charge q goes once around the loop, the total work done on it by the electric field must be equal to q times the emf $\mathcal{E}$. That is, the electric field in the loop *is not conservative,* as we used the term in Section 23.1, because the line integral of $\vec{E}$ around a closed path is not zero. Indeed, this line integral, representing the work done by the induced $\vec{E}$ field per unit charge, is equal to the induced emf $\mathcal{E}$:

$$\oint \vec{E} \cdot d\vec{l} = \mathcal{E} \tag{29.9}$$

From Faraday's law the emf $\mathcal{E}$ is also the negative of the rate of change of magnetic flux through the loop. Thus for this case we can restate Faraday's law as

Faraday's law for a stationary integration path:

$$\underbrace{\oint \vec{E} \cdot d\vec{l}}_{\text{Line integral of electric field around path}} = \underbrace{-\frac{d\Phi_B}{dt}}_{\text{Negative of the time rate of change of magnetic flux through path}} \tag{29.10}$$

Note that Faraday's law is *always* true in the form $\mathcal{E} = -d\Phi_B/dt$; the form given in Eq. (29.10) is valid *only* if the path around which we integrate is *stationary.*

Let's apply Eq. (29.10) to the stationary circular loop in Fig. 29.18b, which we take to have radius r. Because of cylindrical symmetry, $\vec{E}$ has the same magnitude at every point on the circle and is tangent to it at each point. (Symmetry would also permit the field to be *radial,* but then Gauss's law would require the presence of a net charge inside the circle, and there is none.) The line integral in Eq. (29.10) becomes simply the magnitude E times the circumference $2\pi r$ of the loop, $\oint \vec{E} \cdot d\vec{l} = 2\pi rE$, and Eq. (29.10) gives

$$E = \frac{1}{2\pi r}\left|\frac{d\Phi_B}{dt}\right| \tag{29.11}$$

The directions of $\vec{E}$ at points on the loop are shown in Fig. 29.18b. We know that $\vec{E}$ has to have the direction shown when $\vec{B}$ in the solenoid is increasing, because

$\oint \vec{E} \cdot d\vec{l}$ has to be negative when $d\Phi_B/dt$ is positive. The same approach can be used to find the induced electric field *inside* the solenoid when the solenoid $\vec{B}$ field is changing; we leave the details to you (see Exercise 29.37).

Nonelectrostatic Electric Fields

We've learned that Faraday's law, Eq. (29.3), is valid for two rather different situations. In one, an emf is induced by magnetic forces on charges when a conductor moves through a magnetic field. In the other, a time-varying magnetic field induces an electric field and hence an emf; the $\vec{E}$ field is induced even when no conductor is present. This $\vec{E}$ field differs from an electro*static* field in an important way. It is *nonconservative;* the line integral $\oint \vec{E} \cdot d\vec{l}$ around a closed path is not zero, and when a charge moves around a closed path, the field does a nonzero amount of work on it. It follows that for such a field the concept of *potential* has no meaning. We call such a field a **nonelectrostatic field.** In contrast, an *electrostatic* field is *always* conservative, as we discussed in Section 23.1, and always has an associated potential function. Despite this difference, the fundamental effect of *any* electric field is to exert a force $\vec{F} = q\vec{E}$ on a charge q. This relationship is valid whether $\vec{E}$ is conservative and produced by charges or nonconservative and produced by changing magnetic flux.

So a changing magnetic field acts as a source of electric field of a sort that we *cannot* produce with any static charge distribution. What's more, we'll see in Section 29.7 that a changing *electric* field acts as a source of *magnetic* field. We'll explore this symmetry between the two fields in our study of electromagnetic waves in Chapter 32.

If any doubt remains in your mind about the reality of magnetically induced electric fields, consider a few of the many practical applications (**Fig. 29.19**). Pickups in electric guitars use currents induced in stationary pickup coils by the vibration of nearby ferromagnetic strings. Alternators in most cars use rotating magnets to induce currents in stationary coils. Whether we realize it or not, magnetically induced electric fields play an important role in everyday life.

29.19 Applications of induced electric fields. **(a)** This hybrid automobile has both a gasoline engine and an electric motor. As the car comes to a halt, the spinning wheels run the motor backward so that it acts as a generator. The resulting induced current is used to recharge the car's batteries. **(b)** The rotating crankshaft of a piston-engine airplane spins a magnet, inducing an emf in an adjacent coil and generating the spark that ignites fuel in the engine cylinders. This keeps the engine running even if the airplane's other electrical systems fail.

SOLUTION

EXAMPLE 29.11 INDUCED ELECTRIC FIELDS

Suppose the long solenoid in Fig. 29.18a has 500 turns per meter and cross-sectional area $4.0\ \text{cm}^2$. The current in its windings is increasing at $100\ \text{A/s}$. (a) Find the magnitude of the induced emf in the wire loop outside the solenoid. (b) Find the magnitude of the induced electric field within the loop if its radius is 2.0 cm.

SOLUTION

IDENTIFY and SET UP: As in Fig. 29.18b, the increasing magnetic field inside the solenoid causes a change in the magnetic flux through the wire loop and hence induces an electric field $\vec{E}$ around the loop. Our target variables are the induced emf $\mathcal{E}$ and the electric-field magnitude E. We use Eq. (29.8) to determine the emf. The loop and the solenoid share the same central axis. Hence, by symmetry, the electric field is tangent to the loop and has the same magnitude E all the way around its circumference. We can therefore use Eq. (29.9) to find E.

EXECUTE: (a) From Eq. (29.8), the induced emf is

$$\begin{aligned}\mathcal{E} &= -\frac{d\Phi_B}{dt} = -\mu_0 n A \frac{dI}{dt}\\ &= -(4\pi \times 10^{-7}\ \text{Wb/A}\cdot\text{m})(500\ \text{turns/m})\\ &\quad \times (4.0 \times 10^{-4}\ \text{m}^2)(100\ \text{A/s})\\ &= -25 \times 10^{-6}\ \text{Wb/s} = -25 \times 10^{-6}\ \text{V} = -25\ \mu\text{V}\end{aligned}$$

(b) By symmetry the line integral $\oint \vec{E} \cdot d\vec{l}$ has absolute value $2\pi r E$ no matter which direction we integrate around the loop. This is equal to the absolute value of the emf, so

$$E = \frac{|\mathcal{E}|}{2\pi r} = \frac{25 \times 10^{-6}\ \text{V}}{2\pi(2.0 \times 10^{-2}\ \text{m})} = 2.0 \times 10^{-4}\ \text{V/m}$$

EVALUATE: In Fig. 29.18b the magnetic flux *into* the plane of the figure is increasing. According to the right-hand rule for induced emf (Fig. 29.6), a positive emf would be clockwise around the loop; the negative sign of $\mathcal{E}$ shows that the emf is in the counterclockwise direction. Can you also show this by using Lenz's law?

TEST YOUR UNDERSTANDING OF SECTION 29.5 If you wiggle a magnet back and forth in your hand, are you generating an electric field? If so, is this field conservative? ❙

29.6 EDDY CURRENTS

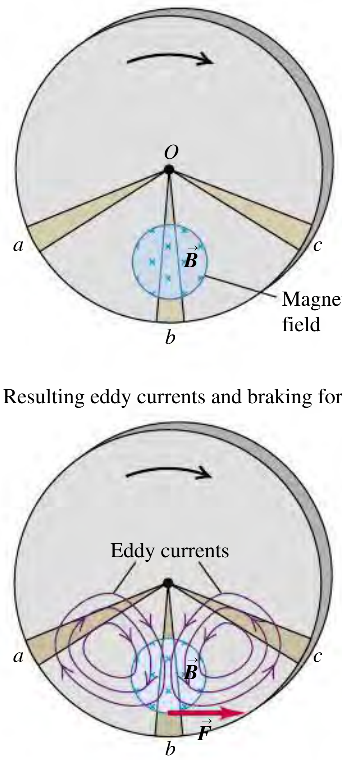
In the examples of induction effects that we have studied, the induced currents have been confined to well-defined paths in conductors and other components forming a circuit. However, many pieces of electrical equipment contain masses of metal moving in magnetic fields or located in changing magnetic fields. In situations like these we can have induced currents that circulate throughout the volume of a material. Because their flow patterns resemble swirling eddies in a river, we call these **eddy currents.**

As an example, consider a metallic disk rotating in a magnetic field perpendicular to the plane of the disk but confined to a limited portion of the disk's area, as shown in **Fig. 29.20a**. Sector Ob is moving across the field and has an emf induced in it. Sectors Oa and Oc are not in the field, but they provide return conducting paths for charges displaced along Ob to return from b to O. The result is a circulation of eddy currents in the disk, somewhat as sketched in Fig. 29.20b.

29.20 Eddy currents induced in a rotating metal disk.

(a) Metal disk rotating through a magnetic field

(b) Resulting eddy currents and braking force



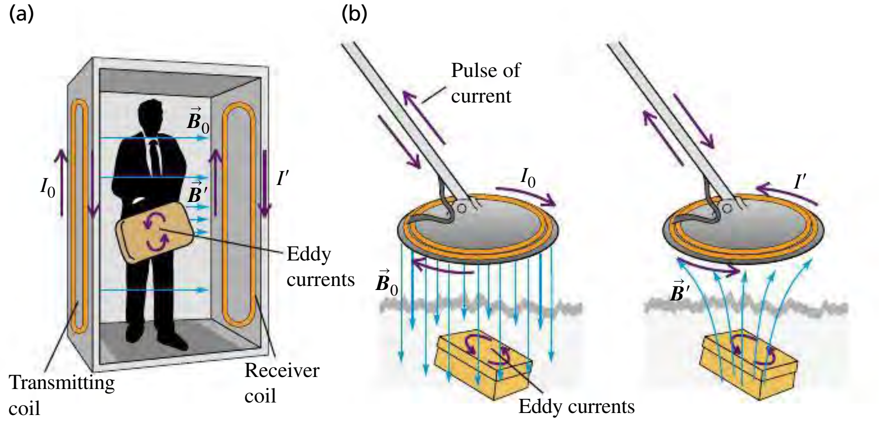
We can use Lenz's law to decide on the direction of the induced current in the neighborhood of sector Ob. This current must experience a magnetic force $\vec{F} = I\vec{L} \times \vec{B}$ that *opposes* the rotation of the disk, and so this force must be to the right in Fig. 29.20b. Since $\vec{B}$ is directed into the plane of the disk, the current and hence $\vec{L}$ have downward components. The return currents lie outside the field, so they do not experience magnetic forces. The interaction between the eddy currents and the field causes a braking action on the disk. Such effects can be used to stop the rotation of a circular saw quickly when the power is turned off. Eddy current braking is used on some electrically powered rapid-transit vehicles. Electromagnets mounted in the cars induce eddy currents in the rails; the resulting magnetic fields cause braking forces on the electromagnets and thus on the cars.

Eddy currents have many other practical uses. In induction furnaces, eddy currents are used to heat materials in completely sealed containers for processes in which it is essential to avoid the slightest contamination of the materials. The metal detectors used at airport security checkpoints (**Fig. 29.21a**) operate by detecting eddy currents induced in metallic objects. Similar devices (Fig. 29.21b) are used to find buried treasure such as bottlecaps and lost pennies.

DEMO

Eddy currents also have undesirable effects. In an alternating-current transformer, coils wrapped around an iron core carry a sinusoidally varying current. The resulting eddy currents in the core waste energy through I^2R heating and set up an unwanted opposing emf in the coils. To minimize these effects, the core

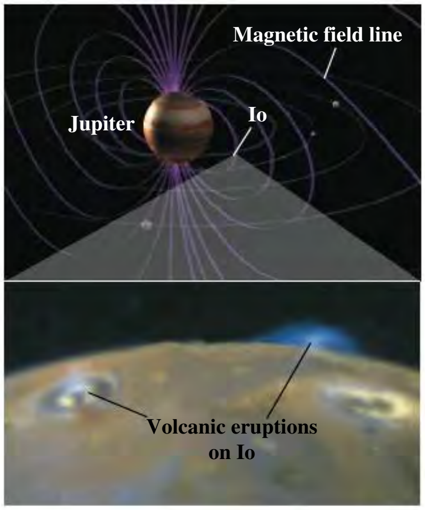
29.21 (a) A metal detector at an airport security checkpoint generates an alternating magnetic field $\vec{B}_0$. This induces eddy currents in a conducting object carried through the detector. The eddy currents in turn produce an alternating magnetic field $\vec{B}'$, which induces a current in the detector's receiver coil. (b) Portable metal detectors work on the same principle.



is designed so that the paths for eddy currents are as narrow as possible. We'll describe how this is done when we discuss transformers in Section 31.6.

TEST YOUR UNDERSTANDING OF SECTION 29.6 Suppose that the magnetic field in Fig. 29.20 were directed out of the plane of the figure and the disk were rotating counterclockwise. Compared to the directions of the force $\vec{F}$ and the eddy currents shown in Fig. 29.20b, what would the new directions be? (i) The force $\vec{F}$ and the eddy currents would both be in the same direction; (ii) the force $\vec{F}$ would be in the same direction, but the eddy currents would be in the opposite direction; (iii) the force $\vec{F}$ would be in the opposite direction, but the eddy currents would be in the same direction; (iv) the force $\vec{F}$ and the eddy currents would be in the opposite directions. ❙

Application Eddy Currents Help Power Io's Volcanoes Jupiter's moon Io is slightly larger than the earth's moon. It moves at more than 60,000 km/h through Jupiter's intense magnetic field (about ten times stronger than the earth's field), which sets up strong eddy currents within Io that dissipate energy at a rate of 10^{12} W. This dissipated energy helps to heat Io's interior and causes volcanic eruptions on its surface, as shown in the lower close-up image. (Gravitational effects from Jupiter cause even more heating.)



29.7 DISPLACEMENT CURRENT AND MAXWELL'S EQUATIONS

We have seen that a varying magnetic field gives rise to an induced electric field. In one of the more remarkable examples of the symmetry of nature, it turns out that a varying *electric* field gives rise to a *magnetic* field. This effect is of tremendous importance, for it turns out to explain the existence of radio waves, gamma rays, and visible light, as well as all other forms of electromagnetic waves.

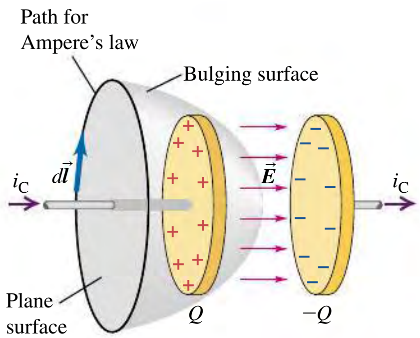
Generalizing Ampere's Law

To see the origin of the relationship between varying electric fields and magnetic fields, let's return to Ampere's law as given in Section 28.6, Eq. (28.20):

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 I_{\text{encl}}$$

The problem with Ampere's law in this form is that it is *incomplete*. To see why, let's consider the process of charging a capacitor (**Fig. 29.22**). Conducting wires lead current i_C into one plate and out of the other; the charge Q increases, and the electric field $\vec{E}$ between the plates increases. The notation i_C indicates *conduction* current to distinguish it from another kind of current we are about to encounter, called *displacement* current i_D. We use lowercase i's and v's to denote instantaneous values of currents and potential differences, respectively, that may vary with time.

29.22 Parallel-plate capacitor being charged. The conduction current through the plane surface is i_C, but there is no conduction current through the surface that bulges out to pass between the plates. The two surfaces have a common boundary, so this difference in I_{encl} leads to an apparent contradiction in applying Ampere's law.



Let's apply Ampere's law to the circular path shown. The integral $\oint \vec{B} \cdot d\vec{l}$ around this path equals $\mu_0 I_{\text{encl}}$. For the plane circular area bounded by the circle, I_{encl} is just the current i_C in the left conductor. But the surface that bulges out to the right is bounded by the same circle, and the current through that surface is zero. So $\oint \vec{B} \cdot d\vec{l}$ is equal to $\mu_0 i_C$, and at the same time it is equal to zero! This is a clear contradiction.

However, something else is happening on the bulged-out surface. As the capacitor charges, the electric field $\vec{E}$ and the electric *flux* Φ_E through the surface are increasing. We can determine their rates of change in terms of the charge and current. The instantaneous charge is $q = Cv$, where C is the capacitance and v is the instantaneous potential difference. For a parallel-plate capacitor, $C = \epsilon_0 A/d$, where A is the plate area and d is the spacing. The potential difference v between plates is $v = Ed$, where E is the electric-field magnitude between plates. (We ignore fringing and assume that $\vec{E}$ is uniform in the region between the plates.) If this region is filled with a material with permittivity ϵ, we replace ϵ_0 by ϵ everywhere; we'll use ϵ in the following discussion.

Substituting these expressions for C and v into $q = Cv$, we can express the capacitor charge q in terms of the electric flux $\Phi_E = EA$ through the surface:

$$q = Cv = \frac{\epsilon A}{d}(Ed) = \epsilon EA = \epsilon\Phi_E \tag{29.12}$$

As the capacitor charges, the rate of change of q is the conduction current, $i_C = dq/dt$. Taking the derivative of Eq. (29.12) with respect to time, we get

$$i_C = \frac{dq}{dt} = \epsilon \frac{d\Phi_E}{dt} \tag{29.13}$$

Stretching our imagination a bit, we invent a fictitious **displacement current** i_D in the region between the plates, defined as

$$\underbrace{i_D}_{\text{Displacement current through an area}} = \underbrace{\epsilon}_{\text{Permittivity of material in area}} \underbrace{\frac{d\Phi_E}{dt}}_{\text{Time rate of change of electric flux through area}} \tag{29.14}$$

That is, we imagine that the changing flux through the curved (bulged-out) surface in Fig. 29.22 is equivalent, in Ampere's law, to a conduction current through that surface. We include this fictitious current, along with the real conduction current i_C, in Ampere's law:

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 (i_C + i_D)_{\text{encl}} \quad \text{(generalized Ampere's law)} \tag{29.15}$$

Ampere's law in this form is obeyed no matter which surface we use in Fig. 29.22. For the flat surface, i_D is zero; for the curved surface, i_C is zero; and i_C for the flat surface equals i_D for the curved surface. Equation (29.15) remains valid in a magnetic material, provided that the magnetization is proportional to the external field and we replace μ_0 by μ.

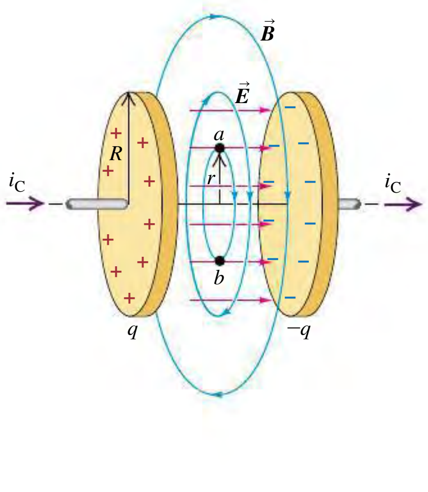
The fictitious displacement current i_D was invented in 1865 by the Scottish physicist James Clerk Maxwell. There is a corresponding *displacement current density* $j_D = i_D/A$; using $\Phi_E = EA$ and dividing Eq. (29.14) by A, we find

$$j_D = \epsilon \frac{dE}{dt} \tag{29.16}$$

We have pulled the concept out of thin air, as Maxwell did, but we see that it enables us to save Ampere's law in situations such as that in Fig. 29.22.

Another benefit of displacement current is that it lets us generalize Kirchhoff's junction rule, discussed in Section 26.2. Considering the left plate of the capacitor, we have conduction current into it but none out of it. But when we include the displacement current, we have conduction current coming in one side and an equal displacement current coming out the other side. With this generalized meaning of the term "current," we can speak of current going *through* the capacitor.

29.23 A capacitor being charged by a current i_C has a displacement current equal to i_C between the plates, with displacement-current density $j_D = \epsilon\, dE/dt$. This can be regarded as the source of the magnetic field between the plates.



The Reality of Displacement Current

You might well ask at this point whether displacement current has any real physical significance or whether it is just a ruse to satisfy Ampere's law and Kirchhoff's junction rule. Here's a fundamental experiment that helps to answer that question. We take a plane circular area between the capacitor plates (**Fig. 29.23**). If displacement current really plays the role in Ampere's law that we have claimed, then there ought to be a magnetic field in the region between the plates while the capacitor is charging. We can use our generalized Ampere's law, including displacement current, to predict what this field should be.

To be specific, let's picture round capacitor plates with radius R. To find the magnetic field at a point in the region between the plates at a distance r from the axis, we apply Ampere's law to a circle of radius r passing through the point, with $r < R$. This circle passes through points a and b in Fig. 29.23. The total current enclosed by the circle is j_D times its area, or $(i_D/\pi R^2)(\pi r^2)$. The integral

$\oint \vec{B} \cdot d\vec{l}$ in Ampere's law is just B times the circumference $2\pi r$ of the circle, and because $i_D = i_C$ for the charging capacitor, Ampere's law becomes

$$\oint \vec{B} \cdot d\vec{l} = 2\pi r B = \mu_0 \frac{r^2}{R^2} i_C \qquad \text{or}$$

$$B = \frac{\mu_0}{2\pi} \frac{r}{R^2} i_C \tag{29.17}$$

This result predicts that in the region between the plates $\vec{B}$ is zero at the axis and increases linearly with distance from the axis. A similar calculation shows that *outside* the region between the plates (that is, for $r > R$), $\vec{B}$ is the same as though the wire were continuous and the plates not present at all.

When we *measure* the magnetic field in this region, we find that it really is there and that it behaves just as Eq. (29.17) predicts. This confirms directly the role of displacement current as a source of magnetic field. It is now established beyond reasonable doubt that Maxwell's displacement current, far from being just an artifice, is a fundamental fact of nature.

Maxwell's Equations of Electromagnetism

We are now in a position to wrap up in a single package *all* of the relationships between electric and magnetic fields and their sources. This package consists of four equations, called **Maxwell's equations.** Maxwell did not discover all of these equations single-handedly (though he did develop the concept of displacement current). But he did put them together and recognized their significance, particularly in predicting the existence of electromagnetic waves.

For now we'll state Maxwell's equations in their simplest form, for the case in which we have charges and currents in otherwise empty space. In Chapter 32 we'll discuss how to modify these equations if a dielectric or a magnetic material is present.

Two of Maxwell's equations involve an integral of $\vec{E}$ or $\vec{B}$ over a closed surface. The first is simply Gauss's law for electric fields, Eq. (22.8):

Gauss's law for $\vec{E}$:

$$\underbrace{\oint \vec{E} \cdot d\vec{A}}_{\text{Flux of electric field through a closed surface}} = \frac{Q_{\text{encl}}}{\epsilon_0} \tag{29.18}$$

Q_{encl}: Charge enclosed by surface; ϵ_0: Electric constant

The second is the analogous relationship for *magnetic* fields, Eq. (27.8):

Gauss's law for $\vec{B}$:

$$\underbrace{\oint \vec{B} \cdot d\vec{A}}_{\text{Flux of magnetic field through any closed surface ...}} = 0 \tag{29.19}$$

0: ... equals zero.

This statement means, among other things, that there are no magnetic monopoles (single magnetic charges) to act as sources of magnetic field.

The third and fourth equations involve a line integral of $\vec{E}$ or $\vec{B}$ around a closed path. Faraday's law states that a changing magnetic flux acts as a source of electric field:

Faraday's law for a stationary integration path:

$$\underbrace{\oint \vec{E} \cdot d\vec{l}}_{\text{Line integral of electric field around path}} = \underbrace{-\frac{d\Phi_B}{dt}}_{\text{Negative of the time rate of change of magnetic flux through path}} \tag{29.20}$$

If there is a changing magnetic field, the line integral in Eq. (29.20)—which must be carried out over a *stationary* closed path—is not zero. Thus the $\vec{E}$ field produced by a changing $\vec{B}$ is not conservative.

The fourth and final equation is Ampere's law including displacement current. It states that both a conduction current and a changing electric flux act as sources of magnetic field:

Ampere's law for a stationary integration path:

$$\underbrace{\oint \vec{B} \cdot d\vec{l}}_{\text{Line integral of magnetic field around path}} = \underbrace{\mu_0}_{\text{Magnetic constant}}\Big(\underbrace{i_C}_{\text{Conduction current through path}} + \underbrace{\epsilon_0}_{\text{Electric constant}}\underbrace{\frac{d\Phi_E}{dt}}_{\text{Time rate of change of electric flux through path}}\Big)_{\text{encl}} \tag{29.21}$$

Displacement current through path

It's worthwhile to look more carefully at the electric field $\vec{E}$ and its role in Maxwell's equations. In general, the total $\vec{E}$ field at a point in space can be the superposition of an electrostatic field $\vec{E}_c$ caused by a distribution of charges at rest and a magnetically induced, nonelectrostatic field $\vec{E}_n$. That is,

$$\vec{E} = \vec{E}_c + \vec{E}_n$$

The electrostatic part $\vec{E}_c$ is *always* conservative, so $\oint \vec{E}_c \cdot d\vec{l} = 0$. This conservative part of the field does not contribute to the integral in Faraday's law, so we can take $\vec{E}$ in Eq. (29.20) to be the *total* electric field $\vec{E}$, including both the part $\vec{E}_c$ due to charges and the magnetically induced part $\vec{E}_n$. Similarly, the nonconservative part $\vec{E}_n$ of the $\vec{E}$ field does not contribute to the integral in Gauss's law, because this part of the field is not caused by static charges. Hence $\oint \vec{E}_n \cdot d\vec{A}$ is always zero. We conclude that in all the Maxwell equations, $\vec{E}$ is the total electric field; these equations don't distinguish between conservative and nonconservative fields.

Symmetry in Maxwell's Equations

There is a remarkable symmetry in Maxwell's four equations. In empty space where there is no charge, the first two equations (Eqs. (29.18) and (29.19)) are identical in form, one containing $\vec{E}$ and the other containing $\vec{B}$ (**Fig. 29.24**). When we compare the second two equations, Eq. (29.20) says that a changing magnetic flux creates an electric field, and Eq. (29.21) says that a changing electric flux creates a magnetic field. In empty space, where there is no conduction current, $i_C = 0$ and the two equations have the same form, apart from a numerical constant and a negative sign, with the roles of $\vec{E}$ and $\vec{B}$ exchanged.

29.24 Maxwell's equations in empty space are highly symmetric.

In empty space there are no charges, so the fluxes of $\vec{E}$ and $\vec{B}$ through any closed surface are equal to zero.

$$\oint \vec{E} \cdot d\vec{A} = 0$$

$$\oint \vec{B} \cdot d\vec{A} = 0$$

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt}$$

$$\oint \vec{B} \cdot d\vec{l} = \mu_0\epsilon_0\frac{d\Phi_E}{dt}$$

In empty space there are no conduction currents, so the line integrals of $\vec{E}$ and $\vec{B}$ around any closed path are related to the rate of change of flux of the other field.

We can rewrite Eqs. (29.20) and (29.21) in a different but equivalent form by introducing the definitions of magnetic flux, $\Phi_B = \int \vec{B} \cdot d\vec{A}$, and electric flux, $\Phi_E = \int \vec{E} \cdot d\vec{A}$, respectively. In empty space, where there is no charge or conduction current, $i_C = 0$ and $Q_{\text{encl}} = 0$, and we have

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d}{dt}\int \vec{B} \cdot d\vec{A} \tag{29.22}$$

$$\oint \vec{B} \cdot d\vec{l} = \mu_0\epsilon_0\frac{d}{dt}\int \vec{E} \cdot d\vec{A} \tag{29.23}$$

Again we notice the symmetry between the roles of $\vec{E}$ and $\vec{B}$ in these expressions.

The most remarkable feature of these equations is that a time-varying field of *either* kind induces a field of the other kind in neighboring regions of space. Maxwell recognized that these relationships predict the existence of electromagnetic disturbances consisting of time-varying electric and magnetic fields that travel or *propagate* from one region of space to another, even if no matter is present in the intervening space. Such disturbances, called *electromagnetic waves,* provide the physical basis for light, radio and television waves, infrared, ultraviolet, and x rays. We will return to this vitally important topic in Chapter 32.

Although it may not be obvious, *all* the basic relationships between fields and their sources are contained in Maxwell's equations. We can derive Coulomb's law from Gauss's law, we can derive the law of Biot and Savart from Ampere's law, and so on. When we add the equation that defines the $\vec{E}$ and $\vec{B}$ fields in terms of the forces that they exert on a charge q, namely,

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \tag{29.24}$$

we have *all* the fundamental relationships of electromagnetism!

Maxwell's equations would have even greater symmetry between the $\vec{E}$ and $\vec{B}$ fields if single magnetic charges (magnetic monopoles) existed. The right side of Eq. (29.19) would contain the total *magnetic* charge enclosed by the surface, and the right side of Eq. (29.20) would include a magnetic monopole current term. However, no magnetic monopoles have yet been found.

In conciseness and generality, Maxwell's equations are in the same league with Newton's laws of motion and the laws of thermodynamics. Indeed, a major goal of science is learning how to express very broad and general relationships in a concise and compact form. Maxwell's synthesis of electromagnetism stands as a towering intellectual achievement, comparable to the Newtonian synthesis we described at the end of Section 13.5 and to the development of relativity and quantum mechanics in the 20th century.

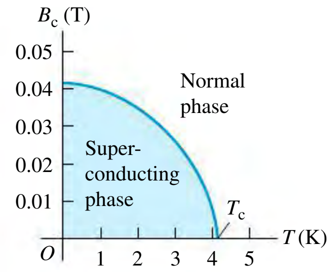
TEST YOUR UNDERSTANDING OF SECTION 29.7 (a) Which of Maxwell's equations explains how a credit card reader works? (b) Which one describes how a wire carrying a steady current generates a magnetic field? ❙

29.8 SUPERCONDUCTIVITY

The most familiar property of a superconductor is the sudden disappearance of all electrical resistance when the material is cooled below a temperature called the *critical temperature,* denoted by T_c. We discussed this behavior and the circumstances of its discovery in Section 25.2. But superconductivity is far more than just the absence of measurable resistance. As we'll see in this section, superconductors also have extraordinary *magnetic* properties.

The first hint of unusual magnetic properties was the discovery that for any superconducting material the critical temperature T_c changes when the material is placed in an externally produced magnetic field $\vec{B}_0$. **Figure 29.25** shows this dependence for mercury, the first element in which superconductivity was observed. As the external field magnitude B_0 increases, the superconducting transition occurs at lower and lower temperature. When B_0 is greater than 0.0412 T, *no* superconducting transition occurs. The minimum magnitude of magnetic field that is needed to eliminate superconductivity at a temperature below T_c is called the *critical field,* denoted by B_c.

29.25 Phase diagram for pure mercury, showing the critical magnetic field B_c and its dependence on temperature. Superconductivity is impossible above the critical temperature T_c. The curves for other superconducting materials are similar but with different numerical values.

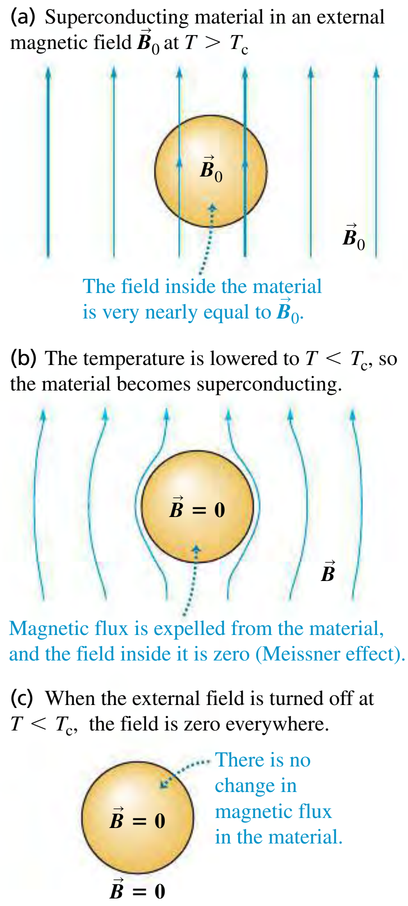


The Meissner Effect

Another aspect of the magnetic behavior of superconductors appears if we place a homogeneous sphere of a superconducting material in a uniform applied magnetic field $\vec{B}_0$ at a temperature T greater than T_c. The material is then in the normal phase, not the superconducting phase (**Fig. 29.26a**). Now we lower the temperature until the superconducting transition occurs. (We assume that the magnitude of $\vec{B}_0$ is not large enough to prevent the phase transition.) What happens to the field?

Measurements of the field outside the sphere show that the field lines become distorted as in Fig. 29.26b. There is no longer any field inside the material, except possibly in a very thin surface layer a hundred or so atoms thick. If a coil is wrapped

29.26 A superconducting material (a) above the critical temperature and (b), (c) below the critical temperature.

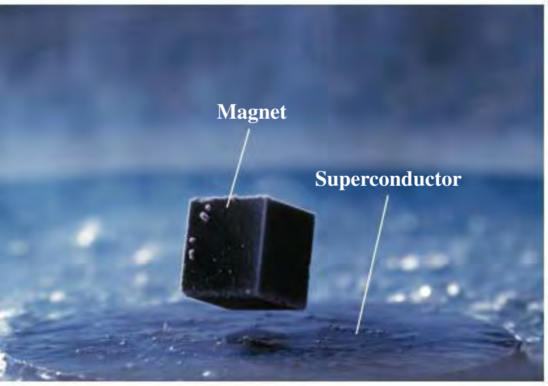


around the sphere, the emf induced in the coil shows that during the superconducting transition the magnetic flux through the coil decreases from its initial value to zero; this is consistent with the absence of field inside the material. Finally, if the field is now turned off while the material is still in its superconducting phase, no emf is induced in the coil, and measurements show no field outside the sphere (Fig. 29.26c).

We conclude that during a superconducting transition in the presence of the field $\vec{B}_0$, all of the magnetic flux is expelled from the bulk of the sphere, and the magnetic flux Φ_B through the coil becomes zero. This expulsion of magnetic flux is called the *Meissner effect*. As Fig. 29.26b shows, this expulsion crowds the magnetic field lines closer together to the side of the sphere, increasing $\vec{B}$ there.

Superconductor Levitation and Other Applications

29.27 A superconductor exerts a repulsive force on a magnet, supporting the magnet in midair.



The diamagnetic nature of a superconductor has some interesting *mechanical* consequences. A paramagnetic or ferromagnetic material is attracted by a permanent magnet because the magnetic dipoles in the material align with the nonuniform magnetic field of the permanent magnet. (We discussed this in Section 27.7.) For a diamagnetic material the magnetization is in the opposite sense, and a diamagnetic material is *repelled* by a permanent magnet. By Newton's third law the magnet is also repelled by the diamagnetic material. **Figure 29.27** shows the repulsion between a specimen of a high-temperature superconductor and a magnet; the magnet is supported ("levitated") by this repulsive magnetic force.

The behavior we have described is characteristic of what are called *type-I superconductors*. There is another class of superconducting materials called *type-II superconductors*. When such a material in the superconducting phase is placed in a magnetic field, the bulk of the material remains superconducting, but thin filaments of material, running parallel to the field, may return to the normal phase. Currents circulate around the boundaries of these filaments, and there *is* magnetic flux inside them. Type-II superconductors are used for electromagnets because they usually have much larger values of B_c than do type-I materials, permitting much larger magnetic fields without destroying the superconducting state. Type-II superconductors have *two* critical magnetic fields: The first, B_{c1}, is the field at which magnetic flux begins to enter the material, forming the filaments just described, and the second, B_{c2}, is the field at which the material becomes normal.

Superconducting electromagnets are in everyday use not only in research laboratories but also in medical MRI (magnetic resonance imaging) scanners. As we described in Section 27.7, scanning a patient through MRI requires a strong magnetic field to align the magnetic dipoles of the patient's atomic nuclei. A steady field of 1.5 T or more is needed, which is very difficult to produce with a conventional electromagnet, since this would require very high currents and hence very high energy losses due to resistance in the electromagnet coils. But with a superconducting electromagnet there is no resistive energy loss, and fields up to 10 T can routinely be attained.

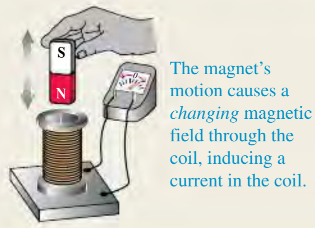
Very sensitive measurements of magnetic fields can be made with superconducting quantum interference devices (SQUIDs), which can detect changes in magnetic flux of less than 10^{-14} Wb; these devices have applications in medicine, geology, and other fields. The number of potential uses for superconductors has increased greatly since the discovery in 1987 of high-temperature superconductors. These materials have critical temperatures that are above the temperature of liquid nitrogen (about 77 K) and so are comparatively easy to attain. Development of practical applications of superconductor science promises to be an exciting chapter in contemporary technology.

CHAPTER 29 SUMMARY

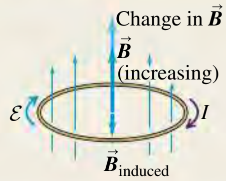
SOLUTIONS TO ALL EXAMPLES

Faraday's law: Faraday's law states that the induced emf in a closed loop equals the negative of the time rate of change of magnetic flux through the loop. This relationship is valid whether the flux change is caused by a changing magnetic field, motion of the loop, or both. (See Examples 29.1–29.6.)

$$\mathcal{E} = -\frac{d\Phi_B}{dt} \qquad (29.3)$$



Lenz's law: Lenz's law states that an induced current or emf always tends to oppose or cancel out the change that caused it. Lenz's law can be derived from Faraday's law and is often easier to use. (See Examples 29.7 and 29.8.)



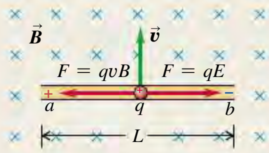
Motional emf: If a conductor moves in a magnetic field, a motional emf is induced. (See Examples 29.9 and 29.10.)

$$\mathcal{E} = vBL \qquad (29.6)$$

(conductor with length L moves in uniform $\vec{B}$ field, $\vec{L}$ and $\vec{v}$ both perpendicular to $\vec{B}$ and to each other)

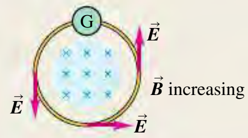
$$\mathcal{E} = \oint (\vec{v} \times \vec{B}) \cdot d\vec{l} \qquad (29.7)$$

(all or part of a closed loop moves in a $\vec{B}$ field)



Induced electric fields: When an emf is induced by a changing magnetic flux through a stationary conductor, there is an induced electric field $\vec{E}$ of nonelectrostatic origin. This field is nonconservative and cannot be associated with a potential. (See Example 29.11.)

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \qquad (29.10)$$



Displacement current and Maxwell's equations: A time-varying electric field generates displacement current i_D, which acts as a source of magnetic field in exactly the same way as conduction current. The relationships between electric and magnetic fields and their sources can be stated compactly in four equations, called Maxwell's equations. Together they form a complete basis for the relationship of $\vec{E}$ and $\vec{B}$ fields to their sources.

$$i_D = \epsilon \frac{d\Phi_E}{dt} \qquad (29.14)$$

(displacement current)

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q_{\text{encl}}}{\epsilon_0} \qquad (29.18)$$

(Gauss's law for $\vec{E}$ fields)

$$\oint \vec{B} \cdot d\vec{A} = 0 \qquad (29.19)$$

(Gauss's law for $\vec{B}$ fields)

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \qquad (29.20)$$

(Faraday's law)

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \left(i_C + \epsilon_0 \frac{d\Phi_E}{dt} \right)_{\text{encl}} \qquad (29.21)$$

(Ampere's law including displacement current)

BRIDGING PROBLEM A FALLING SQUARE LOOP

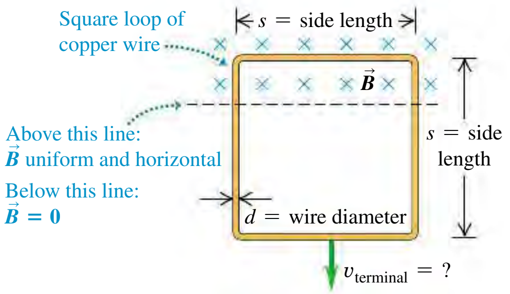
A vertically oriented square loop of copper wire falls from rest in a region in which the field $\vec{B}$ is horizontal, uniform, and perpendicular to the plane of the loop, into a field-free region (**Fig. 29.28**). The side length of the loop is s and the wire diameter is d. The resistivity of copper is ρ_R and the density of copper is ρ_m. If the loop reaches its terminal speed while its upper segment is still in the magnetic-field region, find an expression for the terminal speed.

SOLUTION GUIDE

IDENTIFY and SET UP

1. The motion of the loop through the magnetic field induces an emf and a current in the loop. The field then gives rise to a magnetic force on this current that opposes the downward force of gravity. The loop reaches terminal speed (it no longer accelerates) when the upward magnetic force balances the downward force of gravity.
2. Consider the case in which the entire loop is in the magnetic-field region. Is there an induced emf in this case? If so, what is its direction?
3. Consider the case in which only the upper segment of the loop is in the magnetic-field region. Is there an induced emf in this case? If so, what is its direction?
4. For the case in which there is an induced emf and hence an induced current, what is the direction of the magnetic force on each of the four sides of the loop? What is the direction of the *net* magnetic force on the loop?

29.28 A wire loop falling in a horizontal magnetic field $\vec{B}$. The plane of the loop is perpendicular to $\vec{B}$.



EXECUTE

5. For the case in which the loop is falling at speed v and there is an induced emf, find (i) the emf, (ii) the induced current, and (iii) the magnetic force on the loop in terms of its resistance R.
6. Find R and the mass of the loop in terms of the given information about the loop.
7. Use your results from steps 5 and 6 to find an expression for the terminal speed.

EVALUATE

8. How does the terminal speed depend on the magnetic-field magnitude B? Explain why this makes sense.

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q29.1 A sheet of copper is placed between the poles of an electromagnet with the magnetic field perpendicular to the sheet. When the sheet is pulled out, a considerable force is required, and the force required increases with speed. Explain. Is a force required also when the sheet is inserted between the poles? Explain.

Q29.2 In Fig. 29.8, if the angular speed ω of the loop is doubled, then the frequency with which the induced current changes direction doubles, and the maximum emf also doubles. Why? Does the torque required to turn the loop change? Explain.

Q29.3 Two circular loops lie side by side in the same plane. One is connected to a source that supplies an increasing current; the other is a simple closed ring. Is the induced current in the ring in the same direction as the current in the loop connected to the source, or opposite? What if the current in the first loop is decreasing? Explain.

Q29.4 For Eq. (29.6), show that if v is in meters per second, B in teslas, and L in meters, then the units of the right-hand side of the equation are joules per coulomb or volts (the correct SI units for $\mathcal{E}$).

Q29.5 A long, straight conductor passes through the center of a metal ring, perpendicular to its plane. If the current in the conductor increases, is a current induced in the ring? Explain.

Q29.6 A student asserted that if a permanent magnet is dropped down a vertical copper pipe, it eventually reaches a terminal velocity even if there is no air resistance. Why should this be? Or should it?

Q29.7 An airplane is in level flight over Antarctica, where the magnetic field of the earth is mostly directed upward away from the ground. As viewed by a passenger facing toward the front of the plane, is the left or the right wingtip at higher potential? Does your answer depend on the direction the plane is flying?

Q29.8 Consider the situation in Exercise 29.21. In part (a), find the direction of the force that the large circuit exerts on the small one. Explain how this result is consistent with Lenz's law.

Q29.9 A metal rectangle is close to a long, straight, current-carrying wire, with two of its sides parallel to the wire. If the current in the long wire is decreasing, is the rectangle repelled by or attracted to the wire? Explain why this result is consistent with Lenz's law.

Q29.10 A square conducting loop is in a region of uniform, constant magnetic field. Can the loop be rotated about an axis along one side and no emf be induced in the loop? Discuss, in terms of the orientation of the rotation axis relative to the magnetic-field direction.

Q29.11 Example 29.6 discusses the external force that must be applied to the slidewire to move it at constant speed. If there were a break in the left-hand end of the U-shaped conductor, how much force would be needed to move the slidewire at constant speed? As in the example, you can ignore friction.

Q29.12 In the situation shown in Fig. 29.18, would it be appropriate to ask how much *energy* an electron gains during a complete trip around the wire loop with current I'? Would it be appropriate to ask what *potential difference* the electron moves through during such a complete trip? Explain your answers.

Q29.13 A metal ring is oriented with the plane of its area perpendicular to a spatially uniform magnetic field that increases at a steady rate. If the radius of the ring is doubled, by what factor do (a) the emf induced in the ring and (b) the electric field induced in the ring change?

Q29.14 Small one-cylinder gasoline engines sometimes use a device called a *magneto* to supply current to the spark plug. A permanent magnet is attached to the flywheel, and a stationary coil is mounted adjacent to it. Explain how this device is able to generate current. What happens when the magnet passes the coil?

Q29.15 Does Lenz's law say that the induced current in a metal loop always flows to oppose the magnetic flux through that loop? Explain.

Q29.16 Does Faraday's law say that a large magnetic flux induces a large emf in a coil? Explain.

Q29.17 Can one have a displacement current as well as a conduction current within a conductor? Explain.

Q29.18 Your physics study partner asks you to consider a parallel-plate capacitor that has a dielectric completely filling the volume between the plates. He then claims that Eqs. (29.13) and (29.14) show that the conduction current in the dielectric equals the displacement current in the dielectric. Do you agree? Explain.

Q29.19 Match the mathematical statements of Maxwell's equations as given in Section 29.7 to these verbal statements. (a) Closed electric field lines are evidently produced only by changing magnetic flux. (b) Closed magnetic field lines are produced both by the motion of electric charge and by changing electric flux. (c) Electric field lines can start on positive charges and end on negative charges. (d) Evidently there are no magnetic monopoles on which to start and end magnetic field lines.

Q29.20 If magnetic monopoles existed, the right-hand side of Eq. (29.20) would include a term proportional to the current of magnetic monopoles. Suppose a steady monopole current is moving in a long straight wire. Sketch the *electric* field lines that such a current would produce.

Q29.21 A type-II superconductor in an external field between B_{c1} and B_{c2} has regions that contain magnetic flux and have resistance, and also has superconducting regions. What is the resistance of a long, thin cylinder of such material?

EXERCISES

Section 29.2 Faraday's Law

29.1 • A single loop of wire with an area of $0.0900\ \text{m}^2$ is in a uniform magnetic field that has an initial value of 3.80 T, is perpendicular to the plane of the loop, and is decreasing at a constant rate of $0.190\ \text{T/s}$. (a) What emf is induced in this loop? (b) If the loop has a resistance of $0.600\ \Omega$, find the current induced in the loop.

29.2 •• In a physics laboratory experiment, a coil with 200 turns enclosing an area of $12\ \text{cm}^2$ is rotated in 0.040 s from a position where its plane is perpendicular to the earth's magnetic field to a position where its plane is parallel to the field. The earth's magnetic field at the lab location is 6.0×10^{-5} T. (a) What is the total magnetic flux through the coil before it is rotated? After it is rotated? (b) What is the average emf induced in the coil?

29.3 •• **Search Coils and Credit Cards.** One practical way to measure magnetic field strength uses a small, closely wound coil called a *search coil*. The coil is initially held with its plane perpendicular to a magnetic field. The coil is then either quickly rotated a quarter-turn about a diameter or quickly pulled out of the field. (a) Derive the equation relating the total charge Q that flows through a search coil to the magnetic-field magnitude B. The search coil has N turns, each with area A, and the flux through the coil is decreased from its initial maximum value to zero in a time Δt. The resistance of the coil is R, and the total charge is $Q = I\Delta t$, where I is the average current induced by the change in flux. (b) In a credit card reader, the magnetic strip on the back of a credit card is rapidly "swiped" past a coil within the reader. Explain, using the same ideas that underlie the operation of a search coil, how the reader can decode the information stored in the pattern of magnetization on the strip. (c) Is it necessary that the credit card be "swiped" through the reader at exactly the right speed? Why or why not?

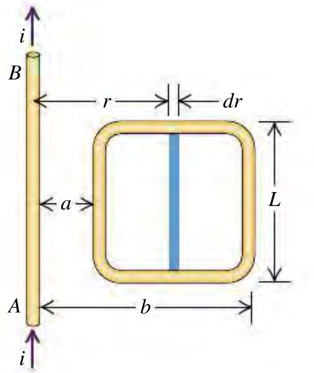
29.4 • A closely wound search coil (see Exercise 29.3) has an area of $3.20\ \text{cm}^2$, 120 turns, and a resistance of $60.0\ \Omega$. It is connected to a charge-measuring instrument whose resistance is $45.0\ \Omega$. When the coil is rotated quickly from a position parallel to a uniform magnetic field to a position perpendicular to the field, the instrument indicates a charge of 3.56×10^{-5} C. What is the magnitude of the field?

29.5 • A circular loop of wire with a radius of 12.0 cm and oriented in the horizontal xy-plane is located in a region of uniform magnetic field. A field of 1.5 T is directed along the positive z-direction, which is upward. (a) If the loop is removed from the field region in a time interval of 2.0 ms, find the average emf that will be induced in the wire loop during the extraction process. (b) If the coil is viewed looking down on it from above, is the induced current in the loop clockwise or counterclockwise?

29.6 • **CALC** A coil 4.00 cm in radius, containing 500 turns, is placed in a uniform magnetic field that varies with time according to $B = (0.0120\ \text{T/s})t + (3.00 \times 10^{-5}\ \text{T/s}^4)t^4$. The coil is connected to a $600\text{-}\Omega$ resistor, and its plane is perpendicular to the magnetic field. You can ignore the resistance of the coil. (a) Find the magnitude of the induced emf in the coil as a function of time. (b) What is the current in the resistor at time $t = 5.00$ s?

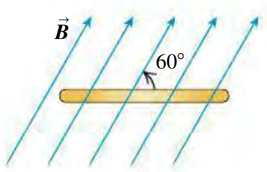
29.7 • **CALC** The current in the long, straight wire AB shown in **Fig. E29.7** is upward and is increasing steadily at a rate di/dt. (a) At an instant when the current is i, what are the magnitude and direction of the field $\vec{B}$ at a distance r to the right of the wire? (b) What is the flux $d\Phi_B$ through the narrow, shaded strip? (c) What is the total flux through the loop? (d) What is the induced emf in the loop? (e) Evaluate the numerical value of the induced emf if $a = 12.0$ cm, $b = 36.0$ cm, $L = 24.0$ cm, and $di/dt = 9.60$ A/s.

Figure **E29.7**



29.8 • CALC A flat, circular, steel loop of radius 75 cm is at rest in a uniform magnetic field, as shown in an edge-on view in **Fig. E29.8**. The field is changing with time, according to $B(t) = (1.4\ \mathrm{T})e^{-(0.057\ \mathrm{s}^{-1})t}$. (a) Find the emf induced in the loop as a function of time. (b) When is the induced emf equal to $\frac{1}{10}$ of its initial value? (c) Find the direction of the current induced in the loop, as viewed from above the loop.

Figure **E29.8**



29.9 • **Shrinking Loop.** A circular loop of flexible iron wire has an initial circumference of 165.0 cm, but its circumference is decreasing at a constant rate of 12.0 cm/s due to a tangential pull on the wire. The loop is in a constant, uniform magnetic field oriented perpendicular to the plane of the loop and with magnitude 0.500 T. (a) Find the emf induced in the loop at the instant when 9.0 s have passed. (b) Find the direction of the induced current in the loop as viewed looking along the direction of the magnetic field.

29.10 • A closely wound rectangular coil of 80 turns has dimensions of 25.0 cm by 40.0 cm. The plane of the coil is rotated from a position where it makes an angle of 37.0° with a magnetic field of 1.70 T to a position perpendicular to the field. The rotation takes 0.0600 s. What is the average emf induced in the coil?

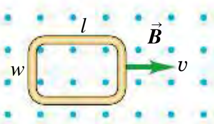
29.11 • CALC In a region of space, a magnetic field points in the $+x$-direction (toward the right). Its magnitude varies with position according to the formula $B_x = B_0 + bx$, where B_0 and b are positive constants, for $x \geq 0$. A flat coil of area A moves with uniform speed v from right to left with the plane of its area always perpendicular to this field. (a) What is the emf induced in this coil while it is to the right of the origin? (b) As viewed from the origin, what is the direction (clockwise or counterclockwise) of the current induced in the coil? (c) If instead the coil moved from left to right, what would be the answers to parts (a) and (b)?

29.12 • In many magnetic resonance imaging (MRI) systems, the magnetic field is produced by a superconducting magnet that must be kept cooled below the superconducting transition temperature. If the cryogenic cooling system fails, the magnet coils may lose their superconductivity and the strength of the magnetic field will rapidly decrease, or *quench*. The dissipation of energy as heat in the now-nonsuperconducting magnet coils can cause a rapid boil-off of the cryogenic liquid (usually liquid helium) that is used for cooling. Consider a superconducting MRI magnet for which the magnetic field decreases from 8.0 T to nearly 0 in 20 s. What is the average emf induced in a circular wedding ring of diameter 2.2 cm if the ring is at the center of the MRI magnet coils and the original magnetic field is perpendicular to the plane that is encircled by the ring?

29.13 •• The armature of a small generator consists of a flat, square coil with 120 turns and sides with a length of 1.60 cm. The coil rotates in a magnetic field of 0.0750 T. What is the angular speed of the coil if the maximum emf produced is 24.0 mV?

29.14 • A flat, rectangular coil of dimensions l and w is pulled with uniform speed v through a uniform magnetic field B with the plane of its area perpendicular to the field (**Fig. E29.14**). (a) Find the emf induced in this coil. (b) If the speed and magnetic field are both tripled, what is the induced emf?

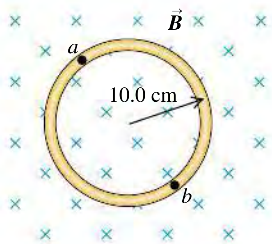
Figure **E29.14**



Section 29.3 Lenz's Law

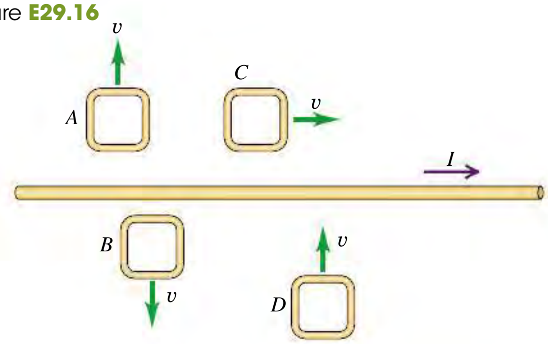
29.15 • A circular loop of wire is in a region of spatially uniform magnetic field, as shown in **Fig. E29.15**. The magnetic field is directed into the plane of the figure. Determine the direction (clockwise or counterclockwise) of the induced current in the loop when (a) B is increasing; (b) B is decreasing; (c) B is constant with value B_0. Explain your reasoning.

Figure **E29.15**



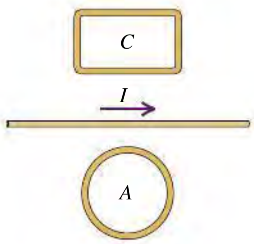
29.16 • The current I in a long, straight wire is constant and is directed toward the right as in **Fig. E29.16**. Conducting loops A, B, C, and D are moving, in the directions shown, near the wire. (a) For each loop, is the direction of the induced current clockwise or counterclockwise, or is the induced current zero? (b) For each loop, what is the direction of the net force that the wire exerts on the loop? Give your reasoning for each answer.

Figure **E29.16**



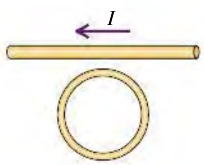
29.17 • Two closed loops A and C are close to a long wire carrying a current I (**Fig. E29.17**). (a) Find the direction (clockwise or counterclockwise) of the current induced in each loop if I is steadily decreasing. (b) While I is decreasing, what is the direction of the net force that the wire exerts on each loop? Explain how you obtain your answer.

Figure **E29.17**



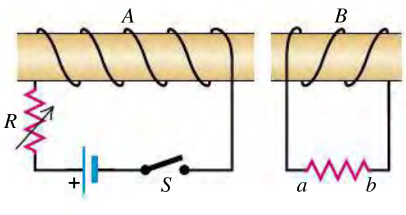
29.18 • The current in **Fig. E29.18** obeys the equation $I(t) = I_0 e^{-bt}$, where $b > 0$. Find the direction (clockwise or counterclockwise) of the current induced in the round coil for $t > 0$.

Figure **E29.18**



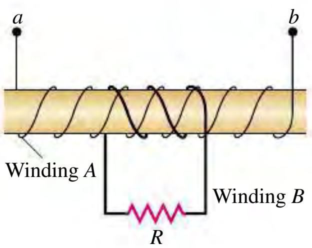
29.19 • Using Lenz's law, determine the direction of the current in resistor ab of **Fig. E29.19** when (a) switch S is opened after having been closed for several minutes; (b) coil B is brought closer to coil A with the switch closed; (c) the resistance of R is decreased while the switch remains closed.

Figure **E29.19**



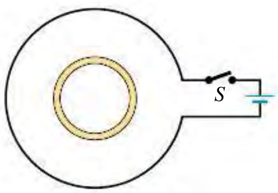
29.20 • A cardboard tube is wrapped with two windings of insulated wire wound in opposite directions, as shown in **Fig. E29.20**. Terminals a and b of winding A may be connected to a battery through a reversing switch. State whether the induced current in the resistor R is from left to right or from right to left in the following circumstances: (a) the current in winding A is from a to b and is increasing; (b) the current in winding A is from b to a and is decreasing; (c) the current in winding A is from b to a and is increasing.

Figure **E29.20**



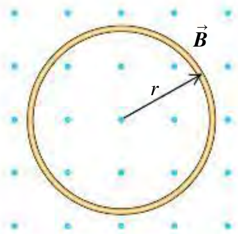
29.21 • A small, circular ring is inside a larger loop that is connected to a battery and a switch (**Fig. E29.21**). Use Lenz's law to find the direction of the current induced in the small ring (a) just after switch S is closed; (b) after S has been closed a long time; (c) just after S has been reopened after it was closed for a long time.

Figure **E29.21**



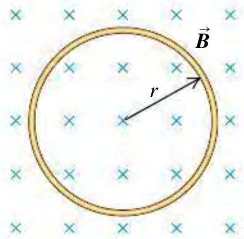
29.22 • A circular loop of wire with radius $r = 0.0480$ m and resistance $R = 0.160\ \Omega$ is in a region of spatially uniform magnetic field, as shown in **Fig. E29.22**. The magnetic field is directed out of the plane of the figure. The magnetic field has an initial value of 8.00 T and is decreasing at a rate of $dB/dt = -0.680$ T/s. (a) Is the induced current in the loop clockwise or counterclockwise? (b) What is the rate at which electrical energy is being dissipated by the resistance of the loop?

Figure **E29.22**



29.23 • **CALC** A circular loop of wire with radius $r = 0.0250$ m and resistance $R = 0.390\ \Omega$ is in a region of spatially uniform magnetic field, as shown in **Fig. E29.23**. The magnetic field is directed into the plane of the figure. At $t = 0$, $B = 0$. The magnetic field then begins increasing, with $B(t) = (0.380\ \text{T/s}^3)t^3$. What is the current in the loop (magnitude and direction) at the instant when $B = 1.33$ T?

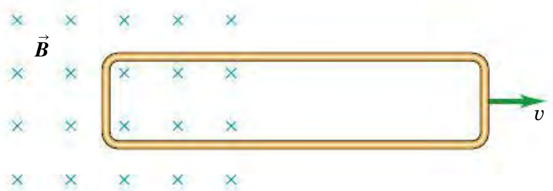
Figure **E29.23**



Section 29.4 Motional Electromotive Force

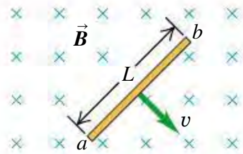
29.24 • A rectangular loop of wire with dimensions 1.50 cm by 8.00 cm and resistance $R = 0.600\ \Omega$ is being pulled to the right out of a region of uniform magnetic field. The magnetic field has magnitude $B = 2.40$ T and is directed into the plane of **Fig. E29.24**. At the instant when the speed of the loop is 3.00 m/s and it is still partially in the field region, what force (magnitude and direction) does the magnetic field exert on the loop?

Figure **E29.24**



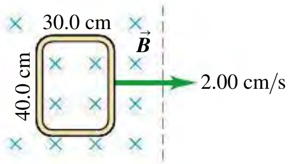
29.25 • In **Fig. E29.25** a conducting rod of length $L = 30.0$ cm moves in a magnetic field $\vec{B}$ of magnitude 0.450 T directed into the plane of the figure. The rod moves with speed $v = 5.00$ m/s in the direction shown. (a) What is the potential difference between the ends of the rod? (b) Which point, a or b, is at higher potential? (c) When the charges in the rod are in equilibrium, what are the magnitude and direction of the electric field within the rod? (d) When the charges in the rod are in equilibrium, which point, a or b, has an excess of positive charge? (e) What is the potential difference across the rod if it moves (i) parallel to ab and (ii) directly out of the page?

Figure **E29.25**



29.26 • A rectangle measuring 30.0 cm by 40.0 cm is located inside a region of a spatially uniform magnetic field of 1.25 T, with the field perpendicular to the plane of the coil (**Fig. E29.26**). The coil is pulled out at a steady rate of 2.00 cm/s traveling perpendicular to the field lines. The region of the field ends abruptly as shown. Find the emf induced in this coil when it is (a) all inside the field; (b) partly inside the field; (c) all outside the field.

Figure **E29.26**

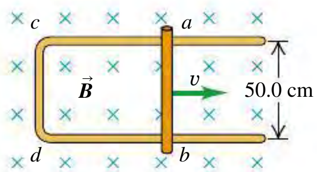


29.27 • **Are Motional emfs a Practical Source of Electricity?** How fast (in m/s and mph) would a 5.00-cm copper bar have to move at right angles to a 0.650-T magnetic field to generate 1.50 V (the same as a AA battery) across its ends? Does this seem like a practical way to generate electricity?

29.28 • **Motional emfs in Transportation.** Airplanes and trains move through the earth's magnetic field at rather high speeds, so it is reasonable to wonder whether this field can have a substantial effect on them. We shall use a typical value of 0.50 G for the earth's field. (a) The French TGV train and the Japanese "bullet train" reach speeds of up to 180 mph moving on tracks about 1.5 m apart. At top speed moving perpendicular to the earth's magnetic field, what potential difference is induced across the tracks as the wheels roll? Does this seem large enough to produce noticeable effects? (b) The Boeing 747-400 aircraft has a wingspan of 64.4 m and a cruising speed of 565 mph. If there is no wind blowing (so that this is also their speed relative to the ground), what is the maximum potential difference that could be induced between the opposite tips of the wings? Does this seem large enough to cause problems with the plane?

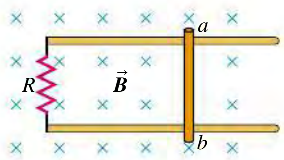
29.29 • The conducting rod ab shown in **Fig. E29.29** makes contact with metal rails ca and db. The apparatus is in a uniform magnetic field of 0.800 T, perpendicular to the plane of the figure. (a) Find the magnitude of the emf induced in the rod when it is moving toward the right with a speed 7.50 m/s. (b) In what direction does the current flow in the rod? (c) If the resistance of the circuit $abdc$ is 1.50 Ω (assumed to be constant), find the force (magnitude and direction) required to keep the rod moving to the right with a constant speed of 7.50 m/s. You can ignore friction. (d) Compare the rate at which mechanical work is done by the force (Fv) with the rate at which thermal energy is developed in the circuit (I^2R).

Figure **E29.29**



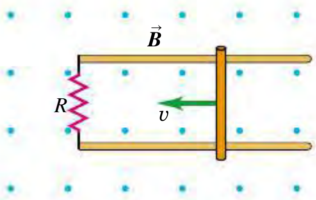
29.30 • A 0.650-m-long metal bar is pulled to the right at a steady 5.0 m/s perpendicular to a uniform, 0.750-T magnetic field. The bar rides on parallel metal rails connected through a 25.0-Ω resistor (**Fig. E29.30**), so the apparatus makes a complete circuit. Ignore the resistance of the bar and the rails. (a) Calculate the magnitude of the emf induced in the circuit. (b) Find the direction of the current induced in the circuit by using (i) the magnetic force on the charges in the moving bar; (ii) Faraday's law; (iii) Lenz's law. (c) Calculate the current through the resistor.

Figure **E29.30**



29.31 • A 0.360-m-long metal bar is pulled to the left by an applied force F. The bar rides on parallel metal rails connected through a 45.0-Ω resistor, as shown in **Fig. E29.31**, so the apparatus makes a complete circuit. You can ignore the resistance of the bar and rails. The circuit is in a uniform 0.650-T magnetic field that is directed out of the plane of the figure. At the instant when the bar is moving to the left at 5.90 m/s, (a) is the induced current in the circuit clockwise or counterclockwise and (b) what is the rate at which the applied force is doing work on the bar?

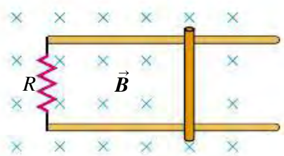
Figure **E29.31**



29.32 • Consider the circuit shown in Fig. E29.31, but with the bar moving to the right with speed v. As in Exercise 29.31, the bar has length 0.360 m, $R = 45.0\ \Omega$, and $B = 0.650$ T. (a) Is the induced current in the circuit clockwise or counterclockwise? (b) At an instant when the 45.0-Ω resistor is dissipating electrical energy at a rate of 0.840 J/s, what is the speed of the bar?

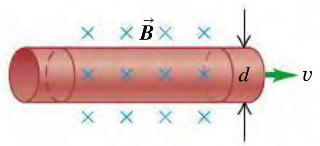
29.33 • A 0.250-m-long bar moves on parallel rails that are connected through a 6.00-Ω resistor, as shown in **Fig. E29.33**, so the apparatus makes a complete circuit. You can ignore the resistance of the bar and rails. The circuit is in a uniform magnetic field $B = 1.20$ T that is directed into the plane of the figure. At an instant when the induced current in the circuit is counterclockwise and equal to 1.75 A, what is the velocity of the bar (magnitude and direction)?

Figure **E29.33**



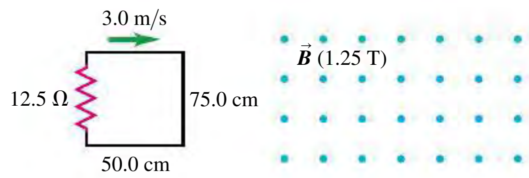
29.34 •• **BIO** **Measuring Blood Flow.** Blood contains positive and negative ions and thus is a conductor. A blood vessel, therefore, can be viewed as an electrical wire. We can even picture the flowing blood as a series of parallel conducting slabs whose thickness is the diameter d of the vessel moving with speed v. (See **Fig. E29.34**.) (a) If the blood vessel is placed in a magnetic field B perpendicular to the vessel, as in the figure, show that the motional potential difference induced across it is $\mathcal{E} = vBd$. (b) If you expect that the blood will be flowing at 15 cm/s for a vessel 5.0 mm in diameter, what strength of magnetic field will you need to produce a potential difference of 1.0 mV? (c) Show that the volume rate of flow (R) of the blood is equal to $R = \pi\mathcal{E}d/4B$. (*Note:* Although the method developed here is useful in measuring the rate of blood flow in a vessel, it is limited to use in surgery because measurement of the potential $\mathcal{E}$ must be made directly across the vessel.)

Figure **E29.34**



29.35 •• A rectangular circuit is moved at a constant velocity of 3.0 m/s into, through, and then out of a uniform 1.25-T magnetic field, as shown in **Fig. E29.35**. The magnetic-field region is considerably wider than 50.0 cm. Find the magnitude and direction (clockwise or counterclockwise) of the current induced in the circuit as it is (a) going into the magnetic field; (b) totally within the magnetic field, but still moving; and (c) moving out of the field. (d) Sketch a graph of the current in this circuit as a function of time, including the preceding three cases.

Figure **E29.35**



Section 29.5 Induced Electric Fields

29.36 • A metal ring 4.50 cm in diameter is placed between the north and south poles of large magnets with the plane of its area perpendicular to the magnetic field. These magnets produce an initial uniform field of 1.12 T between them but are gradually pulled apart, causing this field to remain uniform but decrease steadily at 0.250 T/s. (a) What is the magnitude of the electric field induced in the ring? (b) In which direction (clockwise or counterclockwise) does the current flow as viewed by someone on the south pole of the magnet?

29.37 • The magnetic field within a long, straight solenoid with a circular cross section and radius R is increasing at a rate of dB/dt. (a) What is the rate of change of flux through a circle with radius r_1 inside the solenoid, normal to the axis of the solenoid, and with center on the solenoid axis? (b) Find the magnitude of the induced electric field inside the solenoid, at a distance r_1 from its axis. Show the direction of this field in a diagram. (c) What is the magnitude of the induced electric field *outside* the solenoid, at a distance r_2 from the axis? (d) Graph the magnitude of the induced electric field as a function of the distance r from the axis from $r = 0$ to $r = 2R$. (e) What is the magnitude of the induced emf in a circular turn of radius $R/2$ that has its center on the solenoid axis? (f) What is the magnitude of the induced emf if the radius in part (e) is R? (g) What is the induced emf if the radius in part (e) is $2R$?

29.38 •• A long, thin solenoid has 900 turns per meter and radius 2.50 cm. The current in the solenoid is increasing at a uniform rate of 36.0 A/s. What is the magnitude of the induced electric field at a point near the center of the solenoid and (a) 0.500 cm from the axis of the solenoid; (b) 1.00 cm from the axis of the solenoid?

29.39 •• A long, thin solenoid has 400 turns per meter and radius 1.10 cm. The current in the solenoid is increasing at a uniform rate di/dt. The induced electric field at a point near the center of the solenoid and 3.50 cm from its axis is 8.00×10^{-6} V/m. Calculate di/dt.

29.40 • The magnetic field $\vec{B}$ at all points within the colored circle shown in Fig. E29.15 has an initial magnitude of 0.750 T.

(The circle could represent approximately the space inside a long, thin solenoid.) The magnetic field is directed into the plane of the diagram and is decreasing at the rate of -0.0350 T/s. (a) What is the shape of the field lines of the induced electric field shown in Fig. E29.15, within the colored circle? (b) What are the magnitude and direction of this field at any point on the circular conducting ring with radius 0.100 m? (c) What is the current in the ring if its resistance is 4.00 Ω? (d) What is the emf between points a and b on the ring? (e) If the ring is cut at some point and the ends are separated slightly, what will be the emf between the ends?

29.41 • A long, straight solenoid with a cross-sectional area of 8.00 cm^2 is wound with 90 turns of wire per centimeter, and the windings carry a current of 0.350 A. A second winding of 12 turns encircles the solenoid at its center. The current in the solenoid is turned off such that the magnetic field of the solenoid becomes zero in 0.0400 s. What is the average induced emf in the second winding?

Section 29.7 Displacement Current and Maxwell's Equations

29.42 • A parallel-plate, air-filled capacitor is being charged as in Fig. 29.23. The circular plates have radius 4.00 cm, and at a particular instant the conduction current in the wires is 0.520 A. (a) What is the displacement current density j_D in the air space between the plates? (b) What is the rate at which the electric field between the plates is changing? (c) What is the induced magnetic field between the plates at a distance of 2.00 cm from the axis? (d) At 1.00 cm from the axis?

29.43 • **Displacement Current in a Dielectric.** Suppose that the parallel plates in Fig. 29.23 have an area of 3.00 cm^2 and are separated by a 2.50-mm-thick sheet of dielectric that completely fills the volume between the plates. The dielectric has dielectric constant 4.70. (You can ignore fringing effects.) At a certain instant, the potential difference between the plates is 120 V and the conduction current i_C equals 6.00 mA. At this instant, what are (a) the charge q on each plate; (b) the rate of change of charge on the plates; (c) the displacement current in the dielectric?

29.44 • CALC In Fig. 29.23 the capacitor plates have area 5.00 cm^2 and separation 2.00 mm. The plates are in vacuum. The charging current i_C has a *constant* value of 1.80 mA. At $t = 0$ the charge on the plates is zero. (a) Calculate the charge on the plates, the electric field between the plates, and the potential difference between the plates when $t = 0.500\ \mu$s. (b) Calculate dE/dt, the time rate of change of the electric field between the plates. Does dE/dt vary in time? (c) Calculate the displacement current density j_D between the plates, and from this the total displacement current i_D. How do i_C and i_D compare?

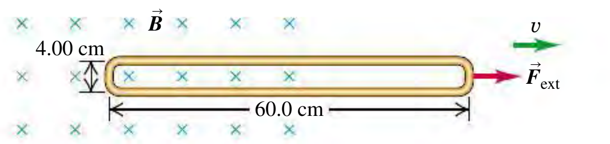
Section 29.8 Superconductivity

29.45 • At temperatures near absolute zero, B_c approaches 0.142 T for vanadium, a type-I superconductor. The normal phase of vanadium has a magnetic susceptibility close to zero. Consider a long, thin vanadium cylinder with its axis parallel to an external magnetic field $\vec{B}_0$ in the $+x$-direction. At points far from the ends of the cylinder, by symmetry, all the magnetic vectors are parallel to the x-axis. At temperatures near absolute zero, what are the resultant magnetic field $\vec{B}$ and the magnetization $\vec{M}$ inside and outside the cylinder (far from the ends) for (a) $\vec{B}_0 = (0.130\text{ T})\hat{\imath}$ and (b) $\vec{B}_0 = (0.260\text{ T})\hat{\imath}$?

PROBLEMS

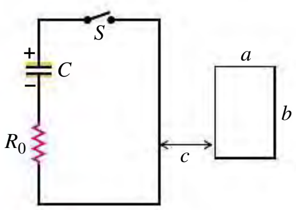
29.46 •• A very long, rectangular loop of wire can slide without friction on a horizontal surface. Initially the loop has part of its area in a region of uniform magnetic field that has magnitude $B = 2.90$ T and is perpendicular to the plane of the loop. The loop has dimensions 4.00 cm by 60.0 cm, mass 24.0 g, and resistance $R = 5.00 \times 10^{-3}\ \Omega$. The loop is initially at rest; then a constant force $F_{ext} = 0.180$ N is applied to the loop to pull it out of the field (**Fig. P29.46**). (a) What is the acceleration of the loop when $v = 3.00$ cm/s? (b) What are the loop's terminal speed and acceleration when the loop is moving at that terminal speed? (c) What is the acceleration of the loop when it is completely out of the magnetic field?

Figure **P29.46**



29.47 •• CP CALC In the circuit shown in **Fig. P29.47**, the capacitor has capacitance $C = 20\ \mu$F and is initially charged to 100 V with the polarity shown. The resistor R_0 has resistance 10 Ω. At time $t = 0$ the switch S is closed. The small circuit is not connected in any way to the large one. The wire of the small circuit has a resistance of 1.0 Ω/m and contains 25 loops. The large circuit is a rectangle 2.0 m by 4.0 m, while the small one has dimensions $a = 10.0$ cm and $b = 20.0$ cm. The distance c is 5.0 cm. (The figure is not drawn to scale.) Both circuits are held stationary. Assume that only the wire nearest the small circuit produces an appreciable magnetic field through it. (a) Find the current in the large circuit 200 μs after S is closed. (b) Find the current in the small circuit 200 μs after S is closed. (*Hint:* See Exercise 29.7.) (c) Find the direction of the current in the small circuit. (d) Justify why we can ignore the magnetic field from all the wires of the large circuit except for the wire closest to the small circuit.

Figure **P29.47**

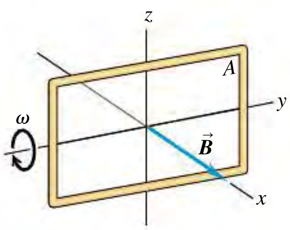


29.48 •• CP CALC In the circuit in Fig. P29.47, an emf of 90.0 V is added in series with the capacitor and the resistor, and the capacitor is initially uncharged. The emf is placed between the capacitor and switch S, with the positive terminal of the emf adjacent to the capacitor. Otherwise, the two circuits are the same as in Problem 29.47. The switch is closed at $t = 0$. When the current in the large circuit is 5.00 A, what are the magnitude and direction of the induced current in the small circuit?

29.49 •• CALC A very long, straight solenoid with a cross-sectional area of 2.00 cm^2 is wound with 90.0 turns of wire per centimeter. Starting at $t = 0$, the current in the solenoid is increasing according to $i(t) = (0.160\text{ A/s}^2)t^2$. A secondary winding of 5 turns encircles the solenoid at its center, such that the secondary winding has the same cross-sectional area as the solenoid. What is the magnitude of the emf induced in the secondary winding at the instant that the current in the solenoid is 3.20 A?

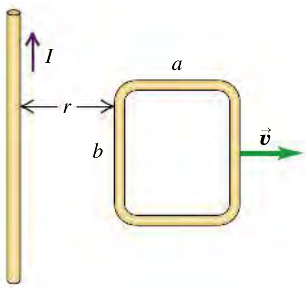
29.50 • Suppose the loop in **Fig. P29.50** is (a) rotated about the y-axis; (b) rotated about the x-axis; (c) rotated about an edge parallel to the z-axis. What is the maximum induced emf in each case if $A = 600\ \text{cm}^2$, $\omega = 35.0\ \text{rad/s}$, and $B = 0.320\ \text{T}$?

Figure **P29.50**



29.51 • In **Fig. P29.51** the loop is being pulled to the right at constant speed v. A constant current I flows in the long wire, in the direction shown. (a) Calculate the magnitude of the net emf $\mathcal{E}$ induced in the loop. Do this two ways: (i) by using Faraday's law of induction (*Hint:* See Exercise 29.7) and (ii) by looking at the emf induced in each segment of the loop due to its motion. (b) Find the direction (clockwise or counterclockwise) of the current induced in the loop. Do this two ways: (i) using Lenz's law and (ii) using the magnetic force on charges in the loop. (c) Check your answer for the emf in part (a) in the following special cases to see whether it is physically reasonable: (i) The loop is stationary; (ii) the loop is very thin, so $a \rightarrow 0$; (iii) the loop gets very far from the wire.

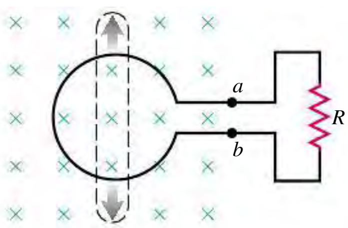
Figure **P29.51**



29.52 • **Make a Generator?** You are shipwrecked on a deserted tropical island. You have some electrical devices that you could operate using a generator but you have no magnets. The earth's magnetic field at your location is horizontal and has magnitude $8.0 \times 10^{-5}\ \text{T}$, and you decide to try to use this field for a generator by rotating a large circular coil of wire at a high rate. You need to produce a peak emf of 9.0 V and estimate that you can rotate the coil at 30 rpm by turning a crank handle. You also decide that to have an acceptable coil resistance, the maximum number of turns the coil can have is 2000. (a) What area must the coil have? (b) If the coil is circular, what is the maximum translational speed of a point on the coil as it rotates? Do you think this device is feasible? Explain.

29.53 • A flexible circular loop 6.50 cm in diameter lies in a magnetic field with magnitude 1.35 T, directed into the plane of the page as shown in **Fig. P29.53**. The loop is pulled at the points indicated by the arrows, forming a loop of zero area in 0.250 s. (a) Find the average induced emf in the circuit. (b) What is the direction of the current in R: from a to b or from b to a? Explain your reasoning.

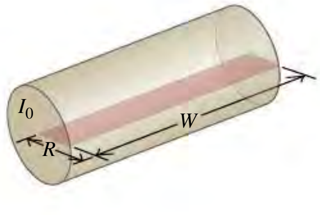
Figure **P29.53**



29.54 ••• **CALC** A conducting rod with length $L = 0.200\ \text{m}$, mass $m = 0.120\ \text{kg}$, and resistance $R = 80.0\ \Omega$ moves without friction on metal rails as shown in Fig. 29.11. A uniform magnetic field with magnitude $B = 1.50\ \text{T}$ is directed into the plane of the figure. The rod is initially at rest, and then a constant force with magnitude $F = 1.90\ \text{N}$ and directed to the right is applied to the rod. How many seconds after the force is applied does the rod reach a speed of 25.0 m/s?

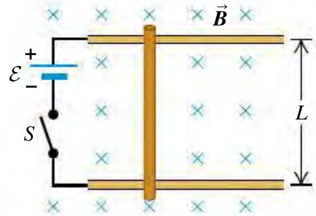
29.55 •• **CALC** A very long, cylindrical wire of radius R carries a current I_0 uniformly distributed across the cross section of the wire. Calculate the magnetic flux through a rectangle that has one side of length W running down the center of the wire and another side of length R, as shown in **Fig. P29.55** (see Exercise 29.7).

Figure **P29.55**



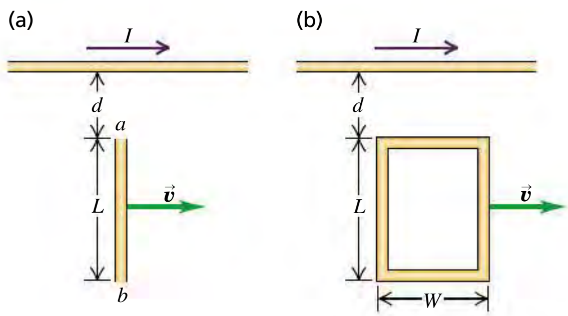
29.56 •• **CP CALC Terminal Speed.** A bar of length $L = 0.36\ \text{m}$ is free to slide without friction on horizontal rails as shown in **Fig. P29.56**. A uniform magnetic field $B = 2.4\ \text{T}$ is directed into the plane of the figure. At one end of the rails there is a battery with emf $\mathcal{E} = 12\ \text{V}$ and a switch S. The bar has mass 0.90 kg and resistance $5.0\ \Omega$; ignore all other resistance in the circuit. The switch is closed at time $t = 0$. (a) Sketch the bar's speed as a function of time. (b) Just after the switch is closed, what is the acceleration of the bar? (c) What is the acceleration of the bar when its speed is 2.0 m/s? (d) What is the bar's terminal speed?

Figure **P29.56**



29.57 • **CALC** The long, straight wire shown in **Fig. P29.57a** carries constant current I. A metal bar with length L is moving at constant velocity $\vec{v}$, as shown in the figure. Point a is a distance d from the wire. (a) Calculate the emf induced in the bar. (b) Which point, a or b, is at higher potential? (c) If the bar is replaced by a rectangular wire loop of resistance R (Fig. P29.57b), what is the magnitude of the current induced in the loop?

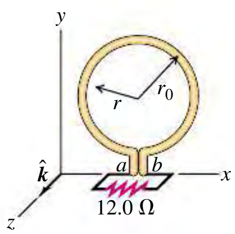
Figure **P29.57**



29.58 • **CALC** A circular conducting ring with radius $r_0 = 0.0420\ \text{m}$ lies in the xy-plane in a region of uniform magnetic field $\vec{B} = B_0[1 - 3(t/t_0)^2 + 2(t/t_0)^3]\hat{k}$. In this expression, $t_0 = 0.0100\ \text{s}$ and is constant, t is time, $\hat{k}$ is the unit vector in the $+z$-direction, and $B_0 = 0.0800\ \text{T}$ and is constant. At points a and b (**Fig. P29.58**) there is a small gap in the ring with wires leading to an external circuit of resistance $R = 12.0\ \Omega$. There is no magnetic field at the location of the external circuit. (a) Derive an expression, as a function of time, for the total magnetic flux Φ_B through the ring. (b) Determine the emf induced in the ring at time

$t = 5.00 \times 10^{-3}$ s. What is the polarity of the emf? (c) Because of the internal resistance of the ring, the current through R at the time given in part (b) is only 3.00 mA. Determine the internal resistance of the ring. (d) Determine the emf in the ring at a time $t = 1.21 \times 10^{-2}$ s. What is the polarity of the emf? (e) Determine the time at which the current through R reverses its direction.

Figure **P29.58**

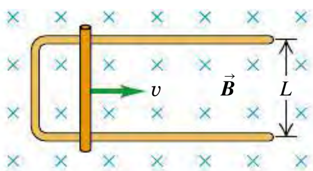


29.59 • **CALC** A slender rod, 0.240 m long, rotates with an angular speed of 8.80 rad/s about an axis through one end and perpendicular to the rod. The plane of rotation of the rod is perpendicular to a uniform magnetic field with a magnitude of 0.650 T. (a) What is the induced emf in the rod? (b) What is the potential difference between its ends? (c) Suppose instead the rod rotates at 8.80 rad/s about an axis through its center and perpendicular to the rod. In this case, what is the potential difference between the ends of the rod? Between the center of the rod and one end?

29.60 •• A 25.0-cm-long metal rod lies in the xy-plane and makes an angle of 36.9° with the positive x-axis and an angle of 53.1° with the positive y-axis. The rod is moving in the $+x$-direction with a speed of 6.80 m/s. The rod is in a uniform magnetic field $\vec{B} = (0.120\ \text{T})\hat{\imath} - (0.220\ \text{T})\hat{\jmath} - (0.0900\ \text{T})\hat{k}$. (a) What is the magnitude of the emf induced in the rod? (b) Indicate in a sketch which end of the rod is at higher potential.

29.61 •• **CP CALC** A rectangular loop with width L and a slide wire with mass m are as shown in **Fig. P29.61**. A uniform magnetic field $\vec{B}$ is directed perpendicular to the plane of the loop into the plane of the figure. The slide wire is given an initial speed of v_0 and then released. There is no friction between the slide wire and the loop, and the resistance of the loop is negligible in comparison to the resistance R of the slide wire. (a) Obtain an expression for F, the magnitude of the force exerted on the wire while it is moving at speed v. (b) Show that the distance x that the wire moves before coming to rest is $x = mv_0R/L^2B^2$.

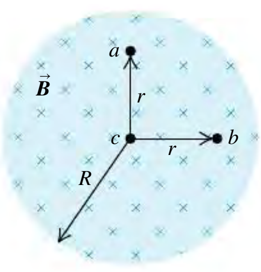
Figure **P29.61**



29.62 • **CALC** An airplane propeller of total length L rotates around its center with angular speed ω in a magnetic field that is perpendicular to the plane of rotation. Modeling the propeller as a thin, uniform bar, find the potential difference between (a) the center and either end of the propeller and (b) the two ends. (c) If the field is the earth's field of 0.50 G and the propeller turns at 220 rpm and is 2.0 m long, what is the potential difference between the middle and either end? It this large enough to be concerned about?

29.63 • The magnetic field $\vec{B}$, at all points within a circular region of radius R, is uniform in space and directed into the plane of the page as shown in **Fig. P29.63**. (The region could be a cross section inside the windings of a long, straight solenoid.) If the magnetic field is increasing at a rate dB/dt, what are the magnitude and direction of the force on a stationary positive point charge q located at points a, b, and c? (Point a is a distance r above the center of the region, point b is a distance r to the right of the center, and point c is at the center of the region.)

Figure **P29.63**

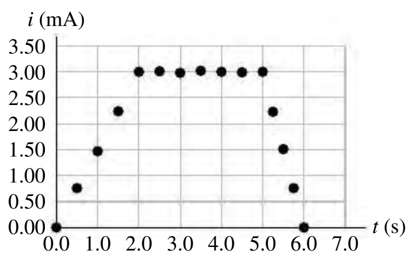


29.64 •• **CP CALC** A capacitor has two parallel plates with area A separated by a distance d. The space between plates is filled with a material having dielectric constant K. The material is not a perfect insulator but has resistivity ρ. The capacitor is initially charged with charge of magnitude Q_0 on each plate that gradually discharges by conduction through the dielectric. (a) Calculate the conduction current density $j_C(t)$ in the dielectric. (b) Show that at any instant the displacement current density in the dielectric is equal in magnitude to the conduction current density but opposite in direction, so the total current density is zero at every instant.

29.65 ••• **CALC** A dielectric of permittivity 3.5×10^{-11} F/m completely fills the volume between two capacitor plates. For $t > 0$ the electric flux through the dielectric is $(8.0 \times 10^3\ \text{V} \cdot \text{m/s}^3)t^3$. The dielectric is ideal and nonmagnetic; the conduction current in the dielectric is zero. At what time does the displacement current in the dielectric equal 21 μA?

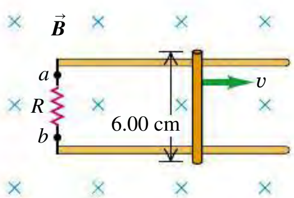
29.66 •• **DATA** You are evaluating the performance of a large electromagnet. The magnetic field of the electromagnet is zero at $t = 0$ and increases as the current through the windings of the electromagnet is increased. You determine the magnetic field as a function of time by measuring the time dependence of the current induced in a small coil that you insert between the poles of the electromagnet, with the plane of the coil parallel to the pole faces as in Fig. 29.5. The coil has 4 turns, a radius of 0.800 cm, and a resistance of 0.250 Ω. You measure the current i in the coil as a function of time t. Your results are shown in **Fig. P29.66**. Throughout your measurements, the current induced in the coil remains in the same direction. Calculate the magnetic field at the location of the coil for (a) $t = 2.00$ s, (b) $t = 5.00$ s, and (c) $t = 6.00$ s.

Figure **P29.66**



29.67 •• **DATA** You are conducting an experiment in which a metal bar of length 6.00 cm and mass 0.200 kg slides without friction on two parallel metal rails (**Fig. P29.67**). A resistor with resistance $R = 0.800\ \Omega$ is connected across one end of the rails so that the bar, rails, and resistor form a complete conducting path. The resistances of the rails and of the bar are much less than R and can be ignored. The entire apparatus is in a uniform magnetic field $\vec{B}$ that is directed into the plane of the figure. You give the bar an initial velocity $v = 20.0$ cm/s to the right and then release it, so that the only force on the bar then is the force exerted by the magnetic field. Using high-speed photography, you measure the magnitude of the acceleration of the bar as a function of its speed. Your results are given in the table:

Figure **P29.67**

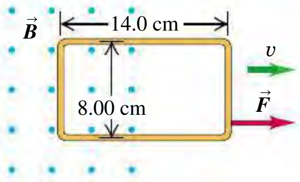


v (cm/s)	20.0	16.0	14.0	12.0	10.0	8.0
a (cm/s^2)	6.2	4.9	4.3	3.7	3.1	2.5

(a) Plot the data as a graph of a versus v. Explain why the data points plotted this way lie close to a straight line, and determine the slope of the best-fit straight line for the data. (b) Use your graph from part (a) to calculate the magnitude B of the magnetic field. (c) While the bar is moving, which end of the resistor, a or b, is at higher potential? (d) How many seconds does it take the speed of the bar to decrease from 20.0 cm/s to 10.0 cm/s?

29.68 ••• DATA You measure the magnitude of the external force $\vec{F}$ that must be applied to a rectangular conducting loop to pull it at constant speed v out of a region of uniform magnetic field $\vec{B}$ that is directed out of the plane of **Fig. P29.68**. The loop has dimensions 14.0 cm by 8.00 cm and resistance $4.00 \times 10^{-3}\ \Omega$; it does not change shape as it moves. The measurements you collect are listed in the table.

Figure **P29.68**



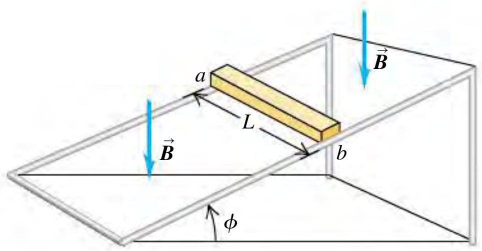
F (N)	0.10	0.21	0.31	0.41	0.52
v (cm/s)	2.0	4.0	6.0	8.0	10.0

(a) Plot the data as a graph of F versus v. Explain why the data points plotted this way lie close to a straight line, and determine the slope of the best-fit straight line for the data. (b) Use your graph from part (a) to calculate the magnitude B of the uniform magnetic field. (c) In Fig. P29.68, is the current induced in the loop clockwise or counterclockwise? (d) At what rate is electrical energy being dissipated in the loop when the speed of the loop is 5.00 cm/s?

CHALLENGE PROBLEMS

29.69 ••• A metal bar with length L, mass m, and resistance R is placed on frictionless metal rails that are inclined at an angle ϕ above the horizontal. The rails have negligible resistance. A uniform magnetic field of magnitude B is directed downward as shown in **Fig. P29.69**. The bar is released from rest and slides down the rails. (a) Is the direction of the current induced in the bar from a to b or from b to a? (b) What is the terminal speed of the bar? (c) What is the induced current in the bar when the terminal speed has been reached? (d) After the terminal speed has been reached, at what rate is electrical energy being converted to thermal energy in the resistance of the bar? (e) After the terminal speed has been reached, at what rate is work being done on the bar by gravity? Compare your answer to that in part (d).

Figure **P29.69**

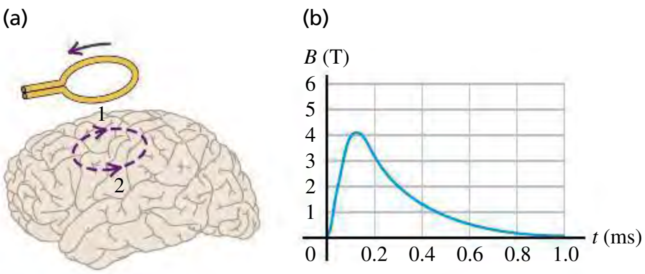


29.70 ••• CP CALC A square, conducting, wire loop of side L, total mass m, and total resistance R initially lies in the horizontal xy-plane, with corners at $(x, y, z) = (0, 0, 0)$, $(0, L, 0)$, $(L, 0, 0)$, and $(L, L, 0)$. There is a uniform, upward magnetic field $\vec{B} = B\hat{k}$ in the space within and around the loop. The side of the loop that extends from $(0, 0, 0)$ to $(L, 0, 0)$ is held in place on the x-axis; the rest of the loop is free to pivot around this axis. When the loop is released, it begins to rotate due to the gravitational torque. (a) Find the *net* torque (magnitude and direction) that acts on the loop when it has rotated through an angle ϕ from its original orientation and is rotating downward at an angular speed ω. (b) Find the angular acceleration of the loop at the instant described in part (a). (c) Compared to the case with zero magnetic field, does it take the loop a longer or shorter time to rotate through 90°? Explain. (d) Is mechanical energy conserved as the loop rotates downward? Explain.

PASSAGE PROBLEMS

BIO **STIMULATING THE BRAIN.** Communication in the nervous system is based on the propagation of electrical signals called *action potentials* along axons, which are extensions of nerve cells (see the Passage Problems in Chapter 26). Action potentials are generated when the electric potential difference across the membrane of the nerve cell changes: Specifically, the inside of the cell becomes more positive. Researchers in clinical medicine and neurobiology cannot stimulate nerves (even noninvasively) at specific locations in conscious human subjects. Using electrodes to apply current to the skin is painful and requires large currents, which could be dangerous.

Anthony Barker and colleagues at the University of Sheffield in England developed a technique called *transcranial magnetic stimulation* (TMS). In this widely used procedure, a coil positioned near the skull produces a time-varying magnetic field that induces in the conductive tissue of the brain (see part (a) of the figure) electric currents that are sufficient to cause action potentials in nerve cells. For example, if the coil is placed near the motor cortex (the region of the brain that controls voluntary movement), scientists can monitor muscle contraction and assess the connections between the brain and the muscles. Part (b) of the figure is a graph of the typical dependence on time t of the magnetic field B produced by the coil.

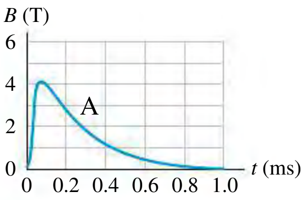


29.71 In part (a) of the figure, a current pulse increases to a peak and then decreases to zero in the direction shown in the stimulating coil. What will be the direction of the induced current (dashed line) in the brain tissue? (a) 1; (b) 2; (c) 1 while the current increases in the stimulating coil, 2 while the current decreases; (d) 2 while the current increases in the stimulating coil, 1 while the current decreases.

29.72 Consider the brain tissue at the level of the dashed line to be a series of concentric circles, each behaving independently of the others. Where will the induced emf be the greatest? (a) At the center of the dashed line; (b) at the periphery of the dashed line; (c) nowhere—it will be the same in all concentric circles; (d) at the center while the stimulating current increases, at the periphery while the current decreases.

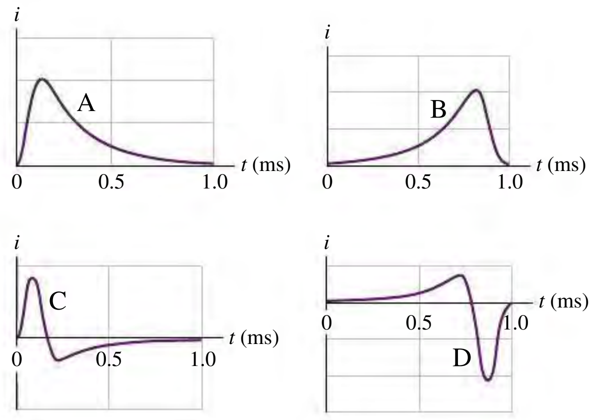
29.73 It may be desirable to increase the maximum induced current in the brain tissue. In **Fig. P29.73**, which time-dependent graph of the magnetic field B in the coil achieves that goal? Assume that everything else remains constant. (a) A; (b) B; (c) either A or B; (d) neither A nor B.

Figure **P29.73**



29.74 Which graph in **Fig. P29.74** best represents the time t dependence of the current i induced in the brain tissue, assuming that this tissue can be modeled as a resistive circuit? (The units of i are arbitrary.) (a) A; (b) B; (c) C; (d) D.

Figure **P29.74**



Answers

Chapter Opening Question ?

iv As the magnetic stripe moves through the card reader, the coded pattern of magnetization in the stripe causes a varying magnetic flux. An electric field is induced, which causes a current in the reader's circuits. If the card does not move, there is no induced current and none of the credit card's information is read.

Test Your Understanding Questions

29.2 **(a) (i), (b) (iii)** In (a), initially there is magnetic flux into the plane of the page, which we call positive. While the loop is being squeezed, the flux is becoming less positive ($d\Phi_B/dt < 0$) and so the induced emf is positive as in Fig. 29.6b ($\mathcal{E} = -d\Phi_B/dt > 0$). If you point the thumb of your right hand into the page, your fingers curl clockwise, so this is the direction of positive induced emf. In (b), since the coil's shape is no longer changing, the magnetic flux is not changing and there is no induced emf.

29.3 **(a) (i), (b) (iii)** In (a), as in the original situation, the magnet and loop are approaching each other and the downward flux through the loop is increasing. Hence the induced emf and induced current are the same. In (b), since the magnet and loop are moving together, the flux through the loop is not changing and no emf is induced.

29.4 **(a) (iii); (b) (i) or (ii); (c) (ii) or (iii)** You will get the maximum motional emf if you hold the rod vertically, so that its length is perpendicular to both the magnetic field and the direction of motion. With this orientation, $\vec{L}$ is parallel to $\vec{v} \times \vec{B}$. If you hold the rod in any horizontal orientation, $\vec{L}$ will be perpendicular to $\vec{v} \times \vec{B}$ and no emf will be induced. If you walk due north or south, $\vec{v} \times \vec{B} = \mathbf{0}$ and no emf will be induced for any orientation of the rod.

29.5 **yes, no** The magnetic field at a fixed position changes as you move the magnet, which induces an electric field. Such induced electric fields are *not* conservative.

29.6 **(iii)** By Lenz's law, the force must oppose the motion of the disk through the magnetic field. Since the disk material is now moving to the right through the field region, the force $\vec{F}$ is to the left—that is, in the opposite direction to that shown in Fig. 29.20b. To produce a leftward magnetic force $\vec{F} = I\vec{L} \times \vec{B}$ on currents moving through a magnetic field $\vec{B}$ directed out of the plane of the figure, the eddy currents must be moving downward in the figure—that is, in the same direction shown in Fig. 29.20b.

29.7 **(a) Faraday's law, (b) Ampere's law** A credit card reader works by inducing currents in the reader's coils as the card's magnetized stripe is swiped (see the answer to the chapter opening question). Ampere's law describes how currents of all kinds (both conduction currents and displacement currents) give rise to magnetic fields.

Bridging Problem

$v_{\text{terminal}} = 16\rho_m\rho_R g/B^2$

? Many traffic lights change when a car rolls up to the intersection. This process works because the car contains
(i) conducting material;
(ii) insulating material that carries a net electric charge;
(iii) ferromagnetic material;
(iv) ferromagnetic material that is already magnetized.

30 INDUCTANCE

LEARNING GOALS

Looking forward at ...

30.1 How a time-varying current in one coil can induce an emf in a second, unconnected coil.

30.2 How to relate the induced emf in a circuit to the rate of change of current in the same circuit.

30.3 How to calculate the energy stored in a magnetic field.

30.4 How to analyze circuits that include both a resistor and an inductor (coil).

30.5 Why electrical oscillations occur in circuits that include both an inductor and a capacitor.

30.6 Why oscillations decay in circuits with an inductor, a resistor, and a capacitor.

Looking back at ...

14.2, 14.3, 14.7 Simple harmonic motion, damped oscillations.

24.1, 24.3 Capacitance, electric-field energy.

26.2, 26.4 Kirchhoff's rules, *R-C* circuits.

28.4, 28.7, 28.8 Magnetic forces between conductors; field of a solenoid; permeability.

29.2, 29.3, 29.7 Faraday's law; Lenz's law; conservative and nonconservative electric fields.

Take a length of copper wire and wrap it around a pencil to form a coil. If you put this coil in a circuit, the coil behaves quite differently than a straight piece of wire. In an ordinary gasoline-powered car, a coil of this kind makes it possible for the 12-volt car battery to provide thousands of volts to the spark plugs in order for the plugs to fire and make the engine run. Other coils are used to keep fluorescent light fixtures shining. Larger coils placed under city streets are used to control the operation of traffic signals. All of these applications, and many others, involve the *induction* effects that we studied in Chapter 29.

A changing current in a coil induces an emf in an adjacent coil. The coupling between the coils is described by their *mutual inductance*. A changing current in a coil also induces an emf in that same coil. Such a coil is called an *inductor,* and the relationship of current to emf is described by the *inductance* (also called *self-inductance*) of the coil. If a coil is initially carrying a current, energy is released when the current decreases; this principle is used in automotive ignition systems. We'll find that this released energy was stored in the magnetic field caused by the current that was initially in the coil, and we'll look at some of the practical applications of magnetic-field energy.

We'll also take a first look at what happens when an inductor is part of a circuit. In Chapter 31 we'll go on to study how inductors behave in alternating-current circuits, and we'll learn why inductors play an essential role in modern electronics.

30.1 MUTUAL INDUCTANCE

In Section 28.4 we considered the magnetic interaction between two wires carrying *steady* currents; the current in one wire causes a magnetic field, which exerts a force on the current in the second wire. But an additional interaction arises between two circuits when there is a *changing* current in one of the circuits.

Consider two neighboring coils of wire, as in **Fig. 30.1**. A current flowing in coil 1 produces a magnetic field $\vec{B}$ and hence a magnetic flux through coil 2. If the current in coil 1 changes, the flux through coil 2 changes as well; according to Faraday's law (Section 29.2), this induces an emf in coil 2. In this way, a change in the current in one circuit can induce a current in a second circuit.

Let's analyze the situation shown in Fig. 30.1 in more detail. We'll use lowercase letters to represent quantities that vary with time; for example, a time-varying current is i, often with a subscript to identify the circuit. In Fig. 30.1 a current i_1 in coil 1 sets up a magnetic field $\vec{B}$, and some of the (blue) field lines pass through coil 2. We denote the magnetic flux through *each* turn of coil 2, caused by the current i_1 in coil 1, as Φ_{B2}. (If the flux is different through different turns of the coil, then Φ_{B2} denotes the *average* flux.) The magnetic field is proportional to i_1, so Φ_{B2} is also proportional to i_1. When i_1 changes, Φ_{B2} changes; this changing flux induces an emf $\mathcal{E}_2$ in coil 2, given by

$$\mathcal{E}_2 = -N_2 \frac{d\Phi_{B2}}{dt} \tag{30.1}$$

We could represent the proportionality of Φ_{B2} and i_1 in the form $\Phi_{B2} = (\text{constant})i_1$, but instead it is more convenient to include the number of turns N_2 in the relationship. Introducing a proportionality constant M_{21}, called the **mutual inductance** of the two coils, we write

$$N_2\Phi_{B2} = M_{21}i_1 \tag{30.2}$$

where Φ_{B2} is the flux through a *single* turn of coil 2. From this,

$$N_2 \frac{d\Phi_{B2}}{dt} = M_{21} \frac{di_1}{dt}$$

and we can rewrite Eq. (30.1) as

$$\mathcal{E}_2 = -M_{21} \frac{di_1}{dt} \tag{30.3}$$

That is, a change in the current i_1 in coil 1 induces an emf in coil 2 that is directly proportional to the rate of change of i_1 (**Fig. 30.2**).

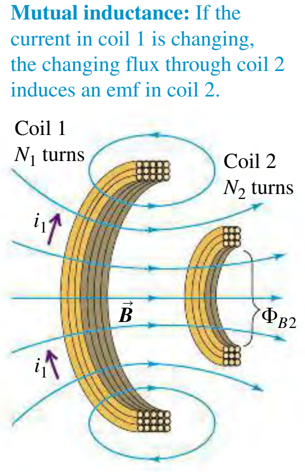
We may also write the definition of mutual inductance, Eq. (30.2), as

$$M_{21} = \frac{N_2\Phi_{B2}}{i_1}$$

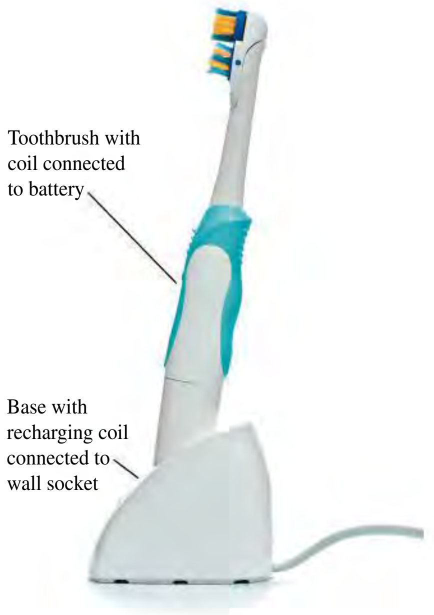
If the coils are in vacuum, the flux Φ_{B2} through each turn of coil 2 is directly proportional to the current i_1. Then the mutual inductance M_{21} is a constant that depends only on the geometry of the two coils (the size, shape, number of turns, and orientation of each coil and the separation between the coils). If a magnetic material is present, M_{21} also depends on the magnetic properties of the material. If the material has nonlinear magnetic properties—that is, if the relative permeability K_m (defined in Section 28.8) is not constant and magnetization is not proportional to magnetic field—then Φ_{B2} is no longer directly proportional to i_1. In that case the mutual inductance also depends on the value of i_1. In this discussion we will assume that any magnetic material present has constant K_m so that flux *is* directly proportional to current and M_{21} depends on geometry only.

We can repeat our discussion for the opposite case in which a changing current i_2 in coil 2 causes a changing flux Φ_{B1} and an emf $\mathcal{E}_1$ in coil 1. It turns out that the corresponding constant M_{12} is *always* equal to M_{21}, even though in general the two coils are not identical and the flux through them is not the same.

30.1 A current i_1 in coil 1 gives rise to a magnetic flux through coil 2.



30.2 This electric toothbrush makes use of mutual inductance. The base contains a coil that is supplied with alternating current from a wall socket. Even though there is no direct electrical contact between the base and the toothbrush, this varying current induces an emf in a coil within the toothbrush itself, recharging the toothbrush battery.



We call this common value simply the mutual inductance, denoted by the symbol M without subscripts; it characterizes completely the induced-emf interaction of two coils. Then

Induced emf in coil 2 — Rate of change of current in coil 1 — Induced emf in coil 1 — Rate of change of current in coil 2

Mutually induced emfs:

$$\mathcal{E}_2 = -M\frac{di_1}{dt} \quad \text{and} \quad \mathcal{E}_1 = -M\frac{di_2}{dt} \tag{30.4}$$

Mutual inductance of coils 1 and 2

The negative signs in Eqs. (30.4) reflect Lenz's law (Section 29.3). The first equation says that a current change in coil 1 causes a flux change through coil 2, inducing an emf in coil 2 that opposes the flux change; in the second equation, the roles of the two coils are interchanged. The mutual inductance M is

Turns in coil 2 — Magnetic flux through each turn of coil 2 — Turns in coil 1 — Magnetic flux through each turn of coil 1

Mutual inductance of coils 1 and 2

$$M = \frac{N_2\Phi_{B2}}{i_1} = \frac{N_1\Phi_{B1}}{i_2} \tag{30.5}$$

Current in coil 1 (causes flux through coil 2) — Current in coil 2 (causes flux through coil 1)

CAUTION Only a time-varying current induces an emf Only a *time-varying* current in a coil can induce an emf and hence a current in a second coil. Equations (30.4) show that the induced emf in each coil is directly proportional to the *rate of change* of the current in the other coil, not to the *value* of the current. A steady current in one coil, no matter how strong, cannot induce a current in a neighboring coil.

The SI unit of mutual inductance is called the **henry** (1 H), in honor of the American physicist Joseph Henry (1797–1878), one of the discoverers of electromagnetic induction. From Eq. (30.5), one henry is equal to one weber per ampere. Other equivalent units, obtained by using Eqs. (30.4), are

$$1\ \text{H} = 1\ \text{Wb/A} = 1\ \text{V}\cdot\text{s/A} = 1\ \Omega\cdot\text{s} = 1\ \text{J/A}^2$$

Just as the farad is a rather large unit of capacitance (see Section 24.1), the henry is a rather large unit of mutual inductance. Typical values of mutual inductance can be in the millihenry (mH) or microhenry (μH) range.

Drawbacks and Uses of Mutual Inductance

Mutual inductance can be a nuisance in electric circuits, since variations in current in one circuit can induce unwanted emfs in other nearby circuits. To minimize these effects, multiple-circuit systems must be designed so that M is as small as possible; for example, two coils would be placed far apart.

Happily, mutual inductance also has many useful applications. A *transformer,* used in alternating-current circuits to raise or lower voltages, is fundamentally no different from the two coils shown in Fig. 30.1. A time-varying alternating current in one coil of the transformer produces an alternating emf in the other coil; the value of M, which depends on the geometry of the coils, determines the amplitude of the induced emf in the second coil and hence the amplitude of the output voltage. (We'll describe transformers in more detail in Chapter 31.)

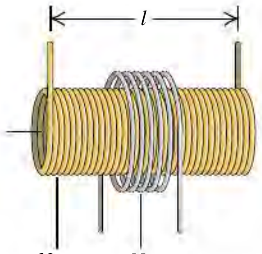
EXAMPLE 30.1 CALCULATING MUTUAL INDUCTANCE

In one form of Tesla coil (a high-voltage generator popular in science museums), a long solenoid with length l and cross-sectional area A is closely wound with N_1 turns of wire. A coil with N_2 turns surrounds it at its center (**Fig. 30.3**). Find the mutual inductance M.

SOLUTION

IDENTIFY and SET UP: Mutual inductance occurs here because a current in either coil sets up a magnetic field that causes a flux through the other coil. From Example 28.9 (Section 28.7) we have an expression [Eq. (28.23)] for the field magnitude B_1 at the center of the solenoid (coil 1) in terms of the solenoid current i_1. This allows us to determine the flux through a cross section of the solenoid. Since there is almost no magnetic field outside a very long solenoid, this is also equal to the flux Φ_{B2} through each turn of the *outer* coil (2). We then use Eq. (30.5), in the form $M = N_2\Phi_{B2}/i_1$, to determine M.

30.3 A long solenoid with cross-sectional area A and N_1 turns is surrounded at its center by a coil with N_2 turns.



EXECUTE: Equation (28.23) is expressed in terms of the number of turns per unit length, which for solenoid (1) is $n_1 = N_1/L$. So

$$B_1 = \mu_0 n_1 i_1 = \frac{\mu_0 N_1 i_1}{l}$$

The flux through a cross section of the solenoid equals B_1A. As we mentioned above, this also equals the flux Φ_{B2} through each turn of the outer coil, independent of its cross-sectional area. From Eq. (30.5), the mutual inductance M is then

$$M = \frac{N_2\Phi_{B2}}{i_1} = \frac{N_2B_1A}{i_1} = \frac{N_2}{i_1}\frac{\mu_0N_1i_1}{l}A = \frac{\mu_0AN_1N_2}{l}$$

EVALUATE: The mutual inductance M of any two coils is proportional to the product N_1N_2 of their numbers of turns. Notice that M depends only on the geometry of the two coils, not on the current.

Here's a numerical example to give you an idea of magnitudes. Suppose $l = 0.50$ m, $A = 10\text{ cm}^2 = 1.0 \times 10^{-3}\text{ m}^2$, $N_1 = 1000$ turns, and $N_2 = 10$ turns. Then

$$M = \frac{(4\pi \times 10^{-7}\text{ Wb/A}\cdot\text{m})(1.0 \times 10^{-3}\text{ m}^2)(1000)(10)}{0.50\text{ m}}$$
$$= 25 \times 10^{-6}\text{ Wb/A} = 25 \times 10^{-6}\text{ H} = 25\ \mu\text{H}$$

EXAMPLE 30.2 EMF DUE TO MUTUAL INDUCTANCE

In Example 30.1, suppose the current i_2 in the outer coil is given by $i_2 = (2.0 \times 10^6\text{ A/s})t$. (Currents in wires can indeed increase this rapidly for brief periods.) (a) At $t = 3.0\ \mu$s, what is the average magnetic flux through each turn of the solenoid (coil 1) due to the current in the outer coil? (b) What is the induced emf in the solenoid?

SOLUTION

IDENTIFY and SET UP: In Example 30.1 we found the mutual inductance by relating the current in the solenoid to the flux produced in the outer coil; to do that, we used Eq. (30.5) in the form $M = N_2\Phi_{B2}/i_1$. Here we are given the current i_2 in the outer coil and want to find the resulting flux Φ_1 in the solenoid. The mutual inductance is the *same* in either case, and we have $M = 25\ \mu$H from Example 30.1. We use Eq. (30.5) in the form $M = N_1\Phi_{B1}/i_2$ to determine the average flux Φ_{B1} through each turn of the solenoid caused by current i_2 in the outer coil. We then use Eqs. (30.4) to find the emf induced in the solenoid by the time variation of i_2.

EXECUTE: (a) At $t = 3.0\ \mu\text{s} = 3.0 \times 10^{-6}$ s, the current in the outer coil is $i_2 = (2.0 \times 10^6\text{ A/s})(3.0 \times 10^{-6}\text{ s}) = 6.0$ A. We solve Eq. (30.5) for the flux Φ_{B1} through each turn of coil 1:

$$\Phi_{B1} = \frac{Mi_2}{N_1} = \frac{(25 \times 10^{-6}\text{ H})(6.0\text{ A})}{1000} = 1.5 \times 10^{-7}\text{ Wb}$$

We emphasize that this is an *average* value; the flux can vary considerably between the center and the ends of the solenoid.

(b) We are given $i_2 = (2.0 \times 10^6\text{ A/s})t$, so $di_2/dt = 2.0 \times 10^6$ A/s; then, from Eqs. (30.4), the induced emf in the solenoid is

$$\mathcal{E}_1 = -M\frac{di_2}{dt} = -(25 \times 10^{-6}\text{ H})(2.0 \times 10^6\text{ A/s}) = -50\text{ V}$$

EVALUATE: This is a substantial induced emf in response to a very rapid current change. In an operating Tesla coil, there is a high-frequency alternating current rather than a continuously increasing current as in this example; both di_2/dt and $\mathcal{E}_1$ alternate as well, with amplitudes that can be thousands of times larger than in this example.

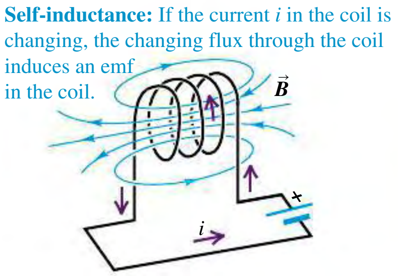
TEST YOUR UNDERSTANDING OF SECTION 30.1 Consider the Tesla coil described in Example 30.1. If you make the solenoid out of twice as much wire, so that it has twice as many turns and is twice as long, how much larger is the mutual inductance? (i) M is four times greater; (ii) M is twice as great; (iii) M is unchanged; (iv) M is $\frac{1}{2}$ as great; (v) M is $\frac{1}{4}$ as great.

30.2 SELF-INDUCTANCE AND INDUCTORS

In our discussion of mutual inductance we considered two separate, independent circuits: A current in one circuit creates a magnetic field that gives rise to a flux through the second circuit. If the current in the first circuit changes, the flux through the second circuit changes and an emf is induced in the second circuit.

An important related effect occurs in a *single* isolated circuit. A current in a circuit sets up a magnetic field that causes a magnetic flux through the *same* circuit; this flux changes when the current changes. Thus any circuit that carries a varying current has an emf induced in it by the variation in *its own* magnetic field. Such an emf is called a **self-induced emf.** By Lenz's law, a self-induced emf opposes the change in the current that caused the emf and so tends to make it more difficult for variations in current to occur. Hence self-induced emfs can be of great importance whenever there is a varying current.

Self-induced emfs can occur in *any* circuit, since there is always some magnetic flux through the closed loop of a current-carrying circuit. But the effect is greatly enhanced if the circuit includes a coil with N turns of wire (**Fig. 30.4**). As a result of the current i, there is an average magnetic flux Φ_B through each turn of the coil. In analogy to Eq. (30.5) we define the **self-inductance** L of the circuit as



30.4 The current i in the circuit causes a magnetic field $\vec{B}$ in the coil and hence a flux through the coil.

Self-inductance (or **inductance**) of a coil

Number of turns in coil

Flux due to current through each turn of coil

$$L = \frac{N\Phi_B}{i} \tag{30.6}$$

Current in coil

When there is no danger of confusion with mutual inductance, the self-inductance is called simply the **inductance.** Comparing Eqs. (30.5) and (30.6), we see that the units of self-inductance are the same as those of mutual inductance; the SI unit of self-inductance is the henry.

If the current i changes, so does the flux Φ_B; from rearranging Eq. (30.6) and differentiating with respect to time, the rates of change are related by

$$N\frac{d\Phi_B}{dt} = L\frac{di}{dt}$$

From Faraday's law for a coil with N turns, Eq. (29.4), the self-induced emf is $\mathcal{E} = -N\,d\Phi_B/dt$, so it follows that

Self-induced emf in a circuit

Inductance of circuit

$$\mathcal{E} = -L\frac{di}{dt} \tag{30.7}$$

Rate of change of current in circuit

The minus sign in Eq. (30.7) is a reflection of Lenz's law; it says that the self-induced emf in a circuit opposes any change in the current in that circuit.

Equation (30.7) states that the self-inductance of a circuit is the magnitude of the self-induced emf per unit rate of change of current. This relationship makes it possible to measure an unknown self-inductance: Change the current at a known rate di/dt, measure the induced emf, and take the ratio to determine L.

Application Inductors, Power Transmission, and Lightning Strikes If lightning strikes part of an electrical power transmission system, it causes a sudden spike in voltage that can damage the components of the system as well as anything connected to that system (for example, home appliances). To minimize these effects, large inductors are incorporated into the transmission system. These use the principle that an inductor opposes and suppresses any rapid changes in the current.

Inductors As Circuit Elements

A circuit device that is designed to have a particular inductance is called an **inductor,** or a *choke*. The usual circuit symbol for an inductor is

Like resistors and capacitors, inductors are among the indispensable circuit elements of modern electronics. Their purpose is to oppose any variations in the current through the circuit. An inductor in a direct-current circuit helps to maintain a steady current despite any fluctuations in the applied emf; in an alternating-current circuit, an inductor tends to suppress variations of the current that are more rapid than desired.

To understand the behavior of circuits containing inductors, we need to develop a general principle analogous to Kirchhoff's loop rule (discussed in Section 26.2). To apply that rule, we go around a conducting loop, measuring potential differences across successive circuit elements as we go. The algebraic sum of these differences around any closed loop must be zero because the electric field produced by charges distributed around the circuit is *conservative*. In Section 29.7 we denoted such a conservative field as $\vec{E}_c$.

When an inductor is included in the circuit, the situation changes. The magnetically induced electric field within the coils of the inductor is *not* conservative; as in Section 29.7, we'll denote it by $\vec{E}_n$. We need to think very carefully about the roles of the various fields. Let's assume we are dealing with an inductor whose coils have negligible resistance. Then a negligibly small electric field is required to make charge move through the coils, so the *total* electric field $\vec{E}_c + \vec{E}_n$ within the coils must be zero, even though neither field is individually zero. Because $\vec{E}_c$ is nonzero, there have to be accumulations of charge on the terminals of the inductor and the surfaces of its conductors to produce this field.

Consider the circuit shown in **Fig. 30.5**; the box contains some combination of batteries and variable resistors that enables us to control the current i in the circuit. According to Faraday's law, Eq. (29.10), the line integral of $\vec{E}_n$ around the circuit is the negative of the rate of change of flux through the circuit, which in turn is given by Eq. (30.7). Combining these two relationships, we get

$$\oint \vec{E}_n \cdot d\vec{l} = -L\frac{di}{dt}$$

where we integrate clockwise around the loop (the direction of the assumed current). But $\vec{E}_n$ is different from zero only within the inductor. Therefore the integral of $\vec{E}_n$ around the whole loop can be replaced by its integral only from a to b through the inductor; that is,

$$\int_a^b \vec{E}_n \cdot d\vec{l} = -L\frac{di}{dt}$$

Next, because $\vec{E}_c + \vec{E}_n = \mathbf{0}$ at each point within the inductor coils, $\vec{E}_n = -\vec{E}_c$. So we can rewrite the above equation as

$$\int_a^b \vec{E}_c \cdot d\vec{l} = L\frac{di}{dt}$$

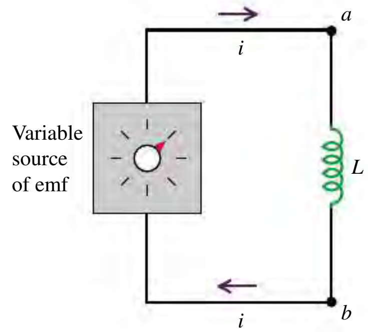
But this integral is just the potential V_{ab} of point a with respect to point b, so

$$V_{ab} = V_a - V_b = L\frac{di}{dt} \qquad (30.8)$$

We conclude that there is a genuine potential difference between the terminals of the inductor, associated with conservative, electrostatic forces, even though the electric field associated with magnetic induction is nonconservative. Thus we are justified in using Kirchhoff's loop rule to analyze circuits that include inductors. Equation (30.8) gives the potential difference across an inductor in a circuit.

CAUTION **Self-induced emf opposes changes in current** Note that the self-induced emf does not oppose the current i itself; rather, it opposes any *change* (di/dt) in the current. Thus the circuit behavior of an inductor is quite different from that of a resistor. **Figure 30.6** compares the behaviors of a resistor and an inductor and summarizes the sign relationships.

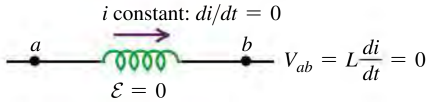
30.5 A circuit containing a source of emf and an inductor. The source is variable, so the current i and its rate of change di/dt can be varied.



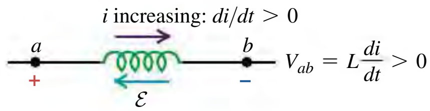
30.6 (a) The potential difference across a resistor depends on the current, whereas (b), (c), (d) the potential difference across an inductor depends on the rate of change of the current.

(a) Resistor with current i flowing from a to b: potential drops from a to b.

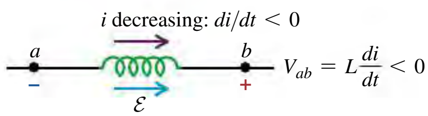
(b) Inductor with *constant* current i flowing from a to b: no potential difference.



(c) Inductor with *increasing* current i flowing from a to b: potential drops from a to b.



(d) Inductor with *decreasing* current i flowing from a to b: potential increases from a to b.



Applications of Inductors

30.7 These fluorescent light tubes are wired in series with an inductor, or ballast, that helps to sustain the current flowing through the tubes.

Because an inductor opposes changes in current, it plays an important role in fluorescent light fixtures (**Fig. 30.7**). In such fixtures, current flows from the wiring into the gas that fills the tube, ionizing the gas and causing it to glow. However, an ionized gas or *plasma* is a highly nonohmic conductor: The greater the current, the more highly ionized the plasma becomes and the lower its resistance. If a sufficiently large voltage is applied to the plasma, the current can grow so much that it damages the circuitry outside the fluorescent tube. To prevent this problem, an inductor or *magnetic ballast* is put in series with the fluorescent tube to keep the current from growing out of bounds.

The ballast also makes it possible for the fluorescent tube to work with the alternating voltage provided by household wiring. This voltage oscillates sinusoidally with a frequency of 60 Hz, so that it goes momentarily to zero 120 times per second. If there were no ballast, the plasma in the fluorescent tube would rapidly deionize when the voltage went to zero and the tube would shut off. With a ballast present, a self-induced emf sustains the current and keeps the tube lit. Magnetic ballasts are also used for this purpose in streetlights (which obtain their light from a glowing mercury or sodium vapor) and in neon lights. (In compact fluorescent lamps, the magnetic ballast is replaced by a more complicated scheme that utilizes transistors, discussed in Chapter 42.)

The self-inductance of a circuit depends on its size, shape, and number of turns. For N turns close together, it is always proportional to N^2. It also depends on the magnetic properties of the material enclosed by the circuit. In the following examples we will assume that the circuit encloses only vacuum (or air, which from the standpoint of magnetism is essentially vacuum). If, however, the flux is concentrated in a region containing a magnetic material with permeability μ, then in the expression for B we must replace μ_0 (the permeability of vacuum) by $\mu = K_m\mu_0$, as discussed in Section 28.8. If the material is diamagnetic or paramagnetic, this replacement makes very little difference, since K_m is very close to 1. If the material is *ferromagnetic,* however, the difference is of crucial importance. A solenoid wound on a soft iron core having $K_m = 5000$ can have an inductance approximately 5000 times as great as that of the same solenoid with an air core. Ferromagnetic-core inductors are very widely used in a variety of electronic and electric-power applications.

With ferromagnetic materials, the magnetization is in general not a linear function of magnetizing current, especially as saturation is approached. As a result, the inductance is not constant but can depend on current in a fairly complicated way. In our discussion we will ignore this complication and assume always that the inductance is constant. This is a reasonable assumption even for a ferromagnetic material if the magnetization remains well below the saturation level.

Because automobiles contain steel, a ferromagnetic material, driving an automobile over a coil causes an appreciable increase in the coil's inductance. ? This effect is used in traffic light sensors, which use a large, current-carrying coil embedded under the road surface near an intersection. The circuitry connected to the coil detects the inductance change as a car drives over. When a preprogrammed number of cars have passed over the coil, the light changes to green to allow the cars through the intersection.

EXAMPLE 30.3 CALCULATING SELF-INDUCTANCE

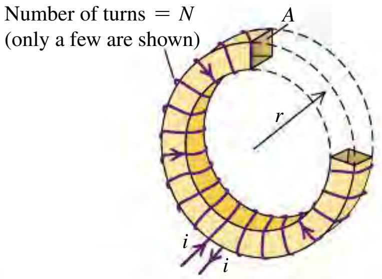
SOLUTION

Determine the self-inductance of a toroidal solenoid with cross-sectional area A and mean radius r, closely wound with N turns of wire on a nonmagnetic core (**Fig. 30.8**). Assume that B is uniform across a cross section (that is, neglect the variation of B with distance from the toroid axis).

SOLUTION

IDENTIFY and SET UP: Our target variable is the self-inductance L of the toroidal solenoid. We can find L by using Eq. (30.6), which requires knowing the flux Φ_B through each turn and the

30.8 Determining the self-inductance of a closely wound toroidal solenoid. For clarity, only a few turns of the winding are shown. Part of the toroid has been cut away to show the cross-sectional area A and radius r.



current i in the coil. For this, we use the results of Example 28.10 (Section 28.7), in which we found the magnetic field in the interior of a toroidal solenoid as a function of the current.

EXECUTE: From Eq. (30.6), the self-inductance is $L = N\Phi_B/i$. From Example 28.10, the field magnitude at a distance r from the toroid axis is $B = \mu_0 Ni/2\pi r$. If we assume that the field has this magnitude over the entire cross-sectional area A, then

$$\Phi_B = BA = \frac{\mu_0 NiA}{2\pi r}$$

The flux Φ_B is the same through each turn, and

$$L = \frac{N\Phi_B}{i} = \frac{\mu_0 N^2 A}{2\pi r} \quad \text{(self-inductance of a toroidal solenoid)}$$

EVALUATE: Suppose $N = 200$ turns, $A = 5.0\ \text{cm}^2 = 5.0 \times 10^{-4}\ \text{m}^2$, and $r = 0.10$ m; then

$$L = \frac{(4\pi \times 10^{-7}\ \text{Wb/A} \cdot \text{m})(200)^2(5.0 \times 10^{-4}\ \text{m}^2)}{2\pi(0.10\ \text{m})}$$
$$= 40 \times 10^{-6}\ \text{H} = 40\ \mu\text{H}$$

EXAMPLE 30.4 CALCULATING SELF-INDUCED EMF

SOLUTION

If the current in the toroidal solenoid in Example 30.3 increases uniformly from 0 to 6.0 A in 3.0 μs, find the magnitude and direction of the self-induced emf.

SOLUTION

IDENTIFY and SET UP: We are given L, the self-inductance, and di/dt, the rate of change of the solenoid current. We find the magnitude of the self-induced emf $\mathcal{E}$ by using Eq. (30.7) and its direction by using Lenz's law.

EXECUTE: We have $di/dt = (6.0\ \text{A})/(3.0 \times 10^{-6}\ \text{s}) = 2.0 \times 10^6\ \text{A/s}$. From Eq. (30.7), the magnitude of the induced emf is

$$|\mathcal{E}| = L\left|\frac{di}{dt}\right| = (40 \times 10^{-6}\ \text{H})(2.0 \times 10^6\ \text{A/s}) = 80\ \text{V}$$

The current is increasing, so according to Lenz's law the direction of the emf is opposite to that of the current. This corresponds to the situation in Fig. 30.6c; the emf is in the direction from b to a, like a battery with a as the + terminal and b the − terminal, tending to oppose the current increase from the external circuit.

EVALUATE: This example shows that even a small inductance L can give rise to a substantial induced emf if the current changes rapidly.

TEST YOUR UNDERSTANDING OF SECTION 30.2 Rank the following inductors in order of the potential difference V_{ab}, from most positive to most negative. In each case the inductor has zero resistance and the current flows from point a through the inductor to point b. (i) The current through a 2.0-μH inductor increases from 1.0 A to 2.0 A in 0.50 s; (ii) the current through a 4.0-μH inductor decreases from 3.0 A to 0 in 2.0 s; (iii) the current through a 1.0-μH inductor remains constant at 4.0 A; (iv) the current through a 1.0-μH inductor increases from 0 to 4.0 A in 0.25 s.

30.3 MAGNETIC-FIELD ENERGY

Establishing a current in an inductor requires an input of energy, and an inductor carrying a current has energy stored in it. Let's see how this comes about. In Fig. 30.5, an increasing current i in the inductor causes an emf $\mathcal{E}$ between its terminals and a corresponding potential difference V_{ab} between the terminals of the source, with point a at higher potential than point b. Thus the source must be adding energy to the inductor, and the instantaneous power P (rate of transfer of energy into the inductor) is $P = V_{ab}i$.

Energy Stored in an Inductor

We can calculate the total energy input U needed to establish a final current I in an inductor with inductance L if the initial current is zero. We assume that the inductor has zero resistance, so no energy is dissipated within the inductor. Let the rate of change of the current i at some instant be di/dt; the current is increasing, so $di/dt > 0$. The voltage between the terminals a and b of the inductor at this instant is $V_{ab} = L\,di/dt$, and the rate P at which energy is being delivered to the inductor (equal to the instantaneous power supplied by the external source) is

$$P = V_{ab}i = Li\frac{di}{dt}$$

The energy dU supplied to the inductor during an infinitesimal time interval dt is $dU = P\,dt$, so

$$dU = Li\,di$$

The total energy U supplied while the current increases from zero to a final value I is

Energy stored in an inductor ········ Inductance ········ Final current

$$U = L\int_0^I i\,di = \tfrac{1}{2}LI^2 \qquad (30.9)$$

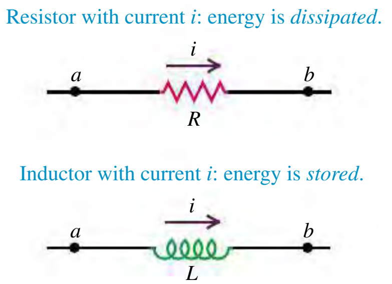
Integral from initial (zero) value of instantaneous current to final value

After the current has reached its final steady value I, $di/dt = 0$ and no more energy is input to the inductor. When there is no current, the stored energy U is zero; when the current is I, the energy is $\frac{1}{2}LI^2$.

When the current decreases from I to zero, the inductor acts as a source that supplies a total amount of energy $\frac{1}{2}LI^2$ to the external circuit. If we interrupt the circuit suddenly by opening a switch, the current decreases very rapidly, the induced emf is very large, and the energy may be dissipated in an arc across the switch contacts.

CAUTION **Energy, resistors, and inductors** Don't confuse the behavior of resistors and inductors where energy is concerned (**Fig. 30.9**). Energy flows into a resistor whenever a current passes through it, whether the current is steady or varying; this energy is dissipated in the form of heat. By contrast, energy flows into an ideal, zero-resistance inductor only when the current in the inductor *increases*. This energy is not dissipated; it is stored in the inductor and released when the current *decreases*. When a steady current flows through an inductor, there is no energy flow in or out.

30.9 A resistor is a device in which energy is irrecoverably dissipated. By contrast, energy stored in a current-carrying inductor can be recovered when the current decreases to zero.



Magnetic Energy Density

The energy in an inductor is actually stored in the magnetic field of the coil, just as the energy of a capacitor is stored in the electric field between its plates. We can develop relationships for magnetic-field energy analogous to those we obtained for electric-field energy in Section 24.3 [Eqs. (24.9) and (24.11)]. We will concentrate on one simple case, the ideal toroidal solenoid. This system has the advantage that its magnetic field is confined completely to a finite region of space within its core. As in Example 30.3, we assume that the cross-sectional area A is small enough that we can pretend that the magnetic field is uniform over the area. The volume V enclosed by the toroidal solenoid is approximately equal to the circumference $2\pi r$ multiplied by the area A: $V = 2\pi rA$. From Example 30.3, the self-inductance of the toroidal solenoid with vacuum within its coils is

$$L = \frac{\mu_0 N^2 A}{2\pi r}$$

From Eq. (30.9), the energy U stored in the toroidal solenoid when the current is I is

$$U = \tfrac{1}{2}LI^2 = \tfrac{1}{2}\frac{\mu_0 N^2 A}{2\pi r}I^2$$

The magnetic field and therefore this energy are localized in the volume $V = 2\pi r A$ enclosed by the windings. The energy *per unit volume,* or *magnetic energy density,* is $u = U/V$:

$$u = \frac{U}{2\pi r A} = \tfrac{1}{2}\mu_0 \frac{N^2 I^2}{(2\pi r)^2}$$

We can express this in terms of the magnitude B of the magnetic field inside the toroidal solenoid. From Eq. (28.24) in Example 28.10 (Section 28.7), this is

$$B = \frac{\mu_0 N I}{2\pi r}$$

and so

$$\frac{N^2 I^2}{(2\pi r)^2} = \frac{B^2}{\mu_0^2}$$

When we substitute this into the above equation for u, we finally find the expression for **magnetic energy density** in vacuum:

Magnetic energy density in vacuum

$$u = \frac{B^2}{2\mu_0} \qquad (30.10)$$

Magnetic-field magnitude (B); Magnetic constant (μ_0)

This is the magnetic analog of the energy per unit volume in an *electric* field in vacuum, $u = \tfrac{1}{2}\epsilon_0 E^2$, which we derived in Section 24.3. As an example, the energy density in the 1.5-T magnetic field of an MRI scanner (see Section 27.7) is $u = B^2/2\mu_0 = (1.5\ \mathrm{T})^2/(2 \times 4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}) = 9.0 \times 10^5\ \mathrm{J/m^3}$.

When the material inside the toroid is not vacuum but a material with (constant) magnetic permeability $\mu = K_m \mu_0$, we replace μ_0 by μ in Eq. (30.10):

Magnetic energy density in a material

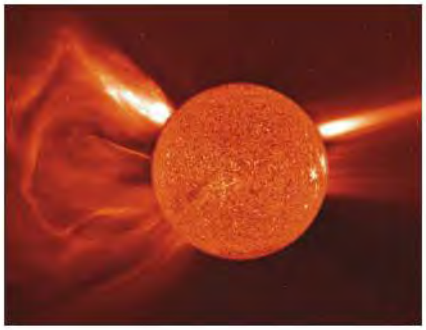
$$u = \frac{B^2}{2\mu} \qquad (30.11)$$

Magnetic-field magnitude (B); Permeability of material (μ)

Although we have derived Eq. (30.11) for only one special situation, it turns out to be the correct expression for the energy per unit volume associated with *any* magnetic-field configuration in a material with constant permeability. For vacuum, Eq. (30.11) reduces to Eq. (30.10). We will use the expressions for electric-field and magnetic-field energy in Chapter 32 when we study the energy associated with electromagnetic waves.

Magnetic-field energy plays an important role in the ignition systems of gasoline-powered automobiles. A primary coil of about 250 turns is connected to the car's battery and produces a strong magnetic field. This coil is surrounded by a secondary coil with some 25,000 turns of very fine wire. When it is time for a spark plug to fire (see Fig. 20.5 in Section 20.3), the current to the primary coil is interrupted, the magnetic field quickly drops to zero, and an emf of tens of thousands of volts is induced in the secondary coil. The energy stored in the magnetic field thus goes into a powerful pulse of current that travels through the secondary coil to the spark plug, generating the spark that ignites the fuel–air mixture in the engine's cylinders (**Fig. 30.10**).

Application A Magnetic Eruption on the Sun This composite of two images of the sun shows a coronal mass ejection, a dramatic event in which about 10^{12} kg (a billion tons) of material from the sun's outer atmosphere is ejected into space at speeds of 500 km/s or faster. Such ejections happen at intervals of a few hours to a few days. These immense eruptions are powered by the energy stored in the sun's magnetic field. Unlike the earth's relatively steady magnetic field, the sun's field is constantly changing, and regions of unusually strong field (and hence unusually high magnetic energy density) frequently form. A coronal mass ejection occurs when the energy stored in such a region is suddenly released.



30.10 The energy required to fire an automobile spark plug is derived from magnetic-field energy stored in the ignition coil.

SOLUTION

EXAMPLE 30.5 STORING ENERGY IN AN INDUCTOR

The electric-power industry would like to find efficient ways to store electrical energy generated during low-demand hours to help meet customer requirements during high-demand hours. Could a large inductor be used? What inductance would be needed to store 1.00 kW·h of energy in a coil carrying a 200-A current?

SOLUTION

IDENTIFY and SET UP: We are given the required amount of stored energy U and the current $I = 200$ A. We use Eq. (30.9) to find the self-inductance L.

EXECUTE: Here we have $I = 200\text{ A}$ and $U = 1.00\text{ kW}\cdot\text{h} = (1.00 \times 10^3\text{ W})(3600\text{ s}) = 3.60 \times 10^6\text{ J}$. Solving Eq. (30.9) for L, we find

$$L = \frac{2U}{I^2} = \frac{2(3.60 \times 10^6\text{ J})}{(200\text{ A})^2} = 180\text{ H}$$

EVALUATE: The required inductance is more than a million times greater than the self-inductance of the toroidal solenoid of Example 30.3. Conventional wires that are to carry 200 A would have to be of large diameter to keep the resistance low and avoid unacceptable energy losses due to I^2R heating. As a result, a 180-H inductor using conventional wire would be very large (room-size). A superconducting inductor could be much smaller, since the resistance of a superconductor is zero and much thinner wires could be used; but the wires would have to be kept at low temperature to remain superconducting, and maintaining this temperature would itself require energy. This scheme is impractical with present technology.

TEST YOUR UNDERSTANDING OF SECTION 30.3 The current in a solenoid is reversed in direction while keeping the same magnitude. (a) Does this change the magnetic field within the solenoid? (b) Does this change the magnetic energy density in the solenoid?

30.4 THE *R-L* CIRCUIT

Let's look at some examples of the circuit behavior of an inductor. One thing is clear already; an inductor in a circuit makes it difficult for rapid changes in current to occur, thanks to the effects of self-induced emf. Equation (30.7) shows that the greater the rate of change of current di/dt, the greater the self-induced emf and the greater the potential difference between the inductor terminals. This equation, together with Kirchhoff's rules (see Section 26.2), gives us the principles we need to analyze circuits containing inductors.

PROBLEM-SOLVING STRATEGY 30.1 INDUCTORS IN CIRCUITS

IDENTIFY *the relevant concepts:* An inductor is just another circuit element, like a source of emf, a resistor, or a capacitor. One key difference is that when an inductor is included in a circuit, all the voltages, currents, and capacitor charges are in general functions of time, not constants as they have been in most of our previous circuit analysis. But even when the voltages and currents vary with time, Kirchhoff's rules (see Section 26.2) hold at each instant of time.

SET UP *the problem* using the following steps:
1. Follow the procedure given in Problem-Solving Strategy 26.2 (Section 26.2). Draw a circuit diagram and label all quantities, known and unknown. Apply the junction rule immediately to express the currents in terms of as few quantities as possible.
2. Determine which quantities are the target variables.

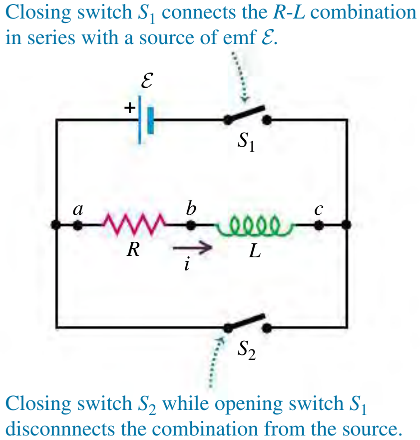
EXECUTE *the solution* as follows:
1. As in Problem-Solving Strategy 26.2, apply Kirchhoff's loop rule to each loop in the circuit.
2. Review the sign rules given in Problem-Solving Strategy 26.2. To get the correct sign for the potential difference between the terminals of an inductor, apply Lenz's law and the sign rule described in Section 30.2 in connection with Eq. (30.7) and Fig. 30.6. In Kirchhoff's loop rule, when we go through an inductor in the *same* direction as the assumed current, we encounter a voltage *drop* equal to $L\,di/dt$, so the corresponding term in the loop equation is $-L\,di/dt$. When we go through an inductor in the *opposite* direction from the assumed current, the potential difference is reversed and the term to use in the loop equation is $+L\,di/dt$.
3. Solve for the target variables.

EVALUATE *your answer:* Check whether your answer is consistent with the behavior of inductors. By Lenz's law, if the current through an inductor is changing, your result should indicate that the potential difference across the inductor opposes the change.

Current Growth in an *R-L* Circuit

We can learn a great deal about inductor behavior by analyzing the circuit of **Fig. 30.11**. A circuit that includes both a resistor and an inductor, and possibly a source of emf, is called an ***R-L* circuit.** The inductor helps to prevent rapid changes in current, which can be useful if a steady current is required but the source has a fluctuating emf. The resistor R may be a separate circuit element, or it may be the resistance of the inductor windings; every real-life inductor has some resistance unless it is made of superconducting wire. By closing switch S_1, we can connect the R-L combination to a source with constant emf $\mathcal{E}$. (We assume that the source has zero internal resistance, so the terminal voltage equals $\mathcal{E}$.)

30.11 An *R-L* circuit.



Suppose both switches are open to begin with, and then at some initial time $t = 0$ we close switch S_1. The current cannot change suddenly from zero to some final value, since di/dt and the induced emf in the inductor would both be infinite. Instead, the current begins to grow at a rate that depends on the value of L in the circuit.

Let i be the current at some time t after switch S_1 is closed, and let di/dt be its rate of change at that time. The potential differences v_{ab} (across the resistor) and v_{bc} (across the inductor) are

$$v_{ab} = iR \quad \text{and} \quad v_{bc} = L\frac{di}{dt}$$

Note that if the current is in the direction shown in Fig. 30.11 and is increasing, then both v_{ab} and v_{bc} are positive; a is at a higher potential than b, which in turn is at a higher potential than c. (Compare to Figs. 30.6a and c.) We apply Kirchhoff's loop rule, starting at the negative terminal and proceeding counterclockwise around the loop:

$$\mathcal{E} - iR - L\frac{di}{dt} = 0 \tag{30.12}$$

Solving this for di/dt, we find that the rate of increase of current is

$$\frac{di}{dt} = \frac{\mathcal{E} - iR}{L} = \frac{\mathcal{E}}{L} - \frac{R}{L}i \tag{30.13}$$

At the instant that switch S_1 is first closed, $i = 0$ and the potential drop across R is zero. The initial rate of change of current is

$$\left(\frac{di}{dt}\right)_{\text{initial}} = \frac{\mathcal{E}}{L}$$

The greater the inductance L, the more slowly the current increases.

As the current increases, the term $(R/L)i$ in Eq. (30.13) also increases, and the *rate* of increase of current given by Eq. (30.13) becomes smaller and smaller. This means that the current is approaching a final, steady-state value I. When the current reaches this value, its rate of increase is zero. Then Eq. (30.13) becomes

$$\left(\frac{di}{dt}\right)_{\text{final}} = 0 = \frac{\mathcal{E}}{L} - \frac{R}{L}I \quad \text{and}$$

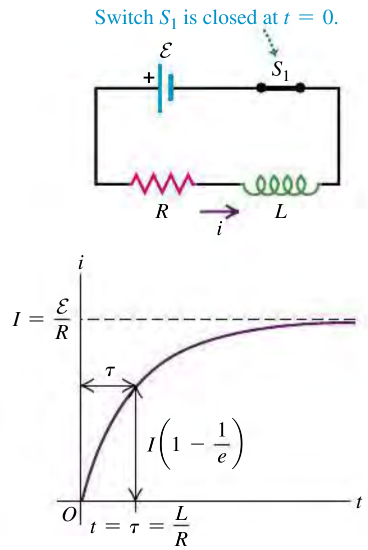
$$I = \frac{\mathcal{E}}{R}$$

The *final* current I does not depend on the inductance L; it is the same as it would be if the resistance R alone were connected to the source with emf $\mathcal{E}$.

Figure 30.12 shows the behavior of the current as a function of time. To derive the equation for this curve (that is, an expression for current as a function of time), we proceed just as we did for the charging capacitor in Section 26.4. First we rearrange Eq. (30.13) to the form

$$\frac{di}{i - (\mathcal{E}/R)} = -\frac{R}{L}dt$$

30.12 Graph of i versus t for growth of current in an *R-L* circuit with an emf in series. The final current is $I = \mathcal{E}/R$; after one time constant τ, the current is $1 - 1/e$ of this value.



This separates the variables, with i on the left side and t on the right. Then we integrate both sides, renaming the integration variables i' and t' so that we can use i and t as the upper limits. (The lower limit for each integral is zero, corresponding to zero current at the initial time $t = 0$.) We get

$$\int_0^i \frac{di'}{i' - (\mathcal{E}/R)} = -\int_0^t \frac{R}{L}dt'$$

$$\ln\left(\frac{i - (\mathcal{E}/R)}{-\mathcal{E}/R}\right) = -\frac{R}{L}t$$

Now we take exponentials of both sides and solve for i. We leave the details for you to work out; the final result is the equation of the curve in Fig. 30.12:

$$i = \frac{\mathcal{E}}{R}\left(1 - e^{-(R/L)t}\right) \qquad \text{(current in an } R\text{-}L \text{ circuit with emf)} \tag{30.14}$$

Taking the derivative of Eq. (30.14), we find

$$\frac{di}{dt} = \frac{\mathcal{E}}{L}e^{-(R/L)t} \tag{30.15}$$

At time $t = 0$, $i = 0$ and $di/dt = \mathcal{E}/L$. As $t \rightarrow \infty$, $i \rightarrow \mathcal{E}/R$ and $di/dt \rightarrow 0$, as we predicted.

As Fig. 30.12 shows, the instantaneous current i first rises rapidly, then increases more slowly and approaches the final value $I = \mathcal{E}/R$ asymptotically. At a time equal to L/R, the current has risen to $(1 - 1/e)$, or about 63%, of its final value. The quantity L/R is therefore a measure of how quickly the current builds toward its final value; this quantity is called the **time constant** for the circuit, denoted by τ:

Time constant for an R-L circuit

$$\tau = \frac{L}{R} \qquad \begin{matrix}\text{Inductance} \\ \text{Resistance}\end{matrix} \tag{30.16}$$

In a time equal to 2τ, the current reaches 86% of its final value; in 5τ, 99.3%; and in 10τ, 99.995%. (Compare the discussion in Section 26.4 of charging a capacitor of capacitance C that was in series with a resistor of resistance R; the time constant for that situation was the product RC.)

The graphs of i versus t have the same general shape for all values of L. For a given value of R, the time constant τ is greater for greater values of L. When L is small, the current rises rapidly to its final value; when L is large, it rises more slowly. For example, if $R = 100\ \Omega$ and $L = 10$ H,

$$\tau = \frac{L}{R} = \frac{10\text{ H}}{100\ \Omega} = 0.10\text{ s}$$

and the current increases to about 63% of its final value in 0.10 s. (Recall that $1\text{ H} = 1\ \Omega \cdot \text{s}$.) But if $L = 0.010$ H, $\tau = 1.0 \times 10^{-4}\text{ s} = 0.10$ ms, and the rise is much more rapid.

Energy considerations offer us additional insight into the behavior of an R-L circuit. The instantaneous rate at which the source delivers energy to the circuit is $P = \mathcal{E}i$. The instantaneous rate at which energy is dissipated in the resistor is i^2R, and the rate at which energy is stored in the inductor is $iv_{bc} = Li\,di/dt$ [or, equivalently, $(d/dt)\left(\frac{1}{2}Li^2\right) = Li\,di/dt$]. When we multiply Eq. (30.12) by i and rearrange, we find

$$\mathcal{E}i = i^2R + Li\frac{di}{dt} \tag{30.17}$$

Of the power $\mathcal{E}i$ supplied by the source, part (i^2R) is dissipated in the resistor and part $(Li\,di/dt)$ goes to store energy in the inductor. This discussion is analogous to our power analysis for a charging capacitor, given at the end of Section 26.4.

EXAMPLE 30.6 ANALYZING AN *R-L* CIRCUIT

SOLUTION

A sensitive electronic device of resistance $R = 175\ \Omega$ is to be connected to a source of emf (of negligible internal resistance) by a switch. The device is designed to operate with a 36-mA current, but to avoid damage to the device, the current can rise to no more than 4.9 mA in the first 58 μs after the switch is closed. An inductor is therefore connected in series with the device, as in Fig. 30.11; the switch in question is S_1. (a) What is the required source emf $\mathcal{E}$? (b) What is the required inductance L? (c) What is the R-L time constant τ?

SOLUTION

IDENTIFY and SET UP: This problem concerns current and current growth in an R-L circuit, so we can use the ideas of this section. Figure 30.12 shows the current i versus the time t that has elapsed since closing S_1. The graph shows that the final current is $I = \mathcal{E}/R$; we are given $R = 175\ \Omega$, so the emf is determined by the requirement that the final current be $I = 36$ mA. The other requirement is that the current be no more than $i = 4.9$ mA at $t = 58\ \mu$s; to satisfy this, we use Eq. (30.14) for the current as a function of time and solve for the inductance, which is the only unknown quantity. Equation (30.16) then tells us the time constant.

EXECUTE: (a) We solve $I = \mathcal{E}/R$ for $\mathcal{E}$:

$$\mathcal{E} = IR = (0.036\ \text{A})(175\ \Omega) = 6.3\ \text{V}$$

(b) To find the required inductance, we solve Eq. (30.14) for L. First we multiply through by $(-R/\mathcal{E})$ and add 1 to both sides:

$$1 - \frac{iR}{\mathcal{E}} = e^{-(R/L)t}$$

Then we take natural logs of both sides, solve for L, and substitute:

$$L = \frac{-Rt}{\ln(1 - iR/\mathcal{E})}$$

$$= \frac{-(175\ \Omega)(58 \times 10^{-6}\ \text{s})}{\ln[1 - (4.9 \times 10^{-3}\ \text{A})(175\ \Omega)/(6.3\ \text{V})]} = 69\ \text{mH}$$

(c) From Eq. (30.16),

$$\tau = \frac{L}{R} = \frac{69 \times 10^{-3}\ \text{H}}{175\ \Omega} = 3.9 \times 10^{-4}\ \text{s} = 390\ \mu\text{s}$$

EVALUATE: Note that 58 μs is much less than the time constant. In 58 μs the current builds up from zero to 4.9 mA, a small fraction of its final value of 36 mA; after 390 μs the current equals $(1 - 1/e)$ of its final value, or about $(0.63)(36\ \text{mA}) = 23$ mA.

Current Decay in an *R-L* Circuit

Now suppose switch S_1 in the circuit of Fig. 30.11 has been closed for a while and the current has reached the value I_0. Resetting our stopwatch to redefine the initial time, we close switch S_2 at time $t = 0$, bypassing the battery. (At the same time we should open S_1 to protect the battery.) The current through R and L does not instantaneously go to zero but decays smoothly, as shown in **Fig. 30.13**. The Kirchhoff's-rule loop equation is obtained from Eq. (30.12) by omitting the $\mathcal{E}$ term. We challenge you to retrace the steps in the above analysis and show that the current i varies with time according to

$$i = I_0 e^{-(R/L)t} \tag{30.18}$$

where I_0 is the initial current at time $t = 0$. The time constant, $\tau = L/R$, is the time for current to decrease to $1/e$, or about 37%, of its original value. In time 2τ it has dropped to 13.5%, in time 5τ to 0.67%, and in 10τ to 0.0045%.

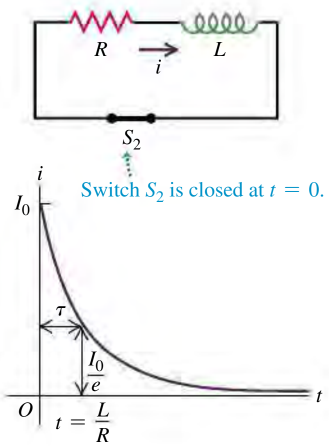
The energy that is needed to maintain the current during this decay is provided by the energy stored in the magnetic field of the inductor. The detailed energy analysis is simpler this time. In place of Eq. (30.17) we have

$$0 = i^2R + Li\frac{di}{dt} \tag{30.19}$$

Now $Li\,di/dt$ is negative; Eq. (30.19) shows that the energy stored in the inductor *decreases* at the same rate i^2R at which energy is dissipated in the resistor.

This entire discussion should look familiar; the situation is very similar to that of a charging and discharging capacitor, analyzed in Section 26.4. It would be a good idea to compare that section with our discussion of the R-L circuit.

30.13 Graph of i versus t for decay of current in an R-L circuit. After one time constant τ, the current is $1/e$ of its initial value.



SOLUTION

EXAMPLE 30.7 ENERGY IN AN *R-L* CIRCUIT

When the current in an *R-L* circuit is decaying, what fraction of the original energy stored in the inductor has been dissipated after 2.3 time constants?

SOLUTION

IDENTIFY and SET UP: This problem concerns current decay in an *R-L* circuit as well as the relationship between the current in an inductor and the amount of stored energy. The current i at any time t is given by Eq. (30.18); the stored energy associated with this current is given by Eq. (30.9), $U = \frac{1}{2}Li^2$.

EXECUTE: From Eq. (30.18), the current i at any time t is

$$i = I_0 e^{-(R/L)t}$$

We substitute this into $U = \frac{1}{2}Li^2$ to obtain an expression for the stored energy at any time:

$$U = \tfrac{1}{2}LI_0^2 e^{-2(R/L)t} = U_0 e^{-2(R/L)t}$$

where $U_0 = \frac{1}{2}LI_0^2$ is the energy at the initial time $t = 0$. When $t = 2.3\tau = 2.3L/R$, we have

$$U = U_0 e^{-2(2.3)} = U_0 e^{-4.6} = 0.010U_0$$

That is, only 0.010 or 1.0% of the energy initially stored in the inductor remains, so 99.0% has been dissipated in the resistor.

EVALUATE: To get a sense of what this result means, consider the *R-L* circuit we analyzed in Example 30.6, for which $\tau = 390\ \mu\text{s}$. With $L = 69\ \text{mH}$ and $I_0 = 36\ \text{mA}$, we have $U_0 = \frac{1}{2}LI_0^2 = \frac{1}{2}(0.069\ \text{H})(0.036\ \text{A})^2 = 4.5 \times 10^{-5}\ \text{J}$. Of this, 99.0% or 4.4×10^{-5} J is dissipated in $2.3(390\ \mu\text{s}) = 9.0 \times 10^{-4}\ \text{s} = 0.90\ \text{ms}$. In other words, this circuit can be almost completely powered off (or powered on) in 0.90 ms, so the minimum time for a complete on–off cycle is 1.8 ms. Even shorter cycle times are required for many purposes, such as in fast switching networks for telecommunications. In such cases a smaller time constant $\tau = L/R$ is needed.

DATA *SPEAKS*

Inductors in Circuits

When students were given a problem involving an *R-L* circuit, more than 23% gave an incorrect response. Common errors:

- Confusion about current and its rate of change. The current i cannot change abruptly in a circuit with an inductor, so i must be a continuous function of time t. However, di/dt can change abruptly (say, when the emf in Fig. 30.11 is connected to the circuit).
- Confusion about initial and final values. When an emf is connected to an *R-L* circuit as in Fig. 30.12, the inductor opposes current change and so $i = 0$ just after the switch is closed. Long after the switch is closed and the current has stabilized, the inductor acts like a simple wire and has no effect.

TEST YOUR UNDERSTANDING OF SECTION 30.4 (a) In Fig. 30.11, what are the algebraic signs of the potential differences v_{ab} and v_{bc} when switch S_1 is closed and switch S_2 is open? (i) $v_{ab} > 0$, $v_{bc} > 0$; (ii) $v_{ab} > 0$, $v_{bc} < 0$; (iii) $v_{ab} < 0$, $v_{bc} > 0$; (iv) $v_{ab} < 0$, $v_{bc} < 0$. (b) What are the signs of v_{ab} and v_{bc} when S_1 is open, S_2 is closed, and current is flowing in the direction shown? (i) $v_{ab} > 0$, $v_{bc} > 0$; (ii) $v_{ab} > 0$, $v_{bc} < 0$; (iii) $v_{ab} < 0$, $v_{bc} > 0$; (iv) $v_{ab} < 0$, $v_{bc} < 0$.

30.5 THE *L-C* CIRCUIT

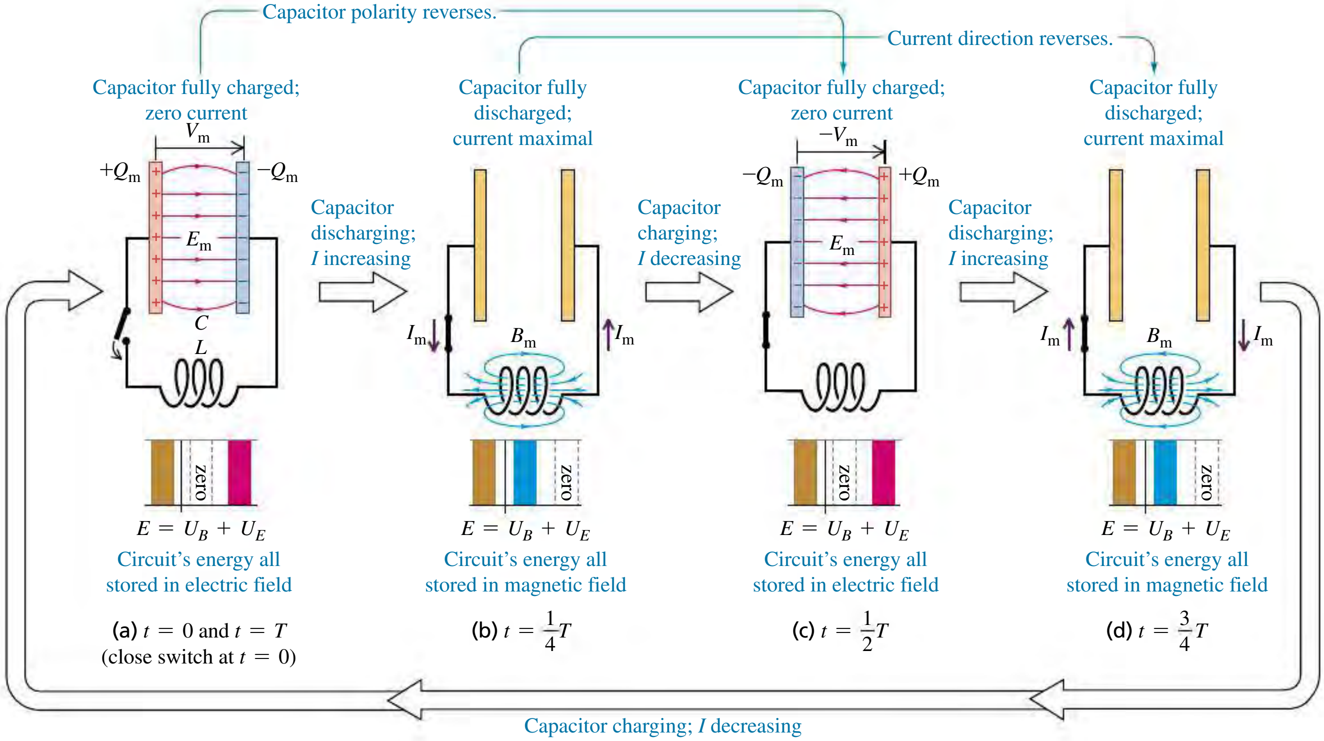
A circuit containing an inductor and a capacitor shows an entirely new mode of behavior, characterized by *oscillating* current and charge. This is in sharp contrast to the *exponential* approach to a steady-state situation that we have seen with both *R-C* and *R-L* circuits. In the ***L-C* circuit** in **Fig. 30.14a** we charge the capacitor to a potential difference V_m and initial charge $Q_m = CV_m$ on its left-hand plate and then close the switch. What happens?

The capacitor begins to discharge through the inductor. Because of the induced emf in the inductor, the current cannot change instantaneously; it starts at zero and eventually builds up to a maximum value I_m. During this buildup the capacitor is discharging. At each instant the capacitor potential equals the induced emf, so as the capacitor discharges, the *rate of change* of current decreases. When the capacitor potential becomes zero, the induced emf is also zero, and the current has leveled off at its maximum value I_m. Figure 30.14b shows this situation; the capacitor has completely discharged. The potential difference between its terminals (and those of the inductor) has decreased to zero, and the current has reached its maximum value I_m.

During the discharge of the capacitor, the increasing current in the inductor has established a magnetic field in the space around it, and the energy that was initially stored in the capacitor's electric field is now stored in the inductor's magnetic field.

Although the capacitor is completely discharged in Fig. 30.14b, the current persists (it cannot change instantaneously), and the capacitor begins to charge with polarity opposite to that in the initial state. As the current decreases, the magnetic field also decreases, inducing an emf in the inductor in the *same* direction as the

30.14 In an oscillating L-C circuit, the charge on the capacitor and the current through the inductor both vary sinusoidally with time. Energy is transferred between magnetic energy in the inductor (U_B) and electrical energy in the capacitor (U_E). As in simple harmonic motion, the total energy E remains constant. (Compare Fig. 14.14 in Section 14.3.)



current; this slows down the decrease of the current. Eventually, the current and the magnetic field reach zero, and the capacitor has been charged in the sense *opposite* to its initial polarity (Fig. 30.14c), with potential difference $-V_m$ and charge $-Q_m$ on its left-hand plate.

The process now repeats in the reverse direction; a little later, the capacitor has again discharged, and there is a current in the inductor in the opposite direction (Fig. 30.14d). Still later, the capacitor charge returns to its original value (Fig. 30.14a), and the whole process repeats. If there are no energy losses, the charges on the capacitor continue to oscillate back and forth indefinitely. This process is called an **electrical oscillation.** (Before you read further, review the analogous case of *mechanical* oscillation in Sections 14.2 and 14.3.)

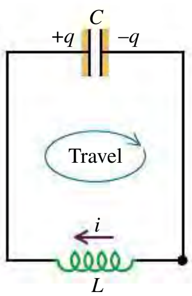
From an energy standpoint the oscillations of an electric circuit transfer energy from the capacitor's electric field to the inductor's magnetic field and back. The *total* energy associated with the circuit is constant. This is analogous to the transfer of energy in an oscillating mechanical system from potential energy to kinetic energy and back, with constant total energy (Section 14.3). As we will see, this analogy goes much further.

Electrical Oscillations in an *L-C* Circuit

To study the flow of charge in detail, we proceed just as we did for the R-L circuit. **Figure 30.15** shows our definitions of q and i.

CAUTION Positive current in an *L-C* circuit After you have examined Fig. 30.14, the positive direction for current in Fig. 30.15 may seem backward. In fact, we chose this direction to simplify the relationship between current and capacitor charge. We define the current at each instant to be $i = dq/dt$, the rate of change of the charge on the left-hand capacitor plate. If the capacitor is initially charged and begins to discharge as in Figs. 30.14a and 30.14b, then $dq/dt < 0$ and the initial current i is negative; the current direction is opposite to the (positive) direction shown in Fig. 30.15.

30.15 Applying Kirchhoff's loop rule to the L-C circuit. The direction of travel around the loop in the loop equation is shown. Just after the circuit is completed and the capacitor first begins to discharge, as in Fig. 30.14a, the current is negative (opposite to the direction shown).



We apply Kirchhoff's loop rule to the circuit in Fig. 30.15. Starting at the lower-right corner of the circuit and adding voltages as we go clockwise around the loop, we obtain

$$-L\frac{di}{dt} - \frac{q}{C} = 0$$

Since $i = dq/dt$, it follows that $di/dt = d^2q/dt^2$. We substitute this expression into the above equation and divide by $-L$ to obtain

$$\frac{d^2q}{dt^2} + \frac{1}{LC}q = 0 \qquad (L\text{-}C \text{ circuit}) \tag{30.20}$$

Equation (30.20) has exactly the same form as the equation we derived for simple harmonic motion in Section 14.2, Eq. (14.4): $d^2x/dt^2 = -(k/m)x$, or

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0$$

In an L-C circuit the capacitor charge q plays the role of the displacement x, and the current $i = dq/dt$ is analogous to the particle's velocity $v_x = dx/dt$. The inductance L is analogous to the mass m, and the reciprocal of the capacitance, $1/C$, is analogous to the force constant k.

Pursuing this analogy, we recall that the angular frequency $\omega = 2\pi f$ of the harmonic oscillator is equal to $(k/m)^{1/2}$ [Eq. (14.10)], and the position is given as a function of time by Eq. (14.13),

$$x = A\cos(\omega t + \phi)$$

where the amplitude A and the phase angle ϕ depend on the initial conditions. In the analogous electrical situation, the capacitor charge q is given by

$$q = Q\cos(\omega t + \phi) \tag{30.21}$$

and the angular frequency ω of oscillation is given by

Angular frequency of oscillation in an L-C circuit

$$\omega = \sqrt{\frac{1}{LC}} \tag{30.22}$$

Capacitance

Inductance

Verify that Eq. (30.21) satisfies the loop equation, Eq. (30.20), when ω has the value given by Eq. (30.22). In doing this, you will find that the instantaneous current $i = dq/dt$ is

$$i = -\omega Q\sin(\omega t + \phi) \tag{30.23}$$

Thus the charge and current in an L-C circuit oscillate sinusoidally with time, with an angular frequency determined by the values of L and C. The ordinary frequency f, the number of cycles per second, is equal to $\omega/2\pi$. The constants Q and ϕ in Eqs. (30.21) and (30.23) are determined by the initial conditions. If at time $t = 0$ the left-hand capacitor plate in Fig. 30.15 has its maximum charge Q and the current i is zero, then $\phi = 0$. If $q = 0$ at $t = 0$, then $\phi = \pm\pi/2$ rad.

Energy in an L-C Circuit

We can also analyze the L-C circuit by using an energy approach. The analogy to simple harmonic motion is equally useful here. In the mechanical problem a body with mass m is attached to a spring with force constant k. Suppose we displace the body a distance A from its equilibrium position and release it from rest at time $t = 0$. The kinetic energy of the system at any later time is $\frac{1}{2}mv_x^2$, and its elastic potential energy is $\frac{1}{2}kx^2$. Because the system is conservative, the sum of these energies equals the initial energy of the system, $\frac{1}{2}kA^2$. We find the velocity v_x at any position x just as we did in Section 14.3, Eq. (14.22):

$$v_x = \pm\sqrt{\frac{k}{m}}\sqrt{A^2 - x^2} \tag{30.24}$$

The L-C circuit is also a conservative system. Again let Q be the maximum capacitor charge. The magnetic-field energy $\frac{1}{2}Li^2$ in the inductor at any time corresponds to the kinetic energy $\frac{1}{2}mv^2$ of the oscillating body, and the electric-field energy $q^2/2C$ in the capacitor corresponds to the elastic potential energy $\frac{1}{2}kx^2$ of the spring. The sum of these energies is the total energy $Q^2/2C$ of the system:

$$\frac{1}{2}Li^2 + \frac{q^2}{2C} = \frac{Q^2}{2C} \tag{30.25}$$

The total energy in the L-C circuit is *constant;* it oscillates between the magnetic and the electric forms, just as the constant total mechanical energy in simple harmonic motion is constant and oscillates between the kinetic and potential forms.

Solving Eq. (30.25) for i, we find that when the charge on the capacitor is q, the current i is

$$i = \pm\sqrt{\frac{1}{LC}}\sqrt{Q^2 - q^2} \tag{30.26}$$

Verify this equation by substituting q from Eq. (30.21) and i from Eq. (30.23). Comparing Eqs. (30.24) and (30.26), we see that current $i = dq/dt$ and charge q are related in the same way as are velocity $v_x = dx/dt$ and position x in the mechanical problem.

Table 30.1 summarizes the analogies between simple harmonic motion and L-C circuit oscillations. The striking parallels shown there are so close that we can solve complicated mechanical problems by setting up analogous electric circuits and measuring the currents and voltages that correspond to the mechanical quantities to be determined. This is the basic principle of many analog computers. This analogy can be extended to *damped* oscillations, which we consider in the next section. In Chapter 31 we will extend the analogy further to include *forced* electrical oscillations, which occur in all alternating-current circuits.

TABLE 30.1 Oscillation of a Mass-Spring System Compared with Electrical Oscillation in an *L-C* Circuit

Mass-Spring System
Kinetic energy $= \frac{1}{2}mv_x^2$
Potential energy $= \frac{1}{2}kx^2$
$\frac{1}{2}mv_x^2 + \frac{1}{2}kx^2 = \frac{1}{2}kA^2$
$v_x = \pm\sqrt{k/m}\sqrt{A^2 - x^2}$
$v_x = dx/dt$
$\omega = \sqrt{\frac{k}{m}}$
$x = A\cos(\omega t + \phi)$
Inductor-Capacitor Circuit
Magnetic energy $= \frac{1}{2}Li^2$
Electrical energy $= q^2/2C$
$\frac{1}{2}Li^2 + q^2/2C = Q^2/2C$
$i = \pm\sqrt{1/LC}\sqrt{Q^2 - q^2}$
$i = dq/dt$
$\omega = \sqrt{\frac{1}{LC}}$
$q = Q\cos(\omega t + \phi)$

EXAMPLE 30.8 AN OSCILLATING CIRCUIT

SOLUTION

A 300-V dc power supply is used to charge a 25-μF capacitor. After the capacitor is fully charged, it is disconnected from the power supply and connected across a 10-mH inductor. The resistance in the circuit is negligible. (a) Find the frequency and period of oscillation of the circuit. (b) Find the capacitor charge and the circuit current 1.2 ms after the inductor and capacitor are connected.

SOLUTION

IDENTIFY and SET UP: Our target variables are the oscillation frequency f and period T, as well as the charge q and current i at a particular time t. We are given the capacitance C and the inductance L, with which we can calculate the frequency and period from Eq. (30.22). We find the charge and current from Eqs. (30.21) and (30.23). Initially the capacitor is fully charged and the current is zero, as in Fig. 30.14a, so the phase angle is $\phi = 0$ [see the discussion that follows Eq. (30.23)].

EXECUTE: (a) The natural angular frequency is

$$\omega = \sqrt{\frac{1}{LC}} = \sqrt{\frac{1}{(10 \times 10^{-3}\ \text{H})(25 \times 10^{-6}\ \text{F})}}$$
$$= 2.0 \times 10^3\ \text{rad/s}$$

The frequency f and period T are then

$$f = \frac{\omega}{2\pi} = \frac{2.0 \times 10^3\ \text{rad/s}}{2\pi\ \text{rad/cycle}} = 320\ \text{Hz}$$

$$T = \frac{1}{f} = \frac{1}{320\ \text{Hz}}$$
$$= 3.1 \times 10^{-3}\ \text{s} = 3.1\ \text{ms}$$

Continued

(b) Since the period of the oscillation is $T = 3.1$ ms, $t = 1.2$ ms equals $0.38T$; this corresponds to a situation intermediate between Fig. 30.14b ($t = T/4$) and Fig. 30.14c ($t = T/2$). Comparing those figures with Fig. 30.15, we expect the capacitor charge q to be negative (that is, there will be negative charge on the left-hand plate of the capacitor) and the current i to be negative as well (that is, current will flow counterclockwise).

To find q, we use Eq. (30.21), $q = Q\cos(\omega t + \phi)$. The charge is maximum at $t = 0$, so $\phi = 0$ and $Q = C\mathcal{E} = (25 \times 10^{-6}\ \text{F}) \times (300\ \text{V}) = 7.5 \times 10^{-3}$ C. Hence Eq. (30.21) becomes

$$q = (7.5 \times 10^{-3}\ \text{C})\cos\omega t$$

At time $t = 1.2 \times 10^{-3}$ s,

$$\omega t = (2.0 \times 10^3\ \text{rad/s})(1.2 \times 10^{-3}\ \text{s}) = 2.4\ \text{rad}$$

$$q = (7.5 \times 10^{-3}\ \text{C})\cos(2.4\ \text{rad}) = -5.5 \times 10^{-3}\ \text{C}$$

From Eq. (30.23), the current i at any time is $i = -\omega Q \sin\omega t$. At $t = 1.2 \times 10^{-3}$ s,

$$i = -(2.0 \times 10^3\ \text{rad/s})(7.5 \times 10^{-3}\ \text{C})\sin(2.4\ \text{rad}) = -10\ \text{A}$$

EVALUATE: The signs of both q and i are negative, as predicted.

EXAMPLE 30.9 ENERGY IN AN OSCILLATING CIRCUIT

SOLUTION

For the L-C circuit of Example 30.8, find the magnetic and electrical energies (a) at $t = 0$ and (b) at $t = 1.2$ ms.

SOLUTION

IDENTIFY and SET UP: We must calculate the magnetic energy U_B (stored in the inductor) and the electrical energy U_E (stored in the capacitor) at two times during the L-C circuit oscillation. From Example 30.8 we know the values of the capacitor charge q and circuit current i for both times. We use them to calculate $U_B = \frac{1}{2}Li^2$ and $U_E = q^2/2C$.

EXECUTE: (a) At $t = 0$ there is no current and $q = Q$. Hence there is no magnetic energy, and all the energy in the circuit is in the form of electrical energy in the capacitor:

$$U_B = \tfrac{1}{2}Li^2 = 0 \qquad U_E = \frac{Q^2}{2C} = \frac{(7.5 \times 10^{-3}\ \text{C})^2}{2(25 \times 10^{-6}\ \text{F})} = 1.1\ \text{J}$$

(b) From Example 30.8, at $t = 1.2$ ms we have $i = -10$ A and $q = -5.5 \times 10^{-3}$ C. Hence

$$U_B = \tfrac{1}{2}Li^2 = \tfrac{1}{2}(10 \times 10^{-3}\ \text{H})(-10\ \text{A})^2 = 0.5\ \text{J}$$

$$U_E = \frac{q^2}{2C} = \frac{(-5.5 \times 10^{-3}\ \text{C})^2}{2(25 \times 10^{-6}\ \text{F})} = 0.6\ \text{J}$$

EVALUATE: The magnetic and electrical energies are the same at $t = 3T/8 = 0.375T$, halfway between the situations in Figs. 30.14b and 30.14c. We saw in Example 30.8 that the time considered in part (b), $t = 1.2$ ms, equals $0.38T$; this is slightly later than $0.375T$, so U_B is slightly less than U_E. At *all* times the *total* energy $E = U_B + U_E$ has the same value, 1.1 J. An L-C circuit without resistance is a conservative system; no energy is dissipated.

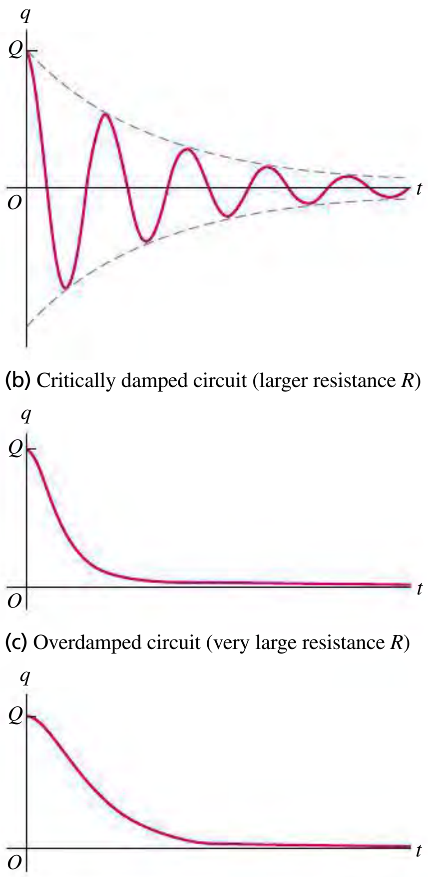
TEST YOUR UNDERSTANDING OF SECTION 30.5 One way to think about the energy stored in an L-C circuit is to say that the circuit elements do positive or negative work on the charges that move back and forth through the circuit. (a) Between stages (a) and (b) in Fig. 30.14, does the capacitor do positive or negative work on the charges? (b) What kind of force (electric or magnetic) does the capacitor exert on the charges to do this work? (c) During this process, does the inductor do positive or negative work on the charges? (d) What kind of force (electric or magnetic) does the inductor exert on the charges?

30.6 THE *L-R-C* SERIES CIRCUIT

In our discussion of the L-C circuit we assumed that there was no *resistance* in the circuit. This is an idealization, of course; every real inductor has resistance in its windings, and there may also be resistance in the connecting wires. Because of resistance, the electromagnetic energy in the circuit is dissipated and converted to other forms, such as internal energy of the circuit materials. Resistance in an electric circuit is analogous to friction in a mechanical system.

Suppose an inductor with inductance L and a resistor of resistance R are connected in series across the terminals of a charged capacitor, forming an ***L-R-C* series circuit.** As before, the capacitor starts to discharge as soon as the circuit is completed. But due to i^2R losses in the resistor, the magnetic-field energy that the inductor acquires when the capacitor is completely discharged is *less* than the original electric-field energy of the capacitor. In the same way, the energy of the capacitor when the magnetic field has decreased to zero is still less and so on.

If the resistance R of the resistor is relatively small, the circuit still oscillates, but with **damped harmonic motion** (**Fig. 30.16a**), and we say that the circuit is **underdamped.** If we increase R, the oscillations die out more rapidly. When R reaches a certain value, the circuit no longer oscillates; it is **critically damped** (Fig. 30.16b). For still larger values of R, the circuit is **overdamped** (Fig. 30.16c), and the capacitor charge approaches zero even more slowly. We used these same terms to describe the behavior of the analogous mechanical system, the damped harmonic oscillator, in Section 14.7.



30.16 Graphs of capacitor charge as a function of time in an *L-R-C* series circuit with initial charge Q.

Analyzing an *L-R-C* Series Circuit

To analyze *L-R-C* series circuit behavior in detail, consider the circuit shown in **Fig. 30.17**. It is like the *L-C* circuit of Fig. 30.15 except for the added resistor R; we also show the source that charges the capacitor initially. The labeling of the positive senses of q and i is the same as for the *L-C* circuit.

First we close the switch in the upward position, connecting the capacitor to a source of emf $\mathcal{E}$ for a long enough time to ensure that the capacitor acquires its final charge $Q = C\mathcal{E}$ and any initial oscillations have died out. Then at time $t = 0$ we flip the switch to the downward position, removing the source from the circuit and placing the capacitor in series with the resistor and inductor. Note that the initial current is negative, opposite to the direction of i shown in Fig. 30.17.

To find how q and i vary with time, we apply Kirchhoff's loop rule. Starting at point a and going around the loop in the direction *abcda,* we obtain

$$-iR - L\frac{di}{dt} - \frac{q}{C} = 0$$

Replacing i with dq/dt and rearranging, we get

$$\frac{d^2q}{dt^2} + \frac{R}{L}\frac{dq}{dt} + \frac{1}{LC}q = 0 \qquad (30.27)$$

Note that when $R = 0$, this reduces to Eq. (30.20) for an *L-C* circuit.

There are general methods for obtaining solutions of Eq. (30.27). The form of the solution is different for the underdamped (small R) and overdamped (large R) cases. When R^2 is less than $4L/C$, the solution has the form

$$q = Ae^{-(R/2L)t}\cos\left(\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\,t + \phi\right) \qquad (30.28)$$

where A and ϕ are constants. You can take the first and second derivatives of this function and show by direct substitution that it does satisfy Eq. (30.27).

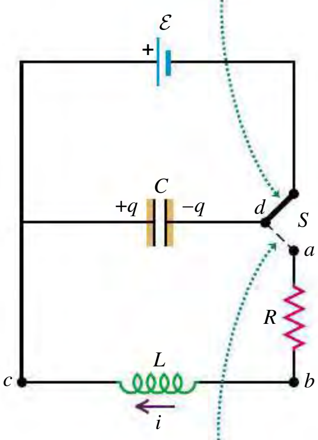
This solution corresponds to the *underdamped* behavior shown in Fig. 30.16a; the function represents a sinusoidal oscillation with an exponentially decaying amplitude. (Note that the exponential factor $e^{-(R/2L)t}$ is *not* the same as the factor $e^{-(R/L)t}$ that we encountered in describing the *R-L* circuit in Section 30.4.) When $R = 0$, Eq. (30.28) reduces to Eq. (30.21) for the oscillations in an *L-C* circuit. If R is not zero, the angular frequency of the oscillation is *less* than $1/(LC)^{1/2}$ because of the term containing R. The angular frequency ω' of the damped oscillations is given by

Angular frequency of underdamped oscillations in an *L-R-C* series circuit

$$\omega' = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \qquad (30.29)$$

Resistance (R); Inductance (L in $4L^2$); Inductance (L in LC); Capacitance (C)

30.17 An *L-R-C* series circuit.



When $R = 0$, this reduces to Eq. (30.22), $\omega = (1/LC)^{1/2}$. As R increases, ω' becomes smaller and smaller. When $R^2 = 4L/C$, the quantity under the radical becomes zero; the system no longer oscillates, and the case of *critical damping* (Fig. 30.16b) has been reached. For still larger values of R the system behaves as in Fig. 30.16c. In this case the circuit is *overdamped,* and q is given as a function of time by the sum of two decreasing exponential functions.

In the *underdamped* case the phase constant ϕ in the cosine function of Eq. (30.28) provides for the possibility of both an initial charge and an initial current at time $t = 0$, analogous to an underdamped harmonic oscillator given both an initial displacement and an initial velocity (see Exercise 30.41).

We emphasize once more that the behavior of the L-R-C series circuit is completely analogous to that of the damped harmonic oscillator, Section 14.7. We invite you to verify, for example, that if you start with Eq. (14.41) and substitute q for x, L for m, $1/C$ for k, and R for the damping constant b, the result is Eq. (30.27). Similarly, the cross-over point between underdamping and overdamping occurs at $b^2 = 4km$ for the mechanical system and at $R^2 = 4L/C$ for the electrical one. Can you find still other aspects of this analogy?

The practical applications of the L-R-C series circuit emerge when we include a sinusoidally varying source of emf in the circuit. This is analogous to the *forced oscillations* that we discussed in Section 14.7, and there are analogous *resonance* effects. Such a circuit is called an *alternating-current (ac) circuit*. The analysis of ac circuits is the principal topic of the next chapter.

EXAMPLE 30.10 AN UNDERDAMPED *L-R-C* SERIES CIRCUIT

SOLUTION

What resistance R is required (in terms of L and C) to give an L-R-C series circuit a frequency that is one-half the undamped frequency?

SOLUTION

IDENTIFY and SET UP: This problem concerns an underdamped L-R-C series circuit (Fig. 30.16a). We want just enough resistance to reduce the oscillation frequency to one-half of the undamped value. Equation (30.29) gives the angular frequency ω' of an underdamped L-R-C series circuit; Eq. (30.22) gives the angular frequency ω of an undamped L-C circuit. We use these two equations to solve for R.

EXECUTE: From Eqs. (30.29) and (30.22), the requirement $\omega' = \omega/2$ yields

$$\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = \frac{1}{2}\sqrt{\frac{1}{LC}}$$

We square both sides and solve for R:

$$R = \sqrt{\frac{3L}{C}}$$

For example, adding $35\ \Omega$ to the circuit of Example 30.8 ($L = 10\text{ mH}$, $C = 25\ \mu\text{F}$) would reduce the frequency from 320 Hz to 160 Hz.

EVALUATE: The circuit becomes critically damped with no oscillations when $R = \sqrt{4L/C}$. Our result for R is smaller than that, as it should be; we want the circuit to be *under*damped.

TEST YOUR UNDERSTANDING OF SECTION 30.6 An L-R-C series circuit includes a $2.0\text{-}\Omega$ resistor. At $t = 0$ the capacitor charge is $2.0\ \mu\text{C}$. For which of the following values of the inductance and capacitance will the charge on the capacitor *not* oscillate? (i) $L = 3.0\ \mu\text{H}$, $C = 6.0\ \mu\text{F}$; (ii) $L = 6.0\ \mu\text{H}$, $C = 3.0\ \mu\text{F}$; (iii) $L = 3.0\ \mu\text{H}$, $C = 3.0\ \mu\text{F}$. ❙

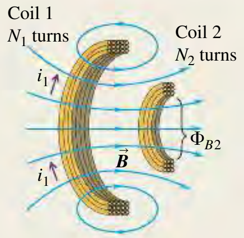
CHAPTER 30 SUMMARY

SOLUTIONS TO ALL EXAMPLES

Mutual inductance: When a changing current i_1 in one circuit causes a changing magnetic flux in a second circuit, an emf $\mathcal{E}_2$ is induced in the second circuit. Likewise, a changing current i_2 in the second circuit induces an emf $\mathcal{E}_1$ in the first circuit. If the circuits are coils of wire with N_1 and N_2 turns, the mutual inductance M can be expressed in terms of the average flux Φ_{B2} through each turn of coil 2 caused by the current i_1 in coil 1, or in terms of the average flux Φ_{B1} through each turn of coil 1 caused by the current i_2 in coil 2. The SI unit of mutual inductance is the henry, abbreviated H. (See Examples 30.1 and 30.2.)

$$\mathcal{E}_2 = -M\frac{di_1}{dt} \quad \text{and} \quad \mathcal{E}_1 = -M\frac{di_2}{dt} \tag{30.4}$$

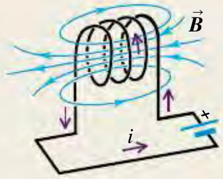
$$M = \frac{N_2\Phi_{B2}}{i_1} = \frac{N_1\Phi_{B1}}{i_2} \tag{30.5}$$



Self-inductance: A changing current i in any circuit causes a self-induced emf $\mathcal{E}$. The inductance (or self-inductance) L depends on the geometry of the circuit and the material surrounding it. The inductance of a coil of N turns is related to the average flux Φ_B through each turn caused by the current i in the coil. An inductor is a circuit device, usually including a coil of wire, intended to have a substantial inductance. (See Examples 30.3 and 30.4.)

$$\mathcal{E} = -L\frac{di}{dt} \tag{30.7}$$

$$L = \frac{N\Phi_B}{i} \tag{30.6}$$

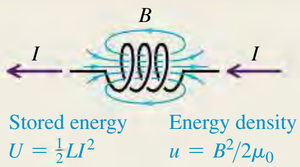


Magnetic-field energy: An inductor with inductance L carrying current I has energy U associated with the inductor's magnetic field. The magnetic energy density u (energy per unit volume) is proportional to the square of the magnetic-field magnitude. (See Example 30.5.)

$$U = \tfrac{1}{2}LI^2 \tag{30.9}$$

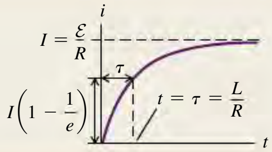
$$u = \frac{B^2}{2\mu_0} \quad \text{(in vacuum)} \tag{30.10}$$

$$u = \frac{B^2}{2\mu} \quad \text{(in a material with magnetic permeability } \mu\text{)} \tag{30.11}$$



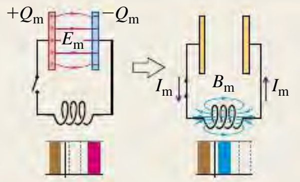
***R-L* circuits:** In a circuit containing a resistor R, an inductor L, and a source of emf, the growth and decay of current are exponential. The time constant τ is the time required for the current to approach within a fraction $1/e$ of its final value. (See Examples 30.6 and 30.7.)

$$\tau = \frac{L}{R} \tag{30.16}$$



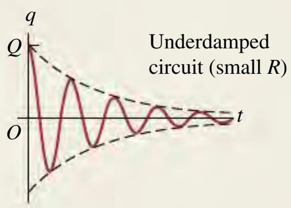
***L-C* circuits:** A circuit that contains inductance L and capacitance C undergoes electrical oscillations with an angular frequency ω that depends on L and C. This is analogous to a mechanical harmonic oscillator, with inductance L analogous to mass m, the reciprocal of capacitance $1/C$ to force constant k, charge q to displacement x, and current i to velocity v_x. (See Examples 30.8 and 30.9.)

$$\omega = \sqrt{\frac{1}{LC}} \tag{30.22}$$



***L-R-C* series circuits:** A circuit that contains inductance, resistance, and capacitance undergoes damped oscillations for sufficiently small resistance. The frequency ω' of damped oscillations depends on the values of L, R, and C. As R increases, the damping increases; if R is greater than a certain value, the behavior becomes overdamped and no longer oscillates. (See Example 30.10.)

$$\omega' = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \tag{30.29}$$



BRIDGING PROBLEM ANALYZING AN L-C CIRCUIT

SOLUTION

An L-C circuit like that shown in Fig. 30.14 consists of a 60.0-mH inductor and a 250-μF capacitor. The initial charge on the capacitor is 6.00 μC, and the initial current in the inductor is 0.400 mA. (a) What is the maximum energy stored in the inductor? (b) What is the maximum current in the inductor? (c) What is the maximum voltage across the capacitor? (d) When the current in the inductor has half its maximum value, what are the energy stored in the inductor and the voltage across the capacitor?

SOLUTION GUIDE

IDENTIFY and SET UP

1. An L-C circuit is a conservative system—there is no resistance to dissipate energy. The energy oscillates between electrical energy in the capacitor and magnetic energy stored in the inductor.
2. Oscillations in an L-C circuit are analogous to the mechanical oscillations of a particle at the end of an ideal spring (see Table 30.1). Compare this problem to the analogous mechanical problem (see Example 14.3 in Section 14.2 and Example 14.4 in Section 14.3).
3. Which key equations are needed to describe the capacitor? To describe the inductor?

EXECUTE

4. Find the initial total energy in the circuit. Use it to determine the maximum energy stored in the inductor during the oscillation.
5. Use the result of step 4 to find the maximum current in the inductor.
6. Use the result of step 4 to find the maximum energy stored in the capacitor during the oscillation. Then use this to find the maximum capacitor voltage.
7. Find the energy in the inductor and the capacitor charge when the current has half the value that you found in step 5.

EVALUATE

8. Initially, what fraction of the total energy is in the inductor? Is it possible to tell whether this is initially increasing or decreasing?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®. MP

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q30.1 In an electric trolley or bus system, the vehicle's motor draws current from an overhead wire by means of a long arm with an attachment at the end that slides along the overhead wire. A brilliant electric spark is often seen when the attachment crosses a junction in the wires where contact is momentarily lost. Explain this phenomenon.

Q30.2 From Eq. (30.5) $1\text{ H} = 1\text{ Wb/A}$, and from Eqs. (30.4) $1\text{ H} = 1\ \Omega \cdot \text{s}$. Show that these two definitions are equivalent.

Q30.3 In Fig. 30.1, if coil 2 is turned 90° so that its axis is vertical, does the mutual inductance increase or decrease? Explain.

Q30.4 The tightly wound toroidal solenoid is one of the few configurations for which it is easy to calculate self-inductance. What features of the toroidal solenoid give it this simplicity?

Q30.5 Two identical, closely wound, circular coils, each having self-inductance L, are placed next to each other, so that they are coaxial and almost touching. If they are connected in series, what is the self-inductance of the combination? What if they are connected in parallel? Can they be connected so that the total inductance is zero? Explain.

Q30.6 Two closely wound circular coils have the same number of turns, but one has twice the radius of the other. How are the self-inductances of the two coils related? Explain your reasoning.

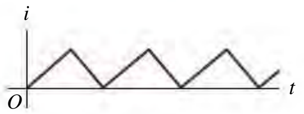
Q30.7 You are to make a resistor by winding a wire around a cylindrical form. To make the inductance as small as possible, it is proposed that you wind half the wire in one direction and the other half in the opposite direction. Would this achieve the desired result? Why or why not?

Q30.8 For the same magnetic field strength B, is the energy density greater in vacuum or in a magnetic material? Explain. Does Eq. (30.11) imply that for a long solenoid in which the current is I the energy stored is proportional to $1/\mu$? And does this mean that for the same current less energy is stored when the solenoid is filled with a ferromagnetic material rather than with air? Explain.

Q30.9 In an R-C circuit, a resistor, an uncharged capacitor, a dc battery, and an open switch are in series. In an R-L circuit, a resistor, an inductor, a dc battery, and an open switch are in series. Compare the behavior of the current in these circuits (a) just after the switch is closed and (b) long after the switch has been closed. In other words, compare the way in which a capacitor and an inductor affect a circuit.

Q30.10 A Differentiating Circuit. The current in a resistanceless inductor is caused to vary with time as shown in the graph of **Fig. Q30.10**. (a) Sketch the pattern that would be observed on the screen of an oscilloscope connected to the terminals of the inductor. (The oscilloscope spot sweeps horizontally across the screen at a constant speed, and its vertical deflection is proportional to the potential difference between the inductor terminals.) (b) Explain why a circuit with an inductor can be described as a "differentiating circuit."

Figure **Q30.10**



Q30.11 In Section 30.5 Kirchhoff's loop rule is applied to an L-C circuit where the capacitor is initially fully charged and the equation $-L\,(di/dt) - (q/C) = 0$ is derived. But as the capacitor starts to discharge, the current increases from zero. The equation says $L\,di/dt = -q/C$, so it says $L\,di/dt$ is negative. Explain how $L\,di/dt$ can be negative when the current is increasing.

Q30.12 In Section 30.5 the relationship $i = dq/dt$ is used in deriving Eq. (30.20). But a flow of current corresponds to a decrease in the charge on the capacitor. Explain, therefore, why this is the correct equation to use in the derivation, rather than $i = -dq/dt$.

Q30.13 In the R-L circuit shown in Fig. 30.11, when switch S_1 is closed, the potential v_{ac} changes suddenly and discontinuously, but the current does not. Explain why the voltage can change suddenly but the current can't.

Q30.14 In the R-L circuit shown in Fig. 30.11, is the current in the resistor always the same as the current in the inductor? How do you know?

Q30.15 Suppose there is a steady current in an inductor. If you attempt to reduce the current to zero instantaneously by quickly opening a switch, an arc can appear at the switch contacts. Why? Is it physically possible to stop the current instantaneously? Explain.

Q30.16 In an L-R-C series circuit, what criteria could be used to decide whether the system is overdamped or underdamped? For example, could we compare the maximum energy stored during one cycle to the energy dissipated during one cycle? Explain.

EXERCISES

Section 30.1 Mutual Inductance

30.1 • Two coils have mutual inductance $M = 3.25 \times 10^{-4}$ H. The current i_1 in the first coil increases at a uniform rate of 830 A/s. (a) What is the magnitude of the induced emf in the second coil? Is it constant? (b) Suppose that the current described is in the second coil rather than the first. What is the magnitude of the induced emf in the first coil?

30.2 • Two coils are wound around the same cylindrical form, like the coils in Example 30.1. When the current in the first coil is decreasing at a rate of -0.242 A/s, the induced emf in the second coil has magnitude 1.65×10^{-3} V. (a) What is the mutual inductance of the pair of coils? (b) If the second coil has 25 turns, what is the flux through each turn when the current in the first coil equals 1.20 A? (c) If the current in the second coil increases at a rate of 0.360 A/s, what is the magnitude of the induced emf in the first coil?

30.3 • A 10.0-cm-long solenoid of diameter 0.400 cm is wound uniformly with 800 turns. A second coil with 50 turns is wound around the solenoid at its center. What is the mutual inductance of the combination of the two coils?

30.4 • A solenoidal coil with 25 turns of wire is wound tightly around another coil with 300 turns (see Example 30.1). The inner solenoid is 25.0 cm long and has a diameter of 2.00 cm. At a certain time, the current in the inner solenoid is 0.120 A and is increasing at a rate of 1.75×10^3 A/s. For this time, calculate: (a) the average magnetic flux through each turn of the inner solenoid; (b) the mutual inductance of the two solenoids; (c) the emf induced in the outer solenoid by the changing current in the inner solenoid.

30.5 • Two toroidal solenoids are wound around the same form so that the magnetic field of one passes through the turns of the other. Solenoid 1 has 700 turns, and solenoid 2 has 400 turns. When the current in solenoid 1 is 6.52 A, the average flux through each turn of solenoid 2 is 0.0320 Wb. (a) What is the mutual inductance of the pair of solenoids? (b) When the current in solenoid 2 is 2.54 A, what is the average flux through each turn of solenoid 1?

30.6 •• A toroidal solenoid with mean radius r and cross-sectional area A is wound uniformly with N_1 turns. A second toroidal solenoid with N_2 turns is wound uniformly on top of the first, so that the two solenoids have the same cross-sectional area and mean radius. (a) What is the mutual inductance of the two solenoids? Assume that the magnetic field of the first solenoid is uniform across the cross section of the two solenoids. (b) If $N_1 = 500$ turns, $N_2 = 300$ turns, $r = 10.0$ cm, and $A = 0.800$ cm^2, what is the value of the mutual inductance?

Section 30.2 Self-Inductance and Inductors

30.7 • A 2.50-mH toroidal solenoid has an average radius of 6.00 cm and a cross-sectional area of 2.00 cm^2. (a) How many coils does it have? (Make the same assumption as in Example 30.3.) (b) At what rate must the current through it change so that a potential difference of 2.00 V is developed across its ends?

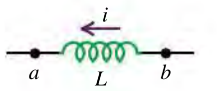
30.8 • A toroidal solenoid has 500 turns, cross-sectional area 6.25 cm^2, and mean radius 4.00 cm. (a) Calculate the coil's self-inductance. (b) If the current decreases uniformly from 5.00 A to 2.00 A in 3.00 ms, calculate the self-induced emf in the coil. (c) The current is directed from terminal a of the coil to terminal b. Is the direction of the induced emf from a to b or from b to a?

30.9 • At the instant when the current in an inductor is increasing at a rate of 0.0640 A/s, the magnitude of the self-induced emf is 0.0160 V. (a) What is the inductance of the inductor? (b) If the inductor is a solenoid with 400 turns, what is the average magnetic flux through each turn when the current is 0.720 A?

30.10 •• When the current in a toroidal solenoid is changing at a rate of 0.0260 A/s, the magnitude of the induced emf is 12.6 mV. When the current equals 1.40 A, the average flux through each turn of the solenoid is 0.00285 Wb. How many turns does the solenoid have?

30.11 • The inductor in **Fig. E30.11** has inductance 0.260 H and carries a current in the direction shown that is decreasing at a uniform rate, $di/dt = -0.0180$ A/s. (a) Find the self-induced emf. (b) Which end of the inductor, a or b, is at a higher potential?

Figure **E30.11**



30.12 • The inductor shown in Fig. E30.11 has inductance 0.260 H and carries a current in the direction shown. The current is changing at a constant rate. (a) The potential between points a and b is $V_{ab} = 1.04$ V, with point a at higher potential. Is the current increasing or decreasing? (b) If the current at $t = 0$ is 12.0 A, what is the current at $t = 2.00$ s?

30.13 •• A toroidal solenoid has mean radius 12.0 cm and cross-sectional area 0.600 cm^2. (a) How many turns does the solenoid have if its inductance is 0.100 mH? (b) What is the resistance of the solenoid if the wire from which it is wound has a resistance per unit length of 0.0760 Ω/m?

30.14 • A long, straight solenoid has 800 turns. When the current in the solenoid is 2.90 A, the average flux through each turn of the solenoid is 3.25×10^{-3} Wb. What must be the magnitude of the rate of change of the current in order for the self-induced emf to equal 6.20 mV?

30.15 •• **Inductance of a Solenoid.** (a) A long, straight solenoid has N turns, uniform cross-sectional area A, and length l. Show that the inductance of this solenoid is given by the equation $L = \mu_0 AN^2/l$. Assume that the magnetic field is uniform inside the solenoid and zero outside. (Your answer is approximate because B is actually smaller at the ends than at the center. For this reason, your answer is actually an upper limit on the inductance.) (b) A metallic laboratory spring is typically 5.00 cm long and 0.150 cm in diameter and has 50 coils. If you connect such a spring in an electric circuit, how much self-inductance must you include for it if you model it as an ideal solenoid?

Section 30.3 Magnetic-Field Energy

30.16 • An inductor used in a dc power supply has an inductance of 12.0 H and a resistance of 180 Ω. It carries a current of 0.500 A. (a) What is the energy stored in the magnetic field? (b) At what rate is thermal energy developed in the inductor? (c) Does your answer to part (b) mean that the magnetic-field energy is decreasing with time? Explain.

30.17 • An air-filled toroidal solenoid has a mean radius of 15.0 cm and a cross-sectional area of 5.00 cm^2. When the current is 12.0 A, the energy stored is 0.390 J. How many turns does the winding have?

30.18 • An air-filled toroidal solenoid has 300 turns of wire, a mean radius of 12.0 cm, and a cross-sectional area of 4.00 cm^2. If the current is 5.00 A, calculate: (a) the magnetic field in the solenoid; (b) the self-inductance of the solenoid; (c) the energy stored in the magnetic field; (d) the energy density in the magnetic field. (e) Check your answer for part (d) by dividing your answer to part (c) by the volume of the solenoid.

30.19 •• A solenoid 25.0 cm long and with a cross-sectional area of 0.500 cm^2 contains 400 turns of wire and carries a current of 80.0 A. Calculate: (a) the magnetic field in the solenoid; (b) the energy density in the magnetic field if the solenoid is filled with air; (c) the total energy contained in the coil's magnetic field (assume the field is uniform); (d) the inductance of the solenoid.

30.20 • It has been proposed to use large inductors as energy storage devices. (a) How much electrical energy is converted to light and thermal energy by a 150-W light bulb in one day? (b) If the amount of energy calculated in part (a) is stored in an inductor in which the current is 80.0 A, what is the inductance?

30.21 •• In a proton accelerator used in elementary particle physics experiments, the trajectories of protons are controlled by bending magnets that produce a magnetic field of 4.80 T. What is the magnetic-field energy in a 10.0-cm^3 volume of space where $B = 4.80$ T?

30.22 • It is proposed to store $1.00\ \text{kW}\cdot\text{h} = 3.60 \times 10^6$ J of electrical energy in a uniform magnetic field with magnitude 0.600 T. (a) What volume (in vacuum) must the magnetic field occupy to store this amount of energy? (b) If instead this amount of energy is to be stored in a volume (in vacuum) equivalent to a cube 40.0 cm on a side, what magnetic field is required?

Section 30.4 The *R-L* Circuit

30.23 • An inductor with an inductance of 2.50 H and a resistance of 8.00 Ω is connected to the terminals of a battery with an emf of 6.00 V and negligible internal resistance. 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 0.500 A; (c) the current 0.250 s after the circuit is closed; (d) the final steady-state current.

30.24 • In Fig. 30.11, $R = 15.0\ \Omega$ and the battery emf is 6.30 V. With switch S_2 open, switch S_1 is closed. After several minutes, S_1 is opened and S_2 is closed. (a) At 2.00 ms after S_1 is opened, the current has decayed to 0.280 A. Calculate the inductance of the coil. (b) How long after S_1 is opened will the current reach 1.00% of its original value?

30.25 • A 35.0-V battery with negligible internal resistance, a 50.0-Ω resistor, and a 1.25-mH inductor with negligible resistance are all connected in series with an open switch. The switch is suddenly closed. (a) How long after closing the switch will the current through the inductor reach one-half of its maximum value? (b) How long after closing the switch will the energy stored in the inductor reach one-half of its maximum value?

30.26 • In Fig. 30.11, switch S_1 is closed while switch S_2 is kept open. The inductance is $L = 0.115$ H, and the resistance is $R = 120\ \Omega$. (a) When the current has reached its final value, the energy stored in the inductor is 0.260 J. What is the emf $\mathcal{E}$ of the battery? (b) After the current has reached its final value, S_1 is opened and S_2 is closed. How much time does it take for the energy stored in the inductor to decrease to 0.130 J, half of the original value?

30.27 • In Fig. 30.11, suppose that $\mathcal{E} = 60.0$ V, $R = 240\ \Omega$, and $L = 0.160$ H. With switch S_2 open, switch S_1 is left closed until a constant current is established. Then S_2 is closed and S_1 opened, taking the battery out of the circuit. (a) What is the initial current in the resistor, just after S_2 is closed and S_1 is opened? (b) What is the current in the resistor at $t = 4.00 \times 10^{-4}$ s? (c) What is the potential difference between points b and c at $t = 4.00 \times 10^{-4}$ s? Which point is at a higher potential? (d) How long does it take the current to decrease to half its initial value?

30.28 • In Fig. 30.11, suppose that $\mathcal{E} = 60.0$ V, $R = 240\ \Omega$, and $L = 0.160$ H. Initially there is no current in the circuit. Switch S_2 is left open, and switch S_1 is closed. (a) Just after S_1 is closed, what are the potential differences v_{ab} and v_{bc}? (b) A long time (many time constants) after S_1 is closed, what are v_{ab} and v_{bc}? (c) What are v_{ab} and v_{bc} at an intermediate time when $i = 0.150$ A?

30.29 • In Fig. 30.11 switch S_1 is closed while switch S_2 is kept open. The inductance is $L = 0.380$ H, the resistance is $R = 48.0\ \Omega$, and the emf of the battery is 18.0 V. At time t after S_1 is closed, the current in the circuit is increasing at a rate of $di/dt = 7.20$ A/s. At this instant what is v_{ab}, the voltage across the resistor?

30.30 •• Consider the circuit in Exercise 30.23. (a) Just after the circuit is completed, at what rate is the battery supplying electrical energy to the circuit? (b) When the current has reached its final steady-state value, how much energy is stored in the inductor? What is the rate at which electrical energy is being dissipated in the resistance of the inductor? What is the rate at which the battery is supplying electrical energy to the circuit?

Section 30.5 The *L-C* Circuit

30.31 • In an L-C circuit, $L = 85.0$ mH and $C = 3.20\ \mu$F. During the oscillations the maximum current in the inductor is 0.850 mA. (a) What is the maximum charge on the capacitor? (b) What is the magnitude of the charge on the capacitor at an instant when the current in the inductor has magnitude 0.500 mA?

30.32 •• A 15.0-μF capacitor is charged by a 150.0-V power supply, then disconnected from the power and connected in series with a 0.280-mH inductor. Calculate: (a) the oscillation frequency of the circuit; (b) the energy stored in the capacitor at time $t = 0$ ms (the moment of connection with the inductor); (c) the energy stored in the inductor at $t = 1.30$ ms.

30.33 • A 7.50-nF capacitor is charged up to 12.0 V, then disconnected from the power supply and connected in series through a coil. The period of oscillation of the circuit is then measured to be 8.60×10^{-5} s. Calculate: (a) the inductance of the coil; (b) the maximum charge on the capacitor; (c) the total energy of the circuit; (d) the maximum current in the circuit.

30.34 •• A 18.0-μF capacitor is placed across a 22.5-V battery for several seconds and is then connected across a 12.0-mH inductor that has no appreciable resistance. (a) After the capacitor and inductor are connected together, find the maximum current in the circuit. When the current is a maximum, what is the charge on the capacitor? (b) How long after the capacitor and inductor are connected together does it take for the capacitor to be completely discharged for the first time? For the second time? (c) Sketch graphs of the charge on the capacitor plates and the current through the inductor as functions of time.

30.35 • ***L-C* Oscillations.** A capacitor with capacitance 6.00×10^{-5} F is charged by connecting it to a 12.0-V battery. The capacitor is disconnected from the battery and connected across an inductor with $L = 1.50$ H. (a) What are the angular frequency ω of the electrical oscillations and the period of these oscillations (the time for one oscillation)? (b) What is the initial charge on the capacitor? (c) How much energy is initially stored in the capacitor? (d) What is the charge on the capacitor 0.0230 s after the connection to the inductor is made? Interpret the sign of your answer. (e) At the time given in part (d), what is the current in the inductor? Interpret the sign of your answer. (f) At the time given in part (d), how much electrical energy is stored in the capacitor and how much is stored in the inductor?

30.36 • **A Radio Tuning Circuit.** The minimum capacitance of a variable capacitor in a radio is 4.18 pF. (a) What is the inductance of a coil connected to this capacitor if the oscillation frequency of the *L-C* circuit is 1600×10^3 Hz, corresponding to one end of the AM radio broadcast band, when the capacitor is set to its minimum capacitance? (b) The frequency at the other end of the broadcast band is 540×10^3 Hz. What is the maximum capacitance of the capacitor if the oscillation frequency is adjustable over the range of the broadcast band?

30.37 •• An *L-C* circuit containing an 80.0-mH inductor and a 1.25-nF capacitor oscillates with a maximum current of 0.750 A. Calculate: (a) the maximum charge on the capacitor and (b) the oscillation frequency of the circuit. (c) Assuming the capacitor had its maximum charge at time $t = 0$, calculate the energy stored in the inductor after 2.50 ms of oscillation.

Section 30.6 The *L-R-C* Series Circuit

30.38 • An *L-R-C* series circuit has $L = 0.600$ H and $C = 3.00\ \mu$F. (a) Calculate the angular frequency of oscillation for the circuit when $R = 0$. (b) What value of R gives critical damping? (c) What is the oscillation frequency ω' when R has half of the value that produces critical damping?

30.39 • An *L-R-C* series circuit has $L = 0.450$ H, $C = 2.50 \times 10^{-5}$ F, and resistance R. (a) What is the angular frequency of the circuit when $R = 0$? (b) What value must R have to give a 5.0% decrease in angular frequency compared to the value calculated in part (a)?

30.40 • An *L-R-C* series circuit has $L = 0.400$ H, $C = 7.00\ \mu$F, and $R = 320\ \Omega$. At $t = 0$ the current is zero and the initial charge on the capacitor is 2.80×10^{-4} C. (a) What are the values of the constants A and ϕ in Eq. (30.28)? (b) How much time does it take for each complete current oscillation after the switch in this circuit is closed? (c) What is the charge on the capacitor after the first complete current oscillation?

30.41 • For the circuit of Fig. 30.17, let $C = 15.0$ nF, $L = 22$ mH, and $R = 75.0\ \Omega$. (a) Calculate the oscillation frequency of the circuit once the capacitor has been charged and the switch has been connected to point a. (b) How long will it take for the amplitude of the oscillation to decay to 10.0% of its original value? (c) What value of R would result in a critically damped circuit?

PROBLEMS

30.42 • An inductor is connected to the terminals of a battery that has an emf of 16.0 V and negligible internal resistance. The current is 4.86 mA at 0.940 ms after the connection is completed. After a long time, the current is 6.45 mA. What are (a) the resistance R of the inductor and (b) the inductance L of the inductor?

30.43 • One solenoid is centered inside another. The outer one has a length of 50.0 cm and contains 6750 coils, while the coaxial inner solenoid is 3.0 cm long and 0.120 cm in diameter and contains 15 coils. The current in the outer solenoid is changing at 49.2 A/s. (a) What is the mutual inductance of these solenoids? (b) Find the emf induced in the inner solenoid.

30.44 •• CALC A coil has 400 turns and self-inductance 7.50 mH. The current in the coil varies with time according to $i = (680\ \text{mA})\cos(\pi t/0.0250\ \text{s})$. (a) What is the maximum emf induced in the coil? (b) What is the maximum average flux through each turn of the coil? (c) At $t = 0.0180$ s, what is the magnitude of the induced emf?

30.45 •• **Solar Magnetic Energy.** Magnetic fields within a sunspot can be as strong as 0.4 T. (By comparison, the earth's magnetic field is about 1/10,000 as strong.) Sunspots can be as large as 25,000 km in radius. The material in a sunspot has a density of about 3×10^{-4} kg/m^3. Assume μ for the sunspot material is μ_0. If 100% of the magnetic-field energy stored in a sunspot could be used to eject the sunspot's material away from the sun's surface, at what speed would that material be ejected? Compare to the sun's escape speed, which is about 6×10^5 m/s. (*Hint:* Calculate the kinetic energy the magnetic field could supply to 1 m^3 of sunspot material.)

30.46 •• CP CALC **A Coaxial Cable.** A small solid conductor with radius a is supported by insulating, nonmagnetic disks on the axis of a thin-walled tube with inner radius b. The inner and outer conductors carry equal currents i in opposite directions. (a) Use Ampere's law to find the magnetic field at any point in the volume between the conductors. (b) Write the expression for the flux $d\Phi_B$ through a narrow strip of length l parallel to the axis, of width dr, at a distance r from the axis of the cable and lying in a plane containing the axis. (c) Integrate your expression from part (b) over the volume between the two conductors to find the total flux produced by a current i in the central conductor. (d) Show that the inductance of a length l of the cable is

$$L = l\frac{\mu_0}{2\pi}\ln\left(\frac{b}{a}\right)$$

(e) Use Eq. (30.9) to calculate the energy stored in the magnetic field for a length l of the cable.

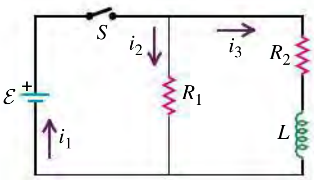
30.47 •• **CP CALC** Consider the coaxial cable of Problem 30.46. The conductors carry equal currents i in opposite directions. (a) Use Ampere's law to find the magnetic field at any point in the volume between the conductors. (b) Use the energy density for a magnetic field, Eq. (30.10), to calculate the energy stored in a thin, cylindrical shell between the two conductors. Let the cylindrical shell have inner radius r, outer radius $r + dr$, and length l. (c) Integrate your result in part (b) over the volume between the two conductors to find the total energy stored in the magnetic field for a length l of the cable. (d) Use your result in part (c) and Eq. (30.9) to calculate the inductance L of a length l of the cable. Compare your result to L calculated in part (d) of Problem 30.46.

30.48 •• **CALC** Consider the circuit in Fig. 30.11 with both switches open. At $t = 0$ switch S_1 is closed while switch S_2 is left open. (a) Use Eq. (30.14) to derive an equation for the rate P_R at which electrical energy is being consumed in the resistor. In terms of $\mathcal{E}$, R, and L, at what value of t is P_R a maximum? What is that maximum value? (b) Use Eqs. (30.14) and (30.15) to derive an equation for P_L, the rate at which energy is being stored in the inductor. (c) What is P_L at $t = 0$ and as $t \to \infty$? (d) In terms of $\mathcal{E}$, R, and L, at what value of t is P_L a maximum? What is that maximum value? (e) Obtain an expression for $P_{\mathcal{E}}$, the rate at which the battery is supplying electrical energy to the circuit. In terms of $\mathcal{E}$, R, and L, at what value of t is $P_{\mathcal{E}}$ a maximum? What is that maximum value?

30.49 • (a) What would have to be the self-inductance of a solenoid for it to store 10.0 J of energy when a 2.00-A current runs through it? (b) If this solenoid's cross-sectional diameter is 4.00 cm, and if you could wrap its coils to a density of 10 coils/mm, how long would the solenoid be? (See Exercise 30.15.) Is this a realistic length for ordinary laboratory use?

30.50 •• **CALC** An inductor with inductance $L = 0.300$ H and negligible resistance is connected to a battery, a switch S, and two resistors, $R_1 = 12.0\ \Omega$ and $R_2 = 16.0\ \Omega$ (**Fig. P30.50**). The battery has emf 96.0 V and negligible internal resistance. S is closed at $t = 0$. (a) What are the currents i_1, i_2, and i_3 just after S is closed? (b) What are i_1, i_2, and i_3 after S has been closed a long time? (c) What is the value of t for which i_3 has half of the final value that you calculated in part (b)? (d) When i_3 has half of its final value, what are i_1 and i_2?

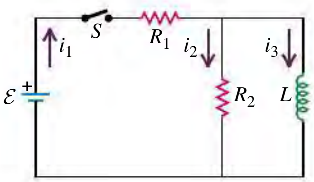
Figure **P30.50**



30.51 • **An Electromagnetic Car Alarm.** Your latest invention is a car alarm that produces sound at a particularly annoying frequency of 3500 Hz. To do this, the car-alarm circuitry must produce an alternating electric current of the same frequency. That's why your design includes an inductor and a capacitor in series. The maximum voltage across the capacitor is to be 12.0 V. To produce a sufficiently loud sound, the capacitor must store 0.0160 J of energy. What values of capacitance and inductance should you choose for your car-alarm circuit?

30.52 ••• **CALC** An inductor with inductance $L = 0.200$ H and negligible resistance is connected to a battery, a switch S, and two resistors, $R_1 = 8.00\ \Omega$ and $R_2 = 6.00\ \Omega$ (**Fig. P30.52**). The battery has emf 48.0 V and negligible internal resistance. S is closed at $t = 0$. (a) What are the currents i_1, i_2, and i_3 just after S is closed? (b) What are i_1, i_2, and i_3 after S has been closed a long time? (c) Apply Kirchhoff's rules to the circuit and obtain a differential equation for $i_3(t)$. Integrate this equation to obtain an equation for i_3 as a function of the time t that has elapsed since S was closed. (d) Use the equation that you derived in part (c) to calculate the value of t for which i_3 has half of the final value that you calculated in part (b). (e) When i_3 has half of its final value, what are i_1 and i_2?

Figure **P30.52**



30.53 • A 7.00-μF capacitor is initially charged to a potential of 16.0 V. It is then connected in series with a 3.75-mH inductor. (a) What is the total energy stored in this circuit? (b) What is the maximum current in the inductor? What is the charge on the capacitor plates at the instant the current in the inductor is maximal?

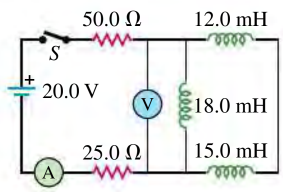
30.54 •• A 6.40-nF capacitor is charged to 24.0 V and then disconnected from the battery in the circuit and connected in series with a coil that has $L = 0.0660$ H and negligible resistance. After the circuit has been completed, there are current oscillations. (a) At an instant when the charge of the capacitor is 0.0800 μC, how much energy is stored in the capacitor and in the inductor, and what is the current in the inductor? (b) At the instant when the charge on the capacitor is 0.0800 μC, what are the voltages across the capacitor and across the inductor, and what is the rate at which current in the inductor is changing?

30.55 • An L-C circuit consists of a 60.0-mH inductor and a 250-μF capacitor. The initial charge on the capacitor is 6.00 μC, and the initial current in the inductor is zero. (a) What is the maximum voltage across the capacitor? (b) What is the maximum current in the inductor? (c) What is the maximum energy stored in the inductor? (d) When the current in the inductor has half its maximum value, what is the charge on the capacitor and what is the energy stored in the inductor?

30.56 •• A charged capacitor with $C = 590\ \mu$F is connected in series to an inductor that has $L = 0.330$ H and negligible resistance. At an instant when the current in the inductor is $i = 2.50$ A, the current is increasing at a rate of $di/dt = 73.0$ A/s. During the current oscillations, what is the maximum voltage across the capacitor?

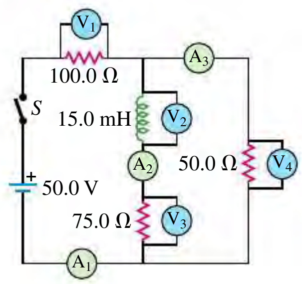
30.57 ••• **CP** In the circuit shown in **Fig. P30.57**, the switch S has been open for a long time and is suddenly closed. Neither the battery nor the inductors have any appreciable resistance. What do the ammeter and the voltmeter read (a) just after S is closed; (b) after S has been closed a very long time; (c) 0.115 ms after S is closed?

Figure **P30.57**



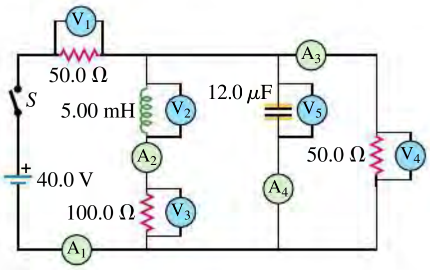
30.58 •• **CP** In the circuit shown in **Fig. P30.58**, find the reading in each ammeter and voltmeter (a) just after switch S is closed and (b) after S has been closed a very long time.

Figure **P30.58**



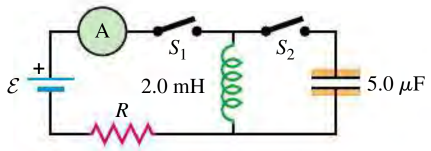
30.59 •• **CP** In the circuit shown in **Fig. P30.59**, switch S is closed at time $t = 0$ with no charge initially on the capacitor. (a) Find the reading of each ammeter and each voltmeter just after S is closed. (b) Find the reading of each meter after a long time has elapsed. (c) Find the maximum charge on the capacitor. (d) Draw a qualitative graph of the reading of voltmeter V_2 as a function of time.

Figure **P30.59**



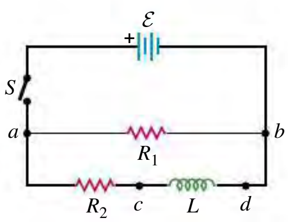
30.60 •• In the circuit shown in **Fig. P30.60**, switch S_1 has been closed for a long enough time so that the current reads a steady 3.50 A. Suddenly, switch S_2 is closed and S_1 is opened at the same instant. (a) What is the maximum charge that the capacitor will receive? (b) What is the current in the inductor at this time?

Figure **P30.60**



30.61 •• **CP** In the circuit shown in **Fig. P30.61**, $\mathcal{E} = 60.0$ V, $R_1 = 40.0\ \Omega$, $R_2 = 25.0\ \Omega$, and $L = 0.300$ H. Switch S is closed at $t = 0$. Just after the switch is closed, (a) what is the potential difference v_{ab} across the resistor R_1; (b) which point, a or b, is at a higher potential; (c) what is the potential difference v_{cd} across the inductor L; (d) which point, c or d, is at a higher potential? The switch is left closed a long time and then opened. Just after the switch is opened, (e) what is the potential difference v_{ab} across the resistor R_1; (f) which point, a or b, is at a higher potential; (g) what is the potential difference v_{cd} across the inductor L; (h) which point, c or d, is at a higher potential?

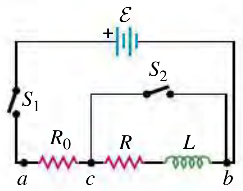
Figure **P30.61**



30.62 •• **CP** In the circuit shown in Fig. P30.61, $\mathcal{E} = 60.0$ V, $R_1 = 40.0\ \Omega$, $R_2 = 25.0\ \Omega$, and $L = 0.300$ H. (a) Switch S is closed. At some time t afterward, the current in the inductor is increasing at a rate of $di/dt = 50.0$ A/s. At this instant, what are the current i_1 through R_1 and the current i_2 through R_2? (*Hint:* Analyze two separate loops: one containing $\mathcal{E}$ and R_1 and the other containing $\mathcal{E}$, R_2, and L.) (b) After the switch has been closed a long time, it is opened again. Just after it is opened, what is the current through R_1?

30.63 •• **CALC** Consider the circuit shown in **Fig. P30.63**. Let $\mathcal{E} = 36.0$ V, $R_0 = 50.0\ \Omega$, $R = 150\ \Omega$, and $L = 4.00$ H. (a) Switch S_1 is closed and switch S_2 is left open. Just after S_1 is closed, what are the current i_0 through R_0 and the potential differences v_{ac} and v_{cb}? (b) After S_1 has been closed a long time (S_2 is still open) so that the current has reached its final, steady value, what are i_0, v_{ac}, and v_{cb}? (c) Find the expressions for i_0, v_{ac}, and v_{cb} as functions of the time t since S_1 was closed. Your results should agree with part (a) when $t = 0$ and with part (b) when $t \rightarrow \infty$. Graph i_0, v_{ac}, and v_{cb} versus time.

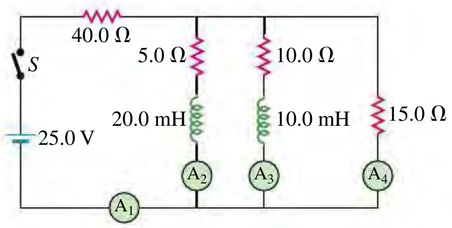
Figure **P30.63**



30.64 •• After the current in the circuit of Fig. P30.63 has reached its final, steady value with switch S_1 closed and S_2 open, switch S_2 is closed, thus short-circuiting the inductor. (Switch S_1 remains closed. See Problem 30.63 for numerical values of the circuit elements.) (a) Just after S_2 is closed, what are v_{ac} and v_{cb}, and what are the currents through R_0, R, and S_2? (b) A long time after S_2 is closed, what are v_{ac} and v_{cb}, and what are the currents through R_0, R, and S_2? (c) Derive expressions for the currents through R_0, R, and S_2 as functions of the time t that has elapsed since S_2 was closed. Your results should agree with part (a) when $t = 0$ and with part (b) when $t \rightarrow \infty$. Graph these three currents versus time.

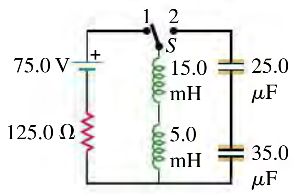
30.65 •• **CP** In the circuit shown in **Fig. P30.65**, switch S is closed at time $t = 0$. (a) Find the reading of each meter just after S is closed. (b) What does each meter read long after S is closed?

Figure **P30.65**



30.66 ••• **CP** In the circuit shown in **Fig. P30.66**, neither the battery nor the inductors have any appreciable resistance, the capacitors are initially uncharged, and the switch S has been in position 1 for a very long time. (a) What is the current in the circuit? (b) The switch is now suddenly flipped to position 2. Find the maximum charge that each capacitor will receive, and how much time after the switch is flipped it will take them to acquire this charge.

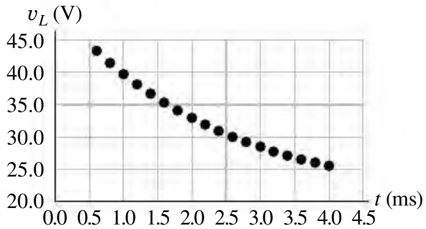
Figure **P30.66**



30.67 •• **DATA** During a summer internship as an electronics technician, you are asked to measure the self-inductance L of a solenoid. You connect the solenoid in series with a 10.0-Ω resistor, a battery that has negligible internal resistance, and a switch. Using an ideal voltmeter, you measure and digitally record the voltage v_L across the solenoid as a function of the time t that has elapsed since the switch is closed. Your measured values are shown in **Fig. P30.67**, where v_L is plotted versus t. In addition, you measure that $v_L = 50.0$ V just after the switch is closed and $v_L = 20.0$ V a long time after it is closed. (a) Apply the loop rule to the circuit and obtain an equation for v_L as a function of t. [*Hint*: Use an analysis similar to that used to derive Eq. (30.15).]

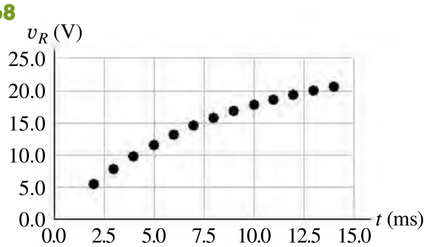
(b) What is the emf $\mathcal{E}$ of the battery? (c) According to your measurements, what is the voltage amplitude across the 10.0-Ω resistor as $t \to \infty$? Use this result to calculate the current in the circuit as $t \to \infty$. (d) What is the resistance R_L of the solenoid? (e) Use the theoretical equation from part (a), Fig. P30.67, and the values of $\mathcal{E}$ and R_L from parts (b) and (d) to calculate L. (*Hint:* According to the equation, what is v_L when $t = \tau$, one time constant? Use Fig. P30.67 to estimate the value of $t = \tau$.)

Figure **P30.67**



30.68 •• **DATA** You are studying a solenoid of unknown resistance and inductance. You connect it in series with a 50.0-Ω resistor, a 25.0-V battery that has negligible internal resistance, and a switch. Using an ideal voltmeter, you measure and digitally record the voltage v_R across the resistor as a function of the time t that has elapsed after the switch is closed. Your measured values are shown in **Fig. P30.68**, where v_R is plotted versus t. In addition, you measure that $v_R = 0$ just after the switch is closed and $v_R = 25.0$ V a long time after it is closed. (a) What is the resistance R_L of the solenoid? (b) Apply the loop rule to the circuit and obtain an equation for v_R as a function of t. (c) According to the equation that you derived in part (b), what is v_R when $t = \tau$, one time constant? Use Fig. P30.68 to estimate the value of $t = \tau$. What is the inductance of the solenoid? (d) How much energy is stored in the inductor a long time after the switch is closed?

Figure **P30.68**



30.69 •• **DATA** To investigate the properties of a large industrial solenoid, you connect the solenoid and a resistor in series with a battery. Switches allow the battery to be replaced by a short circuit across the solenoid and resistor. Therefore Fig. 30.11 applies, with $R = R_{ext} + R_L$, where R_L is the resistance of the solenoid and R_{ext} is the resistance of the series resistor. With switch S_2 open, you close switch S_1 and keep it closed until the current i in the solenoid is constant (Fig. 30.11). Then you close S_2 and open S_1 simultaneously, using a rapid-response switching mechanism. With high-speed electronics you measure the time t_{half} that it takes for the current to decrease to half of its initial value. You repeat this measurement for several values of R_{ext} and obtain these results:

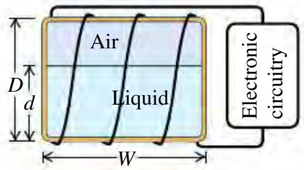
R_{ext} (Ω)	3.0	4.0	5.0	6.0	7.0	8.0	10.0	12.0
t_{half} (s)	0.735	0.654	0.589	0.536	0.491	0.453	0.393	0.347

(a) Graph your data in the form of $1/t_{half}$ versus R_{ext}. Explain why the data points plotted this way fall close to a straight line. (b) Use your graph from part (a) to calculate the resistance R_L and inductance L of the solenoid. (c) If the current in the solenoid is 20.0 A, how much energy is stored there? At what rate is electrical energy being dissipated in the resistance of the solenoid?

CHALLENGE PROBLEMS

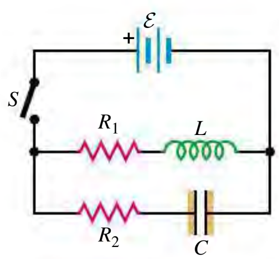
30.70 ••• **CP** **A Volume Gauge.** A tank containing a liquid has turns of wire wrapped around it, causing it to act like an inductor. The liquid content of the tank can be measured by using its inductance to determine the height of the liquid in the tank. The inductance of the tank changes from a value of L_0 corresponding to a relative permeability of 1 when the tank is empty to a value of L_f corresponding to a relative permeability of K_m (the relative permeability of the liquid) when the tank is full. The appropriate electronic circuitry can determine the inductance to five significant figures and thus the effective relative permeability of the combined air and liquid within the rectangular cavity of the tank. The four sides of the tank each have width W and height D (**Fig. P30.70**). The height of the liquid in the tank is d. You can ignore any fringing effects and assume that the relative permeability of the material of which the tank is made can be ignored. (a) Derive an expression for d as a function of L, the inductance corresponding to a certain fluid height, L_0, L_f, and D. (b) What is the inductance (to five significant figures) for a tank $\frac{1}{4}$ full, $\frac{1}{2}$ full, $\frac{3}{4}$ full, and completely full if the tank contains liquid oxygen? Take $L_0 = 0.63000$ H. The magnetic susceptibility of liquid oxygen is $\chi_m = 1.52 \times 10^{-3}$. (c) Repeat part (b) for mercury. The magnetic susceptibility of mercury is given in Table 28.1. (d) For which material is this volume gauge more practical?

Figure **P30.70**



30.71 ••• **CP CALC** Consider the circuit shown in **Fig. P30.71**. Switch S is closed at time $t = 0$, causing a current i_1 through the inductive branch and a current i_2 through the capacitive branch. The initial charge on the capacitor is zero, and the charge at time t is q_2. (a) Derive expressions for i_1, i_2, and q_2 as functions of time. Express your answers in terms of $\mathcal{E}$, L, C, R_1, R_2, and t. For the remainder of the problem let the circuit elements have the following values: $\mathcal{E} = 48$ V, $L = 8.0$ H, $C = 20\ \mu$F, $R_1 = 25$ Ω, and $R_2 = 5000$ Ω. (b) What is the initial current through the inductive branch? What is the initial current through the capacitive branch? (c) What are the currents through the inductive and capacitive branches a long time after the switch has been closed? How long is a "long time"? Explain. (d) At what time t_1 (accurate to two significant figures) will the currents i_1 and i_2 be equal? (*Hint:* You might consider using series expansions for the exponentials.) (e) For the conditions given in part (d), determine i_1. (f) The total current through the battery is $i = i_1 + i_2$. At what time t_2 (accurate to two significant figures) will i equal one-half of its final value? (*Hint:* The numerical work is greatly simplified if one makes suitable approximations. A sketch of i_1 and i_2 versus t may help you decide what approximations are valid.)

Figure **P30.71**



PASSAGE PROBLEMS

BIO QUENCHING AN MRI MAGNET. Magnets carrying very large currents are used to produce the uniform, large-magnitude magnetic fields that are required for *magnetic resonance imaging* (MRI). A typical MRI magnet may be a solenoid that is 2.0 m long and 1.0 m in diameter, has a self-inductance of 4.4 H, and carries a current of 750 A. A normal wire carrying that much current would dissipate a great deal of electrical power as heat, so most MRI magnets are made with coils of superconducting wire cooled by liquid helium at a temperature just under its boiling point (4.2 K). After a current is established in the wire, the power supply is disconnected and the magnet leads are shorted together through a piece of superconductor so that the current flows without resistance as long as the liquid helium keeps the magnet cold.

Under rare circumstances, a small segment of the magnet's wire may lose its superconducting properties and develop resistance. In this segment, electrical energy is converted to thermal energy, which can boil off some of the liquid helium. More of the wire then warms up and loses its superconducting properties, thus dissipating even more energy as heat. Because the latent heat of vaporization of liquid helium is quite low (20.9 kJ/kg), once the wire begins to warm up, all of the liquid helium may boil off rapidly. This event, called a *quench,* can damage the magnet. Also, a large volume of helium gas is generated as the liquid helium boils off, causing an asphyxiation hazard, and the resulting rapid pressure buildup can lead to an explosion. You can see how important it is to keep the wire resistance in an MRI magnet at zero and to have devices that detect a quench and shut down the current immediately.

30.72 How many turns does this typical MRI magnet have? (a) 1100; (b) 3000; (c) 4000; (d) 22,000.

30.73 If a small part of this magnet loses its superconducting properties and the resistance of the magnet wire suddenly rises from 0 to a constant 0.005 Ω, how much time will it take for the current to decrease to half of its initial value? (a) 4.7 min; (b) 10 min; (c) 15 min; (d) 30 min.

30.74 If part of the magnet develops resistance and liquid helium boils away, rendering more and more of the magnet nonsuperconducting, how will this quench affect the time for the current to drop to half of its initial value? (a) The time will be shorter because the resistance will increase; (b) the time will be longer because the resistance will increase; (c) the time will be the same; (d) not enough information is given.

30.75 If all of the magnetic energy stored in this MRI magnet is converted to thermal energy, how much liquid helium will boil off? (a) 27 kg; (b) 38 kg; (c) 60 kg; (d) 110 kg.

Answers

Chapter Opening Question ?

(iii) As explained in Section 30.2, traffic light sensors work by measuring the change in inductance of a coil embedded under the road surface when a car (which contains ferromagnetic material) drives over it.

Test Your Understanding Questions

30.1 (iii) Doubling both the length of the solenoid (l) and the number of turns of wire in the solenoid (N_1) would have *no* effect on the mutual inductance M. Example 30.1 shows that M depends on the ratio of these quantities, which would remain unchanged. This is because the magnetic field produced by the solenoid depends on the number of turns *per unit length,* and the proposed change has no effect on this quantity.

30.2 (iv), (i), (iii), (ii) From Eq. (30.8), the potential difference across the inductor is $V_{ab} = L\,di/dt$. For the four cases we find (i) $V_{ab} = (2.0\,\mu\text{H})(2.0\text{ A} - 1.0\text{ A})/(0.50\text{ s}) = 4.0\,\mu\text{V}$; (ii) $V_{ab} = (4.0\,\mu\text{H})(0 - 3.0\text{ A})/(2.0\text{ s}) = -6.0\,\mu\text{V}$; (iii) $V_{ab} = 0$ because the rate of change of current is zero; and (iv) $V_{ab} = (1.0\,\mu\text{H})(4.0\text{ A} - 0)/(0.25\text{ s}) = 16\,\mu\text{V}$.

30.3 (a) yes, (b) no Reversing the direction of the current has no effect on the magnetic-field magnitude, but it causes the direction of the magnetic field to reverse. It has no effect on the magnetic-field energy density, which is proportional to the square of the *magnitude* of the magnetic field.

30.4 (a) (i), (b) (ii) Recall that v_{ab} is the potential at a minus the potential at b, and similarly for v_{bc}. For either arrangement of the switches, current flows through the resistor from a to b. The upstream end of the resistor is always at the higher potential, so v_{ab} is positive. With S_1 closed and S_2 open, the current through the inductor flows from b to c and is increasing. The self-induced emf opposes this increase and is therefore directed from c toward b, which means that b is at the higher potential. Hence v_{bc} is positive. With S_1 open and S_2 closed, the inductor current again flows from b to c but is now decreasing. The self-induced emf is directed from b to c in an effort to sustain the decaying current, so c is at the higher potential and v_{bc} is negative.

30.5 (a) positive, (b) electric, (c) negative, (d) electric The capacitor loses energy between stages (a) and (b), so it does positive work on the charges. It does this by exerting an electric force that pushes current away from the positively charged left-hand capacitor plate and toward the negatively charged right-hand plate. At the same time, the inductor gains energy and does negative work on the moving charges. Although the inductor stores magnetic energy, the force that the inductor exerts is *electric*. This force comes about from the inductor's self-induced emf (see Section 30.2).

30.6 (i) and (iii) There are no oscillations if $R^2 \geq 4L/C$. In each case $R^2 = (2.0\,\Omega)^2 = 4.0\,\Omega^2$. In case (i) $4L/C = 4(3.0\,\mu\text{H})/(6.0\,\mu\text{F}) = 2.0\,\Omega^2$, so there are no oscillations (the system is overdamped); in case (ii) $4L/C = 4(6.0\,\mu\text{H})/(3.0\,\mu\text{F}) = 8.0\,\Omega^2$, so there are oscillations (the system is underdamped); and in case (iii) $4L/C = 4(3.0\,\mu\text{H})/(3.0\,\mu\text{F}) = 4.0\,\Omega^2$, so there are no oscillations (the system is critically damped).

Bridging Problem

(a) 7.68×10^{-8} J **(b)** 1.60 mA **(c)** 24.8 mV
(d) 1.92×10^{-8} J, 21.5 mV

? Waves from a broadcasting station produce an alternating current in the circuits of a radio (like the one in this classic car). If a radio is tuned to a station at a frequency of 1000 kHz, it will also detect the transmissions from a station broadcasting at (i) 600 kHz; (ii) 800 kHz; (iii) 1200 kHz; (iv) all of these; (v) none of these.

31 ALTERNATING CURRENT

LEARNING GOALS

Looking forward at ...

31.1 How phasors make it easy to describe sinusoidally varying quantities.

31.2 How to use reactance to describe the voltage across a circuit element that carries an alternating current.

31.3 How to analyze an *L-R-C* series circuit with a sinusoidal emf.

31.4 What determines the amount of power flowing into or out of an alternating-current circuit.

31.5 How an *L-R-C* series circuit responds to sinusoidal emfs of different frequencies.

31.6 Why transformers are useful, and how they work.

Looking back at ...

14.2, 14.8 Simple harmonic motion, resonance.

16.5 Resonance and sound.

18.3 Root-mean-square (rms) values.

25.3 Diodes.

26.3 Galvanometers.

28.8 Hysteresis in magnetic materials.

29.2, 29.6, 29.7 Alternating-current generators; eddy currents; displacement current.

30.1, 30.2, 30.5, 30.6 Mutual inductance; voltage across an inductor; *L-C* circuits; *L-R-C* series circuits.

During the 1880s in the United States there was a heated and acrimonious debate between two inventors over the best method of electric-power distribution. Thomas Edison favored direct current (dc)—that is, steady current that does not vary with time. George Westinghouse favored **alternating current (ac),** with sinusoidally varying voltages and currents. He argued that transformers (which we will study in this chapter) can be used to step the voltage up and down with ac but not with dc; low voltages are safer for consumer use, but high voltages and correspondingly low currents are best for long-distance power transmission to minimize i^2R losses in the cables.

Eventually, Westinghouse prevailed, and most present-day household and industrial power distribution systems operate with alternating current. Any appliance that you plug into a wall outlet uses ac. Circuits in modern communication equipment also make extensive use of ac.

In this chapter we will learn how resistors, inductors, and capacitors behave in circuits with sinusoidally varying voltages and currents. Many of the principles that we found useful in Chapter 30 are applicable, along with several new concepts related to the circuit behavior of inductors and capacitors. A key concept in this discussion is *resonance,* which we studied in Chapter 14 for mechanical systems.

31.1 PHASORS AND ALTERNATING CURRENTS

To supply an alternating current to a circuit, a source of alternating emf or voltage is required. An example of such a source is a coil of wire rotating with constant angular velocity in a magnetic field, which we discussed in Example 29.3 (Section 29.2). This develops a sinusoidal alternating emf and is the prototype of the commercial alternating-current generator or *alternator* (see Fig. 29.8).

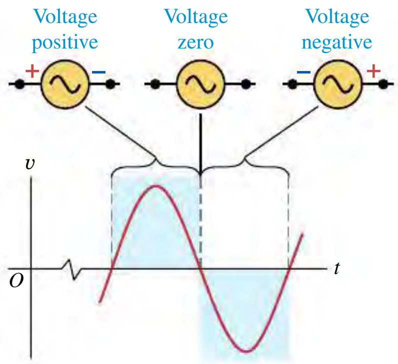
We use the term **ac source** for any device that supplies a sinusoidally varying voltage (potential difference) v or current i. The usual circuit-diagram symbol for an ac source is

A sinusoidal voltage might be described by a function such as

$$v = V\cos\omega t \tag{31.1}$$

In this expression, (lowercase) v is the *instantaneous* potential difference; (uppercase) V is the maximum potential difference, which we call the **voltage amplitude;** and ω is the *angular frequency,* equal to 2π times the frequency f (**Fig. 31.1**).

31.1 The voltage across a sinusoidal ac source.



In the United States and Canada, commercial electric-power distribution systems use a frequency $f = 60$ Hz, corresponding to $\omega = (2\pi \text{ rad})(60 \text{ s}^{-1}) = 377$ rad/s; in much of the rest of the world, $f = 50$ Hz ($\omega = 314$ rad/s) is used. Similarly, a sinusoidal current with a maximum value, or **current amplitude,** of I might be described as

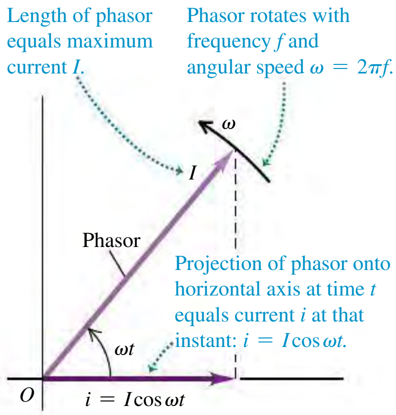
Sinusoidal alternating current:

$$\underset{\text{Instantaneous current}}{i} = \underset{\text{Current amplitude (maximum current)}}{I}\cos\underset{\text{Angular frequency}}{\omega}\underset{\text{Time}}{t} \tag{31.2}$$

Phasor Diagrams

To represent sinusoidally varying voltages and currents, we will use rotating vector diagrams similar to those we used in the study of simple harmonic motion in Section 14.2 (see Figs. 14.5b and 14.6). In these diagrams the instantaneous value of a quantity that varies sinusoidally with time is represented by the *projection* onto a horizontal axis of a vector with a length equal to the amplitude of the quantity. The vector rotates counterclockwise with constant angular speed ω. These rotating vectors are called **phasors,** and diagrams containing them are called **phasor diagrams. Figure 31.2** shows a phasor diagram for the sinusoidal current described by Eq. (31.2). The projection of the phasor onto the horizontal axis at time t is $I\cos\omega t$; this is why we chose to use the cosine function rather than the sine in Eq. (31.2).

31.2 A phasor diagram.



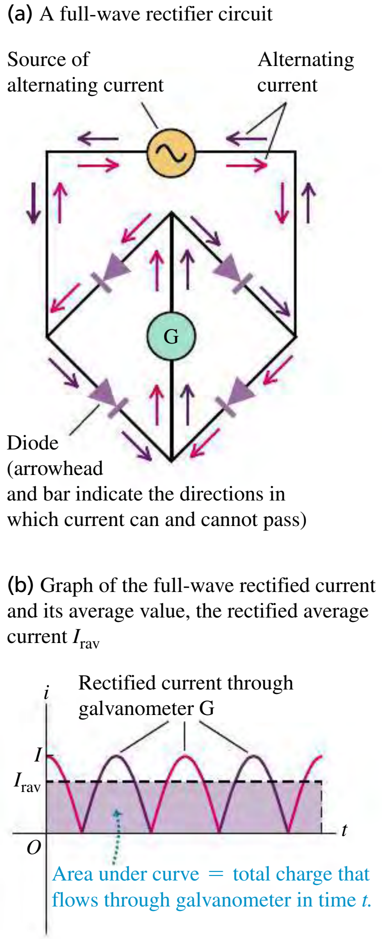
CAUTION **Just what is a phasor?** A phasor isn't a real physical quantity with a direction in space, such as velocity or electric field. Rather, it's a *geometric* entity that helps us describe physical quantities that vary sinusoidally with time. In Section 14.2 we used a single phasor to represent the position of a particle undergoing simple harmonic motion. Here we'll use phasors to *add* sinusoidal voltages and currents. Combining sinusoidal quantities with phase differences then involves vector addition. We'll use phasors in a similar way in Chapters 35 and 36 in our study of interference effects with light.

Rectified Alternating Current

How do we measure a sinusoidally varying current? In Section 26.3 we used a d'Arsonval galvanometer to measure steady currents. But if we pass a *sinusoidal* current through a d'Arsonval meter, the torque on the moving coil varies sinusoidally, with one direction half the time and the opposite direction the other half. The needle may wiggle a little if the frequency is low enough, but its average deflection is zero. Hence a d'Arsonval meter by itself isn't very useful for measuring alternating currents.

To get a measurable one-way current through the meter, we can use *diodes,* which we described in Section 25.3. A diode is a device that conducts better in one direction than in the other; an ideal diode has zero resistance for one

31.3 (a) A full-wave rectifier circuit. (b) Graph of the resulting current through the galvanometer G.



direction of current and infinite resistance for the other. **Figure 31.3a** shows one possible arrangement, called a *full-wave rectifier circuit*. The current through the galvanometer G is always upward, regardless of the direction of the current from the ac source (i.e., which part of the cycle the source is in). The graph in Fig. 31.3b shows the current through G: It pulsates but always has the same direction, and the average meter deflection is *not* zero.

The **rectified average current** I_{rav} is defined so that during any whole number of cycles, the total charge that flows is the same as though the current were constant with a value equal to I_{rav}. The notation I_{rav} and the name *rectified average* current emphasize that this is *not* the average of the original sinusoidal current. In Fig. 31.3b the total charge that flows in time t corresponds to the area under the curve of i versus t (recall that $i = dq/dt$, so q is the integral of t); this area must equal the rectangular area with height I_{rav}. We see that I_{rav} is less than the maximum current I; the two are related by

Rectified average value of a sinusoidal current

$$I_{rav} = \frac{2}{\pi} I = 0.637I \qquad (31.3)$$

Current amplitude

(The factor of $2/\pi$ is the average value of $|\cos \omega t|$ or of $|\sin \omega t|$; see Example 29.4 in Section 29.2.) The galvanometer deflection is proportional to I_{rav}. The galvanometer scale can be calibrated to read I, I_{rav}, or, most commonly, I_{rms} (discussed below).

Root-Mean-Square (rms) Values

A more useful way to describe a quantity that can be either positive or negative is the *root-mean-square (rms) value*. We used rms values in Section 18.3 in connection with the speeds of molecules in a gas. We *square* the instantaneous current i, take the *average* (mean) value of i^2, and finally take the *square root* of that average. This procedure defines the **root-mean-square current,** denoted as I_{rms} (**Fig. 31.4**). Even when i is negative, i^2 is always positive, so I_{rms} is never zero (unless i is zero at every instant).

Here's how we obtain I_{rms} for a sinusoidal current, like that shown in Fig. 31.4. If the instantaneous current is given by $i = I\cos \omega t$, then

$$i^2 = I^2 \cos^2 \omega t$$

Using a double-angle formula from trigonometry,

$$\cos^2 A = \tfrac{1}{2}(1 + \cos 2A)$$

we find

$$i^2 = I^2 \tfrac{1}{2}(1 + \cos 2\omega t) = \tfrac{1}{2}I^2 + \tfrac{1}{2}I^2 \cos 2\omega t$$

The average of $\cos 2\omega t$ is zero because it is positive half the time and negative half the time. Thus the average of i^2 is simply $I^2/2$. The square root of this is I_{rms}:

Root-mean-square (rms) value of a sinusoidal current

$$I_{rms} = \frac{I}{\sqrt{2}} \qquad (31.4)$$

Current amplitude

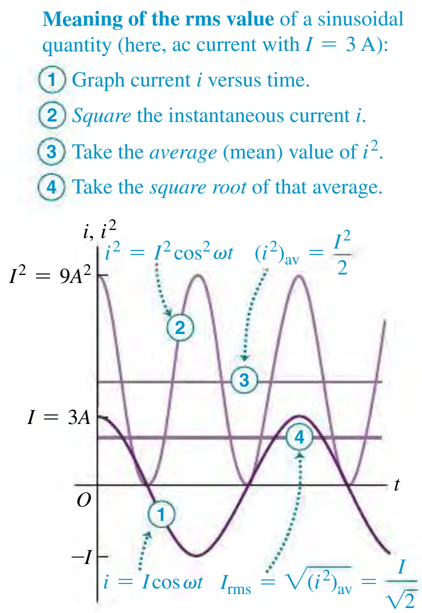
In the same way, the root-mean-square value of a sinusoidal voltage is

Root-mean-square (rms) value of a sinusoidal voltage

$$V_{rms} = \frac{V}{\sqrt{2}} \qquad (31.5)$$

Voltage amplitude (maximum value)

31.4 Calculating the root-mean-square (rms) value of an alternating current.



We can convert a rectifying ammeter into a voltmeter by adding a series resistor, just as for the dc case discussed in Section 26.3. Meters used for ac voltage and current measurements are nearly always calibrated to read rms values, not maximum or rectified average. Voltages and currents in power distribution systems are always described in terms of their rms values. The usual household power supply, "120-volt ac," has an rms voltage of 120 V (**Fig. 31.5**). The voltage amplitude is

$$V = \sqrt{2}V_{\text{rms}}$$
$$= \sqrt{2}(120\text{ V}) = 170\text{ V}$$

31.5 This wall socket delivers a root-mean-square voltage of 120 V. Sixty times per second, the instantaneous voltage across its terminals varies from $(\sqrt{2})(120\text{ V}) = 170\text{ V}$ to -170 V and back again.

EXAMPLE 31.1 CURRENT IN A PERSONAL COMPUTER

The plate on the back of a personal computer says that it draws 2.7 A from a 120-V, 60-Hz line. For this computer, what are (a) the average current, (b) the average of the square of the current, and (c) the current amplitude?

SOLUTION

IDENTIFY and SET UP: This example is about alternating current. In part (a) we find the average, over a complete cycle, of the alternating current. In part (b) we recognize that the 2.7-A current draw of the computer is the rms value I_{rms} —that is, the *square root* of the *mean* (average) of the *square* of the current, $(i^2)_{\text{av}}$. In part (c) we use Eq. (31.4) to relate I_{rms} to the current amplitude.

EXECUTE: (a) The average of *any* sinusoidally varying quantity, over any whole number of cycles, is zero.

(b) We are given $I_{\text{rms}} = 2.7$ A. From the definition of rms value,

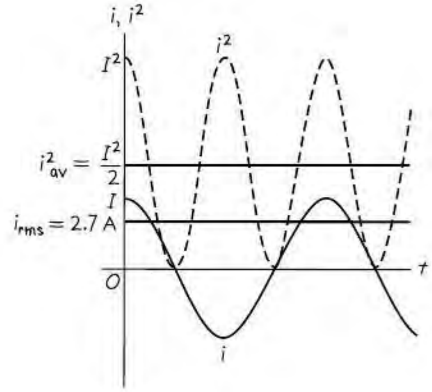
$$I_{\text{rms}} = \sqrt{(i^2)_{\text{av}}} \text{ so } (i^2)_{\text{av}} = (I_{\text{rms}})^2 = (2.7\text{ A})^2 = 7.3\text{ A}^2$$

(c) From Eq. (31.4), the current amplitude I is

$$I = \sqrt{2}I_{\text{rms}} = \sqrt{2}(2.7\text{ A}) = 3.8\text{ A}$$

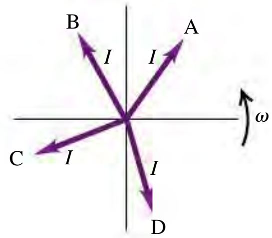
Figure 31.6 shows graphs of i and i^2 versus time t.

31.6 Our graphs of the current i and the square of the current i^2 versus time t.



EVALUATE: Why would we be interested in the average of the square of the current? Recall that the rate at which energy is dissipated in a resistor R equals i^2R. This rate varies if the current is alternating, so it is best described by its average value $(i^2)_{\text{av}}R = I_{\text{rms}}{}^2R$. We'll use this idea in Section 31.4.

TEST YOUR UNDERSTANDING OF SECTION 31.1 The accompanying figure shows four different current phasors with the same angular frequency ω. At the time shown, which phasor corresponds to (a) a positive current that is becoming more positive; (b) a positive current that is decreasing toward zero; (c) a negative current that is becoming more negative; (d) a negative current that is decreasing in magnitude toward zero? ❙



31.2 RESISTANCE AND REACTANCE

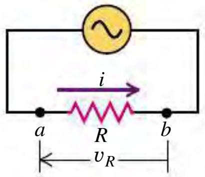
In this section we will derive voltage–current relationships for individual circuit elements—resistors, inductors, and capacitors—carrying a sinusoidal current.

Resistor in an ac Circuit

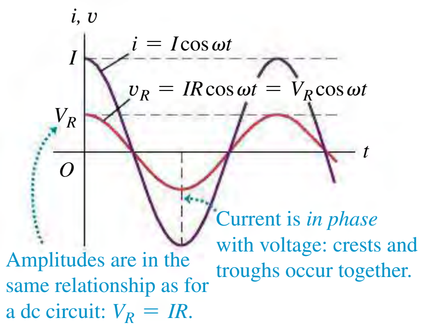
First let's consider a resistor with resistance R through which there is a sinusoidal current given by Eq. (31.2): $i = I\cos\omega t$. The positive direction of current is

31.7 Resistance R connected across an ac source.

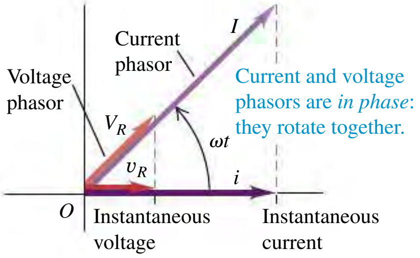
(a) Circuit with ac source and resistor



(b) Graphs of current and voltage versus time



(c) Phasor diagram



counterclockwise around the circuit (**Fig. 31.7a**). The current amplitude (maximum current) is I. From Ohm's law the instantaneous potential v_R of point a with respect to point b (that is, the instantaneous voltage across the resistor) is

$$v_R = iR = (IR)\cos\omega t \tag{31.6}$$

The maximum value of the voltage v_R is V_R, the *voltage amplitude:*

Amplitude of voltage across a resistor, ac circuit $V_R = IR$ (Current amplitude I; Resistance R) (31.7)

Hence we can also write

$$v_R = V_R\cos\omega t \tag{31.8}$$

Both the current i and the voltage v_R are proportional to $\cos\omega t$, so the current is *in phase* with the voltage. Equation (31.7) shows that the current and voltage amplitudes are related in the same way as in a dc circuit.

Figure 31.7b shows graphs of i and v_R as functions of time. The vertical scales for current and voltage are different, so the relative heights of the two curves are not significant. The corresponding phasor diagram is given in Fig. 31.7c. Because i and v_R are *in phase* and have the same frequency, the current and voltage phasors rotate together; they are parallel at each instant. Their projections on the horizontal axis represent the instantaneous current and voltage, respectively.

Inductor in an ac Circuit

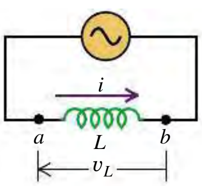
Now we replace the resistor in Fig. 31.7 with a pure inductor with self-inductance L and zero resistance (**Fig. 31.8a**). Again the current is $i = I\cos\omega t$, and the positive direction of current is counterclockwise around the circuit.

Although there is no resistance, there is a potential difference v_L between the inductor terminals a and b because the current varies with time, giving rise to a self-induced emf. The induced emf in the direction of i is given by Eq. (30.7), $\mathcal{E} = -L\,di/dt$; however, the voltage v_L is *not* simply equal to $\mathcal{E}$. To see why, notice that if the current in the inductor is in the positive (counterclockwise) direction from a to b and is increasing, then di/dt is positive and the induced emf is directed to the left to oppose the increase in current; hence point a is at higher potential than is point b. Thus the potential of point a with respect to point b is positive and is given by $v_L = +L\,di/dt$, the *negative* of the induced emf. (Convince yourself that this expression gives the correct sign of v_L in *all* cases, including i counterclockwise and decreasing, i clockwise and increasing, and i clockwise and decreasing; also review Section 30.2.) So

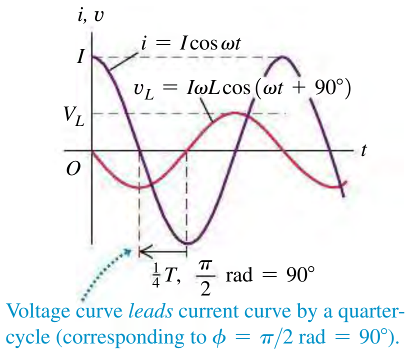
$$v_L = L\frac{di}{dt} = L\frac{d}{dt}(I\cos\omega t) = -I\omega L\sin\omega t \tag{31.9}$$

31.8 Inductance L connected across an ac source.

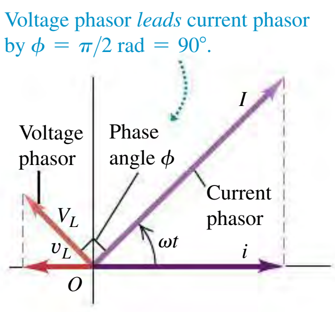
(a) Circuit with ac source and inductor



(b) Graphs of current and voltage versus time



(c) Phasor diagram



The voltage v_L across the inductor at any instant is proportional to the *rate of change* of the current. The points of maximum voltage on the graph correspond to maximum steepness of the current curve, and the points of zero voltage are the points where the current curve has its maximum and minimum values (Fig. 31.8b). The voltage and current are *out of phase* by a quarter-cycle. Since the voltage peaks occur a quarter-cycle earlier than the current peaks, we say that the voltage *leads* the current by 90°. The phasor diagram in Fig. 31.8c also shows this relationship; the voltage phasor is ahead of the current phasor by 90°.

We can also obtain this phase relationship by rewriting Eq. (31.9) with the identity $\cos(A + 90°) = -\sin A$:

$$v_L = I\omega L\cos(\omega t + 90°) \tag{31.10}$$

This result shows that the voltage can be viewed as a cosine function with a "head start" of 90° relative to the current.

As we have done in Eq. (31.10), we will usually describe the phase of the *voltage* relative to the *current,* not the reverse. Thus if the current i in a circuit is

$$i = I\cos\omega t$$

and the voltage v of one point with respect to another is

$$v = V\cos(\omega t + \phi)$$

we call ϕ the **phase angle;** it gives the phase of the *voltage* relative to the *current*. For a pure resistor, $\phi = 0$, and for a pure inductor, $\phi = 90°$.

From Eq. (31.9) or (31.10) the amplitude V_L of the inductor voltage is

$$V_L = I\omega L \tag{31.11}$$

We define the **inductive reactance** X_L of an inductor as

$$X_L = \omega L \qquad \text{(inductive reactance)} \tag{31.12}$$

Using X_L, we can write Eq. (31.11) in a form similar to Eq. (31.7) for a resistor:

Amplitude of voltage across an inductor, ac circuit

$$V_L = I X_L \tag{31.13}$$

Current amplitude; Inductive reactance

Because X_L is the ratio of a voltage and a current, its SI unit is the ohm, the same as for resistance.

CAUTION Inductor voltage and current are not in phase Equation (31.13) relates the *amplitudes* of the oscillating voltage and current for the inductor in Fig. 31.8a. It does *not* say that the voltage at any instant is equal to the current at that instant multiplied by X_L. As Fig. 31.8b shows, the voltage and current are 90° out of phase. Voltage and current are in phase only for resistors, as in Eq. (31.6).

The Meaning of Inductive Reactance

The inductive reactance X_L is really a description of the self-induced emf that opposes any change in the current through the inductor. From Eq. (31.13), for a given current amplitude I the voltage $v_L = +L\,di/dt$ across the inductor and the self-induced emf $\mathcal{E} = -L\,di/dt$ both have an amplitude V_L that is directly proportional to X_L. According to Eq. (31.12), the inductive reactance and self-induced emf increase with more rapid variation in current (that is, increasing angular frequency ω) and increasing inductance L.

If an oscillating voltage of a given amplitude V_L is applied across the inductor terminals, the resulting current will have a smaller amplitude I for larger values of X_L. Since X_L is proportional to frequency, a high-frequency voltage applied to the inductor gives only a small current, while a lower-frequency voltage of the same amplitude gives rise to a larger current. Inductors are used in some circuit applications, such as power supplies and radio-interference filters, to block high frequencies while permitting lower frequencies or dc to pass through. A circuit device that uses an inductor for this purpose is called a *low-pass filter* (see Problem 31.48).

SOLUTION

EXAMPLE 31.2 AN INDUCTOR IN AN AC CIRCUIT

The current amplitude in a pure inductor in a radio receiver is to be 250 μA when the voltage amplitude is 3.60 V at a frequency of 1.60 MHz (at the upper end of the AM broadcast band). (a) What inductive reactance is needed? What inductance? (b) If the voltage amplitude is kept constant, what will be the current amplitude through this inductor at 16.0 MHz? At 160 kHz?

SOLUTION

IDENTIFY and SET UP: The circuit may have other elements, but in this example we don't care: All they do is provide the inductor with an oscillating voltage, so the other elements are lumped into the ac source shown in Fig. 31.8a. We are given the current amplitude I and the voltage amplitude V. Our target variables in part (a) are the inductive reactance X_L at 1.60 MHz and the inductance L, which we find from Eqs. (31.13) and (31.12). Knowing L, we use these equations in part (b) to find X_L and I at any frequency.

EXECUTE: (a) From Eq. (31.13),

$$X_L = \frac{V_L}{I} = \frac{3.60\ \text{V}}{250 \times 10^{-6}\ \text{A}} = 1.44 \times 10^4\ \Omega = 14.4\ \text{k}\Omega$$

From Eq. (31.12), with $\omega = 2\pi f$,

$$L = \frac{X_L}{2\pi f} = \frac{1.44 \times 10^4\ \Omega}{2\pi(1.60 \times 10^6\ \text{Hz})} = 1.43 \times 10^{-3}\ \text{H} = 1.43\ \text{mH}$$

(b) Combining Eqs. (31.12) and (31.13), we find $I = V_L/X_L = V_L/\omega L = V_L/2\pi f L$. Thus the current amplitude is inversely proportional to the frequency f. Since $I = 250\ \mu\text{A}$ at $f = 1.60$ MHz, the current amplitudes at 16.0 MHz $(10f)$ and 160 kHz $= 0.160$ MHz $(f/10)$ will be, respectively, one-tenth as great $(25.0\ \mu\text{A})$ and ten times as great $(2500\ \mu\text{A} = 2.50\ \text{mA})$.

EVALUATE: In general, the lower the frequency of an oscillating voltage applied across an inductor, the greater the amplitude of the resulting oscillating current.

CAUTION Alternating current through a capacitor Charge can't really move through the capacitor because its two plates are insulated from each other. But as the capacitor charges and discharges, there is at each instant a current i into one plate, an equal current out of the other plate, and an equal *displacement* current between the plates. (You should review the discussion of displacement current in Section 29.7.) Thus we often speak about alternating current *through* a capacitor.

Capacitor in an ac Circuit

Finally, we connect a capacitor with capacitance C to the source, as in **Fig. 31.9a**, producing a current $i = I\cos\omega t$ through the capacitor. Again, the positive direction of current is counterclockwise around the circuit.

To find the instantaneous voltage v_C across the capacitor—that is, the potential of point a with respect to point b—we first let q denote the charge on the left-hand plate of the capacitor in Fig. 31.9a (so $-q$ is the charge on the right-hand plate). The current i is related to q by $i = dq/dt$; with this definition, positive current corresponds to an increasing charge on the left-hand capacitor plate. Then

$$i = \frac{dq}{dt} = I\cos\omega t$$

Integrating this, we get

$$q = \frac{I}{\omega}\sin\omega t \tag{31.14}$$

Also, from Eq. (24.1) the charge q equals the voltage v_C multiplied by the capacitance, $q = Cv_C$. Using this in Eq. (31.14), we find

$$v_C = \frac{I}{\omega C}\sin\omega t \tag{31.15}$$

The instantaneous current i is equal to the rate of change dq/dt of the capacitor charge q; since $q = Cv_C$, i is also proportional to the rate of change of voltage. (Compare to an inductor, for which the situation is reversed and v_L is proportional to the rate of change of i.) Figure 31.9b shows v_C and i as functions of t. Because $i = dq/dt = C\,dv_C/dt$, the current has its greatest magnitude when the v_C curve is rising or falling most steeply and is zero when the v_C curve instantaneously levels off at its maximum and minimum values.

The peaks of capacitor voltage occur a quarter-cycle *after* the corresponding current peaks, and we say that the voltage *lags* the current by 90°. The phasor diagram in Fig. 31.9c shows this relationship; the voltage phasor is behind the current phasor by a quarter-cycle, or 90°.

We can also derive this phase difference by rewriting Eq. (31.15) with the identity $\cos(A - 90°) = \sin A$:

$$v_C = \frac{I}{\omega C}\cos(\omega t - 90°) \tag{31.16}$$

This corresponds to a phase angle $\phi = -90°$. This cosine function has a "late start" of 90° compared with the current $i = I\cos\omega t$.

Equations (31.15) and (31.16) show that the voltage *amplitude* V_C is

$$V_C = \frac{I}{\omega C} \tag{31.17}$$

To put this expression in a form similar to Eq. (31.7) for a resistor, $V_R = IR$, we define a quantity X_C, called the **capacitive reactance** of the capacitor, as

$$X_C = \frac{1}{\omega C} \quad \text{(capacitive reactance)} \tag{31.18}$$

Then

Amplitude of voltage across a capacitor, ac circuit

$$V_C = \underbrace{I}_{\text{Current amplitude}}\underbrace{X_C}_{\text{Capacitive reactance}} \tag{31.19}$$

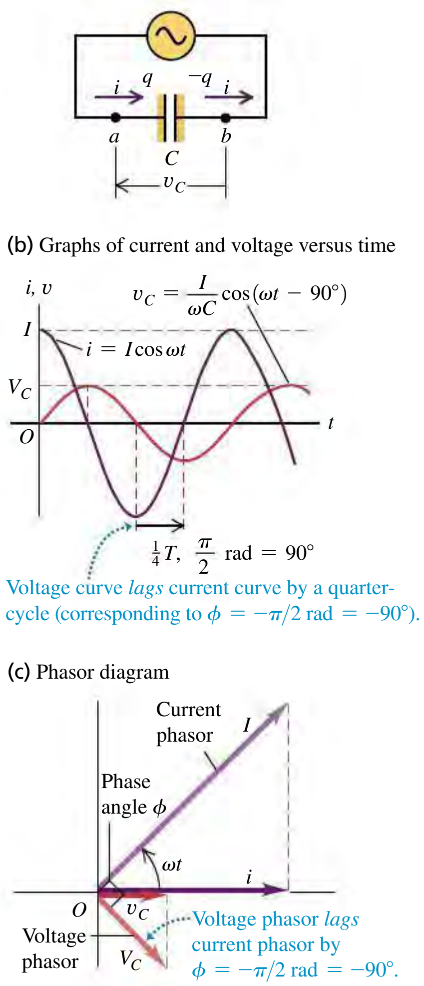
The SI unit of X_C is the ohm, the same as for resistance and inductive reactance, because X_C is the ratio of a voltage and a current.

CAUTION **Capacitor voltage and current are not in phase** Remember that Eq. (31.19) for a capacitor, like Eq. (31.13) for an inductor, is *not* a statement about the instantaneous values of voltage and current. The instantaneous values are 90° out of phase, as Fig. 31.9b shows. Rather, Eq. (31.19) relates the *amplitudes* of voltage and current.

The Meaning of Capacitive Reactance

The capacitive reactance of a capacitor is inversely proportional both to the capacitance C and to the angular frequency ω; the greater the capacitance and the higher the frequency, the *smaller* the capacitive reactance X_C. Capacitors tend to pass high-frequency current and to block low-frequency currents and dc, just the opposite of inductors. A device that preferentially passes signals of high frequency is called a *high-pass filter* (see Problem 31.47).

31.9 Capacitor C connected across an ac source.



(a) Circuit with ac source and capacitor

(b) Graphs of current and voltage versus time

(c) Phasor diagram

SOLUTION

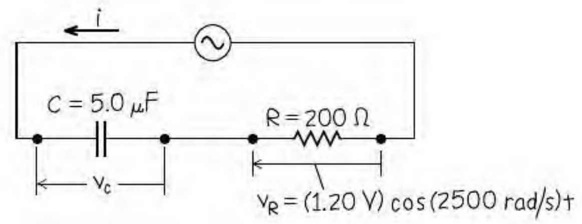
EXAMPLE 31.3 A RESISTOR AND A CAPACITOR IN AN AC CIRCUIT

A 200-Ω resistor is connected in series with a 5.0-μF capacitor. The voltage across the resistor is $v_R = (1.20\text{ V})\cos(2500\text{ rad/s})t$ (**Fig. 31.10**). (a) Derive an expression for the circuit current. (b) Determine the capacitive reactance of the capacitor. (c) Derive an expression for the voltage across the capacitor.

SOLUTION

IDENTIFY and SET UP: Since this is a series circuit, the current is the same through the capacitor as through the resistor. Our target variables are the current i, the capacitive reactance X_C, and the capacitor voltage v_C. We use Eq. (31.6) to find an expression for i in terms of the angular frequency $\omega = 2500$ rad/s, Eq. (31.18) to find X_C, Eq. (31.19) to find the capacitor voltage amplitude V_C, and Eq. (31.16) to write an expression for v_C.

31.10 Our sketch for this problem.



EXECUTE: (a) From Eq. (31.6), $v_R = iR$, we find

$$i = \frac{v_R}{R} = \frac{(1.20\text{ V})\cos(2500\text{ rad/s})t}{200\ \Omega}$$
$$= (6.0 \times 10^{-3}\text{ A})\cos(2500\text{ rad/s})t$$

(b) From Eq. (31.18), the capacitive reactance at $\omega = 2500$ rad/s is

$$X_C = \frac{1}{\omega C} = \frac{1}{(2500\text{ rad/s})(5.0 \times 10^{-6}\text{ F})} = 80\ \Omega$$

(c) From Eq. (31.19), the capacitor voltage amplitude is

$$V_C = IX_C = (6.0 \times 10^{-3}\text{ A})(80\ \Omega) = 0.48\text{ V}$$

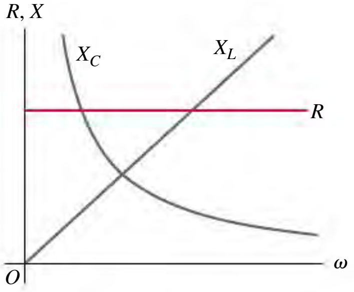
Continued

(The 80-Ω reactance of the capacitor is 40% of the resistor's 200-Ω resistance, so V_C is 40% of V_R.) The instantaneous capacitor voltage is given by Eq. (31.16):

$$v_C = V_C \cos(\omega t - 90°)$$
$$= (0.48 \text{ V}) \cos[(2500 \text{ rad/s})t - \pi/2 \text{ rad}]$$

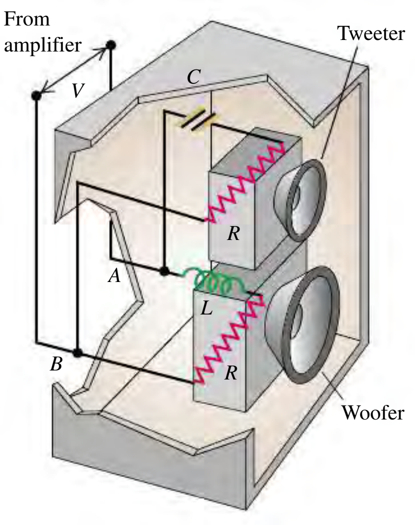
EVALUATE: Although the same *current* passes through both the capacitor and the resistor, the *voltages* across them are different in both amplitude and phase. Note that in the expression for v_C we converted the 90° to $\pi/2$ rad so that all the angular quantities have the same units. In ac circuit analysis, phase angles are often given in degrees, so be careful to convert to radians when necessary.

31.11 Graphs of R, X_L, and X_C as functions of angular frequency ω.

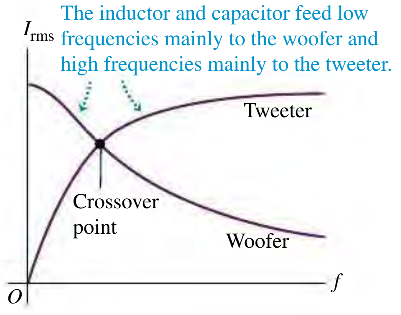


31.12 (a) The two speakers in this loudspeaker system are connected in parallel to the amplifier. (b) Graphs of current amplitude in the tweeter and woofer as functions of frequency for a given amplifier voltage amplitude.

(a) A crossover network in a loudspeaker system



(b) Graphs of rms current as functions of frequency for a given amplifier voltage



Comparing ac Circuit Elements

Table 31.1 summarizes the relationships of voltage and current amplitudes for the three circuit elements we have discussed. Note again that *instantaneous* voltage and current are proportional in a resistor, where there is zero phase difference between v_R and i (see Fig. 31.7b). The instantaneous voltage and current are *not* proportional in an inductor or capacitor, because there is a 90° phase difference in both cases (see Figs. 31.8b and 31.9b).

TABLE 31.1 Circuit Elements with Alternating Current

Circuit Element	Amplitude Relationship	Circuit Quantity	Phase of v
Resistor	$V_R = IR$	R	In phase with i
Inductor	$V_L = IX_L$	$X_L = \omega L$	Leads i by 90°
Capacitor	$V_C = IX_C$	$X_C = 1/\omega C$	Lags i by 90°

Figure 31.11 shows how the resistance of a resistor and the reactances of an inductor and a capacitor vary with angular frequency ω. Resistance R is independent of frequency, while the reactances X_L and X_C are not. If $\omega = 0$, corresponding to a dc circuit, there is *no* current through a capacitor because $X_C \rightarrow \infty$, and there is no inductive effect because $X_L = 0$. In the limit $\omega \rightarrow \infty$, X_L also approaches infinity, and the current through an inductor becomes vanishingly small; recall that the self-induced emf opposes rapid changes in current. In this same limit, X_C and the voltage across a capacitor both approach zero; the current changes direction so rapidly that no charge can build up on either plate.

Figure 31.12 shows an application of the above discussion to a loudspeaker system. Low-frequency sounds are produced by the *woofer,* which is a speaker with large diameter; the *tweeter,* a speaker with smaller diameter, produces high-frequency sounds. In order to route signals of different frequency to the appropriate speaker, the woofer and tweeter are connected in parallel across the amplifier output. The capacitor in the tweeter branch blocks the low-frequency components of sound but passes the higher frequencies; the inductor in the woofer branch does the opposite.

TEST YOUR UNDERSTANDING OF SECTION 31.2 An oscillating voltage of fixed amplitude is applied across a circuit element. If the frequency of this voltage is increased, will the amplitude of the current through the element (i) increase, (ii) decrease, or (iii) remain the same if it is (a) a resistor, (b) an inductor, or (c) a capacitor?

31.3 THE *L-R-C* SERIES CIRCUIT

Many ac circuits used in practical electronic systems involve resistance, inductive reactance, and capacitive reactance. **Figure 31.13a** shows a simple example: a series circuit containing a resistor, an inductor, a capacitor, and an ac source. (In Section 30.6 we studied an *L-R-C* series circuit *without* a source.)

To analyze this circuit, we'll use a phasor diagram that includes the voltage and current phasors for each of the components. Because of Kirchhoff's loop rule, the instantaneous *total* voltage v_{ad} across all three components is equal to the source voltage at that instant. We will show that the phasor representing this total voltage is the *vector sum* of the phasors for the individual voltages.

Figures 31.13b and 31.13c show complete phasor diagrams for the circuit of Fig. 31.13a. We assume that the source supplies a current i given by $i = I\cos\omega t$. Because the circuit elements are connected in series, the current at any instant is the same at every point in the circuit. Thus a *single phasor I*, with length proportional to the current amplitude, represents the current in *all* circuit elements.

As in Section 31.2, we use the symbols v_R, v_L, and v_C for the instantaneous voltages across R, L, and C, and the symbols V_R, V_L, and V_C for the maximum voltages. We denote the instantaneous and maximum *source* voltages by v and V. Then, in Fig. 31.13a, $v = v_{ad}$, $v_R = v_{ab}$, $v_L = v_{bc}$, and $v_C = v_{cd}$.

The potential difference between the terminals of a resistor is *in phase* with the current in the resistor. Its maximum value V_R is given by Eq. (31.7):

$$V_R = IR$$

The phasor V_R in Fig. 31.13b, in phase with the current phasor I, represents the voltage across the resistor. Its projection onto the horizontal axis at any instant gives the instantaneous potential difference v_R.

The voltage across an inductor *leads* the current by 90°. Its voltage amplitude is given by Eq. (31.13):

$$V_L = IX_L$$

The phasor V_L in Fig. 31.13b represents the voltage across the inductor, and its projection onto the horizontal axis at any instant equals v_L.

The voltage across a capacitor *lags* the current by 90°. Its voltage amplitude is given by Eq. (31.19):

$$V_C = IX_C$$

The phasor V_C in Fig. 31.13b represents the voltage across the capacitor, and its projection onto the horizontal axis at any instant equals v_C.

The instantaneous potential difference v between terminals a and d is equal at every instant to the (algebraic) sum of the potential differences v_R, v_L, and v_C. That is, it equals the sum of the *projections* of the phasors V_R, V_L, and V_C. But the sum of the projections of these phasors is equal to the *projection* of their *vector sum*. So the vector sum V must be the phasor that represents the source voltage v and the instantaneous total voltage v_{ad} across the series of elements.

To form this vector sum, we first subtract the phasor V_C from the phasor V_L. (These two phasors always lie along the same line, with opposite directions.) This gives the phasor $V_L - V_C$. This is always at right angles to the phasor V_R, so from the Pythagorean theorem the magnitude of the phasor V is

$$V = \sqrt{V_R^2 + (V_L - V_C)^2} = \sqrt{(IR)^2 + (IX_L - IX_C)^2} \quad \text{or}$$

$$V = I\sqrt{R^2 + (X_L - X_C)^2} \tag{31.20}$$

We define the **impedance** Z of an ac circuit as the ratio of the voltage amplitude across the circuit to the current amplitude in the circuit. From Eq. (31.20) the impedance of the *L-R-C* series circuit is

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{31.21}$$

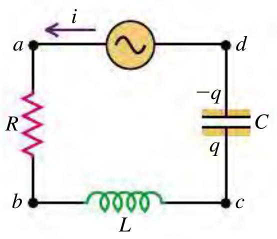
so we can rewrite Eq. (31.20) as

Amplitude of voltage across an ac circuit Current amplitude Impedance of circuit

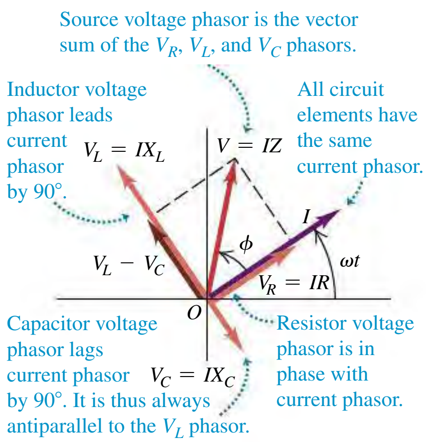
$$V = IZ \tag{31.22}$$

31.13 An *L-R-C* series circuit with an ac source.

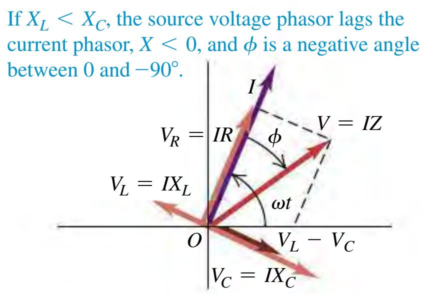
(a) *L-R-C* series circuit



(b) Phasor diagram for the case $X_L > X_C$



(c) Phasor diagram for the case $X_L < X_C$



PhET: Circuit Construction Kit (AC+DC)
PhET: Faraday's Electromagnetic Lab

31.14 This gas-filled glass sphere has an alternating voltage between its surface and the electrode at its center. The glowing streamers show the resulting alternating current that passes through the gas. When you touch the outside of the sphere, your fingertips and the inner surface of the sphere act as the plates of a capacitor, and the sphere and your body together form an *L-R-C* series circuit. The current (which is low enough to be harmless) is drawn to your fingers because the path through your body has a low impedance.

While Eq. (31.21) is valid only for an *L-R-C* series circuit, we can use Eq. (31.22) to define the impedance of *any* network of resistors, inductors, and capacitors as the ratio of the amplitude of the voltage across the network to the current amplitude. The SI unit of impedance is the ohm.

The Meaning of Impedance and Phase Angle

Equation (31.22) has a form similar to $V = IR$, with impedance Z in an ac circuit playing the role of resistance R in a dc circuit. Just as direct current tends to follow the path of least resistance, so alternating current tends to follow the path of lowest impedance (**Fig. 31.14**). Note, however, that impedance is actually a function of R, L, and C, as well as of the angular frequency ω. We can see this by substituting Eq. (31.12) for X_L and Eq. (31.18) for X_C into Eq. (31.21), giving the following complete expression for Z for a series circuit:

Impedance of an *L-R-C* series circuit

$$Z = \sqrt{\overset{\text{Resistance}}{R^2} + [\underset{\text{Angular frequency}}{\omega} \overset{\text{Inductance}}{L} - (1/\underset{\text{Angular frequency}}{\omega} \overset{\text{Capacitance}}{C})]^2} \tag{31.23}$$

Hence for a given amplitude V of the source voltage applied to the circuit, the amplitude $I = V/Z$ of the resulting current will be different at different frequencies. We'll explore this frequency dependence in detail in Section 31.5.

In Fig. 31.13b, the angle ϕ between the voltage and current phasors is the phase angle of the source voltage v with respect to the current i; that is, it is the angle by which the source voltage leads the current. From the diagram,

$$\tan\phi = \frac{V_L - V_C}{V_R} = \frac{I(X_L - X_C)}{IR} = \frac{X_L - X_C}{R}$$

Phase angle of voltage with respect to current in an *L-R-C* series circuit

$$\tan\phi = \frac{\underset{\text{Angular frequency}}{\omega}\overset{\text{Inductance}}{L} - 1/\underset{\text{Angular frequency}}{\omega}\overset{\text{Capacitance}}{C}}{\underset{\text{Resistance}}{R}} \tag{31.24}$$

If the current is $i = I\cos\omega t$, then the source voltage v is

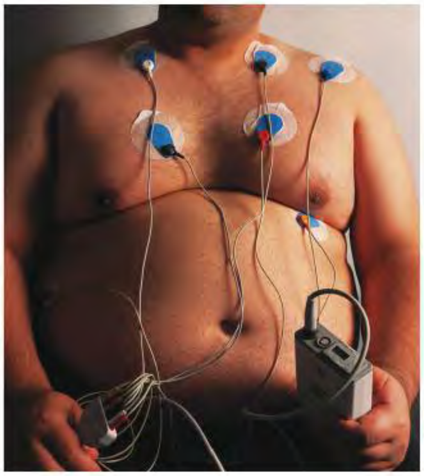
$$v = V\cos(\omega t + \phi) \tag{31.25}$$

Figure 31.13b shows the behavior of an *L-R-C* series circuit in which $X_L > X_C$. Figure 31.13c shows the behavior when $X_L < X_C$; the voltage phasor V lies on the opposite side of the current phasor I and the voltage *lags* the current. In this case, $X_L - X_C$ is *negative,* $\tan\phi$ is negative, and ϕ is a negative angle between $0°$ and $-90°$. Since X_L and X_C depend on frequency, the phase angle ϕ depends on frequency as well. We'll examine the consequences of this in Section 31.5.

All of the expressions that we've developed for an *L-R-C* series circuit are still valid if one of the circuit elements is missing. If the resistor is missing, we set $R = 0$; if the inductor is missing, we set $L = 0$. But if the capacitor is missing, we set $C = \infty$, corresponding to the absence of any potential difference $(v_C = q/C = 0)$ or any capacitive reactance $(X_C = 1/\omega C = 0)$.

In this entire discussion we have described magnitudes of voltages and currents in terms of their *maximum* values, the voltage and current *amplitudes.* But we remarked at the end of Section 31.1 that these quantities are usually described in terms of rms values, not amplitudes. For any sinusoidally varying quantity, the rms value is always $1/\sqrt{2}$ times the amplitude. All the relationships between voltage and current that we have derived in this and the preceding sections are

BIO Application Measuring Body Fat by Bioelectric Impedance Analysis The electrodes attached to this overweight patient's chest are applying a small ac voltage of frequency 50 kHz. The attached instrumentation measures the amplitude and phase angle of the resulting current through the patient's body. These depend on the relative amounts of water and fat along the path followed by the current, and so provide a sensitive measure of body composition.



still valid if we use rms quantities throughout instead of amplitudes. For example, if we divide Eq. (31.22) by $\sqrt{2}$, we get

$$\frac{V}{\sqrt{2}} = \frac{I}{\sqrt{2}}Z$$

which we can rewrite as

$$V_{\mathrm{rms}} = I_{\mathrm{rms}}Z \tag{31.26}$$

We can translate Eqs. (31.7), (31.13), and (31.19) in exactly the same way.

We have considered only ac circuits in which an inductor, a resistor, and a capacitor are in series. You can do a similar analysis for an *L-R-C parallel* circuit; see Problem 31.54.

PROBLEM-SOLVING STRATEGY 31.1 ALTERNATING-CURRENT CIRCUITS

IDENTIFY *the relevant concepts:* In analyzing ac circuits, we can apply all of the concepts used to analyze direct-current circuits, particularly those in Problem-Solving Strategies 26.1 and 26.2. But now we must distinguish between the amplitudes of alternating currents and voltages and their instantaneous values, and among resistance (for resistors), reactance (for inductors or capacitors), and impedance (for composite circuits).

SET UP *the problem* using the following steps:
1. Draw a diagram of the circuit and label all known and unknown quantities.
2. Identify the target variables.

EXECUTE *the solution* as follows:
1. Use the relationships derived in Sections 31.2 and 31.3 to solve for the target variables, using the following hints.
2. It's almost always easiest to work with angular frequency $\omega = 2\pi f$ rather than ordinary frequency f.
3. Keep in mind the following phase relationships: For a resistor, voltage and current are *in phase,* so the corresponding phasors always point in the same direction. For an inductor, the voltage *leads* the current by 90° (i.e., $\phi = +90° = \pi/2$ radians), so the voltage phasor points 90° counterclockwise from the current phasor. For a capacitor, the voltage *lags* the current by 90° (i.e., $\phi = -90° = -\pi/2$ radians), so the voltage phasor points 90° clockwise from the current phasor.
4. Kirchhoff's rules hold *at each instant*. For example, in a series circuit, the instantaneous current is the same in all circuit elements; in a parallel circuit, the instantaneous potential difference is the same across all circuit elements.
5. Inductive reactance, capacitive reactance, and impedance are analogous to resistance; each represents the ratio of voltage amplitude V to current amplitude I in a circuit element or combination of elements. However, phase relationships are crucial. In applying Kirchhoff's loop rule, you must combine the effects of resistance and reactance by *vector* addition of the corresponding voltage phasors, as in Figs. 31.13b and 31.13c. When several circuit elements are in series, for example, you can't *add* all the numerical values of resistance and reactance to get the impedance; that would ignore the phase relationships.

EVALUATE *your answer:* When working with an *L-R-C series* circuit, you can check your results by comparing the values of the inductive and capacitive reactances X_L and X_C. If $X_L > X_C$, then the voltage amplitude across the inductor is greater than that across the capacitor and the phase angle ϕ is positive (between 0° and 90°). If $X_L < X_C$, then the voltage amplitude across the inductor is less than that across the capacitor and the phase angle ϕ is negative (between 0° and −90°).

EXAMPLE 31.4 AN *L-R-C* SERIES CIRCUIT I

SOLUTION

In the series circuit of Fig. 31.13a, suppose $R = 300\ \Omega$, $L = 60$ mH, $C = 0.50\ \mu$F, $V = 50$ V, and $\omega = 10{,}000$ rad/s. Find the reactances X_L and X_C, the impedance Z, the current amplitude I, the phase angle ϕ, and the voltage amplitude across each circuit element.

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas developed in Section 31.2 and this section about the behavior of circuit elements in an ac circuit. We use Eqs. (31.12) and (31.18) to determine X_L and X_C, and Eq. (31.23) to find Z. We then use Eq. (31.22) to find the current amplitude and Eq. (31.24) to find the phase angle. The relationships in Table 31.1 then yield the voltage amplitudes.

EXECUTE: The inductive and capacitive reactances are

$$X_L = \omega L = (10{,}000\ \mathrm{rad/s})(60\ \mathrm{mH}) = 600\ \Omega$$

$$X_C = \frac{1}{\omega C} = \frac{1}{(10{,}000\ \mathrm{rad/s})(0.50 \times 10^{-6}\ \mathrm{F})} = 200\ \Omega$$

The impedance Z of the circuit is then

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{(300\ \Omega)^2 + (600\ \Omega - 200\ \Omega)^2}$$
$$= 500\ \Omega$$

Continued

With source voltage amplitude $V = 50$ V, the current amplitude I and phase angle ϕ are

$$I = \frac{V}{Z} = \frac{50\ \text{V}}{500\ \Omega} = 0.10\ \text{A}$$

$$\phi = \arctan\frac{X_L - X_C}{R} = \arctan\frac{400\ \Omega}{300\ \Omega} = 53°$$

From Table 31.1, the voltage amplitudes V_R, V_L, and V_C across the resistor, inductor, and capacitor, respectively, are

$$V_R = IR = (0.10\ \text{A})(300\ \Omega) = 30\ \text{V}$$
$$V_L = IX_L = (0.10\ \text{A})(600\ \Omega) = 60\ \text{V}$$
$$V_C = IX_C = (0.10\ \text{A})(200\ \Omega) = 20\ \text{V}$$

EVALUATE: As in Fig. 31.13b, $X_L > X_C$; hence the voltage amplitude across the inductor is greater than that across the capacitor and ϕ is positive. The value $\phi = 53°$ means that the voltage *leads* the current by 53°.

Note that the source voltage amplitude $V = 50$ V is *not* equal to the sum of the voltage amplitudes across the separate circuit elements: $50\ \text{V} \neq 30\ \text{V} + 60\ \text{V} + 20\ \text{V}$. Instead, V is the *vector sum* of the V_R, V_L, and V_C phasors. If you draw the phasor diagram like Fig. 31.13b for this particular situation, you'll see that V_R, $V_L - V_C$, and V constitute a 3-4-5 right triangle.

EXAMPLE 31.5 AN *L-R-C* SERIES CIRCUIT II

SOLUTION

For the *L-R-C* series circuit of Example 31.4, find expressions for the time dependence of the instantaneous current i and the instantaneous voltages across the resistor (v_R), inductor (v_L), capacitor (v_C), and ac source (v).

SOLUTION

IDENTIFY and SET UP: We describe the current by using Eq. (31.2), which assumes that the current is maximum at $t = 0$. The voltages are then given by Eq. (31.8) for the resistor, Eq. (31.10) for the inductor, Eq. (31.16) for the capacitor, and Eq. (31.25) for the source.

EXECUTE: The current and the voltages all oscillate with the same angular frequency, $\omega = 10{,}000$ rad/s, and hence with the same period, $2\pi/\omega = 2\pi/(10{,}000\ \text{rad/s}) = 6.3 \times 10^{-4}\ \text{s} = 0.63$ ms. From Eq. (31.2), the current is

$$i = I\cos\omega t = (0.10\ \text{A})\cos(10{,}000\ \text{rad/s})t$$

The resistor voltage is *in phase* with the current, so

$$v_R = V_R\cos\omega t = (30\ \text{V})\cos(10{,}000\ \text{rad/s})t$$

The inductor voltage *leads* the current by 90°, so

$$v_L = V_L\cos(\omega t + 90°) = -V_L\sin\omega t$$
$$= -(60\ \text{V})\sin(10{,}000\ \text{rad/s})t$$

The capacitor voltage *lags* the current by 90°, so

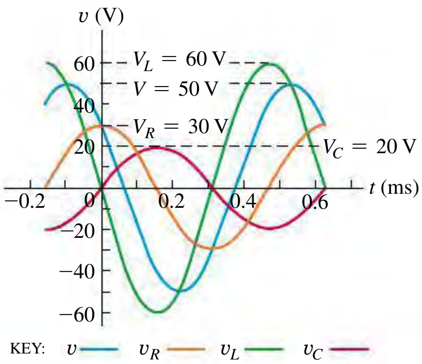
$$v_C = V_C\cos(\omega t - 90°) = V_C\sin\omega t$$
$$= (20\ \text{V})\sin(10{,}000\ \text{rad/s})t$$

We found in Example 31.4 that the source voltage (equal to the voltage across the entire combination of resistor, inductor, and capacitor) *leads* the current by $\phi = 53°$, so

$$v = V\cos(\omega t + \phi)$$
$$= (50\ \text{V})\cos\left[(10{,}000\ \text{rad/s})t + \left(\frac{2\pi\ \text{rad}}{360°}\right)(53°)\right]$$
$$= (50\ \text{V})\cos[(10{,}000\ \text{rad/s})t + 0.93\ \text{rad}]$$

EVALUATE: **Figure 31.15** graphs the four voltages versus time. The inductor voltage has a larger amplitude than the capacitor voltage because $X_L > X_C$. The *instantaneous* source voltage v is always equal to the sum of the instantaneous voltages v_R, v_L, and v_C. You should verify this by measuring the values of the voltages shown in the graph at different values of the time t.

31.15 Graphs of the source voltage v, resistor voltage v_R, inductor voltage v_L, and capacitor voltage v_C as functions of time for the situation of Example 31.4. The current, which is not shown, is in phase with the resistor voltage.



TEST YOUR UNDERSTANDING OF SECTION 31.3 Rank the following ac circuits in order of their current amplitude, from highest to lowest value. (i) The circuit in Example 31.4; (ii) the circuit in Example 31.4 with both the capacitor and inductor removed; (iii) the circuit in Example 31.4 with both the resistor and capacitor removed; (iv) the circuit in Example 31.4 with both the resistor and inductor removed.

31.4 POWER IN ALTERNATING-CURRENT CIRCUITS

Alternating currents play a central role in systems for distributing, converting, and using electrical energy, so it's important to look at power relationships in ac circuits. For an ac circuit with instantaneous current i and current amplitude I, we'll consider an element of that circuit across which the instantaneous potential difference is v with voltage amplitude V. The instantaneous power p delivered to this circuit element is

$$p = vi$$

Let's first see what this means for individual circuit elements. We'll assume in each case that $i = I\cos\omega t$.

Power in a Resistor

Suppose first that the circuit element is a *pure resistor* R, as in Fig. 31.7a; then $v = v_R$ and i are *in phase*. We obtain the graph representing p by multiplying the heights of the graphs of v and i in Fig. 31.7b at each instant. The result is the black curve in **Fig. 31.16a**. The product vi is always positive because v and i are always either both positive or both negative. Hence energy is supplied *to* the resistor at every instant for both directions of i, although the power is not constant.

The power curve for a pure resistor is symmetric about a value equal to one-half its maximum value VI, so the *average power* P_{av} is

$$P_{av} = \tfrac{1}{2}VI \qquad \text{(for a pure resistor)} \qquad (31.27)$$

An equivalent expression is

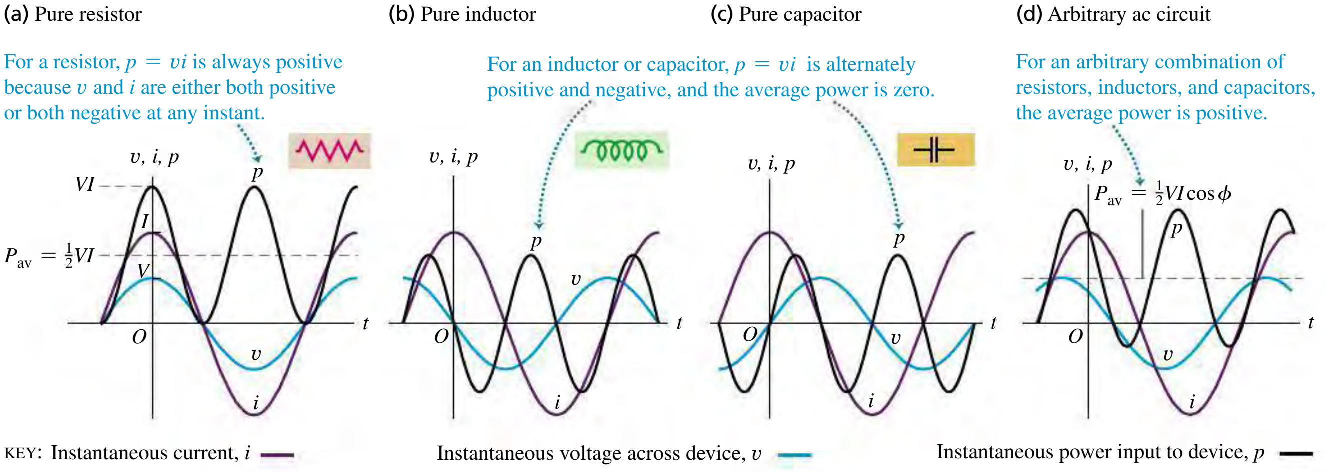
$$P_{av} = \frac{V}{\sqrt{2}}\frac{I}{\sqrt{2}} = V_{rms}I_{rms} \qquad \text{(for a pure resistor)} \qquad (31.28)$$

Also, $V_{rms} = I_{rms}R$, so we can express P_{av} by any of the equivalent forms

$$P_{av} = I_{rms}{}^2R = \frac{V_{rms}{}^2}{R} = V_{rms}I_{rms} \qquad \text{(for a pure resistor)} \qquad (31.29)$$

Note that the expressions in Eq. (31.29) have the same form as the corresponding relationships for a dc circuit, Eq. (25.18). Also note that they are valid only for pure resistors, not for more complicated combinations of circuit elements.

31.16 Graphs of current, voltage, and power as functions of time for (a) a pure resistor, (b) a pure inductor, (c) a pure capacitor, and (d) an arbitrary ac circuit that can have resistance, inductance, and capacitance.



Power in an Inductor

Next we connect the source to a pure inductor L, as in Fig. 31.8a. The voltage $v = v_L$ leads the current i by 90°. When we multiply the curves of v and i, the product vi is *negative* during the half of the cycle when v and i have *opposite* signs. The power curve, shown in Fig. 31.16b, is symmetric about the horizontal axis; it is positive half the time and negative the other half, and the average power is zero. When p is positive, energy is being supplied to set up the magnetic field in the inductor; when p is negative, the field is collapsing and the inductor is returning energy to the source. The net energy transfer over one cycle is zero.

Power in a Capacitor

Finally, we connect the source to a pure capacitor C, as in Fig. 31.9a. The voltage $v = v_C$ lags the current i by 90°. Figure 31.16c shows the power curve; the average power is again zero. Energy is supplied to charge the capacitor and is returned to the source when the capacitor discharges. The net energy transfer over one cycle is again zero.

Power in a General ac Circuit

In *any* ac circuit, with any combination of resistors, capacitors, and inductors, the voltage v across the entire circuit has some phase angle ϕ with respect to the current i. Then the instantaneous power p is given by

$$p = vi = [V\cos(\omega t + \phi)][I\cos\omega t] \tag{31.30}$$

The instantaneous power curve has the form shown in Fig. 31.16d. The area between the positive loops and the horizontal axis is greater than the area between the negative loops and the horizontal axis, and the average power is positive.

We can derive from Eq. (31.30) an expression for the *average* power P_{av} by using the identity for the cosine of the sum of two angles:

$$\begin{aligned} p &= [V(\cos\omega t\cos\phi - \sin\omega t\sin\phi)][I\cos\omega t] \\ &= VI\cos\phi\cos^2\omega t - VI\sin\phi\cos\omega t\sin\omega t \end{aligned}$$

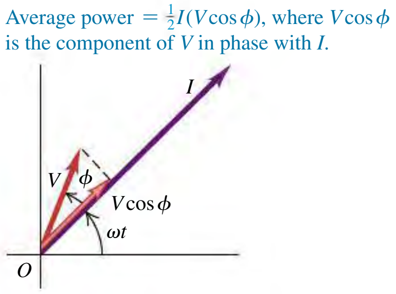
From the discussion in Section 31.1 that led to Eq. (31.4), the average value of $\cos^2\omega t$ (over one cycle) is $\frac{1}{2}$. Furthermore, $\cos\omega t\sin\omega t$ is equal to $\frac{1}{2}\sin 2\omega t$, whose average over a cycle is zero. So the average power P_{av} is

Phase angle of voltage with respect to current

Average power into a general ac circuit

$$P_{av} = \tfrac{1}{2}VI\cos\phi = V_{rms}I_{rms}\cos\phi \tag{31.31}$$

Voltage amplitude; Current amplitude; rms voltage; rms current



31.17 Using phasors to calculate the average power for an arbitrary ac circuit.

Figure 31.17 shows the general relationship of the current and voltage phasors. When v and i are in phase, so $\phi = 0$, the average power equals $\frac{1}{2}VI = V_{rms}I_{rms}$; when v and i are 90° out of phase, the average power is zero. In the general case, when v has a phase angle ϕ with respect to i, the average power equals $\frac{1}{2}I$ multiplied by $V\cos\phi$, the component of the voltage phasor that is *in phase* with the current phasor. For the L-R-C series circuit, Figs. 31.13b and 31.13c show that $V\cos\phi$ equals the voltage amplitude V_R for the resistor; hence Eq. (31.31) is the average power dissipated in the resistor. On average there is no energy flow into or out of the inductor or capacitor, so none of P_{av} goes into either of these circuit elements.

The factor $\cos\phi$ is called the **power factor** of the circuit. For a pure resistance, $\phi = 0$, $\cos\phi = 1$, and $P_{av} = V_{rms}I_{rms}$. For a pure inductor or capacitor, $\phi = \pm 90°$, $\cos\phi = 0$, and $P_{av} = 0$. For an *L-R-C* series circuit the power factor is equal to R/Z; we leave the proof of this statement to you (see Exercise 31.21).

A low power factor (large angle ϕ of lag or lead) is usually undesirable in power circuits. The reason is that for a given potential difference, a large current is needed to supply a given amount of power. This results in large i^2R losses in the transmission lines. Many types of ac machinery draw a *lagging* current; that is, the current drawn by the machinery lags the applied voltage. Hence the voltage leads the current, so $\phi > 0$ and $\cos\phi < 1$. The power factor can be corrected toward the ideal value of 1 by connecting a capacitor in parallel with the load. The current drawn by the capacitor *leads* the voltage (that is, the voltage across the capacitor lags the current), which compensates for the lagging current in the other branch of the circuit. The capacitor itself absorbs no net power from the line.

EXAMPLE 31.6 POWER IN A HAIR DRYER

SOLUTION

An electric hair dryer is rated at 1500 W (the *average* power) at 120 V (the *rms* voltage). Calculate (a) the resistance, (b) the rms current, and (c) the maximum instantaneous power. Assume that the dryer is a pure resistor. (The heating element acts as a resistor.)

SOLUTION

IDENTIFY and SET UP: We are given $P_{av} = 1500$ W and $V_{rms} = 120$ V. Our target variables are the resistance R, the rms current I_{rms}, and the maximum value p_{max} of the instantaneous power p. We solve Eq. (31.29) to find R, Eq. (31.28) to find I_{rms} from V_{rms} and P_{av}, and Eq. (31.30) to find p_{max}.

EXECUTE: (a) From Eq. (31.29), the resistance is

$$R = \frac{V_{rms}{}^2}{P_{av}} = \frac{(120\text{ V})^2}{1500\text{ W}} = 9.6\ \Omega$$

(b) From Eq. (31.28),

$$I_{rms} = \frac{P_{av}}{V_{rms}} = \frac{1500\text{ W}}{120\text{ V}} = 12.5\text{ A}$$

(c) For a pure resistor, the voltage and current are in phase and the phase angle ϕ is zero. Hence from Eq. (31.30), the instantaneous power is $p = VI\cos^2\omega t$ and the maximum instantaneous power is $p_{max} = VI$. From Eq. (31.27), this is twice the average power P_{av}, so

$$p_{max} = VI = 2P_{av} = 2(1500\text{ W}) = 3000\text{ W}$$

EVALUATE: We can use Eq. (31.7) to confirm our result in part (b): $I_{rms} = V_{rms}/R = (120\text{ V})/(9.6\ \Omega) = 12.5$ A. Note that some unscrupulous manufacturers of stereo amplifiers advertise the *peak* power output rather than the lower average value.

EXAMPLE 31.7 POWER IN AN *L-R-C* SERIES CIRCUIT

SOLUTION

For the *L-R-C* series circuit of Example 31.4, (a) calculate the power factor and (b) calculate the average power delivered to the entire circuit and to each circuit element.

SOLUTION

IDENTIFY and SET UP: We can use the results of Example 31.4. The power factor is the cosine of the phase angle ϕ, and we use Eq. (31.31) to find the average power delivered in terms of ϕ and the amplitudes of voltage and current.

EXECUTE: (a) The power factor is $\cos\phi = \cos 53° = 0.60$.

(b) From Eq. (31.31),

$$P_{av} = \tfrac{1}{2}VI\cos\phi = \tfrac{1}{2}(50\text{ V})(0.10\text{ A})(0.60) = 1.5\text{ W}$$

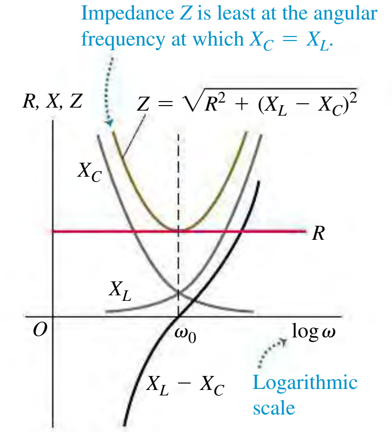
EVALUATE: Although P_{av} is the average power delivered to the *L-R-C* combination, all of this power is dissipated in the *resistor*. As Figs. 31.16b and 31.16c show, the average power delivered to a pure inductor or pure capacitor is always zero.

TEST YOUR UNDERSTANDING OF SECTION 31.4 Figure 31.16d shows that during part of a cycle of oscillation, the instantaneous power delivered to the circuit is negative. This means that energy is being extracted from the circuit. (a) Where is the energy extracted from? (i) The resistor; (ii) the inductor; (iii) the capacitor; (iv) the ac source; (v) more than one of these. (b) Where does the energy go? (i) The resistor; (ii) the inductor; (iii) the capacitor; (iv) the ac source; (v) more than one of these. ❙

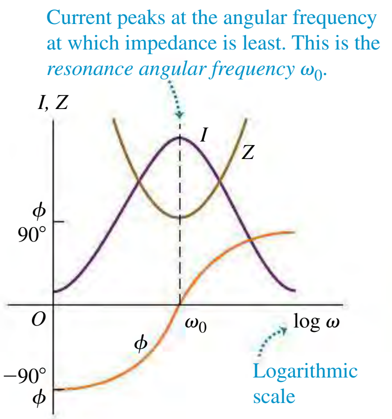
31.5 RESONANCE IN ALTERNATING-CURRENT CIRCUITS

31.18 How variations in the angular frequency of an ac circuit affect **(a)** reactance, resistance, and impedance, and **(b)** impedance, current amplitude, and phase angle.

(a) Reactance, resistance, and impedance as functions of angular frequency



(b) Impedance, current, and phase angle as functions of angular frequency



Much of the practical importance of L-R-C series circuits arises from the way in which such circuits respond to sources of different angular frequency ω. For example, one type of tuning circuit used in radio receivers is simply an L-R-C series circuit. A radio signal of any given frequency produces a current of the same frequency in the receiver circuit, but the amplitude of the current is *greatest* if the signal frequency equals the particular frequency to which the receiver circuit is "tuned." This effect is called *resonance*. The circuit is designed so that signals at other than the tuned frequency produce currents that are too small to make an audible sound come out of the radio's speakers.

To see how an L-R-C series circuit can be used in this way, suppose we connect an ac source with constant voltage amplitude V but adjustable angular frequency ω across an L-R-C series circuit. The current that appears in the circuit has the same angular frequency as the source and a current amplitude $I = V/Z$, where Z is the impedance of the L-R-C series circuit. This impedance depends on the frequency, as Eq. (31.23) shows. **Figure 31.18a** shows graphs of R, X_L, X_C, and Z as functions of ω. We have used a logarithmic angular frequency scale so that we can cover a wide range of frequencies. As the frequency increases, X_L increases and X_C decreases; hence there is always one frequency at which X_L and X_C are equal and $X_L - X_C$ is zero. At this frequency the impedance $Z = \sqrt{R^2 + (X_L - X_C)^2}$ has its *smallest* value, equal simply to the resistance R.

Circuit Behavior at Resonance

As we vary the angular frequency ω of the source, the current amplitude $I = V/Z$ varies as shown in Fig. 31.18b; the *maximum* value of I occurs at the frequency at which the impedance Z is *minimum*. This peaking of the current amplitude at a certain frequency is called **resonance.** The angular frequency ω_0 at which the resonance peak occurs is called the **resonance angular frequency.** At $\omega = \omega_0$ the inductive reactance X_L and capacitive reactance X_C are equal, so $\omega_0 L = 1/\omega_0 C$ and

Resonance angular frequency of an L-R-C series circuit

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad (31.32)$$

Inductance Capacitance

This is equal to the natural angular frequency of oscillation of an L-C circuit, which we derived in Section 30.5, Eq. (30.22). The **resonance frequency** f_0 is $\omega_0/2\pi$. At this frequency, the greatest current appears in the circuit for a given source voltage amplitude; f_0 is the frequency to which the circuit is "tuned."

It's instructive to look at what happens to the *voltages* in an L-R-C series circuit at resonance. The current at any instant is the same in L and C. The voltage across an inductor always *leads* the current by 90°, or $\frac{1}{4}$ cycle, and the voltage across a capacitor always *lags* the current by 90°. Therefore the instantaneous voltages across L and C always differ in phase by 180°, or $\frac{1}{2}$ cycle; they have opposite signs at each instant. At the resonance frequency, and *only* at the resonance frequency, $X_L = X_C$ and the voltage amplitudes $V_L = IX_L$ and $V_C = IX_C$ are *equal;* then the instantaneous voltages across L and C add to zero at each instant, and the *total* voltage v_{bd} across the L-C combination in Fig. 31.13a is exactly zero. The voltage across the resistor is then equal to the source voltage. So at the resonance frequency the circuit behaves as if the inductor and capacitor weren't there at all!

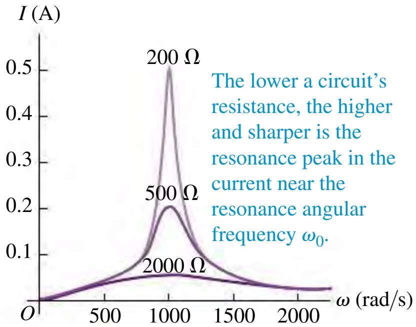
The *phase* of the voltage relative to the current is given by Eq. (31.24). At frequencies below resonance, X_C is greater than X_L; the capacitive reactance dominates, the voltage *lags* the current, and the phase angle ϕ is between 0° and $-90°$. Above resonance, the inductive reactance dominates, the voltage *leads* the current, and the phase angle ϕ is between zero and $+90°$. Figure 31.18b shows this variation of ϕ with angular frequency.

Tailoring an ac Circuit

If we can vary the inductance L or the capacitance C of a circuit, we can also vary the resonance frequency. This is exactly how a radio is "tuned" to receive a particular station. In the early days of radio this was accomplished by the use of capacitors with movable metal plates whose overlap could be varied to change C. (This is what is being done with the radio tuning knob shown in the photograph that opens this chapter.) Another approach is to vary L by using a coil with a ferrite core that slides in or out.

In an L-R-C series circuit the impedance is a minimum and the current is a maximum at the resonance frequency. The middle curve in **Fig. 31.19** is a graph of current as a function of frequency for such a circuit, with source voltage amplitude $V = 100$ V, $L = 2.0$ H, $C = 0.50\ \mu$F, and $R = 500\ \Omega$. This curve, called a *response curve* or *resonance curve,* has a peak at the resonance angular frequency $\omega_0 = \sqrt{LC} = 1000$ rad/s.

31.19 Graph of current amplitude I as a function of angular frequency ω for an L-R-C series circuit with $V = 100$ V, $L = 2.0$ H, $C = 0.50\ \mu$F, and three different values of the resistance R.



The resonance frequency is determined by L and C; what happens when we change R? Figure 31.19 also shows graphs of I as a function of ω for $R = 200\ \Omega$ and for $R = 2000\ \Omega$. The curves are similar for frequencies far away from resonance, where the impedance is dominated by X_L or X_C. But near resonance, where X_L and X_C nearly cancel each other, the curve is higher and more sharply peaked for small values of R and broader and flatter for large values of R. At resonance, $Z = R$ and $I = V/R$, so the maximum height of the curve is inversely proportional to R.

The shape of the response curve is important in the design of radio receiving circuits. The sharply peaked curve is what makes it possible to discriminate between two stations broadcasting on adjacent frequency bands. But if the peak is *too* sharp, some of the information in the received signal is lost, such as the high-frequency sounds in music. The shape of the resonance curve is also related to the overdamped and underdamped oscillations that we described in Section 30.6. A sharply peaked resonance curve corresponds to a small value of R and a lightly damped oscillating system; a broad, flat curve goes with a large value of R and a heavily damped system.

In this section we have discussed resonance in an L-R-C *series* circuit. Resonance can also occur in an ac circuit in which the inductor, resistor, and capacitor are connected in *parallel*. We leave the details to you (see Problem 31.55).

Resonance phenomena occur not just in ac circuits, but in all areas of physics. We discussed examples of resonance in *mechanical* systems in Sections 14.8 and 16.5. The amplitude of a mechanical oscillation peaks when the driving-force frequency is close to a natural frequency of the system; this is analogous to the peaking of the current in an L-R-C series circuit.

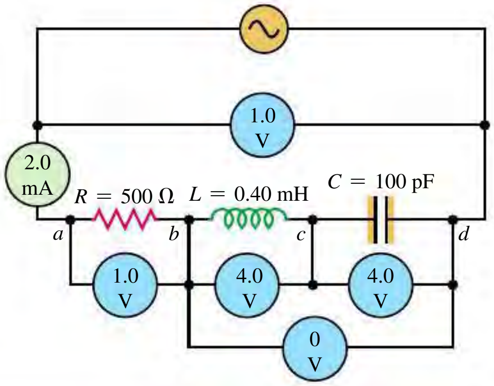
EXAMPLE 31.8 TUNING A RADIO

The series circuit in **Fig. 31.20** is similar to some radio tuning circuits. It is connected to a variable-frequency ac source with an rms terminal voltage of 1.0 V. (a) Find the resonance frequency. At the resonance frequency, find (b) the inductive reactance X_L, the capacitive reactance X_C, and the impedance Z; (c) the rms current I_{rms}; (d) the rms voltage across each circuit element.

SOLUTION

IDENTIFY and SET UP: Figure 31.20 shows an *L-R-C* series circuit, with ideal meters inserted to measure the rms current and voltages, our target variables. Equation (31.32) gives the formula for the resonance angular frequency ω_0, from which we find the resonance frequency f_0. We use Eqs. (31.12) and (31.18) to find X_L and X_C, which are equal at resonance; at resonance, from Eq. (31.23), we have $Z = R$. We use Eqs. (31.7), (31.13), and (31.19) to find the voltages across the circuit elements.

31.20 A radio tuning circuit at resonance. The circles denote rms current and voltages.



EXECUTE: (a) The values of ω_0 and f_0 are

$$\omega_0 = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{(0.40 \times 10^{-3}\,\text{H})(100 \times 10^{-12}\,\text{F})}}$$
$$= 5.0 \times 10^6\ \text{rad/s}$$
$$f_0 = 8.0 \times 10^5\ \text{Hz} = 800\ \text{kHz}$$

This frequency is in the lower part of the AM radio band.

(b) At this frequency,

$$X_L = \omega L = (5.0 \times 10^6\ \text{rad/s})(0.40 \times 10^{-3}\,\text{H}) = 2000\ \Omega$$
$$X_C = \frac{1}{\omega C} = \frac{1}{(5.0 \times 10^6\ \text{rad/s})(100 \times 10^{-12}\,\text{F})} = 2000\ \Omega$$

Since $X_L = X_C$ at resonance as stated above, $Z = R = 500\ \Omega$.

(c) From Eq. (31.26) the rms current at resonance is

$$I_{rms} = \frac{V_{rms}}{Z} = \frac{V_{rms}}{R} = \frac{1.0\ \text{V}}{500\ \Omega} = 0.0020\ \text{A} = 2.0\ \text{mA}$$

(d) The rms potential difference across the resistor is

$$V_{R\text{-rms}} = I_{rms}R = (0.0020\ \text{A})(500\ \Omega) = 1.0\ \text{V}$$

The rms potential differences across the inductor and capacitor are

$$V_{L\text{-rms}} = I_{rms}X_L = (0.0020\ \text{A})(2000\ \Omega) = 4.0\ \text{V}$$
$$V_{C\text{-rms}} = I_{rms}X_C = (0.0020\ \text{A})(2000\ \Omega) = 4.0\ \text{V}$$

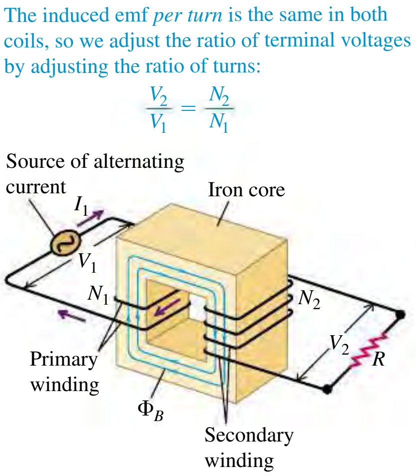
EVALUATE: The potential differences across the inductor and the capacitor have equal rms values and amplitudes, but are 180° out of phase and so add to zero at each instant. Note also that at resonance, $V_{R\text{-rms}}$ is equal to the source voltage V_{rms}, while in this example, $V_{L\text{-rms}}$ and $V_{C\text{-rms}}$ are both considerably *larger* than V_{rms}.

TEST YOUR UNDERSTANDING OF SECTION 31.5 How does the resonance frequency of an *L-R-C* series circuit change if the plates of the capacitor are brought closer together? (i) It increases; (ii) it decreases; (iii) it is unaffected.

31.6 TRANSFORMERS

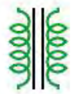
One great advantage of ac over dc for electric-power distribution is that it is much easier to step voltage levels up and down with ac than with dc. For long-distance power transmission it is desirable to use as high a voltage and as small a current as possible; this reduces i^2R losses in the transmission lines, and smaller wires can be used, saving on material costs. Present-day transmission lines operate at rms voltages of the order of 500 kV. However, safety considerations dictate relatively low voltages in generating equipment and in household and industrial power distribution. The standard voltage for household wiring is 120 V in the United States and Canada and 240 V in many other countries. The necessary voltage conversion is accomplished by using **transformers.**

31.21 Schematic diagram of an idealized step-up transformer. The primary is connected to an ac source; the secondary is connected to a device with resistance R.



How Transformers Work

Figure 31.21 shows an idealized transformer. Its key components are two coils or *windings*, electrically insulated from each other but wound on the same core. The core is typically made of a material, such as iron, with a very large relative permeability K_m. This keeps the magnetic field lines due to a current in one winding almost completely within the core. Hence almost all of these field lines pass through the other winding, maximizing the *mutual inductance* of the two windings (see Section 30.1). The winding to which power is supplied is called the **primary;** the winding from which power is delivered is called the **secondary.** The circuit symbol for a transformer with an iron core is



Here's how a transformer works. The ac source causes an alternating current in the primary, which sets up an alternating flux in the core; this induces an emf in each winding, in accordance with Faraday's law. The induced emf in the secondary gives rise to an alternating current in the secondary, and this delivers energy to the device to which the secondary is connected. All currents and emfs have the same frequency as the ac source.

Let's see how the voltage across the secondary can be made larger or smaller in amplitude than the voltage across the primary. We ignore the resistance of the windings and assume that all the magnetic field lines are confined to the iron core, so at any instant the magnetic flux Φ_B is the same in each turn of the primary and secondary windings. The primary winding has N_1 turns and the secondary winding has N_2 turns. When the magnetic flux changes because of changing currents in the two coils, the resulting induced emfs are

$$\mathcal{E}_1 = -N_1\frac{d\Phi_B}{dt} \quad \text{and} \quad \mathcal{E}_2 = -N_2\frac{d\Phi_B}{dt} \tag{31.33}$$

The flux *per turn* Φ_B is the same in both the primary and the secondary, so Eqs. (31.33) show that the induced emf *per turn* is the same in each. The ratio of the secondary emf $\mathcal{E}_2$ to the primary emf $\mathcal{E}_1$ is therefore equal at any instant to the ratio of secondary to primary turns:

$$\frac{\mathcal{E}_2}{\mathcal{E}_1} = \frac{N_2}{N_1} \tag{31.34}$$

Since $\mathcal{E}_1$ and $\mathcal{E}_2$ both oscillate with the same frequency as the ac source, Eq. (31.34) also gives the ratio of the amplitudes or of the rms values of the induced emfs. If the windings have zero resistance, the induced emfs $\mathcal{E}_1$ and $\mathcal{E}_2$ are equal to the terminal voltages across the primary and the secondary, respectively; hence

Terminal voltages in a transformer:

Secondary voltage amplitude or rms value

Number of turns in secondary

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \tag{31.35}$$

Primary voltage amplitude or rms value

Number of turns in primary

By choosing the appropriate turns ratio N_2/N_1, we may obtain any desired secondary voltage from a given primary voltage. If $N_2 > N_1$, as in Fig. 31.21, then $V_2 > V_1$ and we have a *step-up* transformer; if $N_2 < N_1$, then $V_2 < V_1$ and we have a *step-down* transformer. At a power-generating station, step-up transformers are used; the primary is connected to the power source and the secondary is connected to the transmission lines, giving the desired high voltage for transmission. Near the consumer, step-down transformers lower the voltage to a value suitable for use in home or industry (**Fig. 31.22**).

Even the relatively low voltage provided by a household wall socket is too high for many electronic devices, so a further step-down transformer is necessary.

BIO Application Dangers of ac Versus dc Voltages Alternating current at high voltage (above 500 V) is more dangerous than direct current at the same voltage. When a person touches a high-voltage dc source, it usually causes a single muscle contraction that can be strong enough to push the person away from the source. By contrast, touching a high-voltage ac source can cause a continuing muscle contraction that prevents the victim from letting go of the source. Lowering the ac voltage with a transformer reduces the risk of injury.

31.22 The cylindrical can near the top of this power pole is a step-down transformer. It converts the high-voltage ac in the power lines to low-voltage (120 V) ac, which is then distributed to the surrounding homes and businesses.

This is the role of an "ac adapter" such as those used to recharge a mobile phone or laptop computer from line voltage (**Fig. 31.23**).

31.23 An ac adapter like this one converts household ac into low-voltage dc for use in electronic devices. It contains a step-down transformer to change the line voltage to a lower value, typically 3 to 12 V, as well as diodes to convert alternating current to the direct current that small electronic devices require (see Fig. 31.3).

Energy Considerations for Transformers

If the secondary circuit is completed by a resistance R, then the amplitude or rms value of the current in the secondary circuit is $I_2 = V_2/R$. From energy considerations, the power delivered to the primary equals that taken out of the secondary (since there is no resistance in the windings), so

Terminal voltages and currents in a transformer:

$$\underbrace{V_1}_{\text{Primary voltage amplitude or rms value}}\underbrace{I_1}_{\text{Current in primary}} = \underbrace{V_2}_{\text{Secondary voltage amplitude or rms value}}\underbrace{I_2}_{\text{Current in secondary}} \tag{31.36}$$

We can combine Eqs. (31.35) and (31.36) and the relationship $I_2 = V_2/R$ to eliminate V_2 and I_2; we obtain

$$\frac{V_1}{I_1} = \frac{R}{(N_2/N_1)^2} \tag{31.37}$$

This shows that when the secondary circuit is completed through a resistance R, the result is the same as if the *source* had been connected directly to a resistance equal to R divided by the square of the turns ratio, $(N_2/N_1)^2$. In other words, the transformer "transforms" not only voltages and currents, but resistances as well. More generally, we can regard a transformer as "transforming" the *impedance* of the network to which the secondary circuit is completed.

Equation (31.37) has many practical consequences. The power supplied by a source to a resistor depends on the resistances of both the resistor and the source. It can be shown that the power transfer is greatest when the two resistances are *equal*. The same principle applies in both dc and ac circuits. When a high-impedance ac source must be connected to a low-impedance circuit, such as an audio amplifier connected to a loudspeaker, the source impedance can be *matched* to that of the circuit by the use of a transformer with an appropriate turns ratio N_2/N_1.

Real transformers always have some energy losses. (That's why an ac adapter like the one shown in Fig. 31.23 feels warm to the touch after it's been in use for a while; the transformer is heated by the dissipated energy.) The windings have some resistance, leading to i^2R losses. There are also energy losses through hysteresis in the core (see Section 28.8). Hysteresis losses are minimized by the use of soft iron with a narrow hysteresis loop.

Eddy currents (Section 29.6) also cause energy loss in transformers. Consider a section AA through an iron transformer core (**Fig. 31.24a**). Since iron is a conductor, any such section can be pictured as several conducting circuits, one

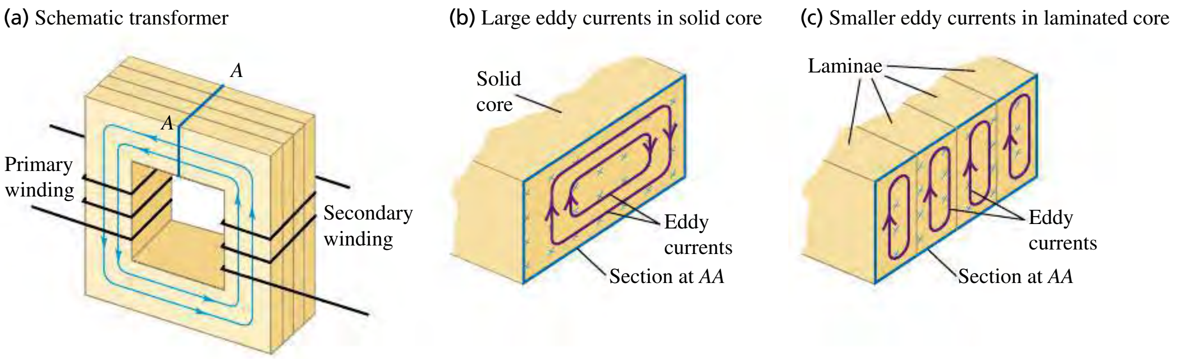
DATA *SPEAKS*

Transformers

When students were given a problem involving transformers, more than 40% gave an incorrect response. Common errors:

- Forgetting that transformers work for *alternating* current only. A transformer works on the principle that a varying current in a primary coil induces a varying current in a secondary coil. It does not work with constant direct current.
- Confusion over voltage and current. The voltage in a transformer coil is *proportional* to the number of turns in that coil. The current in a transformer coil is *inversely proportional* to the number of turns.

31.24 **(a)** Primary and secondary windings in a transformer. **(b)** Eddy currents in the iron core, shown in the cross section at AA. **(c)** Using a laminated core reduces the eddy currents.



within the other (Fig. 31.24b). The flux through each of these circuits is continually changing, so eddy currents circulate in the entire volume of the core, with lines of flow that form planes perpendicular to the flux. These eddy currents waste energy through i^2R heating and themselves set up an opposing flux.

The effects of eddy currents can be minimized by the use of a *laminated* core—that is, one built up of thin sheets, or laminae. The large electrical surface resistance of each lamina, due either to a natural coating of oxide or to an insulating varnish, effectively confines the eddy currents to individual laminae (Fig. 31.24c). The possible eddy-current paths are narrower, the induced emf in each path is smaller, and the eddy currents are greatly reduced. The alternating magnetic field exerts forces on the current-carrying laminae that cause them to vibrate back and forth; this vibration causes the characteristic "hum" of an operating transformer. You can hear this same "hum" from the magnetic ballast of a fluorescent light fixture (see Section 30.2).

Thanks to the use of soft iron cores and lamination, transformer efficiencies are usually well over 90%; in large installations they may reach 99%.

EXAMPLE 31.9 "WAKE UP AND SMELL THE (TRANSFORMER)!"

SOLUTION

A friend returns to the United States from Europe with a 960-W coffeemaker, designed to operate from a 240-V line. (a) What can she do to operate it at the USA-standard 120 V? (b) What current will the coffeemaker draw from the 120-V line? (c) What is the resistance of the coffeemaker? (The voltages are rms values.)

SOLUTION

IDENTIFY and SET UP: Our friend needs a step-up transformer to convert 120-V ac to the 240-V ac that the coffeemaker requires. We use Eq. (31.35) to determine the transformer turns ratio N_2/N_1, $P_{av} = V_{rms}I_{rms}$ for a resistor to find the current draw, and Eq. (31.37) to calculate the resistance.

EXECUTE: (a) To get $V_2 = 240$ V from $V_1 = 120$ V, the required turns ratio is $N_2/N_1 = V_2/V_1 = (240\text{ V})/(120\text{ V}) = 2$. That is, the secondary coil (connected to the coffeemaker) should have twice as many turns as the primary coil (connected to the 120-V line).

(b) We find the rms current I_1 in the 120-V primary by using $P_{av} = V_1I_1$, where P_{av} is the average power drawn by the coffeemaker and hence the power supplied by the 120-V line. (We're assuming that no energy is lost in the transformer.) Hence $I_1 = P_{av}/V_1 = (960\text{ W})/(120\text{ V}) = 8.0$ A. The secondary current is then $I_2 = P_{av}/V_2 = (960\text{ W})/(240\text{ V}) = 4.0$ A.

(c) We have $V_1 = 120$ V, $I_1 = 8.0$ A, and $N_2/N_1 = 2$, so

$$\frac{V_1}{I_1} = \frac{120\text{ V}}{8.0\text{ A}} = 15\ \Omega$$

From Eq. (31.37),

$$R = 2^2(15\ \Omega) = 60\ \Omega$$

EVALUATE: As a check, $V_2/R = (240\text{ V})/(60\ \Omega) = 4.0\text{ A} = I_2$, the same value obtained previously. You can also check this result for R by using the expression $P_{av} = V_2^2/R$ for the power drawn by the coffeemaker.

TEST YOUR UNDERSTANDING OF SECTION 31.6 Each of the following four transformers has 1000 turns in its primary. Rank the transformers from largest to smallest number of turns in the secondary. (i) Converts 120-V ac into 6.0-V ac; (ii) converts 120-V ac into 240-V ac; (iii) converts 240-V ac into 6.0-V ac; (iv) converts 240-V ac into 120-V ac. ❙

CHAPTER 31 SUMMARY

SOLUTIONS TO ALL EXAMPLES

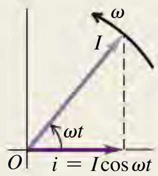
Phasors and alternating current: An alternator or ac source produces an emf that varies sinusoidally with time. A sinusoidal voltage or current can be represented by a phasor, a vector that rotates counterclockwise with constant angular velocity ω equal to the angular frequency of the sinusoidal quantity. Its projection on the horizontal axis at any instant represents the instantaneous value of the quantity.

For a sinusoidal current, the rectified average and rms (root-mean-square) currents are proportional to the current amplitude I. Similarly, the rms value of a sinusoidal voltage is proportional to the voltage amplitude V. (See Example 31.1.)

$$I_{rav} = \frac{2}{\pi} I = 0.637I \quad (31.3)$$

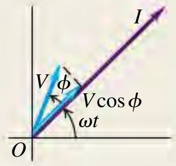
$$I_{rms} = \frac{I}{\sqrt{2}} \quad (31.4)$$

$$V_{rms} = \frac{V}{\sqrt{2}} \quad (31.5)$$



Voltage, current, and phase angle: In general, the instantaneous voltage $v = V\cos(\omega t + \phi)$ between two points in an ac circuit is not in phase with the instantaneous current passing through those points. The quantity ϕ is called the phase angle of the voltage relative to the current.

$$i = I\cos\omega t \quad (31.2)$$

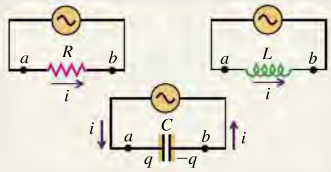


Resistance and reactance: The voltage across a resistor R is in phase with the current. The voltage across an inductor L leads the current by 90° ($\phi = +90°$), while the voltage across a capacitor C lags the current by 90° ($\phi = -90°$). The voltage amplitude across each type of device is proportional to the current amplitude I. An inductor has inductive reactance $X_L = \omega L$, and a capacitor has capacitive reactance $X_C = 1/\omega C$. (See Examples 31.2 and 31.3.)

$$V_R = IR \quad (31.7)$$

$$V_L = IX_L \quad (31.13)$$

$$V_C = IX_C \quad (31.19)$$

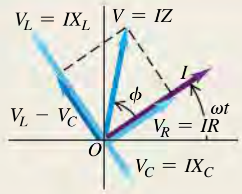


Impedance and the *L-R-C* series circuit: In a general ac circuit, the voltage and current amplitudes are related by the circuit impedance Z. In an L-R-C series circuit, the values of L, R, C, and the angular frequency ω determine the impedance and the phase angle ϕ of the voltage relative to the current. (See Examples 31.4 and 31.5.)

$$V = IZ \quad (31.22)$$

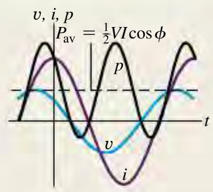
$$Z = \sqrt{R^2 + [\omega L - (1/\omega C)]^2} \quad (31.23)$$

$$\tan\phi = \frac{\omega L - 1/\omega C}{R} \quad (31.24)$$



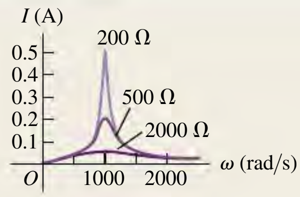
Power in ac circuits: The average power input P_{av} to an ac circuit depends on the voltage and current amplitudes (or, equivalently, their rms values) and the phase angle ϕ of the voltage relative to the current. The quantity $\cos\phi$ is called the power factor. (See Examples 31.6 and 31.7.)

$$P_{av} = \tfrac{1}{2} VI\cos\phi = V_{rms} I_{rms} \cos\phi \quad (31.31)$$



Resonance in ac circuits: In an L-R-C series circuit, the current becomes maximum and the impedance becomes minimum at an angular frequency called the resonance angular frequency. This phenomenon is called resonance. At resonance the voltage and current are in phase, and the impedance Z is equal to the resistance R. (See Example 31.8.)

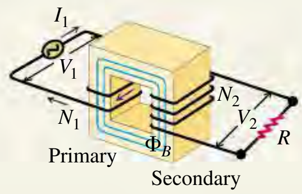
$$\omega_0 = \frac{1}{\sqrt{LC}} \quad (31.32)$$



Transformers: A transformer is used to transform the voltage and current levels in an ac circuit. In an ideal transformer with no energy losses, if the primary winding has N_1 turns and the secondary winding has N_2 turns, the amplitudes (or rms values) of the two voltages are related by Eq. (31.35). The amplitudes (or rms values) of the primary and secondary voltages and currents are related by Eq. (31.36). (See Example 31.9.)

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \tag{31.35}$$

$$V_1 I_1 = V_2 I_2 \tag{31.36}$$



BRIDGING PROBLEM AN ALTERNATING-CURRENT CIRCUIT

SOLUTION

A series circuit like the circuit in Fig. 31.13a consists of a 1.50-mH inductor, a 125-Ω resistor, and a 25.0-nF capacitor connected across an ac source having an rms voltage of 35.0 V and variable frequency. (a) At what angular frequencies will the current amplitude be equal to $\frac{1}{3}$ of its maximum possible value? (b) At the frequencies in part (a), what are the current amplitude and the voltage amplitude across each circuit element (including the ac source)?

SOLUTION GUIDE

IDENTIFY and SET UP

1. The maximum current amplitude occurs at the resonance angular frequency. This problem concerns the angular frequencies at which the current amplitude is one-third of that maximum.
2. Choose the equation that will allow you to find the angular frequencies in question, and choose the equations that you will then use to find the current and voltage amplitudes at each angular frequency.

EXECUTE

3. Find the impedance at the angular frequencies in part (a); then solve for the values of angular frequency.
4. Find the voltage amplitude across the source and the current amplitude for each of the angular frequencies in part (a). (*Hint:* Be careful to distinguish between *amplitude* and *rms value*.)
5. Use the results of steps 3 and 4 to find the reactances at each angular frequency. Then calculate the voltage amplitudes for the resistor, inductor, and capacitor.

EVALUATE

6. Are any voltage amplitudes greater than the voltage amplitude of the source? If so, does this mean your results are in error?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q31.1 Household electric power in most of western Europe is supplied at 240 V, rather than the 120 V that is standard in the United States and Canada. What are the advantages and disadvantages of each system?

Q31.2 The current in an ac power line changes direction 120 times per second, and its average value is zero. Explain how it is possible for power to be transmitted in such a system.

Q31.3 In an ac circuit, why is the average power for an inductor and a capacitor zero, but not for a resistor?

Q31.4 Equation (31.14) was derived by using the relationship $i = dq/dt$ between the current and the charge on the capacitor. In Fig. 31.9a the positive counterclockwise current increases the charge on the capacitor. When the charge on the left plate is positive but decreasing in time, is $i = dq/dt$ still correct or should it be $i = -dq/dt$? Is $i = dq/dt$ still correct when the right-hand plate has positive charge that is increasing or decreasing in magnitude? Explain.

Q31.5 Fluorescent lights often use an inductor, called a ballast, to limit the current through the tubes. Why is it better to use an inductor rather than a resistor for this purpose?

Q31.6 Equation (31.9) says that $v_{ab} = L\,di/dt$ (see Fig. 31.8a). Using Faraday's law, explain why point a is at higher potential than point b when i is in the direction shown in Fig. 31.8a and is increasing in magnitude. When i is counterclockwise and decreasing in magnitude, is $v_{ab} = L\,di/dt$ still correct, or should it be $v_{ab} = -L\,di/dt$? Is $v_{ab} = L\,di/dt$ still correct when i is clockwise and increasing or decreasing in magnitude? Explain.

Q31.7 Is it possible for the power factor of an L-R-C series ac circuit to be zero? Justify your answer on *physical* grounds.

Q31.8 In an L-R-C series circuit, can the instantaneous voltage across the capacitor exceed the source voltage at that same instant? Can this be true for the instantaneous voltage across the inductor? Across the resistor? Explain.

Q31.9 In an L-R-C series circuit, what are the phase angle ϕ and power factor $\cos\phi$ when the resistance is much smaller than the inductive or capacitive reactance and the circuit is operated far from resonance? Explain.

Q31.10 When an L-R-C series circuit is connected across a 120-V ac line, the voltage rating of the capacitor may be exceeded even if it is rated at 200 or 400 V. How can this be?

Q31.11 In Example 31.6 (Section 31.4), a hair dryer is treated as a pure resistor. But because there are coils in the heating element and in the motor that drives the blower fan, a hair dryer also has inductance. Qualitatively, does including an inductance increase or decrease the values of R, I_{rms}, and P?

Q31.12 A light bulb and a parallel-plate capacitor with air between the plates are connected in series to an ac source. What happens to the brightness of the bulb when a dielectric is inserted between the plates of the capacitor? Explain.

Q31.13 A coil of wire wrapped on a hollow tube and a light bulb are connected in series to an ac source. What happens to the brightness of the bulb when an iron rod is inserted in the tube?

Q31.14 A circuit consists of a light bulb, a capacitor, and an inductor connected in series to an ac source. What happens to the brightness of the bulb when the inductor is omitted? When the inductor is left in the circuit but the capacitor is omitted? Explain.

Q31.15 A circuit consists of a light bulb, a capacitor, and an inductor connected in series to an ac source. Is it possible for both the capacitor and the inductor to be removed and the brightness of the bulb to remain the same? Explain.

Q31.16 Can a transformer be used with dc? Explain. What happens if a transformer designed for 120-V ac is connected to a 120-V dc line?

Q31.17 An ideal transformer has N_1 windings in the primary and N_2 windings in its secondary. If you double only the number of secondary windings, by what factor does (a) the voltage amplitude in the secondary change, and (b) the effective resistance of the secondary circuit change?

Q31.18 An inductor, a capacitor, and a resistor are all connected in series across an ac source. If the resistance, inductance, and capacitance are all doubled, by what factor does each of the following quantities change? Indicate whether they increase or decrease: (a) the resonance angular frequency; (b) the inductive reactance; (c) the capacitive reactance. (d) Does the impedance double?

Q31.19 You want to double the resonance angular frequency of an L-R-C series circuit by changing only the *pertinent* circuit elements all by the same factor. (a) Which ones should you change? (b) By what factor should you change them?

EXERCISES

Section 31.1 Phasors and Alternating Currents

31.1 • You have a special light bulb with a *very* delicate wire filament. The wire will break if the current in it ever exceeds 1.50 A, even for an instant. What is the largest root-mean-square current you can run through this bulb?

31.2 • A sinusoidal current $i = I\cos\omega t$ has an rms value $I_{rms} = 2.10$ A. (a) What is the current amplitude? (b) The current is passed through a full-wave rectifier circuit. What is the rectified average current? (c) Which is larger: I_{rms} or I_{rav}? Explain, using graphs of i^2 and of the rectified current.

31.3 • The voltage across the terminals of an ac power supply varies with time according to Eq. (31.1). The voltage amplitude is $V = 45.0$ V. What are (a) the root-mean-square potential difference V_{rms} and (b) the average potential difference V_{av} between the two terminals of the power supply?

Section 31.2 Resistance and Reactance

31.4 • A capacitor is connected across an ac source that has voltage amplitude 60.0 V and frequency 80.0 Hz. (a) What is the phase angle ϕ for the source voltage relative to the current? Does the source voltage lag or lead the current? (b) What is the capacitance C of the capacitor if the current amplitude is 5.30 A?

31.5 • An inductor with $L = 9.50$ mH is connected across an ac source that has voltage amplitude 45.0 V. (a) What is the phase angle ϕ for the source voltage relative to the current? Does the source voltage lag or lead the current? (b) What value for the frequency of the source results in a current amplitude of 3.90 A?

31.6 • A capacitance C and an inductance L are operated at the same angular frequency. (a) At what angular frequency will they have the same reactance? (b) If $L = 5.00$ mH and $C = 3.50\ \mu$F, what is the numerical value of the angular frequency in part (a), and what is the reactance of each element?

31.7 • **Kitchen Capacitance.** The wiring for a refrigerator contains a starter capacitor. A voltage of amplitude 170 V and frequency 60.0 Hz applied across the capacitor is to produce a current amplitude of 0.850 A through the capacitor. What capacitance C is required?

31.8 • (a) Compute the reactance of a 0.450-H inductor at frequencies of 60.0 Hz and 600 Hz. (b) Compute the reactance of a 2.50-μF capacitor at the same frequencies. (c) At what frequency is the reactance of a 0.450-H inductor equal to that of a 2.50-μF capacitor?

31.9 • (a) What is the reactance of a 3.00-H inductor at a frequency of 80.0 Hz? (b) What is the inductance of an inductor whose reactance is 120 Ω at 80.0 Hz? (c) What is the reactance of a 4.00-μF capacitor at a frequency of 80.0 Hz? (d) What is the capacitance of a capacitor whose reactance is 120 Ω at 80.0 Hz?

31.10 • **A Radio Inductor.** You want the current amplitude through a 0.450-mH inductor (part of the circuitry for a radio receiver) to be 1.80 mA when a sinusoidal voltage with amplitude 12.0 V is applied across the inductor. What frequency is required?

31.11 •• A 0.180-H inductor is connected in series with a 90.0-Ω resistor and an ac source. The voltage across the inductor is $v_L = -(12.0\text{ V})\sin[(480\text{ rad/s})t]$. (a) Derive an expression for the voltage v_R across the resistor. (b) What is v_R at $t = 2.00$ ms?

31.12 •• A 250-Ω resistor is connected in series with a 4.80-μF capacitor and an ac source. The voltage across the capacitor is $v_C = (7.60\text{ V})\sin[(120\text{ rad/s})t]$. (a) Determine the capacitive reactance of the capacitor. (b) Derive an expression for the voltage v_R across the resistor.

31.13 •• A 150-Ω resistor is connected in series with a 0.250-H inductor and an ac source. The voltage across the resistor is $v_R = (3.80\text{ V})\cos[(720\text{ rad/s})t]$. (a) Derive an expression for the circuit current. (b) Determine the inductive reactance of the inductor. (c) Derive an expression for the voltage v_L across the inductor.

Section 31.3 The *L-R-C* Series Circuit

31.14 • You have a 200-Ω resistor, a 0.400-H inductor, and a 6.00-μF capacitor. Suppose you take the resistor and inductor and make a series circuit with a voltage source that has voltage amplitude 30.0 V and an angular frequency of 250 rad/s. (a) What is the impedance of the circuit? (b) What is the current amplitude? (c) What are the voltage amplitudes across the resistor and across the inductor? (d) What is the phase angle ϕ of the source voltage with respect to the current? Does the source voltage lag or lead the current? (e) Construct the phasor diagram.

31.15 • The resistor, inductor, capacitor, and voltage source described in Exercise 31.14 are connected to form an *L-R-C* series circuit. (a) What is the impedance of the circuit? (b) What is the current amplitude? (c) What is the phase angle of the source voltage with respect to the current? Does the source voltage lag or lead the current? (d) What are the voltage amplitudes across the resistor, inductor, and capacitor? (e) Explain how it is possible for the voltage amplitude across the capacitor to be greater than the voltage amplitude across the source.

31.16 •• A 200-Ω resistor, 0.900-H inductor, and 6.00-μF capacitor are connected in series across a voltage source that has voltage amplitude 30.0 V and an angular frequency of 250 rad/s. (a) What are v, v_R, v_L, and v_C at $t = 20.0$ ms? Compare $v_R + v_L + v_C$ to v at this instant. (b) What are V_R, V_L, and V_C? Compare V to $V_R + V_L + V_C$. Explain why these two quantities are not equal.

31.17 • In an *L-R-C* series circuit, the rms voltage across the resistor is 30.0 V, across the capacitor it is 90.0 V, and across the inductor it is 50.0 V. What is the rms voltage of the source?

Section 31.4 Po wer in Alternating-Current Circuits

31.18 •• A resistor with $R = 300\ \Omega$ and an inductor are connected in series across an ac source that has voltage amplitude 500 V. The rate at which electrical energy is dissipated in the resistor is 286 W. What is (a) the impedance Z of the circuit; (b) the amplitude of the voltage across the inductor; (c) the power factor?

31.19 • The power of a certain CD player operating at 120 V rms is 20.0 W. Assuming that the CD player behaves like a pure resistor, find (a) the maximum instantaneous power; (b) the rms current; (c) the resistance of this player.

31.20 •• In an *L-R-C* series circuit, the components have the following values: $L = 20.0$ mH, $C = 140$ nF, and $R = 350\ \Omega$. The generator has an rms voltage of 120 V and a frequency of 1.25 kHz. Determine (a) the power supplied by the generator and (b) the power dissipated in the resistor.

31.21 • (a) Show that for an *L-R-C* series circuit the power factor is equal to R/Z. (b) An *L-R-C* series circuit has phase angle -31.5°. The voltage amplitude of the source is 90.0 V. What is the voltage amplitude across the resistor?

31.22 • (a) Use the results of part (a) of Exercise 31.21 to show that the average power delivered by the source in an *L-R-C* series circuit is given by $P_{av} = I_{rms}{}^2 R$. (b) An *L-R-C* series circuit has $R = 96.0\ \Omega$, and the amplitude of the voltage across the resistor is 36.0 V. What is the average power delivered by the source?

31.23 • An *L-R-C* series circuit with $L = 0.120$ H, $R = 240\ \Omega$, and $C = 7.30\ \mu$F carries an rms current of 0.450 A with a frequency of 400 Hz. (a) What are the phase angle and power factor for this circuit? (b) What is the impedance of the circuit? (c) What is the rms voltage of the source? (d) What average power is delivered by the source? (e) What is the average rate at which electrical energy is converted to thermal energy in the resistor? (f) What is the average rate at which electrical energy is dissipated (converted to other forms) in the capacitor? (g) In the inductor?

31.24 •• An *L-R-C* series circuit is connected to a 120-Hz ac source that has $V_{rms} = 80.0$ V. The circuit has a resistance of 75.0 Ω and an impedance at this frequency of 105 Ω. What average power is delivered to the circuit by the source?

31.25 •• A series ac circuit contains a 250-Ω resistor, a 15-mH inductor, a 3.5-μF capacitor, and an ac power source of voltage amplitude 45 V operating at an angular frequency of 360 rad/s. (a) What is the power factor of this circuit? (b) Find the average power delivered to the entire circuit. (c) What is the average power delivered to the resistor, to the capacitor, and to the inductor?

Section 31.5 Resonance in Alternating-Current Circuits

31.26 •• In an *L-R-C* series circuit the source is operated at its resonant angular frequency. At this frequency, the reactance X_C of the capacitor is 200 Ω and the voltage amplitude across the capacitor is 600 V. The circuit has $R = 300\ \Omega$. What is the voltage amplitude of the source?

31.27 • **Analyzing an *L-R-C* Circuit.** You have a 200-Ω resistor, a 0.400-H inductor, a 5.00-μF capacitor, and a variable-frequency ac source with an amplitude of 3.00 V. You connect all four elements together to form a series circuit. (a) At what frequency will the current in the circuit be greatest? What will be the current amplitude at this frequency? (b) What will be the current amplitude at an angular frequency of 400 rad/s? At this frequency, will the source voltage lead or lag the current?

31.28 • An *L-R-C* series circuit is constructed using a 175-Ω resistor, a 12.5-μF capacitor, and an 8.00-mH inductor, all connected across an ac source having a variable frequency and a voltage amplitude of 25.0 V. (a) At what angular frequency will the impedance be smallest, and what is the impedance at this frequency? (b) At the angular frequency in part (a), what is the maximum current through the inductor? (c) At the angular frequency in part (a), find the potential difference across the ac source, the resistor, the capacitor, and the inductor at the instant that the current is equal to one-half its greatest positive value. (d) In part (c), how are the potential differences across the resistor, inductor, and capacitor related to the potential difference across the ac source?

31.29 • In an *L-R-C* series circuit, $R = 300\ \Omega$, $L = 0.400$ H, and $C = 6.00 \times 10^{-8}$ F. When the ac source operates at the resonance frequency of the circuit, the current amplitude is 0.500 A. (a) What is the voltage amplitude of the source? (b) What is the amplitude of the voltage across the resistor, across the inductor, and across the capacitor? (c) What is the average power supplied by the source?

31.30 • An *L-R-C* series circuit consists of a source with voltage amplitude 120 V and angular frequency 50.0 rad/s, a resistor with $R = 400\ \Omega$, an inductor with $L = 3.00$ H, and a capacitor with capacitance C. (a) For what value of C will the current amplitude in the circuit be a maximum? (b) When C has the value calculated in part (a), what is the amplitude of the voltage across the inductor?

31.31 • In an *L-R-C* series circuit, $R = 150\ \Omega$, $L = 0.750$ H, and $C = 0.0180\ \mu$F. The source has voltage amplitude $V = 150$ V and a frequency equal to the resonance frequency of the circuit. (a) What is the power factor? (b) What is the average power delivered by the source? (c) The capacitor is replaced by one with $C = 0.0360\ \mu$F and the source frequency is adjusted to the new resonance value. Then what is the average power delivered by the source?

31.32 • In an *L-R-C* series circuit, $R = 400\ \Omega$, $L = 0.350$ H, and $C = 0.0120\ \mu$F. (a) What is the resonance angular frequency of the circuit? (b) The capacitor can withstand a peak voltage of 670 V. If the voltage source operates at the resonance frequency, what maximum voltage amplitude can it have if the maximum capacitor voltage is not exceeded?

31.33 •• In an *L-R-C* series circuit, $L = 0.280$ H and $C = 4.00\ \mu$F. The voltage amplitude of the source is 120 V. (a) What is the resonance angular frequency of the circuit? (b) When the source operates at the resonance angular frequency, the current amplitude in the circuit is 1.70 A. What is the resistance R of the resistor? (c) At the resonance angular frequency, what are the peak voltages across the inductor, the capacitor, and the resistor?

Section 31.6 Transformers

31.34 • **Off to Europe!** You plan to take your hair dryer to Europe, where the electrical outlets put out 240 V instead of the 120 V seen in the United States. The dryer puts out 1600 W at 120 V. (a) What could you do to operate your dryer via the 240-V line in Europe? (b) What current will your dryer draw from a European outlet? (c) What resistance will your dryer appear to have when operated at 240 V?

31.35 • **A Step-Down Transformer.** A transformer connected to a 120-V (rms) ac line is to supply 12.0 V (rms) to a portable electronic device. The load resistance in the secondary is 5.00 Ω. (a) What should the ratio of primary to secondary turns of the transformer be? (b) What rms current must the secondary supply? (c) What average power is delivered to the load? (d) What resistance connected directly across the 120-V line would draw the same power as the transformer? Show that this is equal to 5.00 Ω times the square of the ratio of primary to secondary turns.

31.36 • **A Step-Up Transformer.** A transformer connected to a 120-V (rms) ac line is to supply 13,000 V (rms) for a neon sign. To reduce shock hazard, a fuse is to be inserted in the primary circuit; the fuse is to blow when the rms current in the secondary circuit exceeds 8.50 mA. (a) What is the ratio of secondary to primary turns of the transformer? (b) What power must be supplied to the transformer when the rms secondary current is 8.50 mA? (c) What current rating should the fuse in the primary circuit have?

PROBLEMS

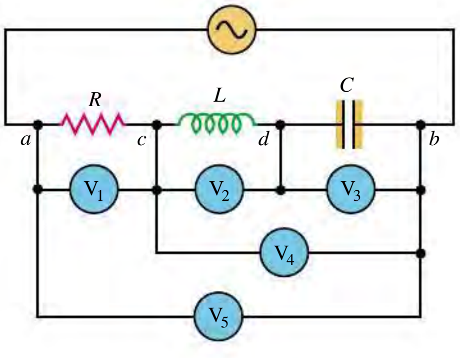
31.37 • A coil has a resistance of 48.0 Ω. At a frequency of 80.0 Hz the voltage across the coil leads the current in it by 52.3°. Determine the inductance of the coil.

31.38 •• When a solenoid is connected to a 48.0-V dc battery that has negligible internal resistance, the current in the solenoid is 5.50 A. When this solenoid is connected to an ac source that has voltage amplitude 48.0 V and angular frequency 20.0 rad/s, the current in the solenoid is 3.60 A. What is the inductance of this solenoid?

31.39 •• An *L-R-C* series circuit has $C = 4.80\ \mu\text{F}$, $L = 0.520$ H, and source voltage amplitude $V = 56.0$ V. The source is operated at the resonance frequency of the circuit. If the voltage across the capacitor has amplitude 80.0 V, what is the value of R for the resistor in the circuit?

31.40 •• Five infinite-impedance voltmeters, calibrated to read rms values, are connected as shown in **Fig. P31.40**. Let $R = 200\ \Omega$, $L = 0.400$ H, $C = 6.00\ \mu\text{F}$, and $V = 30.0$ V. What is the reading of each voltmeter if (a) $\omega = 200$ rad/s and (b) $\omega = 1000$ rad/s?

Figure **P31.40**



31.41 •• **CP** A parallel-plate capacitor having square plates 4.50 cm on each side and 8.00 mm apart is placed in series with the following: an ac source of angular frequency 650 rad/s and voltage amplitude 22.5 V; a 75.0-Ω resistor; and an ideal solenoid that is 9.00 cm long, has a circular cross section 0.500 cm in diameter, and carries 125 coils per centimeter. What is the resonance angular frequency of this circuit? (See Exercise 30.15.)

31.42 •• **CP** A toroidal solenoid has 2900 closely wound turns, cross-sectional area 0.450 cm^2, mean radius 9.00 cm, and resistance $R = 2.80\ \Omega$. Ignore the variation of the magnetic field across the cross section of the solenoid. What is the amplitude of the current in the solenoid if it is connected to an ac source that has voltage amplitude 24.0 V and frequency 495 Hz?

31.43 •• A series circuit has an impedance of 60.0 Ω and a power factor of 0.720 at 50.0 Hz. The source voltage lags the current. (a) What circuit element, an inductor or a capacitor, should be placed in series with the circuit to raise its power factor? (b) What size element will raise the power factor to unity?

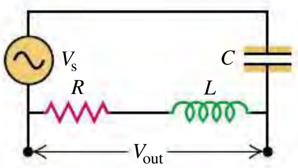
31.44 • A large electromagnetic coil is connected to a 120-Hz ac source. The coil has resistance 400 Ω, and at this source frequency the coil has inductive reactance 250 Ω. (a) What is the inductance of the coil? (b) What must the rms voltage of the source be if the coil is to consume an average electrical power of 450 W?

31.45 •• In an *L-R-C* series circuit, $R = 300\ \Omega$, $X_C = 300\ \Omega$, and $X_L = 500\ \Omega$. The average electrical power consumed in the resistor is 60.0 W. (a) What is the power factor of the circuit? (b) What is the rms voltage of the source?

31.46 • At a frequency ω_1 the reactance of a certain capacitor equals that of a certain inductor. (a) If the frequency is changed to $\omega_2 = 2\omega_1$, what is the ratio of the reactance of the inductor to that of the capacitor? Which reactance is larger? (b) If the frequency is changed to $\omega_3 = \omega_1/3$, what is the ratio of the reactance of the inductor to that of the capacitor? Which reactance is larger? (c) If the capacitor and inductor are placed in series with a resistor of resistance R to form an *L-R-C* series circuit, what will be the resonance angular frequency of the circuit?

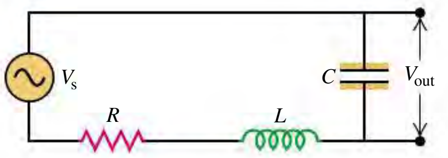
31.47 •• **A High-Pass Filter.** One application of *L-R-C* series circuits is to high-pass or low-pass filters, which filter out either the low- or high-frequency components of a signal. A high-pass filter is shown in **Fig. P31.47**, where the output voltage is taken across the *L-R* combination. (The *L-R* combination represents an inductive coil that also has resistance due to the large length of wire in the coil.) Derive an expression for V_{out}/V_s, the ratio of the output and source voltage amplitudes, as a function of the angular frequency ω of the source. Show that when ω is small, this ratio is proportional to ω and thus is small, and show that the ratio approaches unity in the limit of large frequency.

Figure **P31.47**



31.48 •• **A Low-Pass Filter.** **Figure P31.48** shows a low-pass filter (see Problem 31.47); the output voltage is taken across the capacitor in an *L-R-C* series circuit. Derive an expression for

Figure **P31.48**



V_{out}/V_s, the ratio of the output and source voltage amplitudes, as a function of the angular frequency ω of the source. Show that when ω is large, this ratio is proportional to ω^{-2} and thus is very small, and show that the ratio approaches unity in the limit of small frequency.

31.49 ••• An L-R-C series circuit is connected to an ac source of constant voltage amplitude V and variable angular frequency ω. (a) Show that the current amplitude, as a function of ω, is

$$I = \frac{V}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}}$$

(b) Show that the average power dissipated in the resistor is

$$P = \frac{V^2 R/2}{R^2 + (\omega L - 1/\omega C)^2}$$

(c) Show that I and P are *both* maximum when $\omega = 1/\sqrt{LC}$, the resonance frequency of the circuit. (d) Graph P as a function of ω for $V = 100\text{ V}$, $R = 200\ \Omega$, $L = 2.0\text{ H}$, and $C = 0.50\ \mu\text{F}$. Compare to the light purple curve in Fig. 31.19. Discuss the behavior of I and P in the limits $\omega = 0$ and $\omega \rightarrow \infty$.

31.50 ••• An L-R-C series circuit is connected to an ac source of constant voltage amplitude V and variable angular frequency ω. Using the results of Problem 31.49, find an expression for (a) the amplitude V_L of the voltage across the inductor as a function of ω and (b) the amplitude V_C of the voltage across the capacitor as a function of ω. (c) Graph V_L and V_C as functions of ω for $V = 100\text{ V}$, $R = 200\ \Omega$, $L = 2.0\text{ H}$, and $C = 0.50\ \mu\text{F}$. (d) Discuss the behavior of V_L and V_C in the limits $\omega = 0$ and $\omega \rightarrow \infty$. For what value of ω is $V_L = V_C$? What is the significance of this value of ω?

31.51 •• In an L-R-C series circuit the magnitude of the phase angle is 54.0°, with the source voltage lagging the current. The reactance of the capacitor is 350 Ω, and the resistor resistance is 180 Ω. The average power delivered by the source is 140 W. Find (a) the reactance of the inductor; (b) the rms current; (c) the rms voltage of the source.

31.52 •• In an L-R-C series circuit, the phase angle is 40.0°, with the source voltage leading the current. The reactance of the capacitor is 400 Ω, and the resistance of the resistor is 200 Ω. The average power delivered by the source is 150 W. Find (a) the reactance of the inductor, (b) the rms current, (c) the rms voltage of the source.

31.53 • An L-R-C series circuit has $R = 500\ \Omega$, $L = 2.00\text{ H}$, $C = 0.500\ \mu\text{F}$, and $V = 100\text{ V}$. (a) For $\omega = 800\text{ rad/s}$, calculate V_R, V_L, V_C, and ϕ. Using a single set of axes, graph v, v_R, v_L, and v_C as functions of time. Include two cycles of v on your graph. (b) Repeat part (a) for $\omega = 1000\text{ rad/s}$. (c) Repeat part (a) for $\omega = 1250\text{ rad/s}$.

31.54 •• **The *L-R-C* Parallel Circuit.** A resistor, an inductor, and a capacitor are connected in parallel to an ac source with voltage amplitude V and angular frequency ω. Let the source voltage be given by $v = V\cos\omega t$. (a) Show that each of the instantaneous voltages v_R, v_L, and v_C at any instant is equal to v and that $i = i_R + i_L + i_C$, where i is the current through the source and i_R, i_L, and i_C are the currents through the resistor, inductor, and capacitor, respectively. (b) What are the phases of i_R, i_L, and i_C with respect to v? Use current phasors to represent i, i_R, i_L, and i_C. In a phasor diagram, show the phases of these four currents with respect to v. (c) Use the phasor diagram of part (b) to show that the current amplitude I for the current i through the source is $I = \sqrt{I_R^2 + (I_C - I_L)^2}$. (d) Show that the result of part (c) can be written as $I = V/Z$, with $1/Z = \sqrt{(1/R^2) + [\omega C - (1/\omega L)]^2}$.

31.55 •• The impedance of an L-R-C parallel circuit was derived in Problem 31.54. (a) Show that at the resonance angular frequency $\omega_0 = 1/\sqrt{LC}$, the impedance Z is a maximum and therefore the current through the ac source is a minimum. (b) A 100-Ω resistor, a 0.100-μF capacitor, and a 0.300-H inductor are connected in parallel to a voltage source with amplitude 240 V. What is the resonance angular frequency? For this circuit, what is (c) the maximum current through the source at the resonance frequency; (d) the maximum current in the resistor at resonance; (e) the maximum current in the inductor at resonance; (f) the maximum current in the branch containing the capacitor at resonance?

31.56 •• A 400-Ω resistor and a 6.00-μF capacitor are connected in parallel to an ac generator that supplies an rms voltage of 180 V at an angular frequency of 360 rad/s. Use the results of Problem 31.54. Note that since there is no inductor in this circuit, the $1/\omega L$ term is not present in the expression for $1/Z$. Find (a) the current amplitude in the resistor; (b) the current amplitude in the capacitor; (c) the phase angle of the source current with respect to the source voltage; (d) the amplitude of the current through the generator. (e) Does the source current lag or lead the source voltage?

31.57 ••• An L-R-C series circuit consists of a 2.50-μF capacitor, a 5.00-mH inductor, and a 75.0-Ω resistor connected across an ac source of voltage amplitude 15.0 V having variable frequency. (a) Under what circumstances is the average power delivered to the circuit equal to $\frac{1}{2}V_{rms}I_{rms}$? (b) Under the conditions of part (a), what is the average power delivered to each circuit element and what is the maximum current through the capacitor?

31.58 •• An L-R-C series circuit has $R = 60.0\ \Omega$, $L = 0.800\text{ H}$, and $C = 3.00 \times 10^{-4}\text{ F}$. The ac source has voltage amplitude 90.0 V and angular frequency 120 rad/s. (a) What is the maximum energy stored in the inductor? (b) When the energy stored in the inductor is a maximum, how much energy is stored in the capacitor? (c) What is the maximum energy stored in the capacitor?

31.59 • In an L-R-C series circuit, the source has a voltage amplitude of 120 V, $R = 80.0\ \Omega$, and the reactance of the capacitor is 480 Ω. The voltage amplitude across the capacitor is 360 V. (a) What is the current amplitude in the circuit? (b) What is the impedance? (c) What two values can the reactance of the inductor have? (d) For which of the two values found in part (c) is the angular frequency less than the resonance angular frequency? Explain.

31.60 •• In an L-R-C series ac circuit, the source has a voltage amplitude of 240 V, $R = 90.0\ \Omega$, and the reactance of the inductor is 320 Ω. The voltage amplitude across the resistor is 135 V. (a) What is the current amplitude in the circuit? (b) What is the voltage amplitude across the inductor? (c) What two values can the reactance of the capacitor have? (d) For which of the two values found in part (c) is the angular frequency less than the resonance angular frequency? Explain.

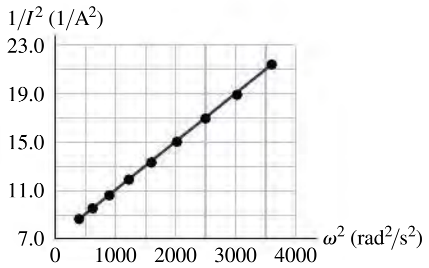
31.61 • A resistance R, capacitance C, and inductance L are connected in series to a voltage source with amplitude V and variable angular frequency ω. If $\omega = \omega_0$, the resonance angular frequency, find (a) the maximum current in the resistor; (b) the maximum voltage across the capacitor; (c) the maximum voltage across the inductor; (d) the maximum energy stored in the capacitor; (e) the maximum energy stored in the inductor. Give your answers in terms of R, C, L, and V.

31.62 •• **The Resonance Width.** Consider an L-R-C series circuit with a 1.80-H inductor, a 0.900-μF capacitor, and a 300-Ω resistor. The source has terminal rms voltage $V_{rms} = 60.0\text{ V}$ and variable angular frequency ω. (a) What is the resonance angular frequency ω_0 of the circuit? (b) What is the rms current through the circuit at resonance, $I_{rms\text{-}0}$? (c) For what two values of the angular frequency, ω_1 and ω_2, is the rms current half the resonance value? (d) The quantity $|\omega_1 - \omega_2|$ defines the *resonance width*. Calculate $I_{\text{rms-0}}$ and the resonance width for $R = 300\ \Omega$, $30.0\ \Omega$, and $3.00\ \Omega$. Describe how your results compare to the discussion in Section 31.5.

31.63 •• An L-R-C series circuit draws 220 W from a 120-V (rms), 50.0-Hz ac line. The power factor is 0.560, and the source voltage leads the current. (a) What is the net resistance R of the circuit? (b) Find the capacitance of the series capacitor that will result in a power factor of unity when it is added to the original circuit. (c) What power will then be drawn from the supply line?

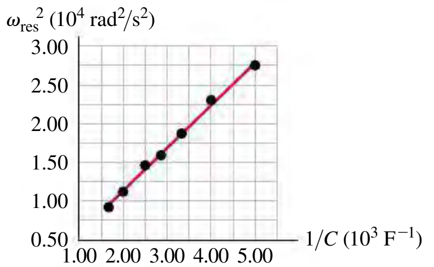
31.64 •• **DATA** A coworker of yours was making measurements of a large solenoid that is connected to an ac voltage source. Unfortunately, she left for vacation before she completed the analysis, and your boss has asked you to finish it. You are given a graph of $1/I^2$ versus ω^2 (**Fig. P31.64**), where I is the current in the circuit and ω is the angular frequency of the source. A note attached to the graph says that the voltage amplitude of the source was kept constant at 12.0 V. Calculate the resistance and inductance of the solenoid.

Figure **P31.64**



31.65 •• **DATA** You are analyzing an ac circuit that contains a solenoid and a capacitor in series with an ac source that has voltage amplitude 90.0 V and angular frequency ω. For different capacitors in the circuit, each with known capacitance, you measure the value of the frequency ω_{res} for which the current in the circuit is a maximum. You plot your measured values on a graph of $\omega_{\text{res}}{}^2$ versus $1/C$ (**Fig. P31.65**). The maximum current for each value of C is the same, you note, and equal to 4.50 A. Calculate the resistance and inductance of the solenoid.

Figure **P31.65**



31.66 •• **DATA** You are given this table of data recorded for a circuit that has a resistor, an inductor with negligible resistance, and a capacitor, all in series with an ac voltage source:

f (Hz)	80	160
Z (Ω)	15	13
ϕ (°)	−71	67

Here f is the frequency of the voltage source, Z is the impedance of the circuit, and ϕ is the phase angle. (a) Use the data at both frequencies to calculate the resistance of the resistor. (*Hint*: Use the results of Exercise 31.21.) Calculate the average of these two values of the resistance, and use the result as the value of R in the rest of the analysis. (b) Use the data at 80 Hz and 160 Hz to calculate the inductance L and capacitance C of the circuit. (c) What is the resonance frequency for the circuit, and what are the impedance and phase angle at the resonance frequency?

CHALLENGE PROBLEMS

31.67 •• **CALC** In an L-R-C series circuit the current is given by $i = I\cos\omega t$. The voltage amplitudes for the resistor, inductor, and capacitor are V_R, V_L, and V_C. (a) Show that the instantaneous power into the resistor is $p_R = V_R I\cos^2\omega t = \frac{1}{2}V_R I(1 + \cos 2\omega t)$. What does this expression give for the average power into the resistor? (b) Show that the instantaneous power into the inductor is $p_L = -V_L I\sin\omega t\cos\omega t = -\frac{1}{2}V_L I\sin 2\omega t$. What does this expression give for the average power into the inductor? (c) Show that the instantaneous power into the capacitor is $p_C = V_C I\sin\omega t\cos\omega t = \frac{1}{2}V_C I\sin 2\omega t$. What does this expression give for the average power into the capacitor? (d) The instantaneous power delivered by the source is shown in Section 31.4 to be $p = VI\cos\omega t(\cos\phi\cos\omega t - \sin\phi\sin\omega t)$. Show that $p_R + p_L + p_C$ equals p at each instant of time.

31.68 ••• **CALC** (a) At what angular frequency is the voltage amplitude across the *resistor* in an L-R-C series circuit at maximum value? (b) At what angular frequency is the voltage amplitude across the *inductor* at maximum value? (c) At what angular frequency is the voltage amplitude across the *capacitor* at maximum value? (You may want to refer to the results of Problem 31.49.)

PASSAGE PROBLEMS

BIO CONVERTING DC TO AC. An individual cell such as an egg cell (an ovum, produced in the ovaries) is commonly organized spatially, as manifested in part by asymmetries in the cell membrane. These asymmetries include nonuniform distributions of ion transport mechanisms, which result in a net electric current entering one region of the membrane and leaving another. These steady cellular currents may regulate cell polarity, leading (in the case of eggs) to embryonic polarity; therefore scientists are interested in measuring them.

These cellular currents move in loops through extracellular fluid. Ohm's law requires that there be voltage differences between any two points in this current-carrying fluid surrounding cells. Although the currents may be significant, the extracellular voltage differences are tiny—on the order of nanovolts. If we can map the voltage differences in the fluid outside a cell, we can calculate the current density by using Ohm's law, assuming that the resistivity of the fluid is known. We cannot measure these voltage differences by spacing two electrodes 10 or 20 μm apart, because the dc impedance (the resistance) of such electrodes is high and the inherent noise in signals detected at the electrodes far exceeds the cellular voltages.

One successful method of measurement uses an electrode with a ball-shaped end made of platinum that is moved sinusoidally between two points in the fluid outside a cell. The electric potential that the electrode measures, with respect to a distant reference electrode, also varies sinusoidally. The dc potential difference between the two extremes (the two points in the fluid) is then converted to a sine-wave ac potential difference. The platinum electrode behaves as a capacitor in series with the resistance of the extracellular fluid. This resistance, called the *access resistance* (R_A), has a value of about $\rho/10a$, where ρ is the resistivity of the fluid (usually expressed in $\Omega \cdot$cm) and a is the radius of the ball electrode. The platinum ball typically has a diameter of 20 μm and a capacitance of 10 nF; the resistivity of many biological fluids is 100 $\Omega \cdot$cm.

31.69 What is the dc impedance of the electrode, assuming that it behaves as an ideal capacitor? (a) 0; (b) infinite; (c) $\sqrt{2} \times 10^4\ \Omega$; (d) $\sqrt{2} \times 10^6\ \Omega$.

31.70 If the electrode oscillates between two points 20 μm apart at a frequency of $(5000/\pi)$Hz, what is the electrode's impedance? (a) 0; (b) infinite; (c) $\sqrt{2} \times 10^4\ \Omega$; (d) $\sqrt{2} \times 10^6\ \Omega$.

31.71 The signal from the oscillating electrode is fed into an amplifier, which reports the measured voltage as an rms value, 1.5 nV. What is the potential difference between the two extremes? (a) 1.5 nV; (b) 3.0 nV; (c) 2.1 nV; (d) 4.2 nV.

31.72 If the frequency at which the electrode is oscillated is increased to a very large value, the electrode's impedance (a) approaches infinity; (b) approaches zero; (c) approaches a constant but nonzero value; (d) does not change.

Answers

Chapter Opening Question ?

(iv) A radio simultaneously detects transmissions at *all* frequencies. However, a radio is an *L-R-C* series circuit, and at any given time it is tuned to have a resonance at just one frequency. Hence the response of the radio to that frequency is much greater than its response to any other frequency, which is why you hear only one broadcasting station through the radio's speaker. (You can sometimes hear a second station if its frequency is sufficiently close to the tuned frequency.)

Test Your Understanding Questions

31.1 (a) D; (b) A; (c) B; (d) C For each phasor, the actual current is represented by the projection of that phasor onto the horizontal axis. The phasors all rotate counterclockwise around the origin with angular speed ω, so at the instant shown the projection of phasor A is positive but trending toward zero; the projection of phasor B is negative and becoming more negative; the projection of phasor C is negative but trending toward zero; and the projection of phasor D is positive and becoming more positive.

31.2 (a) (iii); (b) (ii); (c) (i) For a resistor, $V_R = IR$, so $I = V_R/R$. The voltage amplitude V_R and resistance R do not change with frequency, so the current amplitude I remains constant. For an inductor, $V_L = IX_L = I\omega L$, so $I = V_L/\omega L$. The voltage amplitude V_L and inductance L are constant, so the current amplitude I decreases as the frequency increases. For a capacitor, $V_C = IX_C = I/\omega C$, so $I = V_C\omega C$. The voltage amplitude V_C and capacitance C are constant, so the current amplitude I increases as the frequency increases.

31.3 (iv), (ii), (i), (iii) For the circuit in Example 31.4, $I = V/Z = (50\ \text{V})/(500\ \Omega) = 0.10$ A. If the capacitor and inductor are removed so that only the ac source and resistor remain, the circuit is like that shown in Fig. 31.7a; then $I = V/R = (50\ \text{V})/(300\ \Omega) = 0.17$ A. If the resistor and capacitor are removed so that only the ac source and inductor remain, the circuit is like that shown in Fig. 31.8a; then $I = V/X_L = (50\ \text{V})/(600\ \Omega) = 0.083$ A. Finally, if the resistor and inductor are removed so that only the ac source and capacitor remain, the circuit is like that shown in Fig. 31.9a; then $I = V/X_C = (50\ \text{V})/(200\ \Omega) = 0.25$ A.

31.4 (a) (v); (b) (iv) The energy cannot be extracted from the resistor, since energy is dissipated in a resistor and cannot be recovered. Instead, the energy must be extracted from either the inductor (which stores magnetic-field energy) or the capacitor (which stores electric-field energy). Positive power means that energy is being transferred from the ac source to the circuit, so *negative* power implies that energy is being transferred back into the source.

31.5 (ii) The capacitance C increases if the plate spacing is decreased (see Section 24.1). Hence the resonance frequency $f_0 = \omega_0/2\pi = 1/2\pi\sqrt{LC}$ decreases.

31.6 (ii), (iv), (i), (iii) From Eq. (31.35) the turns ratio is $N_2/N_1 = V_2/V_1$, so the number of turns in the secondary is $N_2 = N_1V_2/V_1$. Hence for the four cases we have (i) $N_2 = (1000)(6.0\ \text{V})/(120\ \text{V}) = 50$ turns; (ii) $N_2 = (1000)(240\ \text{V})/(120\ \text{V}) = 2000$ turns; (iii) $N_2 = (1000)(6.0\ \text{V})/(240\ \text{V}) = 25$ turns; and (iv) $N_2 = (1000)(120\ \text{V})/(240\ \text{V}) = 500$ turns. Note that (i), (iii), and (iv) are step-down transformers with fewer turns in the secondary than in the primary, while (ii) is a step-up transformer with more turns in the secondary than in the primary.

Bridging Problem

(a) 8.35×10^4 rad/s and 3.19×10^5 rad/s

(b) At 8.35×10^4 rad/s: $V_{\text{source}} = 49.5$ V, $I = 0.132$ A, $V_R = 16.5$ V, $V_L = 16.5$ V, $V_C = 63.2$ V.

At 3.19×10^5 rad/s: $V_{\text{source}} = 49.5$ V, $I = 0.132$ A, $V_R = 16.5$ V, $V_L = 63.2$ V, $V_C = 16.5$ V.

? Metal objects reflect not only visible light but also radio waves. This is because at the surface of a metal, (i) the electric-field component parallel to the surface must be zero; (ii) the electric-field component perpendicular to the surface must be zero; (iii) the magnetic-field component parallel to the surface must be zero; (iv) the magnetic-field component perpendicular to the surface must be zero; (v) more than one of these.

32 ELECTROMAGNETIC WAVES

LEARNING GOALS

Looking forward at …

32.1 How electromagnetic waves are generated.

32.2 How and why the speed of light is related to the fundamental constants of electricity and magnetism.

32.3 How to describe the propagation of a sinusoidal electromagnetic wave.

32.4 What determines the amount of energy and momentum carried by an electromagnetic wave.

32.5 How to describe standing electromagnetic waves.

Looking back at …

8.1 Momentum.

15.3, 15.7 Traveling waves and standing waves on a string.

16.4 Standing sound waves.

23.4 Electric field in a conductor.

24.3, 24.4 Electric energy density; permittivity of a dielectric.

28.1, 28.8 Magnetic field of a moving charge; permeability of a dielectric.

29.2, 29.7 Faraday's law and Maxwell's equations.

30.3, 30.5 Magnetic energy density; *L-C* circuits.

What is light? This question has been asked by humans for centuries, but there was no answer until electricity and magnetism were unified into *electromagnetism,* as described by Maxwell's equations. These equations show that a time-varying magnetic field acts as a source of electric field and that a time-varying electric field acts as a source of magnetic field. These $\vec{E}$ and $\vec{B}$ fields can sustain each other, forming an *electromagnetic wave* that propagates through space. Visible light emitted by the glowing filament of a light bulb is one example of an electromagnetic wave; other kinds of electromagnetic waves are produced by wi-fi base stations, x-ray machines, and radioactive nuclei.

In this chapter we'll use Maxwell's equations as the theoretical basis for understanding electromagnetic waves. We'll find that these waves carry both energy and momentum. In sinusoidal electromagnetic waves, the $\vec{E}$ and $\vec{B}$ fields are sinusoidal functions of time and position, with a definite frequency and wavelength. Visible light, radio, x rays, and other types of electromagnetic waves differ only in their frequency and wavelength. Our study of optics in the following chapters will be based in part on the electromagnetic nature of light.

Unlike waves on a string or sound waves in a fluid, electromagnetic waves do not require a material medium; the light that you see coming from the stars at night has traveled without difficulty across tens or hundreds of light-years of (nearly) empty space. Nonetheless, electromagnetic waves and mechanical waves have much in common and are described in much the same language. Before reading further in this chapter, you should review the properties of mechanical waves as discussed in Chapters 15 and 16.

32.1 MAXWELL'S EQUATIONS AND ELECTROMAGNETIC WAVES

In the last several chapters we studied various aspects of electric and magnetic fields. We learned that when the fields don't vary with time, such as an electric field produced by charges at rest or the magnetic field of a steady current, we can analyze the electric and magnetic fields independently without considering interactions between them. But when the fields vary with time, they are no longer independent. Faraday's law (see Section 29.2) tells us that a time-varying magnetic field acts as a source of electric field, as shown by induced emfs in inductors and transformers. Ampere's law, including the displacement current discovered by James Clerk Maxwell (see Section 29.7), shows that a time-varying electric field acts as a source of magnetic field. This mutual interaction between the two fields is summarized in Maxwell's equations, presented in Section 29.7.

Thus, when *either* an electric or a magnetic field is changing with time, a field of the other kind is induced in adjacent regions of space. We are led (as Maxwell was) to consider the possibility of an electromagnetic disturbance, consisting of time-varying electric and magnetic fields, that can propagate through space from one region to another, even when there is no matter in the intervening region. Such a disturbance, if it exists, will have the properties of a *wave,* and an appropriate term is **electromagnetic wave.**

Such waves do exist; radio and television transmission, light, x rays, and many other kinds of radiation are examples of electromagnetic waves. Our goal in this chapter is to see how such waves are explained by the principles of electromagnetism that we have studied thus far and to examine the properties of these waves.

Electricity, Magnetism, and Light

The theoretical understanding of electromagnetic waves actually evolved along a considerably more devious path than the one just outlined. In the early days of electromagnetic theory (the early 19th century), two different units of electric charge were used: one for electrostatics and the other for magnetic phenomena involving currents. In the system of units used at that time, these two units of charge had different physical dimensions. Their *ratio* had units of velocity, and measurements showed that the ratio had a numerical value that was precisely equal to the speed of light, 3.00×10^8 m/s. At the time, physicists regarded this as an extraordinary coincidence and had no idea how to explain it.

In searching to understand this result, Maxwell (**Fig. 32.1**) proved in 1865 that an electromagnetic disturbance should propagate in free space with a speed equal to that of light and hence that light waves were likely to be electromagnetic in nature. At the same time, he discovered that the basic principles of electromagnetism can be expressed in terms of the four equations that we now call **Maxwell's equations,** which we discussed in Section 29.7. These four equations are (1) Gauss's law for electric fields; (2) Gauss's law for magnetic fields, showing the absence of magnetic monopoles; (3) Faraday's law; and (4) Ampere's law, including displacement current:

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q_{\text{encl}}}{\epsilon_0} \qquad \text{(Gauss's law)} \tag{29.18}$$

$$\oint \vec{B} \cdot d\vec{A} = 0 \qquad \text{(Gauss's law for magnetism)} \tag{29.19}$$

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \qquad \text{(Faraday's law)} \tag{29.20}$$

$$\oint \vec{B} \cdot d\vec{l} = \mu_0\left(i_C + \epsilon_0\frac{d\Phi_E}{dt}\right)_{\text{encl}} \qquad \text{(Ampere's law)} \tag{29.21}$$

32.1 The Scottish physicist James Clerk Maxwell (1831–1879) was the first person to truly understand the fundamental nature of light. He also made major contributions to thermodynamics, optics, astronomy, and color photography. Albert Einstein described Maxwell's accomplishments as "the most profound and the most fruitful that physics has experienced since the time of Newton."

These equations apply to electric and magnetic fields *in vacuum*. If a material is present, the electric constant ϵ_0 and magnetic constant μ_0 are replaced by the permittivity ϵ and permeability μ of the material. If the values of ϵ and μ are different at different points in the regions of integration, then ϵ and μ have to be transferred to the left sides of Eqs. (29.18) and (29.21), respectively, and placed inside the integrals. The ϵ in Eq. (29.21) also has to be included in the integral that gives $d\Phi_E/dt$.

According to Maxwell's equations, a point charge at rest produces a static $\vec{E}$ field but no $\vec{B}$ field, whereas a point charge moving with a constant velocity (see Section 28.1) produces both $\vec{E}$ and $\vec{B}$ fields. Maxwell's equations can also be used to show that in order for a point charge to produce electromagnetic waves, the charge must *accelerate*. In fact, in *every* situation where electromagnetic energy is radiated, the source is accelerated charges (**Fig. 32.2**).

32.2 (a) Every mobile phone emits signals in the form of electromagnetic waves that are made by accelerating charges. (b) Power lines carry a strong alternating current, which means that a substantial amount of charge is accelerating back and forth and generating electromagnetic waves. These waves can produce a buzzing sound from your car radio when you drive near the lines.

(a)

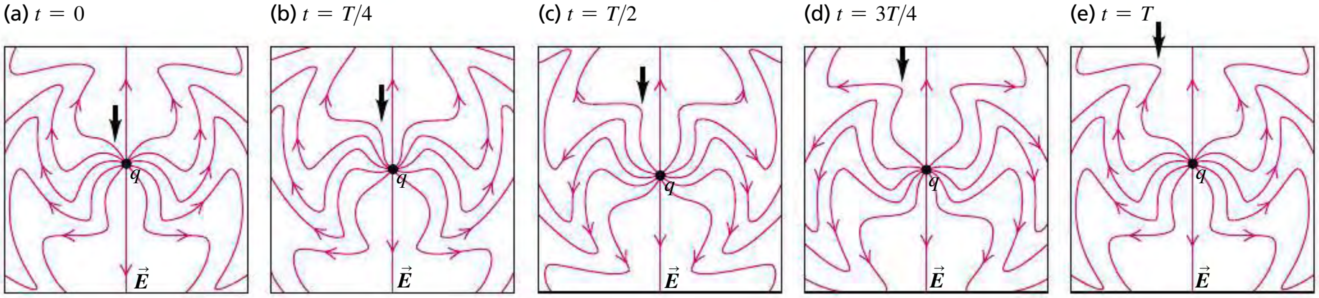
(b)

Generating Electromagnetic Radiation

One way in which a point charge can be made to emit electromagnetic waves is by making it oscillate in simple harmonic motion, so that it has an acceleration at almost every instant (the exception is when the charge is passing through its equilibrium position). **Figure 32.3** shows some of the electric field lines produced by such an oscillating point charge. Field lines are *not* material objects, but you may nonetheless find it helpful to think of them as behaving somewhat like strings that extend from the point charge off to infinity. Oscillating the charge up and down makes waves that propagate outward from the charge along these "strings." Note that the charge does not emit waves equally in all directions; the waves are strongest at 90° to the axis of motion of the charge, while there are *no* waves along this axis. This is just what the "string" picture would lead you to conclude. There is also a *magnetic* disturbance that spreads outward from the charge; this is not shown in Fig. 32.3. Because the electric and magnetic disturbances spread or radiate away from the source, the name **electromagnetic radiation** is used interchangeably with the phrase "electromagnetic waves."

Electromagnetic waves with macroscopic wavelengths were first produced in the laboratory in 1887 by the German physicist Heinrich Hertz (for whom the SI unit of frequency is named). As a source of waves, he used charges oscillating in L-C circuits (see Section 30.5); he detected the resulting electromagnetic waves with other circuits tuned to the same frequency. Hertz also produced electromagnetic *standing waves* and measured the distance between adjacent nodes (one half-wavelength) to determine the wavelength. Knowing the resonant frequency of his circuits, he then found the speed of the waves from the wavelength–frequency relationship $v = \lambda f$. He established that their speed was the same as that of light; this verified Maxwell's theoretical prediction directly.

32.3 Electric field lines of a point charge oscillating in simple harmonic motion, seen at five instants during an oscillation period T. The charge's trajectory is in the plane of the drawings. At $t = 0$ the point charge is at its maximum upward displacement. The arrow shows one "kink" in the lines of $\vec{E}$ as it propagates outward from the point charge. The magnetic field (not shown) contains circles that lie in planes perpendicular to these figures and concentric with the axis of oscillation.



The modern value of the speed of light, c, is 299,792,458 m/s. (Recall from Section 1.3 that this value is the basis of our standard of length: One meter is defined to be the distance that light travels in 1/299,792,458 second.) For our purposes, $c = 3.00 \times 10^8$ m/s is sufficiently accurate.

In the wake of Hertz's discovery, Guglielmo Marconi and others made radio communication a familiar household experience. In a radio *transmitter,* electric charges are made to oscillate along the length of the conducting antenna, producing oscillating field disturbances like those shown in Fig. 32.3. Since many charges oscillate together in the antenna, the disturbances are much stronger than those of a single oscillating charge and can be detected at a much greater distance. In a radio *receiver* the antenna is also a conductor; the fields of the wave emanating from a distant transmitter exert forces on free charges within the receiver antenna, producing an oscillating current that is detected and amplified by the receiver circuitry.

For the remainder of this chapter our concern will be with electromagnetic waves themselves, not with the rather complex problem of how they are produced.

The Electromagnetic Spectrum

The **electromagnetic spectrum** encompasses electromagnetic waves of all frequencies and wavelengths. **Figure 32.4** shows approximate wavelength and frequency ranges for the most commonly encountered portion of the spectrum. Despite vast differences in their uses and means of production, these are all electromagnetic waves with the same propagation speed (in vacuum) $c = 299{,}792{,}458$ m/s. Electromagnetic waves may differ in frequency f and wavelength λ, but the relationship $c = \lambda f$ in vacuum holds for each.

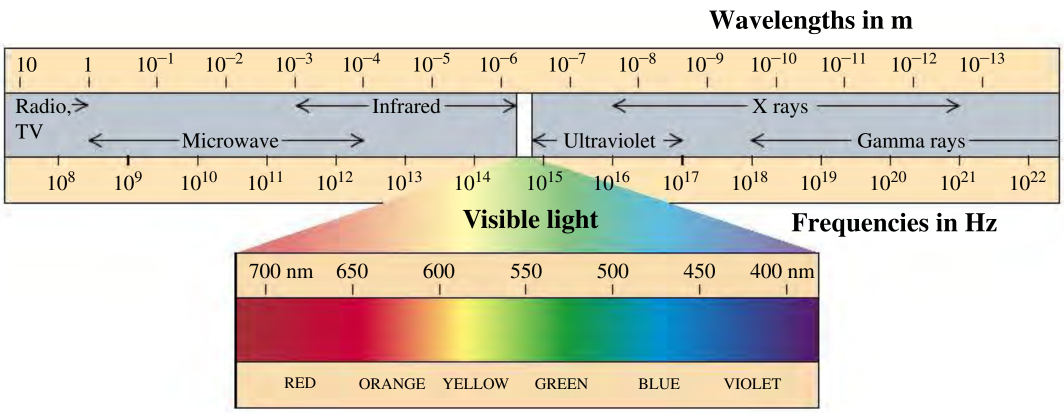
We can detect only a very small segment of this spectrum directly through our sense of sight. We call this range **visible light.** Its wavelengths range from about 380 to 750 nm $(380 \text{ to } 750 \times 10^{-9} \text{ m})$, with corresponding frequencies from about 790 to 400 THz $(7.9 \text{ to } 4.0 \times 10^{14} \text{ Hz})$. Different parts of the visible spectrum evoke in humans the sensations of different colors. **Table 32.1** gives the approximate wavelengths for colors in the visible spectrum.

Ordinary white light includes all visible wavelengths. However, by using special sources or filters, we can select a narrow band of wavelengths within a range of a few nm. Such light is approximately *monochromatic* (single-color) light. Absolutely monochromatic light with only a single wavelength is an unattainable idealization. When we say "monochromatic light with $\lambda = 550$ nm" with reference to a laboratory experiment, we really mean a small band of

TABLE 32.1 Wavelengths of Visible Light

380–450 nm	Violet
450–495 nm	Blue
495–570 nm	Green
570–590 nm	Yellow
590–620 nm	Orange
620–750 nm	Red

32.4 The electromagnetic spectrum. The frequencies and wavelengths found in nature extend over such a wide range that we have to use a logarithmic scale to show all important bands. The boundaries between bands are somewhat arbitrary.



BIO Application Ultraviolet Vision
Many insects and birds can see ultraviolet wavelengths that humans cannot. As an example, the left-hand photo shows how black-eyed Susans (genus *Rudbeckia*) look to us. The right-hand photo (in false color), taken with an ultraviolet-sensitive camera, shows how these same flowers appear to the bees that pollinate them. Note the prominent central spot that is invisible to humans. Similarly, many birds with ultraviolet vision—including budgies, parrots, and peacocks—have ultraviolet patterns on their bodies that make them even more vivid to each other than they appear to us.

wavelengths *around* 550 nm. Light from a *laser* is much more nearly monochromatic than is light obtainable in any other way.

Invisible forms of electromagnetic radiation are no less important than visible light. Our system of global communication, for example, depends on radio waves: AM radio uses waves with frequencies from 5.4×10^5 Hz to 1.6×10^6 Hz, and FM radio broadcasts are at frequencies from 8.8×10^7 Hz to 1.08×10^8 Hz. Microwaves are also used for communication (for example, by mobile phones and wireless networks) and for weather radar (at frequencies near 3×10^9 Hz). Many cameras have a device that emits a beam of infrared radiation; by analyzing the properties of the infrared radiation reflected from the subject, the camera determines the distance to the subject and automatically adjusts the focus. X rays are able to penetrate through flesh, which makes them invaluable in dentistry and medicine. Gamma rays, the shortest-wavelength type of electromagnetic radiation, are used in medicine to destroy cancer cells.

TEST YOUR UNDERSTANDING OF SECTION 32.1 (a) Is it possible to have a purely electric wave propagate through empty space—that is, a wave made up of an electric field but no magnetic field? (b) What about a purely magnetic wave, with a magnetic field but no electric field?

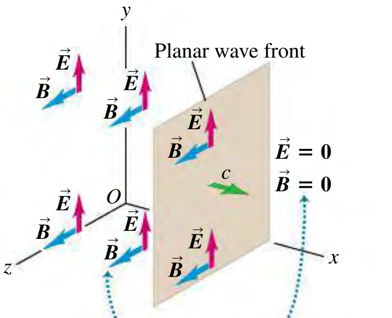
32.2 PLANE ELECTROMAGNETIC WAVES AND THE SPEED OF LIGHT

We are now ready to develop the basic ideas of electromagnetic waves and their relationship to the principles of electromagnetism. Our procedure will be to postulate a simple field configuration that has wavelike behavior. We'll assume an electric field $\vec{E}$ that has only a y-component and a magnetic field $\vec{B}$ with only a z-component, and we'll assume that both fields move together in the $+x$-direction with a speed c that is initially unknown. (As we go along, it will become clear why we choose $\vec{E}$ and $\vec{B}$ to be perpendicular to the direction of propagation as well as to each other.) Then we will test whether these fields are physically possible by asking whether they are consistent with Maxwell's equations, particularly Ampere's law and Faraday's law. We'll find that the answer is yes, provided that c has a particular value. We'll also show that the *wave equation,* which we encountered during our study of mechanical waves in Chapter 15, can be derived from Maxwell's equations.

A Simple Plane Electromagnetic Wave

Using an xyz-coordinate system (**Fig. 32.5**), we imagine that all space is divided into two regions by a plane perpendicular to the x-axis (parallel to the yz-plane). At every point to the left of this plane there are a uniform electric field $\vec{E}$ in the $+y$-direction and a uniform magnetic field $\vec{B}$ in the $+z$-direction, as shown. Furthermore, we suppose that the boundary plane, which we call the *wave front,* moves to the right in the $+x$-direction with a constant speed c, the value of which we'll leave undetermined for now. Thus the $\vec{E}$ and $\vec{B}$ fields travel to the right into previously field-free regions with a definite speed. This is a rudimentary electromagnetic wave. Such a wave, in which at any instant the fields are uniform over any plane perpendicular to the direction of propagation, is called a **plane wave.** In the case shown in Fig. 32.5, the fields are zero for planes to the right of the wave front and have the same values on all planes to the left of the wave front; later we will consider more complex plane waves.

32.5 An electromagnetic wave front. The plane representing the wave front moves to the right (in the positive x-direction) with speed c.



We won't concern ourselves with the problem of actually *producing* such a field configuration. Instead, we simply ask whether it is consistent with the laws of electromagnetism—that is, with all four of Maxwell's equations.

Let us first verify that our wave satisfies Maxwell's first and second equations—that is, Gauss's laws for electric and magnetic fields. To do this, we take as our Gaussian surface a rectangular box with sides parallel to the *xy*-, *xz*-, and *yz*-coordinate planes (**Fig. 32.6**). The box encloses no electric charge. The total electric flux and magnetic flux through the box are both zero, even if part of the box is in the region where $E = B = 0$. This would *not* be the case if $\vec{E}$ or $\vec{B}$ had an *x*-component, parallel to the direction of propagation; if the wave front were inside the box, there would be flux through the left-hand side of the box (at $x = 0$) but not the right-hand side (at $x > 0$). Thus to satisfy Maxwell's first and second equations, the electric and magnetic fields must be perpendicular to the direction of propagation; that is, the wave must be **transverse.**

The next of Maxwell's equations that we'll consider is Faraday's law:

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \qquad (32.1)$$

To test whether our wave satisfies Faraday's law, we apply this law to a rectangle *efgh* that is parallel to the *xy*-plane (**Fig. 32.7a**). As shown in Fig. 32.7b, a cross section in the *xy*-plane, this rectangle has height a and width Δx. At the time shown, the wave front has progressed partway through the rectangle, and $\vec{E}$ is zero along the side *ef*. In applying Faraday's law we take the vector area $d\vec{A}$ of rectangle *efgh* to be in the $+z$-direction. With this choice the right-hand rule requires that we integrate $\vec{E} \cdot d\vec{l}$ *counterclockwise* around the rectangle. At every point on side *ef*, $\vec{E}$ is zero. At every point on sides *fg* and *he*, $\vec{E}$ is either zero or perpendicular to $d\vec{l}$. Only side *gh* contributes to the integral. On this side, $\vec{E}$ and $d\vec{l}$ are opposite, and we find that the left-hand side of Eq. (32.1) is nonzero:

$$\oint \vec{E} \cdot d\vec{l} = -Ea \qquad (32.2)$$

To satisfy Faraday's law, Eq. (32.1), there must be a component of $\vec{B}$ in the *z*-direction (perpendicular to $\vec{E}$) so that there can be a nonzero magnetic flux Φ_B through the rectangle *efgh* and a nonzero derivative $d\Phi_B/dt$. Indeed, in our wave, $\vec{B}$ has *only* a *z*-component. We have assumed that this component is in the *positive* *z*-direction; let's see whether this assumption is consistent with Faraday's law. During a time interval dt the wave front (traveling at speed c) moves a distance $c\,dt$ to the right in Fig. 32.7b, sweeping out an area $ac\,dt$ of the rectangle *efgh*. During this interval the magnetic flux Φ_B through the rectangle *efgh* increases by $d\Phi_B = B(ac\,dt)$, so the rate of change of magnetic flux is

$$\frac{d\Phi_B}{dt} = Bac \qquad (32.3)$$

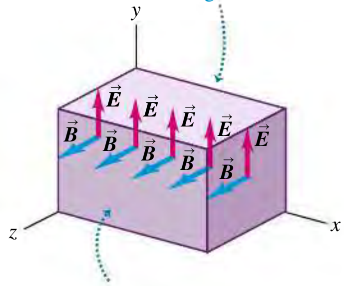
Now we substitute Eqs. (32.2) and (32.3) into Faraday's law, Eq. (32.1); we get $-Ea = -Bac$, so

Electromagnetic wave in vacuum: Electric-field magnitude; Magnetic-field magnitude; Speed of light in vacuum

$$E = cB \qquad (32.4)$$

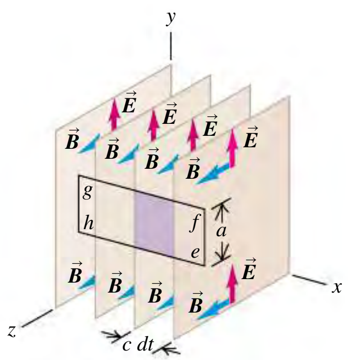
Our wave is consistent with Faraday's law only if the wave speed c and the magnitudes of $\vec{E}$ and $\vec{B}$ are related as in Eq. (32.4). If we had assumed that $\vec{B}$ was in the *negative* *z*-direction, there would have been an additional minus sign in Eq. (32.4); since E, c, and B are all positive magnitudes, no solution would then have been possible. Furthermore, any component of $\vec{B}$ in the *y*-direction (parallel to $\vec{E}$) would not contribute to the changing magnetic flux Φ_B through the rectangle *efgh* (which is parallel to the *xy*-plane) and so would not be part of the wave.

32.6 Gaussian surface for a transverse plane electromagnetic wave.

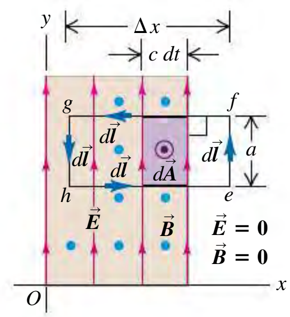


32.7 (a) Applying Faraday's law to a plane wave. (b) In a time dt, the magnetic flux through the rectangle in the *xy*-plane increases by an amount $d\Phi_B$. This increase equals the flux through the shaded rectangle with area $ac\,dt$; that is, $d\Phi_B = Bac\,dt$. Thus $d\Phi_B/dt = Bac$.

(a) In time dt, the wave front moves a distance $c\,dt$ in the $+x$-direction.



(b) Side view of situation in (a)

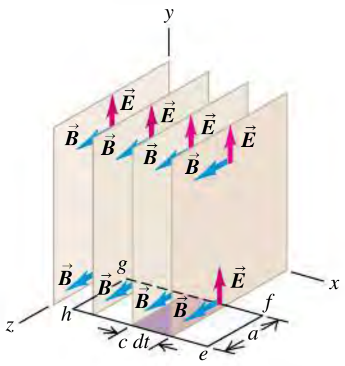


Finally, let's do a similar calculation with Ampere's law, the last of Maxwell's equations. There is no conduction current ($i_C = 0$), so Ampere's law is

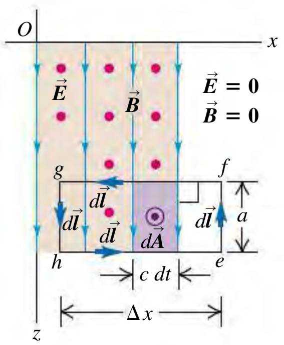
$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt} \tag{32.5}$$

32.8 (a) Applying Ampere's law to a plane wave. (Compare to Fig. 32.7a.) (b) In a time dt, the electric flux through the rectangle in the xz-plane increases by an amount $d\Phi_E$. This increase equals the flux through the shaded rectangle with area $ac\,dt$; that is, $d\Phi_E = Eac\,dt$. Thus $d\Phi_E/dt = Eac$.

(a) In time dt, the wave front moves a distance $c\,dt$ in the $+x$-direction.



(b) Top view of situation in (a)



To check whether our wave is consistent with Ampere's law, we move our rectangle so that it lies in the xz-plane (**Fig. 32.8**), and we again look at the situation at a time when the wave front has traveled partway through the rectangle. We take the vector area $d\vec{A}$ in the $+y$-direction, and so the right-hand rule requires that we integrate $\vec{B} \cdot d\vec{l}$ counterclockwise around the rectangle. The $\vec{B}$ field is zero at every point along side ef, and at each point on sides fg and he it is either zero or perpendicular to $d\vec{l}$. Only side gh, where $\vec{B}$ and $d\vec{l}$ are parallel, contributes to the integral, and

$$\oint \vec{B} \cdot d\vec{l} = Ba \tag{32.6}$$

Hence the left-hand side of Eq. (32.5) is nonzero; the right-hand side must be nonzero as well. Thus $\vec{E}$ must have a y-component (perpendicular to $\vec{B}$) so that the electric flux Φ_E through the rectangle and the time derivative $d\Phi_E/dt$ can be nonzero. Just as we inferred from Faraday's law, we conclude that in an electromagnetic wave, $\vec{E}$ and $\vec{B}$ must be mutually perpendicular.

In a time interval dt the electric flux Φ_E through the rectangle increases by $d\Phi_E = E(ac\,dt)$. Since we chose $d\vec{A}$ to be in the $+y$-direction, this flux change is positive; the rate of change of electric flux is

$$\frac{d\Phi_E}{dt} = Eac \tag{32.7}$$

Substituting Eqs. (32.6) and (32.7) into Ampere's law, Eq. (32.5), we find $Ba = \epsilon_0 \mu_0 Eac$, so

Electromagnetic wave in vacuum:

Magnetic-field magnitude — Electric-field magnitude

$$B = \epsilon_0 \mu_0 c E \tag{32.8}$$

Electric constant — Magnetic constant — Speed of light in vacuum

Our assumed wave obeys Ampere's law only if B, c, and E are related as in Eq. (32.8). The wave must also obey Faraday's law, so Eq. (32.4) must be satisfied as well. This can happen only if $\epsilon_0 \mu_0 c = 1/c$, or

Speed of electromagnetic waves in vacuum

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \tag{32.9}$$

Electric constant — Magnetic constant

Inserting the numerical values of these quantities, we find

$$c = \frac{1}{\sqrt{(8.85 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2})(4\pi \times 10^{-7}\ \mathrm{N/A^2})}}$$
$$= 3.00 \times 10^8\ \mathrm{m/s}$$

Our assumed wave is consistent with all of Maxwell's equations, provided that the wave front moves with the speed given above, which is the speed of light! Recall that the *exact* value of c is defined to be 299,792,458 m/s; the modern value of ϵ_0 is defined to agree with this when used in Eq. (32.9) (see Section 21.3).

Key Properties of Electromagnetic Waves

We chose a simple wave for our study in order to avoid mathematical complications, but this special case illustrates several important features of *all* electromagnetic waves:

1. The wave is *transverse*; both $\vec{E}$ and $\vec{B}$ are perpendicular to the direction of propagation of the wave. The electric and magnetic fields are also perpendicular to each other. The direction of propagation is the direction of the vector product $\vec{E} \times \vec{B}$ (**Fig. 32.9**).
2. There is a definite ratio between the magnitudes of $\vec{E}$ and $\vec{B}$: $E = cB$.
3. The wave travels in vacuum with a definite and unchanging speed.
4. Unlike mechanical waves, which need the particles of a medium such as air to transmit a wave, electromagnetic waves require no medium.

We can generalize this discussion to a more realistic situation. Suppose we have several wave fronts in the form of parallel planes perpendicular to the x-axis, all of which are moving to the right with speed c. Suppose that the $\vec{E}$ and $\vec{B}$ fields are the same at all points within a single region between two planes, but that the fields differ from region to region. The overall wave is a plane wave, but one in which the fields vary in steps along the x-axis. Such a wave could be constructed by superposing several of the simple step waves we have just discussed (shown in Fig. 32.5). This is possible because the $\vec{E}$ and $\vec{B}$ fields obey the superposition principle in waves just as in static situations: When two waves are superposed, the total $\vec{E}$ field at each point is the vector sum of the $\vec{E}$ fields of the individual waves, and similarly for the total $\vec{B}$ field.

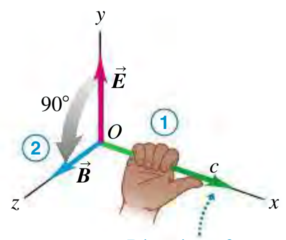
We can extend the above development to show that a wave with fields that vary in steps is also consistent with Ampere's and Faraday's laws, provided that the wave fronts all move with the speed c given by Eq. (32.9). In the limit that we make the individual steps infinitesimally small, we have a wave in which the $\vec{E}$ and $\vec{B}$ fields at any instant vary *continuously* along the x-axis. The entire field pattern moves to the right with speed c. In Section 32.3 we will consider waves in which $\vec{E}$ and $\vec{B}$ are *sinusoidal* functions of x and t. Because at each point the magnitudes of $\vec{E}$ and $\vec{B}$ are related by $E = cB$, the periodic variations of the two fields in any periodic traveling wave must be *in phase*.

Electromagnetic waves have the property of **polarization.** In the above discussion the choice of the y-direction for $\vec{E}$ was arbitrary. We could instead have specified the z-axis for $\vec{E}$; then $\vec{B}$ would have been in the $-y$-direction. A wave in which $\vec{E}$ is always parallel to a certain axis is said to be **linearly polarized** along that axis. More generally, *any* wave traveling in the x-direction can be represented as a superposition of waves linearly polarized in the y- and z-directions. We will study polarization in greater detail in Chapter 33.

32.9 A right-hand rule for electromagnetic waves relates the directions of $\vec{E}$ and $\vec{B}$ and the direction of propagation.

Right-hand rule for an electromagnetic wave:

1. Point the thumb of your right hand in the wave's direction of propagation.
2. Imagine rotating the $\vec{E}$-field vector 90° in the sense your fingers curl. That is the direction of the $\vec{B}$ field.



Direction of propagation = direction of $\vec{E} \times \vec{B}$.

DATA *SPEAKS*

Electromagnetic Waves

When students were given a problem involving electromagnetic waves, more than 29% gave an incorrect response. Common errors:

- Forgetting that, in vacuum, all electromagnetic waves travel at the same speed: c. Since $c = \lambda f$, waves with high frequency f have short wavelength λ but travel at the same speed as waves with low f and long λ.
- Confusion about the directions of $\vec{E}$ and $\vec{B}$ and the propagation direction. The electric and magnetic fields in an electromagnetic wave are always perpendicular to each other. The wave propagates in the direction of $\vec{E} \times \vec{B}$, which is perpendicular to both $\vec{E}$ and $\vec{B}$.

Derivation of the Electromagnetic Wave Equation

Here is an alternative derivation of Eq. (32.9) for the speed of electromagnetic waves. It is more mathematical than our other treatment, but it includes a derivation of the wave equation for electromagnetic waves. This part of the section can be omitted without loss of continuity in the chapter.

During our discussion of mechanical waves in Section 15.3, we showed that a function $y(x, t)$ that represents the displacement of any point in a mechanical wave traveling along the x-axis must satisfy a differential equation, Eq. (15.12):

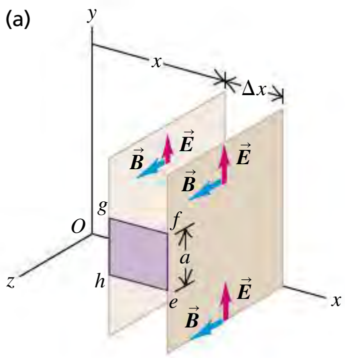
$$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2} \tag{32.10}$$

This equation is called the **wave equation,** and v is the speed of propagation of the wave.

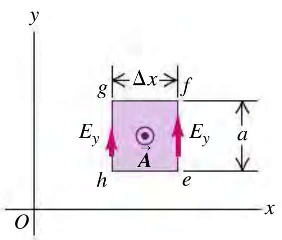
To derive the corresponding equation for an electromagnetic wave, we again consider a plane wave. That is, we assume that at each instant, E_y and B_z are uniform over any plane perpendicular to the x-axis, the direction of propagation. But now we let E_y and B_z vary continuously as we go along the x-axis; then each is a function of x and t. We consider the values of E_y and B_z on two planes perpendicular to the x-axis, one at x and one at $x + \Delta x$.

32.10 Faraday's law applied to a rectangle with height a and width Δx parallel to the xy-plane.

(a)



(b) Side view of the situation in (a)



Following the same procedure as previously, we apply Faraday's law to a rectangle lying parallel to the xy-plane, as in **Fig. 32.10**. This figure is similar to Fig. 32.7. Let the left end gh of the rectangle be at position x, and let the right end ef be at position $(x + \Delta x)$. At time t, the values of E_y on these two sides are $E_y(x, t)$ and $E_y(x + \Delta x, t)$, respectively. When we apply Faraday's law to this rectangle, we find that instead of $\oint \vec{E} \cdot d\vec{l} = -Ea$ as before, we have

$$\oint \vec{E} \cdot d\vec{l} = -E_y(x, t)a + E_y(x + \Delta x, t)a$$

$$= a\left[E_y(x + \Delta x, t) - E_y(x, t)\right] \tag{32.11}$$

To find the magnetic flux Φ_B through this rectangle, we assume that Δx is small enough that B_z is nearly uniform over the rectangle. In that case, $\Phi_B = B_z(x, t)A = B_z(x, t)a\,\Delta x$, and

$$\frac{d\Phi_B}{dt} = \frac{\partial B_z(x, t)}{\partial t} a\,\Delta x$$

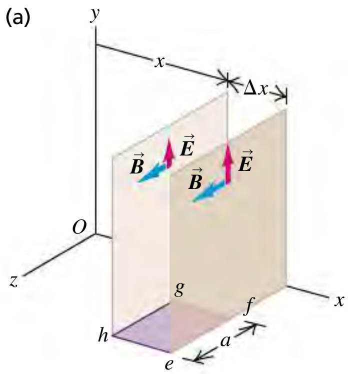
We use partial-derivative notation because B_z is a function of both x and t. When we substitute this expression and Eq. (32.11) into Faraday's law, Eq. (32.1), we get

$$a\left[E_y(x + \Delta x, t) - E_y(x, t)\right] = -\frac{\partial B_z}{\partial t} a\,\Delta x$$

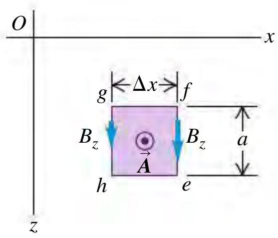
$$\frac{E_y(x + \Delta x, t) - E_y(x, t)}{\Delta x} = -\frac{\partial B_z}{\partial t}$$

32.11 Ampere's law applied to a rectangle with height a and width Δx parallel to the xz-plane.

(a)



(b) Top view of the situation in (a)



Finally, imagine shrinking the rectangle down to a sliver so that Δx approaches zero. When we take the limit of this equation as $\Delta x \rightarrow 0$, we get

$$\frac{\partial E_y(x, t)}{\partial x} = -\frac{\partial B_z(x, t)}{\partial t} \tag{32.12}$$

This equation shows that if there is a time-varying component B_z of magnetic field, there must also be a component E_y of electric field that varies with x, and conversely. We put this relationship on the shelf for now; we'll return to it soon.

Next we apply Ampere's law to the rectangle shown in **Fig. 32.11**. The line integral $\oint \vec{B} \cdot d\vec{l}$ becomes

$$\oint \vec{B} \cdot d\vec{l} = -B_z(x + \Delta x, t)a + B_z(x, t)a \tag{32.13}$$

Again assuming that the rectangle is narrow, we approximate the electric flux Φ_E through it as $\Phi_E = E_y(x, t)A = E_y(x, t)a\,\Delta x$. The rate of change of Φ_E, which we need for Ampere's law, is then

$$\frac{d\Phi_E}{dt} = \frac{\partial E_y(x, t)}{\partial t} a\,\Delta x$$

Now we substitute this expression and Eq. (32.13) into Ampere's law, Eq. (32.5):

$$-B_z(x + \Delta x, t)a + B_z(x, t)a = \epsilon_0 \mu_0 \frac{\partial E_y(x, t)}{\partial t} a\,\Delta x$$

Again we divide both sides by $a\ \Delta x$ and take the limit as $\Delta x \rightarrow 0$. We find

$$-\frac{\partial B_z(x, t)}{\partial x} = \epsilon_0\mu_0\frac{\partial E_y(x, t)}{\partial t} \tag{32.14}$$

Now comes the final step. We take the partial derivatives of both sides of Eq. (32.12) with respect to x, and we take the partial derivatives of both sides of Eq. (32.14) with respect to t. The results are

$$-\frac{\partial^2 E_y(x, t)}{\partial x^2} = \frac{\partial^2 B_z(x, t)}{\partial x \partial t}$$

$$-\frac{\partial^2 B_z(x, t)}{\partial x \partial t} = \epsilon_0\mu_0\frac{\partial^2 E_y(x, t)}{\partial t^2}$$

Combining these two equations to eliminate B_z, we finally find

$$\frac{\partial^2 E_y(x, t)}{\partial x^2} = \epsilon_0\mu_0\frac{\partial^2 E_y(x, t)}{\partial t^2} \quad \text{(electromagnetic wave equation in vacuum)} \tag{32.15}$$

This expression has the same form as the general wave equation, Eq. (32.10). Because the electric field E_y must satisfy this equation, it behaves as a wave with a pattern that travels through space with a definite speed. Furthermore, comparison of Eqs. (32.15) and (32.10) shows that the wave speed v is given by

$$\frac{1}{v^2} = \epsilon_0\mu_0 \quad \text{or} \quad v = \frac{1}{\sqrt{\epsilon_0\mu_0}}$$

This agrees with Eq. (32.9) for the speed c of electromagnetic waves.

We can show that B_z also must satisfy the same wave equation as E_y, Eq. (32.15). To prove this, we take the partial derivative of Eq. (32.12) with respect to t and the partial derivative of Eq. (32.14) with respect to x and combine the results. We leave this derivation for you to carry out.

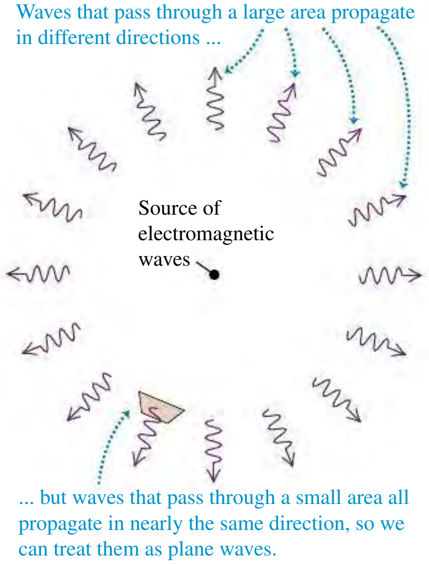
TEST YOUR UNDERSTANDING OF SECTION 32.2 For each of the following electromagnetic waves, state the direction of the magnetic field. (a) The wave is propagating in the positive z-direction, and $\vec{E}$ is in the positive x-direction; (b) the wave is propagating in the positive y-direction, and $\vec{E}$ is in the negative z-direction; (c) the wave is propagating in the negative x-direction, and $\vec{E}$ is in the positive z-direction. ❙

32.3 SINUSOIDAL ELECTROMAGNETIC WAVES

Sinusoidal electromagnetic waves are directly analogous to sinusoidal transverse mechanical waves on a stretched string, which we studied in Section 15.3. In a sinusoidal electromagnetic wave, $\vec{E}$ and $\vec{B}$ at any point in space are sinusoidal functions of time, and at any instant of time the *spatial* variation of the fields is also sinusoidal.

Some sinusoidal electromagnetic waves are *plane waves;* they share with the waves described in Section 32.2 the property that at any instant the fields are uniform over any plane perpendicular to the direction of propagation. The entire pattern travels in the direction of propagation with speed c. The directions of $\vec{E}$ and $\vec{B}$ are perpendicular to the direction of propagation (and to each other), so the wave is *transverse*. Electromagnetic waves produced by an oscillating point charge, shown in Fig. 32.3, are an example of sinusoidal waves that are *not* plane waves. But if we restrict our observations to a relatively small region of space at a sufficiently great distance from the source, even these waves are well approximated by plane waves (**Fig. 32.12**). In the same way, the curved surface of the (nearly) spherical earth appears flat to us because of our small size relative to the earth's radius. In this section we'll restrict our discussion to plane waves.

32.12 Waves passing through a small area at a sufficiently great distance from a source can be treated as plane waves.



The frequency f, the wavelength λ, and the speed of propagation c of any periodic wave are related by the usual wavelength–frequency relationship $c = \lambda f$. If the frequency f is 10^8 Hz (100 MHz), typical of commercial FM radio broadcasts, the wavelength is

$$\lambda = \frac{3 \times 10^8\ \text{m/s}}{10^8\ \text{Hz}} = 3\ \text{m}$$

Figure 32.4 shows the inverse proportionality between wavelength and frequency.

Fields of a Sinusoidal Wave

32.13 Representation of the electric and magnetic fields as functions of x for a linearly polarized sinusoidal plane electromagnetic wave. One wavelength of the wave is shown at time $t = 0$. The fields are shown for only a few points along the x-axis.

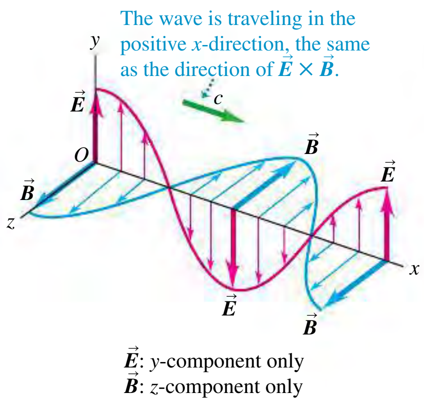


Figure 32.13 shows a linearly polarized sinusoidal electromagnetic wave traveling in the $+x$-direction. The electric and magnetic fields oscillate in phase: $\vec{E}$ is maximum where $\vec{B}$ is maximum and $\vec{E}$ is zero where $\vec{B}$ is zero. Where $\vec{E}$ is in the $+y$-direction, $\vec{B}$ is in the $+z$-direction; where $\vec{E}$ is in the $-y$-direction, $\vec{B}$ is in the $-z$-direction. At all points the vector product $\vec{E} \times \vec{B}$ is in the direction in which the wave is propagating (the $+x$-direction). We mentioned this in Section 32.2 in the list of characteristics of electromagnetic waves.

CAUTION In a plane wave, $\vec{E}$ and $\vec{B}$ are everywhere Figure 32.13 shows $\vec{E}$ and $\vec{B}$ at points on the x-axis only. But, in fact, in a sinusoidal plane wave there are electric and magnetic fields at *all* points in space. Imagine a plane perpendicular to the x-axis (that is, parallel to the yz-plane) at a particular point and time; the fields have the same values at all points in that plane. The values are different on different planes.

We can describe electromagnetic waves by means of *wave functions,* just as we did in Section 15.3 for waves on a string. One form of the wave function for a transverse wave traveling in the $+x$-direction along a stretched string is Eq. (15.7):

$$y(x, t) = A\cos(kx - \omega t)$$

where $y(x, t)$ is the transverse displacement from equilibrium at time t of a point with coordinate x on the string. Here A is the maximum displacement, or *amplitude,* of the wave; ω is its *angular frequency,* equal to 2π times the frequency f; and $k = 2\pi/\lambda$ is the *wave number,* where λ is the wavelength.

MP

PhET: Radio Waves & Electromagnetic Fields

Let $E_y(x, t)$ and $B_z(x, t)$ represent the instantaneous values of the y-component of $\vec{E}$ and the z-component of $\vec{B}$, respectively, in Fig. 32.13, and let E_{max} and B_{max} represent the maximum values, or *amplitudes,* of these fields. The wave functions for the wave are then

$$E_y(x, t) = E_{max}\cos(kx - \omega t) \qquad B_z(x, t) = B_{max}\cos(kx - \omega t) \tag{32.16}$$

We can also write the wave functions in vector form:

CAUTION The symbol k has two meanings Note the two different k's: the unit vector $\hat{k}$ in the z-direction and the wave number k. Don't get these confused!

Sinusoidal electromagnetic plane wave, propagating in $+x$-direction:

Electric field; Electric-field magnitude; Wave number

$$\vec{E}(x, t) = \hat{\jmath}E_{max}\cos(kx - \omega t)$$
$$\vec{B}(x, t) = \hat{k}B_{max}\cos(kx - \omega t) \tag{32.17}$$

Angular frequency; Magnetic field; Magnetic-field magnitude

The sine curves in Fig. 32.13 represent the fields as functions of x at time $t = 0$—that is, $\vec{E}(x, t = 0)$ and $\vec{B}(x, t = 0)$. As the wave travels to the right with speed c, Eqs. (32.16) and (32.17) show that at any point the oscillations of $\vec{E}$ and $\vec{B}$ are *in phase*. From Eq. (32.4) the amplitudes must be related by

Sinusoidal electromagnetic wave in vacuum:

Electric-field amplitude; Magnetic-field amplitude

$$E_{max} = cB_{max} \tag{32.18}$$

Speed of light in vacuum

These amplitude and phase relationships are also required for $E(x, t)$ and $B(x, t)$ to satisfy Eqs. (32.12) and (32.14), which came from Faraday's law and Ampere's law, respectively. Can you verify this statement? (See Problem 32.34.)

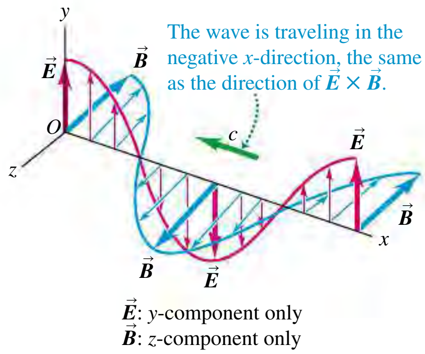
Figure 32.14 shows the $\vec{E}$ and $\vec{B}$ fields of a wave traveling in the *negative* x-direction. At points where $\vec{E}$ is in the positive y-direction, $\vec{B}$ is in the *negative* z-direction; where $\vec{E}$ is in the negative y-direction, $\vec{B}$ is in the *positive* z-direction. As with waves traveling in the $+x$-direction, at any point the oscillations of the $\vec{E}$ and $\vec{B}$ fields of this wave are in phase, and the vector product $\vec{E} \times \vec{B}$ points in the propagation direction. The wave functions for this wave are

$$\vec{E}(x, t) = \hat{\jmath} E_{\max} \cos(kx + \omega t)$$
$$\vec{B}(x, t) = -\hat{k} B_{\max} \cos(kx + \omega t) \qquad (32.19)$$

(sinusoidal electromagnetic plane wave, propagating in $-x$-direction)

The sinusoidal waves shown in both Figs. 32.13 and 32.14 are linearly polarized in the y-direction; the $\vec{E}$ field is always parallel to the y-axis. Example 32.1 concerns a wave that is linearly polarized in the z-direction.

32.14 Representation of one wavelength of a linearly polarized sinusoidal plane electromagnetic wave traveling in the negative x-direction at $t = 0$. The fields are shown only for points along the x-axis. (Compare with Fig. 32.13.)



PROBLEM-SOLVING STRATEGY 32.1 ELECTROMAGNETIC WAVES

IDENTIFY *the relevant concepts:* Many of the same ideas that apply to mechanical waves apply to electromagnetic waves. One difference is that electromagnetic waves are described by two quantities (in this case, electric field $\vec{E}$ and magnetic field $\vec{B}$), rather than by a single quantity, such as the displacement of a string.

SET UP *the problem* using the following steps:

1. Draw a diagram showing the direction of wave propagation and the directions of $\vec{E}$ and $\vec{B}$.
2. Identify the target variables.

EXECUTE *the solution* as follows:

1. Review the treatment of sinusoidal mechanical waves in Chapters 15 and 16, and particularly the four problem-solving strategies suggested there.
2. Keep in mind the basic relationships for periodic waves: $v = \lambda f$ and $\omega = vk$. For electromagnetic waves in vacuum, $v = c$. Distinguish between ordinary frequency f, usually expressed in hertz, and angular frequency $\omega = 2\pi f$, expressed in rad/s. Remember that the wave number is $k = 2\pi/\lambda$.
3. Concentrate on basic relationships, such as those between $\vec{E}$ and $\vec{B}$ (magnitude, direction, and relative phase), how the wave speed is determined, and the transverse nature of the waves.

EVALUATE *your answer:* Check that your result is reasonable. For electromagnetic waves in vacuum, the magnitude of the magnetic field in teslas is much smaller (by a factor of 3.00×10^8) than the magnitude of the electric field in volts per meter. If your answer suggests otherwise, you probably made an error in using the relationship $E = cB$. (We'll see later in this section that this relationship is different for electromagnetic waves in a material medium.)

EXAMPLE 32.1 ELECTRIC AND MAGNETIC FIELDS OF A LASER BEAM

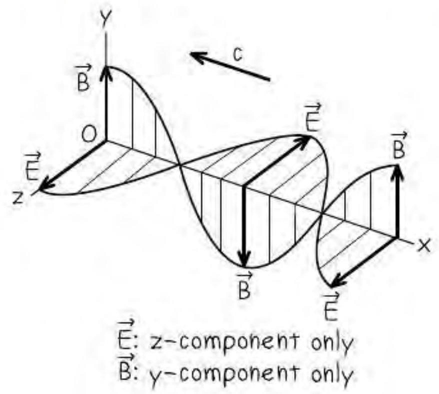
SOLUTION

A carbon dioxide laser emits a sinusoidal electromagnetic wave that travels in vacuum in the negative x-direction. The wavelength is 10.6 μm (in the infrared; see Fig. 32.4) and the $\vec{E}$ field is parallel to the z-axis, with $E_{\max} = 1.5$ MV/m. Write vector equations for $\vec{E}$ and $\vec{B}$ as functions of time and position.

SOLUTION

IDENTIFY and SET UP: Equations (32.19) describe a wave traveling in the negative x-direction with $\vec{E}$ along the y-axis—that is, a wave that is linearly polarized along the y-axis. By contrast, the wave in this example is linearly polarized along the z-axis. At points where $\vec{E}$ is in the positive z-direction, $\vec{B}$ must be in the positive y-direction for the vector product $\vec{E} \times \vec{B}$ to be in the negative x-direction (the direction of propagation). **Figure 32.15** shows a wave that satisfies these requirements.

32.15 Our sketch for this problem.



(Continued)

EXECUTE: A possible pair of wave functions that describe the wave shown in Fig. 32.15 are

$$\vec{E}(x,t) = \hat{k}E_{\max}\cos(kx + \omega t)$$
$$\vec{B}(x,t) = \hat{\jmath}B_{\max}\cos(kx + \omega t)$$

The plus sign in the arguments of the cosine functions indicates that the wave is propagating in the negative x-direction, as it should. Faraday's law requires that $E_{\max} = cB_{\max}$ [Eq. (32.18)], so

$$B_{\max} = \frac{E_{\max}}{c} = \frac{1.5 \times 10^6\ \text{V/m}}{3.0 \times 10^8\ \text{m/s}} = 5.0 \times 10^{-3}\ \text{T}$$

(Recall that $1\ \text{V} = 1\ \text{Wb/s}$ and $1\ \text{Wb/m}^2 = 1\ \text{T}$.)

We have $\lambda = 10.6 \times 10^{-6}$ m, so the wave number and angular frequency are

$$k = \frac{2\pi}{\lambda} = \frac{2\pi\ \text{rad}}{10.6 \times 10^{-6}\ \text{m}} = 5.93 \times 10^5\ \text{rad/m}$$
$$\omega = ck = (3.00 \times 10^8\ \text{m/s})(5.93 \times 10^5\ \text{rad/m}) = 1.78 \times 10^{14}\ \text{rad/s}$$

Substituting these values into the above wave functions, we get

$$\vec{E}(x,t) = \hat{k}(1.5 \times 10^6\ \text{V/m}) \times \cos[(5.93 \times 10^5\ \text{rad/m})x + (1.78 \times 10^{14}\ \text{rad/s})t]$$
$$\vec{B}(x,t) = \hat{\jmath}(5.0 \times 10^{-3}\ \text{T}) \times \cos[(5.93 \times 10^5\ \text{rad/m})x + (1.78 \times 10^{14}\ \text{rad/s})t]$$

EVALUATE: As we expect, the magnitude $B_{\max}$ in teslas is much smaller than the magnitude $E_{\max}$ in volts per meter. To check the directions of $\vec{E}$ and $\vec{B}$, note that $\vec{E} \times \vec{B}$ is in the direction of $\hat{k} \times \hat{\jmath} = -\hat{\imath}$. This is as it should be for a wave that propagates in the negative x-direction.

Our expressions for $\vec{E}(x,t)$ and $\vec{B}(x,t)$ are not the only possible solutions. We could always add a phase angle ϕ to the arguments of the cosine function, so that $kx + \omega t$ would become $kx + \omega t + \phi$. To determine the value of ϕ we would need to know $\vec{E}$ and $\vec{B}$ either as functions of x at a given time t or as functions of t at a given coordinate x. However, the statement of the problem doesn't include this information.

Electromagnetic Waves in Matter

So far, our discussion of electromagnetic waves has been restricted to waves in *vacuum*. But electromagnetic waves can also travel in *matter;* think of light traveling through air, water, or glass. In this subsection we extend our analysis to electromagnetic waves in nonconducting materials—that is, *dielectrics*.

In a dielectric the wave speed is not the same as in vacuum, and we denote it by v instead of c. Faraday's law is unaltered, but in Eq. (32.4), derived from Faraday's law, the speed c is replaced by v. In Ampere's law the displacement current is given not by $\epsilon_0\, d\Phi_E/dt$, where Φ_E is the flux of $\vec{E}$ through a surface, but by $\epsilon\, d\Phi_E/dt = K\epsilon_0\, d\Phi_E/dt$, where K is the dielectric constant and ϵ is the permittivity of the dielectric. (We introduced these quantities in Section 24.4.) Also, the constant μ_0 in Ampere's law must be replaced by $\mu = K_m\mu_0$, where K_m is the relative permeability of the dielectric and μ is its permeability (see Section 28.8). Hence Eqs. (32.4) and (32.8) are replaced by

$$E = vB \qquad \text{and} \qquad B = \epsilon\mu vE \tag{32.20}$$

Following the same procedure as for waves in vacuum, we find that

Speed of electromagnetic waves in a dielectric

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{1}{\sqrt{KK_m}}\frac{1}{\sqrt{\epsilon_0\mu_0}} = \frac{c}{\sqrt{KK_m}} \tag{32.21}$$

Permeability; Speed of light in vacuum; Permittivity; Dielectric constant; Relative permeability; Electric constant; Magnetic constant

32.16 The dielectric constant K of water is about 1.8 for visible light, so the speed of visible light in water is slower than in vacuum by a factor of $1/\sqrt{K} = 1/\sqrt{1.8} = 0.75$.

For most dielectrics the relative permeability K_m is nearly equal to unity (except for insulating ferromagnetic materials). When $K_m \cong 1$, $v = c/\sqrt{K}$. Because K is always greater than unity, the speed v of electromagnetic waves in a nonmagnetic dielectric is always *less* than the speed c in vacuum by a factor of $1/\sqrt{K}$ (**Fig. 32.16**). The ratio of the speed c in vacuum to the speed v in a material is known in optics as the **index of refraction** n of the material. When $K_m \cong 1$,

$$\frac{c}{v} = n = \sqrt{KK_m} \cong \sqrt{K} \tag{32.22}$$

Usually, we can't use the values of K in Table 24.1 in this equation because those values are measured in *constant* electric fields. When the fields oscillate rapidly, there is usually not time for the reorientation of electric dipoles that occurs with steady fields. Values of K with rapidly varying fields are usually much *smaller* than the values in the table. For example, K for water is 80.4 for steady fields but only about 1.8 in the frequency range of visible light. Thus the dielectric "constant" K is actually a function of frequency (the *dielectric function*).

SOLUTION

EXAMPLE 32.2 ELECTROMAGNETIC WAVES IN DIFFERENT MATERIALS

(a) Visiting a jewelry store one evening, you hold a diamond up to the light of a sodium-vapor street lamp. The heated sodium vapor emits yellow light with a frequency of 5.09×10^{14} Hz. Find the wavelength in vacuum and the wave speed and wavelength in diamond, for which $K = 5.84$ and $K_m = 1.00$ at this frequency. (b) A 90.0-MHz radio wave (in the FM radio band) passes from vacuum into an insulating ferrite (a ferromagnetic material used in computer cables to suppress radio interference). Find the wavelength in vacuum and the wave speed and wavelength in the ferrite, for which $K = 10.0$ and $K_m = 1000$ at this frequency.

SOLUTION

IDENTIFY and SET UP: In each case we find the wavelength in vacuum by using $c = \lambda f$. To use the corresponding equation $v = \lambda f$ to find the wavelength in a material medium, we find the speed v of electromagnetic waves in the medium from Eq. (32.21), which relates v to the values of dielectric constant K and relative permeability K_m for the medium.

EXECUTE: (a) The wavelength in vacuum of the sodium light is

$$\lambda_{\text{vacuum}} = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{5.09 \times 10^{14} \text{ Hz}} = 5.89 \times 10^{-7} \text{ m} = 589 \text{ nm}$$

The wave speed and wavelength in diamond are

$$v_{\text{diamond}} = \frac{c}{\sqrt{KK_m}} = \frac{3.00 \times 10^8 \text{ m/s}}{\sqrt{(5.84)(1.00)}} = 1.24 \times 10^8 \text{ m/s}$$

$$\lambda_{\text{diamond}} = \frac{v_{\text{diamond}}}{f} = \frac{1.24 \times 10^8 \text{ m/s}}{5.09 \times 10^{14} \text{ Hz}}$$
$$= 2.44 \times 10^{-7} \text{ m} = 244 \text{ nm}$$

(b) Following the same steps as in part (a), we find

$$\lambda_{\text{vacuum}} = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{90.0 \times 10^6 \text{ Hz}} = 3.33 \text{ m}$$

$$v_{\text{ferrite}} = \frac{c}{\sqrt{KK_m}} = \frac{3.00 \times 10^8 \text{ m/s}}{\sqrt{(10.0)(1000)}} = 3.00 \times 10^6 \text{ m/s}$$

$$\lambda_{\text{ferrite}} = \frac{v_{\text{ferrite}}}{f} = \frac{3.00 \times 10^6 \text{ m/s}}{90.0 \times 10^6 \text{ Hz}} = 3.33 \times 10^{-2} \text{ m} = 3.33 \text{ cm}$$

EVALUATE: The speed of light in transparent materials is typically between $0.2c$ and c; our result in part (a) shows that $v_{\text{diamond}} = 0.414c$. As our results in part (b) show, the speed of electromagnetic waves in dense materials like ferrite (for which $v_{\text{ferrite}} = 0.010c$) can be *far* slower than in vacuum.

TEST YOUR UNDERSTANDING OF SECTION 32.3 The first of Eqs. (32.17) gives the electric field for a plane wave as measured at points along the x-axis. For this plane wave, how does the electric field at points *off* the x-axis differ from the expression in Eqs. (32.17)? (i) The amplitude is different; (ii) the phase is different; (iii) both the amplitude and phase are different; (iv) none of these. ❙

32.4 ENERGY AND MOMENTUM IN ELECTROMAGNETIC WAVES

Electromagnetic waves carry energy; the energy in sunlight is a familiar example. Microwave ovens, radio transmitters, and lasers for eye surgery all make use of this wave energy. To understand how to utilize this energy, it's helpful to derive detailed relationships for the energy in an electromagnetic wave.

We begin with the expressions derived in Sections 24.3 and 30.3 for the **energy densities** in electric and magnetic fields; we suggest that you review those derivations now. Equations (24.11) and (30.10) show that in a region of empty space where $\vec{E}$ and $\vec{B}$ fields are present, the total energy density u is

$$u = \tfrac{1}{2}\epsilon_0 E^2 + \frac{1}{2\mu_0}B^2 \tag{32.23}$$

For electromagnetic waves in vacuum, the magnitudes E and B are related by

$$B = \frac{E}{c} = \sqrt{\epsilon_0 \mu_0}\,E \tag{32.24}$$

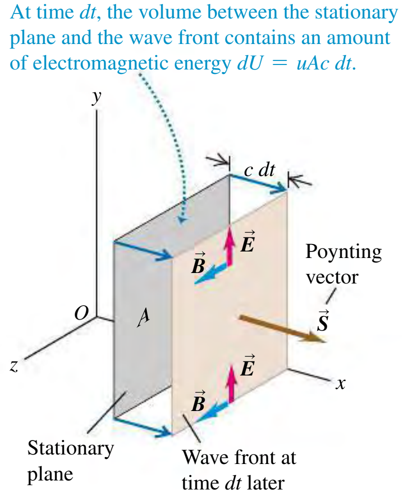
Combining Eqs. (32.23) and (32.24), we can also express the energy density u in a simple electromagnetic wave in vacuum as

$$u = \tfrac{1}{2}\epsilon_0 E^2 + \frac{1}{2\mu_0}\left(\sqrt{\epsilon_0\mu_0}\,E\right)^2 = \epsilon_0 E^2 \qquad (32.25)$$

This shows that in vacuum, the energy density associated with the $\vec{E}$ field in our simple wave is equal to the energy density of the $\vec{B}$ field. In general, the electric-field magnitude E is a function of position and time, as for the sinusoidal wave described by Eqs. (32.16); thus the energy density u of an electromagnetic wave, given by Eq. (32.25), also depends in general on position and time.

Electromagnetic Energy Flow and the Poynting Vector

32.17 A wave front at a time dt after it passes through the stationary plane with area A.



Electromagnetic waves such as those we have described are *traveling* waves that transport energy from one region to another. We can describe this energy transfer in terms of energy transferred *per unit time per unit cross-sectional area,* or *power per unit area,* for an area perpendicular to the direction of wave travel.

To see how the energy flow is related to the fields, consider a stationary plane, perpendicular to the x-axis, that coincides with the wave front at a certain time. In a time dt after this, the wave front moves a distance $dx = c\,dt$ to the right of the plane. Consider an area A on this stationary plane (**Fig. 32.17**). The energy in the space to the right of this area had to pass through the area to reach the new location. The volume dV of the relevant region is the base area A times the length $c\,dt$, and the energy dU in this region is the energy density u times this volume:

$$dU = u\,dV = (\epsilon_0 E^2)(Ac\,dt)$$

This energy passes through the area A in time dt. The energy flow per unit time per unit area, which we will call S, is

$$S = \frac{1}{A}\frac{dU}{dt} = \epsilon_0 c E^2 \qquad \text{(in vacuum)} \qquad (32.26)$$

Using Eqs. (32.4) and (32.9), you can derive the alternative forms

$$S = \frac{\epsilon_0}{\sqrt{\epsilon_0\mu_0}}E^2 = \sqrt{\frac{\epsilon_0}{\mu_0}}E^2 = \frac{EB}{\mu_0} \qquad \text{(in vacuum)} \qquad (32.27)$$

32.18 These rooftop solar panels are tilted to be face-on to the sun—that is, face-on to the Poynting vector of electromagnetic waves from the sun, so that the panels can absorb the maximum amount of wave energy.

The units of S are energy per unit time per unit area, or power per unit area. The SI unit of S is $1\ \mathrm{J/s\cdot m^2}$ or $1\ \mathrm{W/m^2}$.

We can define a *vector* quantity that describes both the magnitude and direction of the energy flow rate. Introduced by the British physicist John Poynting (1852–1914), this quantity is called the **Poynting vector:**

$$\underbrace{\vec{S}}_{\text{Poynting vector in vacuum}} = \frac{1}{\underbrace{\mu_0}_{\text{Magnetic constant}}}\underbrace{\vec{E}}_{\text{Electric field}} \times \underbrace{\vec{B}}_{\text{Magnetic field}} \qquad (32.28)$$

The vector $\vec{S}$ points in the direction of propagation of the wave (**Fig. 32.18**). Since $\vec{E}$ and $\vec{B}$ are perpendicular, the magnitude of $\vec{S}$ is $S = EB/\mu_0$; from Eqs. (32.26) and (32.27) this is the energy flow per unit area and per unit time through a cross-sectional area perpendicular to the propagation direction. The total energy flow per unit time (power, P) out of any closed surface is the integral of $\vec{S}$ over the surface:

$$P = \oint \vec{S} \cdot d\vec{A}$$

For the sinusoidal waves studied in Section 32.3, as well as for other more complex waves, the electric and magnetic fields at any point vary with time, so the Poynting vector at any point is also a function of time. Because the frequencies of typical electromagnetic waves are very high, the time variation of the Poynting vector is so rapid that it's most appropriate to look at its *average* value. The magnitude of the average value of $\vec{S}$ at a point is called the **intensity** of the radiation at that point. The SI unit of intensity is the same as for S, $1\ \mathrm{W/m^2}$.

DEMO

Let's work out the intensity of the sinusoidal wave described by Eqs. (32.17). We first substitute $\vec{E}$ and $\vec{B}$ into Eq. (32.28):

$$\vec{S}(x,t) = \frac{1}{\mu_0}\vec{E}(x,t) \times \vec{B}(x,t)$$
$$= \frac{1}{\mu_0}[\hat{\jmath}E_{\max}\cos(kx - \omega t)] \times [\hat{k}B_{\max}\cos(kx - \omega t)]$$

The vector product of the unit vectors is $\hat{\jmath} \times \hat{k} = \hat{\imath}$ and $\cos^2(kx - \omega t)$ is never negative, so $\vec{S}(x,t)$ always points in the positive x-direction (the direction of wave propagation). The x-component of the Poynting vector is

$$S_x(x,t) = \frac{E_{\max}B_{\max}}{\mu_0}\cos^2(kx - \omega t) = \frac{E_{\max}B_{\max}}{2\mu_0}[1 + \cos 2(kx - \omega t)]$$

The time average value of $\cos 2(kx - \omega t)$ is zero because at any point, it is positive during one half-cycle and negative during the other half. So the average value of the Poynting vector over a full cycle is $\vec{S}_{av} = \hat{\imath}S_{av}$, where

$$S_{av} = \frac{E_{\max}B_{\max}}{2\mu_0}$$

That is, the magnitude of the average value of $\vec{S}$ for a sinusoidal wave (the intensity I of the wave) is $\frac{1}{2}$ the maximum value. You can verify that by using the relationships $E_{\max} = B_{\max}c$ and $\epsilon_0\mu_0 = 1/c^2$, we can express the intensity in several equivalent forms:

Intensity of a sinusoidal electromagnetic wave in vacuum

Electric-field amplitude — Magnetic-field amplitude — Electric constant

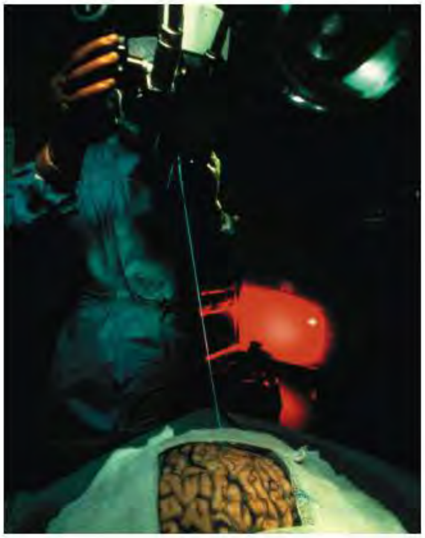
$$I = S_{av} = \frac{E_{\max}B_{\max}}{2\mu_0} = \frac{E_{\max}^{\ 2}}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_{\max}^{\ 2} = \frac{1}{2}\epsilon_0 c E_{\max}^{\ 2} \qquad (32.29)$$

Magnitude of average Poynting vector — Magnetic constant — Speed of light in vacuum

For a wave traveling in the $-x$-direction, represented by Eqs. (32.19), the Poynting vector is in the $-x$-direction at every point, but its magnitude is the same as for a wave traveling in the $+x$-direction. Verifying these statements is left to you.

Throughout this discussion we have considered only electromagnetic waves propagating in vacuum. If the waves are traveling in a dielectric medium, however, the expressions for energy density [Eq. (32.23)], the Poynting vector [Eq. (32.28)], and the intensity of a sinusoidal wave [Eq. (32.29)] must be modified. It turns out that the required modifications are quite simple: Just replace ϵ_0 with the permittivity ϵ of the dielectric, replace μ_0 with the permeability μ of the dielectric, and replace c with the speed v of electromagnetic waves in the dielectric. Remarkably, the energy densities in the $\vec{E}$ and $\vec{B}$ fields are equal even in a dielectric.

BIO Application Laser Surgery Lasers are used widely in medicine as ultra-precise, bloodless "scalpels." They can reach and remove tumors with minimal damage to neighboring healthy tissues, as in the brain surgery shown here. The power output of the laser is typically below 40 W, less than that of a typical light bulb. However, this power is concentrated into a spot from 0.1 to 2.0 mm in diameter, so the intensity of the light (equal to the average value of the Poynting vector) can be as high as 5×10^9 W/m^2.



CAUTION Poynting vector vs. intensity At any point x, the magnitude of the Poynting vector varies with time. Hence, the *instantaneous* rate at which electromagnetic energy in a sinusoidal plane wave arrives at a surface is not constant. This may seem to contradict everyday experience; the light from the sun, a light bulb, or the laser in a grocery-store scanner appears steady and unvarying in strength. In fact the Poynting vector from these sources *does* vary in time, but the variation isn't noticeable because the oscillation frequency is so high (around 5×10^{14} Hz for visible light). All that you sense is the *average* rate at which energy reaches your eye, which is why we commonly use intensity (the average value of S) to describe the strength of electromagnetic radiation.

EXAMPLE 32.3 ENERGY IN A NONSINUSOIDAL WAVE

SOLUTION

For the nonsinusoidal wave described in Section 32.2, suppose that $E = 100$ V/m $= 100$ N/C. Find the value of B, the energy density u, and the rate of energy flow per unit area S.

SOLUTION

IDENTIFY and SET UP: In this wave $\vec{E}$ and $\vec{B}$ are uniform behind the wave front (and zero ahead of it). Hence the target variables B, u, and S must also be uniform behind the wave front. Given the magnitude E, we use Eq. (32.4) to find B, Eq. (32.25) to find u, and Eq. (32.27) to find S. (We cannot use Eq. (32.29), which applies to sinusoidal waves only.)

EXECUTE: From Eq. (32.4),

$$B = \frac{E}{c} = \frac{100\ \text{V/m}}{3.00 \times 10^8\ \text{m/s}} = 3.33 \times 10^{-7}\ \text{T}$$

From Eq. (32.25),

$$u = \epsilon_0 E^2 = (8.85 \times 10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)(100\ \text{N/C})^2$$
$$= 8.85 \times 10^{-8}\ \text{N/m}^2 = 8.85 \times 10^{-8}\ \text{J/m}^3$$

(*Continued*)

The magnitude of the Poynting vector is

$$S = \frac{EB}{\mu_0} = \frac{(100\ \text{V/m})(3.33 \times 10^{-7}\ \text{T})}{4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A}}$$
$$= 26.5\ \text{V}\cdot\text{A/m}^2 = 26.5\ \text{W/m}^2$$

EVALUATE: We can check our result for S by using Eq. (32.26):

$$S = \epsilon_0 c E^2 = (8.85 \times 10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)(3.00 \times 10^8\ \text{m/s}) \times (100\ \text{N/C})^2 = 26.5\ \text{W/m}^2$$

Since $\vec{E}$ and $\vec{B}$ have the same values at all points behind the wave front, u and S likewise have the same value everywhere behind the wave front. In front of the wave front, $\vec{E} = \mathbf{0}$ and $\vec{B} = \mathbf{0}$, and so $u = 0$ and $S = 0$; where there are no fields, there is no field energy.

EXAMPLE 32.4 ENERGY IN A SINUSOIDAL WAVE

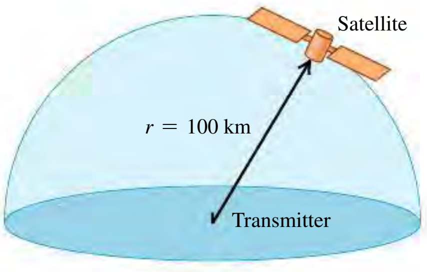
SOLUTION

A radio station on the earth's surface emits a sinusoidal wave with average total power 50 kW (**Fig. 32.19**). Assuming that the transmitter radiates equally in all directions above the ground (which is unlikely in real situations), find the electric-field and magnetic-field amplitudes E_{max} and B_{max} detected by a satellite 100 km from the antenna.

SOLUTION

IDENTIFY and SET UP: We are given the transmitter's average total power P. The intensity I is the average power per unit area; to find I at 100 km from the transmitter we divide P by the surface area of the hemisphere in Fig. 32.19. For a sinusoidal wave, I is also equal to the magnitude of the average value S_{av} of the Poynting vector, so we can use Eq. (32.29) to find E_{max}; Eq. (32.4) yields B_{max}.

32.19 A radio station radiates waves into the hemisphere shown.



EXECUTE: The surface area of a hemisphere of radius $r = 100\ \text{km} = 1.00 \times 10^5$ m is

$$A = 2\pi R^2 = 2\pi(1.00 \times 10^5\ \text{m})^2 = 6.28 \times 10^{10}\ \text{m}^2$$

All the radiated power passes through this surface, so the average power per unit area (that is, the intensity) is

$$I = \frac{P}{A} = \frac{P}{2\pi R^2} = \frac{5.00 \times 10^4\ \text{W}}{6.28 \times 10^{10}\ \text{m}^2} = 7.96 \times 10^{-7}\ \text{W/m}^2$$

From Eq. (32.29), $I = S_{av} = E_{max}{}^2/2\mu_0 c$, so

$$E_{max} = \sqrt{2\mu_0 c S_{av}}$$
$$= \sqrt{2(4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A})(3.00 \times 10^8\ \text{m/s})(7.96 \times 10^{-7}\ \text{W/m}^2)}$$
$$= 2.45 \times 10^{-2}\ \text{V/m}$$

Then from Eq. (32.4),

$$B_{max} = \frac{E_{max}}{c} = 8.17 \times 10^{-11}\ \text{T}$$

EVALUATE: Note that E_{max} is comparable to fields commonly seen in the laboratory, but B_{max} is extremely small in comparison to $\vec{B}$ fields we saw in previous chapters. For this reason, most detectors of electromagnetic radiation respond to the effect of the electric field, not the magnetic field. Loop radio antennas are an exception (see the Bridging Problem at the end of this chapter).

Electromagnetic Momentum Flow and Radiation Pressure

We've shown that electromagnetic waves transport energy. It can also be shown that electromagnetic waves carry *momentum* p, with a corresponding momentum density (momentum dp per volume dV) of magnitude

$$\frac{dp}{dV} = \frac{EB}{\mu_0 c^2} = \frac{S}{c^2} \tag{32.30}$$

This momentum is a property of the field; it is not associated with the mass of a moving particle in the usual sense.

There is also a corresponding momentum flow rate. The volume dV occupied by an electromagnetic wave (speed c) that passes through an area A in time dt is

$dV = Ac\,dt$. When we substitute this into Eq. (32.30) and rearrange, we find that the momentum flow rate per unit area is

Flow rate of electromagnetic momentum

$$\frac{1}{A}\frac{dp}{dt} = \frac{S}{c} = \frac{EB}{\mu_0 c} \tag{32.31}$$

Poynting vector magnitude
Electric-field magnitude
Magnetic-field magnitude
Speed of light in vacuum
Magnetic constant
Momentum transferred per unit surface area per unit time

We obtain the *average* rate of momentum transfer per unit area by replacing S in Eq. (32.31) by $S_{av} = I$.

This momentum is responsible for **radiation pressure.** When an electromagnetic wave is completely absorbed by a surface, the wave's momentum is also transferred to the surface. For simplicity we'll consider a surface perpendicular to the propagation direction. Using the ideas developed in Section 8.1, we see that the rate dp/dt at which momentum is transferred to the absorbing surface equals the *force* on the surface. The average force per unit area due to the wave, or *radiation pressure* p_{rad}, is the average value of dp/dt divided by the absorbing area A. (We use the subscript "rad" to distinguish pressure from momentum, for which the symbol p is also used.) From Eq. (32.31) the radiation pressure is

$$p_{rad} = \frac{S_{av}}{c} = \frac{I}{c} \quad \text{(radiation pressure, wave totally absorbed)} \tag{32.32}$$

If the wave is totally reflected, the momentum change is twice as great, and

$$p_{rad} = \frac{2S_{av}}{c} = \frac{2I}{c} \quad \text{(radiation pressure, wave totally reflected)} \tag{32.33}$$

For example, the value of I (or S_{av}) for direct sunlight, before it passes through the earth's atmosphere, is approximately $1.4\ \text{kW/m}^2$. From Eq. (32.32) the corresponding average pressure on a completely absorbing surface is

$$p_{rad} = \frac{I}{c} = \frac{1.4 \times 10^3\ \text{W/m}^2}{3.0 \times 10^8\ \text{m/s}} = 4.7 \times 10^{-6}\ \text{Pa}$$

From Eq. (32.33) the average pressure on a totally *reflecting* surface is twice this: $2I/c$ or 9.4×10^{-6} Pa. These are very small pressures, of the order of 10^{-10} atm, but they can be measured with sensitive instruments.

The radiation pressure of sunlight is much greater *inside* the sun than at the earth (see Problem 32.37). Inside stars that are much more massive and luminous than the sun, radiation pressure is so great that it substantially augments the gas pressure within the star and so helps to prevent the star from collapsing under its own gravity. In some cases the radiation pressure of stars can have dramatic effects on the material surrounding them (**Fig. 32.20**).

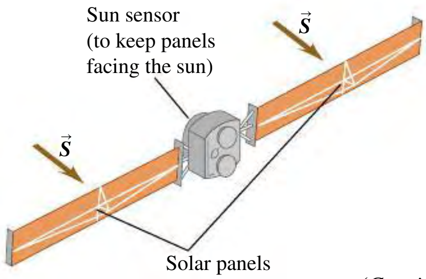
32.20 At the center of this interstellar gas cloud is a group of intensely luminous stars that exert tremendous radiation pressure on their surroundings. Aided by a "wind" of particles emanating from the stars, over the past million years the radiation pressure has carved out a bubble within the cloud 70 light-years across.

EXAMPLE 32.5 POWER AND PRESSURE FROM SUNLIGHT

SOLUTION

An earth-orbiting satellite has solar energy–collecting panels with a total area of $4.0\ \text{m}^2$ (**Fig. 32.21**). If the sun's radiation is perpendicular to the panels and is completely absorbed, find the average solar power absorbed and the average radiation-pressure force.

32.21 Solar panels on a satellite.



SOLUTION

IDENTIFY and SET UP: This problem uses the relationships among intensity, power, radiation pressure, and force. In the previous discussion, we used the intensity I (average power per unit area) of sunlight to find the radiation pressure p_{rad} (force per unit area) of sunlight on a completely absorbing surface. (These values are for

(*Continued*)

points above the atmosphere, which is where the satellite orbits.) Multiplying each value by the area of the solar panels gives the average power absorbed and the net radiation force on the panels.

EXECUTE: The intensity I (power per unit area) is $1.4 \times 10^3\ \mathrm{W/m^2}$. Although the light from the sun is not a simple sinusoidal wave, we can still use the relationship that the average power P is the intensity I times the area A:

$$P = IA = (1.4 \times 10^3\ \mathrm{W/m^2})(4.0\ \mathrm{m^2})$$
$$= 5.6 \times 10^3\ \mathrm{W} = 5.6\ \mathrm{kW}$$

The radiation pressure of sunlight on an absorbing surface is $p_{rad} = 4.7 \times 10^{-6}\ \mathrm{Pa} = 4.7 \times 10^{-6}\ \mathrm{N/m^2}$. The total force F is the pressure p_{rad} times the area A:

$$F = p_{rad}A = (4.7 \times 10^{-6}\ \mathrm{N/m^2})(4.0\ \mathrm{m^2}) = 1.9 \times 10^{-5}\ \mathrm{N}$$

EVALUATE: The absorbed power is quite substantial. Part of it can be used to power the equipment aboard the satellite; the rest goes into heating the panels, either directly or due to inefficiencies in the photocells contained in the panels.

The total radiation force is comparable to the weight (on the earth) of a single grain of salt. Over time, however, this small force can noticeably affect the orbit of a satellite like that in Fig. 32.21, and so radiation pressure must be taken into account.

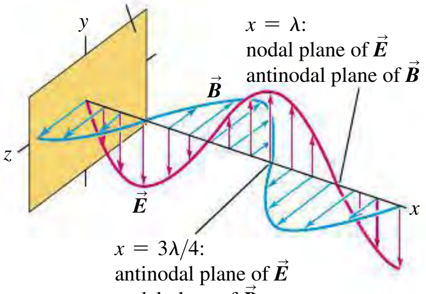
TEST YOUR UNDERSTANDING OF SECTION 32.4 Figure 32.13 shows one wavelength of a sinusoidal electromagnetic wave at time $t = 0$. For which of the following four values of x is (a) the energy density a maximum; (b) the energy density a minimum; (c) the magnitude of the instantaneous (not average) Poynting vector a maximum; (d) the magnitude of the instantaneous (not average) Poynting vector a minimum? (i) $x = 0$; (ii) $x = \lambda/4$; (iii) $x = \lambda/2$; (iv) $x = 3\lambda/4$.

32.5 STANDING ELECTROMAGNETIC WAVES

Electromagnetic waves can be *reflected* by the surface of a conductor (like a polished sheet of metal) or of a dielectric (such as a sheet of glass). The superposition of an incident wave and a reflected wave forms a **standing wave.** The situation is analogous to standing waves on a stretched string, discussed in Section 15.7.

Suppose a sheet of a perfect conductor (zero resistivity) is placed in the yz-plane of **Fig. 32.22** and a linearly polarized electromagnetic wave, traveling in the negative x-direction, strikes it. As we discussed in Section 23.4, $\vec{E}$ cannot have a component parallel to the surface of a perfect conductor. Therefore in the present situation, $\vec{E}$ must be zero everywhere in the yz-plane. The electric field of the *incident* electromagnetic wave is *not* zero at all times in the yz-plane. But this incident wave induces oscillating currents on the surface of the conductor, and these currents give rise to an additional electric field. The *net* electric field, which is the vector sum of this field and the incident $\vec{E}$, *is* zero everywhere inside and on the surface of the conductor.

32.22 Representation of the electric and magnetic fields of a linearly polarized electromagnetic standing wave when $\omega t = 3\pi/4$ rad. In any plane perpendicular to the x-axis, E is maximum (an antinode) where B is zero (a node), and vice versa. As time elapses, the pattern does *not* move along the x-axis; instead, at every point the $\vec{E}$ and $\vec{B}$ vectors simply oscillate.



The currents induced on the surface of the conductor also produce a *reflected* wave that travels out from the plane in the $+x$-direction. Suppose the incident wave is described by the wave functions of Eqs. (32.19) (a sinusoidal wave traveling in the $-x$-direction) and the reflected wave by the negative of Eqs. (32.16) (a sinusoidal wave traveling in the $+x$-direction). We take the *negative* of the wave given by Eqs. (32.16) so that the incident and reflected electric fields cancel at $x = 0$ (the plane of the conductor, where the total electric field must be zero). The superposition principle states that the total $\vec{E}$ field at any point is the vector sum of the $\vec{E}$ fields of the incident and reflected waves, and similarly for the $\vec{B}$ field. Therefore the wave functions for the superposition of the two waves are

$$E_y(x,t) = E_{max}[\cos(kx + \omega t) - \cos(kx - \omega t)]$$
$$B_z(x,t) = B_{max}[-\cos(kx + \omega t) - \cos(kx - \omega t)]$$

We can expand and simplify these expressions by using the identities

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$$

MP

PhET: Microwaves

The results are

$$E_y(x, t) = -2E_{max} \sin kx \sin \omega t \tag{32.34}$$

$$B_z(x, t) = -2B_{max} \cos kx \cos \omega t \tag{32.35}$$

Equation (32.34) is analogous to Eq. (15.28) for a stretched string. We see that at $x = 0$ the electric field $E_y(x = 0, t)$ is *always* zero; this is required by the nature of the ideal conductor, which plays the same role as a fixed point at the end of a string. Furthermore, $E_y(x, t)$ is zero at *all* times at points in those planes perpendicular to the x-axis for which $\sin kx = 0$—that is, $kx = 0, \pi, 2\pi, \ldots$. Since $k = 2\pi/\lambda$, the positions of these planes are

$$x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, \ldots \qquad \text{(nodal planes of } \vec{E}\text{)} \tag{32.36}$$

These planes are called the **nodal planes** of the $\vec{E}$ field; they are the equivalent of the nodes, or nodal points, of a standing wave on a string. Midway between any two adjacent nodal planes is a plane on which $\sin kx = \pm 1$; on each such plane, the magnitude of $E(x, t)$ equals the maximum possible value of $2E_{max}$ twice per oscillation cycle. These are the **antinodal planes** of $\vec{E}$, corresponding to the antinodes of waves on a string.

The total magnetic field is zero at all times at points in planes on which $\cos kx = 0$. These are the nodal planes of $\vec{B}$, and they occur where

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots \qquad \text{(nodal planes of } \vec{B}\text{)} \tag{32.37}$$

There is an antinodal plane of $\vec{B}$ midway between any two adjacent nodal planes.

Figure 32.22 shows a standing-wave pattern at one instant of time. The magnetic field is *not* zero at the conducting surface $(x = 0)$. The surface currents that must be present to make $\vec{E}$ exactly zero at the surface cause magnetic fields at the surface. The nodal planes of each field are separated by one half-wavelength. The nodal planes of $\vec{E}$ are midway between those of $\vec{B}$, and vice versa; hence the nodes of $\vec{E}$ coincide with the antinodes of $\vec{B}$, and conversely. Compare this situation to the distinction between pressure nodes and displacement nodes in Section 16.4.

The total electric field is a *sine* function of t, and the total magnetic field is a *cosine* function of t. The sinusoidal variations of the two fields are therefore 90° out of phase at each point. At times when $\sin \omega t = 0$, the electric field is zero *everywhere,* and the magnetic field is maximum. When $\cos \omega t = 0$, the magnetic field is zero everywhere, and the electric field is maximum. This is in contrast to a wave traveling in one direction, as described by Eqs. (32.16) or (32.19) separately, in which the sinusoidal variations of $\vec{E}$ and $\vec{B}$ at any particular point are *in phase*. You can show that Eqs. (32.34) and (32.35) satisfy the wave equation, Eq. (32.15). You can also show that they satisfy Eqs. (32.12) and (32.14), the equivalents of Faraday's and Ampere's laws.

Standing Waves in a Cavity

Let's now insert a second conducting plane, parallel to the first and a distance L from it, along the $+x$-axis. The cavity between the two planes is analogous to a stretched string held at the points $x = 0$ and $x = L$. Both conducting planes must be nodal planes for $\vec{E}$; a standing wave can exist only when the second plane is placed at one of the positions where $E(x, t) = 0$, so L must be an integer multiple of $\lambda/2$. The wavelengths that satisfy this condition are

$$\lambda_n = \frac{2L}{n} \qquad (n = 1, 2, 3, \ldots) \tag{32.38}$$

The corresponding frequencies are

$$f_n = \frac{c}{\lambda_n} = n\frac{c}{2L} \qquad (n = 1, 2, 3, \ldots) \tag{32.39}$$

Thus there is a set of *normal modes,* each with a characteristic frequency, wave shape, and node pattern (**Fig. 32.23**). By measuring the node positions, we can measure the wavelength. If the frequency is known, the wave speed can be determined. This technique was first used by Hertz in the 1880s in his pioneering investigations of electromagnetic waves.

Conducting surfaces are not the only reflectors of electromagnetic waves. Reflections also occur at an interface between two insulating materials with different dielectric or magnetic properties. The mechanical analog is a junction of two strings with equal tension but different linear mass density. In general, a wave incident on such a boundary surface is partly transmitted into the second material and partly reflected back into the first. For example, light is transmitted through a glass window, but its surfaces also reflect light.

32.23 A typical microwave oven sets up a standing electromagnetic wave with $\lambda = 12.2$ cm, a wavelength that is strongly absorbed by the water in food. Because the wave has nodes spaced $\lambda/2 = 6.1$ cm apart, the food must be rotated while cooking. Otherwise, the portion that lies at a node—where the electric-field amplitude is zero—will remain cold.

EXAMPLE 32.6 INTENSITY IN A STANDING WAVE

Calculate the intensity of the standing wave represented by Eqs. (32.34) and (32.35).

SOLUTION

IDENTIFY and SET UP: The intensity I of the wave is the time-averaged value S_{av} of the magnitude of the Poynting vector $\vec{S}$. To find S_{av}, we first use Eq. (32.28) to find the instantaneous value of $\vec{S}$ and then average it over a whole number of cycles of the wave.

EXECUTE: Using the wave functions of Eqs. (32.34) and (32.35) in Eq. (32.28) for the Poynting vector $\vec{S}$, we find

$$\begin{aligned}\vec{S}(x,t) &= \frac{1}{\mu_0}\vec{E}(x,t) \times \vec{B}(x,t)\\ &= \frac{1}{\mu_0}[-2\hat{\jmath}E_{max}\sin kx \sin \omega t] \times [-2\hat{k}B_{max}\cos kx \cos \omega t]\\ &= \hat{\imath}\frac{E_{max}B_{max}}{\mu_0}(2\sin kx \cos kx)(2\sin \omega t \cos \omega t) = \hat{\imath}S_x(x,t)\end{aligned}$$

Using the identity $\sin 2A = 2\sin A \cos A$, we can rewrite $S_x(x, t)$ as

$$S_x(x,t) = \frac{E_{max}B_{max}\sin 2kx \sin 2\omega t}{\mu_0}$$

The average value of a sine function over any whole number of cycles is zero. Thus *the time average of $\vec{S}$ at any point is zero;* $I = S_{av} = 0$.

EVALUATE: This result is what we should expect. The standing wave is a superposition of two waves with the same frequency and amplitude, traveling in opposite directions. All the energy transferred by one wave is cancelled by an equal amount transferred in the opposite direction by the other wave. When we use electromagnetic waves to transmit power, it is important to avoid reflections that give rise to standing waves.

EXAMPLE 32.7 STANDING WAVES IN A CAVITY

Electromagnetic standing waves are set up in a cavity with two parallel, highly conducting walls 1.50 cm apart. (a) Calculate the longest wavelength λ and lowest frequency f of these standing waves. (b) For a standing wave of this wavelength, where in the cavity does $\vec{E}$ have maximum magnitude? Where is $\vec{E}$ zero? Where does $\vec{B}$ have maximum magnitude? Where is $\vec{B}$ zero?

SOLUTION

IDENTIFY and SET UP: Only certain normal modes are possible for electromagnetic waves in a cavity, just as only certain normal modes are possible for standing waves on a string. The longest possible wavelength and lowest possible frequency correspond to the $n = 1$ mode in Eqs. (32.38) and (32.39); we use these to find λ and f. Equations (32.36) and (32.37) then give the locations of the nodal planes of $\vec{E}$ and $\vec{B}$. The antinodal planes of each field are midway between adjacent nodal planes.

EXECUTE: (a) From Eqs. (32.38) and (32.39), the $n = 1$ wavelength and frequency are

$$\lambda_1 = 2L = 2(1.50\ \text{cm}) = 3.00\ \text{cm}$$

$$f_1 = \frac{c}{2L} = \frac{3.00 \times 10^8\ \text{m/s}}{2(1.50 \times 10^{-2}\ \text{m})} = 1.00 \times 10^{10}\ \text{Hz} = 10\ \text{GHz}$$

(b) With $n = 1$ there is a single half-wavelength between the walls. The electric field has nodal planes ($\vec{E} = \mathbf{0}$) at the walls and an antinodal plane (where $\vec{E}$ has its maximum magnitude) midway between them. The magnetic field has *antinodal* planes at the walls and a nodal plane midway between them.

EVALUATE: One application of such standing waves is to produce an oscillating $\vec{E}$ field of definite frequency, which is used to probe the behavior of a small sample of material placed in the cavity. To subject the sample to the strongest possible field, it should be placed near the center of the cavity, at the antinode of $\vec{E}$.

TEST YOUR UNDERSTANDING OF SECTION 32.5 In the standing wave described in Example 32.7, is there any point in the cavity where the energy density is zero at all times? If so, where? If not, why not?

CHAPTER 32 SUMMARY

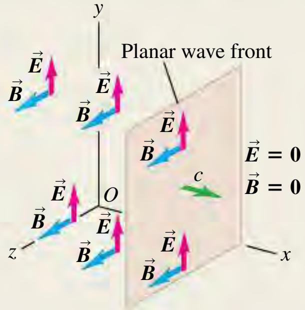
SOLUTIONS TO ALL EXAMPLES

Maxwell's equations and electromagnetic waves: Maxwell's equations predict the existence of electromagnetic waves that propagate in vacuum at the speed of light, c. The electromagnetic spectrum covers frequencies from at least 1 to 10^{24} Hz and a correspondingly broad range of wavelengths. Visible light, with wavelengths from 380 to 750 nm, is a very small part of this spectrum. In a plane wave, $\vec{E}$ and $\vec{B}$ are uniform over any plane perpendicular to the propagation direction. Faraday's law and Ampere's law give relationships between the magnitudes of $\vec{E}$ and $\vec{B}$; requiring that both relationships are satisfied gives an expression for c in terms of ϵ_0 and μ_0. Electromagnetic waves are transverse; the $\vec{E}$ and $\vec{B}$ fields are perpendicular to the direction of propagation and to each other. The direction of propagation is the direction of $\vec{E} \times \vec{B}$.

$$E = cB \quad (32.4)$$

$$B = \epsilon_0 \mu_0 c E \quad (32.8)$$

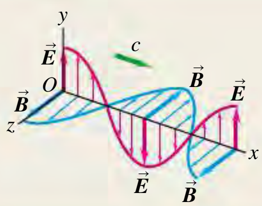
$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \quad (32.9)$$



Sinusoidal electromagnetic waves: Equations (32.17) and (32.18) describe a sinusoidal plane electromagnetic wave traveling in vacuum in the $+x$-direction. If the wave is propagating in the $-x$-direction, replace $kx - \omega t$ by $kx + \omega t$. (See Example 32.1.)

$$\vec{E}(x,t) = \hat{\jmath} E_{\max} \cos(kx - \omega t)$$
$$\vec{B}(x,t) = \hat{k} B_{\max} \cos(kx - \omega t) \quad (32.17)$$

$$E_{\max} = cB_{\max} \quad (32.18)$$



Electromagnetic waves in matter: When an electromagnetic wave travels through a dielectric, the wave speed v is less than the speed of light in vacuum c. (See Example 32.2.)

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{1}{\sqrt{KK_m}}\frac{1}{\sqrt{\epsilon_0\mu_0}} = \frac{c}{\sqrt{KK_m}} \quad (32.21)$$

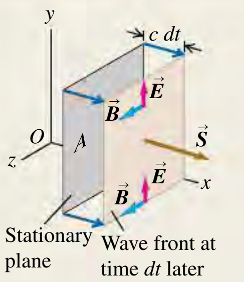
Energy and momentum in electromagnetic waves: The energy flow rate (power per unit area) in an electromagnetic wave in vacuum is given by the Poynting vector $\vec{S}$. The magnitude of the time-averaged value of the Poynting vector is called the intensity I of the wave. Electromagnetic waves also carry momentum. When an electromagnetic wave strikes a surface, it exerts a radiation pressure p_{rad}. If the surface is perpendicular to the wave propagation direction and is totally absorbing, $p_{\text{rad}} = I/c$; if the surface is a perfect reflector, $p_{\text{rad}} = 2I/c$. (See Examples 32.3–32.5.)

$$\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B} \quad (32.28)$$

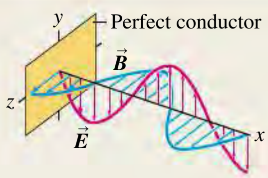
$$I = S_{\text{av}} = \frac{E_{\max}B_{\max}}{2\mu_0} = \frac{E_{\max}{}^2}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_{\max}{}^2 = \frac{1}{2}\epsilon_0 c E_{\max}{}^2 \quad (32.29)$$

$$\frac{1}{A}\frac{dp}{dt} = \frac{S}{c} = \frac{EB}{\mu_0 c} \quad (32.31)$$

(flow rate of electromagnetic momentum)



Standing electromagnetic waves: If a perfect reflecting surface is placed at $x = 0$, the incident and reflected waves form a standing wave. Nodal planes for $\vec{E}$ occur at $kx = 0, \pi, 2\pi, \ldots$, and nodal planes for $\vec{B}$ at $kx = \pi/2, 3\pi/2, 5\pi/2, \ldots$. At each point, the sinusoidal variations of $\vec{E}$ and $\vec{B}$ with time are $90°$ out of phase. (See Examples 32.6 and 32.7.)



BRIDGING PROBLEM DETECTING ELECTROMAGNETIC WAVES

A circular loop of wire can be used as a radio antenna. If an 18.0-cm-diameter antenna is located 2.50 km from a 95.0-MHz source with a total power of 55.0 kW, what is the maximum emf induced in the loop? The orientation of the antenna loop and the polarization of the wave are as shown in **Fig. 32.24**. Assume that the source radiates uniformly in all directions.

SOLUTION GUIDE

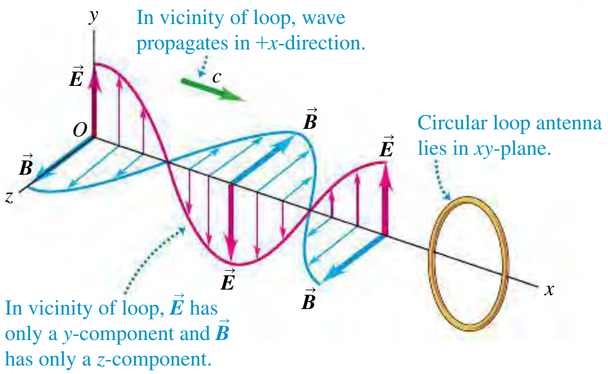
IDENTIFY and SET UP

1. The plane of the antenna loop is perpendicular to the direction of the wave's oscillating magnetic field. This causes a magnetic flux through the loop that varies sinusoidally with time. By Faraday's law, this produces an emf equal in magnitude to the rate of change of the flux. The target variable is the magnitude of this emf.
2. Select the equations that you will need to find (i) the intensity of the wave at the position of the loop, a distance $r = 2.50$ km from the source of power $P = 55.0$ kW; (ii) the amplitude of the sinusoidally varying magnetic field at that position; (iii) the magnetic flux through the loop as a function of time; and (iv) the emf produced by the flux.

EXECUTE

3. Find the wave intensity at the position of the loop.

32.24 Using a circular loop antenna to detect radio waves.



4. Use your result from step 3 to write expressions for the time-dependent magnetic field at this position and the time-dependent magnetic flux through the loop.
5. Use the results of step 4 to find the time-dependent induced emf in the loop. The amplitude of this emf is your target variable.

EVALUATE

6. Is the induced emf large enough to detect? (If it is, a receiver connected to this antenna will pick up signals from the source.)

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q32.1 By measuring the electric and magnetic fields at a point in space where there is an electromagnetic wave, can you determine the direction from which the wave came? Explain.

Q32.2 When driving on the upper level of the Bay Bridge, westbound from Oakland to San Francisco, you can easily pick up a number of radio stations on your car radio. But when driving eastbound on the lower level of the bridge, which has steel girders on either side to support the upper level, the radio reception is much worse. Why is there a difference?

Q32.3 Give several examples of electromagnetic waves that are encountered in everyday life. How are they all alike? How do they differ?

Q32.4 Sometimes neon signs located near a powerful radio station are seen to glow faintly at night, even though they are not turned on. What is happening?

Q32.5 Is polarization a property of all electromagnetic waves, or is it unique to visible light? Can sound waves be polarized? What fundamental distinction in wave properties is involved? Explain.

Q32.6 Suppose that a positive point charge q is initially at rest on the x-axis, in the path of the electromagnetic plane wave described in Section 32.2. Will the charge move after the wave front reaches it? If not, why not? If the charge does move, describe its motion qualitatively. (Remember that $\vec{E}$ and $\vec{B}$ have the same value at all points behind the wave front.)

Q32.7 The light beam from a searchlight may have an electric-field magnitude of 1000 V/m, corresponding to a potential difference of 1500 V between the head and feet of a 1.5-m-tall person on whom the light shines. Does this cause the person to feel a strong electric shock? Why or why not?

Q32.8 For a certain sinusoidal wave of intensity I, the amplitude of the magnetic field is B. What would be the amplitude (in terms of B) in a similar wave of twice the intensity?

Q32.9 The magnetic-field amplitude of the electromagnetic wave from the laser described in Example 32.1 (Section 32.3) is about 100 times greater than the earth's magnetic field. If you illuminate a compass with the light from this laser, would you expect the compass to deflect? Why or why not?

Q32.10 Most automobiles have vertical antennas for receiving radio broadcasts. Explain what this tells you about the direction of polarization of $\vec{E}$ in the radio waves used in broadcasting.

Q32.11 If a light beam carries momentum, should a person holding a flashlight feel a recoil analogous to the recoil of a rifle when it is fired? Why is this recoil not actually observed?

Q32.12 A light source radiates a sinusoidal electromagnetic wave uniformly in all directions. This wave exerts an average pressure p on a perfectly reflecting surface a distance R away from it. What average pressure (in terms of p) would this wave exert on a perfectly absorbing surface that was twice as far from the source?

Q32.13 Does an electromagnetic *standing* wave have energy? Does it have momentum? Are your answers to these questions the same as for a *traveling* wave? Why or why not?

EXERCISES

Section 32.2 Plane Electromagnetic Waves and the Speed of Light

32.1 • (a) How much time does it take light to travel from the moon to the earth, a distance of 384,000 km? (b) Light from the star Sirius takes 8.61 years to reach the earth. What is the distance from earth to Sirius in kilometers?

32.2 • Consider each of the electric- and magnetic-field orientations given next. In each case, what is the direction of propagation of the wave? (a) $\vec{E}$ in the $+x$-direction, $\vec{B}$ in the $+y$-direction; (b) $\vec{E}$ in the $-y$-direction, $\vec{B}$ in the $+x$-direction; (c) $\vec{E}$ in the $+z$-direction, $\vec{B}$ in the $-x$-direction; (d) $\vec{E}$ in the $+y$-direction, $\vec{B}$ in the $-z$-direction.

32.3 • A sinusoidal electromagnetic wave is propagating in vacuum in the $+z$-direction. If at a particular instant and at a certain point in space the electric field is in the $+x$-direction and has magnitude 4.00 V/m, what are the magnitude and direction of the magnetic field of the wave at this same point in space and instant in time?

32.4 • Consider each of the following electric- and magnetic-field orientations. In each case, what is the direction of propagation of the wave? (a) $\vec{E} = E\hat{\imath}, \vec{B} = -B\hat{\jmath}$; (b) $\vec{E} = E\hat{\jmath}, \vec{B} = B\hat{\imath}$; (c) $\vec{E} = -E\hat{k}, \vec{B} = -B\hat{\imath}$; (d) $\vec{E} = E\hat{\imath}, \vec{B} = -B\hat{k}$.

Section 32.3 Sinusoidal Electromagnetic Waves

32.5 • BIO **Medical X rays.** Medical x rays are taken with electromagnetic waves having a wavelength of around 0.10 nm in air. What are the frequency, period, and wave number of such waves?

32.6 • BIO **Ultraviolet Radiation.** There are two categories of ultraviolet light. Ultraviolet A (UVA) has a wavelength ranging from 320 nm to 400 nm. It is necessary for the production of vitamin D. UVB, with a wavelength in vacuum between 280 nm and 320 nm, is more dangerous because it is much more likely to cause skin cancer. (a) Find the frequency ranges of UVA and UVB. (b) What are the ranges of the wave numbers for UVA and UVB?

32.7 • A sinusoidal electromagnetic wave having a magnetic field of amplitude 1.25 μT and a wavelength of 432 nm is traveling in the $+x$-direction through empty space. (a) What is the frequency of this wave? (b) What is the amplitude of the associated electric field? (c) Write the equations for the electric and magnetic fields as functions of x and t in the form of Eqs. (32.17).

32.8 • An electromagnetic wave of wavelength 435 nm is traveling in vacuum in the $-z$-direction. The electric field has amplitude 2.70×10^{-3} V/m and is parallel to the x-axis. What are (a) the frequency and (b) the magnetic-field amplitude? (c) Write the vector equations for $\vec{E}(z, t)$ and $\vec{B}(z, t)$.

32.9 • Consider electromagnetic waves propagating in air. (a) Determine the frequency of a wave with a wavelength of (i) 5.0 km, (ii) 5.0 μm, (iii) 5.0 nm. (b) What is the wavelength (in meters and nanometers) of (i) gamma rays of frequency 6.50×10^{21} Hz and (ii) an AM station radio wave of frequency 590 kHz?

32.10 • The electric field of a sinusoidal electromagnetic wave obeys the equation $E = (375\ \text{V/m})\cos[(1.99 \times 10^7\ \text{rad/m})x + (5.97 \times 10^{15}\ \text{rad/s})t]$. (a) What is the speed of the wave? (b) What are the amplitudes of the electric and magnetic fields of this wave? (c) What are the frequency, wavelength, and period of the wave? Is this light visible to humans?

32.11 • An electromagnetic wave has an electric field given by $\vec{E}(y, t) = (3.10 \times 10^5\ \text{V/m})\hat{k}\cos[ky - (12.65 \times 10^{12}\ \text{rad/s})t]$. (a) In which direction is the wave traveling? (b) What is the wavelength of the wave? (c) Write the vector equation for $\vec{B}(y, t)$.

32.12 • An electromagnetic wave has a magnetic field given by $\vec{B}(x, t) = -(8.25 \times 10^{-9}\ \text{T})\hat{\jmath}\cos[(1.38 \times 10^4\ \text{rad/m})x + \omega t]$. (a) In which direction is the wave traveling? (b) What is the frequency f of the wave? (c) Write the vector equation for $\vec{E}(x, t)$.

32.13 • Radio station WCCO in Minneapolis broadcasts at a frequency of 830 kHz. At a point some distance from the transmitter, the magnetic-field amplitude of the electromagnetic wave from WCCO is 4.82×10^{-11} T. Calculate (a) the wavelength; (b) the wave number; (c) the angular frequency; (d) the electric-field amplitude.

32.14 • An electromagnetic wave with frequency 65.0 Hz travels in an insulating magnetic material that has dielectric constant 3.64 and relative permeability 5.18 at this frequency. The electric field has amplitude 7.20×10^{-3} V/m. (a) What is the speed of propagation of the wave? (b) What is the wavelength of the wave? (c) What is the amplitude of the magnetic field?

32.15 • An electromagnetic wave with frequency 5.70×10^{14} Hz propagates with a speed of 2.17×10^8 m/s in a certain piece of glass. Find (a) the wavelength of the wave in the glass; (b) the wavelength of a wave of the same frequency propagating in air; (c) the index of refraction n of the glass for an electromagnetic wave with this frequency; (d) the dielectric constant for glass at this frequency, assuming that the relative permeability is unity.

Section 32.4 Energy and Momentum in Electromagnetic Waves

32.16 • BIO **High-Energy Cancer Treatment.** Scientists are working on a new technique to kill cancer cells by zapping them with ultrahigh-energy (in the range of 10^{12} W) pulses of light that last for an extremely short time (a few nanoseconds). These short pulses scramble the interior of a cell without causing it to explode, as long pulses would do. We can model a typical such cell as a disk 5.0 μm in diameter, with the pulse lasting for 4.0 ns with an average power of 2.0×10^{12} W. We shall assume that the energy is spread uniformly over the faces of 100 cells for each pulse. (a) How much energy is given to the cell during this pulse? (b) What is the intensity (in W/m^2) delivered to the cell? (c) What are the maximum values of the electric and magnetic fields in the pulse?

32.17 •• **Fields from a Light Bulb.** We can reasonably model a 75-W incandescent light bulb as a sphere 6.0 cm in diameter. Typically, only about 5% of the energy goes to visible light; the rest goes largely to nonvisible infrared radiation. (a) What is the visible-light intensity (in W/m^2) at the surface of the bulb? (b) What are the amplitudes of the electric and magnetic fields at this surface, for a sinusoidal wave with this intensity?

32.18 •• A sinusoidal electromagnetic wave from a radio station passes perpendicularly through an open window that has area 0.500 m^2. At the window, the electric field of the wave has rms value 0.0400 V/m. How much energy does this wave carry through the window during a 30.0-s commercial?

32.19 • A space probe 2.0×10^{10} m from a star measures the total intensity of electromagnetic radiation from the star to be $5.0 \times 10^3\ \mathrm{W/m^2}$. If the star radiates uniformly in all directions, what is its total average power output?

32.20 •• The energy flow to the earth from sunlight is about $1.4\ \mathrm{kW/m^2}$. (a) Find the maximum values of the electric and magnetic fields for a sinusoidal wave of this intensity. (b) The distance from the earth to the sun is about 1.5×10^{11} m. Find the total power radiated by the sun.

32.21 • The intensity of a cylindrical laser beam is $0.800\ \mathrm{W/m^2}$. The cross-sectional area of the beam is $3.0 \times 10^{-4}\ \mathrm{m^2}$ and the intensity is uniform across the cross section of the beam. (a) What is the average power output of the laser? (b) What is the rms value of the electric field in the beam?

32.22 • A sinusoidal electromagnetic wave emitted by a cellular phone has a wavelength of 35.4 cm and an electric-field amplitude of 5.40×10^{-2} V/m at a distance of 250 m from the phone. Calculate (a) the frequency of the wave; (b) the magnetic-field amplitude; (c) the intensity of the wave.

32.23 • A monochromatic light source with power output 60.0 W radiates light of wavelength 700 nm uniformly in all directions. Calculate $E_{\max}$ and $B_{\max}$ for the 700-nm light at a distance of 5.00 m from the source.

32.24 • **Television Broadcasting.** Public television station KQED in San Francisco broadcasts a sinusoidal radio signal at a power of 777 kW. Assume that the wave spreads out uniformly into a hemisphere above the ground. At a home 5.00 km away from the antenna, (a) what average pressure does this wave exert on a totally reflecting surface, (b) what are the amplitudes of the electric and magnetic fields of the wave, and (c) what is the average density of the energy this wave carries? (d) For the energy density in part (c), what percentage is due to the electric field and what percentage is due to the magnetic field?

32.25 •• An intense light source radiates uniformly in all directions. At a distance of 5.0 m from the source, the radiation pressure on a perfectly absorbing surface is 9.0×10^{-6} Pa. What is the total average power output of the source?

32.26 • In the 25-ft Space Simulator facility at NASA's Jet Propulsion Laboratory, a bank of overhead arc lamps can produce light of intensity $2500\ \mathrm{W/m^2}$ at the floor of the facility. (This simulates the intensity of sunlight near the planet Venus.) Find the average radiation pressure (in pascals and in atmospheres) on (a) a totally absorbing section of the floor and (b) a totally reflecting section of the floor. (c) Find the average momentum density (momentum per unit volume) in the light at the floor.

32.27 •• BIO **Laser Safety.** If the eye receives an average intensity greater than $1.0 \times 10^2\ \mathrm{W/m^2}$, damage to the retina can occur. This quantity is called the *damage threshold* of the retina. (a) What is the largest average power (in mW) that a laser beam 1.5 mm in diameter can have and still be considered safe to view head-on? (b) What are the maximum values of the electric and magnetic fields for the beam in part (a)? (c) How much energy would the beam in part (a) deliver per second to the retina? (d) Express the damage threshold in $\mathrm{W/cm^2}$.

32.28 •• A laser beam has diameter 1.20 mm. What is the amplitude of the electric field of the electromagnetic radiation in this beam if the beam exerts a force of 3.8×10^{-9} N on a totally reflecting surface?

32.29 • **Laboratory Lasers.** He–Ne lasers are often used in physics demonstrations. They produce light of wavelength 633 nm and a power of 0.500 mW spread over a cylindrical beam 1.00 mm in diameter (although these quantities can vary). (a) What is the intensity of this laser beam? (b) What are the maximum values of the electric and magnetic fields? (c) What is the average energy density in the laser beam?

Section 32.5 Standing Electromagnetic Waves

32.30 • A standing electromagnetic wave in a certain material has frequency 2.20×10^{10} Hz. The nodal planes of $\vec{B}$ are 4.65 mm apart. Find (a) the wavelength of the wave in this material; (b) the distance between adjacent nodal planes of the $\vec{E}$ field; (c) the speed of propagation of the wave.

32.31 • **Microwave Oven.** The microwaves in a certain microwave oven have a wavelength of 12.2 cm. (a) How wide must this oven be so that it will contain five antinodal planes of the electric field along its width in the standing-wave pattern? (b) What is the frequency of these microwaves? (c) Suppose a manufacturing error occurred and the oven was made 5.0 cm longer than specified in part (a). In this case, what would have to be the frequency of the microwaves for there still to be five antinodal planes of the electric field along the width of the oven?

32.32 • An electromagnetic standing wave in air has frequency 75.0 MHz. (a) What is the distance between nodal planes of the $\vec{E}$ field? (b) What is the distance between a nodal plane of $\vec{E}$ and the closest nodal plane of $\vec{B}$?

PROBLEMS

32.33 •• BIO **Laser Surgery.** Very short pulses of high-intensity laser beams are used to repair detached portions of the retina of the eye. The brief pulses of energy absorbed by the retina weld the detached portions back into place. In one such procedure, a laser beam has a wavelength of 810 nm and delivers 250 mW of power spread over a circular spot 510 μm in diameter. The vitreous humor (the transparent fluid that fills most of the eye) has an index of refraction of 1.34. (a) If the laser pulses are each 1.50 ms long, how much energy is delivered to the retina with each pulse? (b) What average pressure would the pulse of the laser beam exert at normal incidence on a surface in air if the beam is fully absorbed? (c) What are the wavelength and frequency of the laser light inside the vitreous humor of the eye? (d) What are the maximum values of the electric and magnetic fields in the laser beam?

32.34 •• CALC Consider a sinusoidal electromagnetic wave with fields $\vec{E} = E_{\max}\hat{\jmath}\cos(kx - \omega t)$ and $\vec{B} = B_{\max}\hat{k}\cos(kx - \omega t + \phi)$, with $-\pi \le \phi \le \pi$. Show that if $\vec{E}$ and $\vec{B}$ are to satisfy Eqs. (32.12) and (32.14), then $E_{\max} = cB_{\max}$ and $\phi = 0$. (The result $\phi = 0$ means the $\vec{E}$ and $\vec{B}$ fields oscillate in phase.)

32.35 • A satellite 575 km above the earth's surface transmits sinusoidal electromagnetic waves of frequency 92.4 MHz uniformly in all directions, with a power of 25.0 kW. (a) What is the intensity of these waves as they reach a receiver at the surface of the earth directly below the satellite? (b) What are the amplitudes of the electric and magnetic fields at the receiver? (c) If the receiver has a totally absorbing panel measuring 15.0 cm by 40.0 cm oriented with its plane perpendicular to the direction the waves travel, what average force do these waves exert on the panel? Is this force large enough to cause significant effects?

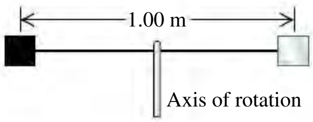
32.36 •• For a sinusoidal electromagnetic wave in vacuum, such as that described by Eq. (32.16), show that the *average* energy density in the electric field is the same as that in the magnetic field.

32.37 • The sun emits energy in the form of electromagnetic waves at a rate of 3.9×10^{26} W. This energy is produced by nuclear reactions deep in the sun's interior. (a) Find the intensity of electromagnetic radiation and the radiation pressure on an absorbing object at the surface of the sun (radius $r = R = 6.96 \times 10^5$ km) and at $r = R/2$, in the sun's interior. Ignore any scattering of the waves as they move radially outward from the center of the sun. Compare to the values given in Section 32.4 for sunlight just before it enters the earth's atmosphere. (b) The gas pressure at the sun's surface is about 1.0×10^4 Pa; at $r = R/2$, the gas pressure is calculated from solar models to be about 4.7×10^{13} Pa. Comparing with your results in part (a), would you expect that radiation pressure is an important factor in determining the structure of the sun? Why or why not?

32.38 • A small helium–neon laser emits red visible light with a power of 5.80 mW in a beam of diameter 2.50 mm. (a) What are the amplitudes of the electric and magnetic fields of this light? (b) What are the average energy densities associated with the electric field and with the magnetic field? (c) What is the total energy contained in a 1.00-m length of the beam?

32.39 •• **CP** Two square reflectors, each 1.50 cm on a side and of mass 4.00 g, are located at opposite ends of a thin, extremely light, 1.00-m rod that can rotate without friction and in vacuum about an axle perpendicular to it through its center (**Fig. P32.39**). These reflectors are small enough to be treated as point masses in moment-of-inertia calculations. Both reflectors are illuminated on one face by a sinusoidal light wave having an electric field of amplitude 1.25 N/C that falls uniformly on both surfaces and always strikes them perpendicular to the plane of their surfaces. One reflector is covered with a perfectly absorbing coating, and the other is covered with a perfectly reflecting coating. What is the angular acceleration of this device?

Figure **P32.39**



32.40 •• A source of sinusoidal electromagnetic waves radiates uniformly in all directions. At a distance of 10.0 m from this source, the amplitude of the electric field is measured to be 3.50 N/C. What is the electric-field amplitude 20.0 cm from the source?

32.41 • **CALC** **CP** A cylindrical conductor with a circular cross section has a radius a and a resistivity ρ and carries a constant current I. (a) What are the magnitude and direction of the electric-field vector $\vec{E}$ at a point just inside the wire at a distance a from the axis? (b) What are the magnitude and direction of the magnetic-field vector $\vec{B}$ at the same point? (c) What are the magnitude and direction of the Poynting vector $\vec{S}$ at the same point? (The direction of $\vec{S}$ is the direction in which electromagnetic energy flows into or out of the conductor.) (d) Use the result in part (c) to find the rate of flow of energy into the volume occupied by a length l of the conductor. (*Hint:* Integrate $\vec{S}$ over the surface of this volume.) Compare your result to the rate of generation of thermal energy in the same volume. Discuss why the energy dissipated in a current-carrying conductor, due to its resistance, can be thought of as entering through the cylindrical sides of the conductor.

32.42 •• **CP** A circular wire loop has a radius of 7.50 cm. A sinusoidal electromagnetic plane wave traveling in air passes through the loop, with the direction of the magnetic field of the wave perpendicular to the plane of the loop. The intensity of the wave at the location of the loop is $0.0275\ \text{W/m}^2$, and the wavelength of the wave is 6.90 m. What is the maximum emf induced in the loop?

32.43 • In a certain experiment, a radio transmitter emits sinusoidal electromagnetic waves of frequency 110.0 MHz in opposite directions inside a narrow cavity with reflectors at both ends, causing a standing-wave pattern to occur. (a) How far apart are the nodal planes of the magnetic field? (b) If the standing-wave pattern is determined to be in its eighth harmonic, how long is the cavity?

32.44 • The 19th-century inventor Nikola Tesla proposed to transmit electric power via sinusoidal electromagnetic waves. Suppose power is to be transmitted in a beam of cross-sectional area 100 m^2. What electric- and magnetic-field amplitudes are required to transmit an amount of power comparable to that handled by modern transmission lines (that carry voltages and currents of the order of 500 kV and 1000 A)?

32.45 •• **CP** **Global Positioning System (GPS).** The GPS network consists of 24 satellites, each of which makes two orbits around the earth per day. Each satellite transmits a 50.0-W (or even less) sinusoidal electromagnetic signal at two frequencies, one of which is 1575.42 MHz. Assume that a satellite transmits half of its power at each frequency and that the waves travel uniformly in a downward hemisphere. (a) What average intensity does a GPS receiver on the ground, directly below the satellite, receive? (*Hint:* First use Newton's laws to find the altitude of the satellite.) (b) What are the amplitudes of the electric and magnetic fields at the GPS receiver in part (a), and how long does it take the signal to reach the receiver? (c) If the receiver is a square panel 1.50 cm on a side that absorbs all of the beam, what average pressure does the signal exert on it? (d) What wavelength must the receiver be tuned to?

32.46 •• **CP** **Solar Sail.** NASA is giving serious consideration to the concept of *solar sailing.* A solar sailcraft uses a large, low-mass sail and the energy and momentum of sunlight for propulsion. (a) Should the sail be absorbing or reflective? Why? (b) The total power output of the sun is 3.9×10^{26} W. How large a sail is necessary to propel a 10,000-kg spacecraft against the gravitational force of the sun? Express your result in square kilometers. (c) Explain why your answer to part (b) is independent of the distance from the sun.

32.47 •• **CP** Interplanetary space contains many small particles referred to as *interplanetary dust.* Radiation pressure from the sun sets a lower limit on the size of such dust particles. To see the origin of this limit, consider a spherical dust particle of radius R and mass density ρ. (a) Write an expression for the gravitational force exerted on this particle by the sun (mass M) when the particle is a distance r from the sun. (b) Let L represent the luminosity of the sun, equal to the rate at which it emits energy in electromagnetic radiation. Find the force exerted on the (totally absorbing) particle due to solar radiation pressure, remembering that the intensity of the sun's radiation also depends on the distance r. The relevant area is the cross-sectional area of the particle, *not* the total surface area of the particle. As part of your answer, explain why this is so. (c) The mass density of a typical interplanetary dust particle is about 3000 kg/m^3. Find the particle radius R such that the gravitational and radiation forces acting on the particle are equal in magnitude. The luminosity of the sun is 3.9×10^{26} W. Does your answer depend on the distance of the particle from the sun? Why or why not? (d) Explain why dust particles with a radius less than that found in part (c) are unlikely to be found in the solar system. [*Hint:* Construct the ratio of the two force expressions found in parts (a) and (b).]

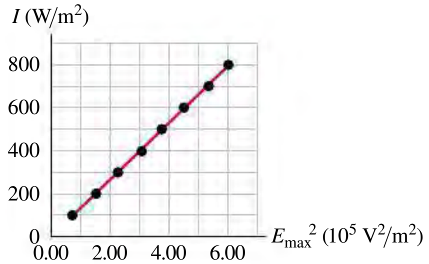
32.48 •• DATA The company where you work has obtained and stored five lasers in a supply room. You have been asked to determine the intensity of the electromagnetic radiation produced by each laser. The lasers are marked with specifications, but unfortunately different information is given for each laser:

Laser A: power $= 2.6$ W; diameter of cylindrical beam $= 2.6$ mm
Laser B: amplitude of electric field $= 480$ V/m
Laser C: amplitude of magnetic field $= 8.7 \times 10^{-6}$ T
Laser D: diameter of cylindrical beam $= 1.8$ mm; force on totally reflecting surface $= 6.0 \times 10^{-8}$ N
Laser E: average energy density in beam $= 3.0 \times 10^{-7}$ J/m^3

Calculate the intensity for each laser, and rank the lasers in order of increasing intensity. Assume that the laser beams have uniform intensity distributions over their cross sections.

32.49 •• DATA Because the speed of light in vacuum (or air) has such a large value, it is very difficult to measure directly. To measure this speed, you conduct an experiment in which you measure the amplitude of the electric field in a laser beam as you change the intensity of the beam. **Figure P32.49** is a graph of the intensity I that you measured versus the square of the amplitude E_{max} of the electric field. The best-fit straight line for your data has a slope of 1.33×10^{-3} J/(V$^2 \cdot$ s). (a) Explain why the data points plotted this way lie close to a straight line. (b) Use this graph to calculate the speed of light in air.

Figure **P32.49**

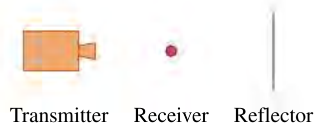


32.50 •• DATA As a physics lab instructor, you conduct an experiment on standing waves of microwaves, similar to the standing waves produced in a microwave oven. A transmitter emits microwaves of frequency f. The waves are reflected by a flat metal reflector, and a receiver measures the waves' electric-field amplitude as a function of position in the standing-wave pattern that is produced between the transmitter and reflector (**Fig. P32.50**). You measure the distance d between points of maximum amplitude (antinodes) of the electric field as a function of the frequency of the waves emitted by the transmitter. You obtain the data given in the table.

f (10^9 Hz)	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0	6.0	8.0
d (cm)	15.2	9.7	7.7	5.8	5.2	4.1	3.8	3.1	2.3	1.7

Use the data to calculate c, the speed of the electromagnetic waves in air. Because each measured value has some experimental error, plot the data in such a way that the data points will lie close to a straight line, and use the slope of that straight line to calculate c.

Figure **P32.50**



CHALLENGE PROBLEMS

32.51 ••• CP Electromagnetic radiation is emitted by accelerating charges. The rate at which energy is emitted from an accelerating charge that has charge q and acceleration a is given by

$$\frac{dE}{dt} = \frac{q^2 a^2}{6\pi\epsilon_0 c^3}$$

where c is the speed of light. (a) Verify that this equation is dimensionally correct. (b) If a proton with a kinetic energy of 6.0 MeV is traveling in a particle accelerator in a circular orbit of radius 0.750 m, what fraction of its energy does it radiate per second? (c) Consider an electron orbiting with the same speed and radius. What fraction of its energy does it radiate per second?

32.52 ••• CP The Classical Hydrogen Atom. The electron in a hydrogen atom can be considered to be in a circular orbit with a radius of 0.0529 nm and a kinetic energy of 13.6 eV. If the electron behaved classically, how much energy would it radiate per second (see Challenge Problem 32.51)? What does this tell you about the use of classical physics in describing the atom?

32.53 ••• CALC Electromagnetic waves propagate much differently in *conductors* than they do in dielectrics or in vacuum. If the resistivity of the conductor is sufficiently low (that is, if it is a sufficiently good conductor), the oscillating electric field of the wave gives rise to an oscillating conduction current that is much larger than the displacement current. In this case, the wave equation for an electric field $\vec{E}(x, t) = E_y(x, t)\hat{\jmath}$ propagating in the $+x$-direction within a conductor is

$$\frac{\partial^2 E_y(x,t)}{\partial x^2} = \frac{\mu}{\rho}\frac{\partial E_y(x,t)}{\partial t}$$

where μ is the permeability of the conductor and ρ is its resistivity. (a) A solution to this wave equation is $E_y(x, t) = E_{max} e^{-k_C x} \cos(k_C x - \omega t)$, where $k_C = \sqrt{\omega\mu/2\rho}$. Verify this by substituting $E_y(x, t)$ into the above wave equation. (b) The exponential term shows that the electric field decreases in amplitude as it propagates. Explain why this happens. (*Hint:* The field does work to move charges within the conductor. The current of these moving charges causes i^2R heating within the conductor, raising its temperature. Where does the energy to do this come from?) (c) Show that the electric-field amplitude decreases by a factor of $1/e$ in a distance $1/k_C = \sqrt{2\rho/\omega\mu}$, and calculate this distance for a radio wave with frequency $f = 1.0$ MHz in copper (resistivity $1.72 \times 10^{-8}\ \Omega \cdot$ m; permeability $\mu = \mu_0$). Since this distance is so short, electromagnetic waves of this frequency can hardly propagate at all into copper. Instead, they are reflected at the surface of the metal. This is why radio waves cannot penetrate through copper or other metals, and why radio reception is poor inside a metal structure.

PASSAGE PROBLEMS

BIO SAFE EXPOSURE TO ELECTROMAGNETIC WAVES. There have been many studies of the effects on humans of electromagnetic waves of various frequencies. Using these studies, the International Commission on Non-Ionizing Radiation Protection (ICNIRP) produced guidelines for limiting exposure to electromagnetic fields, with the goal of protecting against known adverse health effects. At frequencies of 1 Hz to 25 Hz, the maximum exposure level of electric-field amplitude E_{max} for the general public is 14 kV/m. (Different guidelines were created for people who have occupational exposure to radiation.) At frequencies of 25 Hz to 3 kHz, the corresponding E_{max} is $350/f$ V/m, where f is the frequency in kHz. (Source: "ICNIRP Statement on the 'Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic, and Electromagnetic Fields (up to 300 GHz)'," 2009; *Health Physics* 97(3): 257–258.)

32.54 In the United States, household electrical power is provided at a frequency of 60 Hz, so electromagnetic radiation at that frequency is of particular interest. On the basis of the ICNIRP guidelines, what is the maximum intensity of an electromagnetic wave at this frequency to which the general public should be exposed? (a) 7.7 W/m^2; (b) 160 W/m^2; (c) 45 kW/m^2; (d) 260 kW/m^2.

32.55 Doubling the frequency of a wave in the range of 25 Hz to 3 kHz represents what change in the maximum allowed electromagnetic-wave intensity? (a) A factor of 2; (b) a factor of $1/\sqrt{2}$; (c) a factor of $\frac{1}{2}$; (d) a factor of $\frac{1}{4}$.

32.56 The ICNIRP also has guidelines for magnetic-field exposure for the general public. In the frequency range of 25 Hz to 3 kHz, this guideline states that the maximum allowed magnetic-field amplitude is $5/f$ T, where f is the frequency in kHz. Which is a more stringent limit on allowable electromagnetic-wave intensity in this frequency range: the electric-field guideline or the magnetic-field guideline? (a) The magnetic-field guideline, because at a given frequency the allowed magnetic field is smaller than the allowed electric field. (b) The electric field guideline, because at a given frequency the allowed intensity calculated from the electric-field guideline is smaller. (c) It depends on the particular frequency chosen (both guidelines are frequency dependent). (d) Neither—for any given frequency, the guidelines represent the same electromagnetic-wave intensity.

Answers

Chapter Opening Question ?

(i) Metals are reflective because they are good conductors of electricity. When an electromagnetic wave strikes a conductor, the electric field of the wave sets up currents on the conductor surface that generate a reflected wave. For a perfect conductor, the requirement that the electric-field component parallel to the surface must be zero implies that this reflected wave is just as intense as the incident wave. Tarnished metals are less shiny because their surface is oxidized and less conductive; polishing the metal removes the oxide and exposes the conducting metal.

Test Your Understanding Questions

32.1 **(a) no, (b) no** A purely electric wave would have a varying electric field. Such a field necessarily generates a magnetic field through Ampere's law, Eq. (29.21), so a purely electric wave is impossible. In the same way, a purely magnetic wave is impossible: The varying magnetic field in such a wave would automatically give rise to an electric field through Faraday's law, Eq. (29.20).

32.2 **(a) positive y-direction, (b) negative x-direction, (c) positive y-direction** You can verify these answers by using the right-hand rule to show that $\vec{E} \times \vec{B}$ in each case is in the direction of propagation, or by using the rule shown in Fig. 32.9.

32.3 **(iv)** In an ideal electromagnetic plane wave, at any instant the fields are the same anywhere in a plane perpendicular to the direction of propagation. The plane wave described by Eqs. (32.17) is propagating in the x-direction, so the fields depend on the coordinate x and time t but do *not* depend on the coordinates y and z.

32.4 **(a) (i) and (iii), (b) (ii) and (iv), (c) (i) and (iii), (d) (ii) and (iv)** Both the energy density u and the Poynting vector magnitude S are maximum where the $\vec{E}$ and $\vec{B}$ fields have their maximum magnitudes. (The directions of the fields don't matter.) From Fig. 32.13, this occurs at $x = 0$ and $x = \lambda/2$. Both u and S have a minimum value of zero; that occurs where $\vec{E}$ and $\vec{B}$ are both zero. From Fig. 32.13, this occurs at $x = \lambda/4$ and $x = 3\lambda/4$.

32.5 **no** There are places where $\vec{E} = \mathbf{0}$ at all times (at the walls) and the electric energy density $\frac{1}{2}\epsilon_0 E^2$ is always zero. There are also places where $\vec{B} = \mathbf{0}$ at all times (on the plane midway between the walls) and the magnetic energy density $B^2/2\mu_0$ is always zero. However, there are *no* locations where both $\vec{E}$ and $\vec{B}$ are always zero. Hence the energy density at any point in the standing wave is always nonzero.

Bridging Problem

0.0368 V

? When a cut diamond is illuminated with white light, it sparkles brilliantly with a spectrum of vivid colors. These distinctive visual features are a result of (i) light traveling much slower in diamond than in air; (ii) light of different colors traveling at different speeds in diamond; (iii) diamond absorbing light of certain colors; (iv) both (i) and (ii); (v) all of (i), (ii), and (iii).

33 THE NATURE AND PROPAGATION OF LIGHT

LEARNING GOALS

Looking forward at ...

33.1 What light rays are, and how they are related to wave fronts.

33.2 The laws that govern the reflection and refraction of light.

33.3 The circumstances under which light is totally reflected at an interface.

33.4 The consequences of the speed of light in a material being different for different wavelengths.

33.5 How to make polarized light out of ordinary light.

33.6 How the scattering of light explains the blue color of the sky.

33.7 How Huygens's principle helps us analyze reflection and refraction.

Looking back at ...

1.3 Speed of light in vacuum.

21.2 Polarization of a body by an electric field.

29.7 Maxwell's equations.

32.1–32.4 Electromagnetic radiation; plane waves; wave fronts; index of refraction; electromagnetic wave intensity.

Blue lakes, ochre deserts, green forests, and multicolored rainbows can be enjoyed by anyone who has eyes with which to see them. But by studying the branch of physics called **optics,** which deals with the behavior of light and other electromagnetic waves, we can reach a deeper appreciation of the visible world. A knowledge of the properties of light allows us to understand the blue color of the sky and the design of optical devices such as telescopes, microscopes, cameras, eyeglasses, and the human eye. The same basic principles of optics also lie at the heart of modern developments such as the laser, optical fibers, holograms, and new techniques in medical imaging.

The importance of optics to physics, and to science and engineering in general, is so great that we will devote the next four chapters to its study. In this chapter we begin with a study of the laws of reflection and refraction and the concepts of dispersion, polarization, and scattering of light. Along the way we compare the various possible descriptions of light in terms of particles, rays, or waves, and we introduce Huygens's principle, an important link that connects the ray and wave viewpoints. In Chapter 34 we'll use the ray description of light to understand how mirrors and lenses work, and we'll see how mirrors and lenses are used in optical instruments such as cameras, microscopes, and telescopes. We'll explore the wave characteristics of light further in Chapters 35 and 36.

33.1 THE NATURE OF LIGHT

Until the time of Isaac Newton (1642–1727), most scientists thought that light consisted of streams of particles (called *corpuscles*) emitted by light sources. Galileo and others tried (unsuccessfully) to measure the speed of light. Around 1665, evidence of *wave* properties of light began to be discovered. By the early 19th century, evidence that light is a wave had grown very persuasive.

In 1873, James Clerk Maxwell predicted the existence of electromagnetic waves and calculated their speed of propagation, as we learned in Section 32.2. This development, along with the experimental work of Heinrich Hertz starting in 1887, showed conclusively that light is indeed an electromagnetic wave.

The Two Personalities of Light

The wave picture of light is not the whole story, however. Several effects associated with emission and absorption of light reveal a particle aspect, in that the energy carried by light waves is packaged in discrete bundles called *photons* or *quanta*. These apparently contradictory wave and particle properties have been reconciled since 1930 with the development of quantum electrodynamics, a comprehensive theory that includes *both* wave and particle properties. The *propagation* of light is best described by a wave model, but understanding emission and absorption requires a particle approach.

The fundamental sources of all electromagnetic radiation are electric charges in accelerated motion. All bodies emit electromagnetic radiation as a result of thermal motion of their molecules; this radiation, called *thermal radiation,* is a mixture of different wavelengths. At sufficiently high temperatures, all matter emits enough visible light to be self-luminous; a very hot body appears "red-hot" (**Fig. 33.1**) or "white-hot." Thus hot matter in any form is a light source. Familiar examples are a candle flame, hot coals in a campfire, and the coils in an electric toaster oven or room heater.

33.1 An electric heating element emits primarily infrared radiation. But if its temperature is high enough, it also emits a discernible amount of visible light.

Light is also produced during electrical discharges through ionized gases. The bluish light of mercury-arc lamps, the orange-yellow of sodium-vapor lamps, and the various colors of "neon" signs are familiar. A variation of the mercury-arc lamp is the *fluorescent* lamp (see Fig. 30.7). This light source uses a material called a *phosphor* to convert the ultraviolet radiation from a mercury arc into visible light. This conversion makes fluorescent lamps more efficient than incandescent lamps in transforming electrical energy into light.

In most light sources, light is emitted independently by different atoms within the source; in a *laser,* by contrast, atoms are induced to emit light in a cooperative, coherent fashion. The result is a very narrow beam of radiation that can be enormously intense and that is much more nearly *monochromatic,* or single-frequency, than light from any other source. Lasers are used by physicians for microsurgery, in a DVD or Blu-ray player to scan the information recorded on a video disc, in industry to cut through steel and to fuse high-melting-point materials, and in many other applications (**Fig. 33.2**).

33.2 Ophthalmic surgeons use lasers for repairing detached retinas and for cauterizing blood vessels in retinopathy. Pulses of blue-green light from an argon laser are ideal for this purpose, since they pass harmlessly through the transparent part of the eye but are absorbed by red pigments in the retina.

No matter what its source, electromagnetic radiation travels in vacuum at the same speed c. As we saw in Sections 1.3 and 32.1, this speed is defined to be

$$c = 2.99792458 \times 10^8 \text{ m/s}$$

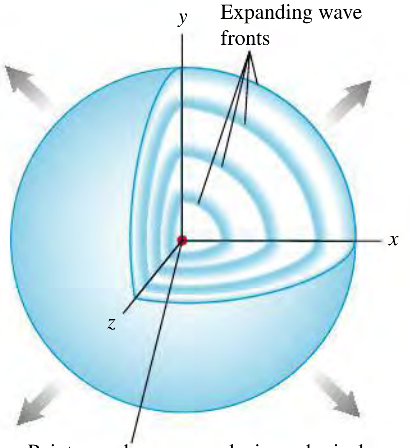
or 3.00×10^8 m/s to three significant figures. The duration of one second is defined by the cesium clock (see Section 1.3), so one meter is defined to be the distance that light travels in 1/299,792,458 s.

Waves, Wave Fronts, and Rays

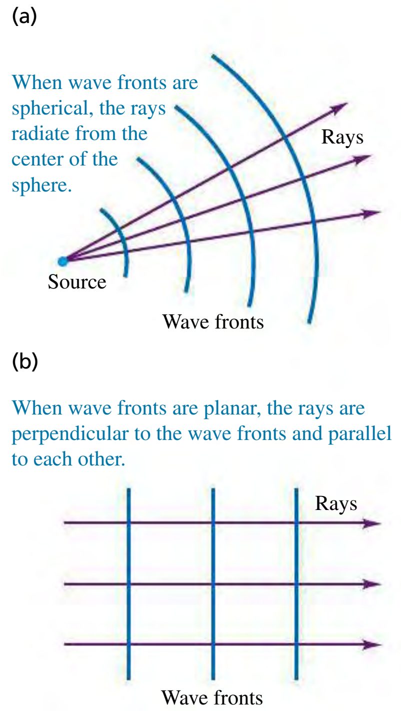
We often use the concept of a **wave front** to describe wave propagation. We introduced this concept in Section 32.2 to describe the leading edge of a wave. More generally, we define a wave front as *the locus of all adjacent points at which the phase of vibration of a physical quantity associated with the wave is the same*. That is, at any instant, all points on a wave front are at the same part of the cycle of their variation.

When we drop a pebble into a calm pool, the expanding circles formed by the wave crests, as well as the circles formed by the wave troughs between them, are wave fronts. Similarly, when sound waves spread out in still air from a

33.3 Spherical wave fronts of sound spread out uniformly in all directions from a point source in a motionless medium, such as still air, that has the same properties in all regions and in all directions. Electromagnetic waves in vacuum also spread out as shown here.



33.4 Wave fronts (blue) and rays (purple).



pointlike source, or when electromagnetic radiation spreads out from a pointlike emitter, any spherical surface that is concentric with the source is a wave front, as shown in **Fig. 33.3**. In diagrams of wave motion we usually draw only parts of a few wave fronts, often choosing consecutive wave fronts that have the same phase and thus are one wavelength apart, such as crests of water waves. Similarly, a diagram for sound waves might show only the "pressure crests," the surfaces over which the pressure is maximum, and a diagram for electromagnetic waves might show only the "crests" on which the electric or magnetic field is maximum.

We will often use diagrams that show the shapes of the wave fronts or their cross sections in some reference plane. For example, when electromagnetic waves are radiated by a small light source, we can represent the wave fronts as spherical surfaces concentric with the source or, as in **Fig. 33.4a**, by the circular intersections of these surfaces with the plane of the diagram. Far away from the source, where the radii of the spheres have become very large, a section of a spherical surface can be considered as a plane, and we have a *plane* wave like those discussed in Sections 32.2 and 32.3 (Fig. 33.4b).

To describe the directions in which light propagates, it's often convenient to represent a light wave by **rays** rather than by wave fronts. In a particle theory of light, rays are the paths of the particles. From the wave viewpoint *a ray is an imaginary line along the direction of travel of the wave*. In Fig. 33.4a the rays are the radii of the spherical wave fronts, and in Fig. 33.4b they are straight lines perpendicular to the wave fronts. When waves travel in a homogeneous isotropic material (a material with the same properties in all regions and in all directions), the rays are always straight lines normal to the wave fronts. At a boundary surface between two materials, such as the surface of a glass plate in air, the wave speed and the direction of a ray may change, but the ray segments in the air and in the glass are straight lines.

The next several chapters will give you many opportunities to see the interplay of the ray, wave, and particle descriptions of light. The branch of optics for which the ray description is adequate is called **geometric optics;** the branch dealing specifically with wave behavior is called **physical optics.** This chapter and the following one are concerned mostly with geometric optics. In Chapters 35 and 36 we will study wave phenomena and physical optics.

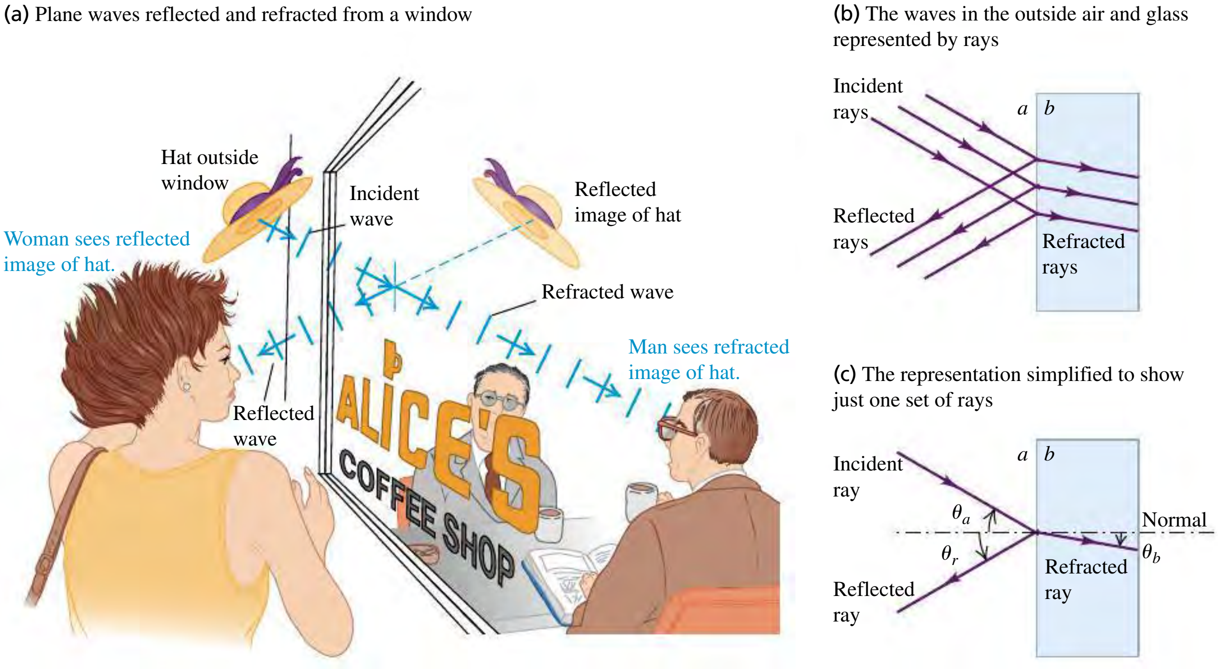
TEST YOUR UNDERSTANDING OF SECTION 33.1 Some crystals are *not* isotropic: Light travels through the crystal at a higher speed in some directions than in others. In a crystal in which light travels at the same speed in the x- and z-directions but faster in the y-direction, what would be the shape of the wave fronts produced by a light source at the origin? (i) Spherical, like those shown in Fig. 33.3; (ii) ellipsoidal, flattened along the y-axis; (iii) ellipsoidal, stretched out along the y-axis. ❙

33.2 REFLECTION AND REFRACTION

In this section we'll use the *ray* model of light to explore two of the most important aspects of light propagation: **reflection** and **refraction.** When a light wave strikes a smooth interface separating two transparent materials (such as air and glass or water and glass), the wave is in general partly *reflected* and partly *refracted* (transmitted) into the second material, as shown in **Fig. 33.5a**. For example, when you look into a restaurant window from the street, you see a reflection of the street scene, but a person inside the restaurant can look out through the window at the same scene as light reaches him by refraction.

The segments of plane waves shown in Fig. 33.5a can be represented by bundles of rays forming *beams* of light (Fig. 33.5b). For simplicity we often draw only one ray in each beam (Fig. 33.5c). Representing these waves in terms of rays is the basis of geometric optics. We begin our study with the behavior of an individual ray.

33.5 (a) A plane wave is in part reflected and in part refracted at the boundary between two media (in this case, air and glass). The light that reaches the inside of the coffee shop is refracted twice, once entering the glass and once exiting the glass. (b), (c) How light behaves at the interface between the air outside the coffee shop (material a) and the glass (material b). For the case shown here, material b has a larger index of refraction than material a ($n_b > n_a$) and the angle θ_b is smaller than θ_a.



We describe the directions of the incident, reflected, and refracted (transmitted) rays at a smooth interface between two optical materials in terms of the angles they make with the *normal* (perpendicular) to the surface at the point of incidence, as shown in Fig. 33.5c. If the interface is rough, both the transmitted light and the reflected light are scattered in various directions, and there is no single angle of transmission or reflection. Reflection at a definite angle from a very smooth surface is called **specular reflection** (from the Latin word for "mirror"); scattered reflection from a rough surface is called **diffuse reflection** (**Fig. 33.6**). Both kinds of reflection can occur with either transparent materials or *opaque* materials that do not transmit light. The vast majority of objects in your environment (including plants, other people, and this book) are visible to you because they reflect light in a diffuse manner from their surfaces. Our primary concern, however, will be with specular reflection from a very smooth surface such as highly polished glass or metal. Unless stated otherwise, when referring to "reflection" we will always mean *specular* reflection.

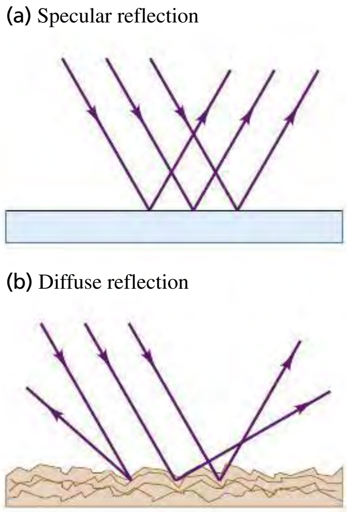
The **index of refraction** of an optical material (also called the **refractive index**), denoted by n, plays a central role in geometric optics:

$$\underbrace{n}_{\text{Index of refraction of an optical material}} = \frac{c}{v} \qquad \begin{array}{l} c\text{: Speed of light in vacuum} \\ v\text{: Speed of light in the material} \end{array} \tag{33.1}$$

Light always travels *more slowly* in a material than in vacuum, so the value of n in anything other than vacuum is always greater than unity. For vacuum, $n = 1$. Since n is a ratio of two speeds, it is a pure number without units. (In Section 32.3 we described the relationship of the value of n to the electric and magnetic properties of a material.)

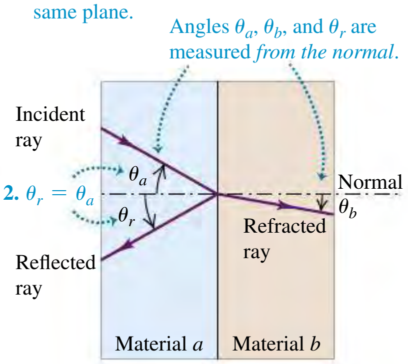
CAUTION **Wave speed and index of refraction** Keep in mind that the wave speed v is *inversely* proportional to the index of refraction n. The greater the index of refraction in a material, the *slower* the wave speed in that material.

33.6 Two types of reflection.



33.7 The laws of reflection and refraction.

1. The incident, reflected, and refracted rays and the normal to the surface all lie in the same plane.



3. When a monochromatic light ray crosses the interface between two given materials a and b, the angles θ_a and θ_b are related to the indexes of refraction of a and b by

$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{n_b}{n_a}$$

The Laws of Reflection and Refraction

Experimental studies of reflection and refraction at a smooth interface between two optical materials lead to the following conclusions (**Fig. 33.7**):

1. **The incident, reflected, and refracted rays and the normal to the surface all lie in the same plane.** This plane, called the **plane of incidence,** is perpendicular to the plane of the boundary surface between the two materials. We always draw ray diagrams so that the incident, reflected, and refracted rays are in the plane of the diagram.
2. **The angle of reflection θ_r is equal to the angle of incidence θ_a for all wavelengths and for any pair of materials.** That is, in Fig. 33.5c,

Law of reflection: Angle of reflection (measured from normal) Angle of incidence (measured from normal)

$$\theta_r = \theta_a \tag{33.2}$$

This relationship, together with the observation that the incident and reflected rays and the normal all lie in the same plane, is called the **law of reflection.**

3. For monochromatic light and for a given pair of materials, a and b, on opposite sides of the interface, **the ratio of the sines of the angles θ_a and θ_b, where both angles are measured from the normal to the surface, is equal to the inverse ratio of the two indexes of refraction:**

$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{n_b}{n_a} \tag{33.3}$$

or

Law of refraction: Angle of incidence (measured from normal) Angle of refraction (measured from normal) Index of refraction for material with incident light Index of refraction for material with refracted light

$$n_a \sin\theta_a = n_b \sin\theta_b \tag{33.4}$$

This result, together with the observation that the incident and refracted rays and the normal all lie in the same plane, is called the **law of refraction** or **Snell's law,** after the Dutch scientist Willebrord Snell (1591–1626). This law was actually first discovered in the 10th century by the Persian scientist Ibn Sahl. The discovery that $n = c/v$ came much later.

While these results were first observed experimentally, they can be derived theoretically from a wave description of light. We do this in Section 33.7.

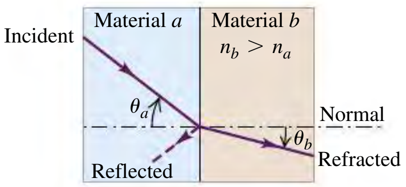
Equations (33.3) and (33.4) show that when a ray passes from one material (a) into another material (b) having a larger index of refraction $(n_b > n_a)$ and hence a slower wave speed, the angle θ_b with the normal is *smaller* in the second material than the angle θ_a in the first; hence the ray is bent *toward* the normal (**Fig. 33.8a**). When the second material has a *smaller* index of refraction than the first material $(n_b < n_a)$ and hence a faster wave speed, the ray is bent *away from* the normal (Fig. 33.8b).

No matter what the materials on either side of the interface, in the case of *normal* incidence the transmitted ray is not bent at all (Fig. 33.8c). In this case $\theta_a = 0$ and $\sin\theta_a = 0$, so from Eq. (33.4) θ_b is also equal to zero; the transmitted ray is also normal to the interface. From Eq. (33.2), θ_r is also equal to zero, so the reflected ray travels back along the same path as the incident ray.

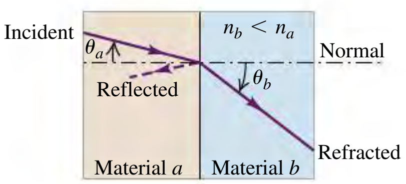
The law of refraction explains why a partially submerged ruler or drinking straw appears bent; light rays coming from below the surface change in direction at the air–water interface, so the rays appear to be coming from a position above

33.8 Refraction and reflection in three cases. (a) Material b has a larger index of refraction than material a. (b) Material b has a smaller index of refraction than material a. (c) The incident light ray is normal to the interface between the materials.

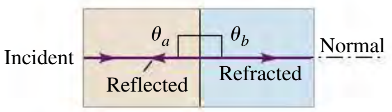
(a) A ray entering a material of *larger* index of refraction bends *toward* the normal.



(b) A ray entering a material of *smaller* index of refraction bends *away from* the normal.



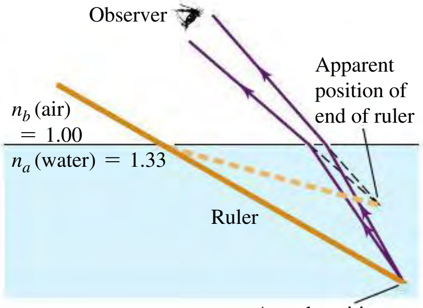
(c) A ray oriented along the normal does not bend, regardless of the materials.



33.9 (a) This ruler is actually straight, but it appears to bend at the surface of the water. (b) Light rays from any submerged object bend away from the normal when they emerge into the air. As seen by an observer above the surface of the water, the object appears to be much closer to the surface than it actually is.

(a) A straight ruler half-immersed in water

(b) Why the ruler appears bent



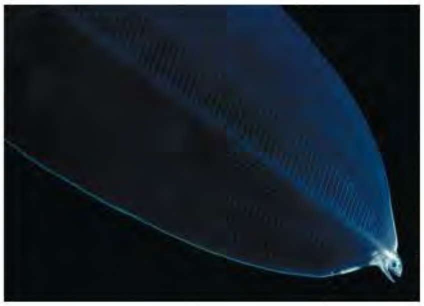
their actual point of origin (**Fig. 33.9**). A similar effect explains the appearance of the setting sun (**Fig. 33.10**).

An important special case is refraction that occurs at an interface between vacuum, for which the index of refraction is unity by definition, and a material. When a ray passes from vacuum into a material (b), so that $n_a = 1$ and $n_b > 1$, the ray is always bent *toward* the normal. When a ray passes from a material into vacuum, so that $n_a > 1$ and $n_b = 1$, the ray is always bent *away from* the normal.

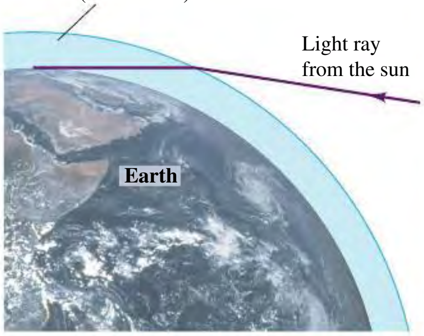
The laws of reflection and refraction apply regardless of which side of the interface the incident ray comes from. If a ray of light approaches the interface in Fig. 33.8a or 33.8b from the right rather than from the left, there are again reflected and refracted rays, and they lie in the same plane as the incident ray and the normal to the surface. Furthermore, the path of a refracted ray is *reversible;* it follows the same path when going from b to a as when going from a to b. [You can verify this by using Eq. (33.4).] Since the reflected and incident angles are the same, the path of a reflected ray is also reversible. That's why when you see someone's eyes in a mirror, they can also see you.

The *intensities* of the reflected and refracted rays depend on the angle of incidence, the two indexes of refraction, and the polarization (that is, the direction of the electric-field vector) of the incident ray. The fraction reflected is smallest at normal incidence ($\theta_a = 0°$), where it is about 4% for an air–glass interface. This fraction increases with increasing angle of incidence to 100% at grazing incidence, when $\theta_a = 90°$. (It's possible to use Maxwell's equations to predict the amplitude, intensity, phase, and polarization states of the reflected and refracted waves. Such an analysis is beyond our scope, however.)

BIO Application Transparency and Index of Refraction An eel in its larval stage is nearly as transparent as the seawater in which it swims. The larva in this photo is nonetheless easy to see because its index of refraction is higher than that of seawater, so that some of the light striking it is reflected instead of transmitted. The larva appears particularly shiny around its edges because the light reaching the camera from those points struck the larva at near-grazing incidence ($\theta_a = 90°$), resulting in almost 100% reflection.



(a)



(b)

33.10 (a) The index of refraction of air is slightly greater than 1, so light rays from the setting sun bend downward when they enter our atmosphere. (The effect is exaggerated in this figure.) (b) Stronger refraction occurs for light coming from the lower limb of the sun (the part that appears closest to the horizon), which passes through denser air in the lower atmosphere. As a result, the setting sun appears flattened vertically. (See Problem 33.51.)

TABLE 33.1 Index of Refraction for Yellow Sodium Light, $\lambda_0 = 589$ nm

Substance	Index of Refraction, n
Solids	
Ice (H_2O)	1.309
Fluorite (CaF_2)	1.434
Polystyrene	1.49
Rock salt (NaCl)	1.544
Quartz (SiO_2)	1.544
Zircon ($ZrO_2 \cdot SiO_2$)	1.923
Diamond (C)	2.417
Fabulite ($SrTiO_3$)	2.409
Rutile (TiO_2)	2.62
Glasses (typical values)	
Crown	1.52
Light flint	1.58
Medium flint	1.62
Dense flint	1.66
Lanthanum flint	1.80
Liquids at 20°C	
Methanol (CH_3OH)	1.329
Water (H_2O)	1.333
Ethanol (C_2H_5OH)	1.36
Carbon tetrachloride (CCl_4)	1.460
Turpentine	1.472
Glycerine	1.473
Benzene	1.501
Carbon disulfide (CS_2)	1.628

The index of refraction depends not only on the substance but also on the wavelength of the light. The dependence on wavelength is called *dispersion;* we will consider it in Section 33.4. Indexes of refraction for several solids and liquids are given in **Table 33.1** for a particular wavelength of yellow light.

The index of refraction of air at standard temperature and pressure is about 1.0003, and we will usually take it to be exactly unity. The index of refraction of a gas increases as its density increases. Most glasses used in optical instruments have indexes of refraction between about 1.5 and 2.0. A few substances have larger indexes; one example is diamond, with 2.417 (see Table 33.1).

Index of Refraction and the Wave Aspects of Light

We have discussed how the direction of a light ray changes when it passes from one material to another material with a different index of refraction. What aspects of the *wave* characteristics of the light change when this happens?

First, the frequency f of the wave does *not* change when passing from one material to another. That is, the number of wave cycles arriving per unit time must equal the number leaving per unit time; this is a statement that the boundary surface cannot create or destroy waves.

Second, the wavelength λ of the wave *is* different in general in different materials. This is because in any material, $v = \lambda f$; since f is the same in any material as in vacuum and v is always less than the wave speed c in vacuum, λ is also correspondingly reduced. Thus the wavelength λ of light in a material is *less than* the wavelength λ_0 of the same light in vacuum. From the above discussion, $f = c/\lambda_0 = v/\lambda$. Combining this with Eq. (33.1), $n = c/v$, we find

$$\underbrace{\lambda}_{\text{Wavelength of light in a material}} = \frac{\overbrace{\lambda_0}^{\text{Wavelength of light in vacuum}}}{\underbrace{n}_{\text{Index of refraction of the material}}} \qquad (33.5)$$

When a wave passes from one material into a second material with larger index of refraction, so $n_b > n_a$, the wave speed decreases. The wavelength $\lambda_b = \lambda_0/n_b$ in the second material is then shorter than the wavelength $\lambda_a = \lambda_0/n_a$ in the first material. If instead the second material has a smaller index of refraction than the first material, so $n_b < n_a$, then the wave speed increases. Then the wavelength λ_b in the second material is longer than the wavelength λ_a in the first material. This makes intuitive sense; the waves get "squeezed" (the wavelength gets shorter) if the wave speed decreases and get "stretched" (the wavelength gets longer) if the wave speed increases.

PROBLEM-SOLVING STRATEGY 33.1 REFLECTION AND REFRACTION

IDENTIFY *the relevant concepts:* Use geometric optics, discussed in this section, whenever light (or electromagnetic radiation of *any* frequency and wavelength) encounters a boundary between materials. In general, part of the light is reflected back into the first material and part is refracted into the second material.

SET UP *the problem* using the following steps:

1. In problems involving rays and angles, start by drawing a large, neat diagram. Label all known angles and indexes of refraction.
2. Identify the target variables.

EXECUTE *the solution* as follows:

1. Apply the laws of reflection, Eq. (33.2), and refraction, Eq. (33.4). Measure angles of incidence, reflection, and refraction with respect to the *normal* to the surface, *never* from the surface itself.
2. Apply geometry or trigonometry in working out angular relationships. Remember that the sum of the acute angles of a right triangle is 90° (they are *complementary*) and the sum of the interior angles in any triangle is 180°.
3. The frequency of the electromagnetic radiation does not change when it moves from one material to another; the wavelength changes in accordance with Eq. (33.5), $\lambda = \lambda_0/n$.

EVALUATE *your answer:* In problems that involve refraction, check that your results are consistent with Snell's law ($n_a \sin\theta_a = n_b \sin\theta_b$). If the second material has a higher index of refraction than the first, the angle of refraction must be *smaller* than the angle of incidence: The refracted ray bends toward the normal. If the first material has the higher index of refraction, the refracted angle must be larger than the incident angle: The refracted ray bends away from the normal.

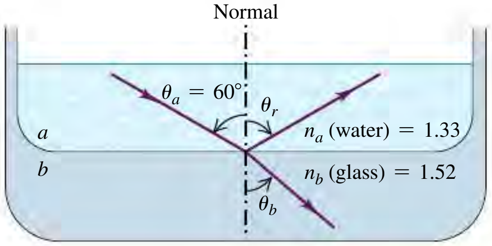
EXAMPLE 33.1 REFLECTION AND REFRACTION

In **Fig. 33.11**, material *a* is water and material *b* is glass with index of refraction 1.52. The incident ray makes an angle of 60.0° with the normal; find the directions of the reflected and refracted rays.

SOLUTION

IDENTIFY and SET UP: This is a problem in geometric optics. We are given the angle of incidence $\theta_a = 60.0°$ and the indexes of refraction $n_a = 1.33$ and $n_b = 1.52$. We must find the angles of reflection and refraction θ_r and θ_b; to do this we use Eqs. (33.2) and (33.4), respectively. Figure 33.11 shows the rays and angles; n_b is slightly greater than n_a, so by Snell's law [Eq. (33.4)] θ_b is slightly smaller than θ_a.

33.11 Reflection and refraction of light passing from water to glass.



EXECUTE: According to Eq. (33.2), the angle the reflected ray makes with the normal is the same as that of the incident ray, so $\theta_r = \theta_a = 60.0°$.

To find the direction of the refracted ray we use Snell's law, Eq. (33.4):

$$n_a \sin\theta_a = n_b \sin\theta_b$$

$$\sin\theta_b = \frac{n_a}{n_b}\sin\theta_a = \frac{1.33}{1.52}\sin 60.0° = 0.758$$

$$\theta_b = \arcsin(0.758) = 49.3°$$

EVALUATE: The second material has a larger refractive index than the first, as in Fig. 33.8a. Hence the refracted ray is bent toward the normal and $\theta_b < \theta_a$.

EXAMPLE 33.2 INDEX OF REFRACTION IN THE EYE

The wavelength of the red light from a helium-neon laser is 633 nm in air but 474 nm in the aqueous humor inside your eyeball. Calculate the index of refraction of the aqueous humor and the speed and frequency of the light in it.

SOLUTION

IDENTIFY and SET UP: The key ideas here are (i) the definition of index of refraction n in terms of the wave speed v in a medium and the speed c in vacuum, and (ii) the relationship between wavelength λ_0 in vacuum and wavelength λ in a medium of index n. We use Eq. (33.1), $n = c/v$; Eq. (33.5), $\lambda = \lambda_0/n$; and $v = \lambda f$.

EXECUTE: The index of refraction of air is very close to unity, so we assume that the wavelength λ_0 in vacuum is the same as that in air, 633 nm. Then from Eq. (33.5),

$$\lambda = \frac{\lambda_0}{n} \qquad n = \frac{\lambda_0}{\lambda} = \frac{633\text{ nm}}{474\text{ nm}} = 1.34$$

This is about the same index of refraction as for water. Then, using $n = c/v$ and $v = \lambda f$, we find

$$v = \frac{c}{n} = \frac{3.00 \times 10^8\text{ m/s}}{1.34} = 2.25 \times 10^8\text{ m/s}$$

$$f = \frac{v}{\lambda} = \frac{2.25 \times 10^8\text{ m/s}}{474 \times 10^{-9}\text{ m}} = 4.74 \times 10^{14}\text{ Hz}$$

EVALUATE: Although the speed and wavelength have different values in air and in the aqueous humor, the *frequency* in air, f_0, is the same as the frequency f in the aqueous humor:

$$f_0 = \frac{c}{\lambda_0} = \frac{3.00 \times 10^8\text{ m/s}}{633 \times 10^{-9}\text{ m}} = 4.74 \times 10^{14}\text{ Hz}$$

When a light wave passes from one material into another, both the wave speed and wavelength change but the wave frequency is unchanged.

EXAMPLE 33.3 A TWICE-REFLECTED RAY

Two mirrors are perpendicular to each other. A ray traveling in a plane perpendicular to both mirrors is reflected from one mirror at *P*, then the other at *Q*, as shown in **Fig. 33.12** (next page). What is the ray's final direction relative to its original direction?

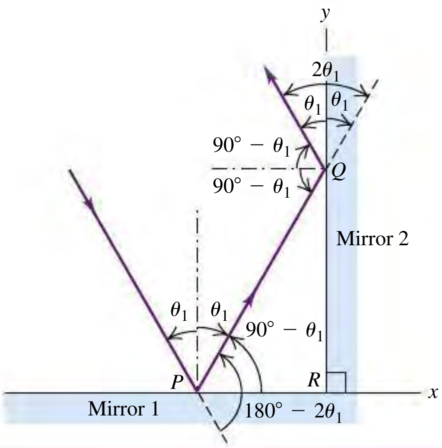
SOLUTION

IDENTIFY and SET UP: This problem involves the law of reflection, which we must apply twice (once for each mirror).

EXECUTE: For mirror 1 the angle of incidence is θ_1, and this equals the angle of reflection. The sum of interior angles in the triangle *PQR* is 180°, so we see that the angles of both incidence and reflection for mirror 2 are $90° - \theta_1$. The total change in direction of the ray after both reflections is therefore $2(90° - \theta_1) + 2\theta_1 = 180°$. That is, the ray's final direction is opposite to its original direction.

Continued

33.12 A ray moving in the *xy*-plane. The first reflection changes the sign of the *y*-component of its velocity, and the second reflection changes the sign of the *x*-component.



EVALUATE: An alternative viewpoint is that reflection reverses the sign of the component of light velocity perpendicular to the surface but leaves the other components unchanged. We invite you to verify this in detail. You should also be able to use this result to show that when a ray of light is successively reflected by three mirrors forming a corner of a cube (a "corner reflector"), its final direction is again opposite to its original direction. This principle is widely used in tail-light lenses and bicycle reflectors to improve their night-time visibility. *Apollo* astronauts placed arrays of corner reflectors on the moon. By use of laser beams reflected from these arrays, the earth–moon distance has been measured to within 0.15 m.

TEST YOUR UNDERSTANDING OF SECTION 33.2 You are standing on the shore of a lake. You spot a tasty fish swimming some distance below the lake surface. (a) If you want to spear the fish, should you aim the spear (i) above, (ii) below, or (iii) directly at the apparent position of the fish? (b) If instead you use a high-power laser to simultaneously kill and cook the fish, should you aim the laser (i) above, (ii) below, or (iii) directly at the apparent position of the fish?

33.3 TOTAL INTERNAL REFLECTION

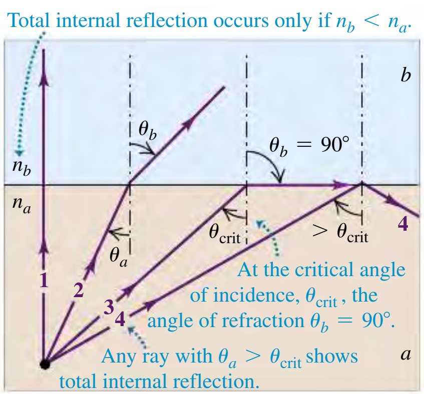
We have described how light is partially reflected and partially transmitted at an interface between two materials with different indexes of refraction. Under certain circumstances, however, *all* of the light can be reflected back from the interface, with none of it being transmitted, even though the second material is transparent. **Figure 33.13a** shows how this can occur. Several rays are shown radiating from a point source in material a with index of refraction n_a. The rays strike the surface of a second material b with index n_b, where $n_a > n_b$. (Materials a and b could be water and air, respectively.) From Snell's law of refraction,

$$\sin\theta_b = \frac{n_a}{n_b}\sin\theta_a$$

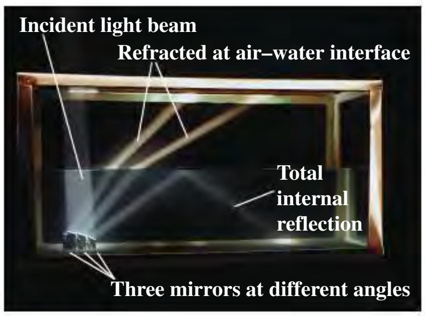
Because n_a/n_b is greater than unity, $\sin\theta_b$ is larger than $\sin\theta_a$; the ray is bent *away from* the normal. Thus there must be some value of θ_a *less than* 90° for which $\sin\theta_b = 1$ and $\theta_b = 90°$. This is shown by ray 3 in the diagram, which emerges just grazing the surface at an angle of refraction of 90°. Compare Fig. 33.13a to the photograph of light rays in Fig. 33.13b.

33.13 (a) Total internal reflection. The angle of incidence for which the angle of refraction is 90° is called the critical angle: This is the case for ray 3. The reflected portions of rays 1, 2, and 3 are omitted for clarity. (b) Rays of laser light enter the water in the fishbowl from above; they are reflected at the bottom by mirrors tilted at slightly different angles. One ray undergoes total internal reflection at the air–water interface.

(a) Total internal reflection



(b) A light beam enters the top left of the tank, then reflects at the bottom from mirrors tilted at different angles. One beam undergoes total internal reflection at the air–water interface.



The angle of incidence for which the refracted ray emerges tangent to the surface is called the **critical angle,** denoted by θ_{crit}. (A more detailed analysis using Maxwell's equations shows that as the incident angle approaches the critical angle, the transmitted intensity approaches zero.) If the angle of incidence is *larger* than the critical angle, $\sin\theta_b$ would have to be greater than unity, which is impossible. Beyond the critical angle, the ray *cannot* pass into the upper material; it is trapped in the lower material and is completely reflected at the boundary surface. This situation, called **total internal reflection,** occurs only when a ray in material a is incident on a second material b whose index of refraction is *smaller* than that of material a (that is, $n_b < n_a$).

We can find the critical angle for two given materials a and b by setting $\theta_b = 90°$ $(\sin\theta_b = 1)$ in Snell's law. We then have

Critical angle for total internal reflection

$$\sin\theta_{crit} = \frac{n_b}{n_a} \tag{33.6}$$

n_b: Index of refraction of second material

n_a: Index of refraction of first material

Total internal reflection will occur if the angle of incidence θ_a is larger than or equal to θ_{crit}.

DATA SPEAKS

Reflection and Refraction

When students were given a problem involving reflection and refraction, more than 55% gave an incorrect response. Common errors:

- Forgetting that the paths of reflected and refracted rays are reversible. If a light ray travels a path from point *A* to point *B*, it will also travel a path from point *B* to point *A*.
- Confusion about angles. The angles of incidence, reflection, and refraction are always measured from the *normal* to the interface between two materials. Furthermore, the angle of refraction cannot exceed 90°.

Applications of Total Internal Reflection

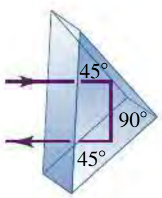
Total internal reflection finds numerous uses in optical technology. As an example, consider glass with index of refraction $n = 1.52$. If light propagating within this glass encounters a glass–air interface, the critical angle is

$$\sin\theta_{crit} = \frac{1}{1.52} = 0.658 \qquad \theta_{crit} = 41.1°$$

The light will be *totally reflected* if it strikes the glass–air surface at an angle of 41.1° or larger. Because the critical angle is slightly smaller than 45°, it is possible to use a prism with angles of 45°−45°−90° as a totally reflecting surface. As reflectors, totally reflecting prisms have some advantages over metallic surfaces such as ordinary coated-glass mirrors. While no metallic surface reflects 100% of the light incident on it, light can be *totally* reflected by a prism. These reflecting properties of a prism are unaffected by tarnishing.

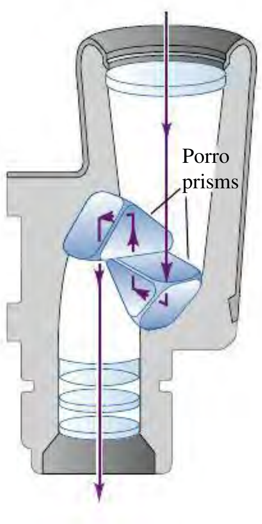
A 45°–45°–90° prism, used as in **Fig. 33.14a**, is called a *Porro* prism. Light enters and leaves at right angles to the hypotenuse and is totally reflected at each of the shorter faces. The total change of direction of the rays is 180°. Binoculars often use combinations of two Porro prisms, as in Fig. 33.14b.

(a) Total internal reflection in a Porro prism



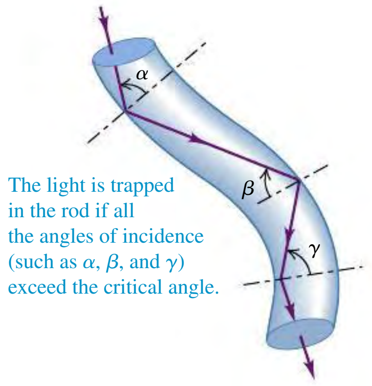
If the incident beam is oriented as shown, total internal reflection occurs on the 45° faces (because, for a glass–air interface, $\theta_{crit} = 41.1$).

(b) Binoculars use Porro prisms to reflect the light to each eyepiece.

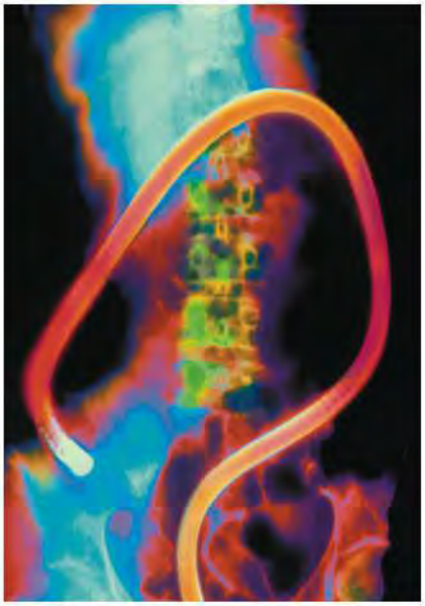


33.14 (a) Total internal reflection in a Porro prism. (b) A combination of two Porro prisms in binoculars.

33.15 A transparent rod with refractive index greater than that of the surrounding material.



33.16 This colored x-ray image of a patient's abdomen shows an endoscope winding through the colon.



When a beam of light enters at one end of a transparent rod (**Fig. 33.15**), the light can be totally reflected internally if the index of refraction of the rod is greater than that of the surrounding material. The light is "trapped" within even a curved rod, provided that the curvature is not too great. A bundle of fine glass or plastic fibers behaves in the same way and has the advantage of being flexible. A bundle may consist of thousands of individual fibers, each of the order of 0.002 to 0.01 mm in diameter. If the fibers are assembled in the bundle so that the relative positions of the ends are the same (or mirror images) at both ends, the bundle can transmit an image.

Fiber-optic devices have found a wide range of medical applications in instruments called *endoscopes,* which can be inserted directly into the bronchial tubes, the bladder, the colon, and other organs for direct visual examination (**Fig. 33.16**). A bundle of fibers can even be enclosed in a hypodermic needle for studying tissues and blood vessels far beneath the skin.

Fiber optics also have applications in communication systems. The rate at which information can be transmitted by a wave (light, radio, or whatever) is proportional to the frequency. To see qualitatively why this is so, consider modulating (modifying) the wave by chopping off some of the wave crests. Suppose each crest represents a binary digit, with a chopped-off crest representing a zero and an unmodified crest representing a one. The number of binary digits we can transmit per unit time is thus proportional to the frequency of the wave. Infrared and visible-light waves have much higher frequency than do radio waves, so a modulated laser beam can transmit an enormous amount of information through a single fiber-optic cable.

Another advantage of optical fibers is that they can be made thinner than conventional copper wire, so more fibers can be bundled together in a cable of a given diameter. Hence more distinct signals (for instance, different phone lines) can be sent over the same cable. Because fiber-optic cables are electrical insulators, they are immune to electrical interference from lightning and other sources, and they don't allow unwanted currents between source and receiver. For these and other reasons, fiber-optic cables play an important role in long-distance telephone, television, and Internet communication.

Total internal reflection also plays an important role in the design of jewelry. The brilliance of diamond is due in large measure to its very high index of refraction $(n = 2.417)$ and correspondingly small critical angle. Light entering a cut diamond is totally internally reflected from facets on its back surface and then emerges from its front surface (see the photograph that opens this chapter). "Imitation diamond" gems, such as cubic zirconia, are made from less expensive crystalline materials with comparable indexes of refraction.

CONCEPTUAL EXAMPLE 33.4 A LEAKY PERISCOPE

SOLUTION

A submarine periscope uses two totally reflecting 45°–45°–90° prisms with total internal reflection on the sides adjacent to the 45° angles. Explain why the periscope will no longer work if it springs a leak and the bottom prism is covered with water.

SOLUTION

The critical angle for water $(n_b = 1.33)$ on glass $(n_a = 1.52)$ is

$$\theta_{\text{crit}} = \arcsin\frac{1.33}{1.52} = 61.0°$$

The 45° angle of incidence for a totally reflecting prism is *smaller* than this new 61° critical angle, so total internal reflection does not occur at the glass–water interface. Most of the light is transmitted into the water, and very little is reflected back into the prism.

TEST YOUR UNDERSTANDING OF SECTION 33.3 In which of the following situations is there total internal reflection? (i) Light propagating in water $(n = 1.33)$ strikes a water–air interface at an incident angle of 70°; (ii) light propagating in glass $(n = 1.52)$ strikes a glass–water interface at an incident angle of 70°; (iii) light propagating in water strikes a water–glass interface at an incident angle of 70°.

33.4 DISPERSION

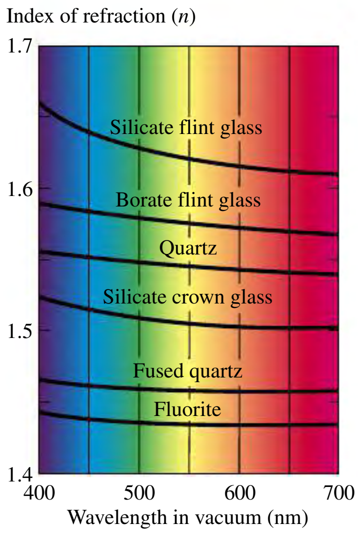
Ordinary white light is a superposition of waves with all visible wavelengths. The speed of light *in vacuum* is the same for all wavelengths, but the speed in a material substance is different for different wavelengths. Therefore the index of refraction of a material depends on wavelength. The dependence of wave speed and index of refraction on wavelength is called **dispersion.**

Figure 33.17 shows the variation of index of refraction n with wavelength for some common optical materials. Note that the horizontal axis of this figure is the wavelength of the light *in vacuum,* λ_0; the wavelength in the material is given by Eq. (33.5), $\lambda = \lambda_0/n$. In most materials the value of n *decreases* with increasing wavelength and decreasing frequency, and thus n *increases* with decreasing wavelength and increasing frequency. In such a material, light of longer wavelength has greater speed than light of shorter wavelength.

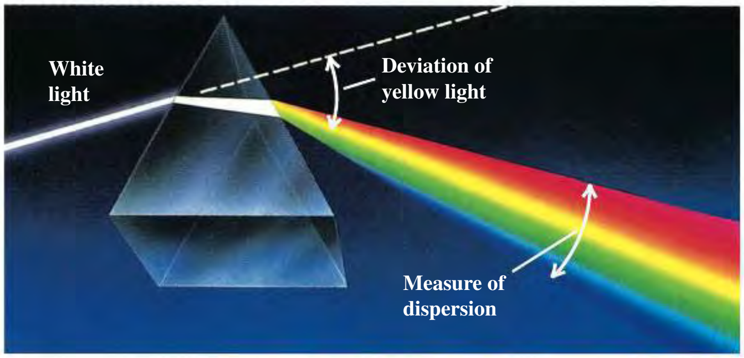
Figure 33.18 shows a ray of white light incident on a prism. The deviation (change of direction) produced by the prism increases with increasing index of refraction and frequency and decreasing wavelength. So violet light is deviated most, and red is deviated least. When it comes out of the prism, the light is spread out into a fan-shaped beam, as shown. The light is said to be *dispersed* into a spectrum. The amount of dispersion depends on the *difference* between the refractive indexes for violet light and for red light. From Fig. 33.17 we can see that for fluorite, the difference between the indexes for red and violet is small, and the dispersion will also be small. A better choice of material for a prism whose purpose is to produce a spectrum would be silicate flint glass, for which there is a larger difference in the value of n between red and violet.

As we mentioned in Section 33.3, the brilliance of diamond is due in part to its unusually large refractive index; another important factor is its large dispersion, which causes white light entering a diamond to emerge as a multicolored spectrum. Crystals of rutile and of strontium titanate, which can be produced synthetically, have about eight times the dispersion of diamond.

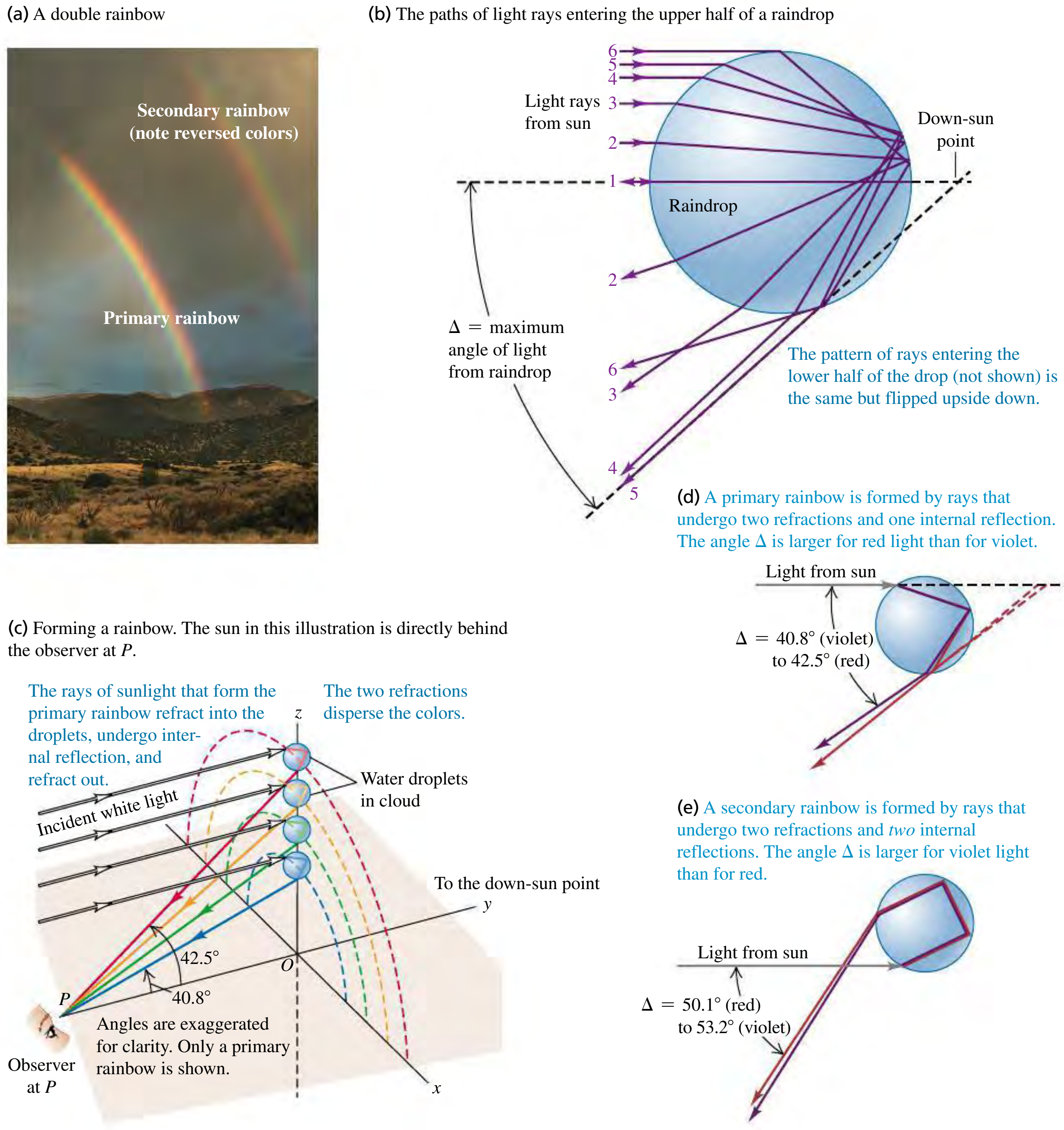
33.17 Variation of index of refraction n with wavelength for different transparent materials. The horizontal axis shows the wavelength λ_0 of the light *in vacuum;* the wavelength in the material is equal to $\lambda = \lambda_0/n$.



33.18 Dispersion of light by a prism. The band of colors is called a spectrum.



33.19 How rainbows form.



Rainbows

When you experience the beauty of a rainbow, as in **Fig. 33.19a**, you are seeing the combined effects of dispersion, refraction, and reflection. Sunlight comes from behind you, enters a water droplet, is (partially) reflected from the back surface of the droplet, and is refracted again upon exiting the droplet (Fig. 33.19b). A light ray that enters the middle of the raindrop is reflected straight back. All other rays exit the raindrop within an angle Δ of that middle ray, with many rays "piling up" at the angle Δ. What you see is a disk of light of angular radius Δ centered on the down-sun point (the point in the sky opposite the sun); due to the "piling up" of light rays, the disk is brightest around its rim, which we see as a rainbow (Fig. 33.19c). Because no light reaches your eye from angles larger than Δ, the sky looks dark outside the rainbow (see Fig. 33.19a). The value of the angle Δ depends on the index of refraction of the water that makes up the raindrops, which in turn depends on the wavelength (Fig. 33.19d). The bright disk of red light is slightly larger than that for orange light, which in turn is slightly larger than that for yellow light, and so on. As a result, you see the rainbow as a band of colors.

In many cases you can see a second, larger rainbow. It is the result of dispersion, refraction, and *two* reflections from the back surface of the droplet (Fig. 33.19e). Each time a light ray hits the back surface, part of the light is refracted out of the drop (not shown in Fig. 33.19); after two such hits, relatively little light is left inside the drop, which is why the secondary rainbow is noticeably fainter than the primary rainbow. Just as a mirror held up to a book reverses the printed letters, so the second reflection reverses the sequence of colors in the secondary rainbow. You can see this effect in Fig. 33.19a.

33.5 POLARIZATION

Polarization is a characteristic of all transverse waves. This chapter is about light, but to introduce some basic polarization concepts, let's go back to the transverse waves on a string that we studied in Chapter 15. For a string that in equilibrium lies along the x-axis, the displacements may be along the y-direction, as in **Fig. 33.20a**. In this case the string always lies in the xy-plane. But the displacements might instead be along the z-axis, as in Fig. 33.20b; then the string always lies in the xz-plane.

When a wave has only y-displacements, we say that it is **linearly polarized** in the y-direction; a wave with only z-displacements is linearly polarized in the z-direction. For mechanical waves we can build a **polarizing filter,** or **polarizer,** that permits only waves with a certain polarization direction to pass. In Fig. 33.20c the string can slide vertically in the slot without friction, but no horizontal motion is possible. This filter passes waves that are polarized in the y-direction but blocks those that are polarized in the z-direction.

This same language can be applied to electromagnetic waves, which also have polarization. As we learned in Chapter 32, an electromagnetic wave is a transverse wave; the fluctuating electric and magnetic fields are perpendicular to each other and to the direction of propagation. We always define the direction of polarization of an electromagnetic wave to be the direction of the *electric*-field vector $\vec{E}$, not the magnetic field, because many common electromagnetic-wave detectors respond to the electric forces on electrons in materials, not the magnetic forces. Thus the electromagnetic wave described by Eq. (32.17),

$$\vec{E}(x, t) = \hat{\jmath}E_{\max}\cos(kx - \omega t)$$

$$\vec{B}(x, t) = \hat{k}B_{\max}\cos(kx - \omega t)$$

is said to be polarized in the y-direction because the electric field has only a y-component.

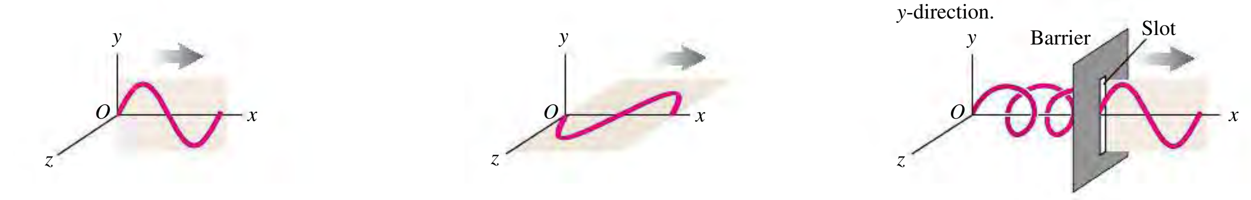
CAUTION **The meaning of "polarization"** It's unfortunate that the same word "polarization" that is used to describe the direction of $\vec{E}$ in an electromagnetic wave is also used to describe the shifting of electric charge within a body, such as in response to a nearby charged body; we described this latter kind of polarization in Section 21.2 (see Fig. 21.7). Don't confuse these two concepts!

33.20 (a), (b) Polarized waves on a string. (c) Making a polarized wave on a string from an unpolarized one using a polarizing filter.

(a) Transverse wave linearly polarized in the y-direction

(b) Transverse wave linearly polarized in the z-direction

(c) The slot functions as a polarizing filter, passing only components polarized in the y-direction.



33.21 (a) Electrons in the red and white broadcast antenna oscillate vertically, producing vertically polarized electromagnetic waves that propagate away from the antenna in the horizontal direction. (The small gray antennas are for relaying cellular phone signals.) (b) No matter how this light bulb is oriented, the random motion of electrons in the filament produces unpolarized light waves.

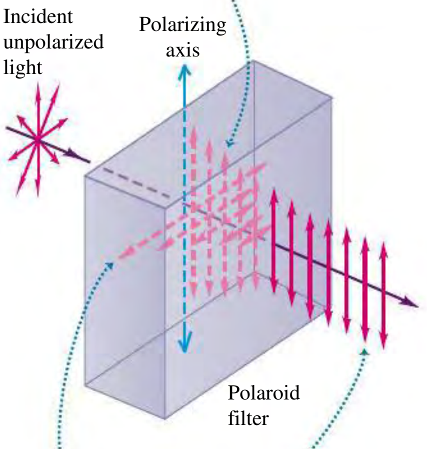
Polarizing Filters

Waves emitted by a radio transmitter are usually linearly polarized. The vertical antennas that are used for radio broadcasting emit waves that, in a horizontal plane around the antenna, are polarized in the vertical direction (parallel to the antenna) (**Fig. 33.21a**).

The situation is different for visible light. Light from incandescent light bulbs and fluorescent light fixtures is *not* polarized (Fig. 33.21b). The "antennas" that radiate light waves are the molecules that make up the sources. The waves emitted by any one molecule may be linearly polarized, like those from a radio antenna. But any actual light source contains a tremendous number of molecules with random orientations, so the emitted light is a random mixture of waves linearly polarized in all possible transverse directions. Such light is called **unpolarized light** or **natural light.** To create polarized light from unpolarized natural light requires a filter that is analogous to the slot for mechanical waves in Fig. 33.20c.

Polarizing filters for electromagnetic waves have different details of construction, depending on the wavelength. For microwaves with a wavelength of a few centimeters, a good polarizer is an array of closely spaced, parallel conducting wires that are insulated from each other. (Think of a barbecue grill with the outer metal ring replaced by an insulating one.) Electrons are free to move along the length of the conducting wires and will do so in response to a wave whose $\vec{E}$ field is parallel to the wires. The resulting currents in the wires dissipate energy by I^2R heating; the dissipated energy comes from the wave, so whatever wave passes through the grid is greatly reduced in amplitude. Waves with $\vec{E}$ oriented perpendicular to the wires pass through almost unaffected, since electrons cannot move through the air between the wires. Hence a wave that passes through such a filter will be predominantly polarized in the direction perpendicular to the wires.

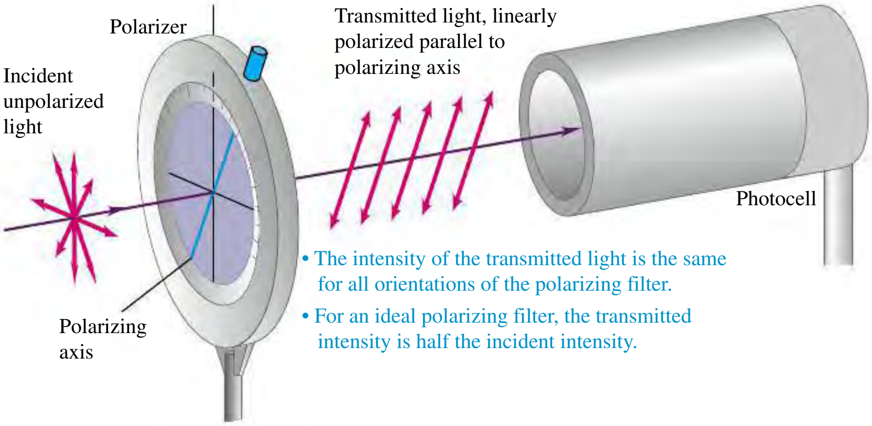
33.22 A Polaroid filter is illuminated by unpolarized natural light (shown by $\vec{E}$ vectors that point in all directions perpendicular to the direction of propagation). The transmitted light is linearly polarized along the polarizing axis (shown by $\vec{E}$ vectors along the polarization direction only).



The most common polarizing filter for visible light is a material known by the trade name Polaroid, widely used for sunglasses and polarizing filters for camera lenses. This material incorporates substances that have **dichroism,** a selective absorption in which one of the polarized components is absorbed much more strongly than the other (**Fig. 33.22**). A Polaroid filter transmits 80% or more of the intensity of a wave that is polarized parallel to the **polarizing axis** of the material, but only 1% or less for waves that are polarized perpendicular to this axis. In one type of Polaroid filter, long-chain molecules within the filter are oriented with their axis perpendicular to the polarizing axis; these molecules preferentially absorb light that is polarized along their length, much like the conducting wires in a polarizing filter for microwaves.

Using Polarizing Filters

An *ideal* polarizing filter (polarizer) passes 100% of the incident light that is polarized parallel to the filter's polarizing axis but completely blocks all light



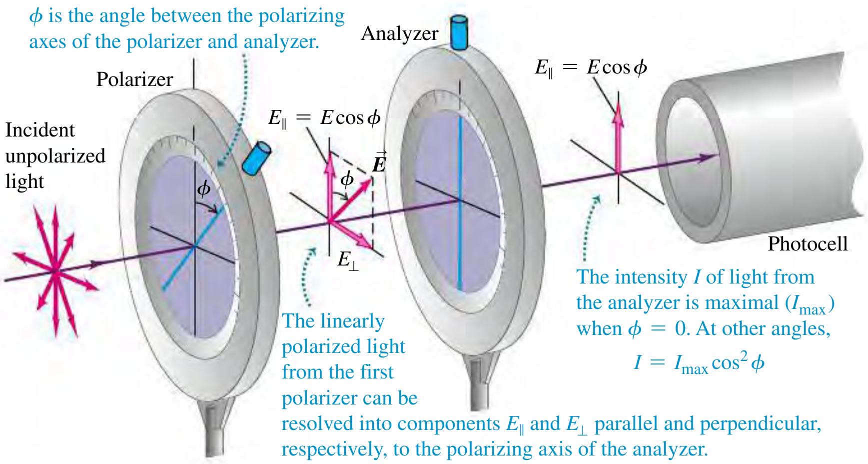
33.23 Unpolarized natural light is incident on the polarizing filter. The photocell measures the intensity of the transmitted linearly polarized light.

that is polarized perpendicular to this axis. Such a device is an unattainable idealization, but the concept is useful in clarifying the basic ideas. In the following discussion we will assume that all polarizing filters are ideal. In **Fig. 33.23** unpolarized light is incident on a flat polarizing filter. The $\vec{E}$ vector of the incident wave can be represented in terms of components parallel and perpendicular to the polarizer axis (shown in blue); only the component of $\vec{E}$ parallel to the polarizing axis is transmitted. Hence the light emerging from the polarizer is linearly polarized parallel to the polarizing axis.

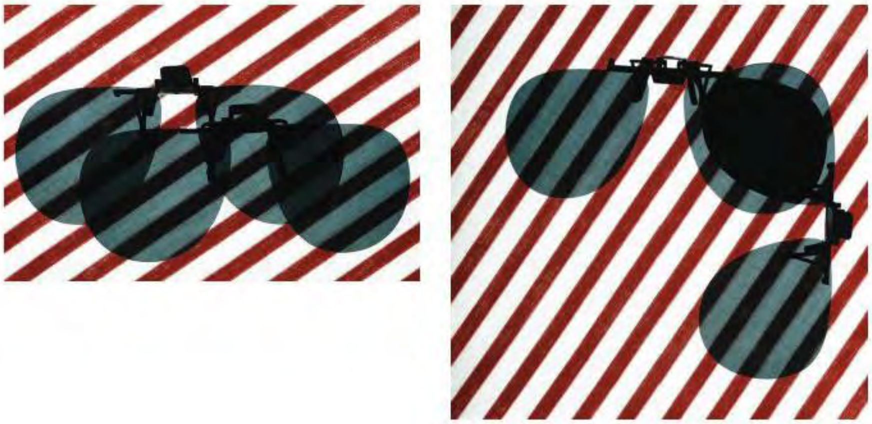
When unpolarized light is incident on an ideal polarizer as in Fig. 33.23, the intensity of the transmitted light is *exactly half* that of the incident unpolarized light, no matter how the polarizing axis is oriented. Here's why: We can resolve the $\vec{E}$ field of the incident wave into a component parallel to the polarizing axis and a component perpendicular to it. Because the incident light is a random mixture of all states of polarization, these two components are, on average, equal. The ideal polarizer transmits only the component that is parallel to the polarizing axis, so half the incident intensity is transmitted.

What happens when the linearly polarized light emerging from a polarizer passes through a second polarizer, or *analyzer,* as in **Fig. 33.24**? Suppose the polarizing axis of the analyzer makes an angle ϕ with the polarizing axis of the first polarizer. We can resolve the linearly polarized light that is transmitted by the first polarizer into two components, as shown in Fig. 33.24, one parallel and the other perpendicular to the axis of the analyzer. Only the parallel component, with amplitude $E\cos\phi$, is transmitted by the analyzer. The transmitted intensity is greatest when $\phi = 0$, and it is zero when the polarizer and analyzer are *crossed*

DEMO



33.24 An ideal analyzer transmits only the electric field component parallel to its transmission direction (that is, its polarizing axis).



33.25 These photos show the view through Polaroid sunglasses whose polarizing axes are (left) aligned ($\phi = 0$) and (right) perpendicular ($\phi = 90°$). The transmitted intensity is greatest when the axes are aligned; it is zero when the axes are perpendicular.

so that $\phi = 90°$ (**Fig. 33.25**). To determine the direction of polarization of the light transmitted by the first polarizer, rotate the analyzer until the photocell in Fig. 33.24 measures zero intensity; the polarization axis of the first polarizer is then perpendicular to that of the analyzer.

To find the transmitted intensity at intermediate values of the angle ϕ, we recall from Section 32.4 that the intensity of an electromagnetic wave is proportional to the *square* of the amplitude of the wave [see Eq. (32.29)]. The ratio of transmitted to incident *amplitude* is $\cos\phi$, so the ratio of transmitted to incident *intensity* is $\cos^2\phi$. Thus the intensity transmitted is

Intensity of polarized light passed through an analyzer

Malus's law: $$I = I_{max}\cos^2\phi \qquad (33.7)$$

Maximum transmitted intensity

Angle between polarization axis of light and polarizing axis of analyzer

This relationship, discovered experimentally by Étienne-Louis Malus in 1809, is called **Malus's law.** Malus's law applies *only* if the incident light passing through the analyzer is already linearly polarized.

PROBLEM-SOLVING STRATEGY 33.2 LINEAR POLARIZATION

IDENTIFY *the relevant concepts:* In all electromagnetic waves, including light waves, the direction of polarization is the direction of the $\vec{E}$ field and is perpendicular to the propagation direction. Problems about polarizers are therefore about the components of $\vec{E}$ parallel and perpendicular to the polarizing axis.

SET UP *the problem* using the following steps:

1. Start by drawing a large, neat diagram. Label all known angles, including the angles of all polarizing axes.
2. Identify the target variables.

EXECUTE *the solution* as follows:

1. Remember that a polarizer lets pass only electric-field components parallel to its polarizing axis.
2. If the incident light is linearly polarized and has amplitude E and intensity I_{max}, the light that passes through an ideal polarizer has amplitude $E\cos\phi$ and intensity $I_{max}\cos^2\phi$, where ϕ is the angle between the incident polarization direction and the filter's polarizing axis.
3. Unpolarized light is a random mixture of all possible polarization states, so on the average it has equal components in any two perpendicular directions. When passed through an ideal polarizer, unpolarized light becomes linearly polarized light with half the incident intensity. Partially linearly polarized light is a superposition of linearly polarized and unpolarized light.
4. The intensity (average power per unit area) of a wave is proportional to the *square* of its amplitude. If you find that two waves differ in amplitude by a certain factor, their intensities differ by the square of that factor.

EVALUATE *your answer:* Check your answer for any obvious errors. If your results say that light emerging from a polarizer has greater intensity than the incident light, something's wrong: A polarizer can't add energy to a light wave.

EXAMPLE 33.5 TWO POLARIZERS IN COMBINATION

In Fig. 33.24 the incident unpolarized light has intensity I_0. Find the intensities transmitted by the first and second polarizers if the angle between the axes of the two filters is 30°.

SOLUTION

IDENTIFY and SET UP: This problem involves a polarizer (a polarizing filter on which unpolarized light shines, producing polarized light) and an analyzer (a second polarizing filter on which the polarized light shines). We are given the intensity I_0 of the incident light and the angle $\phi = 30°$ between the axes of the polarizers. We use Malus's law, Eq. (33.7), to solve for the intensities of the light emerging from each polarizer.

EXECUTE: The incident light is unpolarized, so the intensity of the linearly polarized light transmitted by the first polarizer is $I_0/2$. From Eq. (33.7) with $\phi = 30°$, the second polarizer reduces the intensity by a further factor of $\cos^2 30° = \frac{3}{4}$. Thus the intensity transmitted by the second polarizer is

$$\left(\frac{I_0}{2}\right)\left(\frac{3}{4}\right) = \frac{3}{8}I_0$$

EVALUATE: Note that the intensity decreases after each passage through a polarizer. The only situation in which the transmitted intensity does *not* decrease is if the polarizer is ideal (so it absorbs none of the light that passes through it) and if the incident light is linearly polarized along the polarizing axis, so $\phi = 0$.

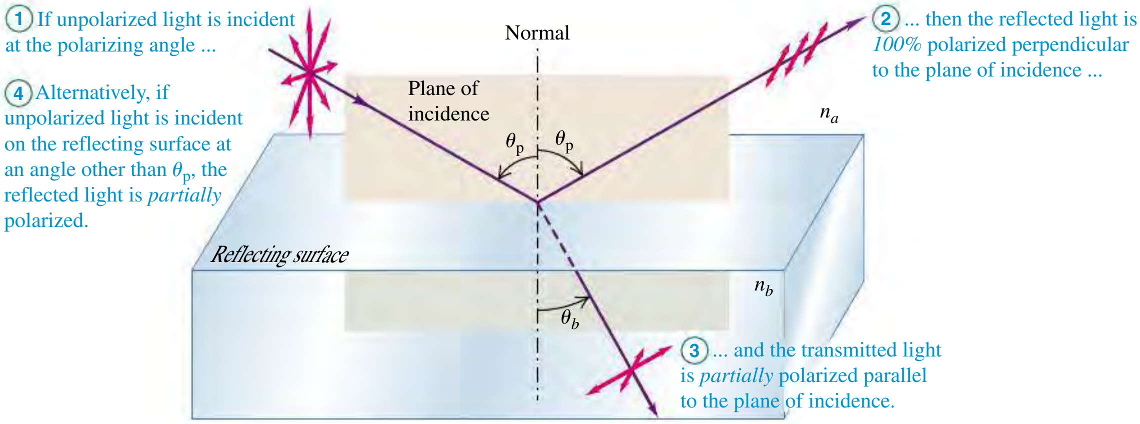
Polarization by Reflection

Unpolarized light can be polarized, either partially or totally, by *reflection*. In **Fig. 33.26**, unpolarized natural light is incident on a reflecting surface between two transparent optical materials. For most angles of incidence, waves for which the electric-field vector $\vec{E}$ is perpendicular to the plane of incidence (that is, parallel to the reflecting surface) are reflected more strongly than those for which $\vec{E}$ lies in this plane. In this case the reflected light is *partially polarized* in the direction perpendicular to the plane of incidence.

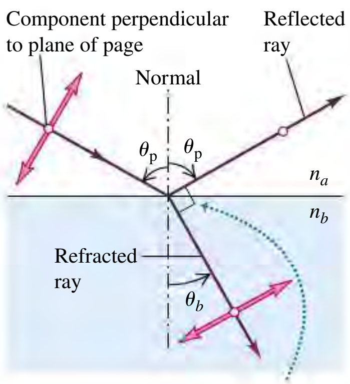
But at one particular angle of incidence, called the **polarizing angle** θ_p, the light for which $\vec{E}$ lies in the plane of incidence is *not reflected at all* but is completely refracted. At this same angle of incidence the light for which $\vec{E}$ is perpendicular to the plane of incidence is partially reflected and partially refracted. The *reflected* light is therefore *completely* polarized perpendicular to the plane of incidence, as shown in Fig. 33.26. The *refracted* (transmitted) light is *partially* polarized parallel to this plane; the refracted light is a mixture of the component parallel to the plane of incidence, all of which is refracted, and the remainder of the perpendicular component.

In 1812 the British scientist Sir David Brewster discovered that when the angle of incidence is equal to the polarizing angle θ_p, the reflected ray and the refracted

33.26 When light is incident on a reflecting surface at the polarizing angle, the reflected light is linearly polarized.



33.27 The significance of the polarizing angle. The open circles represent a component of $\vec{E}$ that is perpendicular to the plane of the figure (the plane of incidence) and parallel to the surface between the two materials.



ray are perpendicular to each other (**Fig. 33.27**). In this case the angle of refraction θ_b equals $90° - \theta_p$. From the law of refraction,

$$n_a \sin\theta_p = n_b \sin\theta_b = n_b \sin(90° - \theta_p) = n_b \cos\theta_p$$

Since $(\sin\theta_p)/(\cos\theta_p) = \tan\theta_p$, we can rewrite this equation as

Brewster's law for the polarizing angle:

Polarizing angle (angle of incidence for which reflected light is 100% polarized)

$$\tan\theta_p = \frac{n_b}{n_a} \tag{33.8}$$

n_b: Index of refraction of second material
n_a: Index of refraction of first material

This relationship is known as **Brewster's law.** Although discovered experimentally, it can also be *derived* from a wave model by using Maxwell's equations.

Polarization by reflection is the reason polarizing filters are widely used in sunglasses (Fig. 33.25). When sunlight is reflected from a horizontal surface, the plane of incidence is vertical, and the reflected light contains a preponderance of light that is polarized in the horizontal direction. When the reflection occurs at a smooth asphalt road surface, it causes unwanted glare. To eliminate this glare, the polarizing axis of the lens material is made vertical, so very little of the horizontally polarized light reflected from the road is transmitted to the eyes. The glasses also reduce the overall intensity of the transmitted light to somewhat less than 50% of the intensity of the unpolarized incident light.

SOLUTION

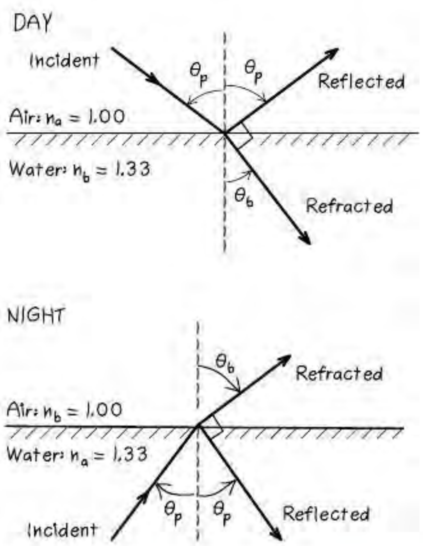
EXAMPLE 33.6 REFLECTION FROM A SWIMMING POOL'S SURFACE

Sunlight reflects off the smooth surface of a swimming pool. (a) For what angle of reflection is the reflected light completely polarized? (b) What is the corresponding angle of refraction? (c) At night, an underwater floodlight is turned on in the pool. Repeat parts (a) and (b) for rays from the floodlight that strike the surface from below.

SOLUTION

IDENTIFY and SET UP: This problem involves polarization by reflection at an air–water interface in parts (a) and (b) and at a water–air interface in part (c). **Figure 33.28** shows our sketches.

33.28 Our sketches for this problem.



For both cases our first target variable is the polarizing angle θ_p, which we find from Brewster's law, Eq. (33.8). For this angle of reflection, the angle of refraction θ_b is the complement of θ_p (that is, $\theta_b = 90° - \theta_p$).

EXECUTE: (a) During the day (shown in the upper part of Fig. 33.28) the light moves in air toward water, so $n_a = 1.00$ (air) and $n_b = 1.33$ (water). From Eq. (33.8),

$$\theta_p = \arctan\frac{n_b}{n_a} = \arctan\frac{1.33}{1.00} = 53.1°$$

(b) The incident light is at the polarizing angle, so the reflected and refracted rays are perpendicular; hence

$$\theta_b = 90° - \theta_p = 90° - 53.1° = 36.9°$$

(c) At night (shown in the lower part of Fig. 33.28) the light moves in water toward air, so now $n_a = 1.33$ and $n_b = 1.00$. Again using Eq. (33.8), we have

$$\theta_p = \arctan\frac{1.00}{1.33} = 36.9°$$

$$\theta_b = 90° - 36.9° = 53.1°$$

EVALUATE: We check our answer in part (b) by using Snell's law, $n_a \sin\theta_a = n_b \sin\theta_b$, to solve for θ_b:

$$\sin\theta_b = \frac{n_a \sin\theta_p}{n_b} = \frac{1.00 \sin 53.1°}{1.33} = 0.600$$

$$\theta_b = \arcsin(0.600) = 36.9°$$

Note that the two polarizing angles found in parts (a) and (c) add to 90°. This is *not* an accident; can you see why?

Circular and Elliptical Polarization

Light and other electromagnetic radiation can also have *circular* or *elliptical* polarization. To introduce these concepts, let's return once more to mechanical waves on a stretched string. In Fig. 33.20, suppose the two linearly polarized waves in parts (a) and (b) are in phase and have equal amplitude. When they are superposed, each point in the string has simultaneous y- and z-displacements of equal magnitude. A little thought shows that the resultant wave lies in a plane oriented at 45° to the y- and z-axes (i.e., in a plane making a 45° angle with the xy- and xz-planes). The amplitude of the resultant wave is larger by a factor of $\sqrt{2}$ than that of either component wave, and the resultant wave is linearly polarized.

But now suppose the two equal-amplitude waves differ in phase by a quarter-cycle. Then the resultant motion of each point corresponds to a superposition of two simple harmonic motions at right angles, with a quarter-cycle phase difference. The y-displacement at a point is greatest at times when the z-displacement is zero, and vice versa. The string as a whole then no longer moves in a single plane. It can be shown that each point on the rope moves in a *circle* in a plane parallel to the yz-plane. Successive points on the rope have successive phase differences, and the overall motion of the string has the appearance of a rotating helix, as shown to the left of the polarizing filter in Fig. 33.20c. Such a superposition of two linearly polarized waves is called **circular polarization.**

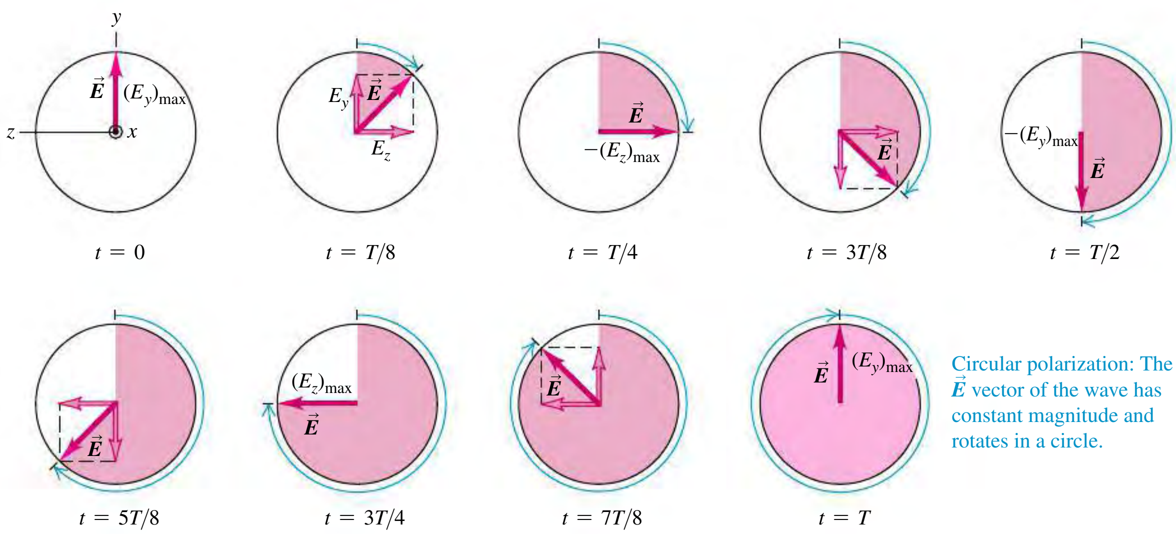
Figure 33.29 shows the analogous situation for an electromagnetic wave. Two sinusoidal waves of equal amplitude, polarized in the y- and z-directions and with a quarter-cycle phase difference, are superposed. The result is a wave in which the $\vec{E}$ vector at each point has a constant magnitude but *rotates* around the direction of propagation. The wave in Fig. 33.29 is propagating toward you and the $\vec{E}$ vector appears to be rotating clockwise, so it is called a *right circularly polarized* electromagnetic wave. If instead the $\vec{E}$ vector of a wave coming toward you appears to be rotating counterclockwise, it is called a *left circularly polarized* electromagnetic wave.

If the phase difference between the two component waves is something other than a quarter-cycle, or if the two component waves have different amplitudes, then each point on the string traces out not a circle but an *ellipse.* The resulting wave is said to be **elliptically polarized.**

For electromagnetic waves with radio frequencies, circular or elliptical polarization can be produced by using two antennas at right angles, fed from the same transmitter but with a phase-shifting network that introduces the appropriate phase difference. For light, the phase shift can be introduced by use of a material

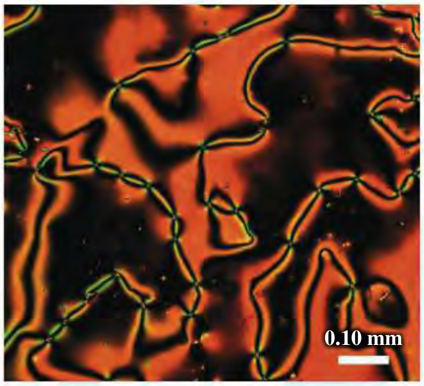
Application Circular Polarization and 3-D Movies The lenses of the special glasses you wear to see a 3-D movie are circular polarizing filters. The lens over one eye allows only right circularly polarized light to pass; the other lens allows only left circularly polarized light to pass. The projector alternately projects the images intended for the left eye and those intended for the right eye. A special filter synchronized with the projector and in front of its lens circularly polarizes the projected light, with alternate polarization for each frame. Hence alternate images go to your left and right eyes, with such a short time interval between them that they produce the illusion of viewing with both eyes simultaneously.

33.29 Circular polarization of an electromagnetic wave moving toward you parallel to the x-axis. The y-component of $\vec{E}$ lags the z-component by a quarter-cycle. This phase difference results in right circular polarization.



Application Birefringence and Liquid Crystal Displays In each pixel of an LCD computer screen is a birefringent material called a liquid crystal. This material is composed of rod-shaped molecules that align to produce a fluid with two different indexes of refraction. The liquid crystal is placed between linear polarizing filters with perpendicular polarizing axes, and the sandwich of filters and liquid crystal is backlighted. The two polarizers by themselves would not transmit light, but like the birefringent object in Fig. 33.30, the liquid crystal allows light to pass through. Varying the voltage across a pixel turns the birefringence effect on and off, changing the pixel from bright to dark and back again.

Microscope image of a liquid crystal



Liquid crystal display

that exhibits *birefringence*—that is, has different indexes of refraction for different directions of polarization. A common example is calcite $(CaCO_3)$. When a calcite crystal is oriented appropriately in a beam of unpolarized light, its refractive index (for a wavelength in vacuum of 589 nm) is 1.658 for one direction of polarization and 1.486 for the perpendicular direction. When two waves with equal amplitude and with perpendicular directions of polarization enter such a material, they travel with different speeds. If they are in phase when they enter the material, then in general they are no longer in phase when they emerge. If the crystal is just thick enough to introduce a quarter-cycle phase difference, then the crystal converts linearly polarized light to circularly polarized light. Such a crystal is called a *quarter-wave plate*. Such a plate also converts circularly polarized light to linearly polarized light. Can you prove this?

Photoelasticity

Some optical materials that are not normally birefringent become so when they are subjected to mechanical stress. This is the basis of the science of *photoelasticity*. Stresses in girders, boiler plates, gear teeth, and cathedral pillars can be analyzed by constructing a transparent model of the object, usually of a plastic material, subjecting it to stress, and examining it between a polarizer and an analyzer in the crossed position. Very complicated stress distributions can be studied by these optical methods.

Figure 33.30 is a photograph of a photoelastic model under stress. The polarized light that enters the model can be thought of as having a component along each of the two directions of the birefringent plastic. Since these two components travel through the plastic at different speeds, the light that emerges from the other side of the model can have a different overall direction of polarization. Hence some of this transmitted light will be able to pass through the analyzer even though its polarization axis is at a 90° angle to the polarizer's axis, and the stressed areas in the plastic will appear as bright spots. The amount of birefringence is different for different wavelengths and hence different colors of light; the color that appears at each location in Fig. 33.30 is that for which the transmitted light is most nearly polarized along the analyzer's polarization axis.

TEST YOUR UNDERSTANDING OF SECTION 33.5 You are taking a photograph of a sunlit office building at sunrise, so the plane of incidence is nearly horizontal. In order to minimize the reflections from the building's windows, you place a polarizing filter on the camera lens. How should you orient the filter? (i) With the polarizing axis vertical; (ii) with the polarizing axis horizontal; (iii) either orientation will minimize the reflections just as well; (iv) neither orientation will have any effect. ❙

33.30 This plastic model of an artificial hip joint was photographed between two polarizing filters (a polarizer and an analyzer) with perpendicular polarizing axes. The colored interference pattern reveals the direction and magnitude of stresses in the model. Engineers use these results to help design the actual hip joint (used in hip replacement surgery), which is made of metal.

33.6 SCATTERING OF LIGHT

The sky is blue. Sunsets are red. Skylight is partially polarized; that's why the sky looks darker from some angles than from others when it is viewed through Polaroid sunglasses. As we will see, a single phenomenon is responsible for all of these effects.

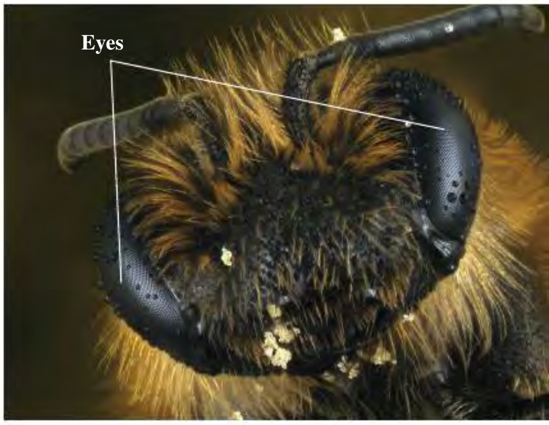
When you look at the daytime sky, the light that you see is sunlight that has been absorbed and then re-radiated in a variety of directions. This process is called **scattering.** (If the earth had no atmosphere, the sky would appear as black in the daytime as it does at night, just as it does to an astronaut in space or on the moon.) **Figure 33.31** shows some of the details of the scattering process. Sunlight, which is unpolarized, comes from the left along the x-axis and passes over an observer looking vertically upward along the y-axis. (We are viewing the situation from the side.) Consider the molecules of the earth's atmosphere located at point O. The electric field in the beam of sunlight sets the electric charges in these molecules into vibration. Since light is a transverse wave, the direction of the electric field in any component of the sunlight lies in the yz-plane, and the motion of the charges takes place in this plane. There is no field, and hence no motion of charges, in the direction of the x-axis.

An incident light wave sets the electric charges in the molecules at point O vibrating along the line of $\vec{E}$. We can resolve this vibration into two components, one along the y-axis and the other along the z-axis. Each component in the incident light produces the equivalent of two molecular "antennas," oscillating with the same frequency as the incident light and lying along the y- and z-axes.

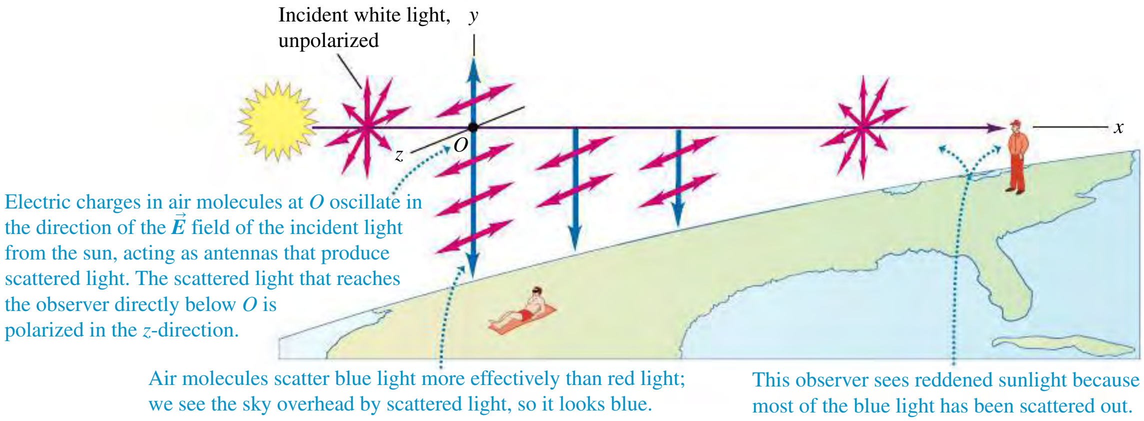
We mentioned in Chapter 32 that an oscillating charge, like those in an antenna, does not radiate in the direction of its oscillation. (See Fig. 32.3 in Section 32.1.) Thus the "antenna" along the y-axis does not send any light to the observer directly below it, although it does emit light in other directions. Therefore the only light reaching this observer comes from the other molecular "antenna," corresponding to the oscillation of charge along the z-axis. This light is linearly polarized, with its electric field along the z-axis (parallel to the "antenna"). The red vectors on the y-axis below point O in Fig. 33.31 show the direction of polarization of the light reaching the observer.

As the original beam of sunlight passes though the atmosphere, its intensity decreases as its energy goes into the scattered light. Detailed analysis of the scattering process shows that the intensity of the light scattered from air molecules increases in proportion to the fourth power of the frequency (inversely to the fourth power of the wavelength). Thus the intensity ratio for the two ends of the visible spectrum is $(750\ \text{nm}/380\ \text{nm})^4 = 15$. Roughly speaking, scattered light contains 15 times as much blue light as red, and that's why the sky is blue.

BIO Application Bee Vision and Polarized Light from the Sky The eyes of a bee can detect the polarization of light. Bees use this ability when they navigate between the hive and food sources. As Fig. 33.31 would suggest, a bee sees unpolarized light if it looks directly toward the sun and sees completely polarized light if it looks 90° away from the sun. These polarizations are unaffected by the presence of clouds, so a bee can navigate relative to the sun even on an overcast day.



33.31 When the sunbathing observer on the left looks up, he sees blue, polarized sunlight that has been scattered by air molecules. The observer on the right sees reddened, unpolarized light when he looks at the sun.



33.32 Clouds are white because they efficiently scatter sunlight of all wavelengths.

Clouds contain a high concentration of suspended water droplets or ice crystals, which also scatter light. Because of this high concentration, light passing through the cloud has many more opportunities for scattering than does light passing through a clear sky. Thus light of *all* wavelengths is eventually scattered out of the cloud, so the cloud looks white (**Fig. 33.32**). Milk looks white for the same reason; the scattering is due to fat globules suspended in the milk.

Near sunset, when sunlight has to travel a long distance through the earth's atmosphere, a substantial fraction of the blue light is removed by scattering. White light minus blue light looks yellow or red. This explains the yellow or red hue that we so often see from the setting sun (and that is seen by the observer at the far right of Fig. 33.31).

33.7 HUYGENS'S PRINCIPLE

The laws of reflection and refraction of light rays, as introduced in Section 33.2, were discovered experimentally long before the wave nature of light was firmly established. However, we can *derive* these laws from wave considerations and show that they are consistent with the wave nature of light.

We begin with a principle called **Huygens's principle.** This principle, stated originally by the Dutch scientist Christiaan Huygens in 1678, is a geometrical method for finding, from the known shape of a wave front at some instant, the shape of the wave front at some later time. Huygens assumed that **every point of a wave front may be considered the source of secondary wavelets that spread out in all directions with a speed equal to the speed of propagation of the wave.** The new wave front at a later time is then found by constructing a surface *tangent* to the secondary wavelets or, as it is called, the *envelope* of the wavelets. All the results that we obtain from Huygens's principle can also be obtained from Maxwell's equations, but Huygens's simple model is easier to use.

33.33 Applying Huygens's principle to wave front AA' to construct a new wave front BB'.

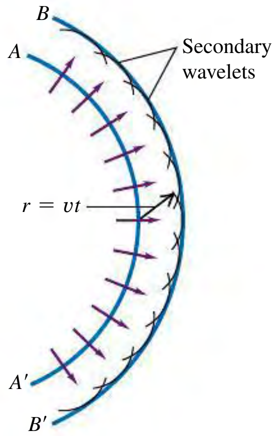
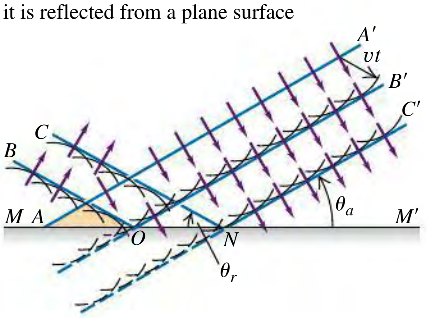


Figure 33.33 illustrates Huygens's principle. The original wave front AA' is traveling outward from a source, as indicated by the arrows. We want to find the shape of the wave front after a time interval t. We assume that v, the speed of propagation of the wave, is the same at all points. Then in time t the wave front travels a distance vt. We construct several circles (traces of spherical wavelets) with radius $r = vt$, centered at points along AA'. The trace of the envelope of these wavelets, which is the new wave front, is the curve BB'.

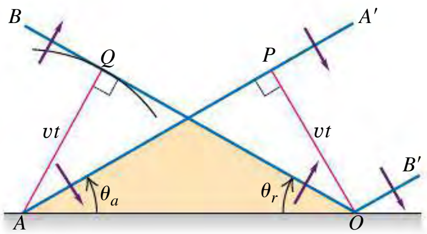
Reflection and Huygens's Principle

33.34 Using Huygens's principle to derive the law of reflection.

(a) Successive positions of a plane wave AA' as it is reflected from a plane surface



(b) Magnified portion of (a)



To derive the law of reflection from Huygens's principle, we consider a plane wave approaching a plane reflecting surface. In **Fig. 33.34a** the lines AA', OB', and NC' represent successive positions of a wave front approaching the surface MM'. Point A on the wave front AA' has just arrived at the reflecting surface. We can use Huygens's principle to find the position of the wave front after a time interval t. With points on AA' as centers, we draw several secondary wavelets with radius vt. The wavelets that originate near the upper end of AA' spread out unhindered, and their envelope gives the portion OB' of the new wave front. If the reflecting surface were not there, the wavelets originating near the lower end of AA' would similarly reach the positions shown by the broken circular arcs. Instead, these wavelets strike the reflecting surface.

The effect of the reflecting surface is to *change the direction* of travel of those wavelets that strike it, so the part of a wavelet that would have penetrated the surface actually lies to the left of it, as shown by the full lines. The first such wavelet is centered at point A; the envelope of all such reflected wavelets is the portion OB of the wave front. The trace of the entire wave front at this instant is the bent line BOB'. A similar construction gives the line CNC' for the wave front after another interval t.

From plane geometry the angle θ_a between the incident *wave front* and the *surface* is the same as that between the incident *ray* and the *normal* to the surface and is therefore the angle of incidence. Similarly, θ_r is the angle of reflection. To find the relationship between these angles, we consider Fig. 33.34b. From O we draw $OP = vt$, perpendicular to AA'. Now OB, by construction, is tangent to a circle of radius vt with center at A. If we draw AQ from A to the point of tangency, the triangles APO and OQA are congruent because they are right triangles with the side AO in common and with $AQ = OP = vt$. The angle θ_a therefore equals the angle θ_r, and we have the law of reflection.

Refraction and Huygens's Principle

We can derive the law of *refraction* by a similar procedure. In **Fig. 33.35a** we consider a wave front, represented by line AA', for which point A has just arrived at the boundary surface SS' between two transparent materials a and b, with indexes of refraction n_a and n_b and wave speeds v_a and v_b. (The *reflected* waves are not shown; they proceed as in Fig. 33.34.) We can apply Huygens's principle to find the position of the refracted wave fronts after a time t.

With points on AA' as centers, we draw several secondary wavelets. Those originating near the upper end of AA' travel with speed v_a and, after a time interval t, are spherical surfaces of radius $v_a t$. The wavelet originating at point A, however, is traveling in the second material b with speed v_b and at time t is a spherical surface of radius $v_b t$. The envelope of the wavelets from the original wave front is the plane whose trace is the bent line BOB'. A similar construction leads to the trace CPC' after a second interval t.

The angles θ_a and θ_b between the surface and the incident and refracted wave fronts are the angle of incidence and the angle of refraction, respectively. To find the relationship between these angles, refer to Fig. 33.35b. We draw $OQ = v_a t$, perpendicular to AQ, and we draw $AB = v_b t$, perpendicular to BO. From the right triangle AOQ,

$$\sin\theta_a = \frac{v_a t}{AO}$$

and from the right triangle AOB,

$$\sin\theta_b = \frac{v_b t}{AO}$$

Combining these, we find

$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{v_a}{v_b} \qquad (33.9)$$

We have defined the index of refraction n of a material as the ratio of the speed of light c in vacuum to its speed v in the material: $n_a = c/v_a$ and $n_b = c/v_b$. Thus

$$\frac{n_b}{n_a} = \frac{c/v_b}{c/v_a} = \frac{v_a}{v_b}$$

and we can rewrite Eq. (33.9) as

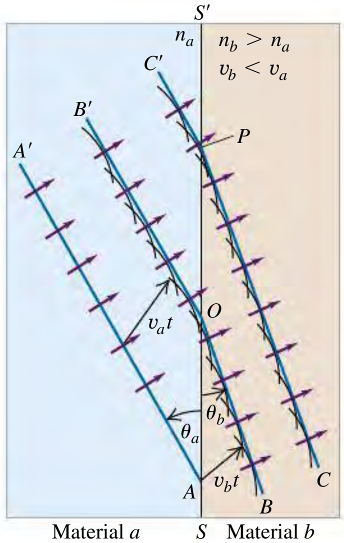
$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{n_b}{n_a} \quad \text{or}$$

$$n_a \sin\theta_a = n_b \sin\theta_b$$

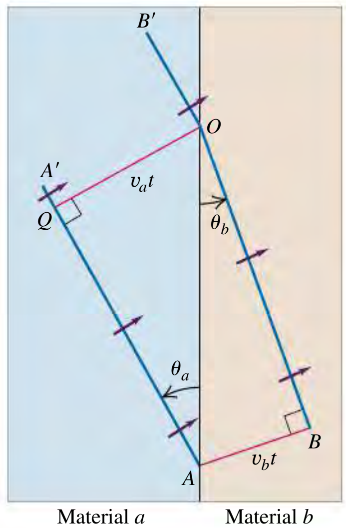
which we recognize as Snell's law, Eq. (33.4). So we have derived Snell's law from a wave theory. Alternatively, we can regard Snell's law as an experimental result that defines the index of refraction of a material; in that case this analysis helps confirm the relationship $v = c/n$ for the speed of light in a material.

33.35 Using Huygens's principle to derive the law of refraction. The case $v_b < v_a$ ($n_b > n_a$) is shown.

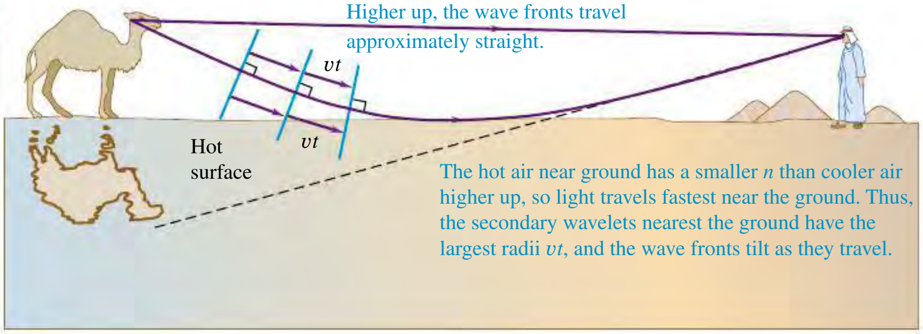
(a) Successive positions of a plane wave AA' as it is refracted by a plane surface



(b) Magnified portion of **(a)**



33.36 How mirages are formed.



Mirages are an example of Huygens's principle in action. When the surface of pavement or desert sand is heated intensely by the sun, a hot, less dense, smaller-n layer of air forms near the surface. The speed of light is slightly greater in the hotter air near the ground, the Huygens wavelets have slightly larger radii, the wave fronts tilt slightly, and rays that were headed toward the surface with a large angle of incidence (near 90°) can be bent up as shown in **Fig. 33.36**. Light farther from the ground is bent less and travels nearly in a straight line. The observer sees the object in its natural position, with an inverted image below it, as though seen in a horizontal reflecting surface. A thirsty traveler can interpret the apparent reflecting surface as a sheet of water.

It is important to keep in mind that Maxwell's equations are the fundamental relationships for electromagnetic wave propagation. But Huygens's principle provides a convenient way to visualize this propagation.

TEST YOUR UNDERSTANDING OF SECTION 33.7 Sound travels faster in warm air than in cold air. Imagine a weather front that runs north-south, with warm air to the west of the front and cold air to the east. A sound wave traveling in a northeast direction in the warm air encounters this front. How will the direction of this sound wave change when it passes into the cold air? (i) The wave direction will deflect toward the north; (ii) the wave direction will deflect toward the east; (iii) the wave direction will be unchanged. ▌

CHAPTER 33 SUMMARY

SOLUTIONS TO ALL EXAMPLES

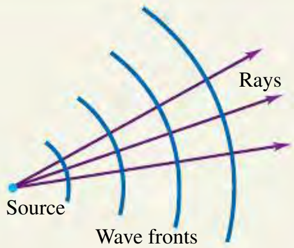
Light and its properties: Light is an electromagnetic wave. When emitted or absorbed, it also shows particle properties. It is emitted by accelerated electric charges.

A wave front is a surface of constant phase; wave fronts move with a speed equal to the propagation speed of the wave. A ray is a line along the direction of propagation, perpendicular to the wave fronts.

When light is transmitted from one material to another, the frequency of the light is unchanged, but the wavelength and wave speed can change. The index of refraction n of a material is the ratio of the speed of light in vacuum c to the speed v in the material. If λ_0 is the wavelength in vacuum, the same wave has a shorter wavelength λ in a medium with index of refraction n. (See Example 33.2.)

$$n = \frac{c}{v} \quad (33.1)$$

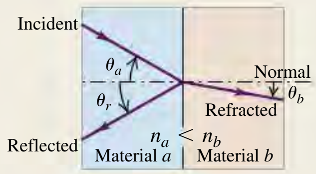
$$\lambda = \frac{\lambda_0}{n} \quad (33.5)$$



Reflection and refraction: At a smooth interface between two optical materials, the incident, reflected, and refracted rays and the normal to the interface all lie in a single plane called the plane of incidence. The law of reflection states that the angles of incidence and reflection are equal. The law of refraction relates the angles of incidence and refraction to the indexes of refraction of the materials. (See Examples 33.1 and 33.3.)

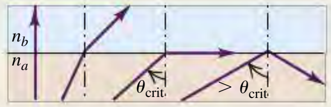
$$\theta_r = \theta_a \quad (33.2)$$
(law of reflection)

$$n_a \sin\theta_a = n_b \sin\theta_b \quad (33.4)$$
(law of refraction)



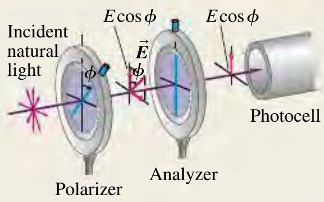
Total internal reflection: When a ray travels in a material of index of refraction n_a toward a material of index $n_b < n_a$, total internal reflection occurs at the interface when the angle of incidence equals or exceeds a critical angle θ_{crit}. (See Example 33.4.)

$$\sin\theta_{crit} = \frac{n_b}{n_a} \quad (33.6)$$



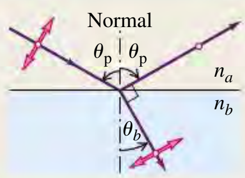
Polarization of light: The direction of polarization of a linearly polarized electromagnetic wave is the direction of the $\vec{E}$ field. A polarizing filter passes waves that are linearly polarized along its polarizing axis and blocks waves polarized perpendicularly to that axis. When polarized light of intensity I_{max} is incident on a polarizing filter used as an analyzer, the intensity I of the light transmitted through the analyzer depends on the angle ϕ between the polarization direction of the incident light and the polarizing axis of the analyzer. (See Example 33.5.)

$$I = I_{max}\cos^2\phi \quad (33.7)$$
(Malus's law)

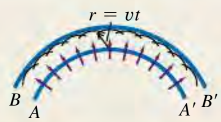


Polarization by reflection: When unpolarized light strikes an interface between two materials, Brewster's law states that the reflected light is completely polarized perpendicular to the plane of incidence (parallel to the interface) if the angle of incidence equals the polarizing angle θ_p. (See Example 33.6.)

$$\tan\theta_p = \frac{n_b}{n_a} \quad (33.8)$$
(Brewster's law)



Huygens's principle: Huygens's principle states that if the position of a wave front at one instant is known, then the position of the front at a later time can be constructed by imagining the front as a source of secondary wavelets. Huygens's principle can be used to derive the laws of reflection and refraction.

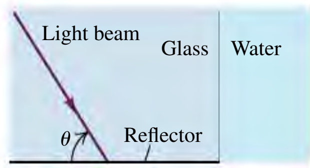


BRIDGING PROBLEM REFLECTION AND REFRACTION

SOLUTION

Figure 33.37 shows a rectangular glass block that has a metal reflector on one face and water on an adjoining face. A light beam strikes the reflector as shown. You gradually increase the angle θ of the light beam. If $\theta \geq 59.2°$, no light enters the water. What is the speed of light in this glass?

33.37 Glass bounded by water and a metal reflector.



SOLUTION GUIDE

IDENTIFY and SET UP

1. Specular reflection occurs where the light ray in the glass strikes the reflector. If no light is to enter the water, we require that there be reflection only and no refraction where this ray strikes the glass–water interface—that is, there must be total internal reflection.
2. The target variable is the speed of light v in the glass, which you can determine from the index of refraction n of the glass. (Table 33.1 gives the index of refraction of water.) Write down the equations you will use to find n and v.

EXECUTE

3. Use the figure to find the angle of incidence of the ray at the glass–water interface.
4. Use the result of step 3 to find n.
5. Use the result of step 4 to find v.

EVALUATE

6. How does the speed of light in the glass compare to the speed in water? Does this make sense?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q33.1 Light requires about 8 minutes to travel from the sun to the earth. Is it delayed appreciably by the earth's atmosphere? Explain.

Q33.2 Sunlight or starlight passing through the earth's atmosphere is always bent toward the vertical. Why? Does this mean that a star is not really where it appears to be? Explain.

Q33.3 A beam of light goes from one material into another. On *physical* grounds, explain *why* the wavelength changes but the frequency and period do not.

Q33.4 A student claimed that, because of atmospheric refraction (see Discussion Question Q33.2), the sun can be seen after it has set and that the day is therefore longer than it would be if the earth had no atmosphere. First, what does she mean by saying that the sun can be seen after it has set? Second, comment on the validity of her conclusion.

Q33.5 When hot air rises from a radiator or heating duct, objects behind it appear to shimmer or waver. What causes this?

Q33.6 Devise straightforward experiments to measure the speed of light in a given glass using (a) Snell's law; (b) total internal reflection; (c) Brewster's law.

Q33.7 Sometimes when looking at a window, you see two reflected images slightly displaced from each other. What causes this?

Q33.8 If you look up from underneath toward the surface of the water in your aquarium, you may see an upside-down reflection of your pet fish in the surface of the water. Explain how this can happen.

Q33.9 A ray of light in air strikes a glass surface. Is there a range of angles for which total internal reflection occurs? Explain.

Q33.10 When light is incident on an interface between two materials, the angle of the refracted ray depends on the wavelength, but the angle of the reflected ray does not. Why should this be?

Q33.11 A salesperson at a bargain counter claims that a certain pair of sunglasses has Polaroid filters; you suspect that the glasses are just tinted plastic. How could you find out for sure?

Q33.12 Does it make sense to talk about the polarization of a *longitudinal* wave, such as a sound wave? Why or why not?

Q33.13 How can you determine the direction of the polarizing axis of a single polarizer?

Q33.14 It has been proposed that automobile windshields and headlights should have polarizing filters to reduce the glare of oncoming lights during night driving. Would this work? How should the polarizing axes be arranged? What advantages would this scheme have? What disadvantages?

Q33.15 When a sheet of plastic food wrap is placed between two crossed polarizers, no light is transmitted. When the sheet is stretched in one direction, some light passes through the crossed polarizers. What is happening?

Q33.16 If you sit on the beach and look at the ocean through Polaroid sunglasses, the glasses help to reduce the glare from sunlight reflecting off the water. But if you lie on your side on the beach, there is little reduction in the glare. Explain why there is a difference.

Q33.17 When unpolarized light is incident on two crossed polarizers, no light is transmitted. A student asserted that if a third polarizer is inserted between the other two, some transmission will occur. Does this make sense? How can adding a third filter *increase* transmission?

Q33.18 For the old "rabbit-ear" style TV antennas, it's possible to alter the quality of reception considerably simply by changing the orientation of the antenna. Why?

Q33.19 In Fig. 33.31, since the light that is scattered out of the incident beam is polarized, why is the transmitted beam not also partially polarized?

Q33.20 You are sunbathing in the late afternoon when the sun is relatively low in the western sky. You are lying flat on your back, looking straight up through Polaroid sunglasses. To minimize the amount of sky light reaching your eyes, how should you lie: with your feet pointing north, east, south, west, or in some other direction? Explain your reasoning.

Q33.21 Light scattered from blue sky is strongly polarized because of the nature of the scattering process described in Section 33.6. But light scattered from white clouds is usually *not* polarized. Why not?

Q33.22 Atmospheric haze is due to water droplets or smoke particles ("smog"). Such haze reduces visibility by scattering light, so that the light from distant objects becomes randomized and images become indistinct. Explain why visibility through haze can be improved by wearing red-tinted sunglasses, which filter out blue light.

Q33.23 The explanation given in Section 33.6 for the color of the setting sun should apply equally well to the *rising* sun, since sunlight travels the same distance through the atmosphere to reach your eyes at either sunrise or sunset. Typically, however, sunsets are redder than sunrises. Why? (*Hint:* Particles of all kinds in the atmosphere contribute to scattering.)

Q33.24 Huygens's principle also applies to sound waves. During the day, the temperature of the atmosphere decreases with increasing altitude above the ground. But at night, when the ground cools, there is a layer of air just above the surface in which the temperature *increases* with altitude. Use this to explain why sound waves from distant sources can be heard more clearly at night than in the daytime. (*Hint:* The speed of sound increases with increasing temperature. Use the ideas displayed in Fig. 33.36 for light.)

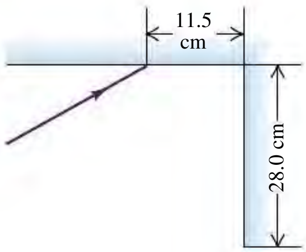
Q33.25 Can water waves be reflected and refracted? Give examples. Does Huygens's principle apply to water waves? Explain.

EXERCISES

Section 33.2 Reflection and Refraction

33.1 • Two plane mirrors intersect at right angles. A laser beam strikes the first of them at a point 11.5 cm from their point of intersection, as shown in **Fig. E33.1**. For what angle of incidence at the first mirror will this ray strike the midpoint of the second mirror (which is 28.0 cm long) after reflecting from the first mirror?

Figure **E33.1**



33.2 • BIO **Light Inside the Eye.** The vitreous humor, a transparent, gelatinous fluid that fills most of the eyeball, has an index of refraction of 1.34. Visible light ranges in wavelength from 380 nm (violet) to 750 nm (red), as measured in air. This light travels through the vitreous humor and strikes the rods and cones at the surface of the retina. What are the ranges of (a) the wavelength, (b) the frequency, and (c) the speed of the light just as it approaches the retina within the vitreous humor?

33.3 • A beam of light has a wavelength of 650 nm in vacuum. (a) What is the speed of this light in a liquid whose index of refraction at this wavelength is 1.47? (b) What is the wavelength of these waves in the liquid?

33.4 • Light with a frequency of 5.80×10^{14} Hz travels in a block of glass that has an index of refraction of 1.52. What is the wavelength of the light (a) in vacuum and (b) in the glass?

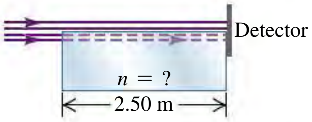
33.5 • A light beam travels at 1.94×10^{8} m/s in quartz. The wavelength of the light in quartz is 355 nm. (a) What is the index of refraction of quartz at this wavelength? (b) If this same light travels through air, what is its wavelength there?

33.6 •• Light of a certain frequency has a wavelength of 526 nm in water. What is the wavelength of this light in benzene?

33.7 •• A parallel beam of light in air makes an angle of 47.5° with the surface of a glass plate having a refractive index of 1.66. (a) What is the angle between the reflected part of the beam and the surface of the glass? (b) What is the angle between the refracted beam and the surface of the glass?

33.8 •• A laser beam shines along the surface of a block of transparent material (see **Fig. E33.8**). Half of the beam goes straight to a detector, while the other half travels through the block and then hits the detector. The time delay between the arrival of the two light beams at the detector is 6.25 ns. What is the index of refraction of this material?

Figure **E33.8**

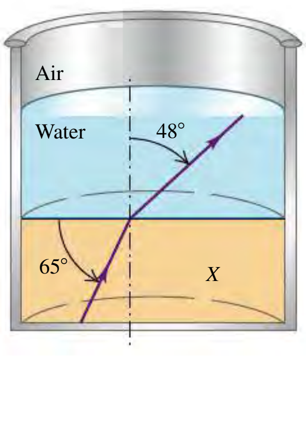


33.9 • Light traveling in air is incident on the surface of a block of plastic at an angle of 62.7° to the normal and is bent so that it makes a 48.1° angle with the normal in the plastic. Find the speed of light in the plastic.

33.10 • (a) A tank containing methanol has walls 2.50 cm thick made of glass of refractive index 1.550. Light from the outside air strikes the glass at a 41.3° angle with the normal to the glass. Find the angle the light makes with the normal in the methanol. (b) The tank is emptied and refilled with an unknown liquid. If light incident at the same angle as in part (a) enters the liquid in the tank at an angle of 20.2° from the normal, what is the refractive index of the unknown liquid?

33.11 •• As shown in **Fig. E33.11**, a layer of water covers a slab of material X in a beaker. A ray of light traveling upward follows the path indicated. Using the information on the figure, find (a) the index of refraction of material X and (b) the angle the light makes with the normal in the *air*.

Figure **E33.11**



33.12 •• A horizontal, parallel-sided plate of glass having a refractive index of 1.52 is in contact with the surface of water in a tank. A ray coming from above in air makes an angle of incidence of 35.0° with the normal to the top surface of the glass. (a) What angle does the ray refracted into the water make with the normal to the surface? (b) What is the dependence of this angle on the refractive index of the glass?

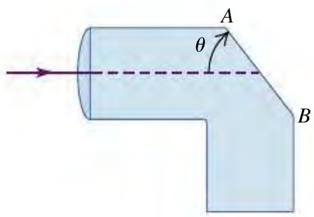
33.13 • A ray of light is incident on a plane surface separating two sheets of glass with refractive indexes 1.70 and 1.58. The angle of incidence is 62.0°, and the ray originates in the glass with $n = 1.70$. Compute the angle of refraction.

33.14 • A ray of light traveling in water is incident on an interface with a flat piece of glass. The wavelength of the light in the water is 726 nm, and its wavelength in the glass is 544 nm. If the ray in water makes an angle of 56.0° with respect to the normal to the interface, what angle does the refracted ray in the glass make with respect to the normal?

Section 33.3 Total Internal Reflection

33.15 • **Light Pipe.** Light enters a solid pipe made of plastic having an index of refraction of 1.60. The light travels parallel to the upper part of the pipe (**Fig. E33.15**). You want to cut the face AB so that all the light will reflect back into the pipe after it first strikes that face. (a) What is the largest that θ can be if the pipe is in air? (b) If the pipe is immersed in water of refractive index 1.33, what is the largest that θ can be?

Figure **E33.15**



33.16 • A flat piece of glass covers the top of a vertical cylinder that is completely filled with water. If a ray of light traveling in the glass is incident on the interface with the water at an angle of $\theta_a = 36.2°$, the ray refracted into the water makes an angle of 49.8° with the normal to the interface. What is the smallest value of the incident angle θ_a for which none of the ray refracts into the water?

33.17 •• The critical angle for total internal reflection at a liquid–air interface is 42.5°. (a) If a ray of light traveling in the liquid has an angle of incidence at the interface of 35.0°, what angle does the refracted ray in the air make with the normal? (b) If a ray of light traveling in air has an angle of incidence at the interface of 35.0°, what angle does the refracted ray in the liquid make with the normal?

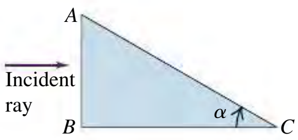
33.18 • A beam of light is traveling inside a solid glass cube that has index of refraction 1.62. It strikes the surface of the cube from the inside. (a) If the cube is in air, at what minimum angle with the normal inside the glass will this light *not* enter the air at this surface? (b) What would be the minimum angle in part (a) if the cube were immersed in water?

33.19 • A ray of light is traveling in a glass cube that is totally immersed in water. You find that if the ray is incident on the glass–water interface at an angle to the normal larger than 48.7°, no light is refracted into the water. What is the refractive index of the glass?

33.20 • At the very end of Wagner's series of operas *Ring of the Nibelung,* Brünnhilde takes the golden ring from the finger of the dead Siegfried and throws it into the Rhine, where it sinks to the bottom of the river. Assuming that the ring is small enough compared to the depth of the river to be treated as a point and that the Rhine is 10.0 m deep where the ring goes in, what is the area of the largest circle at the surface of the water over which light from the ring could escape from the water?

33.21 • Light is incident along the normal on face AB of a glass prism of refractive index 1.52, as shown in **Fig. E33.21**. Find the largest value the angle α can have without any light refracted out of the prism at face AC if (a) the prism is immersed in air and (b) the prism is immersed in water.

Figure **E33.21**

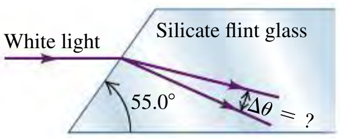


Section 33.4 Dispersion

33.22 • The indexes of refraction for violet light ($\lambda = 400$ nm) and red light ($\lambda = 700$ nm) in diamond are 2.46 and 2.41, respectively. A ray of light traveling through air strikes the diamond surface at an angle of 53.5° to the normal. Calculate the angular separation between these two colors of light in the refracted ray.

33.23 •• A narrow beam of white light strikes one face of a slab of silicate flint glass. The light is traveling parallel to the two adjoining faces, as shown in **Fig. E33.23**. For the transmitted light inside the glass, through what angle $\Delta\theta$ is the portion of the visible spectrum between 400 nm and 700 nm dispersed? (Consult the graph in Fig. 33.17.)

Figure **E33.23**



33.24 • A beam of light strikes a sheet of glass at an angle of 57.0° with the normal in air. You observe that red light makes an angle of 38.1° with the normal in the glass, while violet light makes a 36.7° angle. (a) What are the indexes of refraction of this glass for these colors of light? (b) What are the speeds of red and violet light in the glass?

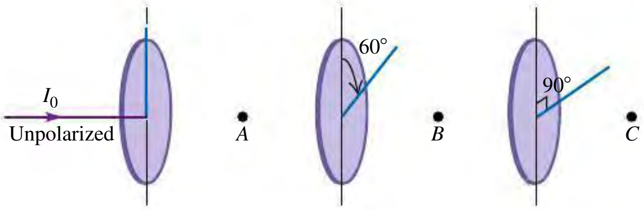
Section 33.5 Polarization

33.25 • Unpolarized light with intensity I_0 is incident on two polarizing filters. The axis of the first filter makes an angle of 60.0° with the vertical, and the axis of the second filter is horizontal. What is the intensity of the light after it has passed through the second filter?

33.26 •• (a) At what angle above the horizontal is the sun if sunlight reflected from the surface of a calm lake is completely polarized? (b) What is the plane of the electric-field vector in the reflected light?

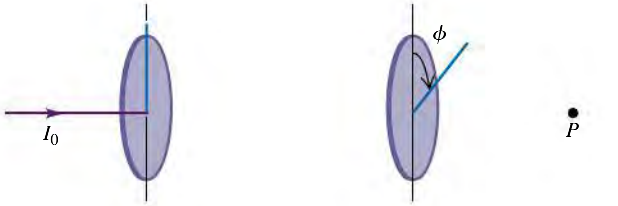
33.27 •• A beam of unpolarized light of intensity I_0 passes through a series of ideal polarizing filters with their polarizing axes turned to various angles as shown in **Fig. E33.27**. (a) What is the light intensity (in terms of I_0) at points A, B, and C? (b) If we remove the middle filter, what will be the light intensity at point C?

Figure **E33.27**



33.28 •• Light of original intensity I_0 passes through two ideal polarizing filters having their polarizing axes oriented as shown in **Fig. E33.28**. You want to adjust the angle ϕ so that the intensity at point P is equal to $I_0/10$. (a) If the original light is unpolarized, what should ϕ be? (b) If the original light is linearly polarized in the same direction as the polarizing axis of the first polarizer the light reaches, what should ϕ be?

Figure **E33.28**



33.29 • A parallel beam of unpolarized light in air is incident at an angle of 54.5° (with respect to the normal) on a plane glass surface. The reflected beam is completely linearly polarized. (a) What is the refractive index of the glass? (b) What is the angle of refraction of the transmitted beam?

33.30 • The refractive index of a certain glass is 1.66. For what incident angle is light reflected from the surface of this glass completely polarized if the glass is immersed in (a) air and (b) water?

33.31 •• A beam of polarized light passes through a polarizing filter. When the angle between the polarizing axis of the filter and the direction of polarization of the light is θ, the intensity of the emerging beam is I. If you now want the intensity to be $I/2$, what should be the angle (in terms of θ) between the polarizing angle of the filter and the original direction of polarization of the light?

33.32 ••• Three polarizing filters are stacked, with the polarizing axis of the second and third filters at 23.0° and 62.0°, respectively, to that of the first. If unpolarized light is incident on the stack, the light has intensity 55.0 W/cm^2 after it passes through the stack. If the incident intensity is kept constant but the second polarizer is removed, what is the intensity of the light after it has passed through the stack?

33.33 •• Unpolarized light of intensity 20.0 W/cm^2 is incident on two polarizing filters. The axis of the first filter is at an angle of 25.0° counterclockwise from the vertical (viewed in the direction the light is traveling), and the axis of the second filter is at 62.0° counterclockwise from the vertical. What is the intensity of the light after it has passed through the second polarizer?

33.34 • **Three Polarizing Filters.** Three polarizing filters are stacked with the polarizing axes of the second and third at 45.0° and 90.0°, respectively, with that of the first. (a) If unpolarized light of intensity I_0 is incident on the stack, find the intensity and state of polarization of light emerging from each filter. (b) If the second filter is removed, what is the intensity of the light emerging from each remaining filter?

Section 33.6 Scattering of Light

33.35 • A beam of white light passes through a uniform thickness of air. If the intensity of the scattered light in the middle of the green part of the visible spectrum is I, find the intensity (in terms of I) of scattered light in the middle of (a) the red part of the spectrum and (b) the violet part of the spectrum. Consult Table 32.1.

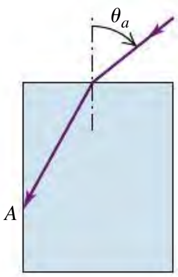
PROBLEMS

33.36 • A light beam is directed parallel to the axis of a hollow cylindrical tube. When the tube contains only air, the light takes 8.72 ns to travel the length of the tube, but when the tube is filled with a transparent jelly, the light takes 1.82 ns longer to travel its length. What is the refractive index of this jelly?

33.37 •• BIO **Heart Sonogram.** Physicians use high-frequency ($f = 1–5$ MHz) sound waves, called ultrasound, to image internal organs. The speed of these ultrasound waves is 1480 m/s in muscle and 344 m/s in air. We define the index of refraction of a material for sound waves to be the ratio of the speed of sound in air to the speed of sound in the material. Snell's law then applies to the refraction of sound waves. (a) At what angle from the normal does an ultrasound beam enter the heart if it leaves the lungs at an angle of 9.73° from the normal to the heart wall? (Assume that the speed of sound in the lungs is 344 m/s.) (b) What is the critical angle for sound waves in air incident on muscle?

33.38 ••• In a physics lab, light with wavelength 490 nm travels in air from a laser to a photocell in 17.0 ns. When a slab of glass 0.840 m thick is placed in the light beam, with the beam incident along the normal to the parallel faces of the slab, it takes the light 21.2 ns to travel from the laser to the photocell. What is the wavelength of the light in the glass?

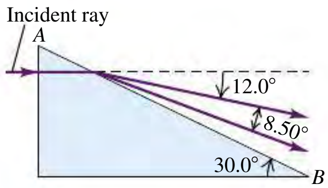
Figure **P33.39**



33.39 •• A ray of light is incident in air on a block of a transparent solid whose index of refraction is n. If $n = 1.38$, what is the *largest* angle of incidence θ_a for which total internal reflection will occur at the vertical face (point A shown in **Fig. P33.39**)?

33.40 • A light ray in air strikes the right-angle prism shown in **Fig. P33.40**. The prism angle at B is 30.0°. This ray consists of two different wavelengths. When it emerges at face AB, it has been split into two different rays that diverge from each other by 8.50°. Find the index of refraction of the prism for each of the two wavelengths.

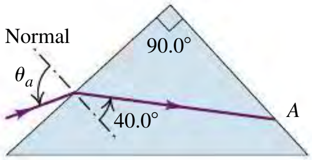
Figure **P33.40**



33.41 •• A ray of light traveling *in* a block of glass ($n = 1.52$) is incident on the top surface at an angle of 57.2° with respect to the normal in the glass. If a layer of oil is placed on the top surface of the glass, the ray is totally reflected. What is the maximum possible index of refraction of the oil?

33.42 •• A ray of light traveling in air is incident at angle θ_a on one face of a 90.0° prism made of glass. Part of the light refracts into the prism and strikes the opposite face at point A (**Fig. P33.42**). If the ray at A is at the critical angle, what is the value of θ_a?

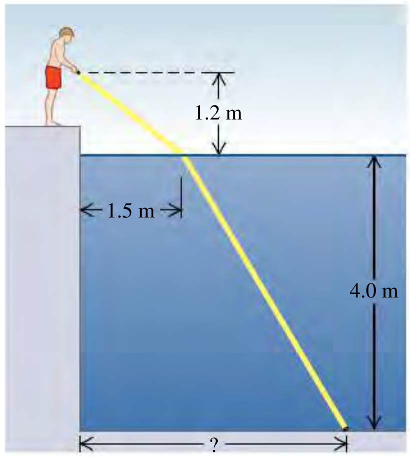
Figure **P33.42**



33.43 ••• A glass plate 2.50 mm thick, with an index of refraction of 1.40, is placed between a point source of light with wavelength 540 nm (in vacuum) and a screen. The distance from source to screen is 1.80 cm. How many wavelengths are there between the source and the screen?

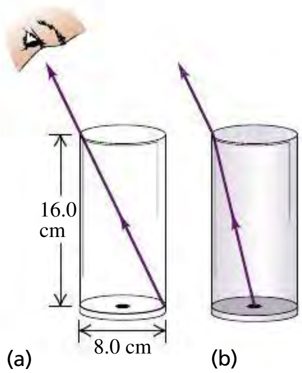
33.44 • After a long day of driving you take a late-night swim in a motel swimming pool. When you go to your room, you realize that you have lost your room key in the pool. You borrow a powerful flashlight and walk around the pool, shining the light into it. The light shines on the key, which is lying on the bottom of the pool, when the flashlight is held 1.2 m above the water surface and is directed at the surface a horizontal distance of 1.5 m from the edge (**Fig. P33.44**). If the water here is 4.0 m deep, how far is the key from the edge of the pool?

Figure **P33.44**



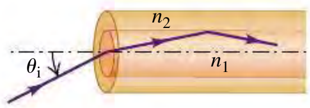
33.45 • You sight along the rim of a glass with vertical sides so that the top rim is lined up with the opposite edge of the bottom (**Fig. P33.45a**). The glass is a thin-walled, hollow cylinder 16.0 cm high. The diameter of the top and bottom of the glass is 8.0 cm. While you keep your eye in the same position, a friend fills the glass with a transparent liquid, and you then see a dime that is lying at the center of the bottom of the glass (Fig. P33.45b). What is the index of refraction of the liquid?

Figure **P33.45**



33.46 •• Optical fibers are constructed with a cylindrical core surrounded by a sheath of cladding material. Common materials used are pure silica ($n_2 = 1.450$) for the cladding and silica doped with germanium ($n_1 = 1.465$) for the core. (a) What is the critical angle θ_{crit} for light traveling in the core and reflecting at the interface with the cladding material? (b) The numerical aperture (NA) is defined as the angle of incidence θ_i at the flat end of the cable for which light is incident on the core–cladding interface at angle θ_{crit} (**Fig. P33.46**). Show that $\sin\theta_i = \sqrt{n_1^2 - n_2^2}$. (c) What is the value of θ_i for $n_1 = 1.465$ and $n_2 = 1.450$?

Figure **P33.46**

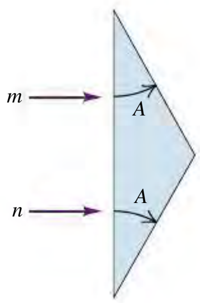


33.47 • A thin layer of ice ($n = 1.309$) floats on the surface of water ($n = 1.333$) in a bucket. A ray of light from the bottom of the bucket travels upward through the water. (a) What is the largest angle with respect to the normal that the ray can make at the ice–water interface and still pass out into the air above the ice? (b) What is this angle after the ice melts?

33.48 •• A 45°–45°–90° prism is immersed in water. A ray of light is incident normally on one of its shorter faces. What is the minimum index of refraction that the prism must have if this ray is to be totally reflected within the glass at the long face of the prism?

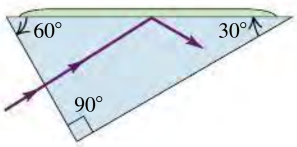
33.49 •• The prism shown in **Fig. P33.49** has a refractive index of 1.66, and the angles A are 25.0°. Two light rays m and n are parallel as they enter the prism. What is the angle between them after they emerge?

Figure **P33.49**



33.50 •• Light is incident normally on the short face of a 30°–60°–90° prism (**Fig. P33.50**). A drop of liquid is placed on the hypotenuse of the prism. If the index of refraction of the prism is 1.56, find the maximum index that the liquid may have for the light to be totally reflected.

Figure **P33.50**

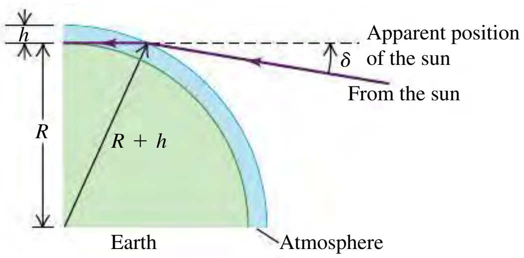


33.51 •• When the sun is either rising or setting and appears to be just on the horizon, it is in fact *below* the horizon. The explanation for this seeming paradox is that light from the sun bends slightly when entering the earth's atmosphere, as shown in **Fig. P33.51**. Since our perception is based on the idea that light travels in straight lines, we perceive the light to be coming from an apparent position that is an angle δ above the sun's true position. (a) Make the simplifying assumptions that the atmosphere has uniform density, and hence uniform index of refraction n, and extends to a height h above the earth's surface, at which point it abruptly stops. Show that the angle δ is given by

$$\delta = \arcsin\left(\frac{nR}{R+h}\right) - \arcsin\left(\frac{R}{R+h}\right)$$

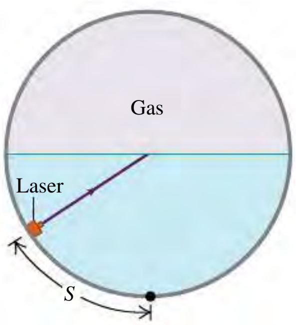
where $R = 6378$ km is the radius of the earth. (b) Calculate δ using $n = 1.0003$ and $h = 20$ km. How does this compare to the angular radius of the sun, which is about one quarter of a degree? (In actuality a light ray from the sun bends gradually, not abruptly, since the density and refractive index of the atmosphere change gradually with altitude.)

Figure **P33.51**



33.52 •• A horizontal cylindrical tank 2.20 m in diameter is half full of water. The space above the water is filled with a pressurized gas of unknown refractive index. A small laser can move along the curved bottom of the water and aims a light beam toward the center of the water surface (**Fig. P33.52**). You observe that when the laser has moved a distance $S = 1.09$ m or more (measured along the curved surface) from the lowest point in the water, no light enters the gas. (a) What is the index of refraction of the gas? (b) What minimum time does it take the light beam to travel from the laser to the rim of the tank when (i) $S > 1.09$ m and (ii) $S < 1.09$ m?

Figure **P33.52**

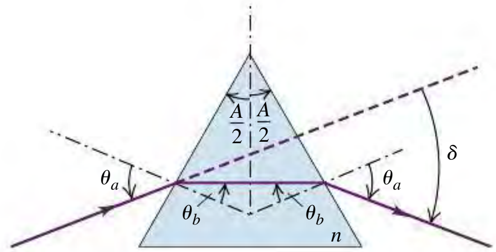


33.53 •• **Angle of Deviation.** The incident angle θ_a shown in **Fig. P33.53** is chosen so that the light passes symmetrically through the prism, which has refractive index n and apex angle A. (a) Show that the angle of deviation δ (the angle between the initial and final directions of the ray) is given by

$$\sin\frac{A+\delta}{2} = n\sin\frac{A}{2}$$

(When the light passes through symmetrically, as shown, the angle of deviation is a minimum.) (b) Use the result of part (a) to find the angle of deviation for a ray of light passing symmetrically through a prism having three equal angles ($A = 60.0°$) and $n = 1.52$. (c) A certain glass has a refractive index of 1.61 for red light (700 nm) and 1.66 for violet light (400 nm). If both colors pass through symmetrically, as described in part (a), and if $A = 60.0°$, find the difference between the angles of deviation for the two colors.

Figure **P33.53**

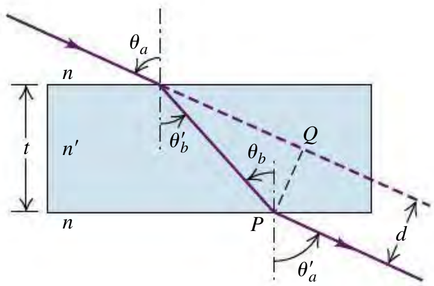


33.54 •• Light is incident in air at an angle θ_a (**Fig. P33.54**) on the upper surface of a transparent plate, the surfaces of the plate being plane and parallel to each other. (a) Prove that $\theta_a = \theta'_a$. (b) Show that this is true for any number of different parallel plates. (c) Prove that the lateral displacement d of the emergent beam is given by the relationship

$$d = t\frac{\sin(\theta_a - \theta'_b)}{\cos\theta'}$$

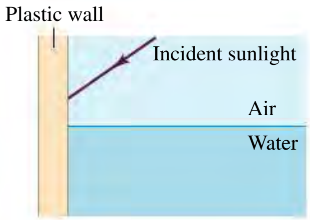
where t is the thickness of the plate. (d) A ray of light is incident at an angle of 66.0° on one surface of a glass plate 2.40 cm thick with an index of refraction of 1.80. The medium on either side of the plate is air. Find the lateral displacement between the incident and emergent rays.

Figure **P33.54**



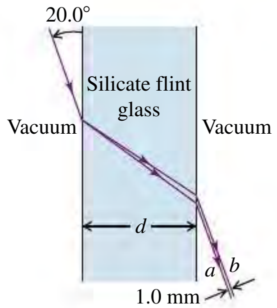
33.55 •• A beam of unpolarized sunlight strikes the vertical plastic wall of a water tank at an unknown angle. Some of the light reflects from the wall and enters the water (**Fig. P33.55**). The refractive index of the plastic wall is 1.61. If the light that has been reflected from the wall into the water is observed to be completely polarized, what angle does this beam make with the normal inside the water?

Figure **P33.55**



33.56 • A thin beam of white light is directed at a flat sheet of silicate flint glass at an angle of 20.0° to the surface of the sheet. Due to dispersion in the glass, the beam is spread out in a spectrum as shown in **Fig. P33.56**. The refractive index of silicate flint glass versus wavelength is graphed in Fig. 33.17. (a) The rays a and b shown in Fig. P33.56 correspond to the extreme wavelengths shown in Fig. 33.17. Which corresponds to red and which to violet? Explain your reasoning. (b) For what thickness d of the glass sheet will the spectrum be 1.0 mm wide, as shown (see Problem 33.54)?

Figure **P33.56**



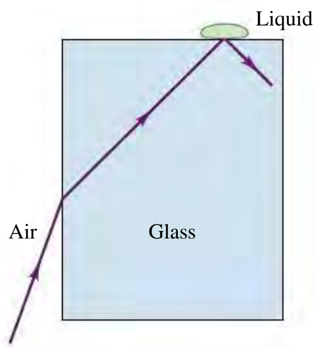
33.57 •• **DATA** In physics lab, you are studying the properties of four transparent liquids. You shine a ray of light (in air) onto the surface of each liquid—A, B, C, and D—one at a time, at a 60.0° angle of incidence; you then measure the angle of refraction. The table gives your data:

Liquid	A	B	C	D
θ_a **(°)**	36.4	40.5	32.1	35.2

The wavelength of the light when it is traveling in air is 589 nm. (a) Find the refractive index of each liquid at this wavelength. Use Table 33.1 to identify each liquid, assuming that all four are listed in the table. (b) For each liquid, what is the dielectric constant K at the frequency of the 589-nm light? For each liquid, the relative permeability (K_m) is very close to unity. (c) What is the frequency of the light in air and in each liquid?

33.58 •• **DATA** Given small samples of three liquids, you are asked to determine their refractive indexes. However, you do not have enough of each liquid to measure the angle of refraction for light refracting from air into the liquid. Instead, for each liquid, you take a rectangular block of glass ($n = 1.52$) and place a drop of the liquid on the top surface of the block. You shine a laser beam with wavelength 638 nm in vacuum at one side of the block and measure the largest angle of incidence θ_a for which there is total internal reflection at the interface between the glass and the liquid (**Fig. P33.58**). Your results are given in the table:

Figure **P33.58**

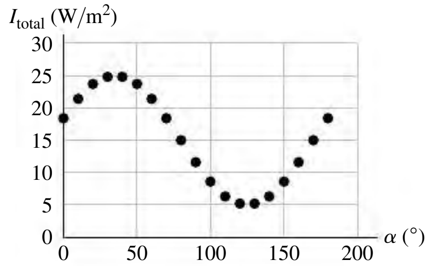


Liquid	A	B	C
θ_a **(°)**	52.0	44.3	36.3

What is the refractive index of each liquid at this wavelength?

33.59 •• **DATA** A beam of light traveling horizontally is made of an unpolarized component with intensity I_0 and a polarized component with intensity I_p. The plane of polarization of the polarized component is oriented at an angle θ with respect to the vertical. **Figure P33.59** is a graph of the total intensity I_{total} after the light passes through a polarizer versus the angle α that the polarizer's axis makes with respect to the vertical. (a) What is the orientation of the polarized component? (That is, what is θ?) (b) What are the values of I_0 and I_p?

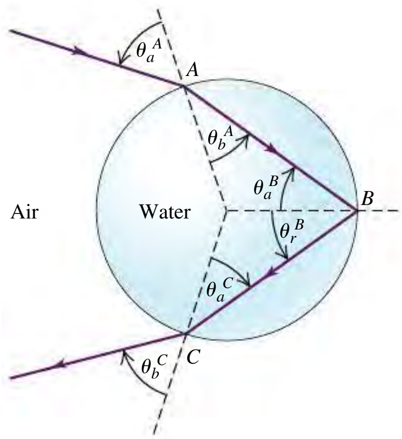
Figure **P33.59**



CHALLENGE PROBLEMS

33.60 ••• **CALC** A rainbow is produced by the reflection of sunlight by spherical drops of water in the air. **Figure P33.60** shows a ray that refracts into a drop at point A, is reflected from the back surface of the drop at point B, and refracts back into the air at point C. The angles of incidence and refraction, θ_a and θ_b, are shown at points A and C, and the angles of incidence and reflection, θ_a and θ_r, are shown at point B. (a) Show that $\theta_a^B = \theta_b^A$, $\theta_a^C = \theta_b^A$, and $\theta_b^C = \theta_a^A$. (b) Show that the angle in radians between the ray before it enters the drop at A and after it exits at C (the total angular deflection of the ray) is $\Delta = 2\theta_a^A - 4\theta_b^A + \pi$. (*Hint:* Find the angular deflections that occur at A, B, and C, and add them to get Δ.) (c) Use Snell's law to write Δ in terms of θ_a^A and n, the refractive index of the water in the drop. (d) A rainbow will form when the angular deflection Δ is *stationary* in the incident angle θ_a^A—that is, when $d\Delta/d\theta_a^A = 0$. If this condition is satisfied, all the rays with incident angles close to θ_a^A will be sent back in the same direction, producing a bright zone in the sky. Let θ_1 be the value of θ_a^A for which this occurs. Show that $\cos^2\theta_1 = \frac{1}{3}(n^2 - 1)$. (*Hint:* You may find the derivative formula $d(\arcsin u(x))/dx = (1 - u^2)^{-1/2}(du/dx)$ helpful.) (e) The index

Figure **P33.60**



of refraction in water is 1.342 for violet light and 1.330 for red light. Use the results of parts (c) and (d) to find θ_1 and Δ for violet and red light. Do your results agree with the angles shown in Fig. 33.19d? When you view the rainbow, which color, red or violet, is higher above the horizon?

33.61 ••• **CALC** A *secondary rainbow* is formed when the incident light undergoes two internal reflections in a spherical drop of water as shown in Fig. 33.19e. (See Challenge Problem 33.60.) (a) In terms of the incident angle θ_a^A and the refractive index n of the drop, what is the angular deflection Δ of the ray? That is, what is the angle between the ray before it enters the drop and after it exits? (b) What is the incident angle θ_2 for which the derivative of Δ with respect to the incident angle θ_a^A is zero? (c) The indexes of refraction for red and violet light in water are given in part (e) of Challenge Problem 33.60. Use the results of parts (a) and (b) to find θ_2 and Δ for violet and red light. Do your results agree with the angles shown in Fig. 33.19e? When you view a secondary rainbow, is red or violet higher above the horizon? Explain.

PASSAGE PROBLEMS

BIO SEEING POLARIZED LIGHT. Some insect eyes have two types of cells that are sensitive to the plane of polarization of light. In a simple model, one cell type (type H) is sensitive to horizontally polarized light only, and the other cell type (type V) is sensitive to vertically polarized light only. To study the responses of these cells, researchers fix the insect in a normal, upright position so that one eye is illuminated by a light source. Then several experiments are carried out.

33.62 First, light with a plane of polarization at 45° to the horizontal shines on the insect. Which statement is true about the two types of cells? (a) Both types detect this light. (b) Neither type detects this light. (c) Only type H detects the light. (d) Only type V detects the light.

33.63 Next, unpolarized light is reflected off a smooth horizontal piece of glass, and the reflected light shines on the insect. Which statement is true about the two types of cells? (a) When the light is directly above the glass, only type V detects the reflected light. (b) When the light is directly above the glass, only type H detects the reflected light. (c) When the light is about 35° above the horizontal, type V responds much more strongly than type H does. (d) When the light is about 35° above the horizontal, type H responds much more strongly than type V does.

33.64 To vary the angle as well as the intensity of polarized light, ordinary unpolarized light is passed through one polarizer with its transmission axis vertical, and then a second polarizer is placed between the first polarizer and the insect. When the light leaving the second polarizer has half the intensity of the original unpolarized light, which statement is true about the two types of cells? (a) Only type H detects this light. (b) Only type V detects this light. (c) Both types detect this light, but type H detects more light. (d) Both types detect this light, but type V detects more light.

Answers

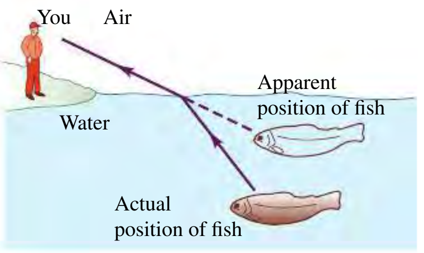
Chapter Opening Question ?

(iv) The brilliance and color of a diamond are due to total internal reflection from its surfaces (Section 33.3) and to dispersion, which spreads this light into a spectrum (Section 33.4).

Test Your Understanding Questions

33.1 (iii) The waves go farther in the y-direction in a given amount of time than in the other directions, so the wave fronts are elongated in the y-direction.

33.2 (a) (ii), (b) (iii) As shown in the figure, light rays coming from the fish bend away from the normal when they pass from the water $(n = 1.33)$ into the air $(n = 1.00)$. As a result, the fish appears to be higher in the water than it actually is. Hence you should aim a spear *below* the apparent position of the fish. If you use a laser beam, you should aim *at* the apparent position of the fish: The beam of laser light takes the same path from you to the fish as ordinary light takes from the fish to you (though in the opposite direction).



33.3 (i), (ii) Total internal reflection can occur only if two conditions are met: n_b must be less than n_a, and the critical angle θ_{crit} (where $\sin\theta_{crit} = n_b/n_a$) must be smaller than the angle of incidence θ_a. In the first two cases both conditions are met: The critical angles are (i) $\theta_{crit} = \sin^{-1}(1/1.33) = 48.8°$ and (ii) $\theta_{crit} = \sin^{-1}(1.33/1.52) = 61.0°$, both of which are smaller than $\theta_a = 70°$. In the third case $n_b = 1.52$ is greater than $n_a = 1.33$, so total internal reflection cannot occur for any incident angle.

33.5 (ii) The sunlight reflected from the windows of the high-rise building is partially polarized in the vertical direction, perpendicular to the horizontal plane of incidence. The Polaroid filter in front of the lens is oriented with its polarizing axis perpendicular to the dominant direction of polarization of the reflected light.

33.7 (ii) Huygens's principle applies to waves of all kinds, including sound waves. Hence this situation is exactly like that shown in Fig. 33.35, with material a representing the warm air, material b representing the cold air in which the waves travel more slowly, and the interface between the materials representing the weather front. North is toward the top of the figure and east is toward the right, so Fig. 33.35 shows that the rays (which indicate the direction of propagation) deflect toward the east.

Bridging Problem

1.93×10^8 m/s

? This surgeon performing microsurgery needs a sharp, magnified view of the surgical site. To obtain this, she's wearing glasses with magnifying lenses that must be (i) at a particular distance from her eye; (ii) at a particular distance from the object being magnified; (iii) both (i) and (ii); (iv) neither (i) nor (ii).

34 GEOMETRIC OPTICS

LEARNING GOALS

Looking forward at ...

34.1 How a plane mirror forms an image.

34.2 Why concave and convex mirrors form images of different kinds.

34.3 How images can be formed by a curved interface between two transparent materials.

34.4 What aspects of a lens determine the type of image that it produces.

34.5 What determines the field of view of a camera lens.

34.6 What causes various defects in human vision, and how they can be corrected.

34.7 The principle of the simple magnifier.

34.8 How microscopes and telescopes work.

Looking back at ...

33.2 Reflection and refraction.

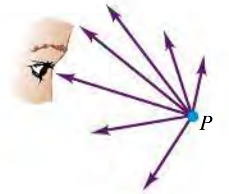
Your reflection in the bathroom mirror, the view of the moon through a telescope, an insect seen through a magnifying lens—all of these are examples of *images*. In each case the object that you're looking at appears to be in a different place than its actual position: Your reflection is on the other side of the mirror, the moon appears to be much closer when seen through a telescope, and an insect seen through a magnifying lens appears *more distant* (so your eye can focus on it easily). In each case, light rays that come from a point on an object are deflected by reflection or refraction (or a combination of the two), so they converge toward or appear to diverge from a point called an *image point*. Our goal in this chapter is to see how this is done and to explore the different kinds of images that can be made with simple optical devices.

To understand images and image formation, all we need are the ray model of light, the laws of reflection and refraction (Section 33.2), and some simple geometry and trigonometry. The key role played by geometry in our analysis explains why we give the name *geometric optics* to the study of how light rays form images. We'll begin our analysis with one of the simplest image-forming optical devices, a plane mirror. We'll go on to study how images are formed by curved mirrors, by refracting surfaces, and by thin lenses. Our results will lay the foundation for understanding many familiar optical instruments, including camera lenses, magnifiers, the human eye, microscopes, and telescopes.

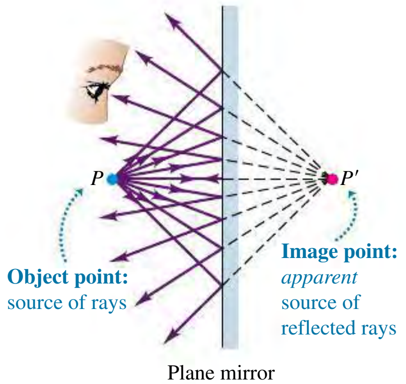
34.1 REFLECTION AND REFRACTION AT A PLANE SURFACE

Before discussing what is meant by an image, we first need the concept of **object** as it is used in optics. By an *object* we mean anything from which light rays radiate. This light could be emitted by the object itself if it is *self-luminous,* like the glowing filament of a light bulb. Alternatively, the light could be emitted by another source (such as a lamp or the sun) and then reflected from the object;

34.1 Light rays radiate from a point object P in all directions. For an observer to see this object directly, there must be no obstruction between the object and the observer's eyes.

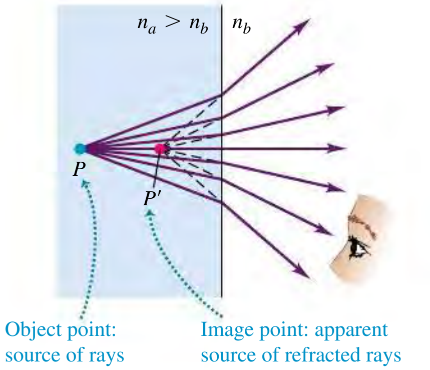


34.2 Light rays from the object at point P are reflected from a plane mirror. The reflected rays entering the eye look as though they had come from image point P'.



34.3 Light rays from the object at point P are refracted at the plane interface. The refracted rays entering the eye look as though they had come from image point P'.

When $n_a > n_b$, P' is closer to the surface than P; for $n_a < n_b$, the reverse is true.



an example is the light you see coming from the pages of this book. **Figure 34.1** shows light rays radiating in all directions from an object at a point P. Note that light rays from the object reach the observer's left and right eyes at different angles; these differences are processed by the observer's brain to infer the *distance* from the observer to the object.

The object in Fig. 34.1 is a **point object** that has no physical extent. Real objects with length, width, and height are called **extended objects.** To start with, we'll consider only an idealized point object, since we can always think of an extended object as being made up of a very large number of point objects.

Suppose some of the rays from the object strike a smooth, plane reflecting surface (**Fig. 34.2**). This could be the surface of a material with a different index of refraction, which reflects part of the incident light, or a polished metal surface that reflects almost 100% of the light that strikes it. We will always draw the reflecting surface as a black line with a shaded area behind it, as in Fig. 34.2. Bathroom mirrors have a thin sheet of glass that lies in front of and protects the reflecting surface; we'll ignore the effects of this thin sheet.

According to the law of reflection, all rays striking the surface are reflected at an angle from the normal equal to the angle of incidence. Since the surface is plane, the normal is in the same direction at all points on the surface, and we have *specular* reflection. After the rays are reflected, their directions are the same as though they had come from point P'. We call point P an *object point* and point P' the corresponding *image point,* and we say that the reflecting surface forms an **image** of point P. An observer who can see only the rays reflected from the surface, and who doesn't know that he's seeing a reflection, *thinks* that the rays originate from the image point P'. The image point is therefore a convenient way to describe the directions of the various reflected rays, just as the object point P describes the directions of the rays arriving at the surface *before* reflection.

If the surface in Fig. 34.2 were *not* smooth, the reflection would be *diffuse*. Rays reflected from different parts of the surface would go in uncorrelated directions (see Fig. 33.6b), and there would be no definite image point P' from which all reflected rays seem to emanate. You can't see your reflection in a tarnished piece of metal because its surface is rough; polishing the metal smoothes the surface so that specular reflection occurs and a reflected image becomes visible.

A plane *refracting* surface also forms an image (**Fig. 34.3**). Rays coming from point P are refracted at the interface between two optical materials. When the angles of incidence are small, the final directions of the rays after refraction are the same as though they had come from an *image point* P' as shown. In Section 33.2 we described how this effect makes underwater objects appear closer to the surface than they really are (see Fig. 33.9).

In both Figs. 34.2 and 34.3 the rays do not actually pass through the image point P'. Indeed, if the mirror in Fig. 34.2 is opaque, there is no light at all on its right side. If the outgoing rays don't actually pass through the image point, we call the image a **virtual image.** Later we will see cases in which the outgoing rays really *do* pass through an image point, and we will call the resulting image a **real image.** The images that are formed on a projection screen, on the electronic sensor in a camera, and on the retina of your eye are real images.

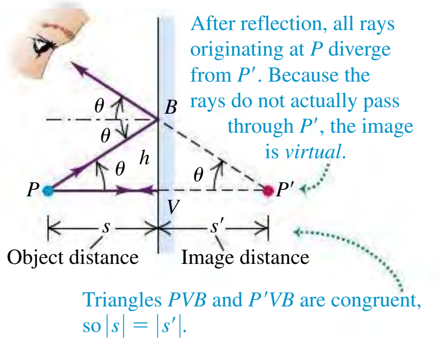
Image Formation by a Plane Mirror

Let's concentrate for now on images produced by *reflection;* we'll return to refraction later in the chapter. **Figure 34.4** shows how to find the precise location of the virtual image P' that a plane mirror forms of an object at P. The diagram shows two rays diverging from an object point P at a distance s to the left of a plane mirror. We call s the **object distance.** The ray PV is perpendicular to the mirror surface, and it returns along its original path.

The ray PB makes an angle θ with PV. It strikes the mirror at an angle of incidence θ and is reflected at an equal angle with the normal. When we extend the two reflected rays backward, they intersect at point P', at a distance s' behind the mirror. We call s' the **image distance.** The line between P and P' is perpendicular to the mirror. The two triangles PVB and $P'VB$ are congruent, so P and P' are at equal distances from the mirror, and s and s' have equal magnitudes. The image point P' is located exactly opposite the object point P as far *behind* the mirror as the object point is from the front of the mirror.

We can repeat the construction of Fig. 34.4 for each ray diverging from P. The directions of *all* the outgoing reflected rays are the same as though they had originated at point P', confirming that P' is the *image* of P. No matter where the observer is located, she will always see the image at the point P'.

34.4 Construction for determining the location of the image formed by a plane mirror. The image point P' is as far behind the mirror as the object point P is in front of it.



Sign Rules

Before we go further, let's introduce some general sign rules. These may seem unnecessarily complicated for the simple case of an image formed by a plane mirror, but we want to state the rules in a form that will be applicable to *all* the situations we will encounter later. These will include image formation by a plane or spherical reflecting or refracting surface, or by a pair of refracting surfaces forming a lens. Here are the rules:

1. **Sign rule for the object distance:** When the object is on the same side of the reflecting or refracting surface as the incoming light, object distance s is positive; otherwise, it is negative.
2. **Sign rule for the image distance:** When the image is on the same side of the reflecting or refracting surface as the outgoing light, image distance s' is positive; otherwise, it is negative.
3. **Sign rule for the radius of curvature of a spherical surface:** When the center of curvature C is on the same side as the outgoing light, the radius of curvature is positive; otherwise, it is negative.

34.5 For both of these situations, the object distance s is positive (rule 1) and the image distance s' is negative (rule 2).

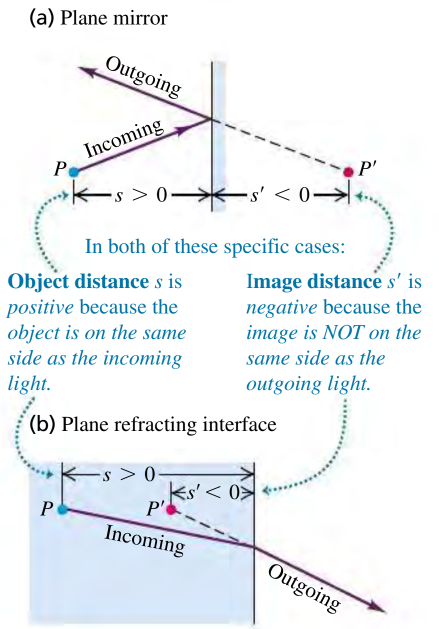


Figure 34.5 illustrates rules 1 and 2 for two different situations. For a mirror the incoming and outgoing sides are always the same; for example, in Figs. 34.2, 34.4, and 34.5a they are both on the left side. For the refracting surfaces in Figs. 34.3 and 34.5b the incoming and outgoing sides are on the left and right sides, respectively, of the interface between the two materials. (Note that other textbooks may use different rules.)

In Figs. 34.4 and 34.5a the object distance s is *positive* because the object point P is on the incoming side (the left side) of the reflecting surface. The image distance s' is *negative* because the image point P' is *not* on the outgoing side (the left side) of the surface. The object and image distances s and s' are related by

$$s = -s' \qquad \text{(plane mirror)} \tag{34.1}$$

For a plane reflecting or refracting surface, the radius of curvature is infinite and not a particularly interesting or useful quantity; in these cases we really don't need sign rule 3. But this rule will be of great importance when we study image formation by *curved* reflecting and refracting surfaces later in the chapter.

34.6 Construction for determining the height of an image formed by reflection at a plane reflecting surface.

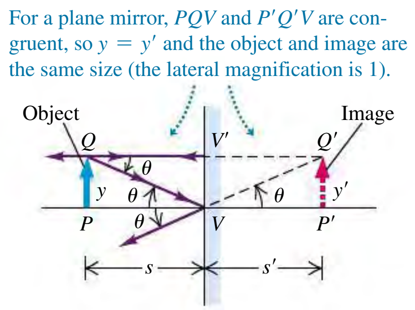


Image of an Extended Object: Plane Mirror

Next we consider an *extended* object with finite size. For simplicity we often consider an object that has only one dimension, like a slender arrow, oriented parallel to the reflecting surface; an example is the arrow PQ in **Fig. 34.6**. The distance from the head to the tail of an arrow oriented in this way is called its *height*. In Fig. 34.6 the height is y. The image formed by such an extended object is an extended image; to each point on the object, there corresponds a point on the image. Two of the rays from Q are shown; *all* the rays from Q appear to diverge from its image point Q' after reflection. The image of the arrow is the line $P'Q'$, with height y'. Other points of the object PQ have image points between P' and Q'. The triangles PQV and $P'Q'V$ are congruent, so the object PQ and image $P'Q'$ have the same size and orientation, and $y = y'$.

The ratio of image height to object height, y'/y, in *any* image-forming situation is called the **lateral magnification** m; that is,

Lateral magnification in an image-forming situation

$$m = \frac{y'}{y} \quad (34.2)$$

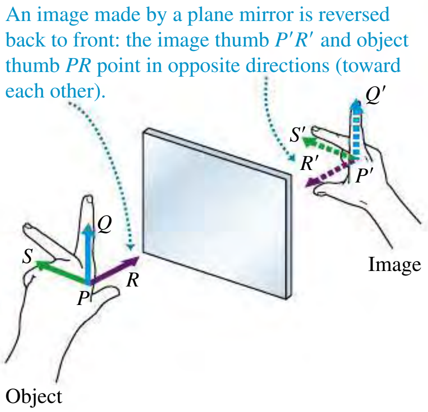
(y': Image height; y: Object height)

For a plane mirror $y = y'$, so the lateral magnification m is unity. When you look at yourself in a plane mirror, your image is the same size as the real you.

In Fig. 34.6 the image arrow points in the *same* direction as the object arrow; we say that the image is **erect.** In this case, y and y' have the same sign, and the lateral magnification m is positive. The image formed by a plane mirror is always erect, so y and y' have both the same magnitude and the same sign; from Eq. (34.2) the lateral magnification of a plane mirror is always $m = +1$. Later we will encounter situations in which the image is **inverted;** that is, the image arrow points in the direction *opposite* to that of the object arrow. For an inverted image, y and y' have *opposite* signs, and the lateral magnification m is *negative*.

The object in Fig. 34.6 has only one dimension. **Figure 34.7** shows a *three-dimensional* object and its three-dimensional virtual image formed by a plane mirror. The object and image are related in the same way as a left hand and a right hand.

34.7 The image formed by a plane mirror is virtual, erect, and reversed. It is the same size as the object.



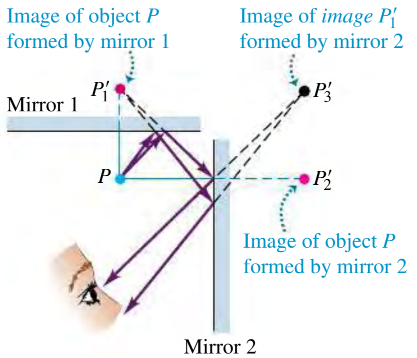
CAUTION **Reflections in a plane mirror** At this point, you may be asking, "Why does a plane mirror reverse images left and right but not top and bottom?" This question is quite misleading! As Fig. 34.7 shows, the up-down image $P'Q'$ and the left-right image $P'S'$ are parallel to their objects and are not reversed at all. Only the front-back image $P'R'$ is reversed relative to PR. Hence it's most correct to say that a plane mirror reverses *back to front*. When an object and its image are related in this way, the image is said to be **reversed;** this means that *only* the front-back dimension is reversed.

The reversed image of a three-dimensional object formed by a plane mirror is the same *size* as the object in all its dimensions. When the transverse dimensions of object and image are in the same direction, the image is erect. Thus a plane mirror always forms an erect but reversed image (**Fig. 34.8**).

34.8 The image formed by a plane mirror is reversed; the image of a right hand is a left hand, and so on. (The hand is resting on a horizontal mirror.) Are images of the letters I, H, and T reversed?

An important property of all images formed by reflecting or refracting surfaces is that an *image* formed by one surface or optical device can serve as the *object* for a second surface or device. **Figure 34.9** shows a simple example. Mirror 1 forms an image P_1' of the object point P, and mirror 2 forms another image P_2', each in the way we have just discussed. But in addition, the image P_1' formed by mirror 1 serves as an object for mirror 2, which then forms an image of this object at point P_3' as shown. Similarly, mirror 1 uses the image P_2' formed by mirror 2 as an object and forms an image of it. We leave it to you to show that this image point is also at P_3'. The idea that an image formed by one device can act as the object for a second device is of great importance. We'll use it later in this chapter to locate the image formed by two successive curved-surface refractions in a lens. We'll also use it to understand image formation by combinations of lenses, as in a microscope or a refracting telescope.

34.9 Images P_1' and P_2' are formed by a single reflection of each ray from the object at P. Image P_3', located by treating either of the other images as an object, is formed by a double reflection of each ray.



TEST YOUR UNDERSTANDING OF SECTION 34.1 If you walk directly toward a plane mirror at a speed v, at what speed does your image approach you? (i) Slower than v; (ii) v; (iii) faster than v but slower than $2v$; (iv) $2v$; (v) faster than $2v$.

34.2 REFLECTION AT A SPHERICAL SURFACE

A plane mirror produces an image that is the same size as the object. But there are many applications for mirrors in which the image and object must be of different sizes. A magnifying mirror used when applying makeup gives an image that is *larger* than the object, and surveillance mirrors (used in stores to help spot shoplifters) give an image that is *smaller* than the object. There are also applications of mirrors in which a *real* image is desired, so light rays do indeed pass through the image point P'. A plane mirror by itself cannot perform any of these tasks. Instead, *curved* mirrors are used.

Image of a Point Object: Spherical Mirror

We'll consider the special (and easily analyzed) case of image formation by a *spherical* mirror. **Figure 34.10a** shows a spherical mirror with radius of curvature R, with its concave side facing the incident light. The **center of curvature** of the surface (the center of the sphere of which the surface is a part) is at C, and the **vertex** of the mirror (the center of the mirror surface) is at V. The line CV is called the **optic axis.** Point P is an object point that lies on the optic axis; for the moment, we assume that the distance from P to V is greater than R.

Ray PV, passing through C, strikes the mirror normally and is reflected back on itself. Ray PB, at an angle α with the axis, strikes the mirror at B, where the angles of incidence and reflection are θ. The reflected ray intersects the axis at point P'. We will show shortly that *all* rays from P intersect the axis at the *same* point P', as in Fig. 34.10b, provided that the angle α is small. Point P' is therefore the *image* of object point P. Unlike the reflected rays in Fig. 34.1, the reflected rays in Fig. 34.10b actually do intersect at point P', then diverge from P' *as if* they had originated at this point. Thus P' is a *real* image.

To see the usefulness of having a real image, suppose that the mirror is in a darkened room in which the only source of light is a self-luminous object at P. If you place a small piece of photographic film at P', all the rays of light coming from point P that reflect off the mirror will strike the same point P' on the film; when developed, the film will show a single bright spot, representing a sharply focused image of the object at point P. This principle is at the heart of most astronomical telescopes, which use large concave mirrors to make photographs of celestial objects. With a *plane* mirror like that in Fig. 34.2, the light rays never actually pass through the image point, and the image can't be recorded on film. Real images are *essential* for photography.

Let's now locate the real image point P' in Fig. 34.10a and prove that all rays from P intersect at P' (provided that their angle with the optic axis is small). The object distance, measured from the vertex V, is s; the image distance, also measured from V, is s'. The signs of s, s', and the radius of curvature R are determined by the sign rules given in Section 34.1. The object point P is on the same side as the incident light, so according to sign rule 1, s is positive. The image point P' is on the same side as the reflected light, so according to sign rule 2, the image distance s' is also positive. The center of curvature C is on the same side as the reflected light, so according to sign rule 3, R, too, is positive; R is always positive when reflection occurs at the *concave* side of a surface (**Fig. 34.11**).

We now use the following theorem from plane geometry: An exterior angle of a triangle equals the sum of the two opposite interior angles. Applying this theorem to triangles PBC and $P'BC$ in Fig. 34.10a, we have

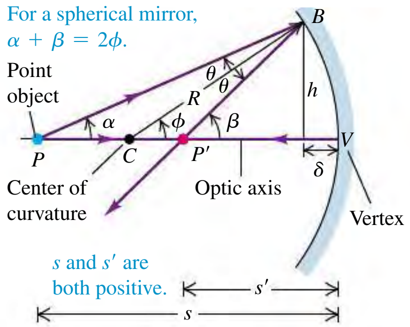
$$\phi = \alpha + \theta \qquad \beta = \phi + \theta$$

Eliminating θ between these equations gives

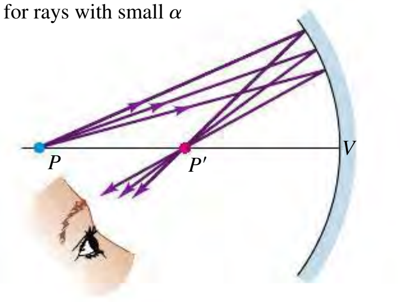
$$\alpha + \beta = 2\phi \tag{34.3}$$

34.10 (a) A concave spherical mirror forms a real image of a point object P on the mirror's optic axis. (b) The eye sees some of the outgoing rays and perceives them as having come from P'.

(a) Construction for finding the position P' of an image formed by a concave spherical mirror

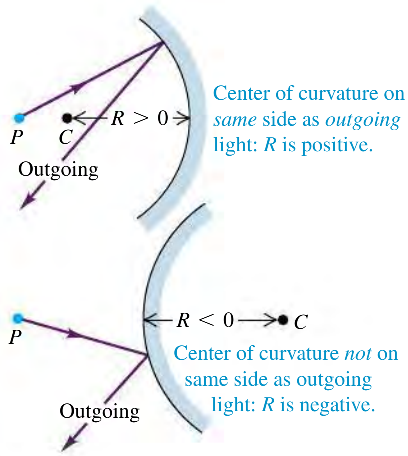


(b) The paraxial approximation, which holds for rays with small α



All rays from P that have a small angle α pass through P', forming a real image.

34.11 The sign rule for the radius of a spherical mirror.



34.12 (a), (b) Soon after the Hubble Space Telescope (HST) was placed in orbit in 1990, it was discovered that the concave primary mirror (also called the *objective mirror*) was too shallow by about $\frac{1}{50}$ the width of a human hair, leading to spherical aberration of the star's image. (c) After corrective optics were installed in 1993, the effects of spherical aberration were almost completely eliminated.

(a) The 2.4-m-diameter primary mirror of the Hubble Space Telescope

(b) A star seen with the original mirror

(c) The same star with corrective optics

We may now compute the image distance s'. Let h represent the height of point B above the optic axis, and let δ represent the short distance from V to the foot of this vertical line. We now write expressions for the tangents of α, β, and ϕ, remembering that s, s', and R are all positive quantities:

$$\tan\alpha = \frac{h}{s-\delta} \qquad \tan\beta = \frac{h}{s'-\delta} \qquad \tan\phi = \frac{h}{R-\delta}$$

These trigonometric equations cannot be solved as simply as the corresponding algebraic equations for a plane mirror. However, *if the angle α is small,* the angles β and ϕ are also small. The tangent of an angle that is much less than one radian is nearly equal to the angle itself (measured in radians), so we can replace $\tan\alpha$ by α, and so on, in the equations above. Also, if α is small, we can ignore the distance δ compared with s', s, and R. So for small angles we have the following approximate relationships:

$$\alpha = \frac{h}{s} \qquad \beta = \frac{h}{s'} \qquad \phi = \frac{h}{R}$$

Substituting these into Eq. (34.3) and dividing by h, we obtain a general relationship among s, s', and R:

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} \qquad \text{(object–image relationship, spherical mirror)} \qquad (34.4)$$

This equation does not contain the angle α. Hence *all* rays from P that make sufficiently small angles with the axis intersect at P' after they are reflected; this verifies our earlier assertion. Such rays, nearly parallel to the axis and close to it, are called **paraxial rays.** (The term **paraxial approximation** is often used for the approximations we have just described.) Since all such reflected light rays converge on the image point, a concave mirror is also called a *converging mirror.*

Be sure you understand that Eq. (34.4), as well as many similar relationships that we will derive later in this chapter and the next, is only *approximately* correct. It results from a calculation containing approximations, and it is valid only for paraxial rays. If we increase the angle α that a ray makes with the optic axis, the point P' where the ray intersects the optic axis moves somewhat closer to the vertex than for a paraxial ray. As a result, a spherical mirror, unlike a plane mirror, does not form a precise point image of a point object; the image is "smeared out." This property of a spherical mirror is called **spherical aberration.** When the primary mirror of the Hubble Space Telescope (**Fig. 34.12a**) was manufactured, tiny errors were made in its shape that led to an unacceptable amount of spherical aberration (Fig. 34.12b). The performance of the telescope improved dramatically after the installation of corrective optics (Fig. 34.12c).

If the radius of curvature becomes infinite ($R = \infty$), the mirror becomes *plane,* and Eq. (34.4) reduces to Eq. (34.1) for a plane reflecting surface.

Focal Point and Focal Length

When the object point P is very far from the spherical mirror ($s = \infty$), the incoming rays are parallel. (The star shown in Fig. 34.12c is an example of such a distant object.) From Eq. (34.4) the image distance s' in this case is given by

$$\frac{1}{\infty} + \frac{1}{s'} = \frac{2}{R} \qquad s' = \frac{R}{2}$$

The situation is shown in **Fig. 34.13a**. The beam of incident parallel rays converges, after reflection from the mirror, to a point F at a distance $R/2$ from the vertex of the mirror. The point F at which the incident parallel rays converge is called the **focal point;** we say that these rays are brought to a focus. The distance from the vertex to the focal point, denoted by f, is called the **focal length.** We see that f is related to the radius of curvature R by

$$f = \frac{R}{2} \qquad \text{(focal length of a spherical mirror)} \tag{34.5}$$

Figure 34.13b shows the opposite situation. Now the *object* is placed at the focal point F, so the object distance is $s = f = R/2$. The image distance s' is again given by Eq. (34.4):

$$\frac{2}{R} + \frac{1}{s'} = \frac{2}{R} \qquad \frac{1}{s'} = 0 \qquad s' = \infty$$

With the object at the focal point, the reflected rays in Fig. 34.13b are parallel to the optic axis; they meet only at a point infinitely far from the mirror, so the image is at infinity.

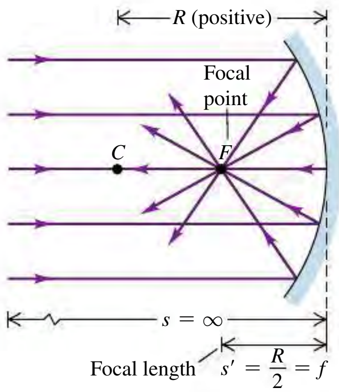
Thus the focal point F of a spherical mirror has the properties that (1) any incoming ray parallel to the optic axis is reflected through the focal point and (2) any incoming ray that passes through the focal point is reflected parallel to the optic axis. For spherical mirrors these statements are true only for paraxial rays. For parabolic mirrors these statements are *exactly* true. Spherical or parabolic mirrors are used in flashlights and headlights to form the light from the bulb into a parallel beam. Some solar-power plants use an array of plane mirrors to simulate an approximately spherical concave mirror; sunlight is collected by the mirrors and directed to the focal point, where a steam boiler is placed. (The concepts of focal point and focal length also apply to lenses, as we'll see in Section 34.4.)

We will usually express the relationship between object and image distances for a mirror, Eq. (34.4), in terms of the focal length f:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad \text{(object–image relationship, spherical mirror)} \tag{34.6}$$

34.13 The focal point and focal length of a concave mirror.

(a) All parallel rays incident on a spherical mirror reflect through the focal point.



(b) Rays diverging from the focal point reflect to form parallel outgoing rays.

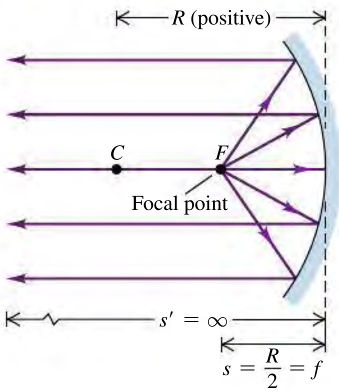
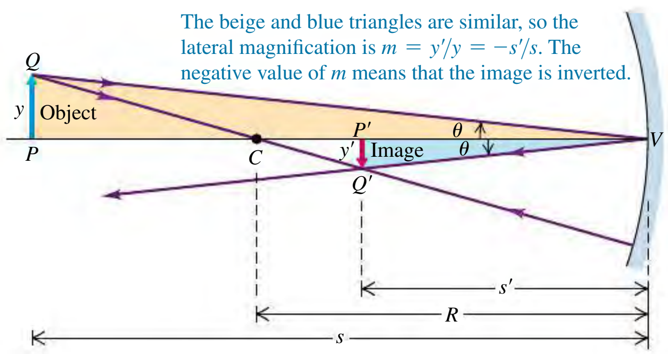


Image of an Extended Object: Spherical Mirror

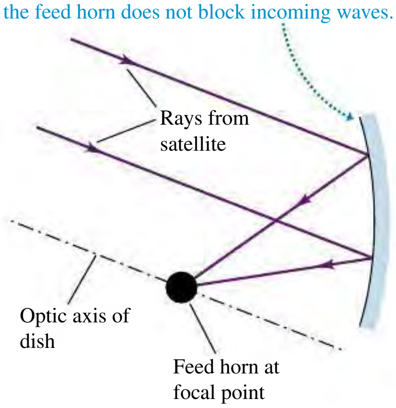
Now suppose we have an object with *finite* size, represented by the arrow PQ in **Fig. 34.14**, perpendicular to the optic axis CV. The image of P formed by paraxial rays is at P'. The object distance for point Q is very nearly equal to that for point P, so the image $P'Q'$ is nearly straight and perpendicular to the axis.



34.14 Construction for determining the position, orientation, and height of an image formed by a concave spherical mirror.

Application Satellite Television Dishes A dish antenna used to receive satellite TV broadcasts is actually a concave parabolic mirror. The waves are of much lower frequency than visible light (1.2 to 1.8×10^{10} Hz compared with 4.0 to 7.9×10^{14} Hz), but the laws of reflection are the same. The transmitter in orbit is so far away that the arriving waves have essentially parallel rays, as in Fig. 34.13a. The dish reflects the waves and brings them to a focus at a feed horn, from which they are "piped" to a decoder that extracts the signal.

Dish = segment of a curved mirror. Only a segment away from the optic axis is used so that the feed horn does not block incoming waves.



Note that the object and image arrows have different sizes, y and y', respectively, and that they have opposite orientation. In Eq. (34.2) we defined the *lateral magnification m* as the ratio of image size y' to object size y:

$$m = \frac{y'}{y}$$

Because triangles PVQ and $P'VQ'$ in Fig. 34.14 are *similar*, we also have the relationship $y/s = -y'/s'$. The negative sign is needed because object and image are on opposite sides of the optic axis; if y is positive, y' is negative. Therefore

$$m = \frac{y'}{y} = -\frac{s'}{s} \quad \text{(lateral magnification, spherical mirror)} \qquad (34.7)$$

If m is positive, the image is erect in comparison to the object; if m is negative, the image is *inverted* relative to the object, as in Fig. 34.14. For a *plane* mirror, $s = -s'$, so $y' = y$ and $m = +1$; since m is positive, the image is erect, and since $|m| = 1$, the image is the same size as the object.

CAUTION Lateral magnification can be less than 1 Although the ratio of image size to object size is called the *lateral magnification*, the image formed by a mirror or lens may be larger than, smaller than, or the same size as the object. If it is smaller, then the lateral magnification is less than unity in absolute value: $|m| < 1$. The image formed by an astronomical telescope mirror or a camera lens is usually *much* smaller than the object. For example, the image of the bright star shown in Fig. 34.12c is just a few millimeters across, while the star itself is hundreds of thousands of kilometers in diameter.

In our discussion of concave mirrors we have so far considered only objects that lie *outside* or at the focal point, so that the object distance s is greater than or equal to the (positive) focal length f. In this case the image point is on the same side of the mirror as the outgoing rays, and the image is real and inverted. If an object is *inside* the focal point of a concave mirror, so that $s < f$, the resulting image is *virtual* (that is, the image point is on the opposite side of the mirror from the object), *erect,* and *larger* than the object. Mirrors used when you apply makeup (referred to at the beginning of this section) are concave mirrors; in use, the distance from the face to the mirror is less than the focal length, and you see an enlarged, erect image. You can prove these statements about concave mirrors by applying Eqs. (34.6) and (34.7). We'll also be able to verify these results later in this section, after we've learned some graphical methods for relating the positions and sizes of the object and the image.

EXAMPLE 34.1 IMAGE FORMATION BY A CONCAVE MIRROR I

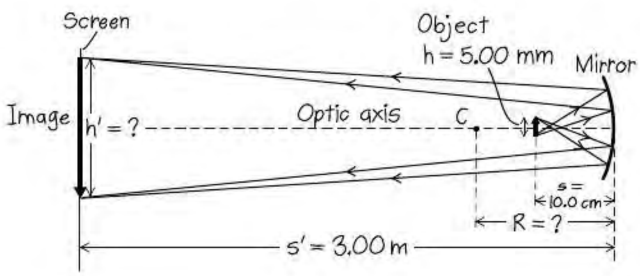
SOLUTION

A concave mirror forms an image, on a wall 3.00 m in front of the mirror, of a headlamp filament 10.0 cm in front of the mirror. (a) What are the radius of curvature and focal length of the mirror? (b) What is the lateral magnification? What is the image height if the object height is 5.00 mm?

SOLUTION

IDENTIFY and SET UP: Figure 34.15 shows our sketch. Our target variables are the radius of curvature R, focal length f, lateral magnification m, and image height y'. We are given the distances from the mirror to the object (s) and from the mirror to the image (s'). We solve Eq. (34.4) for R, and then use Eq. (34.5) to find f. Equation (34.7) yields both m and y'.

34.15 Our sketch for this problem.



EXECUTE: (a) Both the object and the image are on the concave (reflective) side of the mirror, so both s and s' are positive; we have $s = 10.0$ cm and $s' = 300$ cm. We solve Eq. (34.4) for R:

$$\frac{1}{10.0\text{ cm}} + \frac{1}{300\text{ cm}} = \frac{2}{R}$$

$$R = 2\left(\frac{1}{10.0\text{ cm}} + \frac{1}{300\text{ cm}}\right)^{-1} = 19.4\text{ cm}$$

The focal length of the mirror is $f = R/2 = 9.7$ cm.

(b) From Eq. (34.7) the lateral magnification is

$$m = -\frac{s'}{s} = -\frac{300\text{ cm}}{10.0\text{ cm}} = -30.0$$

Because m is negative, the image is inverted. The height of the image is 30.0 times the height of the object, or $(30.0)(5.00\text{ mm}) = 150$ mm.

EVALUATE: Our sketch indicates that the image is inverted; our calculations agree. Note that the object (at $s = 10.0$ cm) is just outside the focal point ($f = 9.7$ cm). This is very similar to what is done in automobile headlights. With the filament close to the focal point, the concave mirror produces a beam of nearly parallel rays.

CONCEPTUAL EXAMPLE 34.2 IMAGE FORMATION BY A CONCAVE MIRROR II

In Example 34.1, suppose that the lower half of the mirror's reflecting surface is covered with nonreflective soot. What effect will this have on the image of the filament?

SOLUTION

It would be natural to guess that the image would now show only half of the filament. But in fact the image will still show the *entire* filament. You can see why by examining Fig. 34.10b. Light rays coming from any object point P are reflected from *all* parts of the mirror and converge on the corresponding image point P'. If part of the mirror surface is made nonreflective (or is removed altogether), rays from the remaining reflective surface still form an image of every part of the object.

Reducing the reflecting area reduces the light energy reaching the image point, however: The image becomes *dimmer*. If the area is reduced by one-half, the image will be one-half as bright. Conversely, *increasing* the reflective area makes the image brighter. To make reasonably bright images of faint stars, astronomical telescopes use mirrors that are up to several meters in diameter (see Fig. 34.12a).

Convex Mirrors

In **Fig. 34.16a** the *convex* side of a spherical mirror faces the incident light. The center of curvature is on the side opposite to the outgoing rays; according to sign rule 3 in Section 34.1, R is negative (see Fig. 34.11). Ray PB is reflected, with the angles of incidence and reflection both equal to θ. The reflected ray, projected backward, intersects the axis at P'. As with a concave mirror, *all* rays from P that are reflected by the mirror diverge from the same point P', provided that the angle α is small. Therefore P' is the image of P. The object distance s is positive, the image distance s' is negative, and the radius of curvature R is *negative* for a *convex* mirror.

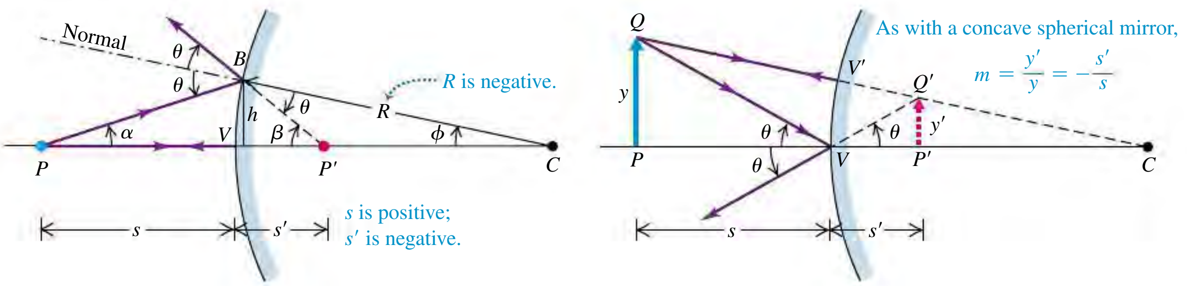
Figure 34.16b shows two rays diverging from the head of the arrow PQ and the virtual image $P'Q'$ of this arrow. The same procedure that we used for a concave mirror can be used to show that for a convex mirror, the expressions for the object–image relationship and the lateral magnification are

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} \quad \text{and} \quad m = \frac{y'}{y} = -\frac{s'}{s}$$

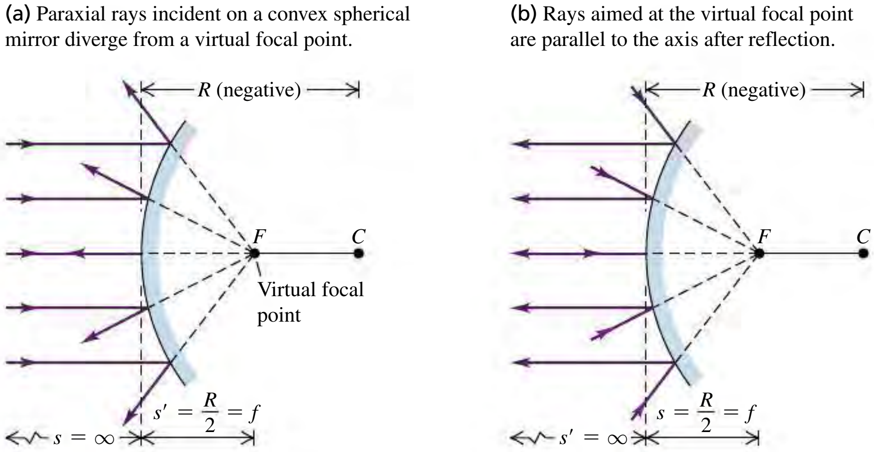
34.16 Image formation by a convex mirror.

(a) Construction for finding the position of an image formed by a convex mirror

(b) Construction for finding the magnification of an image formed by a convex mirror



34.17 The focal point and focal length of a convex mirror.



DATA *SPEAKS*

Image Formation by Mirrors

When students were given a problem involving image formation by mirrors, more than 59% gave an incorrect response. Common errors:

- Not using the law of reflection properly. For a mirror (plane or curved), the incident and reflected rays make the same angle with the normal to the mirror. The reflected ray originates at the point where the incident ray hits the mirror.
- Confusion about lateral magnification. The lateral magnification m depends on only the ratio of image distance s' to object distance s. If s' and s have different values but have the same ratio in two situations, then the value of m is the same.

These expressions are exactly the same as Eqs. (34.4) and (34.7) for a concave mirror. Thus when we use our sign rules consistently, Eqs. (34.4) and (34.7) are valid for both concave and convex mirrors.

When R is negative (convex mirror), incoming rays that are parallel to the optic axis are not reflected through the focal point F. Instead, they diverge as though they had come from the point F at a distance f *behind* the mirror, as shown in **Fig. 34.17a**. In this case, f is the focal length, and F is called a *virtual focal point*. The corresponding image distance s' is negative, so both f and R are negative, and Eq. (34.5), $f = R/2$, holds for convex as well as concave mirrors. In Fig. 34.17b the incoming rays are converging as though they would meet at the virtual focal point F, and they are reflected parallel to the optic axis.

In summary, Eqs. (34.4) through (34.7), the basic relationships for image formation by a spherical mirror, are valid for *both* concave and convex mirrors, provided that we use the sign rules consistently.

EXAMPLE 34.3 SANTA'S IMAGE PROBLEM

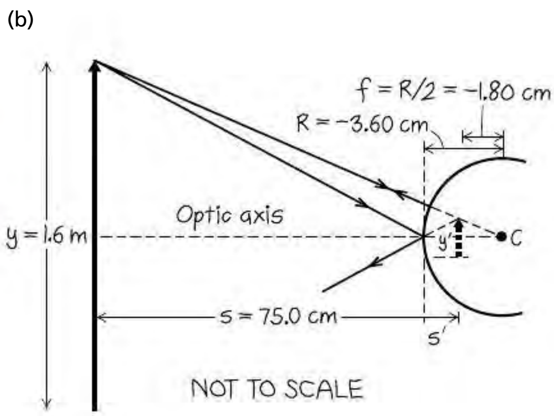
SOLUTION

Santa checks himself for soot, using his reflection in a silvered Christmas tree ornament 0.750 m away (**Fig. 34.18a**). The diameter of the ornament is 7.20 cm. Standard reference texts state that he is a "right jolly old elf," so we estimate his height to be 1.6 m. Where and how tall is the image of Santa formed by the ornament? Is it erect or inverted?

SOLUTION

IDENTIFY and SET UP: Figure 34.18b shows the situation. Santa is the object, and the surface of the ornament closest to him acts as a convex mirror. The relationships among object distance, image distance, focal length, and magnification are the same as

34.18 (a) The ornament forms a virtual, reduced, erect image of Santa. (b) Our sketch of two of the rays forming the image.



for concave mirrors, provided we use the sign rules consistently. The radius of curvature and the focal length of a convex mirror are *negative*. The object distance is $s = 0.750\text{ m} = 75.0\text{ cm}$, and Santa's height is $y = 1.6\text{ m}$. We solve Eq. (34.6) to find the image distance s', and then use Eq. (34.7) to find the lateral magnification m and the image height y'. The sign of m tells us whether the image is erect or inverted.

EXECUTE: The radius of the mirror (half the diameter) is $R = -(7.20\text{ cm})/2 = -3.60\text{ cm}$, and the focal length is $f = R/2 = -1.80\text{ cm}$. From Eq. (34.6),

$$\frac{1}{s'} = \frac{1}{f} - \frac{1}{s} = \frac{1}{-1.80\text{ cm}} - \frac{1}{75.0\text{ cm}}$$

$$s' = -1.76\text{ cm}$$

Because s' is negative, the image is behind the mirror—that is, on the side opposite to the outgoing light (Fig. 34.18b)—and it is virtual. The image is about halfway between the front surface of the ornament and its center.

From Eq. (34.7), the lateral magnification and the image height are

$$m = \frac{y'}{y} = -\frac{s'}{s} = -\frac{-1.76\text{ cm}}{75.0\text{ cm}} = 0.0234$$

$$y' = my = (0.0234)(1.6\text{ m}) = 3.8 \times 10^{-2}\text{ m} = 3.8\text{ cm}$$

EVALUATE: Our sketch indicates that the image is erect so both m and y' are positive; our calculations agree. When the object distance s is positive, a convex mirror *always* forms an erect, virtual, reduced, reversed image. For this reason, convex mirrors are used at blind intersections, for surveillance in stores, and as wide-angle rear-view mirrors for cars and trucks. (Many such mirrors read "Objects in mirror are closer than they appear.")

Graphical Methods for Mirrors

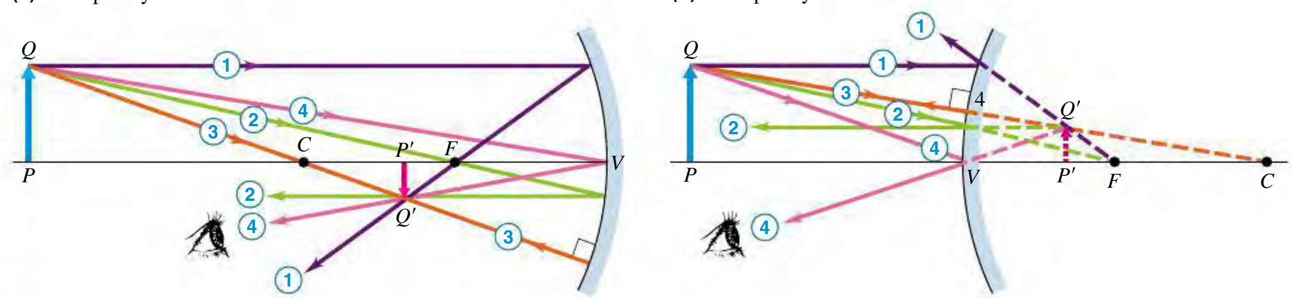
In Examples 34.1 and 34.3, we used Eqs. (34.6) and (34.7) to find the position and size of the image formed by a mirror. We can also determine the properties of the image by a simple *graphical* method. This method consists of finding the point of intersection of a few particular rays that diverge from a point of the object (such as point Q in **Fig. 34.19**) and are reflected by the mirror. Then (ignoring aberrations) *all* rays from this object point that strike the mirror will intersect at the same point. For this construction we always choose an object point that is *not* on the optic axis. Four rays that we can usually draw easily are shown in Fig. 34.19. These are called **principal rays.**

1. *A ray parallel to the axis*, after reflection, passes through the focal point F of a concave mirror or appears to come from the (virtual) focal point of a convex mirror.
2. *A ray through (or proceeding toward) the focal point F* is reflected parallel to the axis.
3. *A ray along the radius* through or away from the center of curvature C intersects the surface normally and is reflected back along its original path.
4. *A ray to the vertex V* is reflected forming equal angles with the optic axis.

34.19 The graphical method of locating an image formed by a spherical mirror. The colors of the rays are for identification only; they do not refer to specific colors of light.

(a) Principal rays for concave mirror

(b) Principal rays for convex mirror



(1) Ray parallel to axis reflects through focal point.

(2) Ray through focal point reflects parallel to axis.

(3) Ray through center of curvature intersects the surface normally and reflects along its original path.

(4) Ray to vertex reflects symmetrically around optic axis.

(1) Reflected parallel ray appears to come from focal point.

(2) Ray toward focal point reflects parallel to axis.

(3) As with concave mirror: Ray radial to center of curvature intersects the surface normally and reflects along its original path.

(4) As with concave mirror: Ray to vertex reflects symmetrically around optic axis.

Once we have found the position of the image point by means of the intersection of any two of these principal rays $(1, 2, 3, 4)$, we can draw the path of any other ray from the object point to the same image point.

CAUTION **Principal rays are not the only rays** Although we've emphasized the principal rays, in fact *any* ray from the object that strikes the mirror will pass through the image point (for a real image) or appear to originate from the image point (for a virtual image). Usually, you need to draw only the principal rays in order to locate the image.

PROBLEM-SOLVING STRATEGY 34.1 IMAGE FORMATION BY MIRRORS

IDENTIFY *the relevant concepts:* Problems involving image formation by mirrors can be solved in two ways: using principal-ray diagrams and using equations. A successful problem solution uses *both* approaches.

SET UP *the problem:* Identify the target variables. One of them is likely to be the focal length, the object distance, or the image distance, with the other two quantities given.

EXECUTE *the solution* as follows:

1. Draw a large, clear principal-ray diagram if you have enough information.
2. Orient your diagram so that incoming rays go from left to right. Draw only the principal rays; color-code them as in Fig. 34.19. If possible, use graph paper or quadrille-ruled paper. Use a ruler and measure distances carefully! A freehand sketch will *not* give good results.
3. If your principal rays don't converge at a real image point, you may have to extend them straight backward to locate a virtual image point, as in Fig. 34.19b. We recommend drawing the extensions with broken lines.
4. Measure the resulting diagram to obtain the magnitudes of the target variables.
5. Solve for the target variables by using Eq. (34.6), $1/s + 1/s' = 1/f$, and the lateral magnification equation, Eq. (34.7), as appropriate. Apply the sign rules given in Section 34.1 to object and image distances, radii of curvature, and object and image heights.
6. Use the sign rules to interpret the results that you deduced from your ray diagram and calculations. Note that the *same* sign rules (given in Section 34.1) work for all four cases in this chapter: reflection and refraction from plane and spherical surfaces.

EVALUATE *your answer:* Check that the results of your calculations agree with your ray-diagram results for image position, image size, and whether the image is real or virtual.

EXAMPLE 34.4 CONCAVE MIRROR WITH VARIOUS OBJECT DISTANCES

SOLUTION

A concave mirror has a radius of curvature with absolute value 20 cm. Find graphically the image of an object in the form of an arrow perpendicular to the axis of the mirror at object distances of (a) 30 cm, (b) 20 cm, (c) 10 cm, and (d) 5 cm. Check the construction by *computing* the size and lateral magnification of each image.

SOLUTION

IDENTIFY and SET UP: We must use graphical methods *and* calculations to analyze the image made by a mirror. The mirror is concave, so its radius of curvature is $R = +20$ cm and its focal length is $f = R/2 = +10$ cm. Our target variables are the image distances s' and lateral magnifications m corresponding to four cases with successively smaller object distances s. In each case we solve Eq. (34.6) for s' and use $m = -s'/s$ to find m.

EXECUTE: **Figure 34.20** shows the principal-ray diagrams for the four cases. Study each of these diagrams carefully and confirm that each numbered ray is drawn in accordance with the rules given earlier (under "Graphical Methods for Mirrors"). Several points are worth noting. First, in case (b) the object and image distances are equal. Ray 3 cannot be drawn in this case because a ray from Q through the center of curvature C does not strike the mirror. In case (c), ray 2 cannot be drawn because a ray from Q through F does not strike the mirror. In this case the outgoing rays are parallel, corresponding to an infinite image distance. In case (d), the outgoing rays diverge; they have been extended backward to the *virtual image point* Q', from which they appear to diverge. Case (d) illustrates the general observation that an object placed inside the focal point of a concave mirror produces a virtual image.

Measurements of the figures, with appropriate scaling, give the following approximate image distances: (a) 15 cm; (b) 20 cm; (c) ∞ or $-\infty$ (because the outgoing rays are parallel and do not converge at any finite distance); (d) -10 cm. To *compute* these distances, we solve Eq. (34.6) for s' and insert $f = 10$ cm:

$$\text{(a)}\ \frac{1}{30\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = 15\text{ cm}$$

$$\text{(b)}\ \frac{1}{20\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = 20\text{ cm}$$

$$\text{(c)}\ \frac{1}{10\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = \infty\ (\text{or } -\infty)$$

$$\text{(d)}\ \frac{1}{5\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = -10\text{ cm}$$

The signs of s' tell us that the image is real in cases (a) and (b) and virtual in case (d).

The lateral magnifications measured from the figures are approximately (a) $-\frac{1}{2}$; (b) -1; (c) ∞ or $-\infty$; (d) $+2$. From Eq. (34.7),

$$\text{(a) } m = -\frac{15\text{ cm}}{30\text{ cm}} = -\frac{1}{2}$$

$$\text{(b) } m = -\frac{20\text{ cm}}{20\text{ cm}} = -1$$

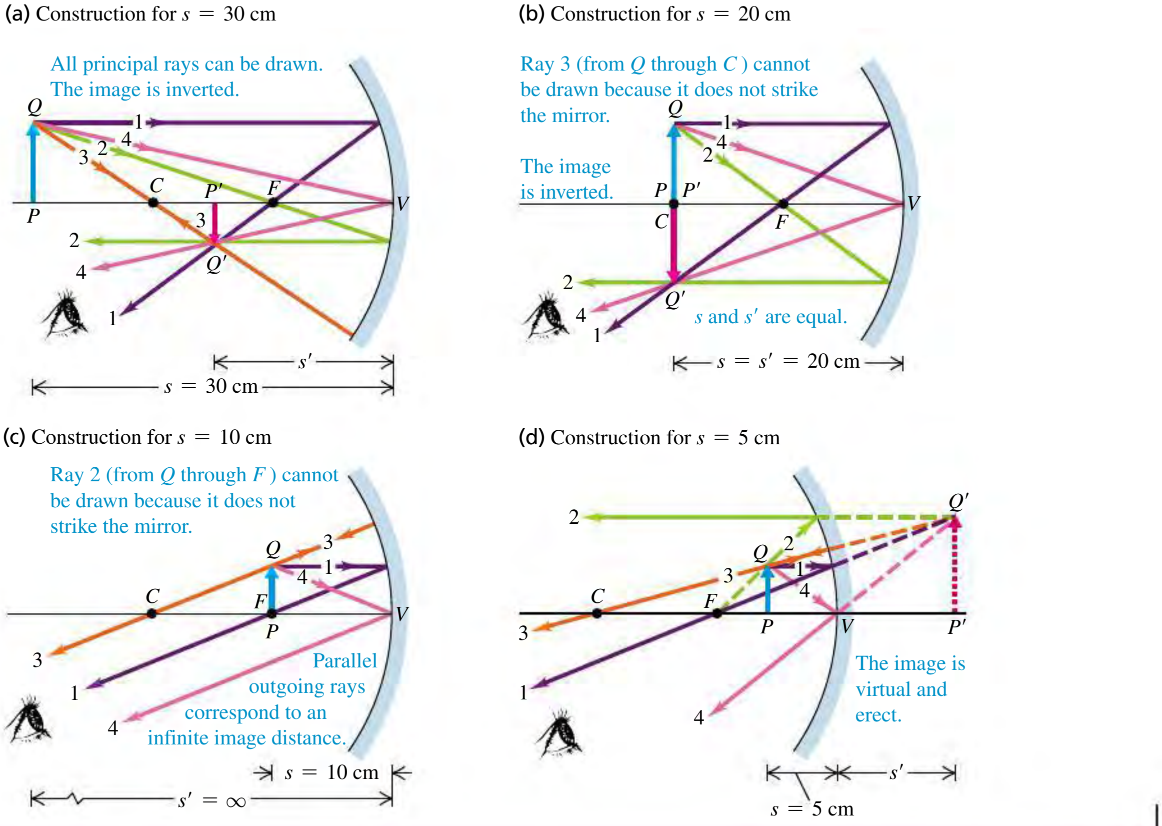
$$\text{(c) } m = -\frac{\infty\text{ cm}}{10\text{ cm}} = -\infty \text{ (or } +\infty\text{)}$$

$$\text{(d) } m = -\frac{-10\text{ cm}}{5\text{ cm}} = +2$$

The signs of m tell us that the image is inverted in cases (a) and (b) and erect in case (d).

EVALUATE: Notice the trend of the results in the four cases. When the object is far from the mirror, as in Fig. 34.20a, the image is smaller than the object, inverted, and real. As the object distance s decreases, the image moves farther from the mirror and gets larger (Fig. 34.20b). When the object is at the focal point, the image is at infinity (Fig. 34.20c). When the object is inside the focal point, the image becomes larger than the object, erect, and virtual (Fig. 34.20d). You can confirm these conclusions by looking at objects reflected in the concave bowl of a shiny metal spoon.

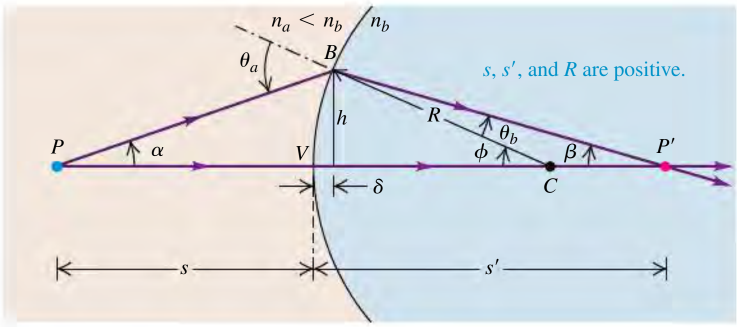
34.20 Using principal-ray diagrams to locate the image $P'Q'$ made by a concave mirror.



TEST YOUR UNDERSTANDING OF SECTION 34.2 A cosmetics mirror is designed so that your reflection appears right-side up and enlarged. (a) Is the mirror concave or convex? (b) To see an enlarged image, what should be the distance from the mirror (of focal length f) to your face? (i) $|f|$; (ii) less than $|f|$; (iii) greater than $|f|$.

34.3 REFRACTION AT A SPHERICAL SURFACE

As we mentioned in Section 34.1, images can be formed by refraction as well as by reflection. To begin with, let's consider refraction at a spherical surface—that is, at a spherical interface between two optical materials with different indexes of refraction. This analysis is directly applicable to some real optical systems, such as the human eye. It also provides a stepping-stone for the analysis of lenses, which usually have *two* spherical (or nearly spherical) surfaces.



34.21 Construction for finding the position of the image point P' of a point object P formed by refraction at a spherical surface. The materials to the left and right of the interface have indexes of refraction n_a and n_b, respectively. In the case shown here, $n_a < n_b$.

Image of a Point Object: Spherical Refracting Surface

In **Fig. 34.21** a spherical surface with radius R forms an interface between two materials with different indexes of refraction n_a and n_b. The surface forms an image P' of an object point P; we want to find how the object and image distances (s and s') are related. We will use the same sign rules that we used for spherical mirrors. The center of curvature C is on the outgoing side of the surface, so R is positive. Ray PV strikes the vertex V and is perpendicular to the surface (that is, to the plane that is tangent to the surface at the point of incidence V). It passes into the second material without deviation. Ray PB, making an angle α with the axis, is incident at an angle θ_a with the normal and is refracted at an angle θ_b. These rays intersect at P', a distance s' to the right of the vertex. The figure is drawn for the case $n_a < n_b$. Both the object and image distances are positive.

We are going to prove that if the angle α is small, *all* rays from P intersect at the same point P', so P' is the *real image* of P. We use much the same approach as we did for spherical mirrors in Section 34.2. We again use the theorem that an exterior angle of a triangle equals the sum of the two opposite interior angles; applying this to the triangles PBC and $P'BC$ gives

$$\theta_a = \alpha + \phi \qquad \phi = \beta + \theta_b \tag{34.8}$$

From the law of refraction,

$$n_a \sin\theta_a = n_b \sin\theta_b$$

Also, the tangents of α, β, and ϕ are

$$\tan\alpha = \frac{h}{s+\delta} \qquad \tan\beta = \frac{h}{s'-\delta} \qquad \tan\phi = \frac{h}{R-\delta} \tag{34.9}$$

For paraxial rays, θ_a and θ_b are both small in comparison to a radian, and we may approximate both the sine and tangent of either of these angles by the angle itself (measured in radians). The law of refraction then gives

$$n_a\theta_a = n_b\theta_b$$

Combining this with the first of Eqs. (34.8), we obtain

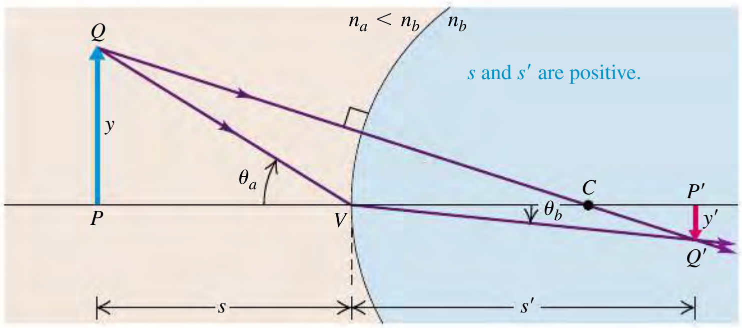
$$\theta_b = \frac{n_a}{n_b}(\alpha + \phi)$$

When we substitute this into the second of Eqs. (34.8), we get

$$n_a\alpha + n_b\beta = (n_b - n_a)\phi \tag{34.10}$$

Now we use the approximations $\tan\alpha = \alpha$, and so on, in Eqs. (34.9) and also ignore the small distance δ; those equations then become

$$\alpha = \frac{h}{s} \qquad \beta = \frac{h}{s'} \qquad \phi = \frac{h}{R}$$



34.22 Construction for determining the height of an image formed by refraction at a spherical surface. In the case shown here, $n_a < n_b$.

Finally, we substitute these into Eq. (34.10) and divide out the common factor h. We obtain

$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R} \qquad \text{(object–image relationship, spherical refracting surface)} \tag{34.11}$$

This equation does not contain the angle α, so the image distance is the same for *all* paraxial rays emanating from P; this proves that P' is the image of P.

To obtain the lateral magnification m for this situation, we use the construction in **Fig. 34.22**. We draw two rays from point Q, one through the center of curvature C and the other incident at the vertex V. From the triangles PQV and $P'Q'V$,

$$\tan\theta_a = \frac{y}{s} \qquad \tan\theta_b = \frac{-y'}{s'}$$

and from the law of refraction,

$$n_a \sin\theta_a = n_b \sin\theta_b$$

For small angles,

$$\tan\theta_a = \sin\theta_a \qquad \tan\theta_b = \sin\theta_b$$

so finally

$$\frac{n_a y}{s} = -\frac{n_b y'}{s'} \qquad \text{or}$$

$$m = \frac{y'}{y} = -\frac{n_a s'}{n_b s} \qquad \text{(lateral magnification, spherical refracting surface)} \tag{34.12}$$

Equations (34.11) and (34.12) apply to both convex and concave refracting surfaces, provided that you use the sign rules consistently. It doesn't matter whether n_b is greater or less than n_a. To verify these statements, construct diagrams like Figs. 34.21 and 34.22 for the following three cases: (i) $R > 0$ and $n_a > n_b$, (ii) $R < 0$ and $n_a < n_b$, and (iii) $R < 0$ and $n_a > n_b$. Then in each case, use your diagram to again derive Eqs. (34.11) and (34.12).

Here's a final note on the sign rule for the radius of curvature R of a surface. For the convex reflecting surface in Fig. 34.16, we considered R negative, but the convex *refracting* surface in Fig. 34.21 has a *positive* value of R. This may seem inconsistent, but it isn't. The rule is that R is positive if the center of curvature C is on the outgoing side of the surface and negative if C is on the other side. For the convex reflecting surface in Fig. 34.16, R is negative because point C is to the right of the surface but outgoing rays are to the left. For the convex refracting surface in Fig. 34.21, R is positive because both C and the outgoing rays are to the right of the surface.

Refraction at a curved surface is one reason gardeners avoid watering plants at midday. As sunlight enters a water drop resting on a leaf (**Fig. 34.23**), the light rays are refracted toward each other as in Figs. 34.21 and 34.22. The sunlight that strikes the leaf is therefore more concentrated and able to cause damage.

34.23 Light rays refract as they pass through the curved surfaces of these water droplets.

An important special case of a spherical refracting surface is a *plane* surface between two optical materials. This corresponds to setting $R = \infty$ in Eq. (34.11). In this case,

$$\frac{n_a}{s} + \frac{n_b}{s'} = 0 \qquad \text{(plane refracting surface)} \tag{34.13}$$

To find the lateral magnification m for this case, we combine this equation with the general relationship, Eq. (34.12), obtaining the simple result

$$m = 1$$

That is, the image formed by a *plane* refracting surface always has the same lateral size as the object, and it is always erect.

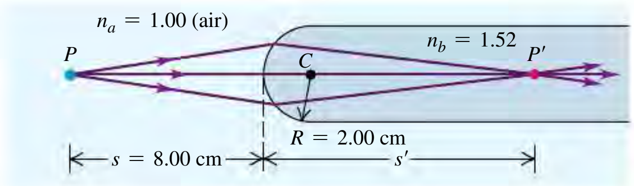
An example of image formation by a plane refracting surface is the appearance of a partly submerged drinking straw or canoe paddle. When viewed from some angles, the submerged part appears to be only about three-quarters of its actual distance below the surface. (We commented on the appearance of a submerged object in Section 33.2; see Fig. 33.9.)

EXAMPLE 34.5 IMAGE FORMATION BY REFRACTION I

SOLUTION

A cylindrical glass rod (**Fig. 34.24**) has index of refraction 1.52. It is surrounded by air. One end is ground to a hemispherical surface with radius $R = 2.00$ cm. A small object is placed on the axis of the rod, 8.00 cm to the left of the vertex. Find (a) the image distance and (b) the lateral magnification.

34.24 The glass rod in air forms a real image.



SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of refraction at a curved surface. Our target variables are the image distance s' and the lateral magnification m. Here material a is air $(n_a = 1.00)$ and material b is the glass of which the rod is made $(n_b = 1.52)$. We are given $s = 8.00\text{ cm}$. The center of curvature of the spherical surface is on the outgoing side of the surface, so the radius is positive: $R = +2.00\text{ cm}$. We solve Eq. (34.11) for s', and we use Eq. (34.12) to find m.

EXECUTE: (a) From Eq. (34.11),

$$\frac{1.00}{8.00\text{ cm}} + \frac{1.52}{s'} = \frac{1.52 - 1.00}{+2.00\text{ cm}}$$

$$s' = +11.3\text{ cm}$$

(b) From Eq. (34.12),

$$m = -\frac{n_a s'}{n_b s} = -\frac{(1.00)(11.3\text{ cm})}{(1.52)(8.00\text{ cm})} = -0.929$$

EVALUATE: Because the image distance s' is positive, the image is formed 11.3 cm to the *right* of the vertex (on the outgoing side), as Fig. 34.24 shows. The value of m tells us that the image is somewhat smaller than the object and that it is inverted. If the object is an arrow 1.000 mm high, pointing upward, the image is an arrow 0.929 mm high, pointing downward.

EXAMPLE 34.6 IMAGE FORMATION BY REFRACTION II

SOLUTION

The glass rod of Example 34.5 is immersed in water, which has index of refraction $n = 1.33$ (**Fig. 34.25**). The object distance is again 8.00 cm. Find the image distance and lateral magnification.

SOLUTION

IDENTIFY and SET UP: The situation is the same as in Example 34.5 except that now $n_a = 1.33$. We again use Eqs. (34.11) and (34.12) to determine s' and m, respectively.

EXECUTE: Our solution of Eq. (34.11) in Example 34.5 yields

$$\frac{1.33}{8.00\text{ cm}} + \frac{1.52}{s'} = \frac{1.52 - 1.33}{+2.00\text{ cm}}$$

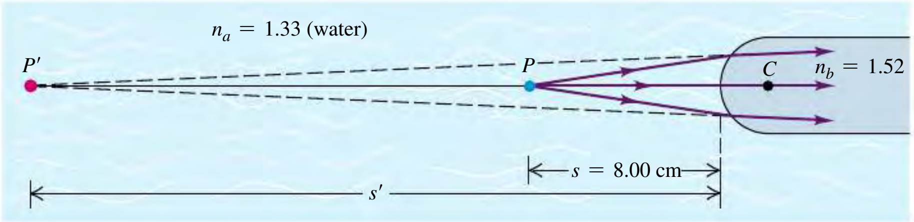
$$s' = -21.3\text{ cm}$$

The lateral magnification in this case is

$$m = \frac{n_a s'}{n_b s} = -\frac{(1.33)(-21.3\text{ cm})}{(1.52)(8.00\text{ cm})} = +2.33$$

EVALUATE: The negative value of s' means that the refracted rays do not converge, but appear to diverge from a point 21.3 cm to the *left* of the vertex. We saw a similar case in the reflection of light from a convex mirror; in both cases we call the result a *virtual image*. The vertical image is erect (because m is positive) and 2.33 times as large as the object.

34.25 When immersed in water, the glass rod forms a virtual image.



EXAMPLE 34.7 APPARENT DEPTH OF A SWIMMING POOL

SOLUTION

If you look straight down into a swimming pool where it is 2.00 m deep, how deep does it appear to be?

SOLUTION

IDENTIFY and SET UP: Figure 34.26 shows the situation. The surface of the water acts as a plane refracting surface. To determine the pool's apparent depth, we imagine an arrow PQ painted on the bottom. The pool's refracting surface forms a virtual image $P'Q'$ of this arrow. We solve Eq. (34.13) to find the image depth s'; that's the pool's apparent depth.

EXECUTE: The object distance is the actual depth of the pool, $s = 2.00$ m. Material a is water $(n_a = 1.33)$ and material b is air $(n_b = 1.00)$. From Eq. (34.13),

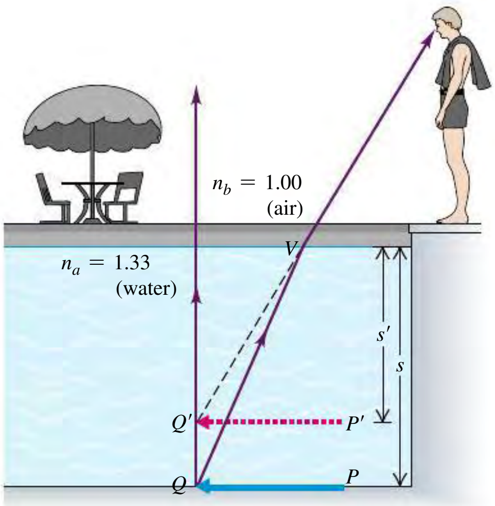
$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{1.33}{2.00\text{ m}} + \frac{1.00}{s'} = 0$$

$$s' = -1.50\text{ m}$$

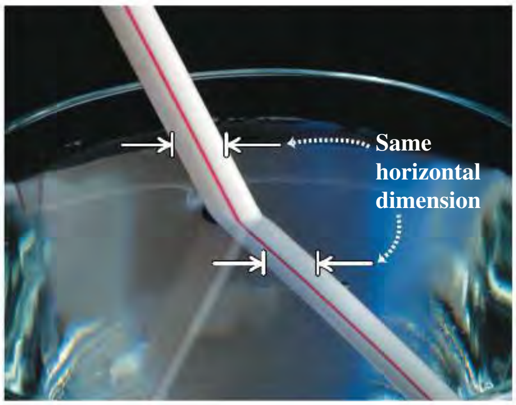
The image distance is negative. By the sign rules in Section 34.1, this means that the image is virtual and on the incoming side of the refracting surface—that is, on the same side as the object, namely underwater. The pool's apparent depth is 1.50 m, or just 75% of its true depth.

EVALUATE: Recall that the lateral magnification for a plane refracting surface is $m = 1$. Hence the image $P'Q'$ of the arrow has the same *horizontal length* as the actual arrow PQ (**Fig. 34.27**). Only its depth is different from that of PQ.

34.26 Arrow $P'Q'$ is the virtual image of underwater arrow PQ. The angles of the ray with the vertical are exaggerated for clarity.



34.27 The submerged portion of this straw appears to be at a shallower depth (closer to the surface) than it actually is.



TEST YOUR UNDERSTANDING OF SECTION 34.3 The water droplets in Fig. 34.23 have radius of curvature R and index of refraction $n = 1.33$. Can they form an image of the sun on the leaf?

34.4 THIN LENSES

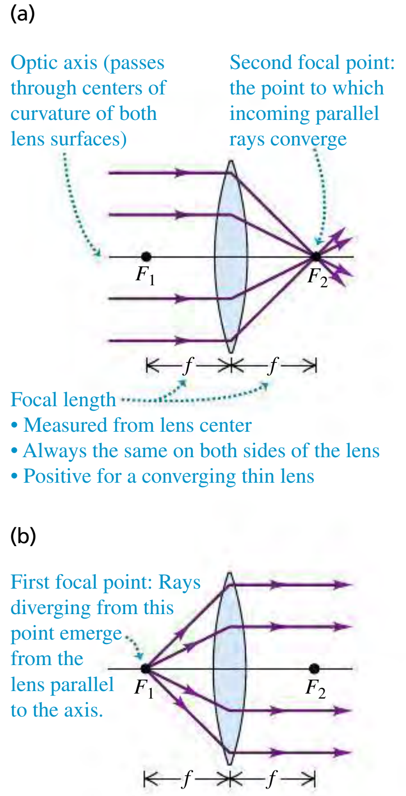
MP

PhET: Geometric Optics

The most familiar and widely used optical device (after the plane mirror) is the *lens.* A lens is an optical system with two refracting surfaces. The simplest lens has two *spherical* surfaces close enough together that we can ignore the distance between them (the thickness of the lens); we call this a **thin lens.** If you wear eyeglasses or contact lenses while reading, you are viewing these words through a pair of thin lenses. Later in this section we'll analyze thin lenses in detail by using the results of Section 34.3 for refraction by a single spherical surface. However, let's first discuss the properties of thin lenses.

Properties of a Lens

34.28 F_1 and F_2 are the first and second focal points of a converging thin lens. The numerical value of f is positive.



A lens of the shape shown in **Fig. 34.28** has an important property: When a beam of rays parallel to the axis passes through the lens, the rays converge to a point F_2 (Fig. 34.28a) and form a real image at that point. Such a lens is called a **converging lens.** Similarly, rays passing through point F_1 emerge from the lens as a beam of parallel rays (Fig. 34.28b). The points F_1 and F_2 are called the first and second *focal points,* and the distance f (measured from the center of the lens) is called the *focal length.* Note the similarities between the two focal points of a converging lens and the single focal point of a concave mirror (see Fig. 34.13). As for a concave mirror, the focal length of a converging lens is defined to be a *positive* quantity, and such a lens is also called a *positive lens.*

The central horizontal line in Fig. 34.28 is called the *optic axis,* as with spherical mirrors. The centers of curvature of the two spherical surfaces lie on and define the optic axis. The two focal lengths in Fig. 34.28, both labeled f, *are always equal* for a thin lens, even when the two sides have different curvatures. We will show this result later in the section, when we derive the relationship of f to the index of refraction of the lens and the radii of curvature of its surfaces.

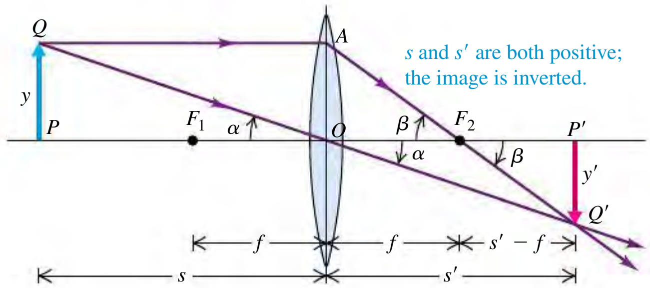
Image of an Extended Object: Converging Lens

Like a concave mirror, a converging lens can form an image of an extended object. **Figure 34.29** shows how to find the position and lateral magnification of an image made by a thin converging lens. Using the same notation and sign rules as before, we let s and s' be the object and image distances, respectively, and let y and y' be the object and image heights. Ray QA, parallel to the optic axis before refraction, passes through the second focal point F_2 after refraction. Ray QOQ' passes undeflected straight through the center of the lens because at the center the two surfaces are parallel and (we have assumed) very close together. There is refraction where the ray enters and leaves the material but no net change in direction.

The two angles labeled α in Fig. 34.29 are equal, so the two right triangles PQO and $P'Q'O$ are *similar* and ratios of corresponding sides are equal. Thus

$$\frac{y}{s} = -\frac{y'}{s'} \quad \text{or} \quad \frac{y'}{y} = -\frac{s'}{s} \tag{34.14}$$

34.29 Construction used to find image position for a thin lens. To emphasize that the lens is assumed to be very thin, the ray QAQ' is shown as bent at the midplane of the lens rather than at the two surfaces and ray QOQ' is shown as a straight line.



(The reason for the negative sign is that the image is below the optic axis and y' is negative.) Also, the two angles labeled β are equal, and the two right triangles OAF_2 and $P'Q'F_2$ are similar, so

$$\frac{y}{f} = -\frac{y'}{s' - f} \quad \text{or}$$

$$\frac{y'}{y} = -\frac{s' - f}{f} \tag{34.15}$$

We now equate Eqs. (34.14) and (34.15), divide by s', and rearrange to obtain

Object–image relationship, thin lens:

$$\underbrace{\frac{1}{s}}_{\text{Object distance}} + \underbrace{\frac{1}{s'}}_{\text{Image distance}} = \underbrace{\frac{1}{f}}_{\text{Focal length of lens}} \tag{34.16}$$

Equation (34.14) also gives the lateral magnification $m = y'/y$ for the lens:

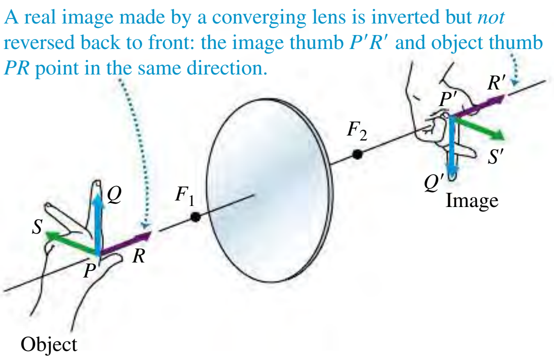
$$m = -\frac{s'}{s} \quad \text{(lateral magnification, thin lens)} \tag{34.17}$$

The negative sign tells us that when s and s' are both positive, as in Fig. 34.29, the image is *inverted,* and y and y' have opposite signs.

Equations (34.16) and (34.17) are the basic equations for thin lenses. They are *exactly* the same as the corresponding equations for spherical mirrors, Eqs. (34.6) and (34.7). As we will see, the same sign rules that we used for spherical mirrors are also applicable to lenses. In particular, consider a lens with a positive focal length (a converging lens). When an object is outside the first focal point F_1 of this lens (that is, when $s > f$), the image distance s' is positive (that is, the image is on the same side as the outgoing rays); this image is real and inverted, as in Fig. 34.29. An object placed inside the first focal point of a converging lens, so that $s < f$, produces an image with a negative value of s'; this image is located on the same side of the lens as the object and is virtual, erect, and larger than the object. You can verify these statements algebraically by using Eqs. (34.16) and (34.17); we'll also verify them in the next section, using graphical methods analogous to those introduced for mirrors in Section 34.2.

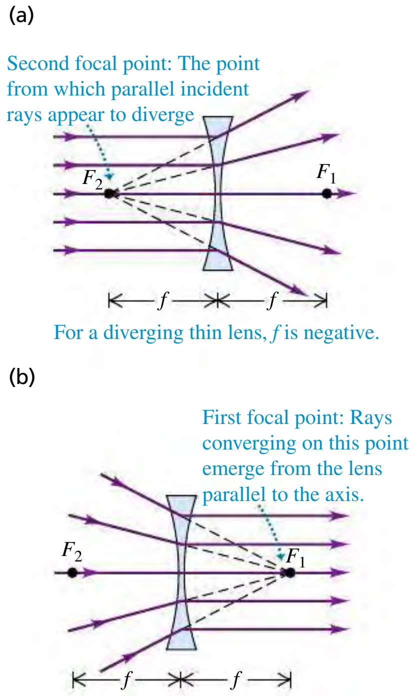
Figure 34.30 shows how a lens forms a three-dimensional image of a three-dimensional object. Point R is nearer the lens than point P. From Eq. (34.16), image point R' is farther from the lens than is image point P', and the image $P'R'$ points in the same direction as the object PR. Arrows $P'S'$ and $P'Q'$ are reversed relative to PS and PQ.

Let's compare Fig. 34.30 with Fig. 34.7, which shows the image formed by a plane *mirror.* We note that the image formed by the lens is inverted, but it is *not* reversed front to back along the optic axis. That is, if the object is a left hand, its image is also a left hand. You can verify this by pointing your left thumb along PR,

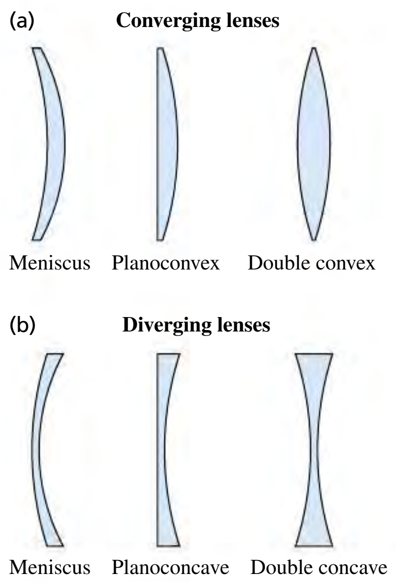


34.30 The image $S'P'Q'R'$ of a three-dimensional object $SPQR$ is not reversed by a lens.

34.31 F_2 and F_1 are the second and first focal points of a diverging thin lens, respectively. The numerical value of f is negative.



34.32 Various types of lenses.



your left forefinger along PQ, and your left middle finger along PS. Then rotate your hand 180°, using your thumb as an axis; this brings the fingers into coincidence with $P'Q'$ and $P'S'$. In other words, an *inverted* image is equivalent to an image that has been rotated by 180° about the lens axis.

Diverging Lenses

So far we have been discussing *converging* lenses. **Figure 34.31** shows a **diverging lens;** the beam of parallel rays incident on this lens *diverges* after refraction. The focal length of a diverging lens is a negative quantity, and the lens is also called a *negative lens.* The focal points of a negative lens are reversed, relative to those of a positive lens. The second focal point, F_2, of a negative lens is the point from which rays that are originally parallel to the axis *appear to diverge* after refraction, as in Fig. 34.31a. Incident rays converging toward the first focal point F_1, as in Fig. 34.31b, emerge from the lens parallel to its axis. Comparing with Section 34.2, you can see that a diverging lens has the same relationship to a converging lens as a convex mirror has to a concave mirror.

Equations (34.16) and (34.17) apply to *both* positive and negative lenses. **Figure 34.32** shows various types of lenses, both converging and diverging. Here's an important observation: *Any lens that is thicker at its center than at its edges is a converging lens with positive f; and any lens that is thicker at its edges than at its center is a diverging lens with negative f* (provided that the lens has a greater index of refraction than the surrounding material). We can prove this by using the *lensmaker's equation,* which it is our next task to derive.

The Lensmaker's Equation

We'll now derive Eq. (34.16) in more detail and at the same time derive the *lensmaker's equation,* which is a relationship among the focal length f, the index of refraction n of the lens, and the radii of curvature R_1 and R_2 of the lens surfaces. We use the principle that an image formed by one reflecting or refracting surface can serve as the object for a second reflecting or refracting surface.

We begin with the somewhat more general problem of two spherical interfaces separating three materials with indexes of refraction n_a, n_b, and n_c, as shown in **Fig. 34.33**. The object and image distances for the first surface are s_1 and s'_1, and those for the second surface are s_2 and s'_2. We assume that the lens is thin, so that the distance t between the two surfaces is small in comparison with the object and image distances and can therefore be ignored. This is usually the case with eyeglass lenses (**Fig. 34.34**). Then s_2 and s'_1 have the same magnitude but opposite sign. For example, if the first image is on the outgoing side of the first surface, s'_1 is positive. But when viewed as an object for the second surface, the first image is *not* on the incoming side of that surface. So we can say that $s_2 = -s'_1$.

We need to use the single-surface equation, Eq. (34.11), twice, once for each surface. The two resulting equations are

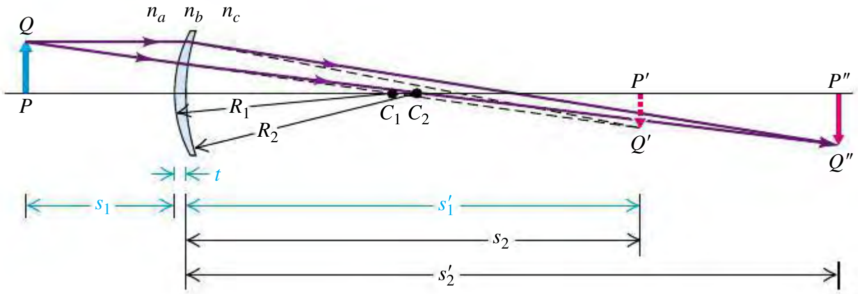
$$\frac{n_a}{s_1} + \frac{n_b}{s'_1} = \frac{n_b - n_a}{R_1}$$

$$\frac{n_b}{s_2} + \frac{n_c}{s'_2} = \frac{n_c - n_b}{R_2}$$

Ordinarily, the first and third materials are air or vacuum, so we set $n_a = n_c = 1$. The second index n_b is that of the lens, which we can call simply n. Substituting these values and the relationship $s_2 = -s'_1$, we get

$$\frac{1}{s_1} + \frac{n}{s'_1} = \frac{n - 1}{R_1}$$

$$-\frac{n}{s'_1} + \frac{1}{s'_2} = \frac{1 - n}{R_2}$$



34.33 The image formed by the first surface of a lens serves as the object for the second surface. The distances s'_1 and s_2 are taken to be equal; this is a good approximation if the lens thickness t is small.

To get a relationship between the initial object position s_1 and the final image position s'_2, we add these two equations. This eliminates the term n/s'_1:

$$\frac{1}{s_1} + \frac{1}{s'_2} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

Finally, thinking of the lens as a single unit, we rename the object distance simply s instead of s_1, and we rename the final image distance s' instead of s'_2:

$$\frac{1}{s} + \frac{1}{s'} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{34.18}$$

Now we compare this with the other thin-lens equation, Eq. (34.16). We see that the object and image distances s and s' appear in exactly the same places in both equations and that the focal length f is given by the **lensmaker's equation:**

Lensmaker's equation for a thin lens:

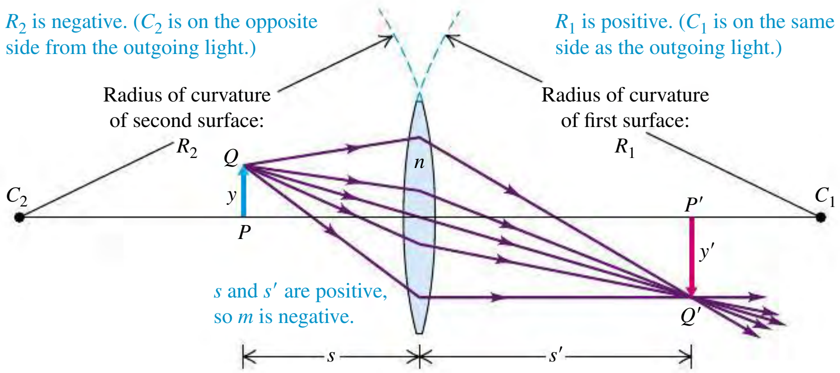
$$\underbrace{\frac{1}{f}}_{\text{Focal length}} = (\overbrace{n}^{\text{Index of refraction of lens material}} - 1)\left(\underbrace{\frac{1}{R_1}}_{\text{Radius of curvature of first surface}} - \underbrace{\frac{1}{R_2}}_{\text{Radius of curvature of second surface}}\right) \tag{34.19}$$

In the process of rederiving the relationship among object distance, image distance, and focal length for a thin lens, we have also derived Eq. (34.19), an expression for the focal length f of a lens in terms of its index of refraction n and the radii of curvature R_1 and R_2 of its surfaces. This can be used to show that all the lenses in Fig. 34.32a are converging lenses with $f > 0$ and that all the lenses in Fig. 34.32b are diverging lenses with $f < 0$.

We use all our sign rules from Section 34.1 with Eqs. (34.18) and (34.19). For example, in **Fig. 34.35**, s, s', and R_1 are positive, but R_2 is negative.

It is not hard to generalize Eq. (34.19) to the situation in which the lens is immersed in a material with an index of refraction greater than unity. We invite you to work out the lensmaker's equation for this more general situation.

34.34 These eyeglass lenses satisfy the thin-lens approximation: Their thickness is small compared to the object and image distances.



34.35 A converging thin lens with a positive focal length f.

We stress that the paraxial approximation is indeed an approximation! Rays that are at sufficiently large angles to the optic axis of a spherical lens will not be brought to the same focus as paraxial rays; this is the same problem of spherical aberration that plagues spherical *mirrors* (see Section 34.2). To avoid this and other limitations of thin spherical lenses, lenses of more complicated shape are used in precision optical instruments.

SOLUTION

EXAMPLE 34.8 DETERMINING THE FOCAL LENGTH OF A LENS

(a) Suppose the absolute values of the radii of curvature of the lens surfaces in Fig. 34.35 are both equal to 10 cm and the index of refraction of the glass is $n = 1.52$. What is the focal length f of the lens? (b) Suppose the lens in Fig. 34.31 also has $n = 1.52$ and the absolute values of the radii of curvature of its lens surfaces are also both equal to 10 cm. What is the focal length of this lens?

SOLUTION

IDENTIFY and SET UP: We are asked to find the focal length f of (a) a lens that is convex on both sides (Fig. 34.35) and (b) a lens that is concave on both sides (Fig. 34.31). In both cases we solve the lensmaker's equation, Eq. (34.19), to determine f. We apply the sign rules given in Section 34.1 to the radii of curvature R_1 and R_2 to take account of whether the surfaces are convex or concave.

EXECUTE: (a) The lens in Fig. 34.35 is *double convex:* The center of curvature of the first surface (C_1) is on the outgoing side of the lens, so R_1 is positive, and the center of curvature of the second surface (C_2) is on the *incoming* side, so R_2 is negative. Hence $R_1 = +10$ cm and $R_2 = -10$ cm. Then from Eq. (34.19),

$$\frac{1}{f} = (1.52 - 1)\left(\frac{1}{+10\text{ cm}} - \frac{1}{-10\text{ cm}}\right)$$

$$f = 9.6\text{ cm}$$

(b) The lens in Fig. 34.31 is *double concave:* The center of curvature of the first surface is on the *incoming* side, so R_1 is negative, and the center of curvature of the second surface is on the outgoing side, so R_2 is positive. Hence in this case $R_1 = -10$ cm and $R_2 = +10$ cm. Again using Eq. (34.19), we get

$$\frac{1}{f} = (1.52 - 1)\left(\frac{1}{-10\text{ cm}} - \frac{1}{+10\text{ cm}}\right)$$

$$f = -9.6\text{ cm}$$

EVALUATE: In part (a) the focal length is *positive,* so this is a converging lens; this makes sense, since the lens is thicker at its center than at its edges. In part (b) the focal length is *negative,* so this is a diverging lens; this also makes sense, since the lens is thicker at its edges than at its center.

Graphical Methods for Lenses

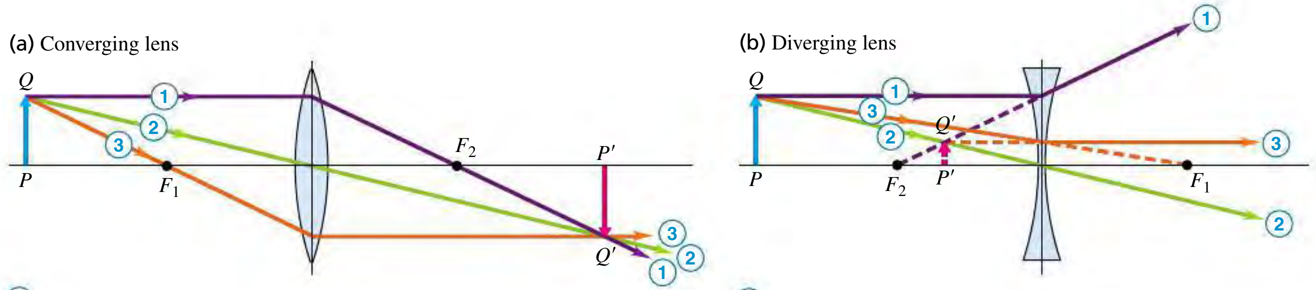
We can determine the position and size of an image formed by a thin lens by using a graphical method very similar to the one we used in Section 34.2 for spherical mirrors. Again we draw a few special rays called *principal rays* that diverge from a point of the object that is *not* on the optic axis. The intersection of these rays, after they pass through the lens, determines the position and size of the image. In using this graphical method, we will consider the entire deviation of a ray as occurring at the midplane of the lens, as shown in **Fig. 34.36**. This is consistent with the assumption that the distance between the lens surfaces is negligible.

The three principal rays whose paths are usually easy to trace for lenses are shown in Fig. 34.36:

1. *A ray parallel to the axis* emerges from the lens in a direction that passes through the second focal point F_2 of a converging lens, or appears to come from the second focal point of a diverging lens.
2. *A ray through the center of the lens* is not appreciably deviated; at the center of the lens the two surfaces are parallel, so this ray emerges at essentially the same angle at which it enters and along essentially the same line.
3. *A ray through (or proceeding toward) the first focal point F_1* emerges parallel to the axis.

When the image is real, the position of the image point is determined by the intersection of any two rays 1, 2, and 3 (Fig. 34.36a). When the image is virtual, we extend the diverging outgoing rays backward to their intersection point to find the image point (Fig. 34.36b).

34.36 The graphical method of locating an image formed by a thin lens. The colors of the rays are for identification only; they do not refer to specific colors of light. (Compare Fig. 34.19 for spherical mirrors.)



(1) Parallel incident ray refracts to pass through second focal point F_2.
(2) Ray through center of lens does not deviate appreciably.
(3) Ray through the first focal point F_1 emerges parallel to the axis.

(1) Parallel incident ray appears after refraction to have come from the second focal point F_2.
(2) Ray through center of lens does not deviate appreciably.
(3) Ray aimed at the first focal point F_1 emerges parallel to the axis.

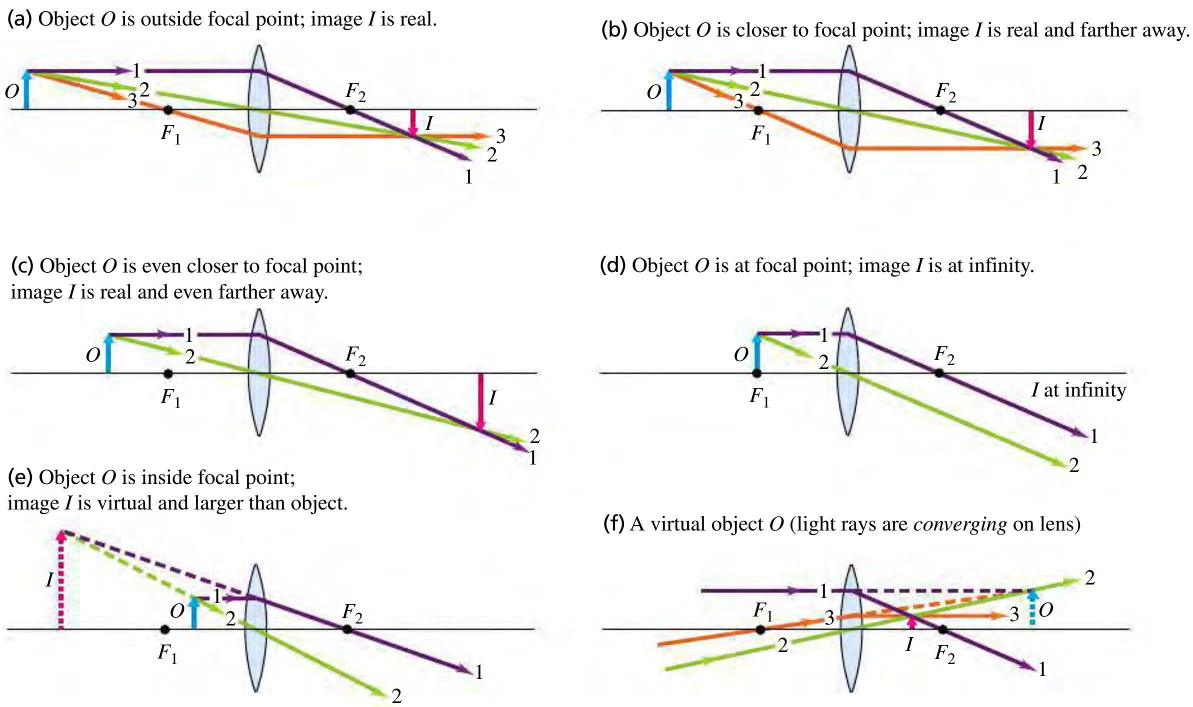
CAUTION **Principal rays are not the only rays** Keep in mind that *any* ray from the object that strikes the lens will pass through the image point (for a real image) or appear to originate from the image point (for a virtual image). (We made a similar comment about image formation by mirrors in Section 34.2.) We've emphasized the principal rays because they're the only ones you need to draw to locate the image.

Figure 34.37 shows principal-ray diagrams for a converging lens for several object distances. We suggest you study each of these diagrams very carefully, comparing each numbered ray with the above description.

DEMO

Parts (a), (b), and (c) of Fig. 34.37 help explain what happens in focusing a camera. For a photograph to be in sharp focus, the electronic sensor or film must be at the position of the real image made by the camera's lens. The image distance increases as the object is brought closer, so the sensor is moved farther behind the lens (i.e., the lens is moved farther in front of the sensor).

34.37 Formation of images by a thin converging lens for various object distances. The principal rays are numbered. (Compare Fig. 34.20 for a concave spherical mirror.)



In Fig. 34.37d the object is at the focal point; ray 3 can't be drawn because it doesn't pass through the lens. In Fig. 34.37e the object distance is less than the focal length. The outgoing rays are divergent, and the image is *virtual;* its position is located by extending the outgoing rays backward, so the image distance s' is negative. Note also that the image is erect and larger than the object. (We'll see the usefulness of this in Section 34.6.) Figure 34.37f corresponds to a *virtual object*. The incoming rays do not diverge from a real object, but are *converging* as though they would meet at the tip of the virtual object O on the right side; the object distance s is negative in this case. The image is real and is located between the lens and the second focal point. This situation can arise if the rays that strike the lens in Fig. 34.37f emerge from another converging lens (not shown) to the left of the figure.

PROBLEM-SOLVING STRATEGY 34.2 IMAGE FORMATION BY THIN LENSES

IDENTIFY *the relevant concepts:* Review Problem-Solving Strategy 34.1 (Section 34.2) for mirrors, which is equally applicable here. As for mirrors, you should use *both* principal-ray diagrams and equations to solve problems that involve image formation by lenses.

SET UP *the problem:* Identify the target variables.

EXECUTE *the solution* as follows:
1. Draw a large principal-ray diagram if you have enough information, using graph paper or quadrille-ruled paper. Orient your diagram so that incoming rays go from left to right. Draw the rays with a ruler, and measure distances carefully.
2. Draw the principal rays so they change direction at the midplane of the lens, as in Fig. 34.36. For a lens there are only three principal rays (compared to four for a mirror). Draw all three whenever possible; the intersection of any two rays determines the image location, but the third ray should pass through the same point.
3. If the outgoing principal rays diverge, extend them backward to find the virtual image point on the *incoming* side of the lens, as in Fig. 34.27e.
4. Solve Eqs. (34.16) and (34.17), as appropriate, for the target variables. Carefully use the sign rules given in Section 34.1.
5. The *image* from a first lens or mirror may serve as the *object* for a second lens or mirror. In finding the object and image distances for this intermediate image, be sure you include the distance between the two elements (lenses and/or mirrors) correctly.

EVALUATE *your answer:* Your calculated results must be consistent with your ray-diagram results. Check that they give the same image position and image size, and that they agree on whether the image is real or virtual.

EXAMPLE 34.9 IMAGE POSITION AND MAGNIFICATION WITH A CONVERGING LENS

SOLUTION

Use ray diagrams to find the image position and magnification for an object at each of the following distances from a converging lens with a focal length of 20 cm: (a) 50 cm; (b) 20 cm; (c) 15 cm; (d) -40 cm. Check your results by calculating the image position and lateral magnification by using Eqs. (34.16) and (34.17), respectively.

SOLUTION

IDENTIFY and SET UP: We are given the focal length $f = 20$ cm and four object distances s. Our target variables are the corresponding image distances s' and lateral magnifications m. We solve Eq. (34.16) for s', and find m from Eq. (34.17), $m = -s'/s$.

EXECUTE: Figures 34.37a, d, e, and f, respectively, show the appropriate principal-ray diagrams. You should be able to reproduce these without referring to the figures. Measuring these diagrams yields the approximate results: $s' = 35$ cm, $-\infty$, -40 cm, and 15 cm, and $m = -\frac{2}{3}$, $+\infty$, $+3$, and $+\frac{1}{3}$, respectively.

Calculating the image distances from Eq. (34.16), we find

$$\text{(a)}\ \frac{1}{50\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = 33.3\text{ cm}$$

$$\text{(b)}\ \frac{1}{20\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = \pm\infty$$

$$\text{(c)}\ \frac{1}{15\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = -60\text{ cm}$$

$$\text{(d)}\ \frac{1}{-40\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = 13.3\text{ cm}$$

The graphical results are fairly close to these except for part (c); the accuracy of the diagram in Fig. 34.37e is limited because the rays extended backward have nearly the same direction.

From Eq. (34.17),

$$\text{(a)}\ m = -\frac{33.3\text{ cm}}{50\text{ cm}} = -\frac{2}{3} \qquad \text{(b)}\ m = -\frac{\pm\infty\text{ cm}}{20\text{ cm}} = \pm\infty$$

$$\text{(c)}\ m = -\frac{-60\text{ cm}}{15\text{ cm}} = +4 \qquad \text{(d)}\ m = -\frac{13.3\text{ cm}}{-40\text{ cm}} = +\frac{1}{3}$$

EVALUATE: Note that the image distance s' is positive in parts (a) and (d) but negative in part (c). This makes sense: The image is real in parts (a) and (d) but virtual in part (c). The light rays that emerge from the lens in part (b) are parallel and never converge, so the image can be regarded as being at either $+\infty$ or $-\infty$.

The values of magnification m tell us that the image is inverted in part (a) and erect in parts (c) and (d), in agreement with the principal-ray diagrams. The infinite value of magnification in part (b) is another way of saying that the image is formed infinitely far away.

EXAMPLE 34.10 IMAGE FORMATION BY A DIVERGING LENS

SOLUTION

A beam of parallel rays spreads out after passing through a thin diverging lens, as if the rays all came from a point 20.0 cm from the center of the lens. You want to use this lens to form an erect, virtual image that is $\frac{1}{3}$ the height of the object. (a) Where should the object be placed? Where will the image be? (b) Draw a principal-ray diagram.

SOLUTION

IDENTIFY and SET UP: The result with parallel rays shows that the focal length is $f = -20$ cm. We want the lateral magnification to be $m = +\frac{1}{3}$ (positive because the image is to be erect). Our target variables are the object distance s and the image distance s'. In part (a), we solve the magnification equation, Eq. (34.17), for s' in terms of s; we then use the object–image relationship, Eq. (34.16), to find s and s' individually.

EXECUTE: (a) From Eq. (34.17), $m = +\frac{1}{3} = -s'/s$, so $s' = -s/3$. We insert this result into Eq. (34.16) and solve for the object distance s:

$$\frac{1}{s} + \frac{1}{-s/3} = \frac{1}{s} - \frac{3}{s} = -\frac{2}{s} = \frac{1}{f}$$

$$s = -2f = -2(-20.0\text{ cm}) = 40.0\text{ cm}$$

The object should be 40.0 cm from the lens. The image distance will be

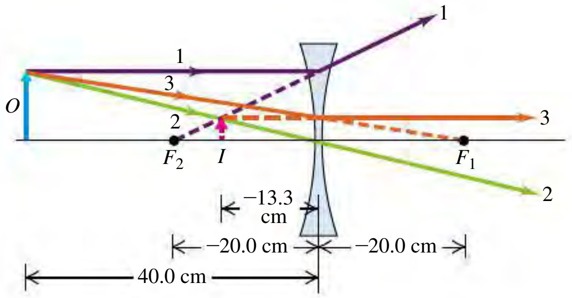
$$s' = -\frac{s}{3} = -\frac{40.0\text{ cm}}{3} = -13.3\text{ cm}$$

The image distance is negative, so the object and image are on the same side of the lens.

(b) **Figure 34.38** is a principal-ray diagram for this problem, with the rays numbered as in Fig. 34.36b.

EVALUATE: You should be able to draw a principal-ray diagram like Fig. 34.38 without referring to the figure. From your diagram, you can confirm our results in part (a) for the object and image distances. You can also check our results for s and s' by substituting them back into Eq. (34.16).

34.38 Principal-ray diagram for an image formed by a thin diverging lens.



EXAMPLE 34.11 AN IMAGE OF AN IMAGE

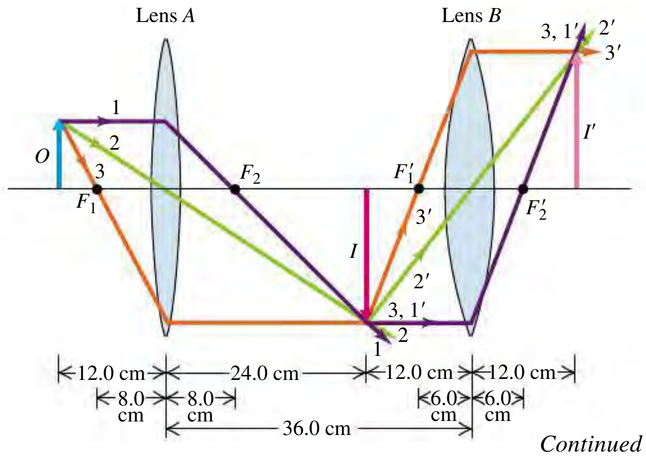
SOLUTION

Converging lenses A and B, of focal lengths 8.0 cm and 6.0 cm, respectively, are placed 36.0 cm apart. Both lenses have the same optic axis. An object 8.0 cm high is placed 12.0 cm to the left of lens A. Find the position, size, and orientation of the image produced by the lenses in combination. (Such combinations are used in telescopes and microscopes, to be discussed in Section 34.7.)

SOLUTION

IDENTIFY and SET UP: Figure 34.39 shows the situation. The object O lies outside the first focal point F_1 of lens A, which therefore produces a real image I. The light rays that strike lens B diverge from this real image just as if I was a material object; image I therefore acts as an *object* for lens B. Our goal is to determine the properties of the image I' made by lens B. We use both ray-diagram and computational methods to do this.

34.39 Principal-ray diagram for a combination of two converging lenses. The first lens (A) makes a real image of the object. This real image acts as an object for the second lens (B).



Continued

EXECUTE: In Fig. 34.39 we have drawn principal rays 1, 2, and 3 from the head of the object arrow O to find the position of the image I made by lens A, and principal rays $1'$, $2'$, and $3'$ from the head of I to find the position of the image I' made by lens B (even though rays $2'$ and $3'$ don't actually exist in this case). The image is inverted *twice,* once by each lens, so the second image I' has the same orientation as the original object.

We first find the position and size of the first image I. Applying Eq. (34.16), $1/s + 1/s' = 1/f$, to lens A gives

$$\frac{1}{12.0 \text{ cm}} + \frac{1}{s'_{I,A}} = \frac{1}{8.0 \text{ cm}} \qquad s'_{I,A} = +24.0 \text{ cm}$$

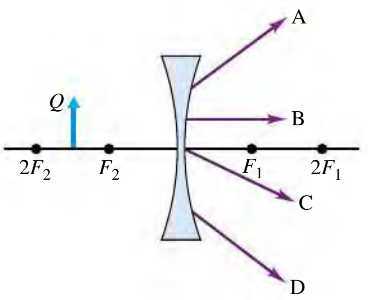
Image I is 24.0 cm to the right of lens A. The lateral magnification is $m_A = -(24.0 \text{ cm})/(12.0 \text{ cm}) = -2.00$, so image I is inverted and twice as tall as object O.

Image I is 36.0 cm $-$ 24.0 cm $=$ 12.0 cm to the left of lens B, so the object distance for lens B is +12.0 cm. Applying Eq. (34.16) to lens B then gives

$$\frac{1}{12.0 \text{ cm}} + \frac{1}{s'_{I',B}} = \frac{1}{6.0 \text{ cm}} \qquad s'_{I',B} = +12.0 \text{ cm}$$

The final image I' is 12.0 cm to the right of lens B. The magnification produced by lens B is $m_B = -(12.0 \text{ cm})/(12.0 \text{ cm}) = -1.00$.

EVALUATE: The value of m_B means that the final image I' is just as large as the first image I but has the opposite orientation. The *overall* magnification is $m_A m_B = (-2.00)(-1.00) = +2.00$. Hence the final image I' is $(2.00)(8.0 \text{ cm}) = 16$ cm tall and has the same orientation as the original object O, just as Fig. 34.39 shows.



TEST YOUR UNDERSTANDING OF SECTION 34.4 A diverging lens and an object are positioned as shown in the accompanying figure. Which of the rays A, B, C, and D could emanate from point Q at the top of the object?

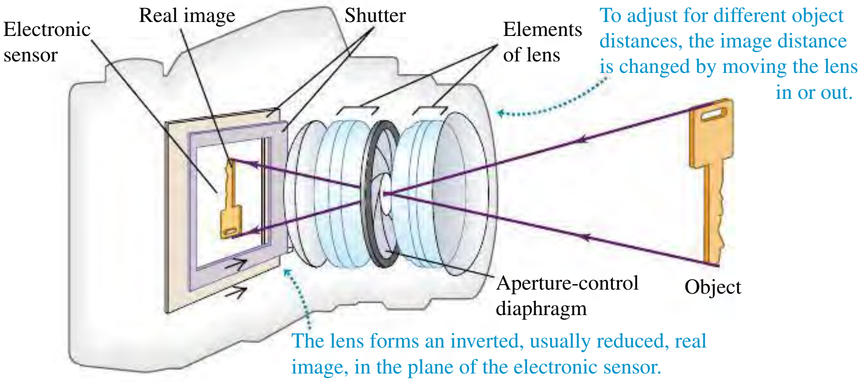
34.5 CAMERAS

The concept of *image,* which is so central to understanding simple mirror and lens systems, plays an equally important role in the analysis of optical instruments (also called *optical devices*). Among the most common optical devices are cameras, which make an image of an object and record it either electronically or on film.

The basic elements of a **camera** are a light-tight box ("camera" is a Latin word meaning "a room or enclosure"), a converging lens, a shutter to open the lens for a prescribed length of time, and a light-sensitive recording medium (**Fig. 34.40**). In digital cameras (including mobile-phone cameras), this is an electronic sensor; in older cameras, it is photographic film. The lens forms an inverted real image on the recording medium of the object being photographed. High-quality camera lenses have several elements, permitting partial correction of various *aberrations,* including the dependence of index of refraction on wavelength and the limitations imposed by the paraxial approximation.

When the camera is in proper *focus*, the position of the recording medium coincides with the position of the real image formed by the lens. The resulting photograph will then be as sharp as possible. With a converging lens, the image distance increases as the object distance decreases (see **Figs. 34.41a**, 34.41b, and 34.41c, and the discussion in Section 34.4). Hence in "focusing" the camera, we move the lens closer to the sensor or film for a distant object and farther from the sensor or film for a nearby object.

34.40 Key elements of a digital camera.



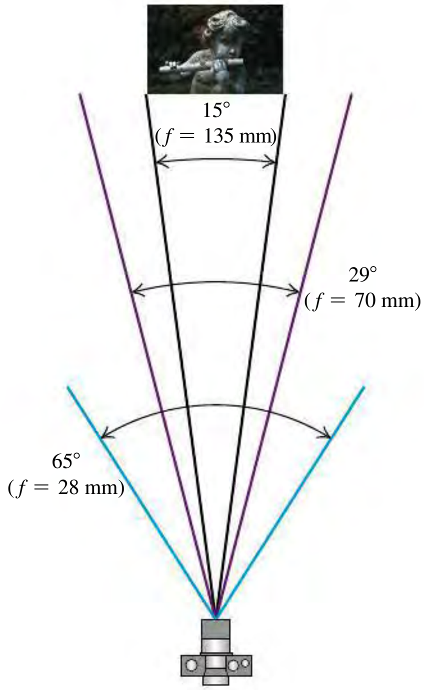
34.41 (a), (b), (c) Three photographs taken with the same camera from the same position, using lenses with focal lengths $f = 28$ mm, 70 mm, and 135 mm. Increasing the focal length increases the image size proportionately. (d) The larger the value of f, the narrower the angle of view. The angles shown here are for a camera with image area 24 mm $\times$ 36 mm (corresponding to 35-mm film) and refer to the angle of view along the 36-mm width of the film.

(a) $f = 28$ mm

(b) $f = 70$ mm

(c) $f = 135$ mm

(d) The angles of view for the photos in (a)–(c)



Camera Lenses: Focal Length

The choice of the focal length f for a camera lens depends on the size of the electronic sensor or film and the desired angle of view. Figure 34.41 shows three photographs taken on 35-mm film with the same camera at the same position, but with lenses of different focal lengths. A lens of long focal length, called a *telephoto* lens, gives a narrow angle of view and a large image of a distant object (such as the statue in Fig. 34.41c); a lens of short focal length gives a small image and a wide angle of view (as in Fig. 34.41a) and is called a *wide-angle* lens. To understand this behavior, recall that the focal length is the distance from the lens to the image when the object is infinitely far away. In general, for *any* object distance, using a lens of longer focal length gives a greater image distance. This also increases the height of the image; as discussed in Section 34.4, the ratio of the image height y' to the object height y (the *lateral magnification*) is equal in absolute value to the ratio of the image distance s' to the object distance s [Eq. (34.17)]:

$$m = \frac{y'}{y} = -\frac{s'}{s}$$

With a lens of short focal length, the ratio s'/s is small, and a distant object gives only a small image. When a lens with a long focal length is used, the image of this same object may entirely cover the area of the electronic sensor or film. Hence the longer the focal length, the narrower the angle of view (Fig. 34.41d).

Camera Lenses: *f*-Number

For a camera to record the image properly, the total light energy per unit area reaching the electronic sensor or film (the "exposure") must fall within certain limits. This is controlled by the *shutter* and the *lens aperture*. The shutter controls the time interval during which light enters the lens. This is usually adjustable in steps corresponding to factors of about 2, often from 1 s to $\frac{1}{1000}$ s.

The intensity of light reaching the sensor or film is proportional to the area viewed by the camera lens and to the effective area of the lens. The size of the area that the lens "sees" is proportional to the square of the angle of view of the lens, and so is roughly proportional to $1/f^2$. The effective area of the lens is controlled by means of an adjustable lens aperture, or *diaphragm,* a nearly circular hole with variable diameter D; hence the effective area is proportional to D^2. Putting these factors together, we see that the intensity of light reaching the sensor or film with a particular lens is proportional to D^2/f^2.

Application Inverting an Inverted Image A camera lens makes an inverted image on the camera's light-sensitive electronic detector. The internal software of the camera then inverts the image again so it appears the right way around on the camera's display. A similar thing happens with your vision: The image formed on the retina of your eye is inverted, but your brain's "software" erects the image so you see the world right-side up.

Photographers commonly express the light-gathering capability of a lens in terms of the ratio f/D, called the ***f*-number** of the lens:

$$f\text{-number of a lens} = \frac{f}{D} \quad \begin{matrix}\leftarrow \text{Focal length of lens} \\ \leftarrow \text{Aperture diameter}\end{matrix} \tag{34.20}$$

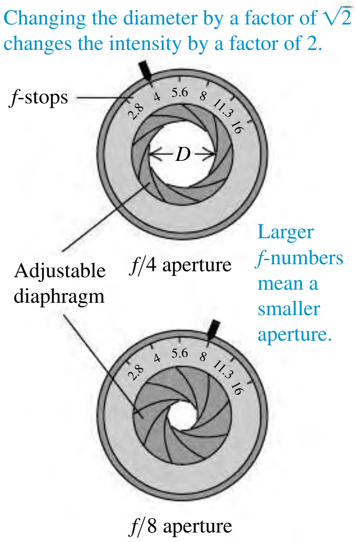
For example, a lens with a focal length $f = 50$ mm and an aperture diameter $D = 25$ mm has an f-number of 2, or "an aperture of $f/2$." The light intensity reaching the sensor or film is *inversely* proportional to the square of the f-number.

For a lens with a variable-diameter aperture, increasing the diameter by a factor of $\sqrt{2}$ changes the f-number by $1/\sqrt{2}$ and increases the intensity at the sensor or film by a factor of 2. Adjustable apertures usually have scales labeled with successive numbers (often called *f-stops*) related by factors of $\sqrt{2}$, such as

$$f/2 \qquad f/2.8 \qquad f/4 \qquad f/5.6 \qquad f/8 \qquad f/11 \qquad f/16$$

and so on. The larger numbers represent smaller apertures and exposures, and each step corresponds to a factor of 2 in intensity (**Fig. 34.42**). The actual *exposure* (total amount of light reaching the sensor or film) is proportional to both the aperture area and the time of exposure. Thus $f/4$ and $\frac{1}{500}$ s, $f/5.6$ and $\frac{1}{250}$ s, and $f/8$ and $\frac{1}{125}$ s all correspond to the same exposure.

34.42 A camera lens with an adjustable diaphragm.

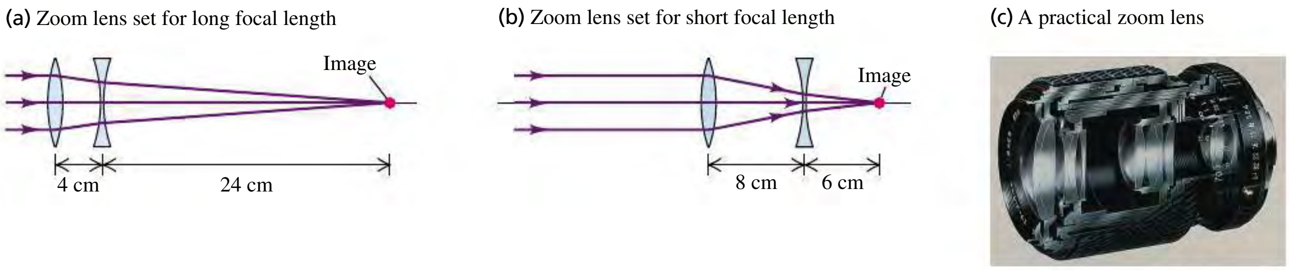


Zoom Lenses and Projectors

Many photographers use a *zoom lens,* which is not a single lens but a complex collection of several lens elements that give a continuously variable focal length, often over a range as great as 10 to 1. **Figures 34.43a** and 34.43b show a simple system with variable focal length, and Fig. 34.43c shows a typical zoom lens for a digital single-lens reflex camera. Zoom lenses give a range of image sizes of a given object. It is an enormously complex problem in optical design to keep the image in focus and maintain a constant f-number while the focal length changes. When you vary the focal length of a typical zoom lens, two groups of elements move within the lens and a diaphragm opens and closes.

A *digital projector* for viewing lecture slides, photos, or movies operates very much like a digital camera in reverse. In the most common type of digital projector, the pixels of data to be projected are shown on a small, transparent liquid-crystal-display (LCD) screen inside the projector and behind the projection lens. A lamp illuminates the LCD screen, which acts as an object for the lens. The lens forms a real, enlarged, inverted image of the LCD screen. Because the image is inverted, the pixels shown on the LCD screen are upside down so that the image on the projection screen appears right-side up.

34.43 A simple zoom lens uses a converging lens and a diverging lens in tandem. (a) When the two lenses are close together, the combination behaves like a single lens of long focal length. (b) If the two lenses are moved farther apart, the combination behaves like a short-focal-length lens. (c) This zoom lens contains twelve elements arranged in four groups.



EXAMPLE 34.12 PHOTOGRAPHIC EXPOSURES

A common telephoto lens for a 35-mm film camera has a focal length of 200 mm; its f-stops range from $f/2.8$ to $f/22$. (a) What is the corresponding range of aperture diameters? (b) What is the corresponding range of image intensities on the film?

SOLUTION

IDENTIFY and SET UP: Part (a) of this problem uses the relationship among lens focal length f, aperture diameter D, and f-number. Part (b) uses the relationship between intensity and aperture diameter. We use Eq. (34.20) to relate D (the target variable) to the f-number and the focal length $f = 200$ mm. The intensity of the light reaching the film is proportional to D^2/f^2; since f is the same in each case, we conclude that the intensity in this case is proportional to D^2, the square of the aperture diameter.

EXECUTE: (a) From Eq. (34.20), the diameter ranges from

$$D = \frac{f}{f\text{-number}} = \frac{200 \text{ mm}}{2.8} = 71 \text{ mm}$$

to

$$D = \frac{200 \text{ mm}}{22} = 9.1 \text{ mm}$$

(b) Because the intensity is proportional to D^2, the ratio of the intensity at $f/2.8$ to the intensity at $f/22$ is

$$\left(\frac{71 \text{ mm}}{9.1 \text{ mm}}\right)^2 = \left(\frac{22}{2.8}\right)^2 = 62 \qquad (\text{about } 2^6)$$

EVALUATE: If the correct exposure time at $f/2.8$ is $\frac{1}{1000}$ s, then the exposure at $f/22$ is $(62)\left(\frac{1}{1000}\text{ s}\right) = \frac{1}{16}$ s to compensate for the lower intensity. In general, the smaller the aperture and the larger the f-number, the longer the required exposure. Nevertheless, many photographers prefer to use small apertures so that only the central part of the lens is used to make the image. This minimizes aberrations that occur near the edges of the lens and gives the sharpest possible image.

TEST YOUR UNDERSTANDING OF SECTION 34.5 When used with 35-mm film (image area 24 mm × 36 mm), a lens with $f = 50$ mm gives a 45° angle of view and is called a "normal lens." When used with an electronic sensor that measures 5 mm × 5 mm, this same lens is (i) a wide-angle lens; (ii) a normal lens; (iii) a telephoto lens. ❙

34.6 THE EYE

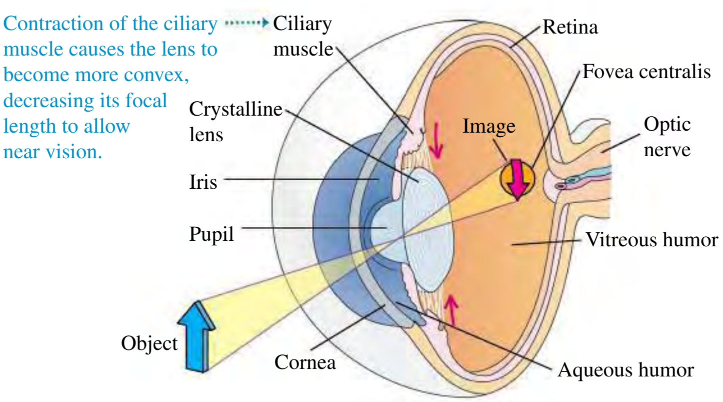
MP

PhET: Color Vision

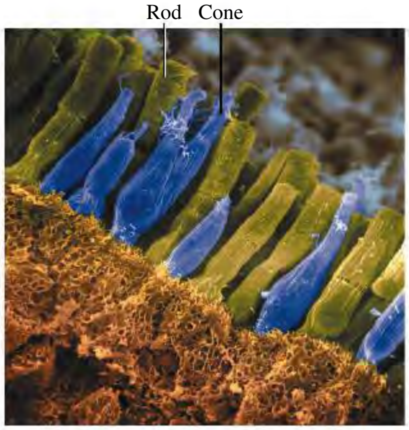
The optical behavior of the eye is similar to that of a camera. **Figure 34.44** shows the essential parts of the human eye, considered as an optical system. The eye is nearly spherical and about 2.5 cm in diameter. The front portion is somewhat more sharply curved and is covered by a tough, transparent membrane called the *cornea*. The region behind the cornea contains a liquid called the *aqueous humor*. Next comes the *crystalline lens,* a capsule containing a fibrous jelly, hard

34.44 (a) The eye. (b) There are two types of light-sensitive cells on the retina. Rods are more sensitive to light than cones, but only the cones are sensitive to differences in color. A typical human eye contains about 1.3×10^8 rods and about 7×10^6 cones.

(a) Diagram of the eye



(b) Scanning electron micrograph showing retinal rods and cones in different colors



BIO Application Focusing in the Animal Kingdom The crystalline lens and ciliary muscle found in humans and other mammals are among a number of focusing mechanisms used by animals. Birds can change the shape not only of their lens but also of the corneal surface. In aquatic animals the corneal surface is not very useful for focusing because its index of refraction is close to that of water. Thus, focusing is accomplished entirely by the lens, which is nearly spherical. Fish focus by using a muscle to move the lens either inward or outward. Whales and dolphins achieve the same effect by filling or emptying a fluid chamber behind the lens to move the lens in or out.

at the center and progressively softer at the outer portions. The crystalline lens is held in place by ligaments that attach it to the ciliary muscle, which encircles it. Behind the lens, the eye is filled with a thin watery jelly called the *vitreous humor*. The indexes of refraction of both the aqueous humor and the vitreous humor are about 1.336, nearly equal to that of water. The crystalline lens, while not homogeneous, has an average index of 1.437. This is not very different from the indexes of the aqueous and vitreous humors. As a result, most of the refraction of light entering the eye occurs at the outer surface of the cornea.

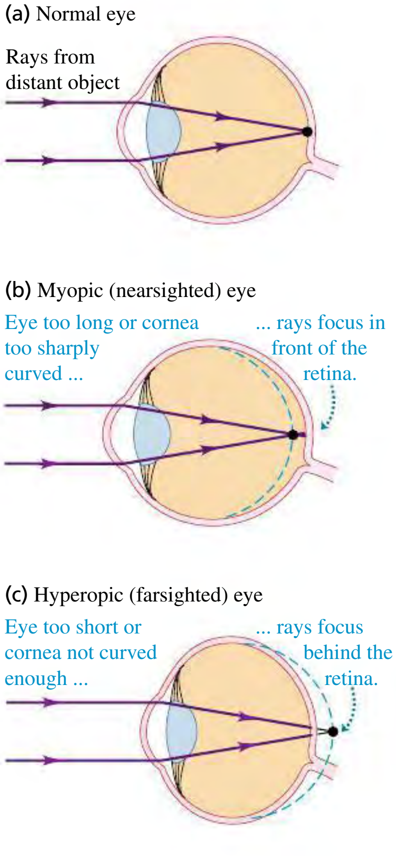
Refraction at the cornea and the surfaces of the lens produces a *real image* of the object being viewed. This image is formed on the light-sensitive *retina,* lining the rear inner surface of the eye. The retina plays the same role as the electronic sensor in a digital camera. The *rods* and *cones* in the retina act like an array of miniature photocells (Fig. 34.44b); they sense the image and transmit it via the *optic nerve* to the brain. Vision is most acute in a small central region called the *fovea centralis,* about 0.25 mm in diameter.

In front of the lens is the *iris*. It contains an aperture with variable diameter called the *pupil,* which opens and closes to adapt to changing light intensity. The receptors of the retina also have intensity adaptation mechanisms.

For an object to be seen sharply, the image must be formed exactly at the location of the retina. The eye adjusts to different object distances s by changing the focal length f of its lens; the lens-to-retina distance, corresponding to s', does not change. (Contrast this with focusing a camera, in which the focal length is fixed and the lens-to-sensor distance is changed.) For the normal eye, an object at infinity is sharply focused when the ciliary muscle is relaxed. To focus sharply on a closer object, the tension in the ciliary muscle surrounding the lens increases, the ciliary muscle contracts, the lens bulges, and the radii of curvature of its surfaces decrease; this decreases f. This process is called *accommodation*.

The extremes of the range over which distinct vision is possible are known as the *far point* and the *near point* of the eye. The far point of a normal eye is at infinity. The position of the near point depends on the amount by which the ciliary muscle can increase the curvature of the crystalline lens. The range of accommodation gradually diminishes with age because the crystalline lens grows throughout a person's life (it is about 50% larger at age 60 than at age 20) and the ciliary muscles are less able to distort a larger lens. For this reason, the near point gradually recedes as one grows older. This recession of the near point is called *presbyopia*. **Table 34.1** shows the approximate position of the near point for an average person at various ages. For example, an average person 50 years of age cannot focus on an object that is closer than about 40 cm.

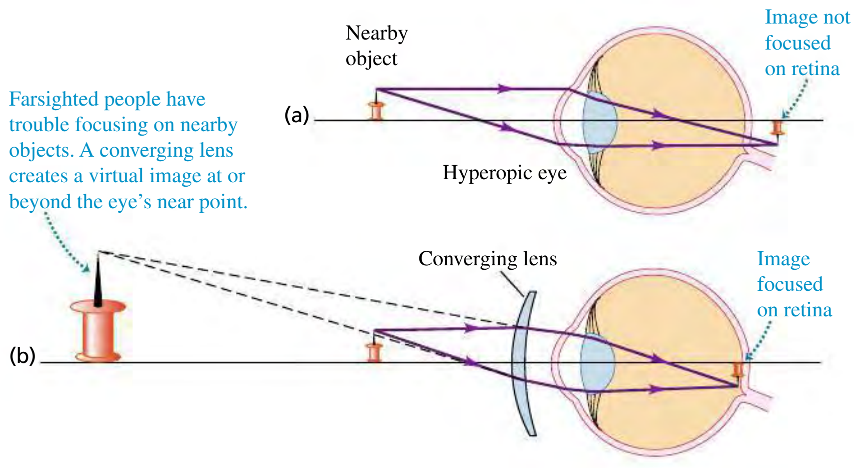
34.45 Refractive errors for (a) a normal eye, (b) a myopic (nearsighted) eye, and (c) a hyperopic (farsighted) eye viewing a very distant object. In each case, the eye is shown with the ciliary muscle relaxed. The dashed blue curve indicates the required position of the retina.



Defects of Vision

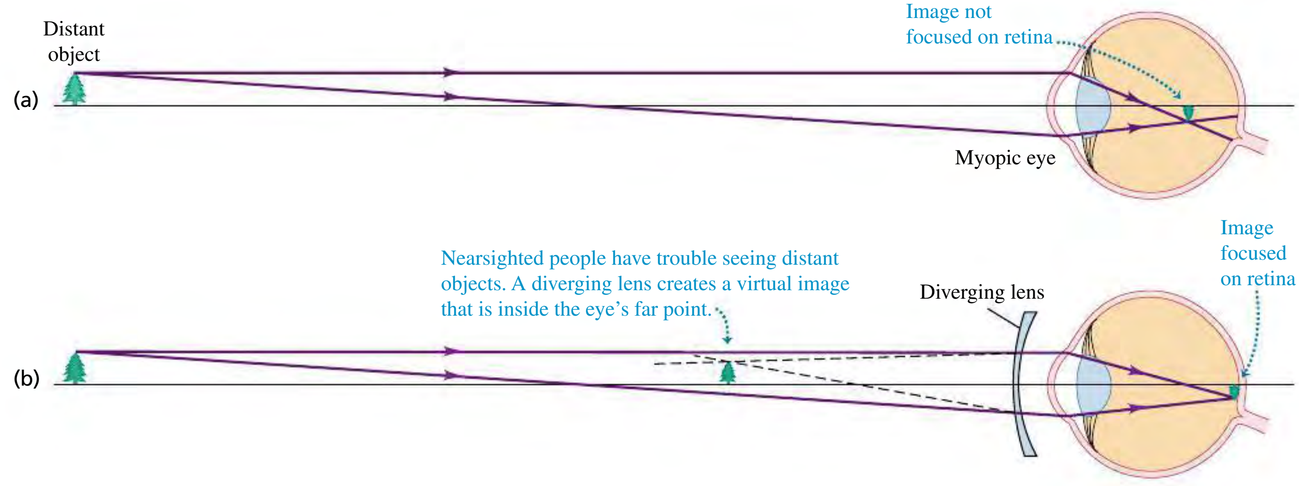
Several common defects of vision result from incorrect distance relationships in the eye. A normal eye forms an image on the retina of an object at infinity when the eye is relaxed (**Fig. 34.45a**). In the *myopic* (nearsighted) eye, the eyeball is too long from front to back in comparison with the radius of curvature of the cornea (or the cornea is too sharply curved), and rays from an object at infinity are focused in front of the retina (Fig. 34.45b). The most distant object for which an image can be formed on the retina is then nearer than infinity. In the *hyperopic* (farsighted) eye, the eyeball is too short or the cornea is not curved enough, and the image of an infinitely distant object is behind the retina (Fig. 34.45c). The myopic eye produces *too much* convergence in a parallel bundle of rays for an image to be formed on the retina; the hyperopic eye, *not enough* convergence.

All of these defects can be corrected by the use of corrective lenses (eyeglasses or contact lenses). The near point of either a presbyopic or a hyperopic eye is *farther* from the eye than normal. To see clearly an object at normal reading distance (often assumed to be 25 cm), we need a lens that forms a virtual image of the object at or beyond the near point. This can be accomplished by a converging (positive) lens (**Fig. 34.46**). In effect the lens moves the object farther



34.46 (a) An uncorrected hyperopic (farsighted) eye. (b) A positive (converging) lens gives the extra convergence needed for a hyperopic eye to focus the image on the retina.

34.47 (a) An uncorrected myopic (nearsighted) eye. (b) A negative (diverging) lens spreads the rays farther apart to compensate for the excessive convergence of the myopic eye.



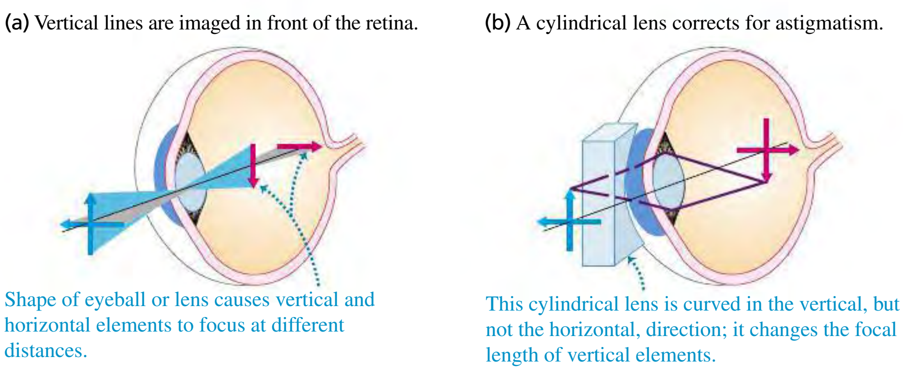
away from the eye to a point where a sharp retinal image can form. Similarly, correcting the myopic eye involves the use of a diverging (negative) lens to move the image closer to the eye than the actual object is (**Fig. 34.47**).

Astigmatism is a different type of defect in which the surface of the cornea is not spherical but rather more sharply curved in one plane than in another. As a result, horizontal lines may be imaged in a different plane from vertical lines (**Fig. 34.48a**). Astigmatism may make it impossible, for example, to focus clearly on both the horizontal and vertical bars of a window at the same time.

Astigmatism can be corrected by use of a lens with a *cylindrical* surface. For example, suppose the curvature of the cornea in a horizontal plane is correct to

TABLE 34.1 Receding of Near Point with Age

Age (years)	Near Point (cm)
10	7
20	10
30	14
40	22
50	40
60	200



34.48 One type of astigmatism and how it is corrected.

focus rays from infinity on the retina but the curvature in the vertical plane is too great to form a sharp retinal image. When a cylindrical lens with its axis horizontal is placed before the eye, the rays in a horizontal plane are unaffected, but the additional divergence of the rays in a vertical plane causes these to be sharply imaged on the retina (Fig. 34.48b).

Lenses for vision correction are usually described in terms of the **power,** defined as the reciprocal of the focal length expressed in meters. The unit of power is the **diopter.** Thus a lens with $f = 0.50$ m has a power of 2.0 diopters, $f = -0.25$ m corresponds to -4.0 diopters, and so on. The numbers on a prescription for glasses are usually powers expressed in diopters. When the correction involves both astigmatism and myopia or hyperopia, there are three numbers: one for the spherical power, one for the cylindrical power, and an angle to describe the orientation of the cylinder axis.

EXAMPLE 34.13 CORRECTING FOR FARSIGHTEDNESS

SOLUTION

The near point of a certain hyperopic eye is 100 cm in front of the eye. Find the focal length and power of the contact lens that will permit the wearer to see clearly an object that is 25 cm in front of the eye.

SOLUTION

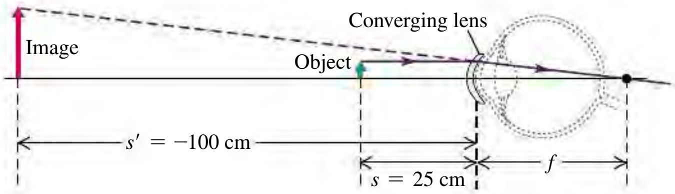
IDENTIFY and SET UP: Figure 34.49 shows the situation. We want the lens to form a virtual image of the object at the near point of the eye, 100 cm from it. The contact lens (which we treat as having negligible thickness) is at the surface of the cornea, so the object distance is $s = 25$ cm. The virtual image is on the incoming side of the contact lens, so the image distance is $s' = -100$ cm. We use Eq. (34.16) to determine the required focal length f of the contact lens; the corresponding power is $1/f$.

EXECUTE: From Eq. (34.16),

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} = \frac{1}{+25\text{ cm}} + \frac{1}{-100\text{ cm}}$$

$$f = +33\text{ cm}$$

34.49 Using a contact lens to correct for farsightedness. For clarity, the eye and contact lens are shown much larger than the scale of the figure; the 2.5-cm diameter of the eye is actually much smaller than the focal length f of the contact lens.



We need a converging lens with focal length $f = 33$ cm and power $1/(0.33\text{ m}) = +3.0$ diopters.

EVALUATE: In this example we used a contact lens to correct hyperopia. Had we used eyeglasses, we would have had to account for the separation between the eye and the eyeglass lens, and a somewhat different power would have been required (see Example 34.14).

EXAMPLE 34.14 CORRECTING FOR NEARSIGHTEDNESS

SOLUTION

The far point of a certain myopic eye is 50 cm in front of the eye. Find the focal length and power of the eyeglass lens that will permit the wearer to see clearly an object at infinity. Assume that the lens is worn 2 cm in front of the eye.

SOLUTION

IDENTIFY and SET UP: Figure 34.50 shows the situation. The far point of a myopic eye is nearer than infinity. To see clearly objects beyond the far point, we need a lens that forms a virtual image of such objects no farther from the eye than the far point. Assume that the virtual image of the object at infinity is formed at the far point, 50 cm in front of the eye (48 cm in front of the eyeglass lens). Then when the object distance is $s = \infty$, we want the image distance to be $s' = -48$ cm. As in Example 34.13, we use the values of s and s' to calculate the required focal length.

EXECUTE: From Eq. (34.16),

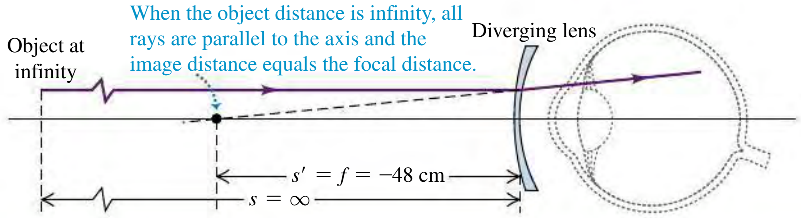
$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} = \frac{1}{\infty} + \frac{1}{-48\text{ cm}}$$

$$f = -48\text{ cm}$$

We need a *diverging* lens with focal length $f = -48$ cm and power $1/(-0.48\text{ m}) = -2.1$ diopters.

EVALUATE: If a *contact* lens were used to correct this myopia, we would need $f = -50$ cm and a power of -2.0 diopters. Can you see why?

34.50 Using an eyeglass lens to correct for nearsightedness. For clarity, the eye and eyeglass lens are shown much larger than the scale of the figure.



TEST YOUR UNDERSTANDING OF SECTION 34.6 A certain eyeglass lens is thin at its center, even thinner at its top and bottom edges, and relatively thick at its left and right edges. What defects of vision is this lens intended to correct? (i) Hyperopia for objects oriented both vertically and horizontally; (ii) myopia for objects oriented both vertically and horizontally; (iii) hyperopia for objects oriented vertically and myopia for objects oriented horizontally; (iv) hyperopia for objects oriented horizontally and myopia for objects oriented vertically. ❙

BIO Application The Telephoto Eyes of Chameleons The crystalline lens of a human eye can change shape but is always a converging (positive) lens. The lens in the eye of a chameleon lizard (family Chamaeleonidae) is different: It can change shape to be either a converging or a *diverging* (negative) lens. When it acts as a diverging lens just behind the cornea (which acts as a converging lens), the combination is like the long-focal-length zoom lens shown in Fig. 34.43a. This "telephoto vision" gives the chameleon a sharp view of potential insect prey.

34.7 THE MAGNIFIER

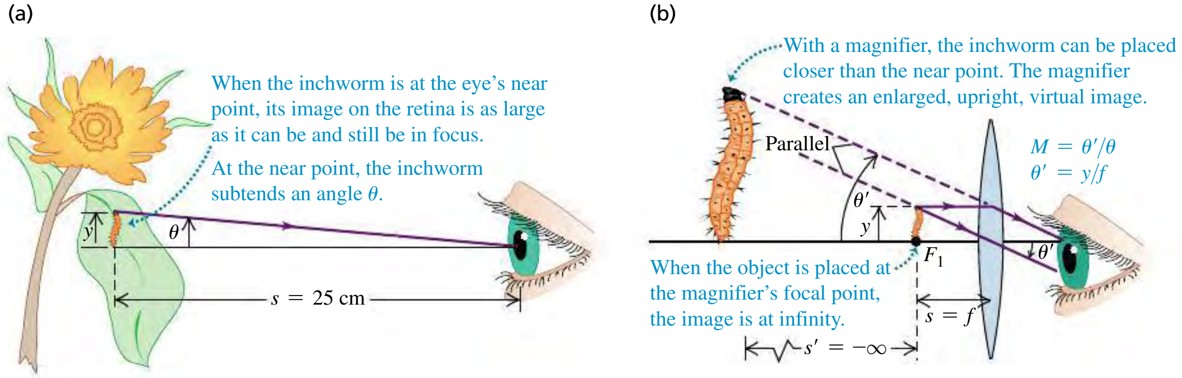
The apparent size of an object is determined by the size of its image on the retina. If the eye is unaided, this size depends on the *angle* θ subtended by the object at the eye, called its **angular size** (**Fig. 34.51a**).

To look closely at a small object such as an insect or a crystal, you bring it close to your eye, making the subtended angle and the retinal image as large as possible. But your eye cannot focus sharply on objects that are closer than the near point, so an object subtends the largest possible viewing angle when it is placed at the near point. In the following discussion we will assume an average viewer for whom the near point is 25 cm from the eye.

A converging lens can be used to form a virtual image that is larger and farther from the eye than the object itself, as shown in Fig. 34.51b. Then the object can be moved closer to the eye, and the angular size of the image may be substantially larger than the angular size of the object at 25 cm without the lens. A lens used in this way is called a **magnifier,** otherwise known as a *magnifying glass* or a *simple magnifier*. The virtual image is most comfortable to view when it is placed at infinity, so that the ciliary muscle of the eye is relaxed; this means that the object is placed at the focal point F_1 of the magnifier. In the following discussion we assume that this is done.

In Fig. 34.51a the object is at the near point, where it subtends an angle θ at the eye. In Fig. 34.51b a magnifier in front of the eye forms an image at infinity,

34.51 (a) The angular size θ is largest when the object is at the near point. (b) The magnifier gives a virtual image at infinity. This virtual image appears to the eye to be a real object subtending a larger angle θ' at the eye.



and the angle subtended at the magnifier is θ'. The usefulness of the magnifier is given by the ratio of the angle θ' (with the magnifier) to the angle θ (without the magnifier). This ratio is called the **angular magnification** M:

$$M = \frac{\theta'}{\theta} \qquad \text{(angular magnification)} \tag{34.21}$$

CAUTION **Angular magnification vs. lateral magnification** Don't confuse *angular* magnification M with *lateral* magnification m. Angular magnification is the ratio of the *angular* size of an image to the angular size of the corresponding object; lateral magnification refers to the ratio of the *height* of an image to the height of the corresponding object. For the situation shown in Fig. 34.51b, the angular magnification is about 3×, since the inchworm subtends an angle about three times larger than that in Fig. 34.51a; hence the inchworm will look about three times larger to the eye. The *lateral* magnification $m = -s'/s$ in Fig. 34.51b is *infinite* because the virtual image is at infinity, but that doesn't mean that the inchworm looks infinitely large through the magnifier! When dealing with a magnifier, M is useful but m is not.

To find the value of M, we first assume that the angles are small enough that each angle (in radians) is equal to its sine and its tangent. Using Fig. 34.51a and drawing the ray in Fig. 34.51b that passes undeviated through the center of the lens, we find that θ and θ' (in radians) are

$$\theta = \frac{y}{25\text{ cm}} \qquad \theta' = \frac{y}{f}$$

Combining these expressions with Eq. (34.21), we find

Angular size of object seen with magnifier

Object height

Angular magnification for a simple magnifier

$$M = \frac{\theta'}{\theta} = \frac{y/f}{y/25\text{ cm}} = \frac{25\text{ cm}}{f} \tag{34.22}$$

Near point

Focal length

Angular size of object seen without magnifier

It may seem that we can make the angular magnification as large as we like by decreasing the focal length f. In fact, the aberrations of a simple double-convex lens set a limit to M of about 3× to 4×. If these aberrations are corrected, the angular magnification may be made as great as 20×. A compound microscope, discussed in the next section, provides even greater magnification.

TEST YOUR UNDERSTANDING OF SECTION 34.7 You are using a magnifier to examine a gem. If you change to a different magnifier with twice the focal length of the first one, you will have to hold the object at (i) twice the distance and the angular magnification will be twice as great; (ii) twice the distance and the angular magnification will be $\frac{1}{2}$ as great; (iii) $\frac{1}{2}$ the distance and the angular magnification will be twice as great; (iv) $\frac{1}{2}$ the distance and the angular magnification will be $\frac{1}{2}$ as great.

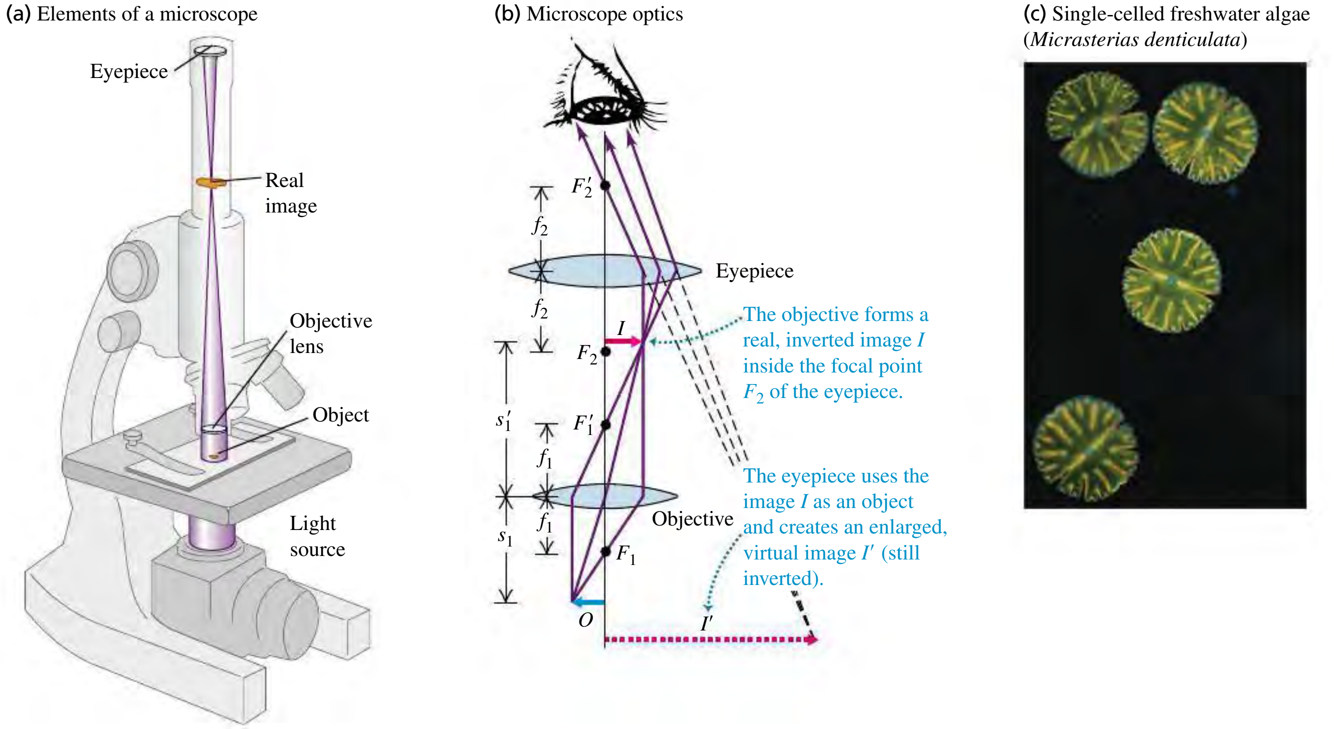
34.8 MICROSCOPES AND TELESCOPES

Cameras, eyeglasses, and magnifiers use a single lens to form an image. Two important optical devices that use *two* lenses are the microscope and the telescope. In each device a primary lens, or *objective,* forms a real image, and a second lens, or *eyepiece,* is used as a magnifier to make an enlarged, virtual image.

Microscopes

Figure 34.52a shows the essential features of a **microscope,** sometimes called a *compound microscope.* To analyze this system, we use the principle that an image formed by one optical element such as a lens or mirror can serve as the object

34.52 (a) Elements of a microscope. (b) The object O is placed just outside the first focal point of the objective (the distance s_1 has been exaggerated for clarity). (c) This microscope image shows single-celled organisms about 2×10^{-4} m (0.2 mm) across. Typical light microscopes can resolve features as small as 2×10^{-7} m, comparable to the wavelength of light.



for a second element. We used this principle in Section 34.4 when we derived the lensmaker's equation by repeated application of the single-surface refraction equation; we used this principle again in Example 34.11 (Section 34.4), in which the image formed by a lens was used as the object of a second lens.

The object O to be viewed is placed just beyond the first focal point F_1 of the **objective,** a converging lens that forms a real and enlarged image I (Fig. 34.52b). In a properly designed instrument this image lies just inside the first focal point F_2 of a second converging lens called the **eyepiece** or *ocular*. (The reason the image should lie just *inside* F_2 is left for you to discover.) The eyepiece acts as a simple magnifier, as discussed in Section 34.7, and forms a final virtual image I' of I. The position of I' may be anywhere between the near and far points of the eye. Both the objective and the eyepiece of an actual microscope are highly corrected compound lenses with several optical elements, but for simplicity we show them here as simple thin lenses.

As for a simple magnifier, what matters when viewing through a microscope is the *angular* magnification M. The overall angular magnification of the compound microscope is the product of two factors. The first factor is the *lateral* magnification m_1 of the objective, which determines the linear size of the real image I; the second factor is the *angular* magnification M_2 of the eyepiece, which relates the angular size of the virtual image seen through the eyepiece to the angular size that the real image I would have if you viewed it *without* the eyepiece. The first of these factors is given by

$$m_1 = -\frac{s_1'}{s_1} \tag{34.23}$$

where s_1 and s_1' are the object and image distances, respectively, for the objective lens. Ordinarily, the object is very close to the focal point, and the resulting image distance s_1' is very great in comparison to the focal length f_1 of the objective lens. Thus s_1 is approximately equal to f_1, and we can write $m_1 = -s_1'/f_1$.

The real image I is close to the focal point F_2 of the eyepiece, so to find the eyepiece angular magnification, we can use Eq. (34.22): $M_2 = (25\text{ cm})/f_2$, where f_2 is the focal length of the eyepiece (considered as a simple lens). The overall angular magnification M of the compound microscope (apart from a negative sign, which is customarily ignored) is the product of the two magnifications:

$$M = m_1 M_2 = \frac{(25\text{ cm})s_1'}{f_1 f_2} \quad \text{(angular magnification for a microscope)} \tag{34.24}$$

where s_1', f_1, and f_2 are measured in centimeters. The final image is inverted with respect to the object. Microscope manufacturers usually specify the values of m_1 and M_2 rather than the focal lengths of the objective and eyepiece.

Equation (34.24) shows that the angular magnification of a microscope can be increased by using an objective of shorter focal length f_1, thereby increasing m_1 and the size of the real image I. Most optical microscopes have a rotating "turret" with three or more objectives of different focal lengths so that the same object can be viewed at different magnifications. The eyepiece should also have a short focal length f_2 to help to maximize the value of M.

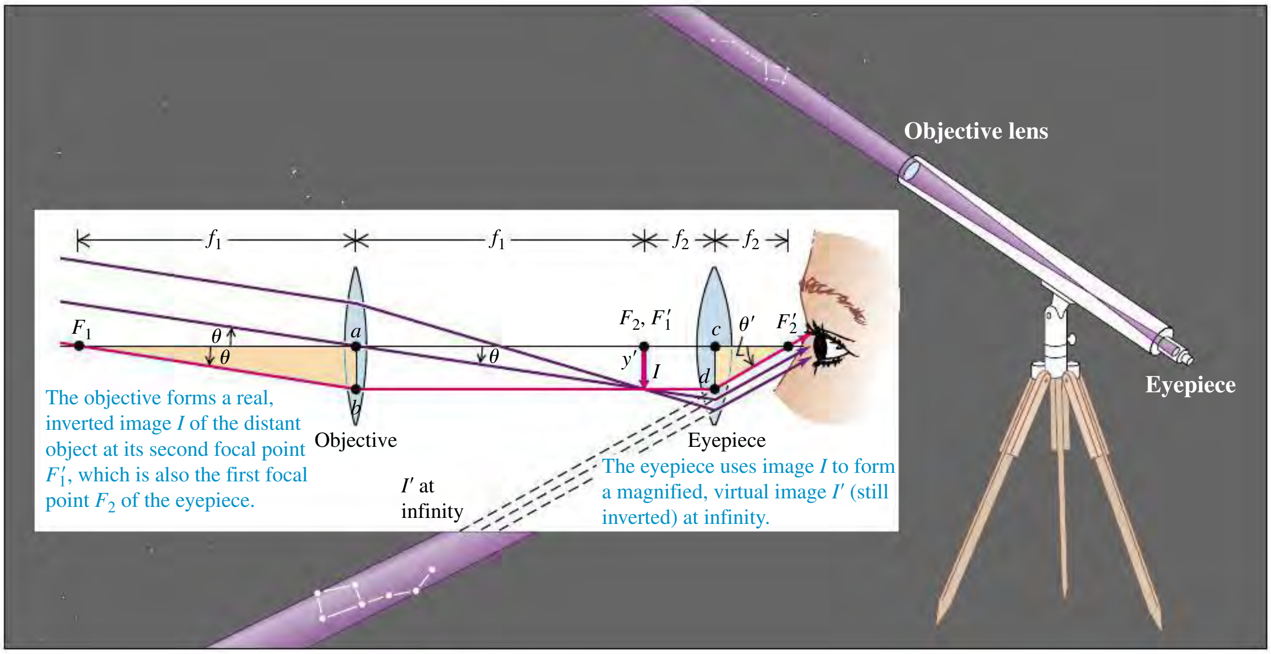
To use a microscope to take a photograph (called a *photomicrograph* or *micrograph*), the eyepiece is removed and a camera placed so that the real image I falls on the camera's electronic sensor or film. Figure 34.52c shows such a photograph. In this case what matters is the *lateral* magnification of the microscope as given by Eq. (34.23).

Telescopes

The optical system of a **telescope** is similar to that of a compound microscope. In both instruments the image formed by an objective is viewed through an eyepiece. The key difference is that the telescope is used to view large objects at large distances and the microscope is used to view small objects close at hand. Another difference is that many telescopes use a curved mirror, not a lens, as an objective.

Figure 34.53 shows an *astronomical telescope*. Because this telescope uses a lens as an objective, it is called a *refracting telescope* or *refractor*. The objective

34.53 Optical system of an astronomical refracting telescope.



lens forms a real, reduced image I of the object. This image is the object for the eyepiece lens, which forms an enlarged, virtual image of I. Objects that are viewed with a telescope are usually so far away from the instrument that the first image I is formed very nearly at the second focal point of the objective lens. If the final image I' formed by the eyepiece is at infinity (for most comfortable viewing by a normal eye), the first image must also be at the first focal point of the eyepiece. The distance between objective and eyepiece, which is the length of the telescope, is therefore the *sum* of the focal lengths of objective and eyepiece, $f_1 + f_2$.

The angular magnification M of a telescope is defined as the ratio of the angle subtended at the eye by the final image I' to the angle subtended at the (unaided) eye by the object. We can express this ratio in terms of the focal lengths of objective and eyepiece. In Fig. 34.53 the ray passing through F_1, the first focal point of the objective, and through F'_2, the second focal point of the eyepiece, is shown in red. The object (not shown) subtends an angle θ at the objective and would subtend essentially the same angle at the unaided eye. Also, since the observer's eye is placed just to the right of the focal point F'_2, the angle subtended at the eye by the final image is very nearly equal to the angle θ'. Because bd is parallel to the optic axis, the distances ab and cd are equal to each other and also to the height y' of the real image I. Because the angles θ and θ' are small, they may be approximated by their tangents. From the right triangles F_1ab and F'_2cd,

$$\theta = \frac{-y'}{f_1}$$

$$\theta' = \frac{y'}{f_2}$$

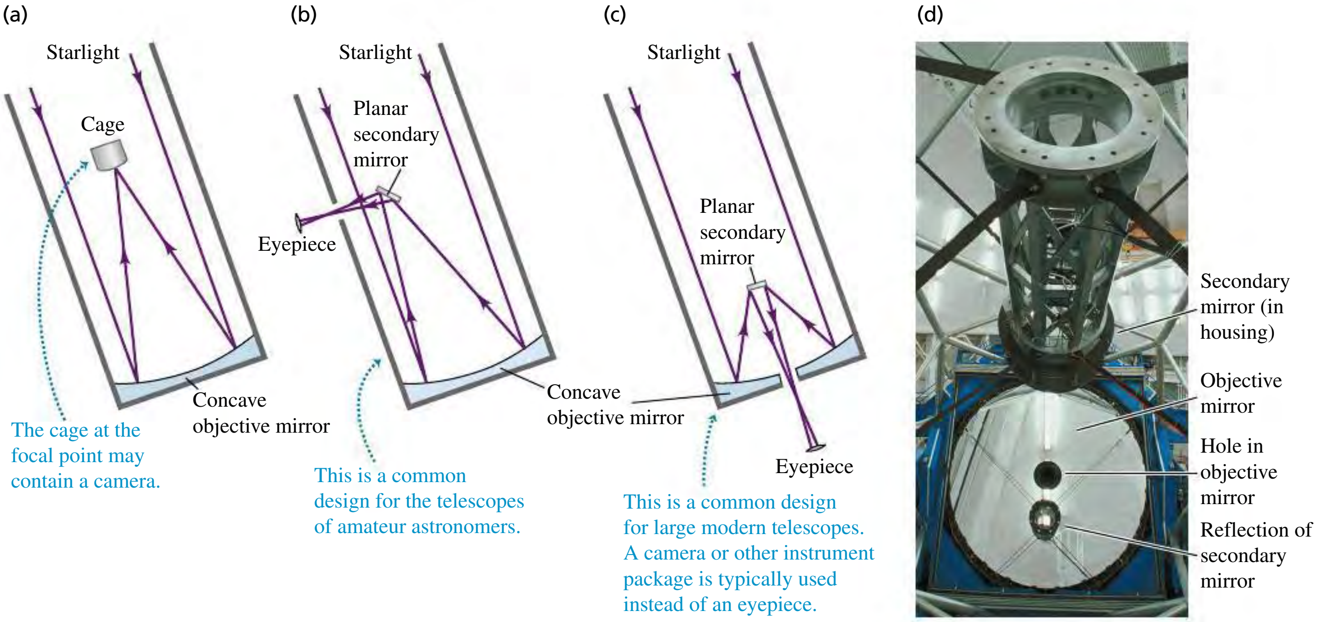
and the angular magnification M is

$$M = \frac{\theta'}{\theta} = -\frac{y'/f_2}{y'/f_1} = -\frac{f_1}{f_2} \qquad \text{(angular magnification for a telescope)} \qquad (34.25)$$

The angular magnification M of a telescope is equal to the ratio of the focal length of the objective to that of the eyepiece. The negative sign shows that the final image is inverted. Equation (34.25) shows that to achieve good angular magnification, a *telescope* should have a *long* objective focal length f_1. By contrast, Eq. (34.24) shows that a *microscope* should have a *short* objective focal length. However, a telescope objective with a long focal length should also have a large diameter D so that the f-number f_1/D will not be too large; as described in Section 34.5, a large f-number means a dim, low-intensity image. Telescopes typically do not have interchangeable objectives; instead, the magnification is varied by using different eyepieces with different focal lengths f_2. Just as for a microscope, smaller values of f_2 give larger angular magnifications.

An inverted image is no particular disadvantage for astronomical observations. When we use a telescope or binoculars—essentially a pair of telescopes mounted side by side—to view objects on the earth, though, we want the image to be right-side up. In prism binoculars, this is accomplished by reflecting the light several times along the path from the objective to the eyepiece. The combined effect of the reflections is to flip the image both horizontally and vertically. Binoculars are usually described by two numbers separated by a multiplication sign, such as 7×50. The first number is the angular magnification M, and the second is the diameter of the objective lenses (in millimeters). The diameter helps to determine the light-gathering capacity of the objective lenses and thus the brightness of the image.

34.54 (a), (b), (c) Three designs for reflecting telescopes. (d) This photo shows the interior of the Gemini North telescope, which uses the design shown in (c). The objective mirror is 8 meters in diameter.



In the *reflecting telescope* (**Fig. 34.54a**) the objective lens is replaced by a concave mirror. In large telescopes this scheme has many advantages. Mirrors are inherently free of chromatic aberrations (dependence of focal length on wavelength), and spherical aberrations (associated with the paraxial approximation) are easier to correct than with a lens. The reflecting surface is sometimes nonspherical. The material of the mirror need not be transparent, and it can be made more rigid than a lens, which has to be supported only at its edges.

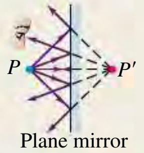
The largest reflecting telescope in the world, the Gran Telescopio Canarias in the Canary Islands, has an objective mirror of overall diameter 10.4 m made up of 36 separate hexagonal reflecting elements.

One challenge in designing reflecting telescopes is that the image is formed in front of the objective mirror, in a region traversed by incoming rays. Isaac Newton devised one solution to this problem. A flat secondary mirror oriented at 45° to the optic axis causes the image to be formed in a hole on the side of the telescope, where it can be magnified with an eyepiece (Fig. 34.54b). Another solution uses a secondary mirror that causes the focused light to pass through a hole in the objective mirror (Fig. 34.54c). Large research telescopes, as well as many amateur telescopes, use this design (Fig. 34.54d).

Like a microscope, when a telescope is used for photography the eyepiece is removed and an electronic sensor is placed at the position of the real image formed by the objective. (Some long-focal-length "lenses" for photography are actually reflecting telescopes used in this way.) Most telescopes used for astronomical research are never used with an eyepiece.

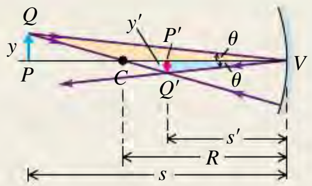
TEST YOUR UNDERSTANDING OF SECTION 34.8 Which gives a lateral magnification of greater absolute value? (i) The objective lens in a microscope (Fig. 34.52); (ii) the objective lens in a refracting telescope (Fig. 34.53); or (iii) not enough information is given to decide. ❙

Reflection or refraction at a plane surface: When rays diverge from an object point P and are reflected or refracted, the directions of the outgoing rays are the same as though they had diverged from a point P' called the image point. If they actually converge at P' and diverge again beyond it, P' is a real image of P; if they only appear to have diverged from P', it is a virtual image. Images can be either erect or inverted.

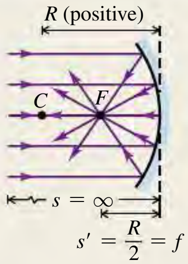


Lateral magnification: The lateral magnification m in any reflecting or refracting situation is defined as the ratio of image height y' to object height y. When m is positive, the image is erect; when m is negative, the image is inverted.

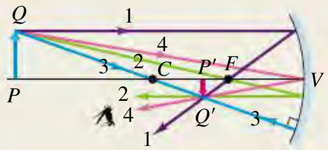
$$m = \frac{y'}{y} \qquad (34.2)$$



Focal point and focal length: The focal point of a mirror is the point where parallel rays converge after reflection from a concave mirror, or the point from which they appear to diverge after reflection from a convex mirror. Rays diverging from the focal point of a concave mirror are parallel after reflection; rays converging toward the focal point of a convex mirror are parallel after reflection. The distance from the focal point to the vertex is called the focal length, denoted as f. The focal points of a lens are defined similarly.



Relating object and image distances: The formulas for object distance s and image distance s' for plane and spherical mirrors and single refracting surfaces are summarized in the table. The equation for a plane surface can be obtained from the corresponding equation for a spherical surface by setting $R = \infty$. (See Examples 34.1–34.7.)



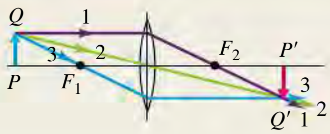
	Plane Mirror	Spherical Mirror	Plane Refracting Surface	Spherical Refracting Surface
Object and image distances	$\frac{1}{s} + \frac{1}{s'} = 0$	$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} = \frac{1}{f}$	$\frac{n_a}{s} + \frac{n_b}{s'} = 0$	$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R}$
Lateral magnification	$m = -\frac{s'}{s} = 1$	$m = -\frac{s'}{s}$	$m = -\frac{n_a s'}{n_b s} = 1$	$m = -\frac{n_a s'}{n_b s}$

Object–image relationships derived in this chapter are valid for only rays close to and nearly parallel to the optic axis; these are called paraxial rays. Nonparaxial rays do not converge precisely to an image point. This effect is called spherical aberration.

Thin lenses: The object–image relationship, given by Eq. (34.16), is the same for a thin lens as for a spherical mirror. Equation (34.19), the lensmaker's equation, relates the focal length of a lens to its index of refraction and the radii of curvature of its surfaces. (See Examples 34.8–34.11.)

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad (34.16)$$

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad (34.19)$$

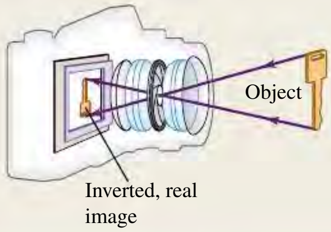


Sign rules: The following sign rules are used with all plane and spherical reflecting and refracting surfaces:

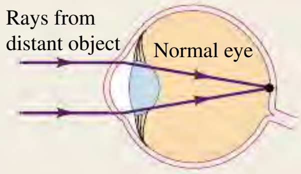
- $s > 0$ when the object is on the incoming side of the surface (a real object); $s < 0$ otherwise.
- $s' > 0$ when the image is on the outgoing side of the surface (a real image); $s' < 0$ otherwise.
- $R > 0$ when the center of curvature is on the outgoing side of the surface; $R < 0$ otherwise.
- $m > 0$ when the image is erect; $m < 0$ when inverted.

Cameras: A camera forms a real, inverted, reduced image of the object being photographed on a light-sensitive surface. The amount of light striking this surface is controlled by the shutter speed and the aperture. The intensity of this light is inversely proportional to the square of the f-number of the lens. (See Example 34.12.)

$$f\text{-number} = \frac{\text{Focal length}}{\text{Aperture diameter}} = \frac{f}{D} \quad (34.20)$$

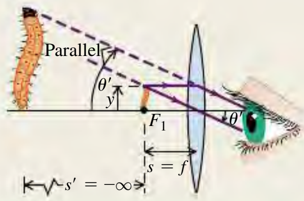


The eye: In the eye, refraction at the surface of the cornea forms a real image on the retina. Adjustment for various object distances is made by squeezing the lens, thereby making it bulge and decreasing its focal length. A nearsighted eye is too long for its lens; a farsighted eye is too short. The power of a corrective lens, in diopters, is the reciprocal of the focal length in meters. (See Examples 34.13 and 34.14.)

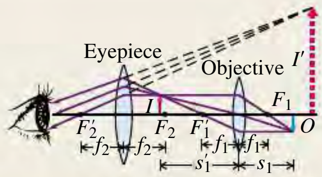


The simple magnifier: The simple magnifier creates a virtual image whose angular size θ' is larger than the angular size θ of the object itself at a distance of 25 cm, the nominal closest distance for comfortable viewing. The angular magnification M of a simple magnifier is the ratio of the angular size of the virtual image to that of the object at this distance.

$$M = \frac{\theta'}{\theta} = \frac{25\text{ cm}}{f} \quad (34.22)$$



Microscopes and telescopes: In a compound microscope, the objective lens forms a first image in the barrel of the instrument, and the eyepiece forms a final virtual image, often at infinity, of the first image. The telescope operates on the same principle, but the object is far away. In a reflecting telescope, the objective lens is replaced by a concave mirror, which eliminates chromatic aberrations.



BRIDGING PROBLEM IMAGE FORMATION BY A WINE GOBLET

SOLUTION

A thick-walled wine goblet can be considered to be a hollow glass sphere with an outer radius of 4.00 cm and an inner radius of 3.40 cm. The index of refraction of the goblet glass is 1.50. (a) A beam of parallel light rays enters the side of the empty goblet along a horizontal radius. Where, if anywhere, will an image be formed? (b) The goblet is filled with white wine ($n = 1.37$). Where is the image formed?

SOLUTION GUIDE

IDENTIFY and SET UP

1. The goblet is *not* a thin lens, so you cannot use the thin-lens formula. Instead, you must think of the inner and outer surfaces of the goblet walls as spherical refracting surfaces. The image formed by one surface serves as the object for the next surface. Draw a diagram that shows the goblet and the light rays that enter it.
2. Choose the appropriate equation that relates the image and object distances for a spherical refracting surface.

EXECUTE

3. For the empty goblet, each refracting surface has glass on one side and air on the other. Find the position of the image formed by the first surface, the outer wall of the goblet. Use this as the object for the second surface (the inner wall of the same side of the goblet) and find the position of the second image. (*Hint:* Be sure to account for the thickness of the goblet wall.)
4. Continue the process of step 3. Consider the refractions at the inner and outer surfaces of the glass on the opposite side of the goblet, and find the position of the final image. (*Hint:* Be sure to account for the distance between the two sides of the goblet.)
5. Repeat steps 3 and 4 for the case in which the goblet is filled with wine.

EVALUATE

6. Are the images real or virtual? How can you tell?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q34.1 A spherical mirror is cut in half horizontally. Will an image be formed by the bottom half of the mirror? If so, where will the image be formed?

Q34.2 For the situation shown in Fig. 34.3, is the image distance s' positive or negative? Is the image real or virtual? Explain your answers.

Q34.3 The laws of optics also apply to electromagnetic waves invisible to the eye. A satellite TV dish is used to detect radio waves coming from orbiting satellites. Why is a curved reflecting surface (a "dish") used? The dish is always concave, never convex; why? The actual radio receiver is placed on an arm and suspended in front of the dish. How far in front of the dish should it be placed?

Q34.4 Explain why the focal length of a *plane* mirror is infinite, and explain what it means for the focal point to be at infinity.

Q34.5 If a spherical mirror is immersed in water, does its focal length change? Explain.

Q34.6 For what range of object positions does a concave spherical mirror form a real image? What about a convex spherical mirror?

Q34.7 When a room has mirrors on two opposite walls, an infinite series of reflections can be seen. Discuss this phenomenon in terms of images. Why do the distant images appear fainter?

Q34.8 For a spherical mirror, if $s = f$, then $s' = \infty$, and the lateral magnification m is infinite. Does this make sense? If so, what does it mean?

Q34.9 You may have noticed a small convex mirror next to your bank's ATM. Why is this mirror convex, as opposed to flat or concave? What considerations determine its radius of curvature?

Q34.10 A student claims that she can start a fire on a sunny day using just the sun's rays and a concave mirror. How is this done? Is the concept of image relevant? Can she do the same thing with a convex mirror? Explain.

Q34.11 A person looks at his reflection in the concave side of a shiny spoon. Is it right side up or inverted? Does it matter how far his face is from the spoon? What if he looks in the convex side? (Try this yourself!)

Q34.12 In Example 34.4 (Section 34.2), there appears to be an ambiguity for the case $s = 10$ cm as to whether s' is $+\infty$ or $-\infty$ and whether the image is erect or inverted. How is this resolved? Or is it?

Q34.13 Suppose that in the situation of Example 34.7 of Section 34.3 (see Fig. 34.26) a vertical arrow 2.00 m tall is painted on the side of the pool beneath the water line. According to the calculations in the example, this arrow would appear to the person shown in Fig. 34.26 to be 1.50 m long. But the discussion following Eq. (34.13) states that the magnification for a plane refracting surface is $m = 1$, which suggests that the arrow would appear to the person to be 2.00 m long. How can you resolve this apparent contradiction?

Q34.14 The bottom of the passenger-side mirror on your car notes, "Objects in mirror are closer than they appear." Is this true? Why?

Q34.15 How could you very quickly make an approximate measurement of the focal length of a converging lens? Could the same method be applied if you wished to use a diverging lens? Explain.

Q34.16 The focal length of a simple lens depends on the color (wavelength) of light passing through it. Why? Is it possible for a lens to have a positive focal length for some colors and negative for others? Explain.

Q34.17 When a converging lens is immersed in water, does its focal length increase or decrease in comparison with the value in air? Explain.

Q34.18 A spherical air bubble in water can function as a lens. Is it a converging or diverging lens? How is its focal length related to its radius?

Q34.19 Can an image formed by one reflecting or refracting surface serve as an object for a second reflection or refraction? Does it matter whether the first image is real or virtual? Explain.

Q34.20 If a piece of photographic film is placed at the location of a real image, the film will record the image. Can this be done with a virtual image? How might one record a virtual image?

Q34.21 According to the discussion in Section 34.2, light rays are reversible. Are the formulas in the table in this chapter's Summary still valid if object and image are interchanged? What does reversibility imply with respect to the *forms* of the various formulas?

Q34.22 You've entered a survival contest that will include building a crude telescope. You are given a large box of lenses. Which two lenses do you pick? How do you quickly identify them?

Q34.23 **BIO** You can't see clearly underwater with the naked eye, but you *can* if you wear a face mask or goggles (with air between your eyes and the mask or goggles). Why is there a difference? Could you instead wear eyeglasses (with water between your eyes and the eyeglasses) in order to see underwater? If so, should the lenses be converging or diverging? Explain.

Q34.24 You take a lens and mask it so that light can pass through only the bottom half of the lens. How does the image formed by the masked lens compare to the image formed before masking?

EXERCISES

Section 34.1 Reflection and Refraction at a Plane Surface

34.1 • A candle 4.85 cm tall is 39.2 cm to the left of a plane mirror. Where is the image formed by the mirror, and what is the height of this image?

34.2 • The image of a tree just covers the length of a plane mirror 4.00 cm tall when the mirror is held 35.0 cm from the eye. The tree is 28.0 m from the mirror. What is its height?

34.3 • A pencil that is 9.0 cm long is held perpendicular to the surface of a plane mirror with the tip of the pencil lead 12.0 cm from the mirror surface and the end of the eraser 21.0 cm from the mirror surface. What is the length of the image of the pencil that is formed by the mirror? Which end of the image is closer to the mirror surface: the tip of the lead or the end of the eraser?

Section 34.2 Reflection at a Spherical Surface

34.4 • A concave mirror has a radius of curvature of 34.0 cm. (a) What is its focal length? (b) If the mirror is immersed in water (refractive index 1.33), what is its focal length?

34.5 • An object 0.600 cm tall is placed 16.5 cm to the left of the vertex of a concave spherical mirror having a radius of curvature of 22.0 cm. (a) Draw a principal-ray diagram showing the formation of the image. (b) Determine the position, size, orientation, and nature (real or virtual) of the image.

34.6 • Repeat Exercise 34.5 for the case in which the mirror is convex.

34.7 •• The diameter of Mars is 6794 km, and its minimum distance from the earth is 5.58×10^7 km. When Mars is at this distance, find the diameter of the image of Mars formed by a spherical, concave telescope mirror with a focal length of 1.75 m.

34.8 •• An object is 18.0 cm from the center of a spherical silvered-glass Christmas tree ornament 6.00 cm in diameter. What are the position and magnification of its image?

34.9 • A coin is placed next to the convex side of a thin spherical glass shell having a radius of curvature of 18.0 cm. Reflection from the surface of the shell forms an image of the 1.5-cm-tall coin that is 6.00 cm behind the glass shell. Where is the coin located? Determine the size, orientation, and nature (real or virtual) of the image.

34.10 • You hold a spherical salad bowl 60 cm in front of your face with the bottom of the bowl facing you. The bowl is made of polished metal with a 35-cm radius of curvature. (a) Where is the image of your 5.0-cm-tall nose located? (b) What are the image's size, orientation, and nature (real or virtual)?

34.11 • A spherical, concave shaving mirror has a radius of curvature of 32.0 cm. (a) What is the magnification of a person's face when it is 12.0 cm to the left of the vertex of the mirror? (b) Where is the image? Is the image real or virtual? (c) Draw a principal-ray diagram showing the formation of the image.

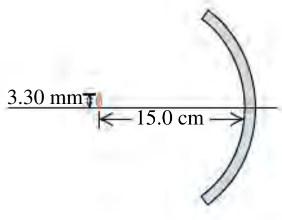
34.12 • For a concave spherical mirror that has focal length $f = +18.0$ cm, what is the distance of an object from the mirror's vertex if the image is real and has the same height as the object?

34.13 • **Dental Mirror.** A dentist uses a curved mirror to view teeth on the upper side of the mouth. Suppose she wants an erect image with a magnification of 2.00 when the mirror is 1.25 cm from a tooth. (Treat this problem as though the object and image lie along a straight line.) (a) What kind of mirror (concave or convex) is needed? Use a ray diagram to decide, without performing any calculations. (b) What must be the focal length and radius of curvature of this mirror? (c) Draw a principal-ray diagram to check your answer in part (b).

34.14 • For a convex spherical mirror that has focal length $f = -12.0$ cm, what is the distance of an object from the mirror's vertex if the height of the image is half the height of the object?

34.15 • The thin glass shell shown in **Fig. E34.15** has a spherical shape with a radius of curvature of 12.0 cm, and both of its surfaces can act as mirrors. A seed 3.30 mm high is placed 15.0 cm from the center of the mirror along the optic axis, as shown in the figure. (a) Calculate the location and height of the image of this seed. (b) Suppose now that the shell is reversed. Find the location and height of the seed's image.

Figure **E34.15**



Section 34.3 Refraction at a Spherical Surface

34.16 •• A tank whose bottom is a mirror is filled with water to a depth of 20.0 cm. A small fish floats motionless 7.0 cm under the surface of the water. (a) What is the apparent depth of the fish when viewed at normal incidence? (b) What is the apparent depth of the image of the fish when viewed at normal incidence?

34.17 • A speck of dirt is embedded 3.50 cm below the surface of a sheet of ice $(n = 1.309)$. What is its apparent depth when viewed at normal incidence?

34.18 • A transparent liquid fills a cylindrical tank to a depth of 3.60 m. There is air above the liquid. You look at normal incidence at a small pebble at the bottom of the tank. The apparent depth of the pebble below the liquid's surface is 2.45 m. What is the refractive index of this liquid?

34.19 • A person swimming 0.80 m below the surface of the water in a swimming pool looks at the diving board that is directly overhead and sees the image of the board that is formed by refraction at the surface of the water. This image is a height of 5.20 m above the swimmer. What is the actual height of the diving board above the surface of the water?

34.20 • A person is lying on a diving board 3.00 m above the surface of the water in a swimming pool. She looks at a penny that is on the bottom of the pool directly below her. To her, the penny appears to be a distance of 7.00 m from her. What is the depth of the water at this point?

34.21 •• **A Spherical Fish Bowl.** A small tropical fish is at the center of a water-filled, spherical fish bowl 28.0 cm in diameter. (a) Find the apparent position and magnification of the fish to an observer outside the bowl. The effect of the thin walls of the bowl may be ignored. (b) A friend advised the owner of the bowl to keep it out of direct sunlight to avoid blinding the fish, which might swim into the focal point of the parallel rays from the sun. Is the focal point actually within the bowl?

34.22 • The left end of a long glass rod 6.00 cm in diameter has a convex hemispherical surface 3.00 cm in radius. The refractive index of the glass is 1.60. Determine the position of the image if an object is placed in air on the axis of the rod at the following distances to the left of the vertex of the curved end: (a) infinitely far, (b) 12.0 cm; (c) 2.00 cm.

34.23 •• The glass rod of Exercise 34.22 is immersed in oil $(n = 1.45)$. An object placed to the left of the rod on the rod's axis is to be imaged 1.20 m inside the rod. How far from the left end of the rod must the object be located to form the image?

34.24 •• The left end of a long glass rod 8.00 cm in diameter, with an index of refraction of 1.60, is ground and polished to a convex hemispherical surface with a radius of 4.00 cm. An object in the form of an arrow 1.50 mm tall, at right angles to the axis of the rod, is located on the axis 24.0 cm to the left of the vertex of the convex surface. Find the position and height of the image of the arrow formed by paraxial rays incident on the convex surface. Is the image erect or inverted?

34.25 •• Repeat Exercise 34.24 for the case in which the end of the rod is ground to a *concave* hemispherical surface with radius 4.00 cm.

34.26 •• The glass rod of Exercise 34.25 is immersed in a liquid. An object 14.0 cm from the vertex of the left end of the rod and on its axis is imaged at a point 9.00 cm from the vertex inside the liquid. What is the index of refraction of the liquid?

Section 34.4 Thin Lenses

34.27 • An insect 3.75 mm tall is placed 22.5 cm to the left of a thin planoconvex lens. The left surface of this lens is flat, the right surface has a radius of curvature of magnitude 13.0 cm, and the index of refraction of the lens material is 1.70. (a) Calculate the location and size of the image this lens forms of the insect.

Is it real or virtual? Erect or inverted? (b) Repeat part (a) if the lens is reversed.

34.28 • A lens forms an image of an object. The object is 16.0 cm from the lens. The image is 12.0 cm from the lens on the same side as the object. (a) What is the focal length of the lens? Is the lens converging or diverging? (b) If the object is 8.50 mm tall, how tall is the image? Is it erect or inverted? (c) Draw a principal-ray diagram.

34.29 • A converging meniscus lens (see Fig. 34.32a) with a refractive index of 1.52 has spherical surfaces whose radii are 7.00 cm and 4.00 cm. What is the position of the image if an object is placed 24.0 cm to the left of the lens? What is the magnification?

34.30 • A converging lens with a focal length of 70.0 cm forms an image of a 3.20-cm-tall real object that is to the left of the lens. The image is 4.50 cm tall and inverted. Where are the object and image located in relation to the lens? Is the image real or virtual?

34.31 •• A converging lens forms an image of an 8.00-mm-tall real object. The image is 12.0 cm to the left of the lens, 3.40 cm tall, and erect. What is the focal length of the lens? Where is the object located?

34.32 • A photographic slide is to the left of a lens. The lens projects an image of the slide onto a wall 6.00 m to the right of the slide. The image is 80.0 times the size of the slide. (a) How far is the slide from the lens? (b) Is the image erect or inverted? (c) What is the focal length of the lens? (d) Is the lens converging or diverging?

34.33 •• A double-convex thin lens has surfaces with equal radii of curvature of magnitude 2.50 cm. Using this lens, you observe that it forms an image of a very distant tree at a distance of 1.87 cm from the lens. What is the index of refraction of the lens?

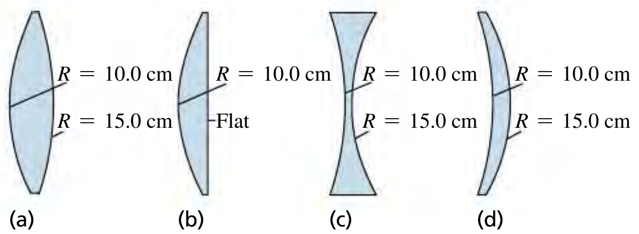
34.34 • A converging lens with a focal length of 9.00 cm forms an image of a 4.00-mm-tall real object that is to the left of the lens. The image is 1.30 cm tall and erect. Where are the object and image located? Is the image real or virtual?

34.35 • BIO **The Cornea As a Simple Lens.** The cornea behaves as a thin lens of focal length approximately 1.8 cm, although this varies a bit. The material of which it is made has an index of refraction of 1.38, and its front surface is convex, with a radius of curvature of 5.0 mm. (a) If this focal length is in air, what is the radius of curvature of the back side of the cornea? (b) The closest distance at which a typical person can focus on an object (called the near point) is about 25 cm, although this varies considerably with age. Where would the cornea focus the image of an 8.0-mm-tall object at the near point? (c) What is the height of the image in part (b)? Is this image real or virtual? Is it erect or inverted? (*Note:* The results obtained here are not strictly accurate because, on one side, the cornea has a fluid with a refractive index different from that of air.)

34.36 •• A lensmaker wants to make a magnifying glass from glass that has an index of refraction $n = 1.55$ and a focal length of 20.0 cm. If the two surfaces of the lens are to have equal radii, what should that radius be?

34.37 • For each thin lens shown in **Fig. E34.37**, calculate the location of the image of an object that is 18.0 cm to the left of the lens. The lens material has a refractive index of 1.50, and the radii of curvature shown are only the magnitudes.

Figure **E34.37**



34.38 • A converging lens with a focal length of 12.0 cm forms a virtual image 8.00 mm tall, 17.0 cm to the right of the lens. Determine the position and size of the object. Is the image erect or inverted? Are the object and image on the same side or opposite sides of the lens? Draw a principal-ray diagram for this situation.

34.39 • Repeat Exercise 34.38 for the case in which the lens is diverging, with a focal length of −48.0 cm.

34.40 • An object is 16.0 cm to the left of a lens. The lens forms an image 36.0 cm to the right of the lens. (a) What is the focal length of the lens? Is the lens converging or diverging? (b) If the object is 8.00 mm tall, how tall is the image? Is it erect or inverted? (c) Draw a principal-ray diagram.

34.41 •• **Combination of Lenses I.** A 1.20-cm-tall object is 50.0 cm to the left of a converging lens of focal length 40.0 cm. A second converging lens, this one having a focal length of 60.0 cm, is located 300.0 cm to the right of the first lens along the same optic axis. (a) Find the location and height of the image (call it I_1) formed by the lens with a focal length of 40.0 cm. (b) I_1 is now the object for the second lens. Find the location and height of the image produced by the second lens. This is the final image produced by the combination of lenses.

34.42 •• **Combination of Lenses II.** Repeat Exercise 34.41 using the same lenses except for the following changes: (a) The second lens is a *diverging* lens having a focal length of magnitude 60.0 cm. (b) The first lens is a *diverging* lens having a focal length of magnitude 40.0 cm. (c) Both lenses are diverging lenses having focal lengths of the same *magnitudes* as in Exercise 34.41.

34.43 •• **Combination of Lenses III.** Two thin lenses with a focal length of magnitude 12.0 cm, the first diverging and the second converging, are located 9.00 cm apart. An object 2.50 mm tall is placed 20.0 cm to the left of the first (diverging) lens. (a) How far from this first lens is the final image formed? (b) Is the final image real or virtual? (c) What is the height of the final image? Is it erect or inverted? (*Hint:* See the preceding two problems.)

34.44 • BIO **The Lens of the Eye.** The crystalline lens of the human eye is a double-convex lens made of material having an index of refraction of 1.44 (although this varies). Its focal length in air is about 8.0 mm, which also varies. We shall assume that the radii of curvature of its two surfaces have the same magnitude. (a) Find the radii of curvature of this lens. (b) If an object 16 cm tall is placed 30.0 cm from the eye lens, where would the lens focus it and how tall would the image be? Is this image real or virtual? Is it erect or inverted? (*Note:* The results obtained here are not strictly accurate because the lens is embedded in fluids having refractive indexes different from that of air.)

Section 34.5 Cameras

34.45 •• A camera lens has a focal length of 200 mm. How far from the lens should the subject for the photo be if the lens is 20.4 cm from the sensor?

34.46 • You wish to project the image of a slide on a screen 9.00 m from the lens of a slide projector. (a) If the slide is placed 15.0 cm from the lens, what focal length lens is required? (b) If the dimensions of the picture on a 35-mm color slide are 24 mm × 36 mm, what is the minimum size of the projector screen required to accommodate the image?

34.47 • When a camera is focused, the lens is moved away from or toward the digital image sensor. If you take a picture of your friend, who is standing 3.90 m from the lens, using a camera with a lens with an 85-mm focal length, how far from the sensor is the lens? Will the whole image of your friend, who is 175 cm tall, fit on a sensor that is 24 mm $\times$ 36 mm?

34.48 • **Zoom Lens.** Consider the simple model of the zoom lens shown in Fig. 34.43a. The converging lens has focal length $f_1 = 12$ cm, and the diverging lens has focal length $f_2 = -12$ cm. The lenses are separated by 4 cm as shown in Fig. 34.43a. (a) For a distant object, where is the image of the converging lens? (b) The image of the converging lens serves as the object for the diverging lens. What is the object distance for the diverging lens? (c) Where is the final image? Compare your answer to Fig. 34.43a. (d) Repeat parts (a), (b), and (c) for the situation shown in Fig. 34.43b, in which the lenses are separated by 8 cm.

34.49 •• A camera lens has a focal length of 180.0 mm and an aperture diameter of 16.36 mm. (a) What is the f-number of the lens? (b) If the correct exposure of a certain scene is $\frac{1}{30}$ s at $f/11$, what is the correct exposure at $f/2.8$?

Section 34.6 The Eye

34.50 •• BIO **Curvature of the Cornea.** In a simplified model of the human eye, the aqueous and vitreous humors and the lens all have a refractive index of 1.40, and all the bending occurs at the cornea, whose vertex is 2.60 cm from the retina. What should be the radius of curvature of the cornea such that the image of an object 40.0 cm from the cornea's vertex is focused on the retina?

34.51 •• BIO (a) Where is the near point of an eye for which a contact lens with a power of +2.75 diopters is prescribed? (b) Where is the far point of an eye for which a contact lens with a power of -1.30 diopters is prescribed for distant vision?

34.52 • BIO **Contact Lenses.** Contact lenses are placed right on the eyeball, so the distance from the eye to an object (or image) is the same as the distance from the lens to that object (or image). A certain person can see distant objects well, but his near point is 45.0 cm from his eyes instead of the usual 25.0 cm. (a) Is this person nearsighted or farsighted? (b) What type of lens (converging or diverging) is needed to correct his vision? (c) If the correcting lenses will be contact lenses, what focal length lens is needed and what is its power in diopters?

34.53 •• BIO **Ordinary Glasses.** Ordinary glasses are worn in front of the eye and usually 2.0 cm in front of the eyeball. Suppose that the person in Exercise 34.52 prefers ordinary glasses to contact lenses. What focal length lenses are needed to correct his vision, and what is their power in diopters?

34.54 • BIO A person can see clearly up close but cannot focus on objects beyond 75.0 cm. She opts for contact lenses to correct her vision. (a) Is she nearsighted or farsighted? (b) What type of lens (converging or diverging) is needed to correct her vision? (c) What focal length contact lens is needed, and what is its power in diopters?

34.55 •• BIO If the person in Exercise 34.54 chooses ordinary glasses over contact lenses, what power lens (in diopters) does she need to correct her vision if the lenses are 2.0 cm in front of the eye?

Section 34.7 The Magnifier

34.56 •• A thin lens with a focal length of 6.00 cm is used as a simple magnifier. (a) What angular magnification is obtainable with the lens if the object is at the focal point? (b) When an object is examined through the lens, how close can it be brought to the lens? Assume that the image viewed by the eye is at the near point, 25.0 cm from the eye, and that the lens is very close to the eye.

34.57 • The focal length of a simple magnifier is 8.00 cm. Assume the magnifier is a thin lens placed very close to the eye. (a) How far in front of the magnifier should an object be placed if the image is formed at the observer's near point, 25.0 cm in front of her eye? (b) If the object is 1.00 mm high, what is the height of its image formed by the magnifier?

34.58 • You want to view through a magnifier an insect that is 2.00 mm long. If the insect is to be at the focal point of the magnifier, what focal length will give the image of the insect an angular size of 0.032 radian?

Section 34.8 Microscopes and Telescopes

34.59 •• The focal length of the eyepiece of a certain microscope is 18.0 mm. The focal length of the objective is 8.00 mm. The distance between objective and eyepiece is 19.7 cm. The final image formed by the eyepiece is at infinity. Treat all lenses as thin. (a) What is the distance from the objective to the object being viewed? (b) What is the magnitude of the linear magnification produced by the objective? (c) What is the overall angular magnification of the microscope?

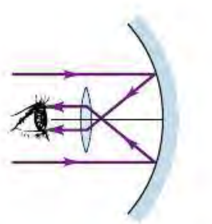
34.60 •• **Resolution of a Microscope.** The image formed by a microscope objective with a focal length of 5.00 mm is 160 mm from its second focal point. The eyepiece has a focal length of 26.0 mm. (a) What is the angular magnification of the microscope? (b) The unaided eye can distinguish two points at its near point as separate if they are about 0.10 mm apart. What is the minimum separation between two points that can be observed (or resolved) through this microscope?

34.61 •• A telescope is constructed from two lenses with focal lengths of 95.0 cm and 15.0 cm, the 95.0-cm lens being used as the objective. Both the object being viewed and the final image are at infinity. (a) Find the angular magnification for the telescope. (b) Find the height of the image formed by the objective of a building 60.0 m tall, 3.00 km away. (c) What is the angular size of the final image as viewed by an eye very close to the eyepiece?

34.62 •• The eyepiece of a refracting telescope (see Fig. 34.53) has a focal length of 9.00 cm. The distance between objective and eyepiece is 1.20 m, and the final image is at infinity. What is the angular magnification of the telescope?

34.63 •• A reflecting telescope (**Fig. E34.63**) is to be made by using a spherical mirror with a radius of curvature of 1.30 m and an eyepiece with a focal length of 1.10 cm. The final image is at infinity. (a) What should the distance between the eyepiece and the mirror vertex be if the object is taken to be at infinity? (b) What will the angular magnification be?

Figure **E34.63**



PROBLEMS

34.64 •• What is the size of the smallest vertical plane mirror in which a woman of height h can see her full-length image?

34.65 • If you run away from a plane mirror at 3.60 m/s, at what speed does your image move away from you?

34.66 • Where must you place an object in front of a concave mirror with radius R so that the image is erect and $2\frac{1}{2}$ times the size of the object? Where is the image?

34.67 •• A concave mirror is to form an image of the filament of a headlight lamp on a screen 8.00 m from the mirror. The filament is 6.00 mm tall, and the image is to be 24.0 cm tall. (a) How far in front of the vertex of the mirror should the filament be placed? (b) What should be the radius of curvature of the mirror?

34.68 • A light bulb is 3.00 m from a wall. You are to use a concave mirror to project an image of the bulb on the wall, with the image 3.50 times the size of the object. How far should the mirror be from the wall? What should its radius of curvature be?

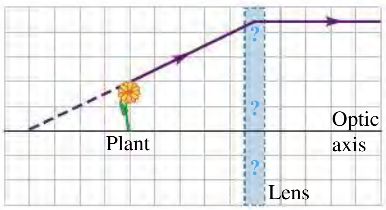
34.69 •• **CP CALC** You are in your car driving on a highway at 25 m/s when you glance in the passenger-side mirror (a convex mirror with radius of curvature 150 cm) and notice a truck approaching. If the image of the truck is approaching the vertex of the mirror at a speed of 1.9 m/s when the truck is 2.0 m from the mirror, what is the speed of the truck relative to the highway?

34.70 •• A layer of benzene $(n = 1.50)$ that is 4.20 cm deep floats on water $(n = 1.33)$ that is 5.70 cm deep. What is the apparent distance from the upper benzene surface to the bottom of the water when you view these layers at normal incidence?

34.71 •• **Rear-View Mirror.** A mirror on the passenger side of your car is convex and has a radius of curvature with magnitude 18.0 cm. (a) Another car is behind your car, 9.00 m from the mirror, and this car is viewed in the mirror by your passenger. If this car is 1.5 m tall, what is the height of the image? (b) The mirror has a warning attached that objects viewed in it are closer than they appear. Why is this so?

34.72 •• **Figure P34.72** shows a small plant near a thin lens. The ray shown is one of the principal rays for the lens. Each square is 2.0 cm along the horizontal direction, but the vertical direction is not to the same scale. Use information from the diagram for the following: (a) Using only the ray shown, decide what type of lens (converging or diverging) this is. (b) What is the focal length of the lens? (c) Locate the image by drawing the other two principal rays. (d) Calculate where the image should be, and compare this result with the graphical solution in part (c).

Figure **P34.72**



34.73 •• **Pinhole Camera.** A pinhole camera is just a rectangular box with a tiny hole in one face. The film is on the face opposite this hole, and that is where the image is formed. The camera forms an image *without* a lens. (a) Make a clear ray diagram to show how a pinhole camera can form an image on the film without using a lens. (*Hint:* Put an object outside the hole, and then draw rays passing through the hole to the opposite side of the box.) (b) A certain pinhole camera is a box that is 25 cm square and 20.0 cm deep, with the hole in the middle of one of the 25 cm × 25 cm faces. If this camera is used to photograph a fierce chicken that is 18 cm high and 1.5 m in front of the camera, how large is the image of this bird on the film? What is the lateral magnification of this camera?

34.74 ••• A microscope is focused on the upper surface of a glass plate. A second plate is then placed over the first. To focus on the bottom surface of the second plate, the microscope must be raised 0.780 mm. To focus on the upper surface, it must be raised another 2.10 mm. Find the index of refraction of the second plate.

34.75 •• What should be the index of refraction of a transparent sphere in order for paraxial rays from an infinitely distant object to be brought to a focus at the vertex of the surface opposite the point of incidence?

34.76 •• **A Glass Rod.** Both ends of a glass rod with index of refraction 1.60 are ground and polished to convex hemispherical surfaces. The radius of curvature at the left end is 6.00 cm, and the radius of curvature at the right end is 12.0 cm. The length of the rod between vertices is 40.0 cm. The object for the surface at the left end is an arrow that lies 23.0 cm to the left of the vertex of this surface. The arrow is 1.50 mm tall and at right angles to the axis. (a) What constitutes the object for the surface at the right end of the rod? (b) What is the object distance for this surface? (c) Is the object for this surface real or virtual? (d) What is the position of the final image? (e) Is the final image real or virtual? Is it erect or inverted with respect to the original object? (f) What is the height of the final image?

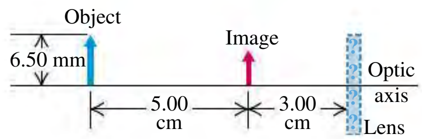
34.77 •• (a) You want to use a lens with a focal length of 35.0 cm to produce a real image of an object, with the height of the image twice the height of the object. What kind of lens do you need, and where should the object be placed? (b) Suppose you want a virtual image of the same object, with the same magnification—what kind of lens do you need, and where should the object be placed?

34.78 •• **Autocollimation.** You place an object alongside a white screen, and a plane mirror is 60.0 cm to the right of the object and the screen, with the surface of the mirror tilted slightly from the perpendicular to the line from object to mirror. You then place a converging lens between the object and the mirror. Light from the object passes through the lens, reflects from the mirror, and passes back through the lens; the image is projected onto the screen. You adjust the distance of the lens from the object until a sharp image of the object is focused on the screen. The lens is then 22.0 cm from the object. Because the screen is alongside the object, the distance from object to lens is the same as the distance from screen to lens. (a) Draw a sketch that shows the locations of the object, lens, plane mirror, and screen. (b) What is the focal length of the lens?

34.79 •• A lens forms a real image that is 214 cm away from the object and $1\frac{2}{3}$ times its height. What kind of lens is this, and what is its focal length?

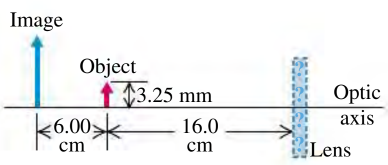
34.80 • **Figure P34.80** shows an object and its image formed by a thin lens. (a) What is the focal length of the lens, and what type of lens (converging or diverging) is it? (b) What is the height of the image? Is it real or virtual?

Figure **P34.80**



34.81 • **Figure P34.81** shows an object and its image formed by a thin lens. (a) What is the focal length of the lens, and what type of lens (converging or diverging) is it? (b) What is the height of the image? Is it real or virtual?

Figure **P34.81**



34.82 ••• A transparent rod 30.0 cm long is cut flat at one end and rounded to a hemispherical surface of radius 10.0 cm at the other end. A small object is embedded within the rod along its axis and halfway between its ends, 15.0 cm from the flat end and 15.0 cm from the vertex of the curved end. When the rod is viewed from its flat end, the apparent depth of the object is 8.20 cm from the flat end. What is its apparent depth when the rod is viewed from its curved end?

34.83 • BIO Focus of the Eye. The cornea of the eye has a radius of curvature of approximately 0.50 cm, and the aqueous humor behind it has an index of refraction of 1.35. The thickness of the cornea itself is small enough that we shall neglect it. The depth of a typical human eye is around 25 mm. (a) What would have to be the radius of curvature of the cornea so that it alone would focus the image of a distant mountain on the retina, which is at the back of the eye opposite the cornea? (b) If the cornea focused the mountain correctly on the retina as described in part (a), would it also focus the text from a computer screen on the retina if that screen were 25 cm in front of the eye? If not, where would it focus that text: in front of or behind the retina? (c) Given that the cornea has a radius of curvature of about 5.0 mm, where does it actually focus the mountain? Is this in front of or behind the retina? Does this help you see why the eye needs help from a lens to complete the task of focusing?

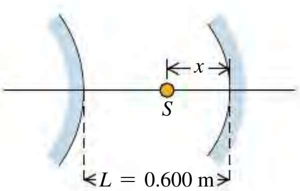
34.84 • The radii of curvature of the surfaces of a thin converging meniscus lens are $R_1 = +12.0$ cm and $R_2 = +28.0$ cm. The index of refraction is 1.60. (a) Compute the position and size of the image of an object in the form of an arrow 5.00 mm tall, perpendicular to the lens axis, 45.0 cm to the left of the lens. (b) A second converging lens with the same focal length is placed 3.15 m to the right of the first. Find the position and size of the final image. Is the final image erect or inverted with respect to the original object? (c) Repeat part (b) except with the second lens 45.0 cm to the right of the first.

34.85 • An object to the left of a lens is imaged by the lens on a screen 30.0 cm to the right of the lens. When the lens is moved 4.00 cm to the right, the screen must be moved 4.00 cm to the left to refocus the image. Determine the focal length of the lens.

34.86 • An object is placed 22.0 cm from a screen. (a) At what two points between object and screen may a converging lens with a 3.00-cm focal length be placed to obtain an image on the screen? (b) What is the magnification of the image for each position of the lens?

34.87 •• A convex mirror and a concave mirror are placed on the same optic axis, separated by a distance $L = 0.600$ m. The radius of curvature of each mirror has a magnitude of 0.360 m. A light source is located a distance x from the concave mirror, as shown in **Fig. P34.87**. (a) What distance x will result in the rays from the source returning to the source after reflecting first from the convex mirror and then from the concave mirror? (b) Repeat part (a), but now let the rays reflect first from the concave mirror and then from the convex one.

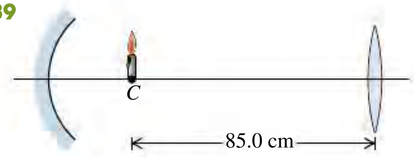
Figure **P34.87**



34.88 •• A screen is placed a distance d to the right of an object. A converging lens with focal length f is placed between the object and the screen. In terms of f, what is the smallest value d can have for an image to be in focus on the screen?

34.89 •• As shown in **Fig. P34.89**, the candle is at the center of curvature of the concave mirror, whose focal length is 10.0 cm. The converging lens has a focal length of 32.0 cm and is 85.0 cm to the right of the candle. The candle is viewed looking through the lens from the right. The lens forms two images of the candle. The first is formed by light passing directly through the lens. The second image is formed from the light that goes from the candle to the mirror, is reflected, and then passes through the lens. (a) For each of these two images, draw a principal-ray diagram that locates the image. (b) For each image, answer the following questions: (i) Where is the image? (ii) Is the image real or virtual? (iii) Is the image erect or inverted with respect to the original object?

Figure **P34.89**



34.90 •• Two Lenses in Contact. (a) Prove that when two thin lenses with focal lengths f_1 and f_2 are placed *in contact*, the focal length f of the combination is given by the relationship

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$$

(b) A converging meniscus lens (see Fig. 34.32a) has an index of refraction of 1.55 and radii of curvature for its surfaces of magnitudes 4.50 cm and 9.00 cm. The concave surface is placed upward and filled with carbon tetrachloride (CCl_4), which has $n = 1.46$. What is the focal length of the CCl_4–glass combination?

34.91 ••• When an object is placed at the proper distance to the left of a converging lens, the image is focused on a screen 30.0 cm to the right of the lens. A diverging lens is now placed 15.0 cm to the right of the converging lens, and it is found that the screen must be moved 19.2 cm farther to the right to obtain a sharp image. What is the focal length of the diverging lens?

34.92 •• (a) Repeat the derivation of Eq. (34.19) for the case in which the lens is totally immersed in a liquid of refractive index n_{liq}. (b) A lens is made of glass that has refractive index 1.60. In air, the lens has focal length +18.0 cm. What is the focal length of this lens if it is totally immersed in a liquid that has refractive index 1.42?

34.93 ••• A convex spherical mirror with a focal length of magnitude 24.0 cm is placed 20.0 cm to the left of a plane mirror. An object 0.250 cm tall is placed midway between the surface of the plane mirror and the vertex of the spherical mirror. The spherical mirror forms multiple images of the object. Where are the two images of the object formed by the spherical mirror that are closest to the spherical mirror, and how tall is each image?

34.94 •• BIO What Is the Smallest Thing We Can See? The smallest object we can resolve with our eye is limited by the size of the light receptor cells in the retina. In order for us to distinguish any detail in an object, its image cannot be any smaller than a single retinal cell. Although the size depends on the type of cell (rod or cone), a diameter of a few microns (μm) is typical near the center of the eye. We shall model the eye as a sphere 2.50 cm in diameter with a single thin lens at the front and the retina at the rear, with light receptor cells 5.0 μm in diameter. (a) What is the smallest object you can resolve at a near point of 25 cm? (b) What angle is subtended by this object at the eye? Express your answer in units of minutes ($1° = 60$ min), and compare it with the typical experimental value of about 1.0 min. (*Note:* There are other limitations, such as the bending of light as it passes through the pupil, but we shall ignore them here.)

34.95 • Three thin lenses, each with a focal length of 40.0 cm, are aligned on a common axis; adjacent lenses are separated by 52.0 cm. Find the position of the image of a small object on the axis, 80.0 cm to the left of the first lens.

34.96 •• A camera with a 90-mm-focal-length lens is focused on an object 1.30 m from the lens. To refocus on an object 6.50 m from the lens, by how much must the distance between the lens and the sensor be changed? To refocus on the more distant object, is the lens moved toward or away from the sensor?

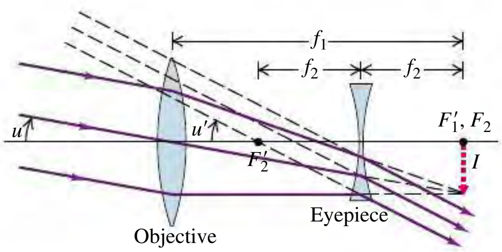
34.97 •• **BIO** In one form of cataract surgery the person's natural lens, which has become cloudy, is replaced by an artificial lens. The refracting properties of the replacement lens can be chosen so that the person's eye focuses on distant objects. But there is no accommodation, and glasses or contact lenses are needed for close vision. What is the power, in diopters, of the corrective contact lenses that will enable a person who has had such surgery to focus on the page of a book at a distance of 24 cm?

34.98 •• **BIO** **A Nearsighted Eye.** A certain very nearsighted person cannot focus on anything farther than 36.0 cm from the eye. Consider the simplified model of the eye described in Exercise 34.50. If the radius of curvature of the cornea is 0.75 cm when the eye is focusing on an object 36.0 cm from the cornea vertex and the indexes of refraction are as described in Exercise 34.50, what is the distance from the cornea vertex to the retina? What does this tell you about the shape of the nearsighted eye?

34.99 •• **BIO** A person with a near point of 85 cm, but excellent distant vision, normally wears corrective glasses. But he loses them while traveling. Fortunately, he has his old pair as a spare. (a) If the lenses of the old pair have a power of $+2.25$ diopters, what is his near point (measured from his eye) when he is wearing the old glasses if they rest 2.0 cm in front of his eye? (b) What would his near point be if his old glasses were contact lenses instead?

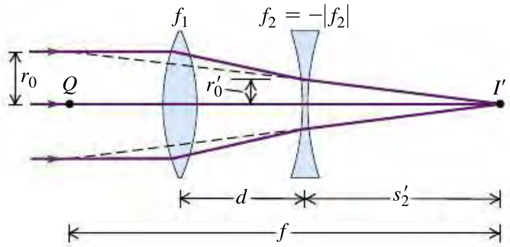
34.100 •• **The Galilean Telescope.** **Figure P34.100** is a diagram of a *Galilean telescope*, or *opera glass*, with both the object and its final image at infinity. The image I serves as a virtual object for the eyepiece. The final image is virtual and erect. (a) Prove that the angular magnification is $M = -f_1/f_2$. (b) A Galilean telescope is to be constructed with the same objective lens as in Exercise 34.61. What focal length should the eyepiece have if this telescope is to have the same magnitude of angular magnification as the one in Exercise 34.61? (c) Compare the lengths of the telescopes.

Figure **P34.100**



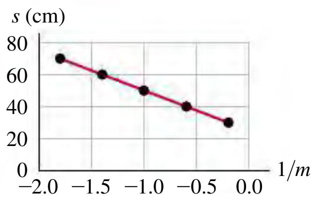
34.101 ••• **Focal Length of a Zoom Lens.** **Figure P34.101** shows a simple version of a zoom lens. The converging lens has focal length f_1 and the diverging lens has focal length $f_2 = -|f_2|$. The two lenses are separated by a variable distance d that is always less than f_1 Also, the magnitude of the focal length of the diverging lens satisfies the inequality $|f_2| > (f_1 - d)$. To determine the effective focal length of the combination lens, consider a bundle of parallel rays of radius r_0 entering the converging lens. (a) Show that the radius of the ray bundle decreases to $r_0' = r_0(f_1 - d)/f_1$ at the point that it enters the diverging lens. (b) Show that the final image I' is formed a distance $s_2' = |f_2|(f_1 - d)/(|f_2| - f_1 + d)$ to the right of the diverging lens. (c) If the rays that emerge from the diverging lens and reach the final image point are extended backward to the left of the diverging lens, they will eventually expand to the original radius r_0 at some point Q. The distance from the final image I' to the point Q is the *effective focal length* f of the lens combination; if the combination were replaced by a single lens of focal length f placed at Q, parallel rays would still be brought to a focus at I'. Show that the effective focal length is given by $f = f_1|f_2|/(|f_2| - f_1 + d)$. (d) If $f_1 = 12.0$ cm, $f_2 = -18.0$ cm, and the separation d is adjustable between 0 and 4.0 cm, find the maximum and minimum focal lengths of the combination. What value of d gives $f = 30.0$ cm?

Figure **P34.101**



34.102 •• **DATA** In setting up an experiment for a high school biology lab, you use a concave spherical mirror to produce real images of a 4.00-mm-tall firefly. The firefly is to the right of the mirror, on the mirror's optic axis, and serves as a real object for the mirror. You want to determine how far the object must be from the mirror's vertex (that is, object distance s) to produce an image of a specified height. First you place a square of white cardboard to the right of the object and find what its distance from the vertex needs to be so that the image is sharply focused on it. Next you measure the height of the sharply focused images for five values of s. For each s value, you then calculate the lateral magnification m. You find that if you graph your data with s on the vertical axis and $1/m$ on the horizontal axis, then your measured points fall close to a straight line. (a) Explain why the data plotted this way should fall close to a straight line. (b) Use the graph in **Fig. P34.102** to calculate the focal length of the mirror. (c) How far from the mirror's vertex should you place the object in order for the image to be real, 8.00 mm tall, and inverted? (d) According to Fig. P34.102, starting from the position that you calculated in part (c), should you move the object closer to the mirror or farther from it to increase the height of the inverted, real image? What distance should you move the object in order to increase the image height from 8.00 mm to 12.00 mm? (e) Explain why $1/m$ approaches zero as s approaches 25 cm. Can you produce a sharp image on the cardboard when $s = 25$ cm? (f) Explain why you can't see sharp images on the cardboard when $s < 25$ cm (and m is positive).

Figure **P34.102**



34.103 •• DATA It is your first day at work as a summer intern at an optics company. Your supervisor hands you a diverging lens and asks you to measure its focal length. You know that with a *converging* lens, you can measure the focal length by placing an object a distance s to the left of the lens, far enough from the lens for the image to be real, and viewing the image on a screen that is to the right of the lens. By adjusting the position of the screen until the image is in sharp focus, you can determine the image distance s' and then use Eq. (34.16) to calculate the focal length f of the lens. But this procedure won't work with a diverging lens—by itself, a diverging lens produces only virtual images, which can't be projected onto a screen. Therefore, to determine the focal length of a diverging lens, you do the following: First you take a *converging* lens and measure that, for an object 20.0 cm to the left of the lens, the image is 29.7 cm to the right of the lens. You then place a *diverging* lens 20.0 cm to the right of the converging lens and measure the final image to be 42.8 cm to the right of the converging lens. Suspecting some inaccuracy in measurement, you repeat the lens-combination measurement with the same object distance for the converging lens but with the diverging lens 25.0 cm to the right of the converging lens. You measure the final image to be 31.6 cm to the right of the converging lens. (a) Use both lens-combination measurements to calculate the focal length of the diverging lens. Take as your best experimental value for the focal length the average of the two values. (b) Which position of the diverging lens, 20.0 cm to the right or 25.0 cm to the right of the converging lens, gives the tallest image?

34.104 •• DATA The science museum where you work is constructing a new display. You are given a glass rod that is surrounded by air and was ground on its left-hand end to form a hemispherical surface there. You must determine the radius of curvature of that surface and the index of refraction of the glass. Remembering the optics portion of your physics course, you place a small object to the left of the rod, on the rod's optic axis, at a distance s from the vertex of the hemispherical surface. You measure the distance s' of the image from the vertex of the surface, with the image being to the right of the vertex. Your measurements are as follows:

s (cm)	22.5	25.0	30.0	35.0	40.0	45.0
s' (cm)	271.6	148.3	89.4	71.1	60.8	53.2

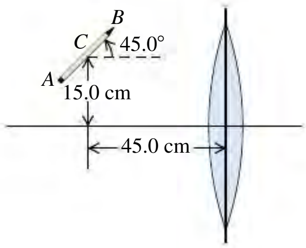
Recalling that the object–image relationships for thin lenses and spherical mirrors include reciprocals of distances, you plot your data as $1/s'$ versus $1/s$. (a) Explain why your data points plotted this way lie close to a straight line. (b) Use the slope and y-intercept of the best-fit straight line to your data to calculate the index of refraction of the glass and the radius of curvature of the hemispherical surface of the rod. (c) Where is the image if the object distance is 15.0 cm?

CHALLENGE PROBLEMS

34.105 ••• CALC (a) For a lens with focal length f, find the smallest distance possible between the object and its real image. (b) Graph the distance between the object and the real image as a function of the distance of the object from the lens. Does your graph agree with the result you found in part (a)?

34.106 ••• An Object at an Angle. A 16.0-cm-long pencil is placed at a 45.0° angle, with its center 15.0 cm above the optic axis and 45.0 cm from a lens with a 20.0-cm focal length as shown in **Fig. P34.106**. (Note that the figure is not drawn to scale.) Assume that the diameter of the lens is large enough for the paraxial approximation to be valid. (a) Where is the image of the pencil? (Give the location of the images of the points A, B, and C on the object, which are located at the eraser, point, and center of the pencil, respectively.) (b) What is the length of the image (that is, the distance between the images of points A and B)? (c) Show the orientation of the image in a sketch.

Figure **P34.106**



34.107 ••• BIO People with normal vision cannot focus their eyes underwater if they aren't wearing a face mask or goggles and there is water in contact with their eyes (see Discussion Question Q34.23). (a) Why not? (b) With the simplified model of the eye described in Exercise 34.50, what corrective lens (specified by focal length as measured in air) would be needed to enable a person underwater to focus an infinitely distant object? (Be careful—the focal length of a lens underwater is *not* the same as in air! See Problem 34.92. Assume that the corrective lens has a refractive index of 1.62 and that the lens is used in eyeglasses, not goggles, so there is water on both sides of the lens. Assume that the eyeglasses are 2.00 cm in front of the eye.)

PASSAGE PROBLEMS

BIO AMPHIBIAN VISION. The eyes of amphibians such as frogs have a much flatter cornea but a more strongly curved (almost spherical) lens than do the eyes of air-dwelling mammals. In mammalian eyes, the shape (and therefore the focal length) of the lens changes to enable the eye to focus at different distances. In amphibian eyes, the shape of the lens doesn't change. Amphibians focus on objects at different distances by using specialized muscles to move the lens closer to or farther from the retina, like the focusing mechanism of a camera. In air, most frogs are nearsighted; correcting the distance vision of a typical frog in air would require contact lenses with a power of about –6.0 D.

34.108 A frog can see an insect clearly at a distance of 10 cm. At that point the effective distance from the lens to the retina is 8 mm. If the insect moves 5 cm farther from the frog, by how much and in which direction does the lens of the frog's eye have to move to keep the insect in focus? (a) 0.02 cm, toward the retina; (b) 0.02 cm, away from the retina; (c) 0.06 cm, toward the retina; (d) 0.06 cm, away from the retina.

34.109 What is the farthest distance at which a typical "nearsighted" frog can see clearly in air? (a) 12 m; (b) 6.0 m; (c) 80 cm; (d) 17 cm.

34.110 Given that frogs are nearsighted in air, which statement is most likely to be true about their vision in water? (a) They are even more nearsighted; because water has a higher index of refraction than air, a frog's ability to focus light increases in water. (b) They are less nearsighted, because the cornea is less effective at refracting light in water than in air. (c) Their vision is no different, because only structures that are internal to the eye can affect the eye's ability to focus. (d) The images projected on the retina are no longer inverted, because the eye in water functions as a diverging lens rather than a converging lens.

34.111 To determine whether a frog can judge distance by means of the amount its lens must move to focus on an object, researchers covered one eye with an opaque material. An insect was placed in front of the frog, and the distance that the frog snapped its tongue out to catch the insect was measured with high-speed video. The experiment was repeated with a contact lens over the eye to determine whether the frog could correctly judge the distance under these conditions. If such an experiment is performed twice, once with a lens of power –9 D and once with a lens of power –15 D, in which case does the frog have to focus at a shorter distance, and why? (a) With the –9-D lens; because the lenses are diverging, the lens with the longer focal length creates an image that is closer to the frog. (b) With the –15-D lens; because the lenses are diverging, the lens with the shorter focal length creates an image that is closer to the frog. (c) With the –9-D lens; because the lenses are converging, the lens with the longer focal length creates a larger real image. (d) With the –15-D lens; because the lenses are converging, the lens with the shorter focal length creates a larger real image.

Answers

Chapter Opening Question ?

(ii) A magnifying lens (simple magnifier) produces a virtual image with a large angular size that is infinitely far away, so you can see it in sharp focus with your eyes relaxed. (A surgeon would not appreciate having to strain her eyes while working.) The object should be at the focal point of the lens, so the object and lens are separated by one focal length. The distance from magnifier to eye is not crucial.

Test Your Understanding Questions

34.1 **(iv)** When you are a distance s from the mirror, your image is a distance s on the other side of the mirror and the distance from you to your image is $2s$. As you move toward the mirror, the distance $2s$ changes at twice the rate of the distance s, so your image moves toward you at speed $2v$.

34.2 **(a) concave, (b) (ii)** A convex mirror always produces an erect image, but that image is smaller than the object (see Fig. 34.16b). Hence a concave mirror must be used. The image will be erect and enlarged only if the distance from the object (your face) to the mirror is less than the focal length of the mirror, as in Fig. 34.20d.

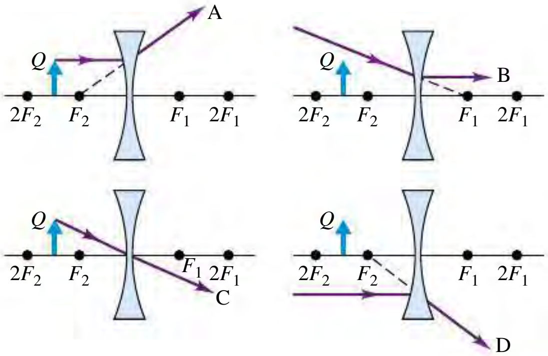
34.3 **no** The sun is very far away, so the object distance is essentially infinite: $s = \infty$ and $1/s = 0$. Material a is air ($n_a = 1.00$) and material b is water ($n_b = 1.33$), so the image position s' is

$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R} \quad \text{or} \quad 0 + \frac{1.33}{s'} = \frac{1.33 - 1.00}{R}$$

$$s' = \frac{1.33}{0.33}R = 4.0R$$

The image would be formed 4.0 drop radii from the front surface of the drop. But since each drop is only a part of a complete sphere, the distance from the front to the back of the drop is less than $2R$. Thus the rays of sunlight never reach the image point, and the drops do not form an image of the sun on the leaf. The rays are nonetheless concentrated and can cause damage to the leaf.

34.4 **A and C** When rays A and D are extended backward, they pass through focal point F_2; thus, before they passed through the lens, they were parallel to the optic axis. The figures show that ray A emanated from point Q, but ray D did not. Ray B is parallel to the optic axis, so before it passed through the lens, it was directed toward focal point F_1. Hence it cannot have come from point Q. Ray C passes through the center of the lens and hence is not deflected by its passage; tracing the ray backward shows that it emanates from point Q.



34.5 **(iii)** The smaller image area of the electronic sensor means that the angle of view is decreased for a given focal length. Individual objects make images of the same size in either case; when a smaller light-sensitive area is used, fewer images fit into the area and the field of view is narrower.

34.6 **(iii)** This lens is designed to correct for a type of astigmatism. Along the vertical axis, the lens is configured as a converging lens; along the horizontal axis, the lens is configured as a diverging lens. Hence the eye is hyperopic (see Fig. 34.46) for objects that are oriented vertically but myopic for objects that are oriented horizontally (see Fig. 34.47). Without correction, the eye focuses vertical objects behind the retina but horizontal objects in front of the retina.

34.7 **(ii)** The object must be held at the focal point, which is twice as far away if the focal length f is twice as great. Equation (34.24) shows that the angular magnification M is inversely proportional to f, so doubling the focal length makes $M \frac{1}{2}$ as great. To improve the magnification, you should use a magnifier with a *shorter* focal length.

34.8 **(i)** The objective lens of a microscope makes enlarged images of small objects, so the absolute value of its lateral magnification m is greater than 1. By contrast, the objective lens of a refracting telescope makes *reduced* images. For example, the moon is thousands of kilometers in diameter, but its image may fit on an electronic sensor a few centimeters across. Thus $|m|$ is much less than 1 for a refracting telescope. (In both cases m is negative because the objective makes an inverted image, which is why the question asks about the absolute value of m.)

Bridging Problem

(a) 29.9 cm to the left of the goblet
(b) 3.73 cm to the right of the goblet

? When white light shines downward on a thin, horizontal layer of oil, light waves reflected from the upper and lower surfaces of the film of oil interfere, producing vibrant colors. The color that you see reflected from a certain spot on the film depends on (i) the film thickness at that spot; (ii) the index of refraction of the oil; (iii) the index of refraction of the material below the oil; (iv) both (i) and (ii); (v) all of (i), (ii), and (iii).

35 INTERFERENCE

LEARNING GOALS

Looking forward at …

35.1 What happens when two waves combine, or interfere, in space.

35.2 How to understand the interference pattern formed by the interference of two coherent light waves.

35.3 How to calculate the intensity at various points in an interference pattern.

35.4 How interference occurs when light reflects from the two surfaces of a thin film.

35.5 How interference makes it possible to measure extremely small distances.

Looking back at …

14.2, 31.1 Phasors.

15.3, 15.6, 15.7 Wave number, wave superposition, standing waves on a string.

16.4 Standing sound waves.

32.1, 32.4, 32.5 Electromagnetic spectrum, wave intensity, standing electromagnetic waves.

An ugly black oil spot on the pavement can become a thing of beauty after a rain, when the oil reflects a rainbow of colors. Multicolored reflections can also be seen from the surfaces of soap bubbles and DVDs. How is it possible for colorless objects to produce these remarkable colors?

In our discussion of lenses, mirrors, and optical instruments we used the model of *geometric optics,* in which we represent light as *rays,* straight lines that are bent at a reflecting or refracting surface. But light is fundamentally a *wave,* and in some situations we have to consider its wave properties explicitly. If two or more light waves of the same frequency overlap at a point, the total effect depends on the *phases* of the waves as well as their amplitudes. The resulting patterns of light are a result of the *wave* nature of light and cannot be understood on the basis of rays. Optical effects that depend on the wave nature of light are grouped under the heading **physical optics.**

In this chapter we'll look at *interference* phenomena that occur when two waves combine. Effects that occur when *many* sources of waves are present are called *diffraction* phenomena; we'll study these in Chapter 36. In that chapter we'll see that diffraction effects occur whenever a wave passes through an aperture or around an obstacle. They are important in practical applications of physical optics such as diffraction gratings, x-ray diffraction, and holography.

While our primary concern is with light, interference and diffraction can occur with waves of *any* kind. As we go along, we'll point out applications to other types of waves such as sound and water waves.

35.1 INTERFERENCE AND COHERENT SOURCES

As we discussed in Chapter 15, the term **interference** refers to any situation in which two or more waves overlap in space. When this occurs, the total wave at any point at any instant of time is governed by the **principle of superposition,** which we introduced in Section 15.6 in the context of waves on a string. This principle also applies to electromagnetic waves and is the most important principle in all of physical optics. The principle of superposition states:

When two or more waves overlap, the resultant displacement at any point and at any instant is found by adding the instantaneous displacements that would be produced at the point by the individual waves if each were present alone.

(In some special situations, such as electromagnetic waves propagating in a crystal, this principle may not apply. A discussion of these is beyond our scope.)

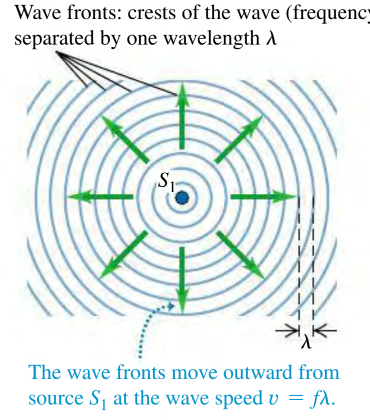
We use the term "displacement" in a general sense. With waves on the surface of a liquid, we mean the actual displacement of the surface above or below its normal level. With sound waves, the term refers to the excess or deficiency of pressure. For electromagnetic waves, we usually mean a specific component of electric or magnetic field.

Interference in Two or Three Dimensions

We have already discussed one important case of interference, in which two identical waves propagating in opposite directions combine to produce a *standing wave*. We saw this in Section 15.7 for transverse waves on a string and in Section 16.4 for longitudinal waves in a fluid filling a pipe; we described the same phenomenon for electromagnetic waves in Section 32.5. In all of these cases the waves propagated along only a single axis: along a string, along the length of a fluid-filled pipe, or along the propagation direction of an electromagnetic plane wave. But light waves can (and do) travel in *two* or *three* dimensions, as can any kind of wave that propagates in a two- or three-dimensional medium. In this section we'll see what happens when we combine waves that spread out in two or three dimensions from a pair of identical wave sources.

Interference effects are most easily seen when we combine *sinusoidal* waves with a single frequency f and wavelength λ. **Figure 35.1** shows a "snapshot" of a *single* source S_1 of sinusoidal waves and some of the wave fronts produced by this source. The figure shows only the wave fronts corresponding to wave *crests,* so the spacing between successive wave fronts is one wavelength. The material surrounding S_1 is uniform, so the wave speed is the same in all directions, and there is no refraction (and hence no bending of the wave fronts). If the waves are two-dimensional, like waves on the surface of a liquid, the circles in Fig. 35.1 represent circular wave fronts; if the waves propagate in three dimensions, the circles represent spherical wave fronts spreading away from S_1.

35.1 A "snapshot" of sinusoidal waves of frequency f and wavelength λ spreading out from source S_1 in all directions.



In optics, sinusoidal waves are characteristic of **monochromatic light** (light of a single color). While it's fairly easy to make water waves or sound waves of a single frequency, common sources of light *do not* emit monochromatic (single-frequency) light. For example, incandescent light bulbs and flames emit a continuous distribution of wavelengths. By far the most nearly monochromatic light source is the *laser.* An example is the helium–neon laser, which emits red light at 632.8 nm with a wavelength range of the order of ± 0.000001 nm, or about one part in 10^9. In this chapter and the next, we will assume that we are working with monochromatic waves (unless we explicitly state otherwise).

Constructive and Destructive Interference

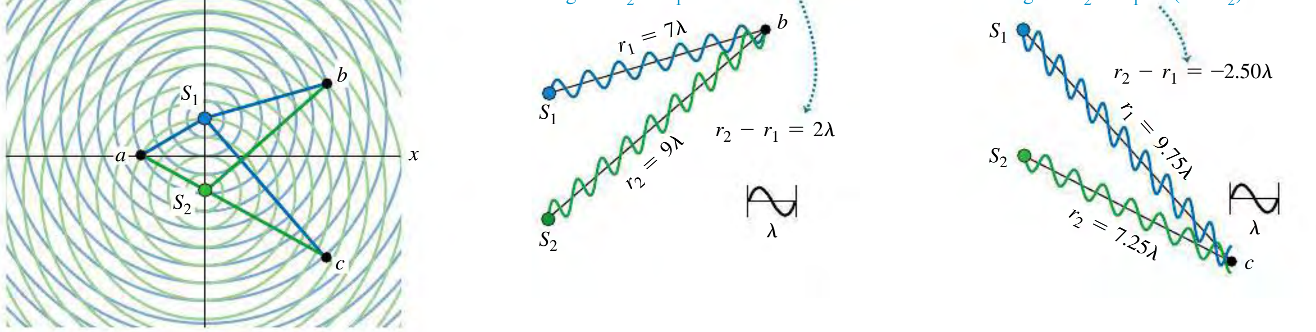
Figure 35.2a (next page) shows two identical sources of monochromatic waves, S_1 and S_2. The two sources produce waves of the same amplitude and the same wavelength λ. In addition, the two sources are permanently *in phase;* they vibrate in unison. They might be two loudspeakers driven by the same amplifier, two radio antennas powered by the same transmitter, or two small slits in an opaque screen, illuminated by the same monochromatic light source. We will see that if

35.2 (a) A "snapshot" of sinusoidal waves spreading out from two coherent sources S_1 and S_2. Constructive interference occurs at point a (equidistant from the two sources) and (b) at point b. (c) Destructive interference occurs at point c.

(a) Two coherent wave sources separated by a distance 4λ

(b) Conditions for constructive interference: Waves interfere constructively if their path lengths differ by an integral number of wavelengths: $r_2 - r_1 = m\lambda$.

(c) Conditions for destructive interference: Waves interfere destructively if their path lengths differ by a half-integral number of wavelengths: $r_2 - r_1 = (m + \frac{1}{2})\lambda$.



there were not a constant phase relationship between the two sources, the phenomena we are about to discuss would not occur. Two monochromatic sources of the same frequency and with a constant phase relationship (not necessarily in phase) are said to be **coherent.** We also use the term *coherent waves* (or, for light waves, *coherent light*) to refer to the waves emitted by two such sources.

If the waves emitted by the two coherent sources are *transverse,* like electromagnetic waves, then we will also assume that the wave disturbances produced by both sources have the same *polarization* (that is, they lie along the same line). For example, the sources S_1 and S_2 in Fig. 35.2a could be two radio antennas in the form of long rods oriented parallel to the z-axis (perpendicular to the plane of the figure); at any point in the xy-plane the waves produced by both antennas have $\vec{E}$ fields with only a z-component. Then we need only a single scalar function to describe each wave; this makes the analysis much easier.

We position the two sources of equal amplitude, equal wavelength, and (if the waves are transverse) the same polarization along the y-axis in Fig. 35.2a, equidistant from the origin. Consider a point a on the x-axis. From symmetry the two distances from S_1 to a and from S_2 to a are *equal*; waves from the two sources thus require equal times to travel to a. Hence waves that leave S_1 and S_2 in phase arrive at a in phase, and the total amplitude at a is *twice* the amplitude of each individual wave. This is true for *any* point on the x-axis.

Similarly, the distance from S_2 to point b is exactly two wavelengths *greater* than the distance from S_1 to b. A wave crest from S_1 arrives at b exactly two cycles earlier than a crest emitted at the same time from S_2, and again the two waves arrive in phase. As at point a, the total amplitude is the sum of the amplitudes of the waves from S_1 and S_2.

In general, when waves from two or more sources arrive at a point *in phase,* they reinforce each other: The amplitude of the resultant wave is the *sum* of the amplitudes of the individual waves. This is called **constructive interference** (Fig. 35.2b). Let the distance from S_1 to any point P be r_1, and let the distance from S_2 to P be r_2. For constructive interference to occur at P, the path difference $r_2 - r_1$ for the two sources must be an integral multiple of the wavelength λ:

$$r_2 - r_1 = m\lambda \qquad (m = 0, \pm 1, \pm 2, \pm 3, \ldots) \qquad \begin{array}{l}\text{(constructive}\\ \text{interference,}\\ \text{sources in phase)}\end{array} \tag{35.1}$$

In Fig. 35.2a, points a and b satisfy Eq. (35.1) with $m = 0$ and $m = +2$, respectively.

Something different occurs at point c in Fig. 35.2a. At this point, the path difference $r_2 - r_1 = -2.50\lambda$, which is a *half-integral* number of wavelengths. Waves from the two sources arrive at point c exactly a half-cycle out of phase. A crest of one wave arrives at the same time as a crest in the opposite direction (a "trough") of the other wave (Fig. 35.2c). The resultant amplitude is the *difference* between the two individual amplitudes. If the individual amplitudes are equal, then the total amplitude is *zero*! This cancellation or partial cancellation of the individual waves is called **destructive interference.** The condition for destructive interference in the situation shown in Fig. 35.2a is

$$r_2 - r_1 = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, \pm 1, \pm 2, \pm 3, \ldots) \quad \begin{array}{l}\text{(destructive}\\ \text{interference,}\\ \text{sources in phase)}\end{array} \tag{35.2}$$

The path difference at point c in Fig. 35.2a satisfies Eq. (35.2) with $m = -3$.

Figure 35.3 shows the same situation as in Fig. 35.2a, but with red curves that show all positions where *constructive* interference occurs. On each curve, the path difference $r_2 - r_1$ is equal to an integer m times the wavelength, as in Eq. (35.1). These curves are called **antinodal curves.** They are directly analogous to *antinodes* in the standing-wave patterns described in Chapters 15 and 16 and Section 32.5. In a standing wave formed by interference between waves propagating in opposite directions, the antinodes are points at which the amplitude is maximum; likewise, the wave amplitude in the situation of Fig. 35.3 is maximum along the antinodal curves. Not shown in Fig. 35.3 are the **nodal curves,** which are the curves that show where *destructive* interference occurs in accordance with Eq. (35.2); these are analogous to the *nodes* in a standing-wave pattern. A nodal curve lies between each two adjacent antinodal curves in Fig. 35.3; one such curve, corresponding to $r_2 - r_1 = -2.50\lambda$, passes through point c.

In some cases, such as two loudspeakers or two radio-transmitter antennas, the interference pattern is three-dimensional. Think of rotating the color curves of Fig. 35.3 around the y-axis; then maximum constructive interference occurs at all points on the resulting surfaces of revolution.

CAUTION Interference patterns are not standing waves In the standing waves described in Sections 15.7, 16.4, and 32.5, the interference is between two waves propagating in opposite directions; there is *no* net energy flow in either direction (the energy in the wave is left "standing"). In the situations shown in Figs. 35.2a and 35.3, there is likewise a stationary pattern of antinodal and nodal curves, but there is a net flow of energy *outward* from the two sources. All that interference does is to "channel" the energy flow so that it is greatest along the antinodal curves and least along the nodal curves.

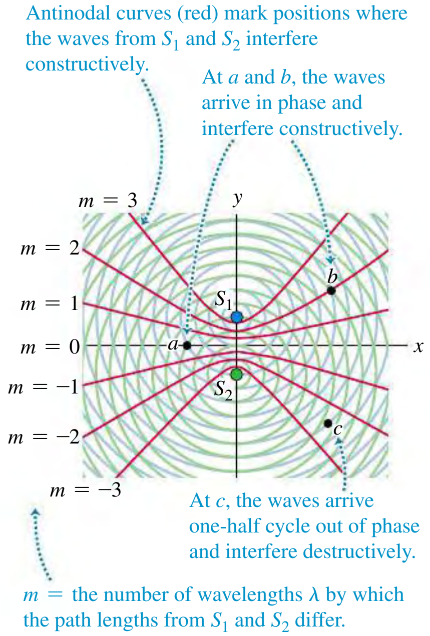
For Eqs. (35.1) and (35.2) to hold, the two sources must have the same wavelength and must *always* be in phase. These conditions are rather easy to satisfy for sound waves. But with *light* waves there is no practical way to achieve a constant phase relationship (coherence) with two independent sources. This is because of the way light is emitted. In ordinary light sources, atoms gain excess energy by thermal agitation or by impact with accelerated electrons. Such an "excited" atom begins to radiate energy and continues until it has lost all the energy it can, typically in a time of the order of 10^{-8} s. The many atoms in a source ordinarily radiate in an unsynchronized and random phase relationship, and the light that is emitted from *two* such sources has no definite phase relationship.

However, the light from a single source can be split so that parts of it emerge from two or more regions of space, forming two or more *secondary sources*. Then any random phase change in the source affects these secondary sources equally and does not change their *relative* phase.

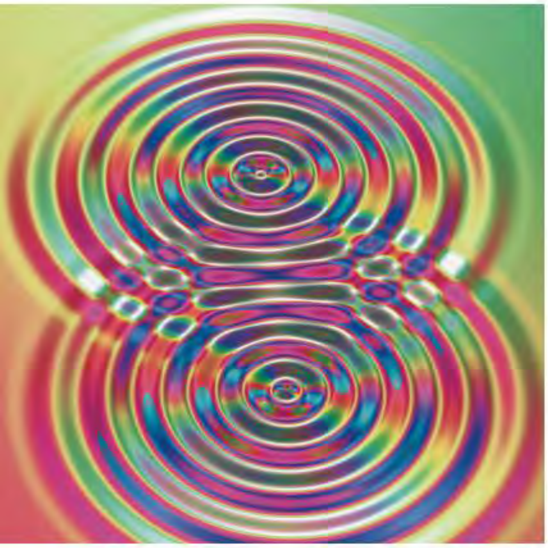
The distinguishing feature of light from a *laser* is that the emission of light from many atoms is synchronized in frequency and phase. As a result, the random phase changes mentioned above occur much less frequently. Definite phase relationships are preserved over correspondingly much greater lengths in the beam, and laser light is much more coherent than ordinary light.

TEST YOUR UNDERSTANDING OF SECTION 35.1 Consider a point in Fig. 35.3 on the positive y-axis above S_1. Does this point lie on (i) an antinodal curve; (ii) a nodal curve; or (iii) neither? (*Hint:* The distance between S_1 and S_2 is 4λ.)

35.3 The same as Fig. 35.2a, but with red antinodal curves (curves of maximum amplitude) superimposed. All points on each curve satisfy Eq. (35.1) with the value of m shown. The nodal curves (not shown) lie between each adjacent pair of antinodal curves.



BIO Application Phase Difference, Path Difference, and Localization in Human Hearing Your auditory system uses the phase differences between sounds received by your left and your right ears for *localization*—determining the direction from which the sounds are coming. For sound waves with frequencies lower than about 800 Hz (which are important in speech and music), the distance between your ears is less than a half-wavelength and the phase difference between the waves detected by each ear is less than a half cycle. Remarkably, your brain can detect this phase difference, determine the corresponding path difference, and use this information to localize the direction of the sound source.



35.4 The concepts of constructive interference and destructive interference apply to these water waves as well as to light waves and sound waves.

35.2 TWO-SOURCE INTERFERENCE OF LIGHT

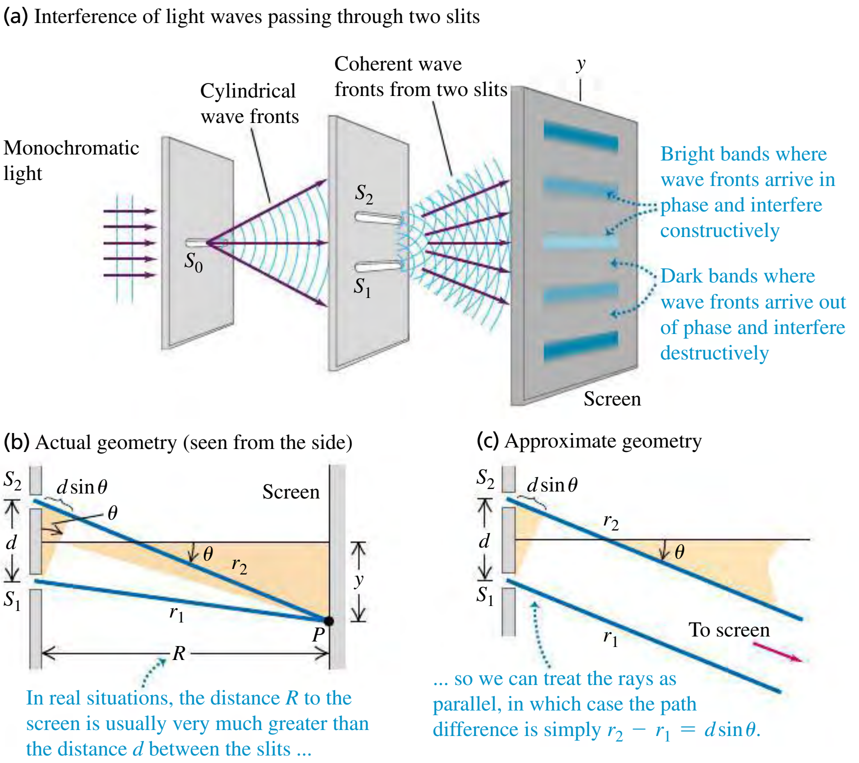
The interference pattern produced by two coherent sources of *water* waves of the same wavelength can be readily seen in a ripple tank with a shallow layer of water (**Fig. 35.4**). This pattern is not directly visible when the interference is between *light* waves, since light traveling in a uniform medium cannot be seen. (Sunlight in a room is made visible by scattering from airborne dust.)

Figure 35.5a shows one of the earliest quantitative experiments to reveal the interference of light from two sources, first performed in 1800 by the English scientist Thomas Young. Let's examine this important experiment in detail. A light source (not shown) emits monochromatic light; however, this light is not suitable for use in an interference experiment because emissions from different parts of an ordinary source are not synchronized. To remedy this, the light is directed at a screen with a narrow slit S_0, 1 μm or so wide. The light emerging from the slit originated from only a small region of the light source; thus slit S_0 behaves more nearly like the idealized source shown in Fig. 35.1. (In modern versions of the experiment, a laser is used as a source of coherent light, and the slit S_0 isn't needed.) The light from slit S_0 falls on a screen with two other narrow slits S_1 and S_2, each 1 μm or so wide and a few tens or hundreds of micrometers apart. Cylindrical wave fronts spread out from slit S_0 and reach slits S_1 and S_2 *in phase* because they travel equal distances from S_0. The waves *emerging* from slits S_1 and S_2 are therefore also always in phase, so S_1 and S_2 are *coherent* sources. The interference of waves from S_1 and S_2 produces a pattern in space like that to the right of the sources in Figs. 35.2a and 35.3.

To visualize the interference pattern, a screen is placed so that the light from S_1 and S_2 falls on it (Fig. 35.5b). The screen will be most brightly illuminated at points P, where the light waves from the slits interfere constructively, and will be darkest at points where the interference is destructive.

To simplify the analysis of Young's experiment, we assume that the distance R from the slits to the screen is so large in comparison to the distance d between the slits that the lines from S_1 and S_2 to P are very nearly parallel, as in Fig. 35.5c.

35.5 (a) Young's experiment to show interference of light passing through two slits. A pattern of bright and dark areas appears on the screen (see Fig. 35.6). (b) Geometrical analysis of Young's experiment. For the case shown, $r_2 > r_1$ and both y and θ are positive. If point P is on the other side of the screen's center, $r_2 < r_1$ and both y and θ are negative. (c) Approximate geometry when the distance R to the screen is much greater than the distance d between the slits.



This is usually the case for experiments with light; the slit separation is typically a few millimeters, while the screen may be a meter or more away. The difference in path length is then given by

$$r_2 - r_1 = d\sin\theta \tag{35.3}$$

where θ is the angle between a line from slits to screen (shown in blue in Fig. 35.5c) and the normal to the plane of the slits (shown as a thin black line).

Constructive and Destructive Two-Slit Interference

We found in Section 35.1 that constructive interference (reinforcement) occurs at points where the path difference is an integral number of wavelengths, $m\lambda$, where $m = 0, \pm1, \pm2, \pm3, \ldots$. So the bright regions on the screen in Fig. 35.5a occur at angles θ for which

Constructive interference, two slits:

Distance between slits, Wavelength

$$d\sin\theta = m\lambda \quad (m = 0, \pm1, \pm2, \ldots) \tag{35.4}$$

Angle of line from slits to mth bright region on screen

Similarly, destructive interference (cancellation) occurs, forming dark regions on the screen, at points for which the path difference is a half-integral number of wavelengths, $\left(m + \frac{1}{2}\right)\lambda$:

Destructive interference, two slits:

Distance between slits, Wavelength

$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, \pm1, \pm2, \ldots) \tag{35.5}$$

Angle of line from slits to mth dark region on screen

Thus the pattern on the screen of Figs. 35.5a and 35.5b is a succession of bright and dark bands, or **interference fringes,** parallel to the slits S_1 and S_2. A photograph of such a pattern is shown in **Fig. 35.6**.

We can derive an expression for the positions of the centers of the bright bands on the screen. In Fig. 35.5b, y is measured from the center of the pattern, corresponding to the distance from the center of Fig. 35.6. Let y_m be the distance from the center of the pattern $(\theta = 0)$ to the center of the mth bright band. Let θ_m be the corresponding value of θ; then

$$y_m = R\tan\theta_m$$

In such experiments, the distances y_m are often much smaller than the distance R from the slits to the screen. Hence θ_m is very small, $\tan\theta_m \approx \sin\theta_m$, and

$$y_m = R\sin\theta_m$$

Combining this with Eq. (35.4), we find that *for small angles only,*

Constructive interference, Young's experiment (small angles only):

Position of mth bright band, Wavelength

$$y_m = R\frac{m\lambda}{d} \quad (m = 0, \pm1, \pm2, \ldots) \tag{35.6}$$

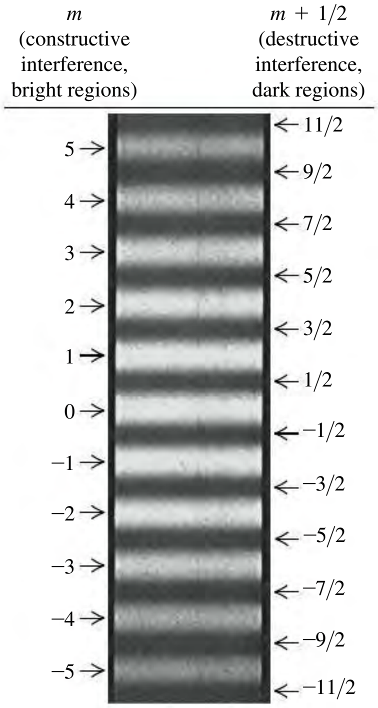
Distance from slits to screen, Distance between slits

We can measure R and d, as well as the positions y_m of the bright fringes, so this experiment provides a direct measurement of the wavelength λ. Young's experiment was in fact the first direct measurement of wavelengths of light.

The distance between adjacent bright bands in the pattern is *inversely* proportional to the distance d between the slits. The closer together the slits are, the more the pattern spreads out. When the slits are far apart, the bands in the pattern are closer together.

MP **PhET:** Wave Interference

35.6 Photograph of interference fringes produced on a screen in Young's double-slit experiment. The center of the pattern is a bright band corresponding to $m = 0$ in Eq. (35.4); this point on the screen is equidistant from the two slits.



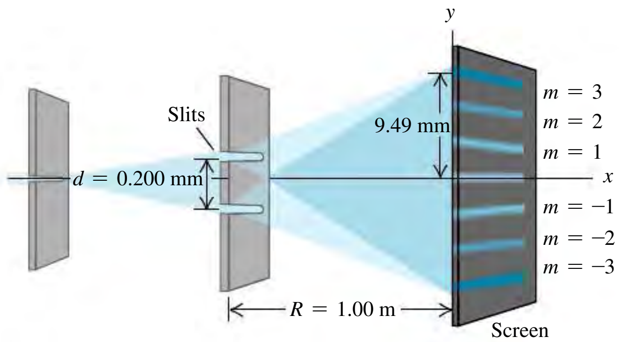
CAUTION **Equation (35.6) is for small angles only** While Eqs. (35.4) and (35.5) are valid for any angle, Eq. (35.6) is valid for *small* angles only. It can be used *only* if the distance R from slits to screen is much greater than the slit separation d and if R is much greater than the distance y_m from the center of the interference pattern to the mth bright band.

While we have described the experiment that Young performed with visible light, the results given in Eqs. (35.4) and (35.5) are valid for *any* type of wave, provided that the resultant wave from two coherent sources is detected at a point that is far away in comparison to the separation d.

EXAMPLE 35.1 TWO-SLIT INTERFERENCE

Figure 35.7 shows a two-slit interference experiment in which the slits are 0.200 mm apart and the screen is 1.00 m from the slits. The $m = 3$ bright fringe in the figure is 9.49 mm from the central fringe. Find the wavelength of the light.

35.7 Using a two-slit interference experiment to measure the wavelength of light.



SOLUTION

IDENTIFY and SET UP: Our target variable in this two-slit interference problem is the wavelength λ. We are given the slit separation $d = 0.200$ mm, the distance from slits to screen $R = 1.00$ m, and the distance $y_3 = 9.49$ mm on the screen from the center of the interference pattern to the $m = 3$ bright fringe. We may use Eq. (35.6) to find λ, since the value of R is so much greater than the values of d or y_3.

EXECUTE: We solve Eq. (35.6) for λ for the case $m = 3$:

$$\lambda = \frac{y_m d}{mR} = \frac{(9.49 \times 10^{-3}\ \text{m})(0.200 \times 10^{-3}\ \text{m})}{(3)(1.00\ \text{m})}$$
$$= 633 \times 10^{-9}\ \text{m} = 633\ \text{nm}$$

EVALUATE: This bright fringe could also correspond to $m = -3$. Can you show that this gives the same result for λ?

EXAMPLE 35.2 BROADCAST PATTERN OF A RADIO STATION

It is often desirable to radiate most of the energy from a radio transmitter in particular directions rather than uniformly in all directions. Pairs or rows of antennas are often used to produce the desired radiation pattern. As an example, consider two identical vertical antennas 400 m apart, operating at 1500 kHz $= 1.5 \times 10^6$ Hz (near the top end of the AM broadcast band) and oscillating in phase. At distances much greater than 400 m, in what directions is the intensity from the two antennas greatest?

SOLUTION

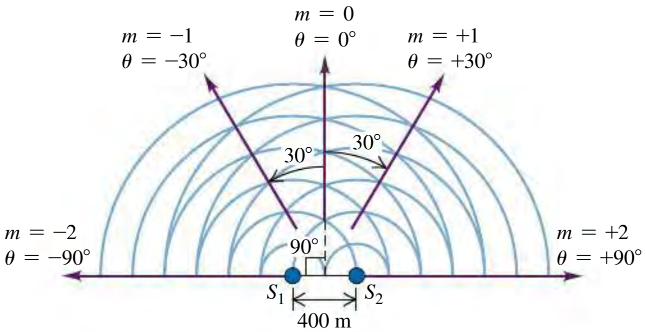
IDENTIFY and SET UP: The antennas, shown in **Fig. 35.8**, correspond to sources S_1 and S_2 in Fig. 35.5. Hence we can apply the ideas of two-slit interference to this problem. Since the resultant wave is detected at distances much greater than $d = 400$ m, we may use Eq. (35.4) to give the directions of the intensity maxima, the values of θ for which the path difference is zero or a whole number of wavelengths.

EXECUTE: The wavelength is $\lambda = c/f = 200$ m. From Eq. (35.4) with $m = 0, \pm 1$, and ± 2, the intensity maxima are given by

$$\sin\theta = \frac{m\lambda}{d} = \frac{m(200\ \text{m})}{400\ \text{m}} = \frac{m}{2} \qquad \theta = 0, \pm 30°, \pm 90°$$

In this example, values of m greater than 2 or less than -2 give values of $\sin\theta$ greater than 1 or less than -1, which is impossible. There is *no* direction for which the path difference is three or more wavelengths, so values of m of ± 3 or beyond have no meaning in this example.

35.8 Two radio antennas broadcasting in phase. The purple arrows indicate the directions of maximum intensity. The waves that are emitted toward the lower half of the figure are not shown.



EVALUATE: We can check our result by calculating the angles for *minimum* intensity, using Eq. (35.5). There should be one intensity minimum between each pair of intensity maxima, just as in Fig. 35.6. From Eq. (35.5), with $m = -2, -1, 0,$ and 1,

$$\sin\theta = \frac{\left(m + \frac{1}{2}\right)\lambda}{d} = \frac{m + \frac{1}{2}}{2} \qquad \theta = \pm 14.5°, \pm 48.6°$$

These angles fall between the angles for intensity maxima, as they should. The angles are not small, so the angles for the minima are *not* exactly halfway between the angles for the maxima.

TEST YOUR UNDERSTANDING OF SECTION 35.2 You shine a tunable laser (whose wavelength can be adjusted by turning a knob) on a pair of closely spaced slits. The light emerging from the two slits produces an interference pattern on a screen like that shown in Fig. 35.6. If you adjust the wavelength so that the laser light changes from red to blue, how will the spacing between bright fringes change? (i) The spacing increases; (ii) the spacing decreases; (iii) the spacing is unchanged; (iv) not enough information is given to decide. ❙

35.3 INTENSITY IN INTERFERENCE PATTERNS

In Section 35.2 we found the positions of maximum and minimum intensity in a two-source interference pattern. Let's now see how to find the intensity at *any* point in the pattern. To do this, we have to combine the two sinusoidally varying fields (from the two sources) at a point P in the radiation pattern, taking proper account of the phase difference of the two waves at point P, which results from the path difference. The intensity is then proportional to the square of the resultant electric-field amplitude, as we learned in Section 32.4.

To calculate the intensity, we will assume, as in Section 35.2, that the waves from the two sources have equal amplitude E and the same polarization. This assumes that the sources are identical and ignores the slight amplitude difference caused by the unequal path lengths (the amplitude decreases with increasing distance from the source). From Eq. (32.29), each source by itself would give an intensity $\frac{1}{2}\epsilon_0 cE^2$ at point P. If the two sources are in phase, then the waves that arrive at P differ in phase by an amount ϕ that is proportional to the difference in their path lengths, $(r_2 - r_1)$. Then we can use the following expressions for the two electric fields superposed at P:

$$E_1(t) = E\cos(\omega t + \phi)$$
$$E_2(t) = E\cos\omega t$$

The superposition of the two fields at P is a sinusoidal function with some amplitude E_P that depends on E and the phase difference ϕ. First we'll work on finding the amplitude E_P if E and ϕ are known. Then we'll find the intensity I of the resultant wave, which is proportional to $E_P{}^2$. Finally, we'll relate the phase difference ϕ to the path difference, which is determined by the geometry of the situation.

Amplitude in Two-Source Interference

To add the two sinusoidal functions with a phase difference, we use the same *phasor* representation that we used for simple harmonic motion (see Section 14.2) and for voltages and currents in ac circuits (see Section 31.1). We suggest that you review these sections now. Each sinusoidal function is represented by a rotating vector (phasor) whose projection on the horizontal axis at any instant represents the instantaneous value of the sinusoidal function.

In **Fig. 35.9**, E_1 is the horizontal component of the phasor representing the wave from source S_1, and E_2 is the horizontal component of the phasor for the wave from S_2. As shown in the diagram, both phasors have the same magnitude E, but E_1 is *ahead* of E_2 in phase by an angle ϕ. Both phasors rotate counterclockwise with constant angular speed ω, and the sum of the projections on the horizontal axis at any time gives the instantaneous value of the total E field at point P. Thus the amplitude E_P of the resultant sinusoidal wave at P is the magnitude of the dark red phasor in the diagram (labeled E_P); this is the *vector sum* of the other two phasors. To find E_P, we use the law of cosines and the trigonometric identity $\cos(\pi - \phi) = -\cos\phi$:

$$E_P^2 = E^2 + E^2 - 2E^2\cos(\pi - \phi)$$
$$= E^2 + E^2 + 2E^2\cos\phi$$

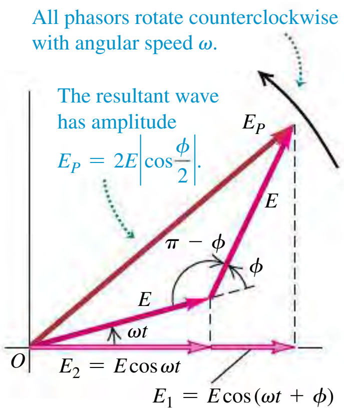
DATA SPEAKS

Two-Source Interference

When students were given a problem involving interference of waves from two sources, more than 34% gave an incorrect response. Common errors:

- Confusion about the kind of sources required to cause interference. For there to be a steady interference pattern from two wave sources, the two sources must be monochromatic, emit waves of the same frequency, and have a fixed phase relationship.
- Confusion about constructive and destructive interference. Constructive interference occurs at points where waves from two sources arrive in phase (the crest of one wave aligns with a crest of the other wave). Destructive interference occurs at points where waves from two sources arrive out of phase (the crest of one wave aligns with a trough of the other wave).

35.9 Phasor diagram for the superposition at a point P of two waves of equal amplitude E with a phase difference ϕ.



Then, using the identity $1 + \cos\phi = 2\cos^2(\phi/2)$, we obtain

$$E_P^2 = 2E^2(1 + \cos\phi) = 4E^2\cos^2\left(\frac{\phi}{2}\right)$$

Electric-field amplitude in two-source interference

$$\underbrace{E_P}_{\text{Electric-field amplitude in two-source interference}} = \underbrace{2E}_{\text{Amplitude of wave from one source}}\left|\cos\frac{\overbrace{\phi}^{\text{Phase difference between waves}}}{2}\right| \tag{35.7}$$

You can also obtain this result without using phasors.

When the two waves are in phase, $\phi = 0$ and $E_P = 2E$. When they are exactly a half-cycle out of phase, $\phi = \pi$ rad $= 180°$, $\cos(\phi/2) = \cos(\pi/2) = 0$, and $E_P = 0$. Thus the superposition of two sinusoidal waves with the same frequency and amplitude but with a phase difference yields a sinusoidal wave with the same frequency and an amplitude between zero and twice the individual amplitudes, depending on the phase difference.

Intensity in Two-Source Interference

To obtain the intensity I at point P, we recall from Section 32.4 that I is equal to the average magnitude of the Poynting vector, S_{av}. For a sinusoidal wave with electric-field amplitude E_P, this is given by Eq. (32.29) with E_{max} replaced by E_P. Thus we can express the intensity in several equivalent forms:

$$I = S_{av} = \frac{E_P^2}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_P^2 = \frac{1}{2}\epsilon_0 c E_P^2 \tag{35.8}$$

The essential content of these expressions is that I is proportional to E_P^2. When we substitute Eq. (35.7) into the last expression in Eq. (35.8), we get

$$I = \frac{1}{2}\epsilon_0 c E_P^2 = 2\epsilon_0 c E^2 \cos^2\frac{\phi}{2} \tag{35.9}$$

In particular, the *maximum* intensity I_0, which occurs at points where the phase difference is zero ($\phi = 0$), is

$$I_0 = 2\epsilon_0 c E^2$$

Note that the maximum intensity I_0 is *four times* (not twice) as great as the intensity $\frac{1}{2}\epsilon_0 c E^2$ from each individual source. Substituting the expression for I_0 into Eq. (35.9), we find

$$\underbrace{I}_{\text{Intensity in two-source interference}} = \underbrace{I_0}_{\text{Maximum intensity}}\cos^2\frac{\overbrace{\phi}^{\text{Phase difference between waves}}}{2} \tag{35.10}$$

The intensity depends on the phase difference ϕ and varies between I_0 and zero. If we average Eq. (35.10) over all possible phase differences, the result is $I_0/2 = \epsilon_0 c E^2$ [the average of $\cos^2(\phi/2)$ is $\frac{1}{2}$]. This is just twice the intensity from each individual source, as we should expect. The total energy output from the two sources isn't changed by the interference effects, but the energy is redistributed (see Section 35.1).

Phase Difference and Path Difference

Our next task is to find the phase difference ϕ between the two fields at any point P. We know that ϕ is proportional to the difference in path length from the two sources to point P. When the path difference is one wavelength, the phase difference is one cycle, and $\phi = 2\pi$ rad $= 360°$. When the path difference is $\lambda/2$, $\phi = \pi$ rad $= 180°$, and so on. That is, the ratio of the phase difference ϕ to 2π is equal to the ratio of the path difference $r_2 - r_1$ to λ:

$$\frac{\phi}{2\pi} = \frac{r_2 - r_1}{\lambda}$$

Path difference   Wave number $= 2\pi/\lambda$

Phase difference in two-source interference $\quad \phi = \frac{2\pi}{\lambda}(r_2 - r_1) = k(r_2 - r_1) \quad$ (35.11)

Wavelength   Distance from source 2   Distance from source 1

We introduced the wave number $k = 2\pi/\lambda$ in Section 15.3.

If the material in the space between the sources and P is anything other than vacuum, we must use the wavelength *in the material* in Eq. (35.11). If λ_0 and k_0 are the wavelength and wave number, respectively, in vacuum and the material has index of refraction n, then

$$\lambda = \frac{\lambda_0}{n} \quad \text{and} \quad k = nk_0 \tag{35.12}$$

Finally, if the point P is far away from the sources in comparison to their separation d, the path difference is given by Eq. (35.3):

$$r_2 - r_1 = d\sin\theta$$

Combining this with Eq. (35.11), we find

$$\phi = k(r_2 - r_1) = kd\sin\theta = \frac{2\pi d}{\lambda}\sin\theta \tag{35.13}$$

When we substitute this into Eq. (35.10), we find

$$I = I_0\cos^2\left(\tfrac{1}{2}kd\sin\theta\right) = I_0\cos^2\left(\frac{\pi d}{\lambda}\sin\theta\right) \quad \text{(intensity far from two sources)} \tag{35.14}$$

Maximum intensity occurs when the cosine has the values ± 1—that is, when

$$\frac{\pi d}{\lambda}\sin\theta = m\pi \qquad (m = 0, \pm 1, \pm 2, \ldots)$$

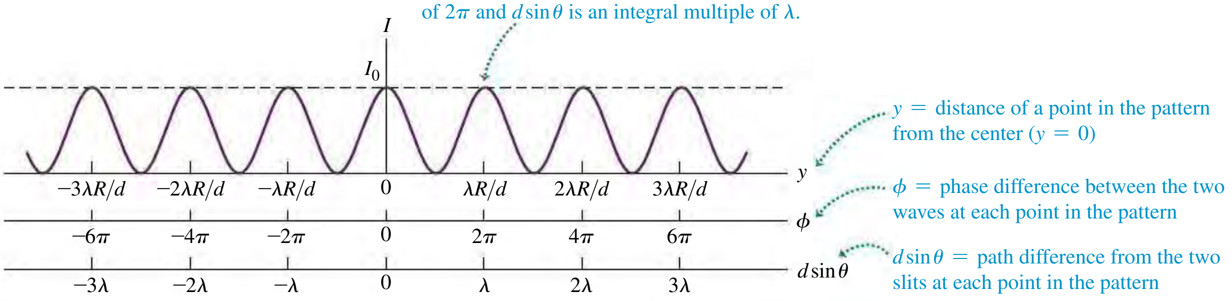
or

$$d\sin\theta = m\lambda$$

in agreement with Eq. (35.4). You can also derive Eq. (35.5) for the zero-intensity directions from Eq. (35.14).

As we noted in Section 35.2, in experiments with light we visualize the interference pattern due to two slits by using a screen placed at a distance R from the slits. We can describe positions on the screen with the coordinate y; the positions of the bright fringes are given by Eq. (35.6), where ordinarily $y \ll R$. In that case, $\sin\theta$ is approximately equal to y/R, and we obtain the following expressions for the intensity at *any* point on the screen as a function of y:

$$I = I_0\cos^2\left(\frac{kdy}{2R}\right) = I_0\cos^2\left(\frac{\pi dy}{\lambda R}\right) \quad \text{(intensity in two-slit interference)} \tag{35.15}$$



35.10 Intensity distribution in the interference pattern from two identical slits.

Figure 35.10 shows a graph of Eq. (35.15); we can compare this with the photographically recorded pattern of Fig. 35.6. All peaks in Fig. 35.10 have the same intensity, while those in Fig. 35.6 fade off as we go away from the center. We'll explore the reasons for this variation in peak intensity in Chapter 36.

EXAMPLE 35.3 A DIRECTIONAL TRANSMITTING ANTENNA ARRAY

SOLUTION

Suppose the two identical radio antennas of Fig. 35.8 are moved to be only 10.0 m apart and the broadcast frequency is increased to $f = 60.0$ MHz. At a distance of 700 m from the point midway between the antennas and in the direction $\theta = 0$ (see Fig. 35.8), the intensity is $I_0 = 0.020\ \text{W/m}^2$. At this same distance, find (a) the intensity in the direction $\theta = 4.0°$; (b) the direction near $\theta = 0$ for which the intensity is $I_0/2$; and (c) the directions in which the intensity is zero.

SOLUTION

IDENTIFY and SET UP: This problem involves the intensity distribution as a function of angle. Because the 700-m distance from the antennas to the point at which the intensity is measured is much greater than the distance $d = 10.0$ m between the antennas, the amplitudes of the waves from the two antennas are very nearly equal. Hence we can use Eq. (35.14) to relate intensity I and angle θ.

EXECUTE: The wavelength is $\lambda = c/f = 5.00$ m. The spacing $d = 10.0$ m between the antennas is just twice the wavelength (as was the case in Example 35.2), so $d/\lambda = 2.00$ and Eq. (35.14) becomes

$$I = I_0\cos^2\left(\frac{\pi d}{\lambda}\sin\theta\right) = I_0\cos^2[(2.00\pi\ \text{rad})\sin\theta]$$

(a) When $\theta = 4.0°$,

$$I = I_0\cos^2[(2.00\pi\ \text{rad})\sin 4.0°] = 0.82I_0$$
$$= (0.82)(0.020\ \text{W/m}^2) = 0.016\ \text{W/m}^2$$

(b) The intensity I equals $I_0/2$ when the cosine in Eq. (35.14) has the value $\pm 1/\sqrt{2}$. The smallest angles at which this occurs correspond to $2.00\pi\sin\theta = \pm\pi/4$ rad, so $\sin\theta = \pm(1/8.00) = \pm 0.125$ and $\theta = \pm 7.2°$.

(c) The intensity is zero when $\cos[(2.00\pi\ \text{rad})\sin\theta] = 0$. This occurs for $2.00\pi\sin\theta = \pm\pi/2, \pm 3\pi/2, \pm 5\pi/2, \ldots$, or $\sin\theta = \pm 0.250, \pm 0.750, \pm 1.25, \ldots$. Values of $\sin\theta$ greater than 1 have no meaning, so the answers are

$$\theta = \pm 14.5°, \pm 48.6°$$

EVALUATE: The condition in part (b) that $I = I_0/2$, so that $(2.00\pi\ \text{rad})\sin\theta = \pm\pi/4$ rad, is also satisfied when $\sin\theta = \pm 0.375, \pm 0.625$, or ± 0.875 so that $\theta = \pm 22.0°, \pm 38.7°$, or $\pm 61.0°$. (Can you verify this?) It would be incorrect to include these angles in the solution, however, because the problem asked for the angle *near* $\theta = 0$ at which $I = I_0/2$. These additional values of θ aren't the ones we're looking for.

TEST YOUR UNDERSTANDING OF SECTION 35.3 A two-slit interference experiment uses coherent light of wavelength 5.00×10^{-7} m. Rank the following points in the interference pattern according to the intensity at each point, from highest to lowest. (i) A point that is closer to one slit than the other by 4.00×10^{-7} m; (ii) a point where the light waves received from the two slits are out of phase by 4.00 rad; (iii) a point that is closer to one slit than the other by 7.50×10^{-7} m; (iv) a point where the light waves received by the two slits are out of phase by 2.00 rad. ❙

35.4 INTERFERENCE IN THIN FILMS

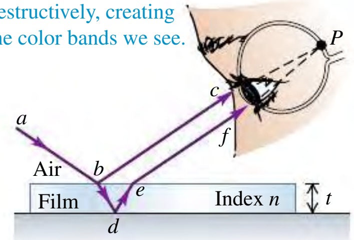
You often see bright bands of color when light reflects from a thin layer of oil floating on water or from a soap bubble (see the photograph that opens this chapter). These are the results of interference. Light waves are reflected from the front and back surfaces of such thin films, and constructive interference between the two reflected waves (with different path lengths) occurs in different places for different wavelengths. **Figure 35.11a** shows the situation. Light shining on the upper surface of a thin film with thickness t is partly reflected at the upper surface (path *abc*). Light *transmitted* through the upper surface is partly reflected at the lower surface (path *abdef*). The two reflected waves come together at point P on the retina of the eye. Depending on the phase relationship, they may interfere constructively or destructively. Different colors have different wavelengths, so the interference may be constructive for some colors and destructive for others. That's why we see colored patterns in the photograph that opens this chapter (which shows a thin film of oil floating on water) and in Fig. 35.11b (which shows thin films of soap solution that make up the bubble walls). The complex shapes of the colored patterns result from variations in the thickness of the film.

35.11 **(a)** A diagram and **(b)** a photograph showing interference of light reflected from a thin film.

(a) Interference between rays reflected from the two surfaces of a thin film

Light reflected from the upper and lower surfaces of the film comes together in the eye at P and undergoes interference.

Some colors interfere constructively and others destructively, creating the color bands we see.



(b) Colorful reflections from a soap bubble

Thin-Film Interference and Phase Shifts During Reflection

Let's look at a simplified situation in which *monochromatic* light reflects from two nearly parallel surfaces at nearly normal incidence. **Figure 35.12** shows two plates of glass separated by a thin wedge, or film, of air. We want to consider interference between the two light waves reflected from the surfaces adjacent to the air wedge. (Reflections also occur at the top surface of the upper plate and the bottom surface of the lower plate; to keep our discussion simple, we won't include these.) The situation is the same as in Fig. 35.11a except that the film (wedge) thickness is not uniform. The path difference between the two waves is just twice the thickness t of the air wedge at each point. At points where $2t$ is an integer number of wavelengths, we expect to see constructive interference and a bright area; where it is a half-integer number of wavelengths, we expect to see destructive interference and a dark area. Where the plates are in contact, there is practically *no* path difference, and we expect a bright area.

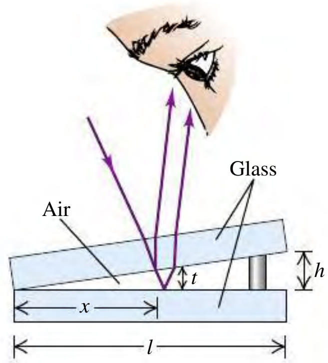
When we carry out the experiment, the bright and dark fringes appear, but they are interchanged! Along the line where the plates are in contact, we find a *dark* fringe, not a bright one. This suggests that one or the other of the reflected waves has undergone a half-cycle phase shift during its reflection. In that case the two waves that are reflected at the line of contact are a half-cycle out of phase even though they have the same path length.

In fact, this phase shift can be predicted from Maxwell's equations and the electromagnetic nature of light. The details of the derivation are beyond our scope, but here is the result. Suppose a light wave with electric-field amplitude E_i is traveling in an optical material with index of refraction n_a. It strikes, at normal incidence, an interface with another optical material with index n_b. The amplitude E_r of the wave reflected from the interface is given by

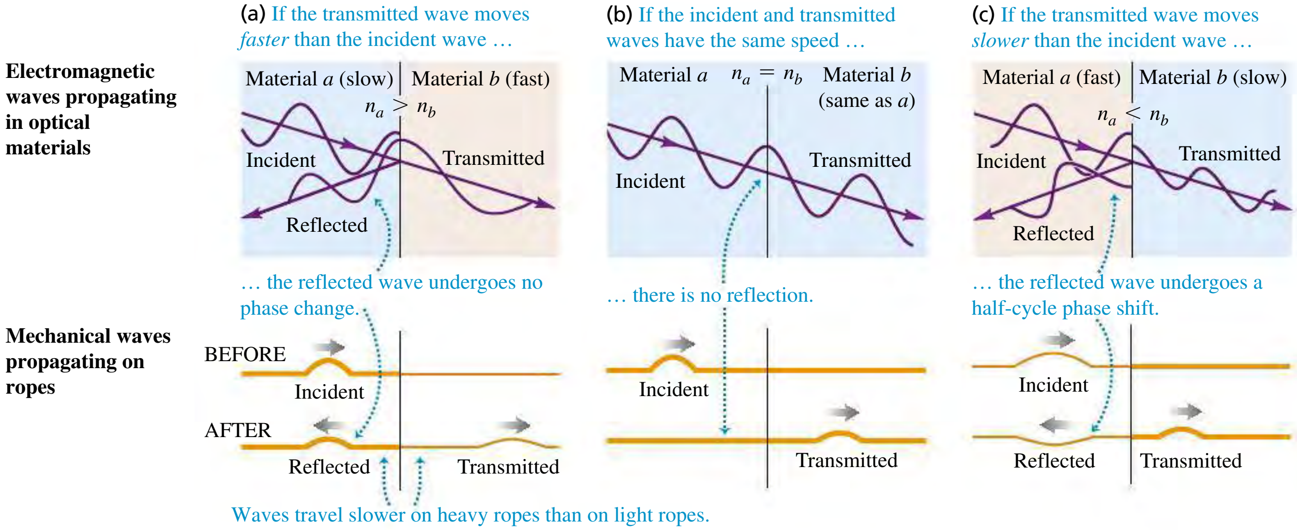
$$E_r = \frac{n_a - n_b}{n_a + n_b} E_i \qquad \text{(normal incidence)} \tag{35.16}$$

This result shows that the incident and reflected amplitudes have the same sign when n_a is larger than n_b and opposite signs when n_b is larger than n_a. Because amplitudes must always be positive or zero, a *negative* value means that the

35.12 Interference between light waves reflected from the two sides of an air wedge separating two glass plates. The angles and the thickness of the air wedge have been exaggerated for clarity; in the text we assume that the light strikes the upper plate at normal incidence and that the distances h and t are much less than l.



35.13 Upper figures: electromagnetic waves striking an interface between optical materials at normal incidence (shown as a small angle for clarity). Lower figures: mechanical wave pulses on ropes.



wave actually undergoes a half-cycle (180°) phase shift. **Figure 35.13** shows three possibilities:

Figure 35.13a: When $n_a > n_b$, light travels more slowly in the first material than in the second. In this case, E_r and E_i have the same sign, and the phase shift of the reflected wave relative to the incident wave is zero. This is analogous to reflection of a transverse mechanical wave on a heavy rope at a point where it is tied to a lighter rope.

Figure 35.13b: When $n_a = n_b$, the amplitude E_r of the reflected wave is zero. In effect there is no interface, so there is *no* reflected wave.

Figure 35.13c: When $n_a < n_b$, light travels more slowly in the second material than in the first. In this case, E_r and E_i have opposite signs, and the phase shift of the reflected wave relative to the incident wave is π rad (a half-cycle). This is analogous to reflection (with inversion) of a transverse mechanical wave on a light rope at a point where it is tied to a heavier rope.

Let's compare with the situation of Fig. 35.12. For the wave reflected from the upper surface of the air wedge, n_a (glass) is greater than n_b, so this wave has zero phase shift. For the wave reflected from the lower surface, n_a (air) is less than n_b (glass), so this wave has a half-cycle phase shift. Waves that are reflected from the line of contact have no path difference to give additional phase shifts, and they interfere destructively; this is what we observe. You can use the above principle to show that for normal incidence, the wave reflected at point b in Fig. 35.11a is shifted by a half-cycle, while the wave reflected at d is not (if there is air below the film).

We can summarize this discussion mathematically. If the film has thickness t, the light is at normal incidence and has wavelength λ in the film; if neither or both of the reflected waves from the two surfaces have a half-cycle reflection phase shift, the conditions for constructive and destructive interference are

Constructive reflection $\quad 2t = m\lambda \quad (m = 0, 1, 2, \ldots) \quad$ (35.17a)

(From thin film, no relative phase shift) $\quad$ Thickness of film $\quad$ Wavelength

Destructive reflection $\quad 2t = \left(m + \frac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \ldots) \quad$ (35.17b)

If *one* of the two waves has a half-cycle reflection phase shift, the conditions for constructive and destructive interference are reversed:

Constructive reflection $$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \ldots) \tag{35.18a}$$

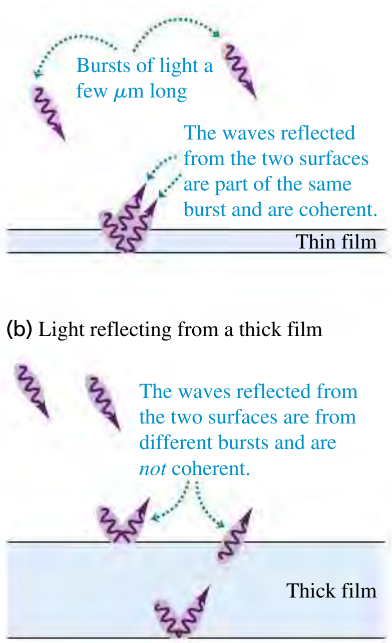
(From thin film, half-cycle phase shift) — t: Thickness of film; λ: Wavelength

Destructive reflection $$2t = m\lambda \quad (m = 0, 1, 2, \ldots) \tag{35.18b}$$

Thin and Thick Films

We have emphasized *thin* films in our discussion because of a principle we introduced in Section 35.1: In order for two waves to cause a steady interference pattern, the waves must be *coherent,* with a definite and constant phase relationship. The sun and light bulbs emit light in a stream of short bursts, each of which is only a few micrometers long ($1 \text{ micrometer} = 1\ \mu m = 10^{-6}$ m). If light reflects from the two surfaces of a thin film, the two reflected waves are part of the same burst (**Fig. 35.14a**). Hence these waves are coherent and interference occurs as we have described. If the film is too thick, however, the two reflected waves will belong to different bursts (Fig. 35.14b). There is no definite phase relationship between different light bursts, so the two waves are incoherent and there is no fixed interference pattern. That's why you see interference colors in light reflected from a soap bubble a few micrometers thick (see Fig. 35.11b), but you do *not* see such colors in the light reflected from a pane of window glass with a thickness of a few millimeters (a thousand times greater).

35.14 **(a)** Light reflecting from a thin film produces a steady interference pattern, but **(b)** light reflecting from a thick film does not.



PROBLEM-SOLVING STRATEGY 35.1 INTERFERENCE IN THIN FILMS

IDENTIFY *the relevant concepts:* Problems with thin films involve interference of two waves, one reflected from the film's front surface and one reflected from the back surface. Typically you will be asked to relate the wavelength, the film thickness, and the index of refraction of the film.

SET UP *the problem* using the following steps:
1. Make a drawing showing the geometry of the film. Identify the materials that adjoin the film; their properties determine whether one or both of the reflected waves have a half-cycle phase shift.
2. Identify the target variable.

EXECUTE *the solution* as follows:
1. Apply the rule for phase changes to each reflected wave: There is a half-cycle phase shift when $n_b > n_a$ and none when $n_b < n_a$.
2. If *neither* reflected wave undergoes a phase shift, or if *both* do, use Eqs. (35.17). If only one reflected wave undergoes a phase shift, use Eqs. (35.18).
3. Solve the resulting equation for the target variable. Use the wavelength $\lambda = \lambda_0/n$ of light *in the film* in your calculations, where n is the index of refraction of the film. (For air, $n = 1.000$ to four-figure precision.)
4. If you are asked about the wave that is transmitted through the film, remember that minimum intensity in the reflected wave corresponds to maximum *transmitted* intensity, and vice versa.

EVALUATE *your answer:* Interpret your results by examining what would happen if the wavelength were changed or if the film had a different thickness.

EXAMPLE 35.4 THIN-FILM INTERFERENCE I

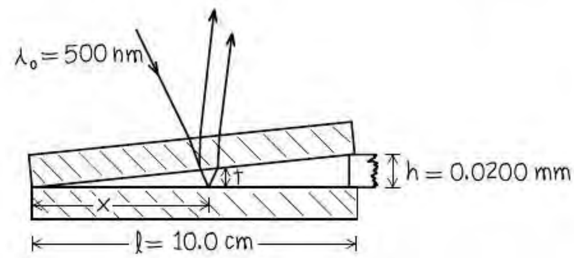
SOLUTION

Suppose the two glass plates in Fig. 35.12 are two microscope slides 10.0 cm long. At one end they are in contact; at the other end they are separated by a piece of paper 0.0200 mm thick. What is the spacing of the interference fringes seen by reflection? Is the fringe at the line of contact bright or dark? Assume monochromatic light with a wavelength in air of $\lambda = \lambda_0 = 500$ nm.

SOLUTION

IDENTIFY and SET UP: **Figure 35.15** depicts the situation. We'll consider only interference between the light reflected from the

35.15 Our sketch for this problem.



Continued

upper and lower surfaces of the air wedge between the microscope slides. [The top slide has a relatively great thickness, about 1 mm, so we can ignore interference between the light reflected from its upper and lower surfaces (see Fig. 35.14b).] Light travels more slowly in the glass of the slides than it does in air. Hence the wave reflected from the upper surface of the air wedge has no phase shift (see Fig. 35.13a), while the wave reflected from the lower surface has a half-cycle phase shift (see Fig. 35.13c).

EXECUTE: Since only one of the reflected waves undergoes a phase shift, the condition for *destructive* interference (a dark fringe) is Eq. (35.18b):

$$2t = m\lambda_0 \qquad (m = 0, 1, 2, \ldots)$$

From similar triangles in Fig. 35.15 the thickness t of the air wedge at each point is proportional to the distance x from the line of contact:

$$\frac{t}{x} = \frac{h}{l}$$

Combining this with Eq. (35.18b), we find

$$\frac{2xh}{l} = m\lambda_0$$

$$x = m\frac{l\lambda_0}{2h} = m\frac{(0.100\text{ m})(500 \times 10^{-9}\text{ m})}{(2)(0.0200 \times 10^{-3}\text{ m})} = m(1.25\text{ mm})$$

Successive dark fringes, corresponding to $m = 1, 2, 3, \ldots$, are spaced 1.25 mm apart. Substituting $m = 0$ into this equation gives $x = 0$, which is where the two slides touch (at the left-hand side of Fig. 35.15). Hence there is a dark fringe at the line of contact.

EVALUATE: Our result shows that the fringe spacing is proportional to the wavelength of the light used; the fringes would be farther apart with red light (larger λ_0) than with blue light (smaller λ_0). If we use white light, the reflected light at any point is a mixture of wavelengths for which constructive interference occurs; the wavelengths that interfere destructively are weak or absent in the reflected light. (This same effect explains the colors seen when a soap bubble is illuminated by white light, as in Fig. 35.11b).

EXAMPLE 35.5 THIN-FILM INTERFERENCE II

SOLUTION

Suppose the glass plates of Example 35.4 have $n = 1.52$ and the space between plates contains water ($n = 1.33$) instead of air. What happens now?

SOLUTION

IDENTIFY and SET UP: The index of refraction of the water wedge is still less than that of the glass on either side of it, so the phase shifts are the same as in Example 35.4. Once again we use Eq. (35.18b) to find the positions of the dark fringes; the only difference is that the wavelength λ in this equation is now the wavelength in water instead of in air.

EXECUTE: In the film of water ($n = 1.33$), the wavelength is $\lambda = \lambda_0/n = (500\text{ nm})/(1.33) = 376$ nm. When we replace λ_0 by λ in the expression from Example 35.4 for the position x of the mth dark fringe, we find that the fringe spacing is reduced by the same factor of 1.33 and is equal to 0.940 mm. There is still a dark fringe at the line of contact.

EVALUATE: Can you see that to obtain the same fringe spacing as in Example 35.4, the dimension h in Fig. 35.15 would have to be reduced to $(0.0200\text{ mm})/1.33 = 0.0150$ mm? This shows that what matters in thin-film interference is the *ratio* t/λ between film thickness and wavelength. [Consider Eqs. (35.17) and (35.18).]

EXAMPLE 35.6 THIN-FILM INTERFERENCE III

SOLUTION

Suppose the upper of the two plates of Example 35.4 is a plastic with $n = 1.40$, the wedge is filled with a silicone grease with $n = 1.50$, and the bottom plate is a dense flint glass with $n = 1.60$. What happens now?

SOLUTION

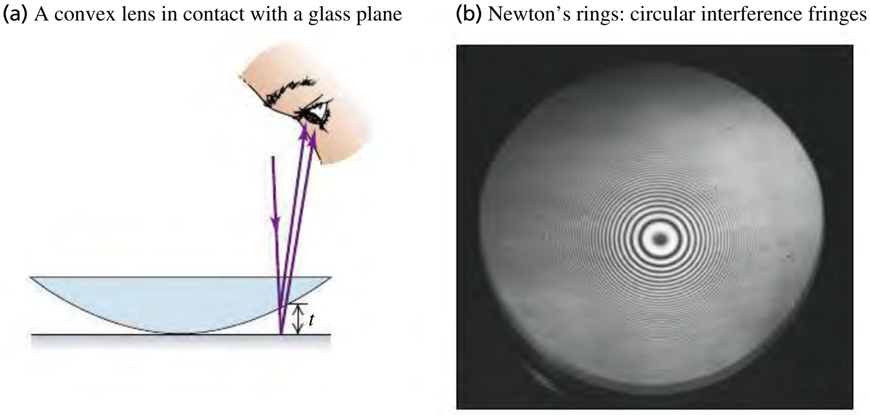
IDENTIFY and SET UP: The geometry is again the same as shown in Fig. 35.15, but now half-cycle phase shifts occur at *both* surfaces of the grease wedge (see Fig. 35.13c). Hence there is no *relative* phase shift and we must use Eq. (35.17b) to find the positions of the dark fringes.

EXECUTE: The value of λ to use in Eq. (35.17b) is the wavelength in the silicone grease, $\lambda = \lambda_0/n = (500\text{ nm})/1.50 = 333$ nm. You can readily show that the fringe spacing is 0.833 mm. Note that the two reflected waves from the line of contact are in phase (they both undergo the same phase shift), so the line of contact is at a *bright* fringe.

EVALUATE: What would happen if you carefully removed the upper microscope slide so that the grease wedge retained its shape? There would still be half-cycle phase changes at the upper and lower surfaces of the wedge, so the pattern of fringes would be the same as with the upper slide present.

Newton's Rings

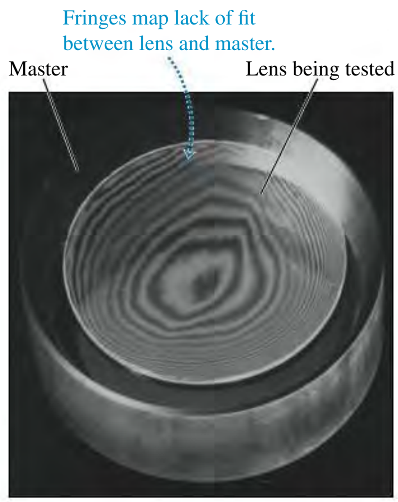
Figure 35.16a shows the convex surface of a lens in contact with a plane glass plate. A thin film of air is formed between the two surfaces. When you view the setup with monochromatic light, you see circular interference fringes (Fig. 35.16b). These were studied by Newton and are called **Newton's rings.**



35.16 (a) Air film between a convex lens and a plane surface. The thickness of the film t increases from zero as we move out from the center, giving (b) a series of alternating dark and bright rings for monochromatic light.

We can use interference fringes to compare the surfaces of two optical parts by placing the two in contact and observing the interference fringes. **Figure 35.17** is a photograph made during the grinding of a telescope objective lens. The lower, larger-diameter, thicker disk is the correctly shaped master, and the smaller, upper disk is the lens under test. The "contour lines" are Newton's interference fringes; each one indicates an additional distance between the specimen and the master of one half-wavelength. At 10 lines from the center spot the distance between the two surfaces is five wavelengths, or about 0.003 mm. This isn't very good; high-quality lenses are routinely ground with a precision of less than one wavelength. The surface of the primary mirror of the Hubble Space Telescope was ground to a precision of better than $\frac{1}{50}$ wavelength. Unfortunately, it was ground to incorrect specifications, creating one of the most precise errors in the history of optical technology (see Section 34.2).

35.17 The surface of a telescope objective lens under inspection during manufacture.



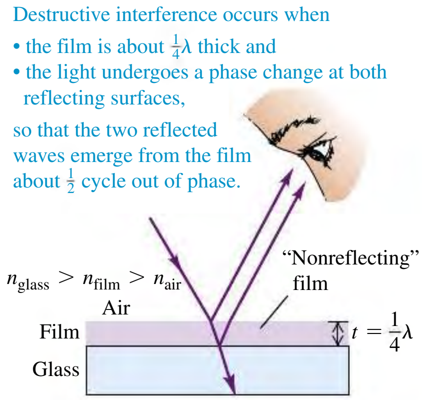
Nonreflective and Reflective Coatings

Nonreflective coatings for lens surfaces make use of thin-film interference. A thin layer or film of hard transparent material with an index of refraction smaller than that of the glass is deposited on the lens surface, as in **Fig. 35.18**. Light is reflected from both surfaces of the layer. In both reflections the light is reflected from a medium of greater index than that in which it is traveling, so the same phase change occurs in both reflections. If the film thickness is a quarter (one-fourth) of the wavelength *in the film* (assuming normal incidence), the total path difference is a half-wavelength. Light reflected from the first surface is then a half-cycle out of phase with light reflected from the second, and there is destructive interference.

The thickness of the nonreflective coating can be a quarter-wavelength for only one particular wavelength. This is usually chosen in the central yellow-green portion of the spectrum ($\lambda = 550$ nm), where the eye is most sensitive. Then there is somewhat more reflection at both longer (red) and shorter (blue) wavelengths, and the reflected light has a purple hue. The overall reflection from a lens or prism surface can be reduced in this way from 4–5% to less than 1%. This also increases the net amount of light that is *transmitted* through the lens, since light that is not reflected will be transmitted. The same principle is used to minimize reflection from silicon photovoltaic solar cells ($n = 3.5$) by use of a thin surface layer of silicon monoxide (SiO, $n = 1.45$); this helps to increase the amount of light that actually reaches the solar cells.

If a quarter-wavelength thickness of a material with an index of refraction *greater* than that of glass is deposited on glass, then the reflectivity is *increased,* and the deposited material is called a **reflective coating.** In this case there is a half-cycle phase shift at the air–film interface but none at the film–glass interface, and reflections from the two sides of the film interfere constructively.

35.18 A nonreflective coating has an index of refraction intermediate between those of glass and air.



For example, a coating with refractive index 2.5 causes 38% of the incident energy to be reflected, compared with 4% or so with no coating. By use of multiple-layer coatings, it is possible to achieve nearly 100% transmission or reflection for particular wavelengths. Some practical applications of these coatings are for color separation in television cameras and for infrared “heat reflectors” in motion-picture projectors, solar cells, and astronauts’ visors.

SOLUTION

EXAMPLE 35.7 A NONREFLECTIVE COATING

A common lens coating material is magnesium fluoride (MgF_2), with $n = 1.38$. What thickness should a nonreflective coating have for 550-nm light if it is applied to glass with $n = 1.52$?

SOLUTION

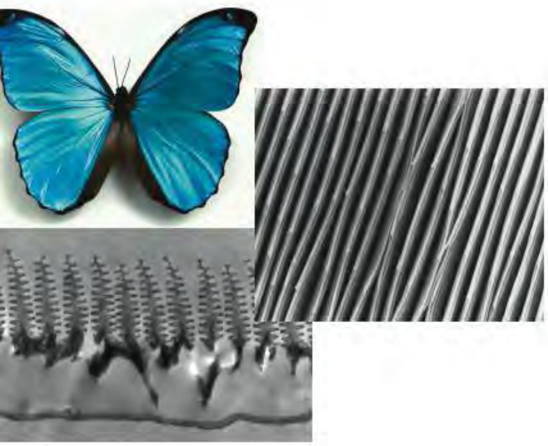
IDENTIFY and SET UP: This coating is of the sort shown in Fig. 35.18. The thickness must be one-quarter of the wavelength of this light *in the coating*.

EXECUTE: The wavelength in air is $\lambda_0 = 550$ nm, so its wavelength in the MgF_2 coating is $\lambda = \lambda_0/n = (550\text{ nm})/1.38 = 400$ nm. The coating thickness should be one-quarter of this, or $\lambda/4 =$ 100 nm.

EVALUATE: This is a very thin film, no more than a few hundred molecules thick. Note that this coating is *reflective* for light whose wavelength is *twice* the coating thickness; light of that wavelength reflected from the coating’s lower surface travels one wavelength farther than light reflected from the upper surface, so the two waves are in phase and interfere constructively. This occurs for light with a wavelength in MgF_2 of 200 nm and a wavelength in air of $(200\text{ nm})(1.38) = 276$ nm. This is an ultraviolet wavelength (see Section 32.1), so designers of optical lenses with nonreflective coatings need not worry about such enhanced reflection.

TEST YOUR UNDERSTANDING OF SECTION 35.4 A thin layer of benzene $(n = 1.501)$ lies on top of a sheet of fluorite $(n = 1.434)$. It is illuminated from above with light whose wavelength in benzene is 400 nm. Which of the following possible thicknesses of the benzene layer will maximize the brightness of the reflected light? (i) 100 nm; (ii) 200 nm; (iii) 300 nm; (iv) 400 nm. ❙

BIO Application Interference and Butterfly Wings Many of the most brilliant colors in the animal world are created by *interference* rather than by pigments. These photos show the butterfly *Morpho rhetenor* and the microscopic scales that cover the upper surfaces of its wings. The scales have a profusion of tiny ridges (middle photo); these carry regularly spaced flanges (bottom photo) that function as reflectors. These are spaced so that the reflections interfere constructively for blue light. The multilayered structure reflects 70% of the blue light that strikes it, giving the wings a mirrorlike brilliance. (The undersides of the wings do not have these structures and are a dull brown.)

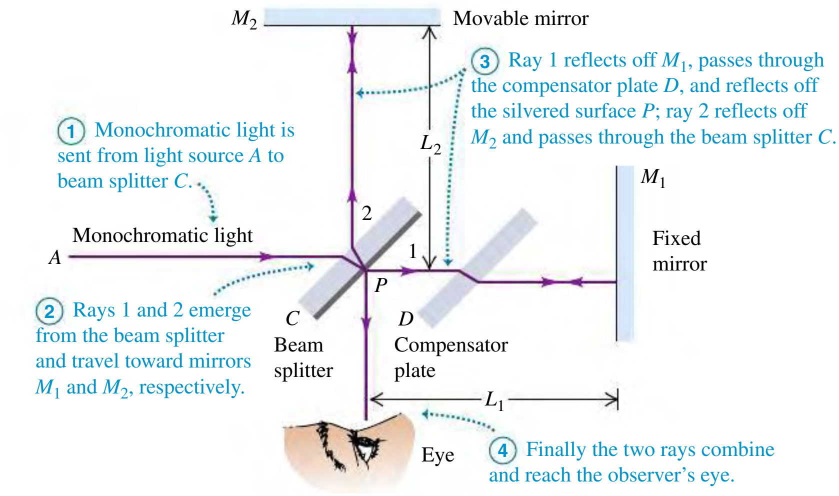


35.5 THE MICHELSON INTERFEROMETER

An important experimental device that uses interference is the **Michelson interferometer.** Michelson interferometers are used to make precise measurements of wavelengths and of very small distances, such as the minute changes in thickness of an axon when a nerve impulse propagates along its length. Like the Young two-slit experiment, a Michelson interferometer takes monochromatic light from a single source and divides it into two waves that follow different paths. In Young’s experiment, this is done by sending part of the light through one slit and part through another; in a Michelson interferometer a device called a *beam splitter* is used. Interference occurs in both experiments when the two light waves are recombined.

How a Michelson Interferometer Works

Figure 35.19 shows the principal components of a Michelson interferometer. A ray of light from a monochromatic source A strikes the beam splitter C, which is a glass plate with a thin coating of silver on its right side. Part of the light (ray 1) passes through the silvered surface and the compensator plate D and is reflected from mirror M_1. It then returns through D and is reflected from the silvered surface of C to the observer. The remainder of the light (ray 2) is reflected from the silvered surface at point P to the mirror M_2 and back through C to the observer’s eye. The purpose of the compensator plate D is to ensure that rays 1 and 2 pass through the same thickness of glass; plate D is cut from the same piece of glass as plate C, so their thicknesses are identical to within a fraction of a wavelength.



35.19 A schematic Michelson interferometer. The observer sees an interference pattern that results from the difference in path lengths for rays 1 and 2.

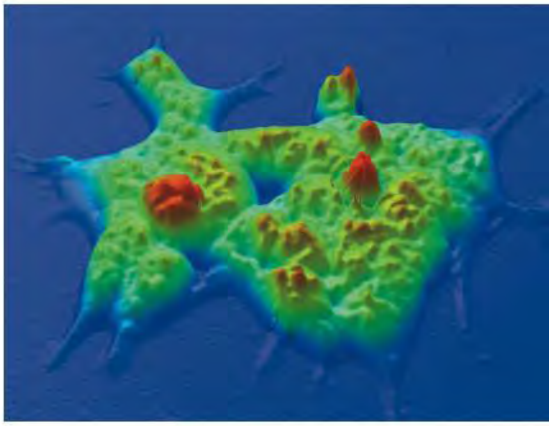
The whole apparatus in Fig. 35.19 is mounted on a very rigid frame, and the position of mirror M_2 can be adjusted with a fine, very accurate micrometer screw. If the distances L_1 and L_2 are exactly equal and the mirrors M_1 and M_2 are exactly at right angles, the virtual image of M_1 formed by reflection at the silvered surface of plate C coincides with mirror M_2. If L_1 and L_2 are *not* exactly equal, the image of M_1 is displaced slightly from M_2; and if the mirrors are not exactly perpendicular, the image of M_1 makes a slight angle with M_2. Then the mirror M_2 and the virtual image of M_1 play the same roles as the two surfaces of a wedge-shaped thin film (see Section 35.4), and light reflected from these surfaces forms the same sort of interference fringes.

Suppose the angle between mirror M_2 and the virtual image of M_1 is just large enough that five or six vertical fringes are present in the field of view. If we now move the mirror M_2 slowly either backward or forward a distance $\lambda/2$, the difference in path length between rays 1 and 2 changes by λ, and each fringe moves to the left or right a distance equal to the fringe spacing. If we observe the fringe positions through a telescope with a crosshair eyepiece and m fringes cross the crosshairs when we move the mirror a distance y, then

$$y = m\frac{\lambda}{2} \qquad \text{or} \qquad \lambda = \frac{2y}{m} \tag{35.19}$$

If m is several thousand, the distance y is large enough that it can be measured with good accuracy, and we can obtain an accurate value for the wavelength λ. Alternatively, if the wavelength is known, a distance y can be measured by simply counting fringes when M_2 is moved by this distance. In this way, distances that are comparable to a wavelength of light can be measured with relative ease.

BIO Application Imaging Cells with a Michelson Interferometer This false-color image of a human colon cancer cell was made by using a microscope that was mated to a Michelson interferometer. The cell is in one arm of the interferometer, and light passing through the cell undergoes a phase shift that depends on the cell thickness and the organelles within the cell. The fringe pattern can then be used to construct a three-dimensional view of the cell. Scientists have used this technique to observe how different types of cells behave when prodded by microscopic probes. Cancer cells turn out to be "softer" than normal cells, a distinction that may make cancer stem cells easier to identify.



The Michelson-Morley Experiment

The original application of the Michelson interferometer was to the historic **Michelson-Morley experiment.** Before the electromagnetic theory of light became established, most physicists thought that the propagation of light waves occurred in a medium called the **ether,** which was believed to permeate all space. In 1887 the American scientists Albert Michelson and Edward Morley used the Michelson interferometer in an attempt to detect the motion of the earth through the ether. Suppose the interferometer in Fig. 35.19 is moving from left to right relative to the ether. According to the ether theory, this would lead to changes in the speed of light in the portions of the path shown as horizontal lines in the figure. There would be fringe shifts relative to the positions that the fringes would have if the instrument were at rest in the ether. Then when the entire instrument was rotated 90°, the other portions of the paths would be similarly affected, giving a fringe shift in the opposite direction.

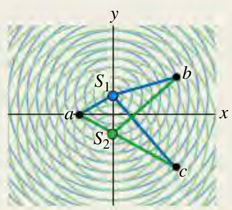
Michelson and Morley expected that the motion of the earth through the ether would cause a shift of about four-tenths of a fringe when the instrument was rotated. The shift that was actually observed was less than a hundredth of a fringe and, within the limits of experimental uncertainty, appeared to be exactly zero. Despite its orbital motion around the sun, the earth appeared to be *at rest* relative to the ether. This negative result baffled physicists until 1905, when Albert Einstein developed the special theory of relativity (which we will study in detail in Chapter 37). Einstein postulated that the speed of a light wave in vacuum has the same magnitude c relative to *all* inertial reference frames, no matter what their velocity may be relative to each other. The presumed ether then plays no role, and the concept of an ether has been abandoned.

TEST YOUR UNDERSTANDING OF SECTION 35.5 You are observing the pattern of fringes in a Michelson interferometer like that shown in Fig. 35.19. If you change the index of refraction (but not the thickness) of the compensator plate, will the pattern change? ❙

CHAPTER 35 SUMMARY

SOLUTIONS TO ALL EXAMPLES

Interference and coherent sources: Monochromatic light is light with a single frequency. Coherence is a definite, unchanging phase relationship between two waves. The overlap of waves from two coherent sources of monochromatic light forms an interference pattern. The principle of superposition states that the total wave disturbance at any point is the sum of the disturbances from the separate waves.

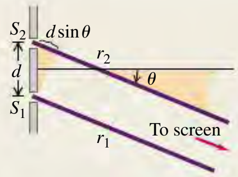


Two-source interference of light: When two sources are in phase, constructive interference occurs where the difference in path length from the two sources is zero or an integer number of wavelengths; destructive interference occurs where the path difference is a half-integer number of wavelengths. If two sources separated by a distance d are both very far from a point P, and the line from the sources to P makes an angle θ with the line perpendicular to the line of the sources, then the condition for constructive interference at P is Eq. (35.4). The condition for destructive interference is Eq. (35.5). When θ is very small, the position y_m of the mth bright fringe on a screen located a distance R from the sources is given by Eq. (35.6). (See Examples 35.1 and 35.2.)

$$d\sin\theta = m\lambda \quad (m = 0, \pm 1, \pm 2, \ldots) \tag{35.4}$$
(constructive interference)

$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, \pm 1, \pm 2, \ldots) \tag{35.5}$$
(destructive interference)

$$y_m = R\frac{m\lambda}{d} \quad (m = 0, \pm 1, \pm 2, \ldots) \tag{35.6}$$
(bright fringes)

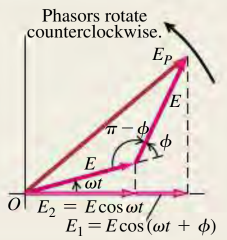


Intensity in interference patterns: When two sinusoidal waves with equal amplitude E and phase difference ϕ are superimposed, the resultant amplitude E_P and intensity I are given by Eqs. (35.7) and (35.10), respectively. If the two sources emit in phase, the phase difference ϕ at a point P (located a distance r_1 from source 1 and a distance r_2 from source 2) is directly proportional to the path difference $r_2 - r_1$. (See Example 35.3.)

$$E_P = 2E\left|\cos\frac{\phi}{2}\right| \tag{35.7}$$

$$I = I_0\cos^2\frac{\phi}{2} \tag{35.10}$$

$$\phi = \frac{2\pi}{\lambda}(r_2 - r_1) = k(r_2 - r_1) \tag{35.11}$$



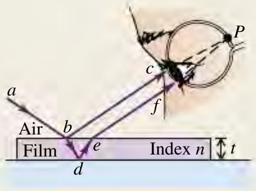
Interference in thin films: When light is reflected from both sides of a thin film of thickness t and no phase shift occurs at either surface, constructive interference between the reflected waves occurs when $2t$ is equal to an integral number of wavelengths. If a half-cycle phase shift occurs at one surface, this is the condition for destructive interference. A half-cycle phase shift occurs during reflection whenever the index of refraction in the second material is greater than that in the first. (See Examples 35.4–35.7.)

$$2t = m\lambda \quad (m = 0, 1, 2, \dots) \tag{35.17a}$$
(constructive reflection from thin film, no relative phase shift)

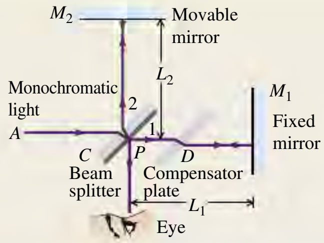
$$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \dots) \tag{35.17b}$$
(destructive reflection from thin film, no relative phase shift)

$$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \dots) \tag{35.18a}$$
(constructive reflection from thin film, half-cycle phase shift)

$$2t = m\lambda \quad (m = 0, 1, 2, \dots) \tag{35.18b}$$
(destructive reflection from thin film, half-cycle phase shift)



Michelson interferometer: The Michelson interferometer uses a monochromatic light source and can be used for high-precision measurements of wavelengths. Its original purpose was to detect motion of the earth relative to a hypothetical ether, the supposed medium for electromagnetic waves. The ether has never been detected, and the concept has been abandoned; the speed of light is the same relative to all observers. This is part of the foundation of the special theory of relativity.



BRIDGING PROBLEM THIN-FILM INTERFERENCE IN AN OIL SLICK

SOLUTION

An oil tanker spills a large amount of oil ($n = 1.45$) into the sea ($n = 1.33$). (a) If you look down onto the oil spill from overhead, what predominant wavelength of light do you see at a point where the oil is 380 nm thick? What color is the light? (*Hint:* See Table 32.1.) (b) In the water under the slick, what visible wavelength (as measured in air) is predominant in the transmitted light at the same place in the slick as in part (a)?

SOLUTION GUIDE

IDENTIFY and SET UP

1. The oil layer acts as a thin film, so we must consider interference between light that reflects from the top and bottom surfaces of the oil. If a wavelength is prominent in the *transmitted* light, there is destructive interference for that wavelength in the *reflected* light.
2. Choose the appropriate interference equations that relate the thickness of the oil film and the wavelength of light. Take account of the indexes of refraction of the air, oil, and water.

EXECUTE

3. For part (a), find the wavelengths for which there is constructive interference as seen from above the oil film. Which of these are in the visible spectrum?
4. For part (b), find the visible wavelength for which there is destructive interference as seen from above the film. (This will ensure that there is substantial transmitted light at the wavelength.)

EVALUATE

5. If a diver below the water's surface shines a light up at the bottom of the oil film, at what wavelengths would there be constructive interference in the light that reflects back downward?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q35.1 A two-slit interference experiment is set up, and the fringes are displayed on a screen. Then the whole apparatus is immersed in the nearest swimming pool. How does the fringe pattern change?

Q35.2 Could an experiment similar to Young's two-slit experiment be performed with sound? How might this be carried out? Does it matter that sound waves are longitudinal and electromagnetic waves are transverse? Explain.

Q35.3 Monochromatic coherent light passing through two thin slits is viewed on a distant screen. Are the bright fringes equally spaced on the screen? If so, why? If not, which ones are closest to being equally spaced?

Q35.4 In a two-slit interference pattern on a distant screen, are the bright fringes midway between the dark fringes? Is this ever a good approximation?

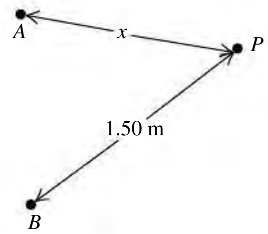
Q35.5 Would the headlights of a distant car form a two-source interference pattern? If so, how might it be observed? If not, why not?
Q35.6 The two sources S_1 and S_2 shown in Fig. 35.3 emit waves of the same wavelength λ and are in phase with each other. Suppose S_1 is a weaker source, so that the waves emitted by S_1 have half the amplitude of the waves emitted by S_2. How would this affect the positions of the antinodal lines and nodal lines? Would there be total reinforcement at points on the antinodal curves? Would there be total cancellation at points on the nodal curves? Explain your answers.
Q35.7 Could the Young two-slit interference experiment be performed with gamma rays? If not, why not? If so, discuss differences in the experimental design compared to the experiment with visible light.
Q35.8 Coherent red light illuminates two narrow slits that are 25 cm apart. Will a two-slit interference pattern be observed when the light from the slits falls on a screen? Explain.
Q35.9 Coherent light with wavelength λ falls on two narrow slits separated by a distance d. If d is less than some minimum value, no dark fringes are observed. Explain. In terms of λ, what is this minimum value of d?
Q35.10 A fellow student, who values memorizing equations above understanding them, combines Eqs. (35.4) and (35.13) to "prove" that ϕ can *only* equal $2\pi m$. How would you explain to this student that ϕ can have values other than $2\pi m$?
Q35.11 If the monochromatic light shown in Fig. 35.5a were replaced by white light, would a two-slit interference pattern be seen on the screen? Explain.
Q35.12 In using the superposition principle to calculate intensities in interference patterns, could you add the intensities of the waves instead of their amplitudes? Explain.
Q35.13 A glass windowpane with a thin film of water on it reflects less than when it is perfectly dry. Why?
Q35.14 A *very* thin soap film ($n = 1.33$), whose thickness is much less than a wavelength of visible light, looks black; it appears to reflect no light at all. Why? By contrast, an equally thin layer of soapy water ($n = 1.33$) on glass ($n = 1.50$) appears quite shiny. Why is there a difference?
Q35.15 Interference can occur in thin films. Why is it important that the films be *thin*? Why don't you get these effects with a relatively *thick* film? Where should you put the dividing line between "thin" and "thick"? Explain your reasoning.
Q35.16 If we shine white light on an air wedge like that shown in Fig. 35.12, the colors that are weak in the light *reflected* from any point along the wedge are strong in the light *transmitted* through the wedge. Explain why this should be so.
Q35.17 Monochromatic light is directed at normal incidence on a thin film. There is destructive interference for the reflected light, so the intensity of the reflected light is very low. What happened to the energy of the incident light?
Q35.18 When a thin oil film spreads out on a puddle of water, the thinnest part of the film looks dark in the resulting interference pattern. What does this tell you about the relative magnitudes of the refractive indexes of oil and water?

EXERCISES

Section 35.1 Interference and Coherent Sources

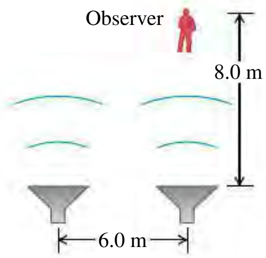
35.1 • Two small stereo speakers A and B that are 1.40 m apart are sending out sound of wavelength 34 cm in all directions and all in phase. A person at point P starts out equidistant from both speakers and walks so that he is always 1.50 m from speaker B (**Fig. E35.1**). For what values of x will the sound this person hears be (a) maximally reinforced, (b) cancelled? Limit your solution to the cases where $x \leq 1.50$ m.

Figure **E35.1**



35.2 •• Two speakers that are 15.0 m apart produce in-phase sound waves of frequency 250.0 Hz in a room where the speed of sound is 340.0 m/s. A woman starts out at the midpoint between the two speakers. The room's walls and ceiling are covered with absorbers to eliminate reflections, and she listens with only one ear for best precision. (a) What does she hear: constructive or destructive interference? Why? (b) She now walks slowly toward one of the speakers. How far from the center must she walk before she first hears the sound reach a minimum intensity? (c) How far from the center must she walk before she first hears the sound maximally enhanced?
35.3 •• A radio transmitting station operating at a frequency of 120 MHz has two identical antennas that radiate in phase. Antenna B is 9.00 m to the right of antenna A. Consider point P between the antennas and along the line connecting them, a horizontal distance x to the right of antenna A. For what values of x will constructive interference occur at point P?
35.4 • **Radio Interference.** Two radio antennas A and B radiate in phase. Antenna B is 120 m to the right of antenna A. Consider point Q along the extension of the line connecting the antennas, a horizontal distance of 40 m to the right of antenna B. The frequency, and hence the wavelength, of the emitted waves can be varied. (a) What is the longest wavelength for which there will be destructive interference at point Q? (b) What is the longest wavelength for which there will be constructive interference at point Q?
35.5 • Two speakers, emitting identical sound waves of wavelength 2.0 m in phase with each other, and an observer are located as shown in **Fig. E35.5**. (a) At the observer's location, what is the path difference for waves from the two speakers? (b) Will the sound waves interfere constructively or destructively at the observer's location—or something in between constructive and destructive? (c) Suppose the observer now increases her distance from the closest speaker to 17.0 m, staying directly in front of the same speaker as initially. Answer the questions of parts (a) and (b) for this new situation.

Figure **E35.5**



35.6 • Two light sources can be adjusted to emit monochromatic light of any visible wavelength. The two sources are coherent, 2.04 μm apart, and in line with an observer, so that one source is 2.04 μm farther from the observer than the other. (a) For what visible wavelengths (380 to 750 nm) will the observer see the brightest light, owing to constructive interference? (b) How would your answers to part (a) be affected if the two sources were not in line with the observer, but were still arranged so that one source is 2.04 μm farther away from the observer than the other? (c) For what visible wavelengths will there be *destructive* interference at the location of the observer?

Section 35.2 Two-Source Interference of Light

35.7 • Young's experiment is performed with light from excited helium atoms ($\lambda = 502$ nm). Fringes are measured carefully on a screen 1.20 m away from the double slit, and the center of the 20th fringe (not counting the central bright fringe) is found to be 10.6 mm from the center of the central bright fringe. What is the separation of the two slits?

35.8 •• Coherent light with wavelength 450 nm falls on a pair of slits. On a screen 1.80 m away, the distance between dark fringes is 3.90 mm. What is the slit separation?

35.9 •• Two slits spaced 0.450 mm apart are placed 75.0 cm from a screen. What is the distance between the second and third dark lines of the interference pattern on the screen when the slits are illuminated with coherent light with a wavelength of 500 nm?

35.10 •• If the entire apparatus of Exercise 35.9 (slits, screen, and space in between) is immersed in water, what then is the distance between the second and third dark lines?

35.11 •• Two thin parallel slits that are 0.0116 mm apart are illuminated by a laser beam of wavelength 585 nm. (a) On a very large distant screen, what is the *total* number of bright fringes (those indicating complete constructive interference), including the central fringe and those on both sides of it? Solve this problem without calculating all the angles! (*Hint:* What is the largest that $\sin\theta$ can be? What does this tell you is the largest value of m?) (b) At what angle, relative to the original direction of the beam, will the fringe that is most distant from the central bright fringe occur?

35.12 • Coherent light with wavelength 400 nm passes through two very narrow slits that are separated by 0.200 mm, and the interference pattern is observed on a screen 4.00 m from the slits. (a) What is the width (in mm) of the central interference maximum? (b) What is the width of the first-order bright fringe?

35.13 •• Two very narrow slits are spaced 1.80 μm apart and are placed 35.0 cm from a screen. What is the distance between the first and second dark lines of the interference pattern when the slits are illuminated with coherent light with $\lambda = 550$ nm? (*Hint:* The angle θ in Eq. (35.5) is *not* small.)

35.14 •• Coherent light that contains two wavelengths, 660 nm (red) and 470 nm (blue), passes through two narrow slits that are separated by 0.300 mm. Their interference pattern is observed on a screen 4.00 m from the slits. What is the distance on the screen between the first-order bright fringes for the two wavelengths?

35.15 •• Coherent light with wavelength 600 nm passes through two very narrow slits and the interference pattern is observed on a screen 3.00 m from the slits. The first-order bright fringe is at 4.84 mm from the center of the central bright fringe. For what wavelength of light will the first-order dark fringe be observed at this same point on the screen?

35.16 •• Coherent light of frequency 6.32×10^{14} Hz passes through two thin slits and falls on a screen 85.0 cm away. You observe that the third bright fringe occurs at ± 3.11 cm on either side of the central bright fringe. (a) How far apart are the two slits? (b) At what distance from the central bright fringe will the third dark fringe occur?

Section 35.3 Intensity in Interference Patterns

35.17 •• In a two-slit interference pattern, the intensity at the peak of the central maximum is I_0. (a) At a point in the pattern where the phase difference between the waves from the two slits is 60.0°, what is the intensity? (b) What is the path difference for 480-nm light from the two slits at a point where the phase difference is 60.0°?

35.18 • Coherent sources A and B emit electromagnetic waves with wavelength 2.00 cm. Point P is 4.86 m from A and 5.24 m from B. What is the phase difference at P between these two waves?

35.19 • Coherent light with wavelength 500 nm passes through narrow slits separated by 0.340 mm. At a distance from the slits large compared to their separation, what is the phase difference (in radians) in the light from the two slits at an angle of 23.0° from the centerline?

35.20 • Two slits spaced 0.260 mm apart are 0.900 m from a screen and illuminated by coherent light of wavelength 660 nm. The intensity at the center of the central maximum ($\theta = 0°$) is I_0. What is the distance on the screen from the center of the central maximum (a) to the first minimum; (b) to the point where the intensity has fallen to $I_0/2$?

35.21 • Consider two antennas separated by 9.00 m that radiate in phase at 120 MHz, as described in Exercise 35.3. A receiver placed 150 m from both antennas measures an intensity I_0. The receiver is moved so that it is 1.8 m closer to one antenna than to the other. (a) What is the phase difference ϕ between the two radio waves produced by this path difference? (b) In terms of I_0, what is the intensity measured by the receiver at its new position?

35.22 •• Two slits spaced 0.0720 mm apart are 0.800 m from a screen. Coherent light of wavelength λ passes through the two slits. In their interference pattern on the screen, the distance from the center of the central maximum to the first minimum is 3.00 mm. If the intensity at the peak of the central maximum is 0.0600 W/m^2, what is the intensity at points on the screen that are (a) 2.00 mm and (b) 1.50 mm from the center of the central maximum?

Section 35.4 Interference in Thin Films

35.23 • What is the thinnest film of a coating with $n = 1.42$ on glass ($n = 1.52$) for which destructive interference of the red component (650 nm) of an incident white light beam in air can take place by reflection?

35.24 •• **Nonglare Glass.** When viewing a piece of art that is behind glass, one often is affected by the light that is reflected off the front of the glass (called *glare*), which can make it difficult to see the art clearly. One solution is to coat the outer surface of the glass with a film to cancel part of the glare. (a) If the glass has a refractive index of 1.62 and you use TiO_2, which has an index of refraction of 2.62, as the coating, what is the minimum film thickness that will cancel light of wavelength 505 nm? (b) If this coating is too thin to stand up to wear, what other thickness would also work? Find only the three thinnest ones.

35.25 •• Two rectangular pieces of plane glass are laid one upon the other on a table. A thin strip of paper is placed between them at one edge so that a very thin wedge of air is formed. The plates are illuminated at normal incidence by 546-nm light from a mercury-vapor lamp. Interference fringes are formed, with 15.0 fringes per centimeter. Find the angle of the wedge.

35.26 •• A plate of glass 9.00 cm long is placed in contact with a second plate and is held at a small angle with it by a metal strip 0.0800 mm thick placed under one end. The space between the plates is filled with air. The glass is illuminated from above with light having a wavelength in air of 656 nm. How many interference fringes are observed per centimeter in the reflected light?

35.27 •• A uniform film of TiO_2, 1036 nm thick and having index of refraction 2.62, is spread uniformly over the surface of crown glass of refractive index 1.52. Light of wavelength 520.0 nm falls at normal incidence onto the film from air. You want to increase the thickness of this film so that the reflected light cancels. (a) What is the *minimum* thickness of TiO_2 that you must *add* so the reflected light cancels as desired? (b) After you make the adjustment in part (a), what is the path difference between the light reflected off the top of the film and the light that cancels it after traveling through the film? Express your answer in (i) nanometers and (ii) wavelengths of the light in the TiO_2 film.

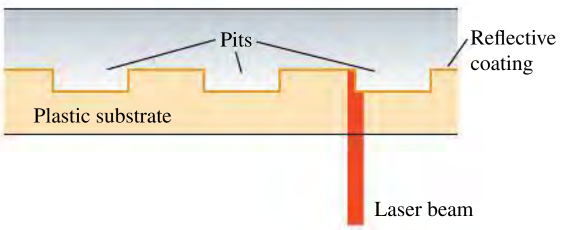
35.28 • A plastic film with index of refraction 1.70 is applied to the surface of a car window to increase the reflectivity and thus to keep the car's interior cooler. The window glass has index of refraction 1.52. (a) What minimum thickness is required if light of wavelength 550 nm in air reflected from the two sides of the film is to interfere constructively? (b) Coatings as thin as that calculated in part (a) are difficult to manufacture and install. What is the next greater thickness for which constructive interference will also occur?

35.29 • The walls of a soap bubble have about the same index of refraction as that of plain water, $n = 1.33$. There is air both inside and outside the bubble. (a) What wavelength (in air) of visible light is most strongly reflected from a point on a soap bubble where its wall is 290 nm thick? To what color does this correspond (see Fig. 32.4 and Table 32.1)? (b) Repeat part (a) for a wall thickness of 340 nm.

35.30 •• A researcher measures the thickness of a layer of benzene $(n = 1.50)$ floating on water by shining monochromatic light onto the film and varying the wavelength of the light. She finds that light of wavelength 575 nm is reflected most strongly from the film. What does she calculate for the minimum thickness of the film?

35.31 •• **Compact Disc Player.** A compact disc (CD) is read from the bottom by a semiconductor laser with wavelength 790 nm passing through a plastic substrate of refractive index 1.8. When the beam encounters a pit, part of the beam is reflected from the pit and part from the flat region between the pits, so these two beams interfere with each other (**Fig. E35.31**). What must the minimum pit depth be so that the part of the beam reflected from a pit cancels the part of the beam reflected from the flat region? (It is this cancellation that allows the player to recognize the beginning and end of a pit.)

Figure **E35.31**



35.32 • What is the thinnest soap film (excluding the case of zero thickness) that appears black when illuminated with light with wavelength 480 nm? The index of refraction of the film is 1.33, and there is air on both sides of the film.

Section 35.5 The Michelson Interferometer

35.33 • How far must the mirror M_2 (see Fig. 35.19) of the Michelson interferometer be moved so that 1800 fringes of He-Ne laser light $(\lambda = 633 \text{ nm})$ move across a line in the field of view?

35.34 • Jan first uses a Michelson interferometer with the 606-nm light from a krypton-86 lamp. He displaces the movable mirror away from him, counting 818 fringes moving across a line in his field of view. Then Linda replaces the krypton lamp with filtered 502-nm light from a helium lamp and displaces the movable mirror toward her. She also counts 818 fringes, but they move across the line in her field of view opposite to the direction they moved for Jan. Assume that both Jan and Linda counted to 818 correctly. (a) What distance did each person move the mirror? (b) What is the resultant displacement of the mirror?

PROBLEMS

35.35 •• One round face of a 3.25-m, solid, cylindrical plastic pipe is covered with a thin black coating that completely blocks light. The opposite face is covered with a fluorescent coating that glows when it is struck by light. Two straight, thin, parallel scratches, 0.225 mm apart, are made in the center of the black face. When laser light of wavelength 632.8 nm shines through the slits perpendicular to the black face, you find that the central bright fringe on the opposite face is 5.82 mm wide, measured between the dark fringes that border it on either side. What is the index of refraction of the plastic?

35.36 ••• Newton's rings are visible when a planoconvex lens is placed on a flat glass surface. For a particular lens with an index of refraction of $n = 1.50$ and a glass plate with an index of $n = 1.80$, the diameter of the third bright ring is 0.640 mm. If water $(n = 1.33)$ now fills the space between the lens and the glass plate, what is the new diameter of this ring? Assume the radius of curvature of the lens is much greater than the wavelength of the light.

35.37 • BIO **Coating Eyeglass Lenses.** Eyeglass lenses can be coated on the *inner* surfaces to reduce the reflection of stray light to the eye. If the lenses are medium flint glass of refractive index 1.62 and the coating is fluorite of refractive index 1.432, (a) what minimum thickness of film is needed on the lenses to cancel light of wavelength 550 nm reflected toward the eye at normal incidence? (b) Will any other wavelengths of visible light be cancelled or enhanced in the reflected light?

35.38 •• BIO **Sensitive Eyes.** After an eye examination, you put some eyedrops on your sensitive eyes. The cornea (the front part of the eye) has an index of refraction of 1.38, while the eyedrops have a refractive index of 1.45. After you put in the drops, your friends notice that your eyes look red, because red light of wavelength 600 nm has been reinforced in the reflected light. (a) What is the minimum thickness of the film of eyedrops on your cornea? (b) Will any other wavelengths of visible light be reinforced in the reflected light? Will any be cancelled? (c) Suppose you had contact lenses, so that the eyedrops went on them instead of on your corneas. If the refractive index of the lens material is 1.50 and the layer of eyedrops has the same thickness as in part (a), what wavelengths of visible light will be reinforced? What wavelengths will be cancelled?

35.39 •• Two flat plates of glass with parallel faces are on a table, one plate on the other. Each plate is 11.0 cm long and has a refractive index of 1.55. A very thin sheet of metal foil is inserted under the end of the upper plate to raise it slightly at that end, in a manner similar to that discussed in Example 35.4. When you view the glass plates from above with reflected white light, you observe that, at 1.15 mm from the line where the sheets are in contact, the violet light of wavelength 400.0 nm is enhanced in this reflected light, but no visible light is enhanced closer to the line of contact. (a) How far from the line of contact will green light (of wavelength 550.0 nm) and orange light (of wavelength 600.0 nm) first be enhanced? (b) How far from the line of contact will the violet, green, and orange light again be enhanced in the reflected light? (c) How thick is the metal foil holding the ends of the plates apart?

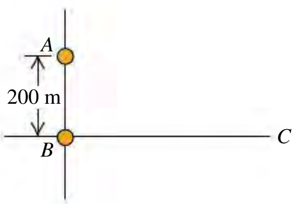
35.40 •• In a setup similar to that of Problem 35.39, the glass has an index of refraction of 1.53, the plates are each 8.00 cm long, and the metal foil is 0.015 mm thick. The space between the plates is filled with a jelly whose refractive index is not known precisely, but is known to be greater than that of the glass. When you illuminate these plates from above with light of wavelength 525 nm, you observe a series of equally spaced dark fringes in the reflected light. You measure the spacing of these fringes and find that there are 10 of them every 6.33 mm. What is the index of refraction of the jelly?

35.41 ••• Suppose you illuminate two thin slits by monochromatic coherent light in air and find that they produce their first interference *minima* at $\pm 35.20°$ on either side of the central bright spot. You then immerse these slits in a transparent liquid and illuminate them with the same light. Now you find that the first minima occur at $\pm 19.46°$ instead. What is the index of refraction of this liquid?

35.42 •• **CP CALC** A very thin sheet of brass contains two thin parallel slits. When a laser beam shines on these slits at normal incidence and room temperature (20.0°C), the first interference dark fringes occur at $\pm 26.6°$ from the original direction of the laser beam when viewed from some distance. If this sheet is now slowly heated to 135°C, by how many degrees do these dark fringes change position? Do they move closer together or farther apart? See Table 17.1 for pertinent information, and ignore any effects that might occur due to a change in the thickness of the slits. (*Hint:* Thermal expansion normally produces very small changes in length, so you can use differentials to find the change in the angle.)

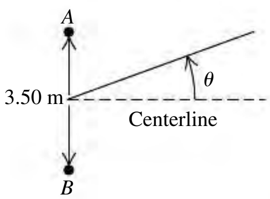
35.43 •• Two radio antennas radiating in phase are located at points A and B, 200 m apart (**Fig. P35.43**). The radio waves have a frequency of 5.80 MHz. A radio receiver is moved out from point B along a line perpendicular to the line connecting A and B (line BC shown in Fig. P35.43). At what distances from B will there be *destructive* interference? (*Note:* The distance of the receiver from the sources is not large in comparison to the separation of the sources, so Eq. (35.5) does not apply.)

Figure **P35.43**



35.44 •• Two speakers A and B are 3.50 m apart, and each one is emitting a frequency of 444 Hz. However, because of signal delays in the cables, speaker A is one-fourth of a period *ahead* of speaker B. For points far from the speakers, find all the angles relative to the centerline (**Fig. P35.44**) at which the sound from these speakers cancels. Include angles on *both* sides of the centerline. The speed of sound is 340 m/s.

Figure **P35.44**



35.45 •• **CP** A thin uniform film of refractive index 1.750 is placed on a sheet of glass of refractive index 1.50. At room temperature (20.0°C), this film is just thick enough for light with wavelength 582.4 nm reflected off the top of the film to be cancelled by light reflected from the top of the glass. After the glass is placed in an oven and slowly heated to 170°C, you find that the film cancels reflected light with wavelength 588.5 nm. What is the coefficient of linear expansion of the film? (Ignore any changes in the refractive index of the film due to the temperature change.)

35.46 ••• **GPS Transmission.** The GPS (Global Positioning System) satellites are approximately 5.18 m across and transmit two low-power signals, one of which is at 1575.42 MHz (in the UHF band). In a series of laboratory tests on the satellite, you put two 1575.42-MHz UHF transmitters at opposite ends of the satellite. These broadcast in phase uniformly in all directions. You measure the intensity at points on a circle that is several hundred meters in radius and centered on the satellite. You measure angles on this circle relative to a point that lies along the centerline of the satellite (that is, the perpendicular bisector of a line that extends from one transmitter to the other). At this point on the circle, the measured intensity is 2.00 W/m^2. (a) At how many other angles in the range $0° < \theta < 90°$ is the intensity also 2.00 W/m^2? (b) Find the four smallest angles in the range $0° < \theta < 90°$ for which the intensity is 2.00 W/m^2. (c) What is the intensity at a point on the circle at an angle of 4.65° from the centerline?

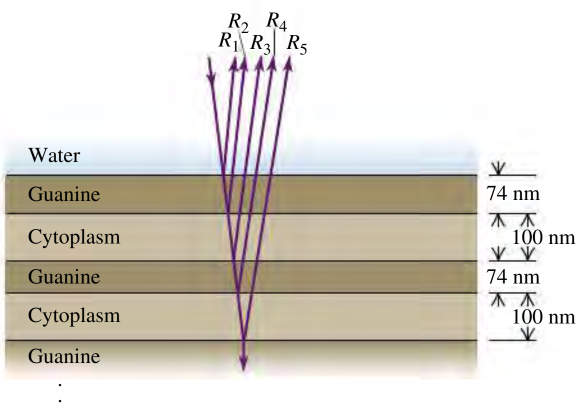
35.47 •• White light reflects at normal incidence from the top and bottom surfaces of a glass plate ($n = 1.52$). There is air above and below the plate. Constructive interference is observed for light whose wavelength in air is 477.0 nm. What is the thickness of the plate if the next longer wavelength for which there is constructive interference is 540.6 nm?

35.48 •• Laser light of wavelength 510 nm is traveling in air and shines at normal incidence onto the flat end of a transparent plastic rod that has $n = 1.30$. The end of the rod has a thin coating of a transparent material that has refractive index 1.65. What is the minimum (nonzero) thickness of the coating (a) for which there is maximum transmission of the light into the rod; (b) for which transmission into the rod is minimized?

35.49 •• Red light with wavelength 700 nm is passed through a two-slit apparatus. At the same time, monochromatic visible light with another wavelength passes through the same apparatus. As a result, most of the pattern that appears on the screen is a mixture of two colors; however, the center of the third bright fringe ($m = 3$) of the red light appears pure red, with none of the other color. What are the possible wavelengths of the second type of visible light? Do you need to know the slit spacing to answer this question? Why or why not?

35.50 •• **BIO** **Reflective Coatings and Herring.** Herring and related fish have a brilliant silvery appearance that camouflages them while they are swimming in a sunlit ocean. The silveriness is due to *platelets* attached to the surfaces of these fish. Each platelet is made up of several alternating layers of crystalline guanine ($n = 1.80$) and of cytoplasm ($n = 1.333$, the same as water), with a guanine layer on the outside in contact with the surrounding water (**Fig. P35.50**). In one typical platelet, the guanine

Figure **P35.50**

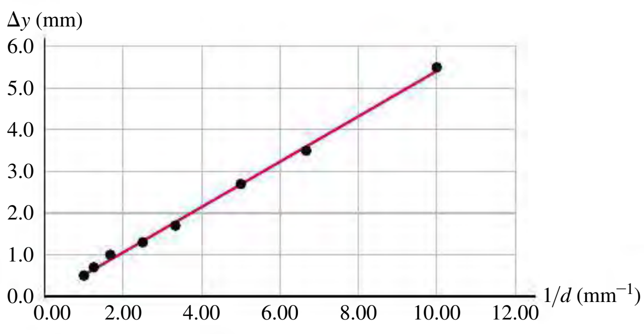


layers are 74 nm thick and the cytoplasm layers are 100 nm thick. (a) For light striking the platelet surface at normal incidence, for which vacuum wavelengths of visible light will all of the reflections R_1, R_2, R_3, R_4, and R_5, shown in Fig. P35.50, be approximately in phase? If white light is shone on this platelet, what color will be most strongly reflected (see Fig. 32.4)? The surface of a herring has very many platelets side by side with layers of different thickness, so that *all* visible wavelengths are reflected. (b) Explain why such a "stack" of layers is more reflective than a single layer of guanine with cytoplasm underneath. (A stack of five guanine layers separated by cytoplasm layers reflects more than 80% of incident light at the wavelength for which it is "tuned.") (c) The color that is most strongly reflected from a platelet depends on the angle at which it is viewed. Explain why this should be so. (You can see these changes in color by examining a herring from different angles. Most of the platelets on these fish are oriented in the same way, so that they are vertical when the fish is swimming.)

35.51 •• After a laser beam passes through two thin parallel slits, the first completely dark fringes occur at $\pm 19.0°$ with the original direction of the beam, as viewed on a screen far from the slits. (a) What is the ratio of the distance between the slits to the wavelength of the light illuminating the slits? (b) What is the smallest angle, relative to the original direction of the laser beam, at which the intensity of the light is $\frac{1}{10}$ the maximum intensity on the screen?

35.52 •• **DATA** In your summer job at an optics company, you are asked to measure the wavelength λ of the light that is produced by a laser. To do so, you pass the laser light through two narrow slits that are separated by a distance d. You observe the interference pattern on a screen that is 0.900 m from the slits and measure the separation Δy between adjacent bright fringes in the portion of the pattern that is near the center of the screen. Using a microscope, you measure d. But both Δy and d are small and difficult to measure accurately, so you repeat the measurements for several pairs of slits, each with a different value of d. Your results are shown in **Fig. P35.52**, where you have plotted Δy versus $1/d$. The line in the graph is the best-fit straight line for the data. (a) Explain why the data points plotted this way fall close to a straight line. (b) Use Fig. P35.52 to calculate λ.

Figure **P35.52**



35.53 •• **DATA** Short-wave radio antennas A and B are connected to the same transmitter and emit coherent waves in phase and with the same frequency f. You must determine the value of f and the placement of the antennas that produce a maximum intensity through constructive interference at a receiving antenna that is located at point P, which is at the corner of your garage. First you place antenna A at a point 240.0 m due east of P. Next you place antenna B on the line that connects A and P, a distance x due east of P, where $x < 240.0$ m. Then you measure that a maximum in the total intensity from the two antennas occurs when $x = 210.0$ m, 216.0 m, and 222.0 m. You don't investigate smaller or larger values of x. (Treat the antennas as point sources.) (a) What is the frequency f of the waves that are emitted by the antennas? (b) What is the greatest value of x, with $x < 240.0$ m, for which the interference at P is destructive?

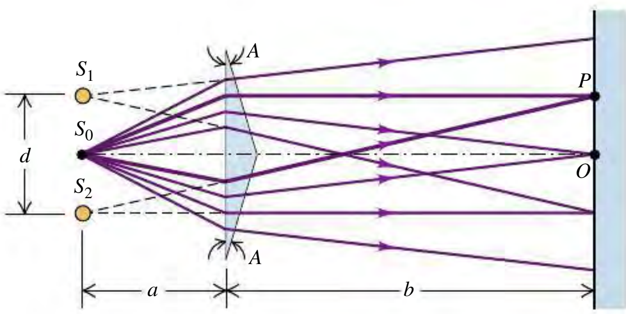
35.54 •• **DATA** In your research lab, a very thin, flat piece of glass with refractive index 1.40 and uniform thickness covers the opening of a chamber that holds a gas sample. The refractive indexes of the gases on either side of the glass are very close to unity. To determine the thickness of the glass, you shine coherent light of wavelength λ_0 in vacuum at normal incidence onto the surface of the glass. When $\lambda_0 = 496$ nm, constructive interference occurs for light that is reflected at the two surfaces of the glass. You find that the next shorter wavelength in vacuum for which there is constructive interference is 386 nm. (a) Use these measurements to calculate the thickness of the glass. (b) What is the longest wavelength in vacuum for which there is constructive interference for the reflected light?

CHALLENGE PROBLEMS

35.55 ••• **CP** The index of refraction of a glass rod is 1.48 at $T = 20.0°C$ and varies linearly with temperature, with a coefficient of $2.50 \times 10^{-5}/C°$. The coefficient of linear expansion of the glass is $5.00 \times 10^{-6}/C°$. At 20.0°C the length of the rod is 3.00 cm. A Michelson interferometer has this glass rod in one arm, and the rod is being heated so that its temperature increases at a rate of 5.00 C°/min. The light source has wavelength $\lambda = 589$ nm, and the rod initially is at $T = 20.0°C$. How many fringes cross the field of view each minute?

35.56 ••• **CP** **Figure P35.56** shows an interferometer known as *Fresnel's biprism*. The magnitude of the prism angle A is extremely small. (a) If S_0 is a very narrow source slit, show that the separation of the two virtual coherent sources S_1 and S_2 is given by $d = 2aA(n - 1)$, where n is the index of refraction of the material of the prism. (b) Calculate the spacing of the fringes of green light with wavelength 500 nm on a screen 2.00 m from the biprism. Take $a = 0.200$ m, $A = 3.50$ mrad, and $n = 1.50$.

Figure **P35.56**



PASSAGE PROBLEMS

INTERFERENCE AND SOUND WAVES. Interference occurs with not only light waves but also all frequencies of electromagnetic waves and all other types of waves, such as sound and water waves. Suppose that your physics professor sets up two sound speakers in the front of your classroom and uses an electronic oscillator to produce sound waves of a single frequency. When she turns the oscillator on (take this to be its original setting), you and many students hear a loud tone while other students hear nothing. (The speed of sound in air is 340 m/s.)

35.57 The professor then adjusts the apparatus. The frequency that you hear does not change, but the loudness decreases. Now all of your fellow students can hear the tone. What did the professor do? (a) She turned off the oscillator. (b) She turned down the volume of the speakers. (c) She changed the phase relationship of the speakers. (d) She disconnected one speaker.

35.58 The professor returns the apparatus to the original setting. She then adjusts the speakers again. All of the students who had heard nothing originally now hear a loud tone, while you and the others who had originally heard the loud tone hear nothing. What did the professor do? (a) She turned off the oscillator. (b) She turned down the volume of the speakers. (c) She changed the phase relationship of the speakers. (d) She disconnected one speaker.

35.59 The professor again returns the apparatus to its original setting, so you again hear the original loud tone. She then slowly moves one speaker away from you until it reaches a point at which you can no longer hear the tone. If she has moved the speaker by 0.34 m (farther from you), what is the frequency of the tone? (a) 1000 Hz; (b) 2000 Hz; (c) 500 Hz; (d) 250 Hz.

35.60 The professor once again returns the apparatus to its original setting, but now she adjusts the oscillator to produce sound waves of half the original frequency. What happens? (a) The students who originally heard a loud tone again hear a loud tone, and the students who originally heard nothing still hear nothing. (b) The students who originally heard a loud tone now hear nothing, and the students who originally heard nothing now hear a loud tone. (c) Some of the students who originally heard a loud tone again hear a loud tone, but others in that group now hear nothing. (d) Among the students who originally heard nothing, some still hear nothing but others now hear a loud tone.

Answers

Chapter Opening Question ?

(v) The colors appear as a result of constructive interference between light waves reflected from the upper and lower surfaces of the oil film. The wavelength of light for which the most constructive interference occurs at a point, and hence the color that appears the brightest at that point, depends on (1) the thickness of the film (which determines the path difference between light waves that reflect off the two surfaces), (2) the oil's index of refraction (which gives the wavelength of light in the oil a different value than in air), and (3) the index of refraction of the material below the oil (which determines whether the wave that reflects from the lower surface undergoes a half-cycle phase shift). (See Examples 35.4, 35.5, and 35.6 in Section 35.4.)

Test Your Understanding Questions

35.1 **(i)** At any point P on the positive y-axis above S_1, the distance r_2 from S_2 to P is greater than the distance r_1 from S_1 to P by 4λ. This corresponds to $m = 4$ in Eq. (35.1), the equation for constructive interference. Hence all such points make up an antinodal curve.

35.2 **(ii)** Blue light has a shorter wavelength than red light (see Section 32.1). Equation (35.6) tells us that the distance y_m from the center of the pattern to the mth bright fringe is proportional to the wavelength λ. Hence all of the fringes will move toward the center of the pattern as the wavelength decreases, and the spacing between fringes will decrease.

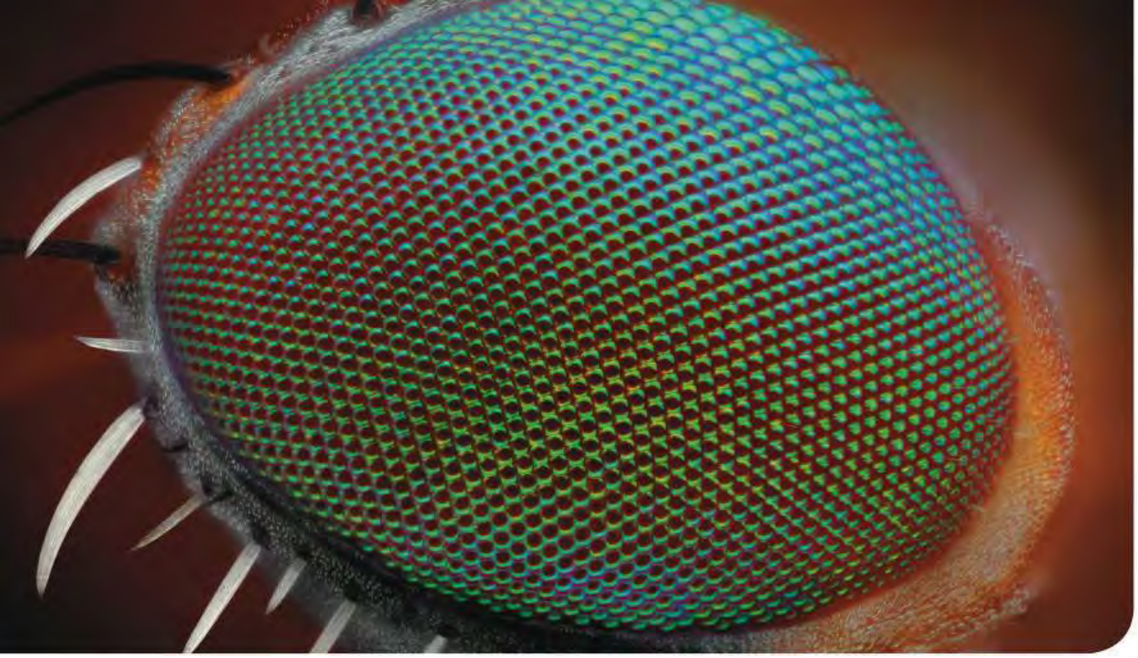
35.3 **(i), (iv), (ii), (iii)** In cases (i) and (iii) we are given the wavelength λ and path difference $d\sin\theta$. Hence we use Eq. (35.14), $I = I_0\cos^2[(\pi d\sin\theta)/\lambda]$. In parts (ii) and (iii) we are given the phase difference ϕ and we use Eq. (35.10), $I = I_0\cos^2(\phi/2)$. We find:
(i) $I = I_0\cos^2[\pi(4.00 \times 10^{-7}\text{ m})/(5.00 \times 10^{-7}\text{ m})] = I_0\cos^2(0.800\pi\text{ rad}) = 0.655I_0$;
(ii) $I = I_0\cos^2[(4.00\text{ rad})/2] = I_0\cos^2(2.00\text{ rad}) = 0.173I_0$;
(iii) $I = I_0\cos^2[\pi(7.50 \times 10^{-7}\text{ m})/(5.00 \times 10^{-7}\text{ m})] = I_0\cos^2(1.50\pi\text{ rad}) = 0$;
(iv) $I = I_0\cos^2[(2.00\text{ rad})/2] = I_0\cos^2(1.00\text{ rad}) = 0.292I_0$.

35.4 **(i) and (iii)** Benzene has a larger index of refraction than air, so light that reflects off the upper surface of the benzene undergoes a half-cycle phase shift. Fluorite has a *smaller* index of refraction than benzene, so light that reflects off the benzene–fluorite interface does not undergo a phase shift. Hence the equation for constructive reflection is Eq. (35.18a), $2t = \left(m + \frac{1}{2}\right)\lambda$, which we can rewrite as $t = \left(m + \frac{1}{2}\right)\lambda/2 = \left(m + \frac{1}{2}\right)(400\text{ mm})/2 = 100\text{ nm}, 300\text{ nm}, 500\text{ nm}, \ldots$.

35.5 **yes** Changing the index of refraction changes the wavelength of the light inside the compensator plate, and so changes the number of wavelengths within the thickness of the plate. Hence this has the same effect as changing the distance L_1 from the beam splitter to mirror M_1, which would change the interference pattern.

Bridging Problem

(a) 441 nm
(b) 551 nm



? Flies have *compound* eyes with thousands of miniature lenses. The overall diameter of the eye is about 1 mm, but each lens is only about 20 μm in diameter and produces an individual image of a small region in the fly's field of view. Compared to the resolving power of the human eye (in which the light-gathering region is about 16 mm across), the ability of a fly's eye to resolve small details is (i) worse because the lenses are so small; (ii) worse because the eye as a whole is so small; (iii) better because the lenses are so small; (iv) better because the eye as a whole is so small; (v) about the same.

36 DIFFRACTION

LEARNING GOALS

Looking forward at …

36.1 What happens when coherent light shines on an object with an edge or aperture.

36.2 How to understand the diffraction pattern formed when coherent light passes through a narrow slit.

36.3 How to calculate the intensity at various points in a single-slit diffraction pattern.

36.4 What happens when coherent light shines on an array of narrow, closely spaced slits.

36.5 How scientists use diffraction gratings for precise measurements of wavelength.

36.6 How x-ray diffraction reveals the arrangement of atoms in a crystal.

36.7 How diffraction sets limits on the smallest details that can be seen with an optical system.

36.8 How holograms work.

Looking back at …

33.4, 33.7 Prisms and dispersion; Huygens's principle.

34.4, 34.5 Image formation by a lens; *f*-number.

35.1–35.3 Coherent light, two-slit interference, and phasors.

Everyone is used to the idea that sound bends around corners. If sound didn't behave this way, you couldn't hear a police siren that's out of sight or the speech of a person whose back is turned to you. But *light* can bend around corners as well. When light from a point source falls on a straightedge and casts a shadow, the edge of the shadow is never perfectly sharp. Some light appears in the area that we expect to be in the shadow, and we find alternating bright and dark fringes in the illuminated area. In general, light emerging from apertures doesn't precisely follow the predictions of the straight-line ray model of geometric optics.

The reason for these effects is that light, like sound, has wave characteristics. In Chapter 35 we studied the interference patterns that can arise when two light waves are combined. In this chapter we'll investigate interference effects due to combining *many* light waves. Such effects are referred to as *diffraction*. The behavior of waves after they pass through an aperture is an example of diffraction; each infinitesimal part of the aperture acts as a source of waves, and these waves interfere, producing a pattern of bright and dark fringes.

Similar patterns appear when light emerges from *arrays* of apertures. The nature of these patterns depends on the color of the light and the size and spacing of the apertures. Examples of this effect include the colors of iridescent butterflies and the "rainbow" you see reflected from the surface of a compact disc. We'll explore similar effects with x rays that are used to study the atomic structure of solids and liquids. Finally, we'll look at the physics of a *hologram,* a special kind of interference pattern used to form three-dimensional images.

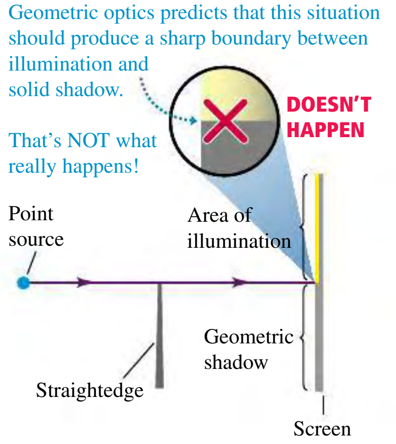
36.1 FRESNEL AND FRAUNHOFER DIFFRACTION

According to geometric optics, when an opaque object is placed between a point light source and a screen, as in **Fig. 36.1**, the shadow of the object forms a perfectly sharp line. No light at all strikes the screen at points within the shadow, and the area outside the shadow is illuminated nearly uniformly. But as we saw in Chapter 35, the *wave* nature of light causes effects that can't be understood with geometric optics. An important class of such effects occurs when light strikes a barrier that has an aperture or an edge. The interference patterns formed in such a situation are grouped under the heading **diffraction.**

Figure 36.2 shows an example of diffraction. The photograph in Fig. 36.2a was made by placing a razor blade halfway between a pinhole, illuminated by monochromatic light, and a photographic film. The film recorded the shadow cast by the blade. Figure 36.2b is an enlargement of a region near the shadow of the right edge of the blade. The position of the *geometric* shadow line is indicated by arrows. The area outside the geometric shadow is bordered by alternating bright and dark bands. There is some light in the shadow region, although this is not very visible in the photograph. The first bright band in Fig. 36.2b, just to the right of the geometric shadow, is considerably brighter than in the region of uniform illumination to the extreme right. This simple experiment gives us some idea of the richness and complexity of diffraction.

We don't often observe diffraction patterns such as Fig. 36.2 in everyday life because most ordinary light sources are neither monochromatic nor point sources. If we use a white frosted light bulb instead of a point source to illuminate the razor blade in Fig. 36.2, each wavelength of the light from every point of the bulb forms its own diffraction pattern, but the patterns overlap so much that we can't see any individual pattern.

36.1 A point source of light illuminates a straightedge.

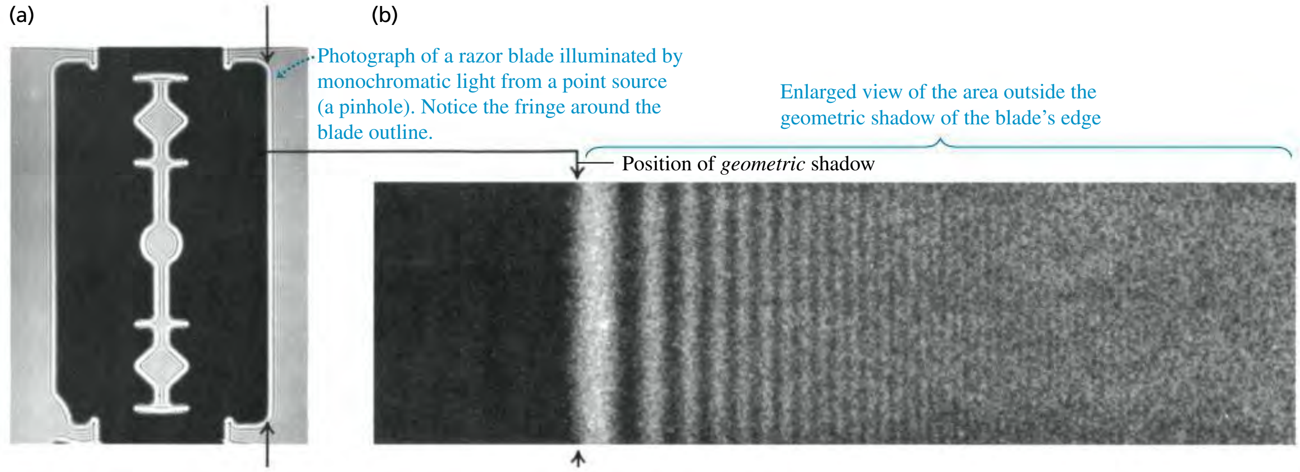


Diffraction and Huygens's Principle

We can use Huygens's principle (see Section 33.7) to analyze diffraction patterns. This principle states that we can consider every point of a wave front as a source of secondary wavelets. These spread out in all directions with a speed equal to the speed of propagation of the wave. The position of the wave front at any later time is the *envelope* of the secondary wavelets at that time. To find the resultant displacement at any point, we use the superposition principle to combine all the individual displacements produced by these secondary waves.

In Fig. 36.1, both the point source and the screen are relatively close to the obstacle forming the diffraction pattern. This situation is described as *near-field diffraction* or **Fresnel diffraction,** pronounced "Freh-nell" (after the French scientist Augustin Jean Fresnel, 1788–1827). By contrast, we use the term **Fraunhofer diffraction** (after the German physicist Joseph von Fraunhofer, 1787–1826) for situations in which the source, obstacle, and screen are far enough apart that we can consider all lines from the source to the obstacle to be parallel, and can likewise consider all lines from the obstacle to a given point on the screen to be parallel. We will restrict the following discussion

36.2 An example of diffraction.



to Fraunhofer diffraction, which is usually simpler to analyze in detail than Fresnel diffraction.

Diffraction is sometimes described as "the bending of light around an obstacle." But the process that causes diffraction is present in the propagation of *every* wave. When part of the wave is cut off by some obstacle, we observe diffraction effects that result from interference of the remaining parts of the wave fronts. Optical instruments typically use only a limited portion of a wave; for example, a telescope uses only the part of a wave that is admitted by its objective lens or mirror. Thus diffraction plays a role in nearly all optical phenomena.

Finally, we emphasize that there is no fundamental distinction between *interference* and *diffraction*. In Chapter 35 we used the term *interference* for effects involving waves from a small number of sources, usually two. *Diffraction* usually involves a *continuous* distribution of Huygens's wavelets across the area of an aperture, or a very large number of sources or apertures. But both interference and diffraction are consequences of superposition and Huygens's principle.

TEST YOUR UNDERSTANDING OF SECTION 36.1 Can *sound* waves undergo diffraction around an edge? ❙

MP

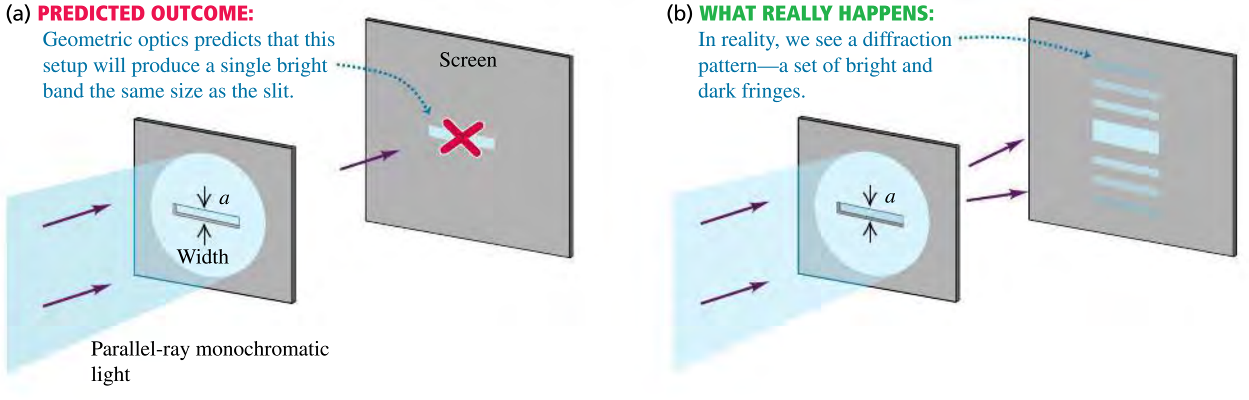
PhET: Wave Interference

36.2 DIFFRACTION FROM A SINGLE SLIT

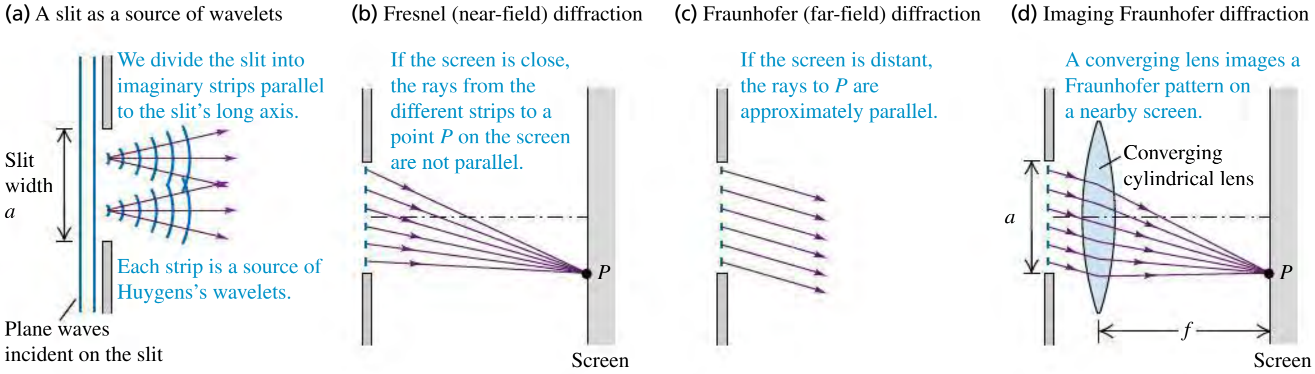
In this section we'll discuss the diffraction pattern formed by plane-wave (parallel-ray) monochromatic light when it emerges from a long, narrow slit, as shown in **Fig. 36.3**. We call the narrow dimension the *width,* even though in this figure it is a vertical dimension.

According to geometric optics, the transmitted beam should have the same cross section as the slit, as in Fig. 36.3a. What is *actually* observed is the pattern shown in Fig. 36.3b. The beam spreads out vertically after passing through the slit. The diffraction pattern consists of a central bright band, which may be much broader than the width of the slit, bordered by alternating dark and bright bands with rapidly decreasing intensity. About 85% of the power in the transmitted beam is in the central bright band, whose width is *inversely* proportional to the slit width. In general, the narrower the slit, the broader the entire diffraction pattern. (The *horizontal* spreading of the beam in Fig. 36.3b is negligible because the horizontal dimension of the slit is relatively large.) You can observe a similar diffraction pattern by looking at a point source, such as a distant street light, through a narrow slit formed between your two thumbs held in front of your eye; the retina of your eye acts as the screen.

36.3 (a) The "shadow" of a horizontal slit as incorrectly predicted by geometric optics. (b) A horizontal slit actually produces a diffraction pattern. The slit width has been greatly exaggerated.



36.4 Diffraction by a single rectangular slit. The long sides of the slit are perpendicular to the figure.

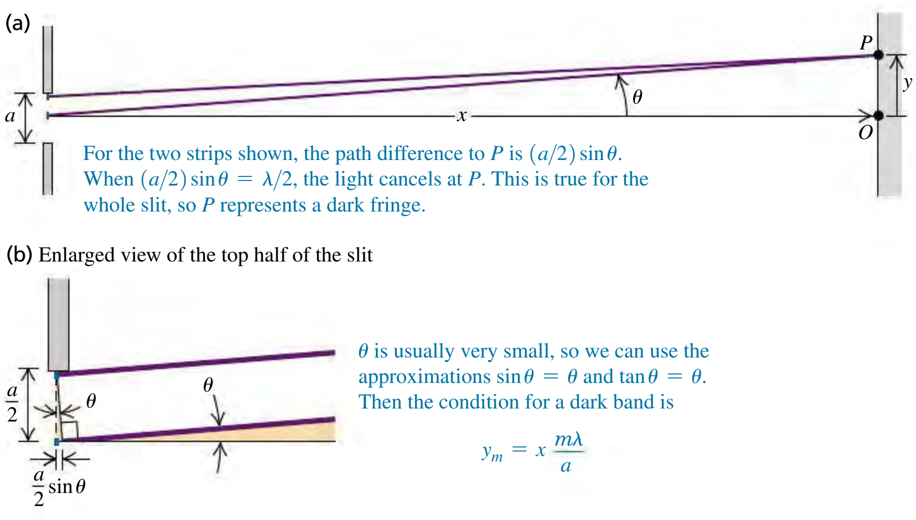


Single-Slit Diffraction: Locating the Dark Fringes

Figure 36.4 shows a side view of the same setup; the long sides of the slit are perpendicular to the figure, and plane waves are incident on the slit from the left. According to Huygens's principle, each element of area of the slit opening can be considered as a source of secondary waves. In particular, imagine dividing the slit into several narrow strips of equal width, parallel to the long edges and perpendicular to the page. Figure 36.4a shows two such strips. Cylindrical secondary wavelets, shown in cross section, spread out from each strip.

In Fig. 36.4b a screen is placed to the right of the slit. We can calculate the resultant intensity at a point P on the screen by adding the contributions from the individual wavelets, taking proper account of their various phases and amplitudes. It's easiest to do this calculation if we assume that the screen is far enough away that all the rays from various parts of the slit to a particular point P on the screen are parallel, as in Fig. 36.4c. An equivalent situation is Fig. 36.4d, in which the rays to the lens are parallel and the lens forms a reduced image of the same pattern that would be formed on an infinitely distant screen without the lens. We might expect that the various light paths through the lens would introduce additional phase shifts, but in fact it can be shown that all the paths have *equal* phase shifts, so this is not a problem.

The situation of Fig. 36.4b is Fresnel diffraction; those in Figs. 36.4c and 36.4d, where the outgoing rays are considered parallel, are Fraunhofer diffraction. We can derive quite simply the most important characteristics of the Fraunhofer diffraction pattern from a single slit. First consider two narrow strips, one just below the top edge of the drawing of the slit and one at its center, shown in end view in **Fig. 36.5**. The difference in path length to point P is $(a/2)\sin\theta$, where a is the slit width and θ is the angle between the perpendicular to the slit



36.5 Side view of a horizontal slit. When the distance x to the screen is much greater than the slit width a, the rays from a distance $a/2$ apart may be considered parallel.

and a line from the center of the slit to P. Suppose this path difference happens to be equal to $\lambda/2$; then light from these two strips arrives at point P with a half-cycle phase difference, and cancellation occurs.

Similarly, light from two strips immediately *below* the two in the figure also arrives at P a half-cycle out of phase. In fact, the light from *every* strip in the top half of the slit cancels out the light from a corresponding strip in the bottom half. Hence the combined light from the entire slit completely cancels at P, giving a dark fringe in the interference pattern. A dark fringe occurs whenever

$$\frac{a}{2}\sin\theta = \pm\frac{\lambda}{2} \qquad \text{or} \qquad \sin\theta = \pm\frac{\lambda}{a} \tag{36.1}$$

The plus-or-minus $(\pm)$ sign in Eq. (36.1) says that there are symmetric dark fringes above and below point O in Fig. 36.5a. The upper fringe $(\theta > 0)$ occurs at a point P where light from the bottom half of the slit travels $\lambda/2$ farther to P than does light from the top half; the lower fringe $(\theta < 0)$ occurs where light from the *top* half travels $\lambda/2$ farther than light from the *bottom* half.

We may also divide the slit into quarters, sixths, and so on, and use the above argument to show that a dark fringe occurs whenever $\sin\theta = \pm 2\lambda/a$, $\pm 3\lambda/a$, and so on. Thus the condition for a *dark* fringe is

Dark fringes, single-slit diffraction:

Angle of line from center of slit to mth dark fringe on screen

$$\sin\theta = \frac{m\lambda}{a} \quad (m = \pm 1, \pm 2, \pm 3, \dots) \tag{36.2}$$

Slit width Wavelength

For example, if the slit width is equal to ten wavelengths $(a = 10\lambda)$, dark fringes occur at $\sin\theta = \pm\frac{1}{10}, \pm\frac{2}{10}, \pm\frac{3}{10}, \dots$. Between the dark fringes are bright fringes. Note that $\sin\theta = 0$ corresponds to a *bright* band; in this case, light from the entire slit arrives at P in phase. Thus it would be wrong to put $m = 0$ in Eq. (36.2).

With light, the wavelength λ is of the order of 500 nm $= 5 \times 10^{-7}$ m. This is often much smaller than the slit width a; a typical slit width is 10^{-2} cm $= 10^{-4}$ m. Therefore the values of θ in Eq. (36.2) are often so small that the approximation $\sin\theta \approx \theta$ (where θ is in radians) is a very good one. In that case we can rewrite this equation as

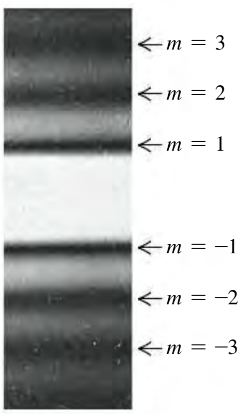
$$\theta = \frac{m\lambda}{a} \qquad (m = \pm 1, \pm 2, \pm 3, \dots) \qquad \text{(for small angles } \theta \text{ in radians)}$$

Also, if the distance from slit to screen is x, as in Fig. 36.5a, and the vertical distance of the mth dark band from the center of the pattern is y_m, then $\tan\theta = y_m/x$. For small θ we may also approximate $\tan\theta$ by θ (in radians). We then find

$$y_m = x\frac{m\lambda}{a} \qquad (\text{for } y_m \ll x) \tag{36.3}$$

Figure 36.6 is a photograph of a single-slit diffraction pattern with the $m = \pm 1$, ± 2, and ± 3 minima labeled. The central bright fringe is wider than the other bright fringes; in the small-angle approximation used in Eq. (36.3), it is exactly twice as wide.

36.6 Photograph of the Fraunhofer diffraction pattern of a single horizontal slit.

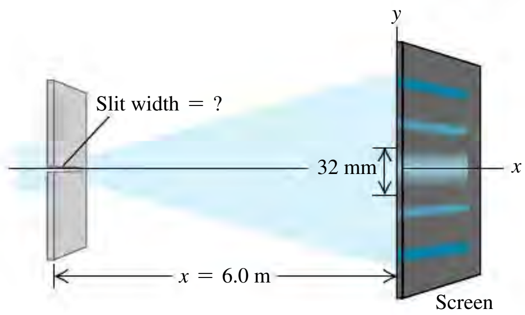


CAUTION Single-slit diffraction vs. two-slit interference Equation (36.3) has the same form as the equation for the two-slit pattern, Eq. (35.6), except that in Eq. (36.3) we use x rather than R for the distance to the screen. But Eq. (36.3) gives the positions of the *dark* fringes in a *single-slit* pattern rather than the *bright* fringes in a *double-slit* pattern. Also, $m = 0$ in Eq. (36.2) is *not* a dark fringe. Be careful!

EXAMPLE 36.1 SINGLE-SLIT DIFFRACTION

You pass 633-nm laser light through a narrow slit and observe the diffraction pattern on a screen 6.0 m away. The distance on the screen between the centers of the first minima on either side of the central bright fringe is 32 mm (**Fig. 36.7**). How wide is the slit?

36.7 A single-slit diffraction experiment.



SOLUTION

IDENTIFY and SET UP: This problem involves the relationship between the positions of dark fringes in a single-slit diffraction pattern and the slit width a (our target variable). The distances between fringes on the screen are much smaller than the slit-to-screen distance, so the angle θ shown in Fig. 36.5a is very small and we can use Eq. (36.3) to solve for a.

EXECUTE: The first minimum corresponds to $m = 1$ in Eq. (36.3). The distance y_1 from the central maximum to the first minimum on either side is half the distance between the two first minima, so $y_1 = (32\text{ mm})/2 = 16$ mm. Solving Eq. (36.3) for a, we find

$$a = \frac{x\lambda}{y_1} = \frac{(6.0\text{ m})(633 \times 10^{-9}\text{ m})}{16 \times 10^{-3}\text{ m}} = 2.4 \times 10^{-4}\text{ m} = 0.24\text{ mm}$$

EVALUATE: The angle θ is small only if the wavelength is small compared to the slit width. Since $\lambda = 633\text{ nm} = 6.33 \times 10^{-7}$ m and we have found $a = 0.24\text{ mm} = 2.4 \times 10^{-4}$ m, our result is consistent with this: The wavelength is $(6.33 \times 10^{-7}\text{ m})/(2.4 \times 10^{-4}\text{ m}) = 0.0026$ as large as the slit width. Can you show that the distance between the *second* minima on either side is $2(32\text{ mm}) = 64$ mm, and so on?

TEST YOUR UNDERSTANDING OF SECTION 36.2 Rank the following single-slit diffraction experiments in order of the size of the angle from the center of the diffraction pattern to the first dark fringe, from largest to smallest: (i) Wavelength 400 nm, slit width 0.20 mm; (ii) wavelength 600 nm, slit width 0.20 mm; (iii) wavelength 400 nm, slit width 0.30 mm; (iv) wavelength 600 nm, slit width 0.30 mm.

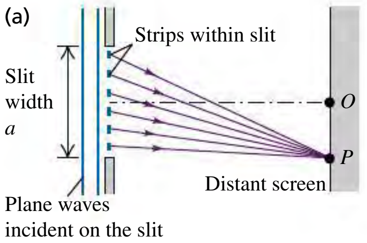
36.3 INTENSITY IN THE SINGLE-SLIT PATTERN

We can derive an expression for the intensity distribution for the single-slit diffraction pattern by the same phasor-addition method that we used in Section 35.3 for the two-slit interference pattern. We again imagine a plane wave front at the slit subdivided into a large number of strips. We superpose the contributions of the Huygens wavelets from all the strips at a point P on a distant screen at an angle θ from the normal to the slit plane (**Fig. 36.8a**, next page). To do this, we use a phasor to represent the sinusoidally varying $\vec{E}$ field from each strip. The magnitude of the vector sum of the phasors at each point P is the amplitude E_P of the total $\vec{E}$ field at that point. The intensity at P is proportional to $E_P{}^2$.

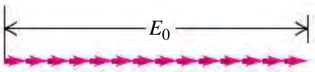
At the point O shown in Fig. 36.8a, corresponding to the center of the pattern where $\theta = 0$, there are negligible path differences for $x \gg a$; the phasors are all essentially *in phase* (that is, have the same direction). In Fig. 36.8b we draw the phasors at time $t = 0$ and denote the resultant amplitude at O by E_0. In this illustration we have divided the slit into 14 strips.

36.8 Using phasor diagrams to find the amplitude of the $\vec{E}$ field in single-slit diffraction. Each phasor represents the $\vec{E}$ field from a single strip within the slit.

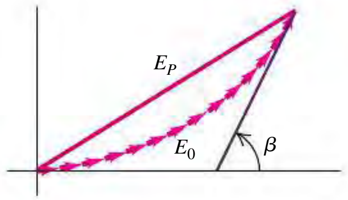
(a)



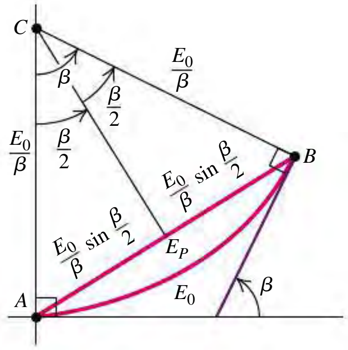
(b) At the center of the diffraction pattern (point O), the phasors from all strips within the slit are in phase.



(c) Phasor diagram at a point slightly off the center of the pattern; β = total phase difference between the first and last phasors.



(d) As in (c), but in the limit that the slit is subdivided into infinitely many strips



Now consider wavelets arriving from different strips at point P in Fig. 36.8a, at an angle θ from point O. Because of the differences in path length, there are now phase differences between wavelets coming from adjacent strips; the corresponding phasor diagram is shown in Fig. 36.8c. The vector sum of the phasors is now part of the perimeter of a many-sided polygon, and E_P, the amplitude of the resultant electric field at P, is the *chord*. The angle β is the total phase difference between the wave received at P from the top strip of Fig. 36.8a and the wave received at P from the bottom strip.

We may imagine dividing the slit into narrower and narrower strips. In the limit that there is an infinite number of infinitesimally narrow strips, the curved trail of phasors becomes an *arc of a circle* (Fig. 36.8d), with arc length equal to the length E_0 in Fig. 36.8b. The center C of this arc is found by constructing perpendiculars at A and B. From the relationship among arc length, radius, and angle, the radius of the arc is E_0/β; the amplitude E_P of the resultant electric field at P is equal to the chord AB, which is $2(E_0/\beta)\sin(\beta/2)$. (Note that β *must* be in radians!) We then have

$$E_P = E_0\frac{\sin(\beta/2)}{\beta/2} \qquad \text{(amplitude in single-slit diffraction)} \qquad (36.4)$$

The intensity at each point on the screen is proportional to the square of the amplitude given by Eq. (36.4). If I_0 is the intensity in the straight-ahead direction where $\theta = 0$ and $\beta = 0$, then the intensity I at any point is

$$I = I_0\left[\frac{\sin(\beta/2)}{\beta/2}\right]^2 \qquad \text{(intensity in single-slit diffraction)} \qquad (36.5)$$

We can express the phase difference β in terms of geometric quantities, as we did for the two-slit pattern. From Eq. (35.11) the phase difference is $2\pi/\lambda$ times the path difference. Figure 36.5 shows that the path difference between the ray from the top of the slit and the ray from the middle of the slit is $(a/2)\sin\theta$. The path difference between the rays from the top of the slit and the bottom of the slit is twice this, so

$$\beta = \frac{2\pi}{\lambda}a\sin\theta \qquad (36.6)$$

and Eq. (36.5) becomes

Angle of line from center of slit to position on screen

Intensity in single-slit diffraction

$$I = I_0\left\{\frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda}\right\}^2 \qquad (36.7)$$

Intensity at $\theta = 0$ Slit width   Wavelength

This equation expresses the intensity directly in terms of the angle θ. In many calculations it is easier first to calculate the phase angle β, from Eq. (36.6), and then to use Eq. (36.5).

Equation (36.7) is plotted in **Fig. 36.9a**. Note that the central intensity peak is much larger than any of the others. This means that most of the power in the wave remains within an angle θ from the perpendicular to the slit, where $\sin\theta = \lambda/a$ (the first diffraction minimum). You can see this easily in Fig. 36.9b, which is a photograph of water waves undergoing single-slit diffraction. Note also that the peak intensities in Fig. 36.9a decrease rapidly as we go away from the center of the pattern. (Compare Fig. 36.6, which shows a single-slit diffraction pattern for light.)

The dark fringes in the pattern are the places where $I = 0$. These occur at points for which the numerator of Eq. (36.5) is zero so that β is a multiple of 2π. From Eq. (36.6) this corresponds to

$$\frac{a\sin\theta}{\lambda} = m \qquad (m = \pm 1, \pm 2, \ldots)$$

$$\sin\theta = \frac{m\lambda}{a} \qquad (m = \pm 1,\ \pm 2, \ldots) \tag{36.8}$$

This agrees with our previous result, Eq. (36.2). Note again that $\beta = 0$ (corresponding to $\theta = 0$) is *not* a minimum. Equation (36.5) is indeterminate at $\beta = 0$, but we can evaluate the limit as $\beta \rightarrow 0$ by using L'Hôpital's rule. We find that at $\beta = 0$, $I = I_0$, as we should expect.

Intensity Maxima in the Single-Slit Pattern

We can also use Eq. (36.5) to calculate the positions of the peaks, or *intensity maxima*, and the intensities at these peaks. This is not quite as simple as it may appear. We might expect the peaks to occur where the sine function reaches the value ± 1—namely, where $\beta = \pm\pi,\ \pm 3\pi,\ \pm 5\pi$, or in general,

$$\beta \approx \pm(2m + 1)\pi \qquad (m = 0, 1, 2, \ldots) \tag{36.9}$$

This is *approximately* correct, but because of the factor $(\beta/2)^2$ in the denominator of Eq. (36.5), the maxima don't occur precisely at these points. When we take the derivative of Eq. (36.5) with respect to β and set it equal to zero to try to find the maxima and minima, we get a transcendental equation that has to be solved numerically. In fact there is *no* maximum near $\beta = \pm\pi$. The first maxima on either side of the central maximum, near $\beta = \pm 3\pi$, actually occur at $\pm 2.860\pi$. The second side maxima, near $\beta = \pm 5\pi$, are actually at $\pm 4.918\pi$, and so on. The error in Eq. (36.9) vanishes in the limit of large m—that is, for intensity maxima far from the center of the pattern.

To find the intensities at the side maxima, we substitute these values of β back into Eq. (36.5). Using the approximate expression in Eq. (36.9), we get

$$I_m \approx \frac{I_0}{\left(m + \frac{1}{2}\right)^2\pi^2} \tag{36.10}$$

where I_m is the intensity of the mth side maximum and I_0 is the intensity of the central maximum. Equation (36.10) gives the series of intensities

$$0.0450I_0 \qquad 0.0162I_0 \qquad 0.0083I_0$$

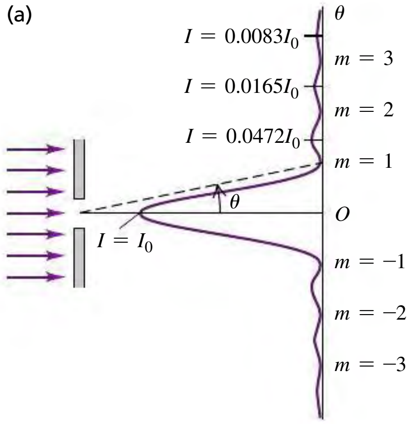
and so on. As we have pointed out, this equation is only approximately correct. The actual intensities of the side maxima turn out to be

$$0.0472I_0 \qquad 0.0165I_0 \qquad 0.0083I_0 \qquad \ldots$$

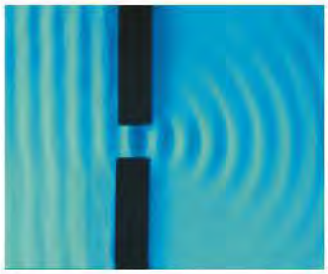
These intensities decrease very rapidly, as Fig. 36.9a also shows. Even the first side maxima have less than 5% of the intensity of the central maximum.

36.9 (a) Intensity versus angle in single-slit diffraction. The values of m label intensity minima given by Eq. (36.8). Most of the wave power goes into the central intensity peak (between the $m = 1$ and $m = -1$ intensity minima). (b) These water waves passing through a small aperture behave exactly like light waves in single-slit diffraction. Only the diffracted waves within the central intensity peak are visible; the waves at larger angles are too faint to see.

(a)



(b)



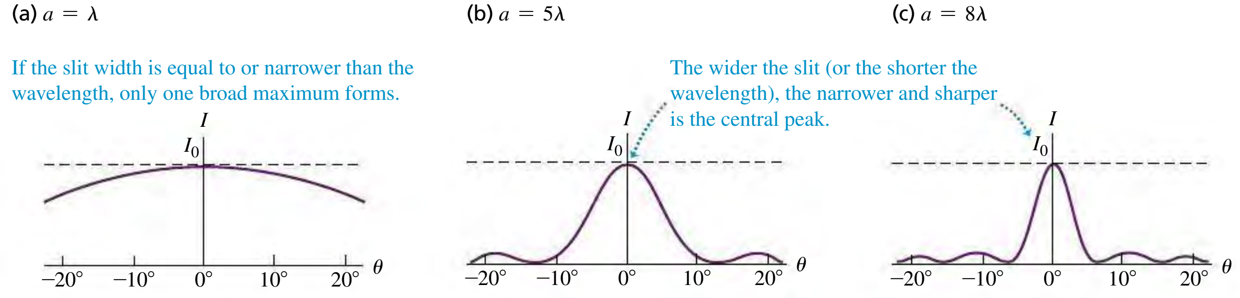
DATA *SPEAKS*

Single-Slit Diffraction

When students were given a problem involving diffraction of waves through a slit, more than 30% gave an incorrect response. Common errors:

- Confusion about the positions of dark fringes. Equation (36.2) gives the angle from the mth dark fringe to the center of the diffraction pattern—*not* the angle from a dark fringe on one side of the pattern to the corresponding dark fringe on the other side.
- Confusion about how the slit width a and wavelength λ affect the width of the diffraction pattern. Decreasing a or increasing λ makes the pattern broader; increasing a or decreasing λ makes the pattern narrower.

36.10 The single-slit diffraction pattern depends on the ratio of the slit width a to the wavelength λ.



Width of the Single-Slit Pattern

For small angles the angular spread of the diffraction pattern is inversely proportional to the ratio of the slit width a to the wavelength λ. **Figure 36.10** shows graphs of intensity I as a function of the angle θ for three values of the ratio a/λ.

With light waves, the wavelength λ is often much smaller than the slit width a, and the values of θ in Eqs. (36.6) and (36.7) are so small that the approximation $\sin\theta = \theta$ is very good. With this approximation the position θ_1 of the first $(m = 1)$ minimum, corresponding to $\beta/2 = \pi$, is, from Eq. (36.7),

$$\theta_1 = \frac{\lambda}{a} \tag{36.11}$$

This characterizes the width (angular spread) of the central maximum, and we see that it is *inversely* proportional to the slit width a. When the small-angle approximation is valid, the central maximum is exactly twice as wide as each side maximum. When a is of the order of a centimeter or more, θ_1 is so small that we can consider practically all the light to be concentrated at the geometrical focus. But when a is less than λ, the central maximum spreads over 180°, and the fringe pattern is not seen at all.

It's important to keep in mind that diffraction occurs for *all* kinds of waves, not just light. Sound waves undergo diffraction when they pass through a slit or aperture such as an ordinary doorway. The sound waves used in speech have wavelengths of about a meter or greater, and a typical doorway is less than 1 m wide; in this situation, a is less than λ, and the central intensity maximum extends over 180°. This is why the sounds coming through an open doorway can easily be heard by an eavesdropper hiding out of sight around the corner. In the same way, sound waves can bend around the head of an instructor who faces the blackboard while lecturing (**Fig. 36.11**). By contrast, there is essentially no diffraction of visible light through a doorway because the width a is very much greater than the wavelength λ (of order 5×10^{-7} m). You can *hear* around corners because typical sound waves have relatively long wavelengths; you cannot *see* around corners because the wavelength of visible light is very short.

36.11 The sound waves used in speech have a long wavelength (about 1 m) and can easily bend around this instructor's head. By contrast, light waves have very short wavelengths and undergo very little diffraction. Hence you can't *see* around his head!

EXAMPLE 36.2 SINGLE-SLIT DIFFRACTION: INTENSITY I

SOLUTION

(a) The intensity at the center of a single-slit diffraction pattern is I_0. What is the intensity at a point in the pattern where there is a 66-radian phase difference between wavelets from the two edges of the slit? (b) If this point is 7.0° away from the central maximum, how many wavelengths wide is the slit?

SOLUTION

IDENTIFY and SET UP: In our analysis of Fig. 36.8 we used the symbol β for the phase difference between wavelets from the two edges of the slit. In part (a) we use Eq. (36.5) to find the intensity I at the point in the pattern where $\beta = 66$ rad. In part (b) we need to find the slit width a as a multiple of the wavelength λ so our target variable is a/λ. We are given the angular position θ of the point where $\beta = 66$ rad, so we can use Eq. (36.6) to solve for a/λ.

EXECUTE: (a) We have $\beta/2 = 33$ rad, so from Eq. (36.5),

$$I = I_0\left[\frac{\sin(33\text{ rad})}{33\text{ rad}}\right]^2 = (9.2 \times 10^{-4})I_0$$

(b) From Eq. (36.6),

$$\frac{a}{\lambda} = \frac{\beta}{2\pi \sin\theta} = \frac{66 \text{ rad}}{(2\pi \text{ rad})\sin 7.0°} = 86$$

For example, for 550-nm light the slit width is $a = (86)(550 \text{ nm}) = 4.7 \times 10^{-5} \text{ m} = 0.047 \text{ mm}$, or roughly $\frac{1}{20}$ mm.

EVALUATE: To what point in the diffraction pattern does this value of β correspond? To find out, note that $\beta = 66$ rad is approximately equal to 21π. This is an odd multiple of π, corresponding to the form $(2m + 1)\pi$ found in Eq. (36.9) for the intensity *maxima*. Hence $\beta = 66$ rad corresponds to a point near the tenth ($m = 10$) maximum. This is well beyond the range shown in Fig. 36.9a, which shows only maxima out to $m = \pm 3$.

EXAMPLE 36.3 SINGLE-SLIT DIFFRACTION: INTENSITY II

SOLUTION

In the experiment described in Example 36.1 (Section 36.2), the intensity at the center of the pattern is I_0. What is the intensity at a point on the screen 3.0 mm from the center of the pattern?

SOLUTION

IDENTIFY and SET UP: This is similar to Example 36.2, except that we are not given the value of the phase difference β at the point in question. We use geometry to determine the angle θ for our point and then use Eq. (36.7) to find the intensity I (the target variable).

EXECUTE: Referring to Fig. 36.5a, we have $y = 3.0$ mm and $x = 6.0$ m, so $\tan\theta = y/x = (3.0 \times 10^{-3} \text{ m})/(6.0 \text{ m}) = 5.0 \times 10^{-4}$. This is so small that the values of $\tan\theta$, $\sin\theta$, and θ (in radians) are all nearly the same. Then, using Eq. (36.7),

$$\frac{\pi a \sin\theta}{\lambda} = \frac{\pi(2.4 \times 10^{-4} \text{ m})(5.0 \times 10^{-4})}{6.33 \times 10^{-7} \text{ m}} = 0.60$$

$$I = I_0\left(\frac{\sin 0.60}{0.60}\right)^2 = 0.89 I_0$$

EVALUATE: Figure 36.9a shows that an intensity this high can occur only within the central intensity maximum. This checks out; from Example 36.1, the first intensity minimum ($m = 1$ in Fig. 36.9a) is $(32 \text{ mm})/2 = 16$ mm from the center of the pattern, so the point in question here at $y = 3$ mm does, indeed, lie within the central maximum.

TEST YOUR UNDERSTANDING OF SECTION 36.3 Coherent electromagnetic radiation is sent through a slit of width 0.0100 mm. For which of the following wavelengths will there be *no* points in the diffraction pattern where the intensity is zero? (i) Blue light of wavelength 500 nm; (ii) infrared light of wavelength 10.6 μm; (iii) microwaves of wavelength 1.00 mm; (iv) ultraviolet light of wavelength 50.0 nm. ❙

36.4 MULTIPLE SLITS

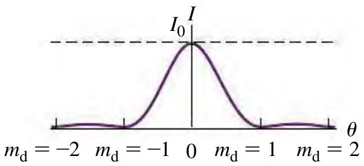
In Sections 35.2 and 35.3 we analyzed interference from two point sources and from two very narrow slits; in this analysis we ignored effects due to the finite (that is, nonzero) slit width. In Sections 36.2 and 36.3 we considered the diffraction effects that occur when light passes through a single slit of finite width. Additional interesting effects occur when we have two slits with finite width or when there are several very narrow slits.

Two Slits of Finite Width

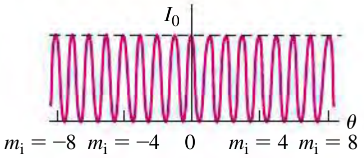
Let's take another look at the two-slit pattern in the more realistic case in which the slits have finite width. If the slits are narrow in comparison to the wavelength, we can assume that light from each slit spreads out uniformly in all directions to the right of the slit. We used this assumption in Section 35.3 to calculate the interference pattern described by Eq. (35.10) or (35.15), consisting of a series of equally spaced, equally intense maxima. However, when the slits have finite width, the peaks in the two-slit interference pattern are modulated by the single-slit diffraction pattern characteristic of the width of each slit.

36.12 Finding the intensity pattern for two slits of finite width.

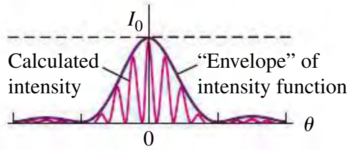
(a) Single-slit diffraction pattern for a slit width a



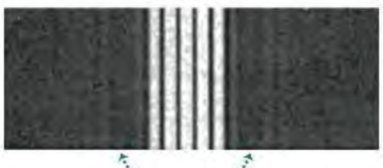
(b) Two-slit interference pattern for narrow slits whose separation d is four times the width of the slit in **(a)**



(c) Calculated intensity pattern for two slits of width a and separation $d = 4a$, including both interference and diffraction effects



(d) Photograph of the pattern calculated in **(c)**



For $d = 4a$, every fourth interference maximum at the sides ($m_i = \pm 4, \pm 8, \ldots$) is missing.

Figure 36.12a shows the intensity in a single-slit diffraction pattern with slit width a. The *diffraction minima* are labeled by the integer $m_d = \pm 1, \pm 2, \ldots$ ("d" for "diffraction"). Figure 36.12b shows the pattern formed by two very narrow slits with distance d between slits, where d is four times as great as the single-slit width a in Fig. 36.12a; that is, $d = 4a$. The *interference maxima* are labeled by the integer $m_i = 0, \pm 1, \pm 2, \ldots$ ("i" for "interference"). We note that the spacing between adjacent minima in the single-slit pattern is four times as great as in the two-slit pattern. Now suppose we widen each of the narrow slits to the same width a as that of the single slit in Fig. 36.12a. Figure 36.12c shows the pattern from two slits with width a, separated by a distance (between centers) $d = 4a$. The effect of the finite width of the slits is to superimpose the two patterns—that is, to multiply the two intensities at each point. The two-slit peaks are in the same positions as before, but their intensities are modulated by the single-slit pattern, which acts as an "envelope" for the intensity function. The expression for the intensity shown in Fig. 36.12c is proportional to the product of the two-slit and single-slit expressions, Eqs. (35.10) and (36.5):

$$I = I_0 \cos^2 \frac{\phi}{2} \left[\frac{\sin(\beta/2)}{\beta/2} \right]^2 \qquad \text{(two slits of finite width)} \qquad (36.12)$$

where, as before,

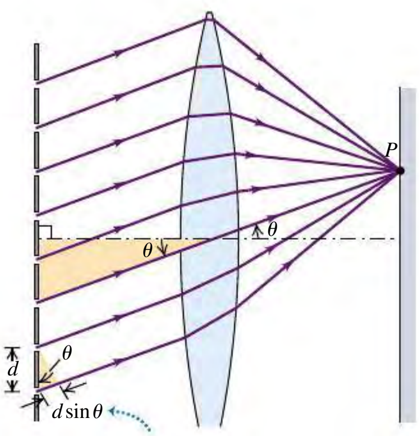
$$\phi = \frac{2\pi d}{\lambda} \sin\theta \qquad \beta = \frac{2\pi a}{\lambda} \sin\theta$$

In Fig. 36.12c, every fourth interference maximum at the sides is *missing* because these interference maxima ($m_i = \pm 4, \pm 8, \ldots$) coincide with diffraction minima ($m_d = \pm 1, \pm 2, \ldots$). This can also be seen in Fig. 36.12d, which is a photograph of an actual pattern with $d = 4a$. You should be able to convince yourself that there will be "missing" maxima whenever d is an integer multiple of a.

Figures 36.12c and 36.12d show that as you move away from the central bright maximum of the two-slit pattern, the intensity of the maxima decreases. This is a result of the single-slit modulating pattern shown in Fig. 36.12a; mathematically, the decrease in intensity arises from the factor $(\beta/2)^2$ in the denominator of Eq. (36.12). You can also see this decrease in Fig. 35.6 (Section 35.2). The narrower the slits, the broader the single-slit pattern (as in Fig. 36.10) and the slower the decrease in intensity from one interference maximum to the next.

Shall we call the pattern in Fig. 36.12d *interference* or *diffraction*? It's really both, since it results from the superposition of waves coming from various parts of the two apertures.

36.13 Multiple-slit diffraction. Here a lens is used to give a Fraunhofer pattern on a nearby screen, as in Fig. 36.4d.



Maxima occur where the path difference for adjacent slits is a whole number of wavelengths: $d \sin\theta = m\lambda$.

Several Slits

Next let's consider patterns produced by *several* very narrow slits. As we will see, systems of narrow slits are of tremendous practical importance in *spectroscopy,* the determination of the particular wavelengths of light coming from a source. Assume that each slit is narrow in comparison to the wavelength, so its diffraction pattern spreads out nearly uniformly. **Figure 36.13** shows an array of eight narrow slits, with distance d between adjacent slits. Constructive interference occurs for rays at angle θ to the normal that arrive at point P with a path difference between adjacent slits equal to an integer number of wavelengths:

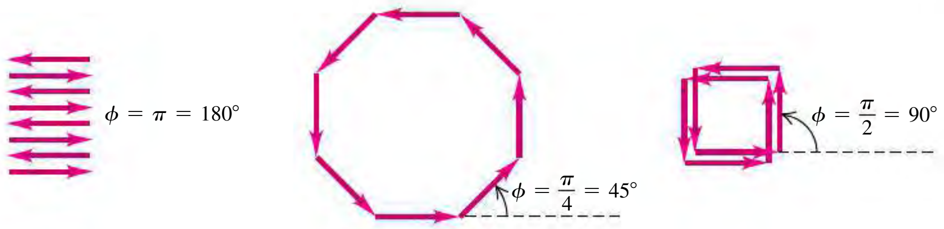
$$d \sin\theta = m\lambda \qquad (m = 0, \pm 1, \pm 2, \ldots)$$

This means that reinforcement occurs when the phase difference ϕ at P for light from adjacent slits is an integer multiple of 2π. That is, the maxima in the pattern occur at the *same* positions as for *two* slits with the same spacing.

What happens *between* the maxima is different with multiple slits, however. In the two-slit pattern, there is exactly one intensity minimum located midway between each pair of maxima, corresponding to angles for which the phase

36.14 Phasor diagrams for light passing through eight narrow slits. Intensity maxima occur when the phase difference $\phi = 0, 2\pi, 4\pi, \ldots$. Between the maxima at $\phi = 0$ and $\phi = 2\pi$ are seven minima, corresponding to $\phi = \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2$, and $7\pi/4$. Can you draw phasor diagrams for the other minima?

(a) Phasor diagram for $\phi = \pi$ (b) Phasor diagram for $\phi = \frac{\pi}{4}$ (c) Phasor diagram for $\phi = \frac{\pi}{2}$



difference between waves from the two sources is π, 3π, 5π, and so on. In the eight-slit pattern these are also minima because the light from adjacent slits cancels out in pairs, corresponding to the phasor diagram in **Fig. 36.14a**. But these are not the only minima in the eight-slit pattern. For example, when the phase difference ϕ from adjacent sources is $\pi/4$, the phasor diagram is as shown in Fig. 36.14b; the total (resultant) phasor is zero, and the intensity is zero. When $\phi = \pi/2$, we get the phasor diagram of Fig. 36.14c, and again both the total phasor and the intensity are zero. More generally, the intensity with eight slits is zero whenever ϕ is an integer multiple of $\pi/4$, *except* when ϕ is a multiple of 2π. Thus there are seven minima for every maximum.

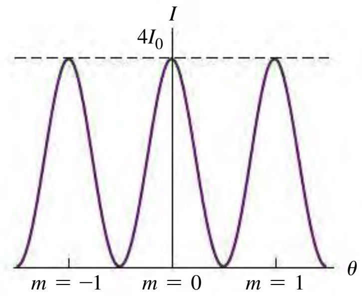
Figure 36.15b shows the result of a detailed calculation of the eight-slit pattern. The large maxima, called *principal maxima,* are in the same positions as for the two-slit pattern of Fig. 36.15a but are much narrower. If the phase difference ϕ between adjacent slits is slightly different from a multiple of 2π, the waves from slits 1 and 2 will be only a little out of phase; however, the phase difference between slits 1 and 3 will be greater, that between slits 1 and 4 will be greater still, and so on. This leads to a partial cancellation for angles that are only slightly different from the angle for a maximum, giving the narrow maxima in Fig. 36.15b. The maxima are even narrower with 16 slits (Fig. 36.15c).

You should show that when there are N slits, there are $(N - 1)$ minima between each pair of principal maxima and a minimum occurs whenever ϕ is an integral multiple of $2\pi/N$ (except when ϕ is an integral multiple of 2π, which gives a principal maximum). There are small *secondary* intensity maxima between the minima; these become smaller in comparison to the principal maxima as N increases. The greater the value of N, the narrower the principal maxima become. From an energy standpoint the total power in the entire pattern is proportional to N. The height of each principal maximum is proportional to N^2, so from energy conservation the width of each principal maximum must be proportional to $1/N$. In the next section we'll see why the details of the multiple-slit pattern are of great practical importance.

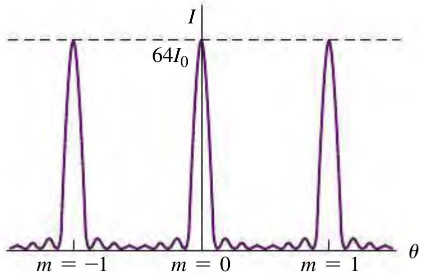
TEST YOUR UNDERSTANDING OF SECTION 36.4 Suppose two slits, each of width a, are separated by a distance $d = 2.5a$. Are there any missing maxima in the interference pattern produced by these slits? If so, which are missing? If not, why not?

36.15 Interference patterns for N equally spaced, very narrow slits. (a) Two slits. (b) Eight slits. (c) Sixteen slits. The vertical scales are different for each graph; the maximum intensity is I_0 for a single slit and $N^2 I_0$ for N slits. The width of each peak is proportional to $1/N$.

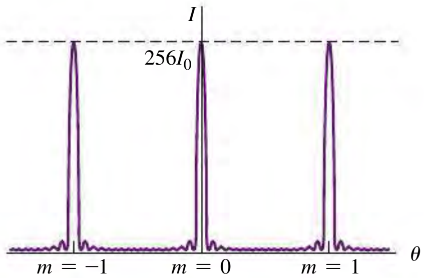
(a) $N = 2$: two slits produce one minimum between adjacent maxima.



(b) $N = 8$: eight slits produce taller, narrower maxima in the same locations, separated by seven minima.



(c) $N = 16$: with 16 slits, the maxima are even taller and narrower, with more intervening minima.

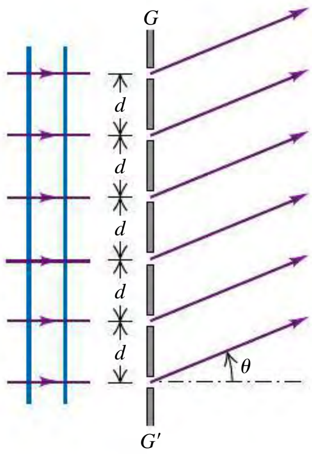


36.5 THE DIFFRACTION GRATING

We have just seen that increasing the number of slits in an interference experiment (while keeping the spacing of adjacent slits constant) gives interference patterns in which the maxima are in the same positions, but progressively narrower, than with two slits. Because these maxima are so narrow, their angular position, and hence the wavelength, can be measured to very high precision. As we will see, this effect has many important applications.

An array of a large number of parallel slits, all with the same width a and spaced equal distances d between centers, is called a **diffraction grating.** The first one was constructed by Fraunhofer using fine wires. Gratings can be made by using a diamond point to scratch many equally spaced grooves on a glass or metal surface, or by photographic reduction of a pattern of black and white stripes on paper. For a grating, what we have been calling *slits* are often called *rulings* or *lines*.

36.16 A portion of a transmission diffraction grating. The separation between the centers of adjacent slits is d.



In **Fig. 36.16**, GG' is a cross section of a *transmission grating;* the slits are perpendicular to the plane of the page, and an interference pattern is formed by the light that is transmitted through the slits. The diagram shows only six slits; an actual grating may contain several thousand. The spacing d between centers of adjacent slits is called the *grating spacing.* A plane monochromatic wave is incident normally on the grating from the left side. We assume far-field (Fraunhofer) conditions; that is, the pattern is formed on a screen that is far enough away that all rays emerging from the grating and going to a particular point on the screen can be considered to be parallel.

We found in Section 36.4 that the principal intensity maxima with multiple slits occur in the same directions as for the two-slit pattern. These are the directions for which the path difference for adjacent slits is an integer number of wavelengths. So the positions of the maxima are once again

Intensity maxima, multiple slits:

$$\underbrace{d}_{\text{Distance between slits}}\sin\underbrace{\theta}_{\substack{\text{Angle of line from center of slit array}\\ \text{to } m\text{th bright region on screen}}} = m\underbrace{\lambda}_{\text{Wavelength}} \quad (m = 0, \pm 1, \pm 2, \dots) \qquad (36.13)$$

The intensity patterns for two, eight, and 16 slits displayed in Fig. 36.15 show the progressive increase in sharpness of the maxima as the number of slits increases.

When a grating containing hundreds or thousands of slits is illuminated by a beam of parallel rays of monochromatic light, the pattern is a series of very sharp lines at angles determined by Eq. (36.13). The $m = \pm 1$ lines are called the *first-order lines,* the $m = \pm 2$ lines the *second-order lines*, and so on. If the grating is illuminated by white light with a continuous distribution of wavelengths, each value of m corresponds to a continuous spectrum in the pattern. The angle for each wavelength is determined by Eq. (36.13); for a given value of m, long wavelengths (the red end of the spectrum) lie at larger angles (that is, are deviated more from the straight-ahead direction) than do the shorter wavelengths at the violet end of the spectrum.

As Eq. (36.13) shows, the sines of the deviation angles of the maxima are proportional to the ratio λ/d. For substantial deviation to occur, the grating spacing d should be of the same order of magnitude as the wavelength λ. Gratings for use with visible light (λ from 400 to 700 nm) usually have about 1000 slits per millimeter; the value of d is the *reciprocal* of the number of slits per unit length, so d is of the order of $\frac{1}{1000}$ mm $= 1000$ nm.

36.17 Microscopic pits on the surface of this DVD act as a reflection grating, splitting white light into its component colors.

In a *reflection grating,* the array of equally spaced slits shown in Fig. 36.16 is replaced by an array of equally spaced ridges or grooves on a reflective screen. The reflected light has maximum intensity at angles where the phase difference between light waves reflected from adjacent ridges or grooves is an integral multiple of 2π. If light of wavelength λ is incident normally on a reflection grating with a spacing d between adjacent ridges or grooves, the *reflected* angles at which intensity maxima occur are given by Eq. (36.13).

The rainbow-colored reflections from the surface of a DVD are a reflection-grating effect (**Fig. 36.17**). The "grooves" are tiny pits 0.12 μm deep in the surface of the disc, with a uniform radial spacing of 0.74 μm $= 740$ nm. Information is coded on the DVD by varying the *length* of the pits. The reflection-grating aspect of the disc is merely an aesthetic side benefit.

EXAMPLE 36.4 WIDTH OF A GRATING SPECTRUM

SOLUTION

The wavelengths of the visible spectrum are approximately 380 nm (violet) to 750 nm (red). (a) Find the angular limits of the first-order visible spectrum produced by a plane grating with 600 slits per millimeter when white light falls normally on the grating. (b) Do the first-order and second-order spectra overlap? What about the second-order and third-order spectra? Do your answers depend on the grating spacing?

SOLUTION

IDENTIFY and SET UP: We must find the angles spanned by the visible spectrum in the first-, second-, and third-order spectra. These correspond to $m = 1$, 2, and 3 in Eq. (36.13).

EXECUTE: (a) The grating spacing is

$$d = \frac{1}{600 \text{ slits/mm}} = 1.67 \times 10^{-6} \text{ m}$$

We solve Eq. (36.13) for θ:

$$\theta = \arcsin\frac{m\lambda}{d}$$

Then for $m = 1$, the angular deviations θ_{v1} and θ_{r1} for violet and red light, respectively, are

$$\theta_{v1} = \arcsin\left(\frac{380 \times 10^{-9} \text{ m}}{1.67 \times 10^{-6} \text{ m}}\right) = 13.2°$$

$$\theta_{r1} = \arcsin\left(\frac{750 \times 10^{-9} \text{ m}}{1.67 \times 10^{-6} \text{ m}}\right) = 26.7°$$

That is, the first-order visible spectrum appears with deflection angles from $\theta_{v1} = 13.2°$ (violet) to $\theta_{r1} = 26.7°$ (red).

(b) With $m = 2$ and $m = 3$, our equation $\theta = \arcsin(m\lambda/d)$ for 380-mm violet light yields

$$\theta_{v2} = \arcsin\left(\frac{2(380 \times 10^{-9} \text{ m})}{1.67 \times 10^{-6} \text{ m}}\right) = 27.1°$$

$$\theta_{v3} = \arcsin\left(\frac{3(380 \times 10^{-9} \text{ m})}{1.67 \times 10^{-6} \text{ m}}\right) = 43.0°$$

For 750-nm red light, this same equation gives

$$\theta_{r2} = \arcsin\left(\frac{2(750 \times 10^{-9} \text{ m})}{1.67 \times 10^{-6} \text{ m}}\right) = 63.9°$$

$$\theta_{r3} = \arcsin\left(\frac{3(750 \times 10^{-9} \text{ m})}{1.67 \times 10^{-6} \text{ m}}\right) = \arcsin(1.35) = \text{undefined}$$

Hence the second-order spectrum extends from 27.1° to 63.9° and the third-order spectrum extends from 43.0° to 90° (the largest possible value of θ). The undefined value of θ_{r3} means that the third-order spectrum reaches $\theta = 90° = \arcsin(1)$ at a wavelength shorter than 750 nm; you should be able to show that this happens for $\lambda = 557$ nm. Hence the first-order spectrum (from 13.2° to 26.7°) does not overlap with the second-order spectrum, but the second- and third-order spectra do overlap. You can convince yourself that this is true for any value of the grating spacing d.

EVALUATE: The fundamental reason the first-order and second-order visible spectra don't overlap is that the human eye is sensitive to only a narrow range of wavelengths. Can you show that if the eye could detect wavelengths from 380 nm to 900 nm (in the near-infrared range), the first and second orders *would* overlap?

Grating Spectrographs

Diffraction gratings are widely used to measure the spectrum of light emitted by a source, a process called *spectroscopy* or *spectrometry*. Light incident on a grating of known spacing is dispersed into a spectrum. The angles of deviation of the maxima are then measured, and Eq. (36.13) is used to compute the wavelength. With a grating that has many slits, very sharp maxima are produced, and the angle of deviation (and hence the wavelength) can be measured very precisely.

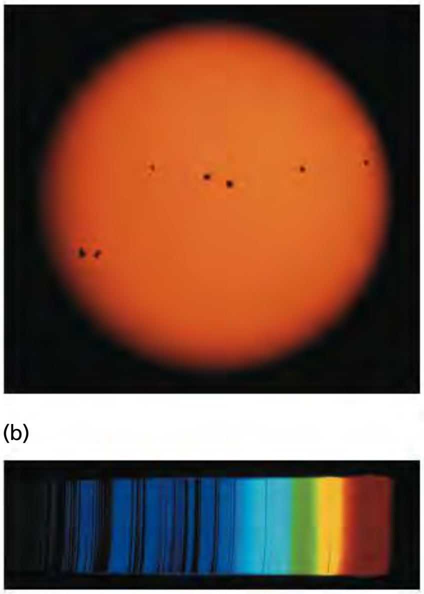
An important application of this technique is to astronomy. As light generated within the sun passes through the sun's atmosphere, certain wavelengths are selectively absorbed. The result is that the spectrum of sunlight produced by a diffraction grating has dark *absorption lines* (**Fig. 36.18**). Experiments in the laboratory show that different types of atoms and ions absorb light of different wavelengths. By comparing these laboratory results with the wavelengths of absorption lines in the spectrum of sunlight, astronomers can deduce the chemical composition of the sun's atmosphere. The same technique is used to make chemical assays of galaxies that are millions of light-years away.

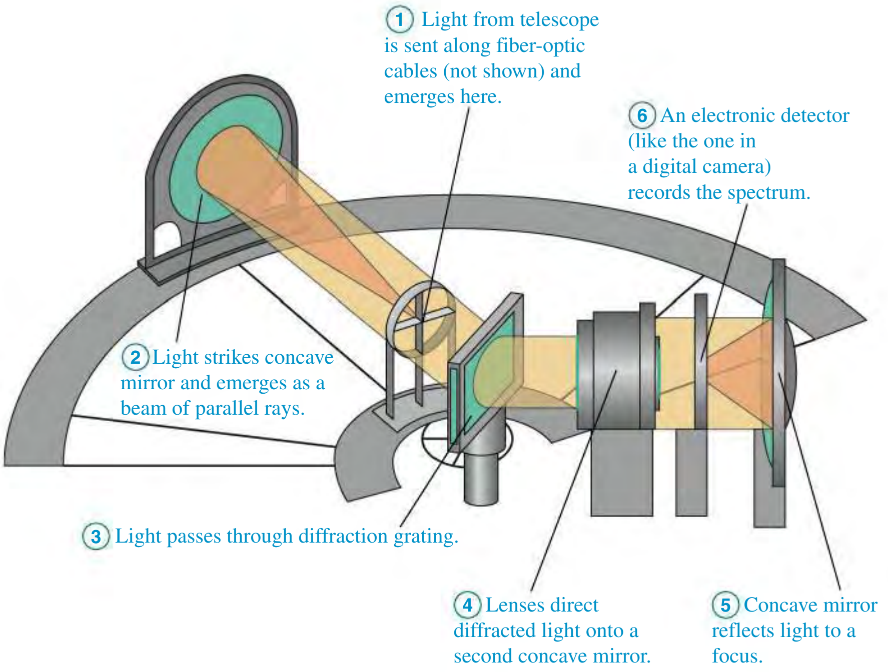
Figure 36.19, next page, shows one design for a *grating spectrograph* used in astronomy. A transmission grating is used in the figure; in other setups, a reflection grating is used. In older designs a prism was used rather than a grating, and a spectrum was formed by dispersion (see Section 33.4) rather than diffraction. However, there is no simple relationship between wavelength and angle of deviation for a prism, prisms absorb some of the light that passes through them, and they are less effective for many nonvisible wavelengths that are important in astronomy. For these and other reasons, gratings are preferred in precision applications.

36.18 (a) A visible-light photograph of the sun. (b) Sunlight is dispersed into a spectrum by a diffraction grating. Specific wavelengths are absorbed as sunlight passes through the sun's atmosphere, leaving dark lines in the spectrum.

(a)

(b)





36.19 A schematic diagram of a diffraction-grating spectrograph for use in astronomy. Note that the light does not strike the grating normal to its surface, so the intensity maxima are given by a somewhat different expression than Eq. (36.13).

BIO Application Detecting DNA with Diffraction Diffraction gratings are used in a common piece of laboratory equipment known as a spectrophotometer. Light shining across a diffraction grating is dispersed into its component wavelengths. A slit is used to block all but a very narrow range of wavelengths, producing a beam of almost perfectly monochromatic light. The instrument then measures how much of that light is absorbed by a solution of biological molecules. For example, the sample tube shown here contains a solution of DNA, which is transparent to visible light but which strongly absorbs ultraviolet light with a wavelength of exactly 260 nm. Therefore, by illuminating the sample with 260-nm light and measuring the amount absorbed, we can determine the concentration of DNA in the solution.

Resolution of a Grating Spectrograph

In spectroscopy it is often important to distinguish slightly differing wavelengths. The minimum wavelength difference $\Delta\lambda$ that can be distinguished by a spectrograph is described by the **chromatic resolving power** R, defined as

$$R = \frac{\lambda}{\Delta\lambda} \quad \text{(chromatic resolving power)} \qquad (36.14)$$

As an example, when sodium atoms are heated, they emit strongly at the yellow wavelengths 589.00 nm and 589.59 nm. A spectrograph that can barely distinguish these two lines in the spectrum (called the *sodium doublet*) has a chromatic resolving power $R = (589.00\text{ nm})/(0.59\text{ nm}) = 1000$. (You can see these wavelengths when boiling water on a gas range. If the water boils over onto the flame, dissolved sodium from table salt emits a burst of yellow light.)

We can derive an expression for the resolving power of a diffraction grating used in a spectrograph. Two different wavelengths give diffraction maxima at slightly different angles. As a reasonable (though arbitrary) criterion, let's assume that we can distinguish them as two separate peaks if the maximum of one coincides with the first minimum of the other.

From our discussion in Section 36.4 the mth-order maximum occurs when the phase difference ϕ for adjacent slits is $\phi = 2\pi m$. The first minimum beside that maximum occurs when $\phi = 2\pi m + 2\pi/N$, where N is the number of slits. The phase difference is also given by $\phi = (2\pi d \sin\theta)/\lambda$, so the angular interval $d\theta$ corresponding to a small increment $d\phi$ in the phase shift can be obtained from the differential of this equation:

$$d\phi = \frac{2\pi d \cos\theta\, d\theta}{\lambda}$$

When $d\phi = 2\pi/N$, this corresponds to the angular interval $d\theta$ between a maximum and the first adjacent minimum. Thus $d\theta$ is given by

$$\frac{2\pi}{N} = \frac{2\pi d \cos\theta\, d\theta}{\lambda} \quad \text{or} \quad d\cos\theta\, d\theta = \frac{\lambda}{N}$$

CAUTION Watch out for different uses of the symbol *d* Don't confuse the slit spacing d with the differential "d" in the angular interval $d\theta$ or in the phase shift increment $d\phi$!

Now we need to find the angular spacing $d\theta$ between maxima for two slightly different wavelengths. The positions of these maxima are given by $d\sin\theta = m\lambda$, and the differential of this equation gives

$$d\cos\theta\, d\theta = m\, d\lambda$$

According to our criterion, the limit or resolution is reached when these two angular spacings are equal. Equating the two expressions for the quantity $(d\cos\theta\, d\theta)$, we find

$$\frac{\lambda}{N} = m\, d\lambda \qquad \text{and} \qquad \frac{\lambda}{d\lambda} = Nm$$

If $\Delta\lambda$ is small, we can replace $d\lambda$ by $\Delta\lambda$, and the resolving power R is

$$R = \frac{\lambda}{\Delta\lambda} = Nm \qquad (36.15)$$

The greater the number of slits N, the better the resolution; also, the higher the order m of the diffraction-pattern maximum that we use, the better the resolution.

TEST YOUR UNDERSTANDING OF SECTION 36.5 What minimum number of slits would be required in a grating to resolve the sodium doublet in the fourth order? (i) 250; (ii) 400; (iii) 1000; (iv) 4000. ❙

36.6 X-RAY DIFFRACTION

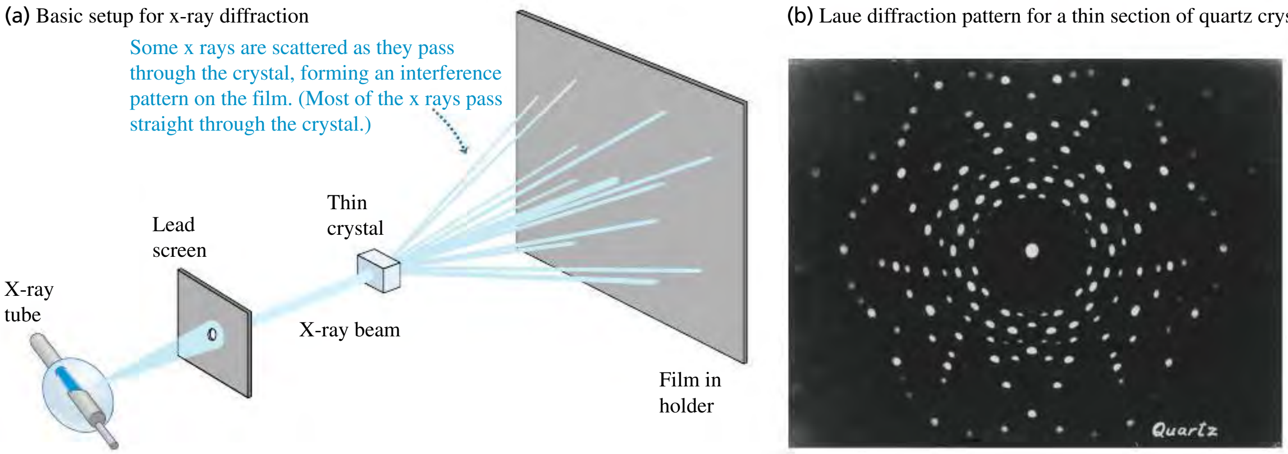
X rays were discovered by Wilhelm Röntgen (1845–1923) in 1895, and early experiments suggested that they were electromagnetic waves with wavelengths of the order of 10^{-10} m. At about the same time, the idea began to emerge that in a crystalline solid the atoms are arranged in a regular repeating pattern, with spacing between adjacent atoms also of the order of 10^{-10} m. Putting these two ideas together, Max von Laue (1879–1960) proposed in 1912 that a crystal might serve as a kind of three-dimensional diffraction grating for x rays. That is, a beam of x rays might be scattered (that is, absorbed and re-emitted) by the individual atoms in a crystal, and the scattered waves might interfere just like waves from a diffraction grating.

The first **x-ray diffraction** experiments were performed in 1912 by Friedrich, Knipping, and von Laue, using the experimental setup shown in **Fig. 36.20a**. The scattered x rays *did* form an interference pattern, which they recorded on photographic film. Figure 36.20b is a photograph of such a pattern. These experiments

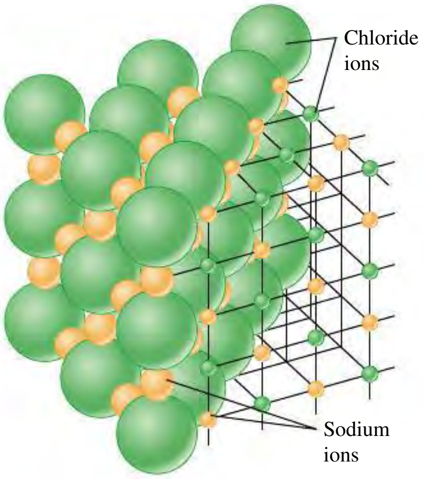
36.20 (a) An x-ray diffraction experiment. (b) Diffraction pattern (or *Laue pattern*) formed by directing a beam of x rays at a thin section of quartz crystal.

(a) Basic setup for x-ray diffraction

(b) Laue diffraction pattern for a thin section of quartz crystal



36.21 Model of the arrangement of ions in a crystal of NaCl (table salt). The spacing of adjacent atoms is 0.282 nm. (The electron clouds of the atoms actually overlap slightly.)



verified that x rays *are* waves, or at least have wavelike properties, and also that the atoms in a crystal *are* arranged in a regular pattern (**Fig. 36.21**). Since that time, x-ray diffraction has proved to be an invaluable research tool, both for measuring x-ray wavelengths and for studying the structure of crystals and complex molecules.

A Simple Model of X-Ray Diffraction

To better understand x-ray diffraction, we consider first a two-dimensional scattering situation, as shown in **Fig. 36.22a**, in which a plane wave is incident on a rectangular array of scattering centers. The situation might be a ripple tank with an array of small posts or x rays incident on an array of atoms. In the case of electromagnetic waves, the wave induces an oscillating electric dipole moment in each scatterer. These dipoles act like little antennas, emitting scattered waves. The resulting interference pattern is the superposition of all these scattered waves. The situation is different from that with a diffraction grating, in which the waves from all the slits are emitted *in phase* (for a plane wave at normal incidence). Here the scattered waves are *not* all in phase because their distances from the *source* are different. To compute the interference pattern, we have to consider the *total* path differences for the scattered waves, including the distances from source to scatterer and from scatterer to observer.

As Fig. 36.22b shows, the path length from source to observer is the same for all the scatterers in a single row if the two angles θ_a and θ_r are equal. Scattered radiation from *adjacent* rows is *also* in phase if the path difference for adjacent rows is an integer number of wavelengths. Figure 36.22c shows that this path difference is $2d\sin\theta$, where θ is the common value of θ_a and θ_r. Therefore the conditions for radiation from the *entire array* to reach the observer in phase are (1) the angle of incidence must equal the angle of scattering and (2) the path difference for adjacent rows must equal $m\lambda$, where m is an integer. We can express the second condition, called the **Bragg condition** in honor of x-ray diffraction pioneers Sir William Bragg and his son Laurence Bragg, as

Bragg condition for constructive interference from an array:

Distance between adjacent rows in array; Wavelength

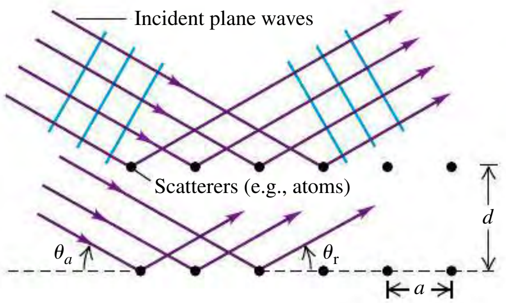
$$2d\sin\theta = m\lambda \quad (m = 1, 2, 3, \ldots) \tag{36.16}$$

Angle of line from surface of array to *m*th bright region on screen

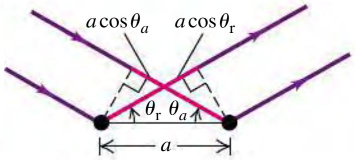
CAUTION Scattering from an array In Eq. (36.16) the angle θ is measured with respect to the *surface* of the crystal rather than with respect to the *normal* to the plane of an array of slits or a grating. Also, note that the path difference in Eq. (36.16) is $2d\sin\theta$, not $d\sin\theta$ as in Eq. (36.13) for a diffraction grating.

36.22 A two-dimensional model of scattering from a rectangular array. The distance between adjacent atoms in a horizontal row is a; the distance between adjacent rows is d. The angles in (b) are measured from the *surface* of the array, not from its normal.

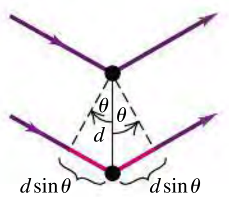
(a) Scattering of waves from a rectangular array

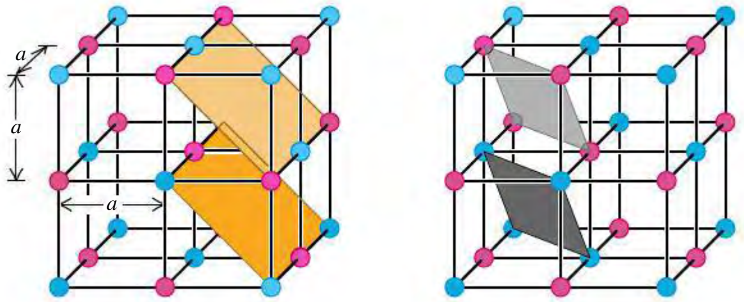


(b) Scattering from adjacent atoms in a row
Interference from adjacent atoms in a row is constructive when the path lengths $a\cos\theta_a$ and $a\cos\theta_r$ are equal, so that the angle of incidence θ_a equals the angle of reflection (scattering) θ_r.



(c) Scattering from atoms in adjacent rows
Interference from atoms in adjacent rows is constructive when the path difference $2d\sin\theta$ is an integral number of wavelengths, as in Eq. (36.16).





36.23 A cubic crystal and two different families of crystal planes. There are also three sets of planes parallel to the cube faces, with spacing a.

In directions for which Eq. (36.16) is satisfied, we see a strong maximum in the interference pattern. We can describe this interference in terms of *reflections* of the wave from the horizontal rows of scatterers in Fig. 36.22a. Strong reflection (constructive interference) occurs at angles such that the incident and scattered angles are equal and Eq. (36.16) is satisfied. Since $\sin\theta$ can never be greater than 1, Eq. (36.16) says that to have constructive interference the quantity $m\lambda$ must be less than $2d$ and so λ must be less than $2d/m$. For example, the value of d in an NaCl crystal (see Fig. 36.21) is only 0.282 nm. Hence to have the mth-order maximum present in the diffraction pattern, λ must be less than $2(0.282\text{ nm})/m$; that is, $\lambda < 0.564$ nm for $m = 1$, $\lambda < 0.282$ nm for $m = 2$, $\lambda < 0.188$ nm for $m = 3$, and so on. These are all x-ray wavelengths (see Fig. 32.4), which is why x rays are used for studying crystal structure.

We can extend this discussion to a three-dimensional array by considering *planes* of scatterers instead of *rows*. **Figure 36.23** shows two different sets of parallel planes that pass through all the scatterers. Waves from all the scatterers in a given plane interfere constructively if the angles of incidence and scattering are equal. There is also constructive interference between planes when Eq. (36.16) is satisfied, where d is now the distance between adjacent planes. Because there are many different sets of parallel planes, there are also many values of d and many sets of angles that give constructive interference for the whole crystal lattice. This phenomenon is called **Bragg reflection.**

CAUTION **Bragg *reflection* is really Bragg *interference*** While we are using the term *reflection,* remember that we are dealing with an *interference* effect. The reflections from various planes are closely analogous to interference effects in thin films (see Section 35.4).

As Fig. 36.20b shows, in x-ray diffraction there is nearly complete cancellation in all but certain very specific directions in which constructive interference occurs and forms bright spots. Such a pattern is usually called an x-ray *diffraction* pattern, although *interference* pattern might be more appropriate.

We can determine the wavelength of x rays by examining the diffraction pattern for a crystal of known structure and known spacing between atoms, just as we determined wavelengths of visible light by measuring patterns from slits or gratings. (The spacing between atoms in simple crystals of known structure, such as sodium chloride, can be found from the density of the crystal and Avogadro's number.) Then, once we know the x-ray wavelength, we can use x-ray diffraction to explore the structure and determine the spacing between atoms in crystals with unknown structure.

X-ray diffraction is by far the most important experimental tool in the investigation of crystal structure of solids. X-ray diffraction also plays an important role in studies of the structures of liquids and of organic molecules. It has been one of the chief experimental techniques in working out the double-helix structure of DNA (**Fig. 36.24**) and subsequent advances in molecular genetics.

36.24 The British scientist Rosalind Franklin made this groundbreaking x-ray diffraction image of DNA in 1953. The dark bands arranged in a cross provided the first evidence of the helical structure of the DNA molecule.

SOLUTION

EXAMPLE 36.5 X-RAY DIFFRACTION

You direct a beam of 0.154-nm x rays at certain planes of a silicon crystal. As you increase the angle of incidence of the beam from zero, the first strong interference maximum occurs when the beam makes an angle of 34.5° with the planes. (a) How far apart are the planes? (b) Will you find other interference maxima from these planes at greater angles of incidence?

SOLUTION

IDENTIFY and SET UP: This problem involves Bragg reflection of x rays from the planes of a crystal. In part (a) we use the Bragg condition, Eq. (36.16), to find the distance d between adjacent planes from the known wavelength $\lambda = 0.154$ nm and angle of incidence $\theta = 34.5°$ for the $m = 1$ interference maximum. Given the value of d, we use the Bragg condition again in part (b) to find the values of θ for interference maxima corresponding to other values of m.

EXECUTE: (a) We solve Eq. (36.16) for d and set $m = 1$:

$$d = \frac{m\lambda}{2\sin\theta} = \frac{(1)(0.154\text{ nm})}{2\sin 34.5°} = 0.136\text{ nm}$$

This is the distance between adjacent planes.

(b) To calculate other angles, we solve Eq. (36.16) for $\sin\theta$:

$$\sin\theta = \frac{m\lambda}{2d} = m\frac{0.154\text{ nm}}{2(0.136\text{ nm})} = m(0.566)$$

Values of m of 2 or greater give values of $\sin\theta$ greater than unity, which is impossible. Hence there are *no* other angles for interference maxima for this particular set of crystal planes.

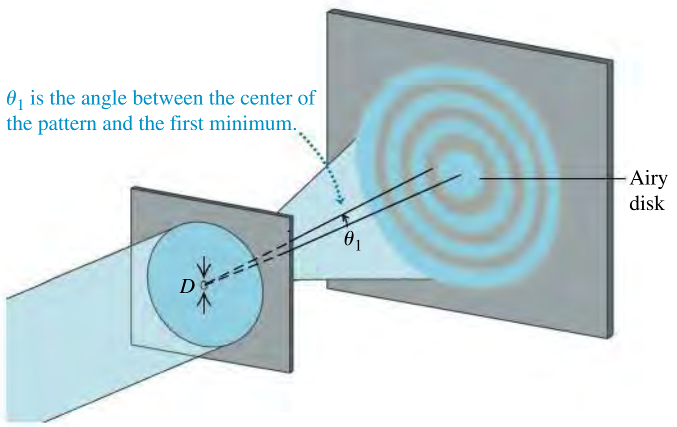
EVALUATE: Our result in part (b) shows that there *would* be a second interference maximum if the quantity $2\lambda/2d = \lambda/d$ were less than 1. This would be the case if the wavelength of the x rays were less than $d = 0.136$ nm. How short would the wavelength need to be to have *three* interference maxima?

TEST YOUR UNDERSTANDING OF SECTION 36.6 You are doing an x-ray diffraction experiment with a crystal in which the atomic planes are 0.200 nm apart. You are using x rays of wavelength 0.0900 nm. What is the highest-order maximum present in the diffraction pattern? (i) Third; (ii) fourth; (iii) fifth; (iv) sixth; (v) seventh.

36.7 CIRCULAR APERTURES AND RESOLVING POWER

We have studied in detail the diffraction patterns formed by long, thin slits or arrays of slits. But an aperture of *any* shape forms a diffraction pattern. The diffraction pattern formed by a *circular* aperture is of special interest because of its role in limiting how well an optical instrument can resolve fine details. In principle, we could compute the intensity at any point P in the diffraction pattern by dividing the area of the aperture into small elements, finding the resulting wave amplitude and phase at P, and then integrating over the aperture area to find the resultant amplitude and intensity at P. In practice, the integration cannot be carried out in terms of elementary functions. We will simply *describe* the pattern and quote a few relevant numbers.

The diffraction pattern formed by a circular aperture consists of a central bright spot surrounded by a series of bright and dark rings, as **Fig. 36.25** shows.



36.25 Diffraction pattern formed by a circular aperture of diameter D. The pattern consists of a central bright spot and alternating dark and bright rings. The angular radius θ_1 of the first dark ring is shown. (This diagram is not drawn to scale.)

We can describe the pattern in terms of the angle θ, representing the angular radius of each ring. The angular radius θ_1 of the first *dark* ring is given by

Diffraction by a circular aperture:

Angular radius of first dark ring = angular radius of Airy disk

$$\sin\theta_1 = 1.22\frac{\overset{\text{Wavelength}}{\lambda}}{\underset{\text{Aperture diameter}}{D}} \tag{36.17}$$

The angular radii of the next two dark rings are given by

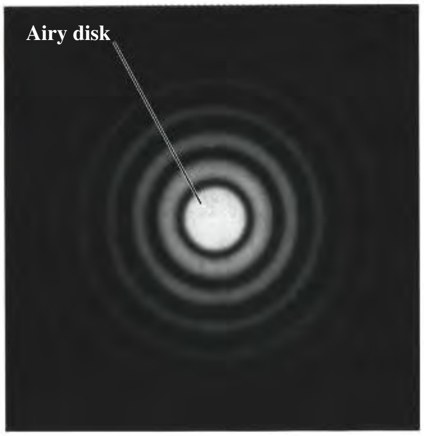
$$\sin\theta_2 = 2.23\frac{\lambda}{D} \qquad \sin\theta_3 = 3.24\frac{\lambda}{D} \tag{36.18}$$

The central bright spot is called the **Airy disk,** in honor of Sir George Airy (1801–1892), who first derived the expression for the intensity in the pattern. The angular radius of the Airy disk is that of the first dark ring, given by Eq. (36.17). The angular radii of the first three *bright* rings outside the Airy disk are

$$\sin\theta = 1.63\frac{\lambda}{D}, \qquad 2.68\frac{\lambda}{D}, \qquad 3.70\frac{\lambda}{D} \tag{36.19}$$

The intensities in the bright rings drop off very quickly with increasing angle. When D is much larger than the wavelength λ, the usual case for optical instruments, the peak intensity in the first ring is only 1.7% of the value at the center of the Airy disk, and the peak intensity of the second ring is only 0.4%. Most (85%) of the light energy falls within the Airy disk. **Figure 36.26** shows a diffraction pattern from a circular aperture.

36.26 Photograph of the diffraction pattern formed by a circular aperture.



Diffraction and Image Formation

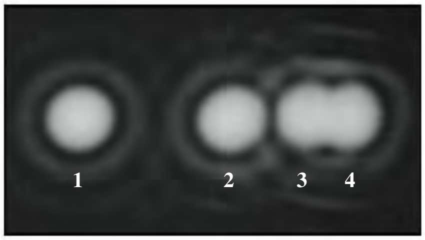
Diffraction has far-reaching implications for image formation by lenses and mirrors. In our study of optical instruments in Chapter 34 we assumed that a lens with focal length f focuses a parallel beam (plane wave) to a *point* at a distance f from the lens. We now see that what we get is *not* a point but the diffraction pattern just described. If we have two point objects, their images are not two points but two diffraction patterns. When the objects are close together, their diffraction patterns overlap; if they are close enough, their patterns overlap almost completely and cannot be distinguished. The effect is shown in **Fig. 36.27**, which presents the patterns for four very small "point" sources of light. In Fig. 36.27a the image of source 1 is well separated from the others, but the images of the sources 3 and 4 have merged. In Fig. 36.27b, with a larger aperture diameter and hence smaller Airy disks, images 3 and 4 are better resolved. In Fig. 36.27c, with a still larger aperture, they are well resolved.

A widely used criterion for resolution of two point objects, proposed by the English physicist Lord Rayleigh (1842–1919) and called **Rayleigh's criterion,** is that the objects are just barely resolved (that is, distinguishable) if the center of one diffraction pattern coincides with the first minimum of the other. In that case the angular separation of the image centers is given by Eq. (36.17). The angular separation of the *objects* is the same as that of the *images* made by a telescope, microscope, or other optical device. So two point objects are barely resolved when their angular separation is given by Eq. (36.17).

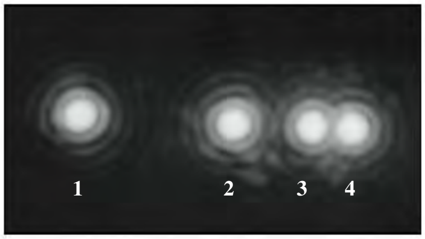
The minimum separation of two objects that can just be resolved by an optical instrument is called the **limit of resolution** of the instrument. The smaller the limit of resolution, the greater the *resolution,* or **resolving power,** of the instrument. Diffraction sets the ultimate limits on resolution of lenses. *Geometric* optics may make it seem that we can make images as large as we like. Eventually, though, we always reach a point at which the image becomes larger but does not gain in detail. The images in Fig. 36.27 would not become sharper if enlarged.

36.27 Diffraction patterns of four very small ("point") sources of light. The photographs were made with a circular aperture in front of the lens. (a) The aperture is so small that the patterns of sources 3 and 4 overlap and are barely resolved by Rayleigh's criterion. Increasing the size of the aperture decreases the size of the diffraction patterns, as shown in (b) and (c).

(a) Small aperture



(b) Medium aperture



(c) Large aperture

Application Bigger Telescope, Better Resolution The large aperture diameter of very large telescopes minimizes diffraction effects. The effective diameter of a telescope can be increased by using arrays of smaller telescopes. The Very Large Array (VLA) in New Mexico is a collection of 27 radio telescopes, each 25 m in diameter, that can be spread out in a Y-shaped arrangement 36 km across. Hence the effective aperture diameter is 36 km, giving the VLA a limit of resolution of 5×10^{-8} rad at a radio wavelength of 1.5 cm. If your eye had this angular resolution, you could read the "20/20" line on an eye chart more than 30 km away!

CAUTION Resolving power vs. chromatic resolving power Don't confuse the resolving power of an optical instrument with the *chromatic* resolving power of a grating (Section 36.5). Resolving power refers to the ability to distinguish the images of objects that appear close to each other, when looking either through an optical instrument or at a photograph made with the instrument. Chromatic resolving power describes how well different wavelengths can be distinguished in a spectrum formed by a diffraction grating.

Rayleigh's criterion combined with Eq. (36.17) shows that resolution (resolving power) improves with larger diameter; it also improves with shorter wavelengths. Ultraviolet microscopes have higher resolution than visible-light microscopes. In electron microscopes the resolution is limited by the wavelengths associated with the electrons, which have wavelike aspects (to be discussed further in Chapter 39). These wavelengths can be made 100,000 times smaller than wavelengths of visible light, with a corresponding gain in resolution. Resolving power also explains the difference in storage capacity between DVDs (introduced in 1995) and Blu-ray discs (introduced in 2003). Information is stored in both of these in a series of tiny pits. In order not to lose information in the scanning process, the scanning optics must be able to resolve two adjacent pits so that they do not seem to blend into a single pit (see sources 3 and 4 in Fig. 36.27). The blue scanning laser used in a Blu-ray player has a shorter wavelength (405 nm) and hence better resolving power than the 650-nm red laser in a DVD player. Hence pits can be spaced closer together in a Blu-ray disc than in a DVD, and more information can be stored on a disc of the same size (50 gigabytes on a Blu-ray disc versus 4.7 gigabytes on a DVD).

EXAMPLE 36.6 RESOLVING POWER OF A CAMERA LENS

SOLUTION

A camera lens with focal length $f = 50$ mm and maximum aperture $f/2$ forms an image of an object 9.0 m away. (a) If the resolution is limited by diffraction, what is the minimum distance between two points on the object that are barely resolved? What is the corresponding distance between image points? (b) How does the situation change if the lens is "stopped down" to $f/16$? Use $\lambda = 500$ nm in both cases.

SOLUTION

IDENTIFY and SET UP: This example uses the ideas about resolving power, image formation by a lens (Section 34.4), and f-number (Section 34.5). From Eq. (34.20), the f-number of a lens is its focal length f divided by the aperture diameter D. We use this equation to determine D and then use Eq. (36.17) (the Rayleigh criterion) to find the angular separation θ between two barely resolved points on the object. We then use the geometry of image formation by a lens to determine the distance y between those points and the distance y' between the corresponding image points.

EXECUTE: (a) The aperture diameter is $D = f/(f\text{-number}) = (50\text{ mm})/2 = 25\text{ mm} = 25 \times 10^{-3}\text{ m}$. From Eq. (36.17) the angular separation θ of two object points that are barely resolved is

$$\theta \approx \sin\theta = 1.22\frac{\lambda}{D} = 1.22\frac{500 \times 10^{-9}\text{ m}}{25 \times 10^{-3}\text{ m}} = 2.4 \times 10^{-5}\text{ rad}$$

We know from our thin-lens analysis in Section 34.4 that, apart from sign, $y/s = y'/s'$ [see Eq. (34.14)]. Thus the angular separations of the object points and the corresponding image points are both equal to θ. Because the object distance s is much greater than the focal length $f = 50$ mm, the image distance s' is approximately equal to f. Thus

$$\frac{y}{9.0\text{ m}} = 2.4 \times 10^{-5} \qquad y = 2.2 \times 10^{-4}\text{ m} = 0.22\text{ mm}$$

$$\frac{y'}{50\text{ mm}} = 2.4 \times 10^{-5} \qquad y' = 1.2 \times 10^{-3}\text{ mm}$$

$$= 0.0012\text{ mm} \approx \tfrac{1}{800}\text{ mm}$$

(b) The aperture diameter is now $(50\text{ mm})/16$, or one-eighth as large as before. The angular separation between barely resolved points is eight times as great, and the values of y and y' are also eight times as great as before:

$$y = 1.8\text{ mm} \qquad y' = 0.0096\text{ mm} = \tfrac{1}{100}\text{ mm}$$

Only the best camera lenses can approach this resolving power.

EVALUATE: Many photographers use the smallest possible aperture for maximum sharpness, since lens aberrations cause light rays that are far from the optic axis to converge to a different image point than do rays near the axis. But as this example shows, diffraction effects become more significant at small apertures. One cause of fuzzy images has to be balanced against another.

TEST YOUR UNDERSTANDING OF SECTION 36.7 You have been asked to compare four proposals for telescopes to be placed in orbit above the blurring effects of the earth's atmosphere. Rank the proposed telescopes in order of their ability to resolve small details, from best to worst. (i) A radio telescope 100 m in diameter observing at a wavelength of 21 cm; (ii) an optical telescope 2.0 m in diameter observing at a wavelength of 500 nm; (iii) an ultraviolet telescope 1.0 m in diameter observing at a wavelength of 100 nm; (iv) an infrared telescope 2.0 m in diameter observing at a wavelength of 10 μm. ▌

36.8 HOLOGRAPHY

Holography is a technique for recording and reproducing an image of an object through the use of interference effects. Unlike the two-dimensional images recorded by an ordinary photograph or television system, a holographic image is truly three-dimensional. Such an image can be viewed from different directions to reveal different sides and from various distances to reveal changing perspective. If you had never seen a hologram, you wouldn't believe it was possible!

Figure 36.28a shows the basic procedure for making a hologram. We illuminate the object to be holographed with monochromatic light, and we place a photographic film so that it is struck by scattered light from the object and also by direct light from the source. In practice, the light source must be a laser, for reasons we will discuss later. Interference between the direct and scattered light forms a complex interference pattern that is recorded on the film.

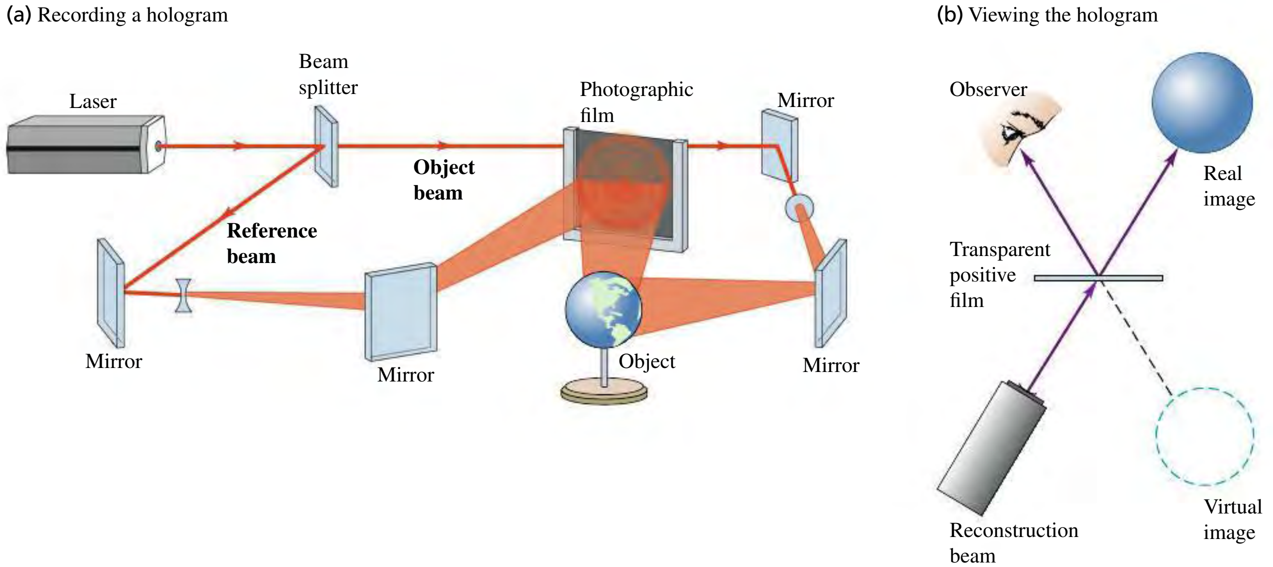
To form the images, we simply project light through the developed film (Fig. 36.28b). Two images are formed: a virtual image on the side of the film nearer the source and a real image on the opposite side.

Holography and Interference Patterns

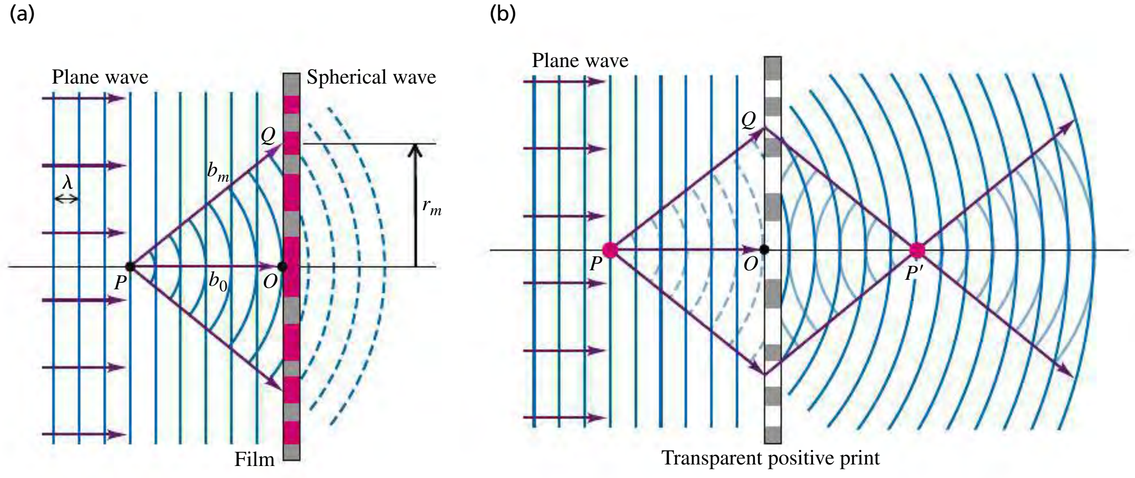
A complete analysis of holography is beyond our scope, but we can gain some insight into the process by looking at how a single point is holographed and imaged. Consider the interference pattern that is formed on a sheet of photographic negative film by the superposition of an incident plane wave and a spherical wave, as shown in **Fig. 36.29a**, next page. The spherical wave originates at a point source P at a distance b_0 from the film; P may in fact be a small object that scatters part of the incident plane wave. We assume that the two waves are monochromatic and coherent and that the phase relationship is such that constructive interference occurs at point O on the diagram. Then constructive interference will *also* occur at any point Q on the film that is farther from P than O is by an integer number of wavelengths. That is, if $b_m - b_0 = m\lambda$, where m is an integer, then constructive interference occurs. The points where this condition is satisfied

BIO Application The Airy Disk in an Eagle's Eye Diffraction by the pupil of an eye limits resolving power. In a human's eye, the maximum pupil diameter D is about 5 mm; in an eagle's eye, D is about 9 mm. From Eq. (36.17) this means that an eagle's eye has superior resolution: A distant point source of light produces an Airy disk on an eagle's retina that is only about 5/9 the angular size of the disk produced on the retina of the human eye. (If our eye produces an image like Fig. 36.27b, an eagle's eye produces one like Fig. 36.27c.) To record the fine details of this high-resolution image, the light-sensitive cones in an eagle's retina are smaller and more closely packed than those in a human retina.

36.28 (a) A hologram is the record on film of the interference pattern formed with light from the coherent source and light scattered from the object. (b) Images are formed when light is projected through the hologram. The observer sees the virtual image formed behind the hologram.



36.29 (a) Constructive interference of the plane and spherical waves occurs in the plane of the film at every point Q for which the distance b_m from P is greater than the distance b_0 from P to O by an integral number of wavelengths $m\lambda$. For the point Q shown, $m = 2$. (b) When a plane wave strikes a transparent positive print of the developed film, the diffracted wave consists of a wave converging to P' and then diverging again and a diverging wave that appears to originate at P. These waves form the real and virtual images, respectively.



form circles on the film centered at O, with radii r_m given by

$$b_m - b_0 = \sqrt{b_0^2 + r_m^2} - b_0 = m\lambda \qquad (m = 1, 2, 3, \ldots) \tag{36.20}$$

Solving this for r_m^2, we find

$$r_m^2 = \lambda(2mb_0 + m^2\lambda)$$

Ordinarily, b_0 is very much larger than λ, so we ignore the second term in parentheses and obtain

$$r_m = \sqrt{2m\lambda b_0} \qquad (m = 1, 2, 3, \ldots) \tag{36.21}$$

The interference pattern consists of a series of concentric bright circular fringes with radii given by Eq. (36.21). Between these bright fringes are dark fringes.

Now we develop the film and make a transparent positive print, so the bright-fringe areas have the greatest transparency on the film. Then we illuminate it with monochromatic plane-wave light of the same wavelength λ that we used initially. In Fig. 36.29b, consider a point P' at a distance b_0 along the axis from the film. The centers of successive bright fringes differ in their distances from P' by an integer number of wavelengths, and therefore a strong *maximum* in the diffracted wave occurs at P'. That is, light converges to P' and then diverges from it on the opposite side. Therefore P' is a *real image* of point P.

This is not the entire diffracted wave, however. The interference of the wavelets that spread out from all the transparent areas forms a second spherical wave that is diverging rather than converging. When this wave is traced back behind the film in Fig. 36.29b, it appears to be spreading out from point P. Thus the total diffracted wave from the hologram is a superposition of a spherical wave converging to form a real image at P' and a spherical wave that diverges as though it had come from the virtual image point P.

Because of the principle of superposition for waves, what is true for the imaging of a single point is also true for the imaging of any number of points. The film records the superposed interference pattern from the various points, and when light is projected through the film, the various image points are reproduced simultaneously. Thus the images of an extended object can be recorded and reproduced just as for a single point object. **Figure 36.30** shows photographs of a holographic image from two different angles, showing the changing perspective in this three-dimensional image.

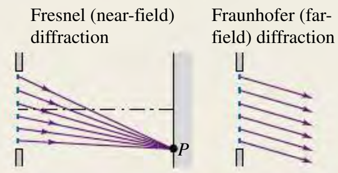
36.30 Two views of the same hologram seen from different angles.

In making a hologram, we have to overcome two practical problems. First, the light used must be *coherent* over distances that are large in comparison to the dimensions of the object and its distance from the film. Ordinary light sources *do not* satisfy this requirement, for reasons that we discussed in Section 35.1. Therefore laser light is essential for making a hologram. (Ordinary white light can be used for *viewing* certain types of hologram, such as those used on credit cards.) Second, extreme mechanical stability is needed. If any relative motion of source, object, or film occurs during exposure, even by as much as a quarter of a wavelength, the interference pattern on the film is blurred enough to prevent satisfactory image formation. These obstacles are not insurmountable, however, and holography has become important in research, entertainment, and a wide variety of technological applications.

CHAPTER 36 SUMMARY

SOLUTIONS TO ALL EXAMPLES

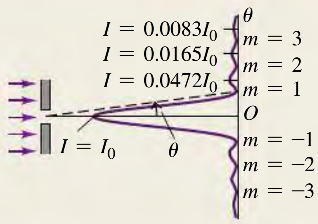
Fresnel and Fraunhofer diffraction: Diffraction occurs when light passes through an aperture or around an edge. When the source and the observer are so far away from the obstructing surface that the outgoing rays can be considered parallel, it is called Fraunhofer diffraction. When the source or the observer is relatively close to the obstructing surface, it is Fresnel diffraction.



Single-slit diffraction: Monochromatic light sent through a narrow slit of width a produces a diffraction pattern on a distant screen. Equation (36.2) gives the condition for destructive interference (a dark fringe) at a point P in the pattern at angle θ. Equation (36.7) gives the intensity in the pattern as a function of θ. (See Examples 36.1–36.3.)

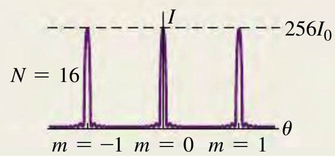
$$\sin\theta = \frac{m\lambda}{a} \quad (m = \pm1, \pm2, \pm3, \ldots) \tag{36.2}$$

$$I = I_0\left\{\frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda}\right\}^2 \tag{36.7}$$



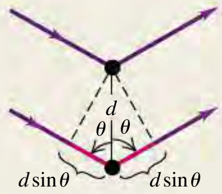
Diffraction gratings: A diffraction grating consists of a large number of thin parallel slits, spaced a distance d apart. The condition for maximum intensity in the interference pattern is the same as for the two-source pattern, but the maxima for the grating are very sharp and narrow. (See Example 36.4.)

$$d\sin\theta = m\lambda \quad (m = 0, \pm1, \pm2, \pm3, \ldots) \tag{36.13}$$



X-ray diffraction: A crystal serves as a three-dimensional diffraction grating for x rays with wavelengths of the same order of magnitude as the spacing between atoms in the crystal. For a set of crystal planes spaced a distance d apart, constructive interference occurs when the angles of incidence and scattering (measured from the crystal planes) are equal and when the Bragg condition [Eq. (36.16)] is satisfied. (See Example 36.5.)

$$2d\sin\theta = m\lambda \quad (m = 1, 2, 3, \ldots) \tag{36.16}$$



Circular apertures and resolving power: The diffraction pattern from a circular aperture of diameter D consists of a central bright spot, called the Airy disk, and a series of concentric dark and bright rings. Equation (36.17) gives the angular radius θ_1 of the first dark ring, equal to the angular size of the Airy disk. Diffraction sets the ultimate limit on resolution (image sharpness) of optical instruments. According to Rayleigh's criterion, two point objects are just barely resolved when their angular separation θ is given by Eq. (36.17). (See Example 36.6.)

$$\sin\theta_1 = 1.22\frac{\lambda}{D} \tag{36.17}$$

BRIDGING PROBLEM OBSERVING THE EXPANDING UNIVERSE

SOLUTION

An astronomer who is studying the light from a galaxy has identified the spectrum of hydrogen but finds that the wavelengths are somewhat shifted from those found in the laboratory. In the lab, the H_α line in the hydrogen spectrum has a wavelength of 656.3 nm. The astronomer is using a transmission diffraction grating having 5758 lines/cm in the first order and finds that the first bright fringe for the H_α line occurs at $\pm 23.41°$ from the central spot. How fast is the galaxy moving? Express your answer in m/s and as a percentage of the speed of light. Is the galaxy moving toward us or away from us?

SOLUTION GUIDE

IDENTIFY and SET UP

1. You can use the information about the grating to find the wavelength of the H_α line in the galaxy's spectrum.
2. In Section 16.8 we learned about the Doppler effect for electromagnetic radiation: The frequency that we receive from a moving source, such as the galaxy, is different from the frequency that is emitted. Equation (16.30) relates the emitted frequency, the received frequency, and the velocity of the source (the target variable). The equation $c = f\lambda$ relates the frequency f and wavelength λ through the speed of light c.

EXECUTE

3. Find the wavelength of the H_α spectral line in the received light.
4. Rewrite Eq. (16.30) as a formula for the velocity v of the galaxy in terms of the received wavelength and the wavelength emitted by the source.
5. Solve for v. Express it in m/s and as a percentage of c, and decide whether the galaxy is moving toward us or moving away.

EVALUATE

6. Is your answer consistent with the relative sizes of the received wavelength and the emitted wavelength?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q36.1 Why can we readily observe diffraction effects for sound waves and water waves, but not for light? Is this because light travels so much faster than these other waves? Explain.

Q36.2 What is the difference between Fresnel and Fraunhofer diffraction? Are they different *physical* processes? Explain.

Q36.3 You use a lens of diameter D and light of wavelength λ and frequency f to form an image of two closely spaced and distant objects. Which of the following will increase the resolving power? (a) Use a lens with a smaller diameter; (b) use light of higher frequency; (c) use light of longer wavelength. In each case justify your answer.

Q36.4 Light of wavelength λ and frequency f passes through a single slit of width a. The diffraction pattern is observed on a screen a distance x from the slit. Which of the following will *decrease* the width of the central maximum? (a) Decrease the slit width; (b) decrease the frequency f of the light; (c) decrease the wavelength λ of the light; (d) decrease the distance x of the screen from the slit. In each case justify your answer.

Q36.5 In a diffraction experiment with waves of wavelength λ, there will be *no* intensity minima (that is, no dark fringes) if the slit width is small enough. What is the maximum slit width for which this occurs? Explain your answer.

Q36.6 An interference pattern is produced by four parallel and equally spaced narrow slits. By drawing appropriate phasor diagrams, explain why there is an interference minimum when the phase difference ϕ from adjacent slits is (a) $\pi/2$; (b) π; (c) $3\pi/2$. In each case, for which pairs of slits is there totally destructive interference?

Q36.7 Phasor Diagram for Eight Slits. An interference pattern is produced by eight equally spaced narrow slits. The caption for Fig. 36.14 claims that minima occur for $\phi = 3\pi/4$, $\pi/4$, $3\pi/2$, and $7\pi/4$. Draw the phasor diagram for each of these four cases, and explain why each diagram proves that there is in fact a minimum. In each case, for which pairs of slits is there totally destructive interference?

Q36.8 A rainbow ordinarily shows a range of colors (see Section 33.4). But if the water droplets that form the rainbow are small enough, the rainbow will appear white. Explain why, using diffraction ideas. How small do you think the raindrops would have to be for this to occur?

Q36.9 Some loudspeaker horns for outdoor concerts (at which the entire audience is seated on the ground) are wider vertically than horizontally. Use diffraction ideas to explain why this is more efficient at spreading the sound uniformly over the audience than either a square speaker horn or a horn that is wider horizontally than vertically. Would this still be the case if the audience were seated at different elevations, as in an amphitheater? Why or why not?

Q36.10 Figure 31.12 (Section 31.2) shows a loudspeaker system. Low-frequency sounds are produced by the *woofer,* which is a speaker with large diameter; the *tweeter,* a speaker with smaller diameter, produces high-frequency sounds. Use diffraction ideas to explain why the tweeter is more effective for distributing high-frequency sounds uniformly over a room than is the woofer.

Q36.11 Information is stored on an audio compact disc, CD-ROM, or DVD disc in a series of pits on the disc. These pits are scanned by a laser beam. An important limitation on the amount of information that can be stored on such a disc is the width of the laser beam. Explain why this should be, and explain how using a shorter-wavelength laser allows more information to be stored on a disc of the same size.

Q36.12 With which color of light can the Hubble Space Telescope see finer detail in a distant astronomical object: red, blue, or ultraviolet? Explain your answer.

Q36.13 At the end of Section 36.4, the following statements were made about an array of N slits. Explain, using phasor diagrams, why each statement is true. (a) A minimum occurs whenever ϕ is an integral multiple of $2\pi/N$, except when ϕ is an integral multiple of 2π (which gives a principal maximum). (b) There are $(N-1)$ minima between each pair of principal maxima.

Q36.14 Could x-ray diffraction effects with crystals be observed by using visible light instead of x rays? Why or why not?

Q36.15 Why is a diffraction grating better than a two-slit setup for measuring wavelengths of light?

Q36.16 One sometimes sees rows of evenly spaced radio antenna towers. A student remarked that these act like diffraction gratings. What did she mean? Why would one want them to act like a diffraction grating?

Q36.17 If a hologram is made using 600-nm light and then viewed with 500-nm light, how will the images look compared to those observed when viewed with 600-nm light? Explain.

Q36.18 A hologram is made using 600-nm light and then viewed by using white light from an incandescent bulb. What will be seen? Explain.

Q36.19 Ordinary photographic film reverses black and white, in the sense that the most brightly illuminated areas become blackest upon development (hence the term *negative*). Suppose a hologram negative is viewed directly, without making a positive transparency. How will the resulting images differ from those obtained with the positive? Explain.

EXERCISES

Section 36.2 Diffraction from a Single Slit

36.1 •• Monochromatic light from a distant source is incident on a slit 0.750 mm wide. On a screen 2.00 m away, the distance from the central maximum of the diffraction pattern to the first minimum is measured to be 1.35 mm. Calculate the wavelength of the light.

36.2 • Parallel rays of green mercury light with a wavelength of 546 nm pass through a slit covering a lens with a focal length of 60.0 cm. In the focal plane of the lens, the distance from the central maximum to the first minimum is 8.65 mm. What is the width of the slit?

36.3 •• Light of wavelength 585 nm falls on a slit 0.0666 mm wide. (a) On a very large and distant screen, how many *totally* dark fringes (indicating complete cancellation) will there be, including both sides of the central bright spot? Solve this problem *without* calculating all the angles! (*Hint:* What is the largest that $\sin\theta$ can be? What does this tell you is the largest that m can be?) (b) At what angle will the dark fringe that is most distant from the central bright fringe occur?

36.4 • Light of wavelength 633 nm from a distant source is incident on a slit 0.750 mm wide, and the resulting diffraction pattern is observed on a screen 3.50 m away. What is the distance between the two dark fringes on either side of the central bright fringe?

36.5 •• Diffraction occurs for all types of waves, including sound waves. High-frequency sound from a distant source with wavelength 9.00 cm passes through a slit 12.0 cm wide. A microphone is placed 8.00 m directly in front of the center of the slit, corresponding to point O in Fig. 36.5a. The microphone is then moved in a direction perpendicular to the line from the center of the slit to point O. At what distances from O will the intensity detected by the microphone be zero?

36.6 • **CP Tsunami!** On December 26, 2004, a violent earthquake of magnitude 9.1 occurred off the coast of Sumatra. This quake triggered a huge tsunami (similar to a tidal wave) that killed more than 150,000 people. Scientists observing the wave on the open ocean measured the time between crests to be 1.0 h and the speed of the wave to be 800 km/h. Computer models of the evolution of this enormous wave showed that it bent around the continents and spread to all the oceans of the earth. When the wave reached the gaps between continents, it diffracted between them as through a slit. (a) What was the wavelength of this tsunami? (b) The distance between the southern tip of Africa and northern Antarctica is about 4500 km, while the distance between the southern end of Australia and Antarctica is about 3700 km. As an approximation, we can model this wave's behavior by using Fraunhofer diffraction. Find the smallest angle away from the central maximum for which the waves would cancel after going through each of these continental gaps.

36.7 •• **CP** A series of parallel linear water wave fronts are traveling directly toward the shore at 15.0 cm/s on an otherwise placid lake. A long concrete barrier that runs parallel to the shore at a distance of 3.20 m away has a hole in it. You count the wave crests and observe that 75.0 of them pass by each minute, and you also observe that no waves reach the shore at ± 61.3 cm from the point directly opposite the hole, but waves do reach the shore everywhere within this distance. (a) How wide is the hole in the barrier? (b) At what other angles do you find no waves hitting the shore?

36.8 • Monochromatic electromagnetic radiation with wavelength λ from a distant source passes through a slit. The diffraction pattern is observed on a screen 2.50 m from the slit. If the width of the central maximum is 6.00 mm, what is the slit width a if the wavelength is (a) 500 nm (visible light); (b) 50.0 μm (infrared radiation); (c) 0.500 nm (x rays)?

36.9 •• **Doorway Diffraction.** Sound of frequency 1250 Hz leaves a room through a 1.00-m-wide doorway (see Exercise 36.5). At which angles relative to the centerline perpendicular to the doorway will someone outside the room hear no sound? Use 344 m/s for the speed of sound in air and assume that the source and listener are both far enough from the doorway for Fraunhofer diffraction to apply. You can ignore effects of reflections.

36.10 • **CP** Light waves, for which the electric field is given by $E_y(x, t) = E_{max} \sin[(1.40 \times 10^7 \text{ m}^{-1})x - \omega t]$, pass through a slit and produce the first dark bands at $\pm 28.6°$ from the center of the diffraction pattern. (a) What is the frequency of this light? (b) How wide is the slit? (c) At which angles will other dark bands occur?

36.11 •• Red light of wavelength 633 nm from a helium–neon laser passes through a slit 0.350 mm wide. The diffraction pattern is observed on a screen 3.00 m away. Define the width of a bright fringe as the distance between the minima on either side. (a) What is the width of the central bright fringe? (b) What is the width of the first bright fringe on either side of the central one?

Section 36.3 Intensity in the Single-Slit Pattern

36.12 •• Public Radio station KXPR-FM in Sacramento broadcasts at 88.9 MHz. The radio waves pass between two tall skyscrapers that are 15.0 m apart along their closest walls. (a) At what horizontal angles, relative to the original direction of the waves, will a distant antenna not receive any signal from this station? (b) If the maximum intensity is 3.50 W/m^2 at the antenna, what is the intensity at $\pm 5.00°$ from the center of the central maximum at the distant antenna?

36.13 •• Monochromatic light of wavelength 580 nm passes through a single slit and the diffraction pattern is observed on a screen. Both the source and screen are far enough from the slit for Fraunhofer diffraction to apply. (a) If the first diffraction minima are at $\pm 90.0°$, so the central maximum completely fills the screen, what is the width of the slit? (b) For the width of the slit as calculated in part (a), what is the ratio of the intensity at $\theta = 45.0°$ to the intensity at $\theta = 0$?

36.14 •• Monochromatic light of wavelength $\lambda = 620$ nm from a distant source passes through a slit 0.450 mm wide. The diffraction pattern is observed on a screen 3.00 m from the slit. In terms of the intensity I_0 at the peak of the central maximum, what is the intensity of the light at the screen the following distances from the center of the central maximum: (a) 1.00 mm; (b) 3.00 mm; (c) 5.00 mm?

36.15 •• A slit 0.240 mm wide is illuminated by parallel light rays of wavelength 540 nm. The diffraction pattern is observed on a screen that is 3.00 m from the slit. The intensity at the center of the central maximum ($\theta = 0°$) is 6.00×10^{-6} W/m^2. (a) What is the distance on the screen from the center of the central maximum to the first minimum? (b) What is the intensity at a point on the screen midway between the center of the central maximum and the first minimum?

36.16 • Monochromatic light of wavelength 592 nm from a distant source passes through a slit that is 0.0290 mm wide. In the resulting diffraction pattern, the intensity at the center of the central maximum ($\theta = 0°$) is 4.00×10^{-5} W/m^2. What is the intensity at a point on the screen that corresponds to $\theta = 1.20°$?

36.17 •• A single-slit diffraction pattern is formed by monochromatic electromagnetic radiation from a distant source passing through a slit 0.105 mm wide. At the point in the pattern 3.25° from the center of the central maximum, the total phase difference between wavelets from the top and bottom of the slit is 56.0 rad. (a) What is the wavelength of the radiation? (b) What is the intensity at this point, if the intensity at the center of the central maximum is I_0?

Section 36.4 Multiple Slits

36.18 • Parallel rays of monochromatic light with wavelength 568 nm illuminate two identical slits and produce an interference pattern on a screen that is 75.0 cm from the slits. The centers of the slits are 0.640 mm apart and the width of each slit is 0.434 mm. If the intensity at the center of the central maximum is 5.00×10^{-4} W/m^2, what is the intensity at a point on the screen that is 0.900 mm from the center of the central maximum?

36.19 • **Number of Fringes in a Diffraction Maximum.** In Fig. 36.12c the central diffraction maximum contains exactly seven interference fringes, and in this case $d/a = 4$. (a) What must the ratio d/a be if the central maximum contains exactly five fringes? (b) In the case considered in part (a), how many fringes are contained within the first diffraction maximum on one side of the central maximum?

36.20 •• **Diffraction and Interference Combined.** Consider the interference pattern produced by two parallel slits of width a and separation d, in which $d = 3a$. The slits are illuminated by normally incident light of wavelength λ. (a) First we ignore diffraction effects due to the slit width. At what angles θ from the central maximum will the next four maxima in the two-slit interference pattern occur? Your answer will be in terms of d and λ. (b) Now we include the effects of diffraction. If the intensity at $\theta = 0°$ is I_0, what is the intensity at each of the angles in part (a)? (c) Which double-slit interference maxima are missing in the pattern? (d) Compare your results to those illustrated in Fig. 36.12c. In what ways are your results different?

36.21 •• An interference pattern is produced by light of wavelength 580 nm from a distant source incident on two identical parallel slits separated by a distance (between centers) of 0.530 mm. (a) If the slits are very narrow, what would be the angular positions of the first-order and second-order, two-slit interference maxima? (b) Let the slits have width 0.320 mm. In terms of the intensity I_0 at the center of the central maximum, what is the intensity at each of the angular positions in part (a)?

36.22 •• Laser light of wavelength 500.0 nm illuminates two identical slits, producing an interference pattern on a screen 90.0 cm from the slits. The bright bands are 1.00 cm apart, and the third bright bands on either side of the central maximum are missing in the pattern. Find the width and the separation of the two slits.

Section 36.5 The Diffraction Grating

36.23 • When laser light of wavelength 632.8 nm passes through a diffraction grating, the first bright spots occur at $\pm 17.8°$ from the central maximum. (a) What is the line density (in lines/cm) of this grating? (b) How many additional bright spots are there beyond the first bright spots, and at what angles do they occur?

36.24 •• Monochromatic light is at normal incidence on a plane transmission grating. The first-order maximum in the interference pattern is at an angle of 11.3°. What is the angular position of the fourth-order maximum?

36.25 • If a diffraction grating produces its third-order bright band at an angle of 78.4° for light of wavelength 681 nm, find (a) the number of slits per centimeter for the grating and (b) the angular location of the first-order and second-order bright bands. (c) Will there be a fourth-order bright band? Explain.

36.26 • If a diffraction grating produces a third-order bright spot for red light (of wavelength 700 nm) at 65.0° from the central maximum, at what angle will the second-order bright spot be for violet light (of wavelength 400 nm)?

36.27 • Visible light passes through a diffraction grating that has 900 slits/cm, and the interference pattern is observed on a screen that is 2.50 m from the grating. (a) Is the angular position of the first-order spectrum small enough for $\sin\theta \approx \theta$ to be a good approximation? (b) In the first-order spectrum, the maxima for two different wavelengths are separated on the screen by 3.00 mm. What is the difference in these wavelengths?

36.28 • The wavelength range of the visible spectrum is approximately 380–750 nm. White light falls at normal incidence on a diffraction grating that has 350 slits/mm. Find the angular width of the visible spectrum in (a) the first order and (b) the third order. (*Note:* An advantage of working in higher orders is the greater angular spread and better resolution. A disadvantage is the overlapping of different orders, as shown in Example 36.4.)

36.29 • (a) What is the wavelength of light that is deviated in the first order through an angle of 13.5° by a transmission grating having 5000 slits/cm? (b) What is the second-order deviation of this wavelength? Assume normal incidence.

36.30 •• **CDs and DVDs as Diffraction Gratings.** A laser beam of wavelength $\lambda = 632.8$ nm shines at normal incidence on the reflective side of a compact disc. (a) The tracks of tiny pits in which information is coded onto the CD are 1.60 μm apart. For what angles of reflection (measured from the normal) will the intensity of light be maximum? (b) On a DVD, the tracks are only 0.740 μm apart. Repeat the calculation of part (a) for the DVD.

36.31 • A typical laboratory diffraction grating has 5.00×10^3 lines/cm, and these lines are contained in a 3.50-cm width of grating. (a) What is the chromatic resolving power of such a grating in the first order? (b) Could this grating resolve the lines of the sodium doublet (see Section 36.5) in the first order? (c) While doing spectral analysis of a star, you are using this grating in the *second* order to resolve spectral lines that are very close to the 587.8002-nm spectral line of iron. (i) For wavelengths longer than the iron line, what is the shortest wavelength you could distinguish from the iron line? (ii) For wavelengths shorter than the iron line, what is the longest wavelength you could distinguish from the iron line? (iii) What is the range of wavelengths you could *not* distinguish from the iron line?

36.32 • **Identifying Isotopes by Spectra.** Different isotopes of the same element emit light at slightly different wavelengths. A wavelength in the emission spectrum of a hydrogen atom is 656.45 nm; for deuterium, the corresponding wavelength is 656.27 nm. (a) What minimum number of slits is required to resolve these two wavelengths in second order? (b) If the grating has 500.00 slits/mm, find the angles and angular separation of these two wavelengths in the second order.

36.33 • The light from an iron arc includes many different wavelengths. Two of these are at $\lambda = 587.9782$ nm and $\lambda = 587.8002$ nm. You wish to resolve these spectral lines in first order using a grating 1.20 cm in length. What minimum number of slits per centimeter must the grating have?

Section 36.6 X-Ray Diffraction

36.34 • If the planes of a crystal are 3.50 Å (1 Å $= 10^{-10}$ m $=$ 1 Ångstrom unit) apart, (a) what wavelength of electromagnetic waves is needed so that the first strong interference maximum in the Bragg reflection occurs when the waves strike the planes at an angle of 22.0°, and in what part of the electromagnetic spectrum do these waves lie? (See Fig. 32.4.) (b) At what other angles will strong interference maxima occur?

36.35 • X rays of wavelength 0.0850 nm are scattered from the atoms of a crystal. The second-order maximum in the Bragg reflection occurs when the angle θ in Fig. 36.22 is 21.5°. What is the spacing between adjacent atomic planes in the crystal?

36.36 • Monochromatic x rays are incident on a crystal for which the spacing of the atomic planes is 0.440 nm. The first-order maximum in the Bragg reflection occurs when the incident and reflected x rays make an angle of 39.4° with the crystal planes. What is the wavelength of the x rays?

Section 36.7 Circular Apertures and Resolving Power

36.37 •• Monochromatic light with wavelength 620 nm passes through a circular aperture with diameter 7.4 μm. The resulting diffraction pattern is observed on a screen that is 4.5 m from the aperture. What is the diameter of the Airy disk on the screen?

36.38 •• Monochromatic light with wavelength 490 nm passes through a circular aperture, and a diffraction pattern is observed on a screen that is 1.20 m from the aperture. If the distance on the screen between the first and second dark rings is 1.65 mm, what is the diameter of the aperture?

36.39 • Two satellites at an altitude of 1200 km are separated by 28 km. If they broadcast 3.6-cm microwaves, what minimum receiving-dish diameter is needed to resolve (by Rayleigh's criterion) the two transmissions?

36.40 • BIO If you can read the bottom row of your doctor's eye chart, your eye has a resolving power of 1 arcminute, equal to $\frac{1}{60}$ degree. If this resolving power is diffraction limited, to what effective diameter of your eye's optical system does this correspond? Use Rayleigh's criterion and assume $\lambda = 550$ nm.

36.41 •• The VLBA (Very Long Baseline Array) uses a number of individual radio telescopes to make one unit having an equivalent diameter of about 8000 km. When this radio telescope is focusing radio waves of wavelength 2.0 cm, what would have to be the diameter of the mirror of a visible-light telescope focusing light of wavelength 550 nm so that the visible-light telescope has the same resolution as the radio telescope?

36.42 •• **Searching for Planets Around Other Stars.** If an optical telescope focusing light of wavelength 550 nm has a perfectly ground mirror, what would the minimum mirror diameter have to be so that the telescope could resolve a Jupiter-size planet around our nearest star, Alpha Centauri, which is about 4.3 light-years from earth? (Consult Appendix F.)

36.43 •• **Hubble Versus Arecibo.** The Hubble Space Telescope has an aperture of 2.4 m and focuses visible light (380–750 nm). The Arecibo radio telescope in Puerto Rico is 305 m (1000 ft) in diameter (it is built in a mountain valley) and focuses radio waves of wavelength 75 cm. (a) Under optimal viewing conditions, what is the smallest crater that each of these telescopes could resolve on our moon? (b) If the Hubble Space Telescope were to be converted to surveillance use, what is the highest orbit above the surface of the earth it could have and still be able to resolve the license plate (not the letters, just the plate) of a car on the ground? Assume optimal viewing conditions, so that the resolution is diffraction limited.

36.44 • **Photography.** A wildlife photographer uses a moderate telephoto lens of focal length 135 mm and maximum aperture $f/4.00$ to photograph a bear that is 11.5 m away. Assume the wavelength is 550 nm. (a) What is the width of the smallest feature on the bear that this lens can resolve if it is opened to its maximum aperture? (b) If, to gain depth of field, the photographer stops the lens down to $f/22.0$, what would be the width of the smallest resolvable feature on the bear?

36.45 • **Observing Jupiter.** You are asked to design a space telescope for earth orbit. When Jupiter is 5.93×10^8 km away (its closest approach to the earth), the telescope is to resolve, by Rayleigh's criterion, features on Jupiter that are 250 km apart. What minimum-diameter mirror is required? Assume a wavelength of 500 nm.

PROBLEMS

36.46 •• Coherent monochromatic light of wavelength λ passes through a narrow slit of width a, and a diffraction pattern is observed on a screen that is a distance x from the slit. On the screen, the width w of the central diffraction maximum is twice the distance x. What is the ratio a/λ of the width of the slit to the wavelength of the light?

36.47 •• BIO **Thickness of Human Hair.** Although we have discussed single-slit diffraction only for a slit, a similar result holds when light bends around a straight, thin object, such as a strand of hair. In that case, a is the width of the strand. From actual laboratory measurements on a human hair, it was found that when a beam of light of wavelength 632.8 nm was shone on a single strand of hair, and the diffracted light was viewed on a screen 1.25 m away, the first dark fringes on either side of the central bright spot were 5.22 cm apart. How thick was this strand of hair?

36.48 •• CP A loudspeaker with a diaphragm that vibrates at 960 Hz is traveling at 80.0 m/s directly toward a pair of holes in a very large wall. The speed of sound in the region is 344 m/s. Far from the wall, you observe that the sound coming through the openings first cancels at $\pm 11.4°$ with respect to the direction in which the speaker is moving. (a) How far apart are the two openings? (b) At what angles would the sound first cancel if the source stopped moving?

36.49 ••• Laser light of wavelength 632.8 nm falls normally on a slit that is 0.0250 mm wide. The transmitted light is viewed on a distant screen where the intensity at the center of the central bright fringe is 8.50 W/m^2. (a) Find the maximum number of totally dark fringes on the screen, assuming the screen is large enough to show them all. (b) At what angle does the dark fringe that is most distant from the center occur? (c) What is the maximum intensity of the bright fringe that occurs immediately before the dark fringe in part (b)? Approximate the angle at which this fringe occurs by assuming it is midway between the angles to the dark fringes on either side of it.

36.50 • **Grating Design.** Your boss asks you to design a diffraction grating that will disperse the first-order visible spectrum through an angular range of 27.0°. (See Example 36.4 in Section 36.5.) (a) What must be the number of slits per centimeter for this grating? (b) At what angles will the first-order visible spectrum begin and end?

36.51 • **Measuring Refractive Index.** A thin slit illuminated by light of frequency f produces its first dark band at $\pm 38.2°$ in air. When the entire apparatus (slit, screen, and space in between) is immersed in an unknown transparent liquid, the slit's first dark bands occur instead at $\pm 21.6°$. Find the refractive index of the liquid.

36.52 •• **Underwater Photography.** An underwater camera has a lens with focal length in air of 35.0 mm and a maximum aperture of $f/2.80$. The film it uses has an emulsion that is sensitive to light of frequency 6.00×10^{14} Hz. If the photographer takes a picture of an object 2.75 m in front of the camera with the lens wide open, what is the width of the smallest resolvable detail on the subject if the object is (a) a fish underwater with the camera in the water and (b) a person on the beach with the camera out of the water?

36.53 ••• CALC The intensity of light in the Fraunhofer diffraction pattern of a single slit is given by Eq. (36.5). Let $\gamma = \beta/2$. (a) Show that the equation for the values of γ at which I is a maximum is $\tan\gamma = \gamma$. (b) Determine the two smallest positive values of γ that are solutions of this equation. (*Hint:* You can use a trial-and-error procedure. Guess a value of γ and adjust your guess to bring $\tan\gamma$ closer to γ. A graphical solution of the equation is very helpful in locating the solutions approximately, to get good initial guesses.) (c) What are the positive values of γ for the first, second, and third minima on one side of the central maximum? Are the γ values in part (b) precisely halfway between the γ values for adjacent minima? (d) If $a = 12\lambda$, what are the angles θ (in degrees) that locate the first minimum, the first maximum beyond the central maximum, and the second minimum?

36.54 •• A slit 0.360 mm wide is illuminated by parallel rays of light that have a wavelength of 540 nm. The diffraction pattern is observed on a screen that is 1.20 m from the slit. The intensity at the center of the central maximum $(\theta = 0°)$ is I_0. (a) What is the distance on the screen from the center of the central maximum to the first minimum? (b) What is the distance on the screen from the center of the central maximum to the point where the intensity has fallen to $I_0/2$?

36.55 •• CP CALC In a large vacuum chamber, monochromatic laser light passes through a narrow slit in a thin aluminum plate and forms a diffraction pattern on a screen that is 0.620 m from the slit. When the aluminum plate has a temperature of 20.0°C, the width of the central maximum in the diffraction pattern is 2.75 mm. What is the change in the width of the central maximum when the temperature of the plate is raised to 520.0°C? Does the width of the central diffraction maximum increase or decrease when the temperature is increased?

36.56 •• CP In a laboratory, light from a particular spectrum line of helium passes through a diffraction grating and the second-order maximum is at 18.9° from the center of the central bright fringe. The same grating is then used for light from a distant galaxy that is moving away from the earth with a speed of 2.65×10^7 m/s. For the light from the galaxy, what is the angular location of the second-order maximum for the same spectral line as was observed in the lab? (See Section 16.8.)

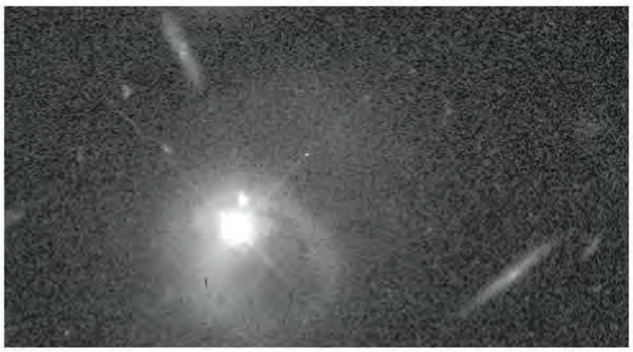
36.57 • What is the longest wavelength that can be observed in the third order for a transmission grating having 9200 slits/cm? Assume normal incidence.

36.58 •• It has been proposed to use an array of infrared telescopes spread over thousands of kilometers of space to observe planets orbiting other stars. Consider such an array that has an effective diameter of 6000 km and observes infrared radiation at a wavelength of 10 μm. If it is used to observe a planet orbiting the star 70 Virginis, which is 59 light-years from our solar system, what is the size of the smallest details that the array might resolve on the planet? How does this compare to the diameter of the planet, which is assumed to be similar to that of Jupiter $(1.40 \times 10^5$ km$)$? (Although the planet of 70 Virginis is thought to be at least 6.6 times more massive than Jupiter, its radius is probably not too different from that of Jupiter. Such large planets are thought to be composed primarily of gases, not rocky material, and hence can be greatly compressed by the mutual gravitational attraction of different parts of the planet.)

36.59 • A diffraction grating has 650 slits/mm. What is the highest order that contains the entire visible spectrum? (The wavelength range of the visible spectrum is approximately 380–750 nm.)

36.60 •• *Quasars,* an abbreviation for *quasi-stellar radio sources,* are distant objects that look like stars through a telescope but that emit far more electromagnetic radiation than an entire normal galaxy of stars. An example is the bright object below and to the left of center in **Fig. P36.60**; the other elongated objects in this image are normal galaxies. The leading model for the structure of a quasar is a galaxy with a supermassive black hole at its center. In this model, the radiation is emitted by interstellar gas and dust within the galaxy as this material falls toward the black hole. The radiation is thought to emanate from a region just a few light-years in diameter. (The diffuse glow surrounding the bright quasar shown in Fig. P36.60 is thought to be this quasar's host galaxy.) To investigate this model of quasars and to study other exotic astronomical objects, the Russian Space Agency plans to place a radio telescope in an orbit that extends to 77,000 km from the earth. When the signals from this telescope are combined with signals from the ground-based telescopes of the VLBA, the resolution will be that of a single radio telescope 77,000 km in diameter. What is the size of the smallest detail that this arrangement could resolve in quasar 3C 405, which is 7.2×10^8 light-years from earth, using radio waves at a frequency of 1665 MHz? (*Hint:* Use Rayleigh's criterion.) Give your answer in light-years and in kilometers.

Figure **P36.60**

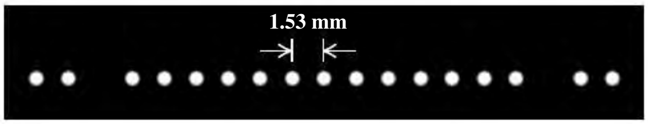


36.61 •• A glass sheet is covered by a very thin opaque coating. In the middle of this sheet there is a thin scratch 0.00125 mm thick. The sheet is totally immersed beneath the surface of a liquid. Parallel rays of monochromatic coherent light with wavelength 612 nm in air strike the sheet perpendicular to its surface and pass through the scratch. A screen is placed in the liquid a distance of 30.0 cm away from the sheet and parallel to it. You observe that the first dark fringes on either side of the central bright fringe on the screen are 22.4 cm apart. What is the refractive index of the liquid?

36.62 •• BIO **Resolution of the Eye.** The maximum resolution of the eye depends on the diameter of the opening of the pupil (a diffraction effect) and the size of the retinal cells. The size of the retinal cells (about 5.0 μm in diameter) limits the size of an object at the near point (25 cm) of the eye to a height of about 50 μm. (To get a reasonable estimate without having to go through complicated calculations, we shall ignore the effect of the fluid in the eye.) (a) Given that the diameter of the human pupil is about 2.0 mm, does the Rayleigh criterion allow us to resolve a 50-μm-tall object at 25 cm from the eye with light of wavelength 550 nm? (b) According to the Rayleigh criterion, what is the shortest object we could resolve at the 25-cm near point with light of wavelength 550 nm? (c) What angle would the object in part (b) subtend at the eye? Express your answer in minutes $(60 \text{ min} = 1°)$, and compare it with the experimental value of about 1 min. (d) Which effect is more important in limiting the resolution of our eyes: diffraction or the size of the retinal cells?

36.63 •• DATA While researching the use of laser pointers, you conduct a diffraction experiment with two thin parallel slits. Your result is the pattern of closely spaced bright and dark fringes shown in **Fig. P36.63**. (Only the central portion of the pattern is shown.) You measure that the bright spots are equally spaced at 1.53 mm center to center (except for the missing spots) on a screen that is 2.50 m from the slits. The light source was a helium–neon laser producing a wavelength of 632.8 nm. (a) How far apart are the two slits? (b) How wide is each one?

Figure **P36.63**



36.64 •• DATA Your physics study partner tells you that the width of the central bright band in a single-slit diffraction pattern is inversely proportional to the width of the slit. This means that the width of the central maximum increases when the width of the slit decreases. The claim seems counterintuitive to you, so you make measurements to test it. You shine monochromatic laser light with wavelength λ onto a very narrow slit of width a and measure the width w of the central maximum in the diffraction pattern that is produced on a screen 1.50 m from the slit. (By "width," you mean the distance on the screen between the two minima on either side of the central maximum.) Your measurements are given in the table.

a (μm)	0.78	0.91	1.04	1.82	3.12	5.20	7.80	10.40	15.60
w (m)	2.68	2.09	1.73	0.89	0.51	0.30	0.20	0.15	0.10

(a) If w is inversely proportional to a, then the product aw is constant, independent of a. For the data in the table, graph aw versus a. Explain why aw is not constant for smaller values of a. (b) Use your graph in part (a) to calculate the wavelength λ of the laser light. (c) What is the angular position of the first minimum in the diffraction pattern for (i) $a = 0.78$ μm and (ii) $a = 15.60$ μm?

36.65 •• DATA At the metal fabrication company where you work, you are asked to measure the diameter D of a very small circular hole in a thin, vertical metal plate. To do so, you pass coherent monochromatic light with wavelength 562 nm through the hole and observe the diffraction pattern on a screen that is a distance x from the hole. You measure the radius r of the first dark ring in the diffraction pattern (see Fig. 36.26). You make the measurements for four values of x. Your results are given in the table.

x (m)	1.00	1.50	2.00	2.50
r (cm)	5.6	8.5	11.6	14.1

(a) Use each set of measurements to calculate D. Because the measurements contain some error, calculate the average of the four values of D and take that to be your reported result. (b) For $x = 1.00$ m, what are the radii of the second and third dark rings in the diffraction pattern?

CHALLENGE PROBLEMS

36.66 ••• Intensity Pattern of N Slits. (a) Consider an arrangement of N slits with a distance d between adjacent slits. The slits emit coherently and in phase at wavelength λ. Show that at a time t, the electric field at a distant point P is

$$\begin{aligned} E_P(t) = {} & E_0\cos(kR - \omega t) + E_0\cos(kR - \omega t + \phi) \\ & + E_0\cos(kR - \omega t + 2\phi) + \cdots \\ & + E_0\cos(kR - \omega t + (N-1)\phi) \end{aligned}$$

where E_0 is the amplitude at P of the electric field due to an individual slit, $\phi = (2\pi d\sin\theta)/\lambda$, θ is the angle of the rays reaching P (as measured from the perpendicular bisector of the slit arrangement), and R is the distance from P to the most distant slit. In this problem, assume that R is much larger than d. (b) To carry out the sum in part (a), it is convenient to use the complex-number relationship $e^{iz} = \cos z + i\sin z$, where $i = \sqrt{-1}$. In this expression, $\cos z$ is the *real part* of the complex number e^{iz}, and $\sin z$ is its *imaginary part*. Show that the electric field $E_P(t)$ is equal to the real part of the complex quantity

$$\sum_{n=0}^{N-1} E_0 e^{i(kR-\omega t+n\phi)}$$

(c) Using the properties of the exponential function that $e^A e^B = e^{(A+B)}$ and $(e^A)^n = e^{nA}$, show that the sum in part (b) can be written as

$$\begin{aligned} & E_0\left(\frac{e^{iN\phi}-1}{e^{i\phi}-1}\right)e^{i(kR-\omega t)} \\ & = E_0\left(\frac{e^{iN\phi/2}-e^{-iN\phi/2}}{e^{i\phi/2}-e^{-i\phi/2}}\right)e^{i[kR-\omega t+(N-1)\phi/2]} \end{aligned}$$

Then, using the relationship $e^{iz} = \cos z + i\sin z$, show that the (real) electric field at point P is

$$E_P(t) = \left[E_0\frac{\sin(N\phi/2)}{\sin(\phi/2)}\right]\cos[kR - \omega t + (N-1)\phi/2]$$

The quantity in the first square brackets in this expression is the amplitude of the electric field at P. (d) Use the result for the electric-field amplitude in part (c) to show that the intensity at an angle θ is

$$I = I_0\left[\frac{\sin(N\phi/2)}{\sin(\phi/2)}\right]^2$$

where I_0 is the maximum intensity for an individual slit. (e) Check the result in part (d) for the case $N = 2$. It will help to recall that $\sin 2A = 2\sin A\cos A$. Explain why your result differs from Eq. (35.10), the expression for the intensity in two-source interference, by a factor of 4. (*Hint:* Is I_0 defined in the same way in both expressions?)

36.67 ••• CALC Intensity Pattern of N Slits, Continued. Part (d) of Challenge Problem 36.66 gives an expression for the intensity in the interference pattern of N identical slits. Use this result to verify the following statements. (a) The maximum intensity in the pattern is $N^2 I_0$. (b) The principal maximum at the center of the pattern extends from $\phi = -2\pi/N$ to $\phi = 2\pi/N$, so its width is inversely proportional to $1/N$. (c) A minimum occurs whenever ϕ is an integral multiple of $2\pi/N$, except when ϕ is an integral multiple of 2π (which gives a principal maximum). (d) There are $(N-1)$ minima between each pair of principal maxima. (e) Halfway between two principal maxima, the intensity can be no greater than I_0; that is, it can be no greater than $1/N^2$ times the intensity at a principal maximum.

36.68 ••• CALC It is possible to calculate the intensity in the single-slit Fraunhofer diffraction pattern *without* using the phasor method of Section 36.3. Let y' represent the position of a point within the slit of width a in Fig. 36.5a, with $y' = 0$ at the center of the slit so that the slit extends from $y' = -a/2$ to $y' = a/2$. We imagine dividing the slit up into infinitesimal strips of width dy', each of which acts as a source of secondary wavelets. (a) The amplitude of the total wave at the point O on the distant screen in Fig. 36.5a is E_0. Explain why the amplitude of the wavelet from each infinitesimal strip within the slit is $E_0(dy'/a)$, so that the electric field of the wavelet a distance x from the infinitesimal strip is $dE = E_0(dy'/a)\sin(kx - \omega t)$. (b) Explain why the wavelet from each strip as detected at point P in Fig. 36.5a can be expressed as

$$dE = E_0\frac{dy'}{a}\sin[k(D - y'\sin\theta) - \omega t]$$

where D is the distance from the center of the slit to point P and $k = 2\pi/\lambda$. (c) By integrating the contributions dE from all parts of the slit, show that the total wave detected at point P is

$$\begin{aligned} E &= E_0\sin(kD - \omega t)\frac{\sin[ka(\sin\theta)/2]}{ka(\sin\theta)/2} \\ &= E_0\sin(kD - \omega t)\frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda} \end{aligned}$$

(The trigonometric identities in Appendix B will be useful.) Show that at $\theta = 0°$, corresponding to point O in Fig. 36.5a, the wave is $E = E_0\sin(kD - \omega t)$ and has amplitude E_0, as stated in part (a). (d) Use the result of part (c) to show that if the intensity at point O is I_0, then the intensity at a point P is given by Eq. (36.7).

PASSAGE PROBLEMS

BRAGG REFLECTION ON A DIFFERENT SCALE. A *colloid* consists of particles of one type of substance dispersed in another substance. Suspensions of electrically charged microspheres (microscopic spheres, such as polystyrene) in a liquid such as water can form a colloidal crystal when the microspheres arrange themselves in a regular repeating pattern under the influence of the electrostatic force. Colloidal crystals can selectively manipulate different wavelengths of visible light. Just as we can study crystalline solids by using Bragg reflection of x rays, we can study colloidal crystals through Bragg scattering of visible light from the regular arrangement of charged microspheres. Because the light is traveling through a liquid when it experiences the path differences that lead to constructive interference, it is the wavelength in the liquid that determines the angles at which Bragg reflections are seen. In one experiment, laser light with a wavelength in vacuum of 650 nm is passed through a sample of charged polystyrene spheres in water. A strong interference maximum is then observed when the incident and reflected beams make an angle of 39° with the colloidal crystal planes.

36.69 Why is visible light, which has much longer wavelengths than x rays do, used for Bragg reflection experiments on colloidal crystals? (a) The microspheres are suspended in a liquid, and it is more difficult for x rays to penetrate liquid than it is for visible light. (b) The irregular spacing of the microspheres allows the longer-wavelength visible light to produce more destructive interference than can x rays. (c) The microspheres are much larger than atoms in a crystalline solid, and in order to get interference maxima at reasonably large angles, the wavelength must be much longer than the size of the individual scatterers. (d) The microspheres are spaced more widely than atoms in a crystalline solid, and in order to get interference maxima at reasonably large angles, the wavelength must be comparable to the spacing between scattering planes.

36.70 What plane spacing in the colloidal crystal could produce the maximum in this experiment? (a) 390 nm; (b) 520 nm; (c) 650 nm; (d) 780 nm.

36.71 When the light is passed through the bottom of the sample container, the interference maximum is observed to be at 41°; when it is passed through the top, the corresponding maximum is at 37°. What is the best explanation for this observation? (a) The microspheres are more tightly packed at the bottom, because they tend to settle in the suspension. (b) The microspheres are more tightly packed at the top, because they tend to float to the top of the suspension. (c) The increased pressure at the bottom makes the microspheres smaller there. (d) The maximum at the bottom corresponds to $m = 2$, whereas the maximum at the top corresponds to $m = 1$.

Answers

Chapter Opening Question ?

(i) For an optical system that uses a lens, the ability to resolve fine details—its resolving power, or resolution—improves as the lens diameter D increases (Section 36.7). Each miniature lens in a fly's eye produces its own image, so these images have very poor resolution, compared to those produced by a human eye, because the lens is so small. However, a fly's eye is much better than a human eye at detecting movement.

Test Your Understanding Questions

36.1 yes When you hear the voice of someone standing around a corner, you are hearing sound waves that underwent diffraction. If there were no diffraction or reflection of sound, you could hear sounds only from objects that were in plain view.

36.2 (ii), (i) and (iv) (tie), (iii) The angle θ of the first dark fringe is given by Eq. (36.2) with $m = 1$, or $\sin\theta = \lambda/a$. The larger the value of the ratio λ/a, the larger the value of $\sin\theta$ and hence the value of θ. The ratio λ/a in each case is
(i) $(400\text{ nm})/(0.20\text{ mm}) = (4.0 \times 10^{-7}\text{ m})/(2.0 \times 10^{-4}\text{ m}) = 2.0 \times 10^{-3}$;
(ii) $(600\text{ nm})/(0.20\text{ mm}) = (6.0 \times 10^{-7}\text{ m})/(2.0 \times 10^{-4}\text{ m}) = 3.0 \times 10^{-3}$;
(iii) $(400\text{ nm})/(0.30\text{ mm}) = (4.0 \times 10^{-7}\text{ m})/(3.0 \times 10^{-4}\text{ m}) = 1.3 \times 10^{-3}$;
(iv) $(600\text{ nm})/(0.30\text{ mm}) = (6.0 \times 10^{-7}\text{ m})/(3.0 \times 10^{-4}\text{ m}) = 2.0 \times 10^{-3}$.

36.3 (ii) and (iii) If the slit width a is less than the wavelength λ, there are no points in the diffraction pattern at which the intensity is zero (see Fig. 36.10a). The slit width is $0.0100\text{ mm} = 1.00 \times 10^{-5}\text{ m}$, so this condition is satisfied for (ii) ($\lambda = 10.6\ \mu\text{m} = 1.06 \times 10^{-5}\text{ m}$) and (iii) ($\lambda = 1.00\text{ mm} = 1.00 \times 10^{-3}\text{ m}$) but not for (i) ($\lambda = 500\text{ nm} = 5.00 \times 10^{-7}\text{ m}$) or (iv) ($\lambda = 50.0\text{ nm} = 5.00 \times 10^{-8}\text{ m}$).

36.4 yes; $m_i = \pm 5, \pm 10, \ldots$ A "missing maximum" satisfies both $d\sin\theta = m_i\lambda$ (the condition for an interference maximum) and $a\sin\theta = m_d\lambda$ (the condition for a diffraction minimum). Substituting $d = 2.5a$, we can combine these two conditions into the relationship $m_i = 2.5m_d$. This is satisfied for $m_i = \pm 5$ and $m_d = \pm 2$ (the fifth interference maximum is missing because it coincides with the second diffraction minimum), $m_i = \pm 10$ and $m_d = \pm 4$ (the tenth interference maximum is missing because it coincides with the fourth diffraction minimum), and so on.

36.5 (i) As described in the text, the resolving power needed is $R = Nm = 1000$. In the first order $(m = 1)$ we need $N = 1000$ slits, but in the fourth order $(m = 4)$ we need only $N = R/m = 1000/4 = 250$ slits. (These numbers are only approximate because of the arbitrary nature of our criterion for resolution and because real gratings always have slight imperfections in the shapes and spacings of the slits.)

36.6 (ii) The angular position of the mth maximum is given by Eq. (36.16), $2d\sin\theta = m\lambda$. This gives $m = (2d\sin\theta)/\lambda$. The sine function can never be greater than 1, so the largest value of m in the pattern can be no greater than $2d/\lambda = 2(0.200\text{ nm})/(0.0900\text{ nm}) = 4.44$. Since m must be an integer, the highest-order maximum in the pattern is $m = 4$ (fourth order). The $m = 5, 6, 7, \ldots$ maxima do not appear.

36.7 (iii), (ii), (iv), (i) Rayleigh's criterion combined with Eq. (36.17) shows that the smaller the value of the ratio λ/D, the better the resolving power of a telescope of diameter D. For the four telescopes, this ratio is equal to (i) $(21\text{ cm})/(100\text{ m}) = (0.21\text{ m})/(100\text{ m}) = 2.1 \times 10^{-3}$; (ii) $(500\text{ nm})/(2.0\text{ m}) = (5.0 \times 10^{-7}\text{ m})/(2.0\text{ m}) = 2.5 \times 10^{-7}$; (iii) $(100\text{ nm})/(1.0\text{ m}) = (1.0 \times 10^{-7}\text{ m})/(1.0\text{ m}) = 1.0 \times 10^{-7}$; (iv) $(10\ \mu\text{m})/(2.0\text{ m}) = (1.0 \times 10^{-5}\text{ m})/(2.0\text{ m}) = 5.0 \times 10^{-6}$.

Bridging Problem

1.501×10^{7} m/s or 5.00% of c; away from us

? At Brookhaven National Laboratory in New York, atomic nuclei are accelerated to 99.995% of the ultimate speed limit of the universe—the speed of light, *c*. Compared to the kinetic energy of a nucleus moving at 99.000% of *c*, the kinetic energy of the same nucleus moving at 99.995% of *c* is about (i) 0.001% greater; (ii) 0.1% greater; (iii) 1% greater; (iv) 2% greater; (v) 16 times greater.

37 RELATIVITY

LEARNING GOALS

Looking forward at …

37.1 The two postulates of Einstein's special theory of relativity, and what motivates these postulates.

37.2 Why different observers can disagree about whether two events are simultaneous.

37.3 How relativity predicts that moving clocks run slow, and what experimental evidence confirms this.

37.4 How the length of an object changes due to the object's motion.

37.5 How the velocity of an object depends on the frame of reference from which it is observed.

37.6 How the frequency of a light wave has different values for different observers.

37.7 How the theory of relativity modifies the relationship between velocity and momentum.

37.8 How to solve problems involving work and kinetic energy for particles moving at relativistic speeds.

37.9 Some of the key concepts of Einstein's general theory of relativity.

Looking back at …

3.4, 3.5 Motion in a circle, relative velocity.

4.2 Inertial frames of reference.

16.8 Doppler effect for sound.

29.1 Electromagnetic induction.

32.2 Maxwell's equations and the speed of light.

35.5 Michelson-Morley experiment.

In 1905, Albert Einstein—then an unknown 25-year-old assistant in the Swiss patent office—published four papers of extraordinary importance. One was an analysis of Brownian motion; a second (for which he was awarded the Nobel Prize) was on the photoelectric effect. In the last two, Einstein introduced his **special theory of relativity,** proposing drastic revisions in the Newtonian concepts of space and time.

Einstein based the special theory of relativity on two postulates. One states that the laws of physics are the same in all inertial frames of reference; the other states that the speed of light in vacuum is the same in all inertial frames. These innocent-sounding propositions have far-reaching implications. Here are three: (1) Events that are simultaneous for one observer may not be simultaneous for another. (2) When two observers moving relative to each other measure a time interval or a length, they may not get the same results. (3) For the conservation principles for momentum and energy to be valid in all inertial systems, Newton's second law and the equations for momentum and kinetic energy have to be revised.

Relativity has important consequences in *all* areas of physics, including electromagnetism, atomic and nuclear physics, and high-energy physics. Although many of the results derived in this chapter may run counter to your intuition, the theory is in solid agreement with experimental observations.

37.1 INVARIANCE OF PHYSICAL LAWS

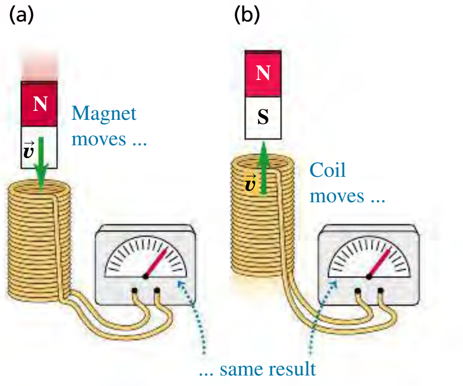
Let's take a look at the two postulates that make up the special theory of relativity. Both postulates describe what is seen by an observer in an *inertial frame of reference,* which we introduced in Section 4.2. The theory is "special" in the sense that it applies to observers in such special reference frames.

Einstein's First Postulate

Einstein's first postulate, called the **principle of relativity,** states: **The laws of physics are the same in every inertial frame of reference.** If the laws differed, that difference could distinguish one inertial frame from the others or make one frame more "correct" than another. As an example, suppose you watch two children playing catch with a ball while the three of you are aboard a train moving with constant velocity. Your observations of the motion *of the ball,* no matter how carefully done, can't tell you how fast (or whether) the train is moving. This is because Newton's laws of motion are the same in every inertial frame.

Another example is the electromotive force (emf) induced in a coil of wire by a nearby moving permanent magnet. In the frame of reference in which the *coil* is stationary (**Fig. 37.1a**), the moving magnet causes a change of magnetic flux through the coil, and this induces an emf. In a different frame of reference in which the *magnet* is stationary (Fig. 37.1b), the motion of the coil through a magnetic field induces the emf. According to the principle of relativity, both of these frames of reference are equally valid. Hence the same emf must be induced in both situations shown in Fig. 37.1. As we saw in Section 29.1, this is indeed the case, so Faraday's law is consistent with the principle of relativity. Indeed, *all* of the laws of electromagnetism are the same in every inertial frame of reference.

37.1 The same emf is induced in the coil whether (a) the magnet moves relative to the coil or (b) the coil moves relative to the magnet.



Equally significant is the prediction of the speed of electromagnetic radiation, derived from Maxwell's equations (see Section 32.2). According to this analysis, light and all other electromagnetic waves travel in vacuum with a constant speed, now defined to equal exactly 299,792,458 m/s. (We often use the approximate value $c = 3.00 \times 10^8$ m/s.) As we will see, the speed of light in vacuum plays a central role in the theory of relativity.

Einstein's Second Postulate

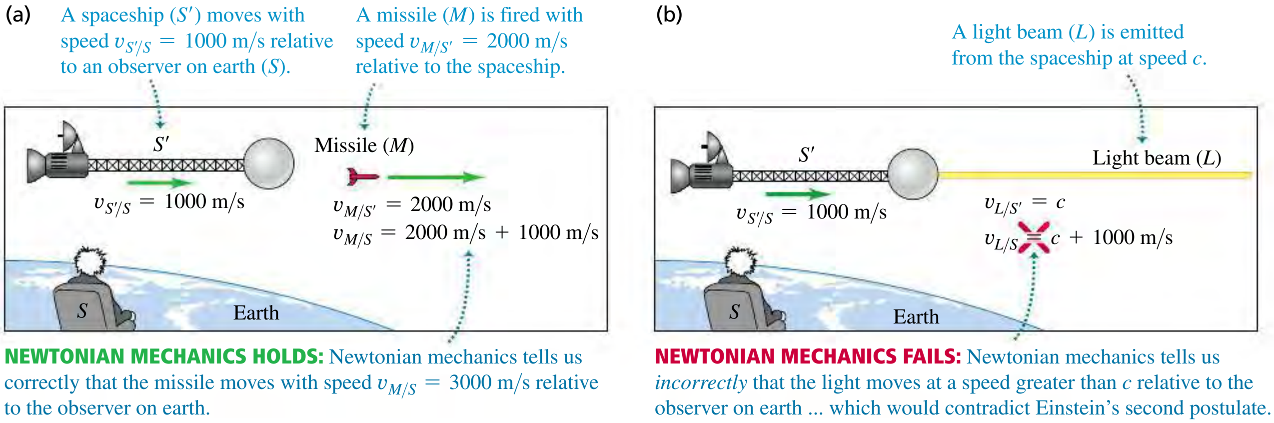
During the 19th century, most physicists believed that light traveled through a hypothetical medium called the *ether,* just as sound waves travel through air. If so, the speed of light measured by observers would depend on their motion relative to the ether and would therefore be different in different directions. The Michelson-Morley experiment, described in Section 35.5, was an effort to detect motion of the earth relative to the ether.

Einstein's conceptual leap was to recognize that if Maxwell's equations are valid in all inertial frames, then the speed of light in vacuum should also be the same in all frames and in all directions. In fact, Michelson and Morley detected *no* ether motion across the earth, and the ether concept has been discarded. Although Einstein may not have known about this negative result, it supported his bold hypothesis. We call this **Einstein's second postulate: The speed of light in vacuum is the same in all inertial frames of reference and is independent of the motion of the source.**

Let's think about what this means. Suppose two observers measure the speed of light in vacuum. One is at rest with respect to the light source, and the other is moving away from it. Both are in inertial frames of reference. According to the principle of relativity, the two observers must obtain the same result, despite the fact that one is moving with respect to the other.

If this seems too easy, consider the following situation. A spacecraft moving past the earth at 1000 m/s fires a missile straight ahead with a speed of 2000 m/s (relative to the spacecraft) (**Fig. 37.2**, next page). What is the missile's speed relative to the earth? Simple, you say; this is an elementary problem in relative velocity (see Section 3.5). The correct answer, according to Newtonian mechanics, is 3000 m/s. But now suppose the spacecraft turns on a searchlight, pointing in the same direction in which the missile was fired. An observer on the spacecraft measures the speed of light emitted by the searchlight and obtains the value c. According to Einstein's second postulate, the motion of the light after it has left the source cannot depend on the motion of the source. So the observer on earth who measures the speed of this same light must also obtain the value c, *not* $c + 1000$ m/s. This result contradicts our elementary notion of relative velocities, and it may not appear to agree with common sense. But "common sense" is intuition based on everyday experience, and this does not usually include measurements of the speed of light.

37.2 (a) Newtonian mechanics makes correct predictions about relatively slow-moving objects; (b) it makes incorrect predictions about the behavior of light.



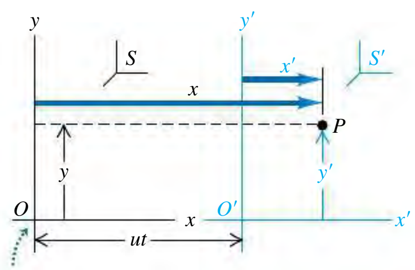
The Ultimate Speed Limit

Einstein's second postulate immediately implies the following result: **It is impossible for an inertial observer to travel at c, the speed of light in vacuum.**

We can prove this by showing that travel at c implies a logical contradiction. Suppose that the spacecraft S' in Fig. 37.2b is moving at the speed of light relative to an observer on the earth, so that $v_{S'/S} = c$. If the spacecraft turns on a headlight, the second postulate now asserts that the earth observer S measures the headlight beam to be also moving at c. Thus this observer measures that the headlight beam and the spacecraft move together and are always at the same point in space. But Einstein's second postulate also asserts that the headlight beam moves at a speed c relative to the spacecraft, so they *cannot* be at the same point in space. This contradictory result can be avoided only if it is impossible for an inertial observer, such as a passenger on the spacecraft, to move at c. As we go through our discussion of relativity, you may find yourself asking the question Einstein asked himself as a 16-year-old student, "What would I see if I were traveling at the speed of light?" Einstein realized only years later that his question's basic flaw was that he could *not* travel at c.

The Galilean Coordinate Transformation

37.3 The position of particle P can be described by the coordinates x and y in frame of reference S or by x' and y' in frame S'.



Let's restate this argument symbolically, using two inertial frames of reference, labeled S for the observer on earth and S' for the moving spacecraft, as shown in **Fig. 37.3**. To keep things as simple as possible, we have omitted the z-axes. The x-axes of the two frames lie along the same line, but the origin O' of frame S' moves relative to the origin O of frame S with constant velocity u along the common x-x'-axis. We on earth set our clocks so that the two origins coincide at time $t = 0$, so their separation at a later time t is ut.

CAUTION Choose your inertial frame coordinates wisely Many of the equations derived in this chapter are true *only* if you define your inertial reference frames as stated in the preceding paragraph. The positive x-direction must be the direction in which the origin O' moves relative to the origin O. In Fig. 37.3 this direction is to the right; if instead O' moves to the left relative to O, you must define the positive x-direction to be to the left.

Now think about how we describe the motion of a particle P. This might be an exploratory vehicle launched from the spacecraft or a pulse of light from a laser. We can describe the *position* of this particle by using the earth coordinates

(x, y, z) in S or the spacecraft coordinates (x', y', z') in S'. Figure 37.3 shows that these are simply related by

$$x = x' + ut \qquad y = y' \qquad z = z' \qquad \text{(Galilean coordinate transformation)} \tag{37.1}$$

These equations, based on the familiar Newtonian notions of space and time, are called the **Galilean coordinate transformation.**

If particle P moves in the x-direction, its instantaneous velocity v_x as measured by an observer stationary in S is $v_x = dx/dt$. Its velocity v'_x as measured by an observer stationary in S' is $v'_x = dx'/dt$. We can derive a relationship between v_x and v'_x by taking the derivative with respect to t of the first of Eqs. (37.1):

$$\frac{dx}{dt} = \frac{dx'}{dt} + u$$

Now dx/dt is the velocity v_x measured in S, and dx'/dt is the velocity v'_x measured in S', so we get the *Galilean velocity transformation* for one-dimensional motion:

$$v_x = v'_x + u \qquad \text{(Galilean velocity transformation)} \tag{37.2}$$

Although the notation differs, this result agrees with our discussion of relative velocities in Section 3.5.

Now here's the fundamental problem. Applied to the speed of light in vacuum, Eq. (37.2) says that $c = c' + u$. Einstein's second postulate, supported subsequently by a wealth of experimental evidence, says that $c = c'$. This is a genuine inconsistency, not an illusion, and it demands resolution. If we accept this postulate, we are forced to conclude that Eqs. (37.1) and (37.2) *cannot* be precisely correct, despite our convincing derivation. These equations have to be modified to bring them into harmony with this principle.

The resolution involves some very fundamental modifications in our kinematic concepts. The first idea to be changed is the seemingly obvious assumption that the observers in frames S and S' use the same *time scale*, formally stated as $t = t'$. Alas, we are about to show that this everyday assumption cannot be correct; the two observers *must* have different time scales. We must define the velocity v' in frame S' as $v' = dx'/dt'$, not as dx'/dt; the two quantities are not the same. The difficulty lies in the concept of *simultaneity*, which is our next topic. A careful analysis of simultaneity will help us develop the appropriate modifications of our notions about space and time.

TEST YOUR UNDERSTANDING OF SECTION 37.1 As a high-speed spaceship flies past you, it fires a strobe light that sends out a pulse of light in all directions. An observer aboard the spaceship measures a spherical wave front that spreads away from the spaceship with the same speed c in all directions. (a) What is the shape of the wave front that *you* measure? (i) Spherical; (ii) ellipsoidal, with the longest axis of the ellipsoid along the direction of the spaceship's motion; (iii) ellipsoidal, with the shortest axis of the ellipsoid along the direction of the spaceship's motion; (iv) not enough information is given to decide. (b) As measured by you, does the wave front remain centered on the spaceship?

37.2 RELATIVITY OF SIMULTANEITY

Measuring times and time intervals involves the concept of **simultaneity.** In a given frame of reference, an **event** is an occurrence that has a definite position and time (**Fig. 37.4**). When you say that you awoke at seven o'clock, you mean that two events (your awakening and your clock showing 7:00) occurred *simultaneously*. The fundamental problem in measuring time intervals is this: In general, two events that are simultaneous in one frame of reference are *not* simultaneous in a second frame that is moving relative to the first, even if both are inertial frames.

37.4 An event has a definite position and time—for instance, on the pavement directly below the center of the Eiffel Tower at midnight on New Year's Eve.

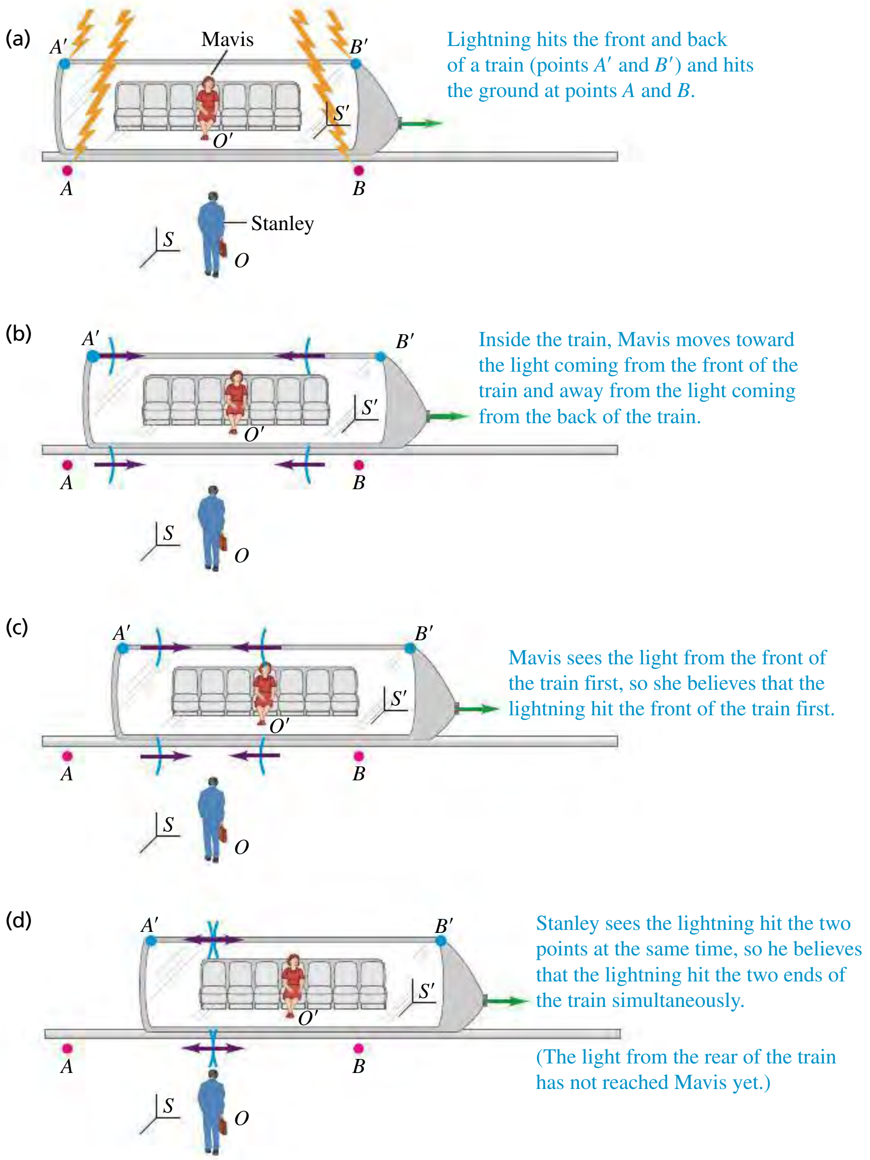
A Thought Experiment in Simultaneity

This may seem to be contrary to common sense. To illustrate the point, here is a version of one of Einstein's *thought experiments*—mental experiments that follow concepts to their logical conclusions. Imagine a train moving with a speed comparable to c, with uniform velocity (**Fig. 37.5**). Two lightning bolts strike a passenger car, one near each end. Each bolt leaves a mark on the car and one on the ground at the instant the bolt hits. The points on the ground are labeled A and B in the figure, and the corresponding points on the car are A' and B'. Stanley is stationary on the ground at O, midway between A and B. Mavis is moving with the train at O' in the middle of the passenger car, midway between A' and B'. Both Stanley and Mavis see both light flashes emitted from the points where the lightning strikes.

Suppose the two wave fronts from the lightning strikes reach Stanley at O simultaneously. He knows that he is the same distance from B and A, so Stanley concludes that the two bolts struck B and A simultaneously. Mavis agrees that the two wave fronts reached Stanley at the same time, but she disagrees that the flashes were emitted simultaneously.

Stanley and Mavis agree that the two wave fronts do not reach Mavis at the same time. Mavis at O' is moving to the right with the train, so she runs into

37.5 A thought experiment in simultaneity.



the wave front from B' *before* the wave front from A' catches up to her. However, because she is in the middle of the passenger car equidistant from A' and B', her observation is that both wave fronts took the same time to reach her because both moved the same distance at the same speed c. (Recall that the speed of each wave front with respect to *either* observer is c.) Thus she concludes that the lightning bolt at B' struck *before* the one at A'. Stanley at O measures the two events to be simultaneous, but Mavis at O' does not! *Whether or not two events at different x-axis locations are simultaneous depends on the state of motion of the observer.*

You may want to argue that in this example the lightning bolts really *are* simultaneous and that if Mavis at O' could communicate with the distant points without the time delay caused by the finite speed of light, she would realize this. But that would be erroneous; the finite speed of information transmission is not the real issue. If O' is midway between A' and B', then in her frame of reference the time for a signal to travel from A' to O' is the same as that from B' to O'. Two signals arrive simultaneously at O' only if they were emitted simultaneously at A' and B'. In this example they *do not* arrive simultaneously at O', and so Mavis must conclude that the events at A' and B' were *not* simultaneous.

Furthermore, there is no basis for saying that Stanley is right and Mavis is wrong, or vice versa. According to the principle of relativity, no inertial frame of reference is more correct than any other in the formulation of physical laws. Each observer is correct *in his or her own frame of reference*. In other words, simultaneity is not an absolute concept. Whether two events are simultaneous depends on the frame of reference. As we mentioned at the beginning of this section, simultaneity plays an essential role in measuring time intervals. It follows that *the time interval between two events may be different in different frames of reference*. So our next task is to learn how to compare time intervals in different frames of reference.

TEST YOUR UNDERSTANDING OF SECTION 37.2 Stanley, who works for the rail system shown in Fig. 37.5, has carefully synchronized the clocks at all of the rail stations. At the moment that Stanley measures all of the clocks striking noon, Mavis is on a high-speed passenger car traveling from Ogdenville toward North Haverbrook. According to Mavis, when the Ogdenville clock strikes noon, what time is it in North Haverbrook? (i) Noon; (ii) before noon; (iii) after noon.

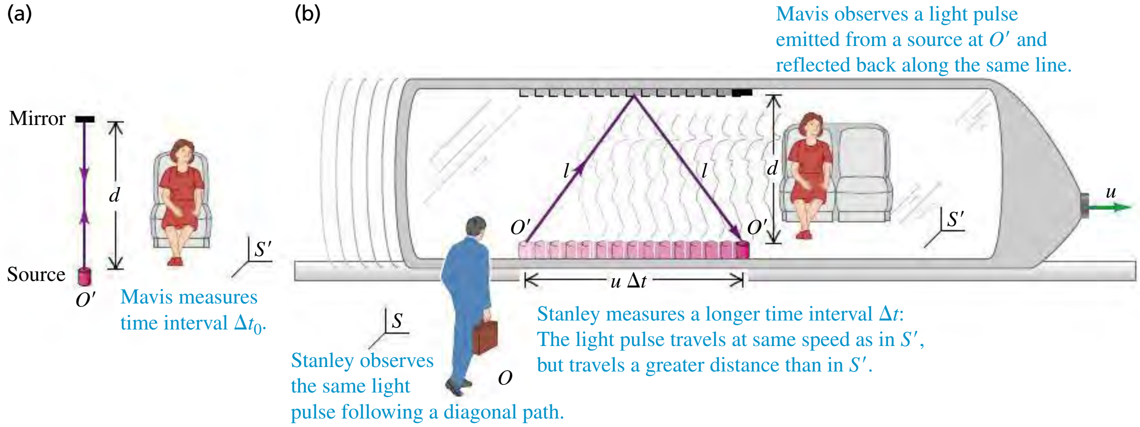
37.3 RELATIVITY OF TIME INTERVALS

We can derive a quantitative relationship between time intervals in different coordinate systems. To do this, let's consider another thought experiment. As before, a frame of reference S' moves along the common x-x'-axis with constant speed u relative to a frame S. As discussed in Section 37.1, u must be less than the speed of light c. Mavis, who is riding along with frame S', measures the time interval between two events that occur at the *same* point in space. Event 1 is when a flash of light from a light source leaves O'. Event 2 is when the flash returns to O', having been reflected from a mirror a distance d away, as shown in **Fig. 37.6a** (next page). We label the time interval Δt_0, using the subscript zero as a reminder that the apparatus is at rest, with zero velocity, in frame S'. The flash of light moves a total distance $2d$, so the time interval is

$$\Delta t_0 = \frac{2d}{c} \tag{37.3}$$

The round-trip time measured by Stanley in frame S is a different interval Δt; in his frame of reference the two events occur at *different* points in space.

37.6 (a) Mavis, in frame of reference S', observes a light pulse emitted from a source at O' and reflected back along the same line. (b) How Stanley (in frame of reference S) and Mavis observe the same light pulse. The positions of O' at the times of departure and return of the pulse are shown.



During the time Δt, the source moves relative to S a distance $u\ \Delta t$ (Fig. 37.6b). In S' the round-trip distance is $2d$ perpendicular to the relative velocity, but the round-trip distance in S is the longer distance $2l$, where

$$l = \sqrt{d^2 + \left(\frac{u\ \Delta t}{2}\right)^2}$$

In writing this expression, we have assumed that both observers measure the same distance d. We will justify this assumption in the next section. The speed of light is the same for both observers, so the round-trip time measured in S is

$$\Delta t = \frac{2l}{c} = \frac{2}{c}\sqrt{d^2 + \left(\frac{u\ \Delta t}{2}\right)^2} \tag{37.4}$$

We would like to have a relationship between Δt and Δt_0 that is independent of d. To get this, we solve Eq. (37.3) for d and substitute the result into Eq. (37.4):

$$\Delta t = \frac{2}{c}\sqrt{\left(\frac{c\ \Delta t_0}{2}\right)^2 + \left(\frac{u\ \Delta t}{2}\right)^2} \tag{37.5}$$

Now we square this and solve for Δt; the result is

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}}$$

Since the quantity $\sqrt{1 - u^2/c^2}$ is less than 1, Δt is greater than Δt_0: Thus Stanley measures a *longer* round-trip time for the light pulse than does Mavis.

Time Dilation and Proper Time

We may generalize this important result. Suppose that in a particular frame of reference, two events occur at the same point in space. If these events are two ticks of a clock, then this is the frame of reference at which the clock is at rest. We call this the *rest frame* of the clock. There is only one frame of reference in which a clock is at rest, and there are infinitely many in which it is moving. Therefore the time interval measured between two events (such as two ticks of the clock) that occur at the same point in a particular frame is a more fundamental quantity than the interval between events at different points. We use the term **proper time** to describe the time interval between two events that occur *at the same point*.

Let Δt_0 be the proper time between the two events—that is, the time as measured by an observer at rest in the frame in which the events occur at the same point. Then our above result says that an observer in a second frame moving with constant speed u relative to the rest frame will measure the time interval to be Δt, where

Proper time between two events (measured in rest frame)

Time dilation: $$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} \qquad (37.6)$$

Speed of light in vacuum

Speed of second frame relative to rest frame

Time interval between same events measured in second frame of reference

We recall that no inertial observer can travel at $u = c$ and we note that $\sqrt{1 - u^2/c^2}$ is imaginary for $u > c$. Thus Eq. (37.6) gives sensible results only when $u < c$. The denominator of Eq. (37.6) is always smaller than 1, so Δt is always *larger* than Δt_0. Thus we call this effect **time dilation.**

Think of an old-fashioned pendulum clock that has one second between ticks, as measured by Mavis in the clock's rest frame; this is Δt_0. If the clock's rest frame is moving relative to Stanley, he measures a time between ticks Δt that is longer than one second. In brief, *observers measure any clock to run slow if it moves relative to them* (**Fig. 37.7**). Note that this conclusion is a direct result of the fact that the speed of light in vacuum is the same in both frames of reference.

The quantity $1/\sqrt{1 - u^2/c^2}$ in Eq. (37.6) is called the **Lorentz factor.** It appears often in relativity and is denoted by the symbol γ (the Greek letter gamma):

Lorentz factor $$\gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \qquad (37.7)$$

Speed of light in vacuum

Speed of one frame of reference relative to another

In terms of this symbol, we can express the time dilation formula, Eq. (37.6), as

Proper time between two events (measured in rest frame)

Time dilation: $$\Delta t = \gamma \Delta t_0 \qquad (37.8)$$

Lorentz factor relating the two frames

Time interval between same events measured in second frame of reference

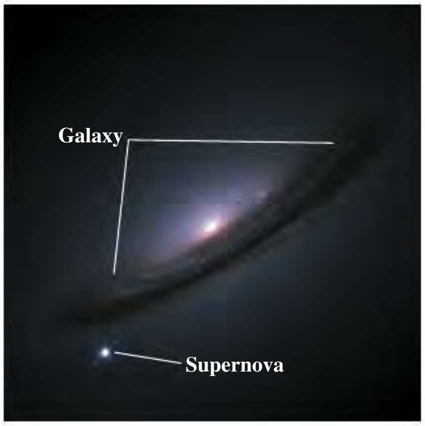
As a further simplification, u/c is sometimes given the symbol β (the Greek letter beta); then $\gamma = 1/\sqrt{1 - \beta^2}$.

Figure 37.8 shows a graph of γ as a function of the relative speed u of two frames of reference. When u is very small compared to c, u^2/c^2 is much smaller than 1 and γ is very nearly *equal* to 1. In that limit, Eqs. (37.6) and (37.8) approach the Newtonian relationship $\Delta t = \Delta t_0$, corresponding to the same time interval in all frames of reference.

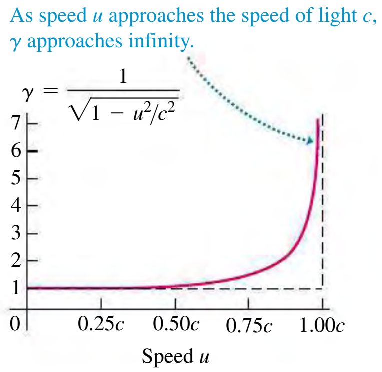
If the relative speed u is great enough that γ is appreciably greater than 1, the speed is said to be *relativistic;* if the difference between γ and 1 is negligibly small, the speed u is called *nonrelativistic*. Thus $u = 6.00 \times 10^7$ m/s $= 0.200c$ (for which $\gamma = 1.02$) is a relativistic speed, but $u = 6.00 \times 10^4$ m/s $= 0.000200c$ (for which $\gamma = 1.00000002$) is a nonrelativistic speed.

CAUTION Measuring time intervals It is important to note that the time interval Δt in Eq. (37.6) involves events that occur *at different space points* in the frame of reference S. Note also that any differences between Δt and the proper time Δt_0 are *not* caused by differences in the times required for light to travel from those space points to an observer at rest in S. We assume that our observer can correct for differences in light transit times, just as an astronomer who's observing the sun understands that an event seen now on earth actually occurred 500 s ago on the sun's surface. Alternatively, we can use *two* observers, one stationary at the location of the first event and the other at the second, each with his or her own clock. We can synchronize these two clocks without difficulty, as long as they are at rest in the same frame of reference. For example, we could send a light pulse simultaneously to the two clocks from a point midway between them. When the pulses arrive, the observers set their clocks to a prearranged time. (But clocks that are synchronized in one frame of reference *are not* in general synchronized in any other frame.)

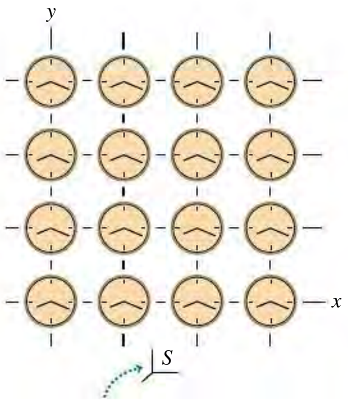
37.7 This image shows an exploding star, called a *supernova,* within a distant galaxy. The brightness of a typical supernova decays at a certain rate. But supernovae that are moving away from us at a substantial fraction of the speed of light decay more slowly, in accordance with Eq. (37.6). The decaying supernova is a moving "clock" that runs slow.



37.8 The Lorentz factor $\gamma = 1/\sqrt{1 - u^2/c^2}$ as a function of the relative speed u of two frames of reference.



37.9 A frame of reference pictured as a coordinate system with a grid of synchronized clocks.



In thought experiments, it's often helpful to imagine many observers with synchronized clocks at rest at various points in a particular frame of reference. We can picture a frame of reference as a coordinate grid with lots of synchronized clocks distributed around it, as suggested by **Fig. 37.9**. Only when a clock is moving relative to a given frame of reference do we have to watch for ambiguities of synchronization or simultaneity.

Throughout this chapter we will frequently use phrases like "Stanley *observes* that Mavis passes the point $x = 5.00$ m, $y = 0$, $z = 0$ at time 2.00 s." This means that Stanley is using a grid of clocks in his frame of reference, like the grid shown in Fig. 37.9, to record the time of an event. We could restate the phrase as "When Mavis passes the point at $x = 5.00$ m, $y = 0$, $z = 0$, the clock at that location in Stanley's frame of reference reads 2.00 s." We will avoid using phrases like "Stanley *sees* that Mavis is at a certain point at a certain time," because there is a time delay for light to travel to Stanley's eye from the position of an event.

PROBLEM-SOLVING STRATEGY 37.1 TIME DILATION

IDENTIFY *the relevant concepts:* The concept of time dilation is used whenever we compare the time intervals between events as measured by observers in different inertial frames of reference.

SET UP *the problem* using the following steps:
1. First decide what two events define the beginning and the end of the time interval. Then identify the two frames of reference in which the time interval is measured.
2. Identify the target variable.

EXECUTE *the solution* as follows:
1. In many problems, the time interval as measured in one frame of reference is the *proper* time Δt_0. This is the time interval between two events in a frame of reference in which the two events occur at the same point in space. In a second frame of reference that has a speed u relative to that first frame, there is a longer time interval Δt between the same two events. In this second frame the two events occur at different points. You will need to decide in which frame the time interval is Δt_0 and in which frame it is Δt.
2. Use Eq. (37.6) or (37.8) to relate Δt_0 and Δt, and then solve for the target variable.

EVALUATE *your answer:* Note that Δt is never smaller than Δt_0, and u is never greater than c. If your results suggest otherwise, you need to rethink your calculation.

EXAMPLE 37.1 TIME DILATION AT 0.990*c*

SOLUTION

High-energy subatomic particles coming from space interact with atoms in the earth's upper atmosphere, in some cases producing unstable particles called *muons*. A muon decays into other particles with a mean lifetime of $2.20\ \mu s = 2.20 \times 10^{-6}$ s as measured in a reference frame in which it is at rest. If a muon is moving at $0.990c$ relative to the earth, what will an observer on earth measure its mean lifetime to be?

SOLUTION

IDENTIFY and SET UP: The muon's lifetime is the time interval between two events: the production of the muon and its subsequent decay. Our target variable is the lifetime in your frame of reference on earth, which we call frame S. We are given the lifetime in a frame S' in which the muon is at rest; this is its *proper* lifetime, $\Delta t_0 = 2.20\ \mu s$. The relative speed of these two frames is $u = 0.990c$. We use Eq. (37.6) to relate the lifetimes in the two frames.

EXECUTE: The muon moves relative to the earth between the two events, so the two events occur at different positions as measured in S and the time interval in that frame is Δt (the target variable). From Eq. (37.6),

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} = \frac{2.20\ \mu s}{\sqrt{1 - (0.990)^2}} = 15.6\ \mu s$$

EVALUATE: Our result predicts that the mean lifetime of the muon in the earth frame (Δt) is about seven times longer than in the muon's frame (Δt_0). This prediction has been verified experimentally; indeed, this was the first experimental confirmation of the time dilation formula, Eq. (37.6).

SOLUTION

EXAMPLE 37.2 TIME DILATION AT AIRLINER SPEEDS

An airplane flies from San Francisco to New York (about 4800 km, or 4.80×10^6 m) at a steady speed of 300 m/s (about 670 mi/h). How much time does the trip take, as measured by an observer on the ground? By an observer in the plane?

SOLUTION

IDENTIFY and SET UP: Here we're interested in the time interval between the airplane departing from San Francisco and landing in New York. The target variables are the time intervals as measured in the frame of reference of the ground S and in the frame of reference of the airplane S'.

EXECUTE: As measured in S the two events occur at different positions (San Francisco and New York), so the time interval measured by ground observers corresponds to Δt in Eq. (37.6). To find it, we simply divide the distance by the speed $u = 300$ m/s:

$$\Delta t = \frac{4.80 \times 10^6 \text{ m}}{300 \text{ m/s}} = 1.60 \times 10^4 \text{ s} \quad (\text{about } 4\tfrac{1}{2} \text{ hours})$$

In the airplane's frame S', San Francisco and New York passing under the plane occur at the same point (the position of the plane). Hence the time interval in the airplane is a proper time, corresponding to Δt_0 in Eq. (37.6). We have

$$\frac{u^2}{c^2} = \frac{(300 \text{ m/s})^2}{(3.00 \times 10^8 \text{ m/s})^2} = 1.00 \times 10^{-12}$$

From Eq. (37.6),

$$\Delta t_0 = (1.60 \times 10^4 \text{ s})\sqrt{1 - 1.00 \times 10^{-12}}$$

The square root can't be evaluated with adequate precision with an ordinary calculator. But we can approximate it using the binomial theorem (see Appendix B):

$$(1 - 1.00 \times 10^{-12})^{1/2} = 1 - \left(\tfrac{1}{2}\right)(1.00 \times 10^{-12}) + \cdots$$

The remaining terms are of the order of 10^{-24} or smaller and can be discarded. The approximate result for Δt_0 is

$$\Delta t_0 = (1.60 \times 10^4 \text{ s})(1 - 0.50 \times 10^{-12})$$

The proper time Δt_0, measured in the airplane, is very slightly less (by less than one part in 10^{12}) than the time measured on the ground.

EVALUATE: We don't notice such effects in everyday life. But present-day atomic clocks (see Section 1.3) can attain a precision of about one part in 10^{13}. A cesium clock traveling a long distance in an airliner has been used to measure this effect and thereby verify Eq. (37.6) even at speeds much less than c.

SOLUTION

EXAMPLE 37.3 JUST WHEN IS IT PROPER?

Mavis boards a spaceship and then zips past Stanley on earth at a relative speed of $0.600c$. At the instant she passes him, they both start timers. (a) A short time later Stanley measures that Mavis has traveled 9.00×10^7 m beyond him and is passing a space station. What does Stanley's timer read as she passes the space station? What does Mavis's timer read? (b) Stanley starts to blink just as Mavis flies past him, and Mavis measures that the blink takes 0.400 s from beginning to end. According to Stanley, what is the duration of his blink?

SOLUTION

IDENTIFY and SET UP: This problem involves time dilation for two *different* sets of events measured in Stanley's frame of reference (which we call S) and in Mavis's frame of reference (which we call S'). The two events of interest in part (a) are when Mavis passes Stanley and when Mavis passes the space station; the target variables are the time intervals between these two events as measured in S and in S'. The two events in part (b) are the start and finish of Stanley's blink; the target variable is the time interval between these two events as measured in S.

EXECUTE: (a) The two events, Mavis passing the earth and Mavis passing the space station, occur at different positions in Stanley's frame but at the same position in Mavis's frame. Hence Stanley measures time interval Δt, while Mavis measures the *proper* time Δt_0. As measured by Stanley, Mavis moves at $0.600c = 0.600(3.00 \times 10^8 \text{ m/s}) = 1.80 \times 10^8$ m/s and travels 9.00×10^7 m in time $\Delta t = (9.00 \times 10^7 \text{ m})/(1.80 \times 10^8 \text{ m/s}) = 0.500$ s. From Eq. (37.6), Mavis's timer reads an elapsed time of

$$\Delta t_0 = \Delta t\sqrt{1 - u^2/c^2} = 0.500 \text{ s}\sqrt{1 - (0.600)^2} = 0.400 \text{ s}$$

(b) It is tempting to answer that Stanley's blink lasts 0.500 s in his frame. But this is wrong, because we are now considering a *different* pair of events than in part (a). The start and finish of Stanley's blink occur at the same point in his frame S but at different positions in Mavis's frame S', so the time interval of 0.400 s that she measures between these events is equal to Δt. The duration of the blink measured on Stanley's timer is the proper time Δt_0:

$$\Delta t_0 = \Delta t\sqrt{1 - u^2/c^2} = 0.400 \text{ s}\sqrt{1 - (0.600)^2} = 0.320 \text{ s}$$

EVALUATE: This example illustrates the relativity of simultaneity. In Mavis's frame she passes the space station at the same instant that Stanley finishes his blink, 0.400 s after she passed Stanley. Hence these two events are simultaneous to Mavis in frame S'. But these two events are *not* simultaneous to Stanley in his frame S: According to his timer, he finishes his blink after 0.320 s and Mavis passes the space station after 0.500 s.

Application Which One's the Grandmother? The answer to this question may seem obvious, but it could depend on which person had traveled to a distant destination at relativistic speeds. Imagine that a 20-year-old woman had given birth to a child and then immediately left on a 100-light-year trip (50 light-years out and 50 light-years back) at 99.5% the speed of light. Because of time dilation for the traveler, only 10 years would pass, and she would be 30 years old when she returned, even though 100 years had passed by for people on earth. Meanwhile, the child she left behind at home could have had a baby 20 years after her departure, and this grandchild would now be 80 years old!

The Twin Paradox

Equations (37.6) and (37.8) for time dilation suggest an apparent paradox called the **twin paradox.** Consider identical twin astronauts named Eartha and Astrid. Eartha remains on earth while her twin Astrid takes off on a high-speed trip through the galaxy. Because of time dilation, Eartha observes Astrid's heartbeat and all other life processes proceeding more slowly than her own. Thus to Eartha, Astrid ages more slowly; when Astrid returns to earth she is younger (has aged less) than Eartha.

Here is the paradox: All inertial frames are equivalent. Can't Astrid make exactly the same arguments to conclude that Eartha is in fact the younger? Then each twin measures the other to be younger when they're back together, and that's a paradox.

To resolve the paradox, note that the twins are *not* identical in all respects. While Eartha remains in an approximately inertial frame at all times, Astrid must *accelerate* with respect to that frame during parts of her trip in order to leave, turn around, and return to earth. Eartha's reference frame is always approximately inertial; Astrid's is often far from inertial. Thus there is a real physical difference between the circumstances of the two twins. Careful analysis shows that Eartha is correct; when Astrid returns, she *is* younger than Eartha.

TEST YOUR UNDERSTANDING OF SECTION 37.3 Samir (who is standing on the ground) starts his stopwatch at the instant that Maria flies past him in her spaceship at a speed of $0.600c$. At the same instant, Maria starts her stopwatch. (a) As measured in Samir's frame of reference, what is the reading on Maria's stopwatch at the instant that Samir's stopwatch reads 10.0 s? (i) 10.0 s; (ii) less than 10.0 s; (iii) more than 10.0 s. (b) As measured in Maria's frame of reference, what is the reading on Samir's stopwatch at the instant that Maria's stopwatch reads 10.0 s? (i) 10.0 s; (ii) less than 10.0 s; (iii) more than 10.0 s. ❙

37.4 RELATIVITY OF LENGTH

Not only does the time interval between two events depend on the observer's frame of reference, but the *distance* between two points may also depend on the observer's frame of reference. The concept of simultaneity is involved. Suppose you want to measure the length of a moving car. One way is to have two assistants make marks on the pavement at the positions of the front and rear bumpers. Then you measure the distance between the marks. But your assistants have to make their marks *at the same time*. If one marks the position of the front bumper at one time and the other marks the position of the rear bumper half a second later, you won't get the car's true length. Since we've learned that simultaneity isn't an absolute concept, we have to proceed with caution.

Lengths Parallel to the Relative Motion

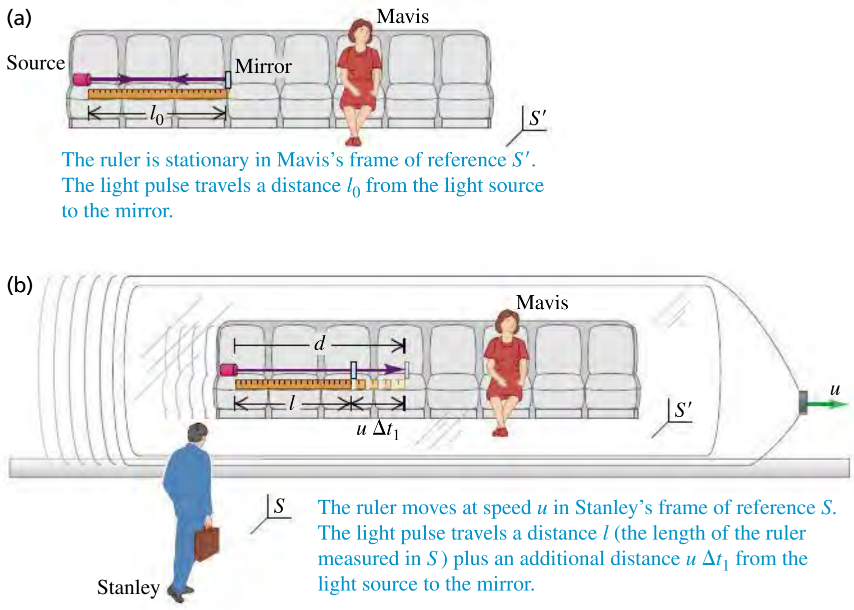
To develop a relationship between lengths that are measured parallel to the direction of motion in various coordinate systems, we consider another thought experiment. We attach a light source to one end of a ruler and a mirror to the other end. The ruler is at rest in reference frame S', and its length in this frame is l_0 (**Fig. 37.10a**). Then the time Δt_0 required for a light pulse to make the round trip from source to mirror and back is

$$\Delta t_0 = \frac{2l_0}{c} \tag{37.9}$$

This is a *proper* time interval because departure and return occur at the same point in S'.

In reference frame S the ruler is moving to the right with speed u during this travel of the light pulse (Fig. 37.10b). The length of the ruler in S is l, and the time of travel from source to mirror, as measured in S, is Δt_1. During this interval the ruler, with source and mirror attached, moves a distance $u\,\Delta t_1$. The total length of path d from source to mirror is not l, but rather

$$d = l + u\,\Delta t_1 \tag{37.10}$$



37.10 (a) A ruler is at rest in Mavis's frame S'. A light pulse is emitted from a source at one end of the ruler, reflected by a mirror at the other end, and returned to the source position. (b) Motion of the light pulse as measured in Stanley's frame S.

The light pulse travels with speed c, so it is also true that

$$d = c\,\Delta t_1 \tag{37.11}$$

Combining Eqs. (37.10) and (37.11) to eliminate d, we find

$$c\,\Delta t_1 = l + u\,\Delta t_1 \quad \text{or}$$

$$\Delta t_1 = \frac{l}{c - u} \tag{37.12}$$

(Dividing the distance l by $c - u$ does *not* mean that light travels with speed $c - u$, but rather that the distance the pulse travels in S is greater than l.)

In the same way we can show that the time Δt_2 for the return trip from mirror to source is

$$\Delta t_2 = \frac{l}{c + u} \tag{37.13}$$

The *total* time $\Delta t = \Delta t_1 + \Delta t_2$ for the round trip, as measured in S, is

$$\Delta t = \frac{l}{c - u} + \frac{l}{c + u} = \frac{2l}{c(1 - u^2/c^2)} \tag{37.14}$$

We also know that Δt and Δt_0 are related by Eq. (37.6) because Δt_0 is a proper time in S'. Thus Eq. (37.9) for the round-trip time in the rest frame S' of the ruler becomes

$$\Delta t\sqrt{1 - \frac{u^2}{c^2}} = \frac{2l_0}{c} \tag{37.15}$$

Finally, we combine Eqs. (37.14) and (37.15) to eliminate Δt and simplify:

Length contraction:

Proper length of object (measured in rest frame)

$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = \frac{l_0}{\gamma} \tag{37.16}$$

Speed of second frame relative to rest frame

Lorentz factor relating the two frames

Length in second frame of reference moving parallel to object's length

Speed of light in vacuum

[We have used the Lorentz factor γ defined in Eq. (37.7).] Thus the length l measured in S, in which the ruler is moving, is *shorter* than the length l_0 measured in its rest frame S'.

CAUTION **Length contraction is real** This is *not* an optical illusion! The ruler really is shorter in reference frame S than it is in S'.

37.11 The speed at which electrons traverse the 3-km beam line of the SLAC National Accelerator Laboratory is slower than c by less than 1 cm/s. As measured in the reference frame of such an electron, the beam line (which extends from the top to the bottom of this photograph) is only about 15 cm long!

A length measured in the frame in which the body is at rest (the rest frame of the body) is called a **proper length;** thus l_0 is a proper length in S', and the length measured in any other frame moving relative to S' is *less than* l_0. This effect is called **length contraction.**

When u is very small in comparison to c, γ approaches 1. Thus in the limit of small speeds we approach the Newtonian relationship $l = l_0$. This and the corresponding result for time dilation show that Eqs. (37.1), the Galilean coordinate transformation, are usually sufficiently accurate for relative speeds much smaller than c. If u is a reasonable fraction of c, however, the quantity $\sqrt{1 - u^2/c^2}$ can be appreciably less than 1. Then l can be substantially smaller than l_0, and the effects of length contraction can be substantial (**Fig. 37.11**).

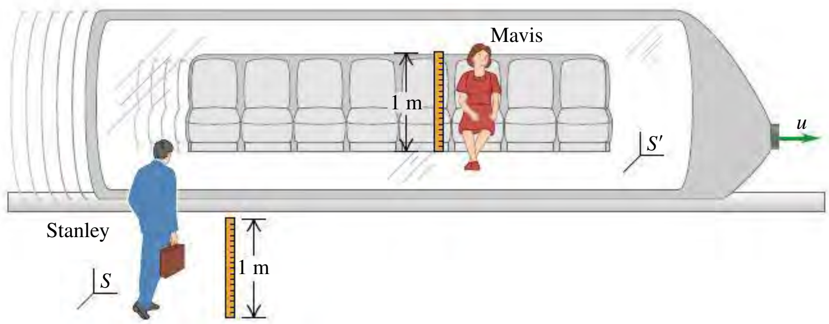
Lengths Perpendicular to the Relative Motion

We have derived Eq. (37.16) for lengths measured in the direction *parallel* to the relative motion of the two frames of reference. Lengths that are measured *perpendicular* to the direction of motion are *not* contracted. To prove this, consider two identical meter sticks. One stick is at rest in frame S and lies along the positive y-axis with one end at O, the origin of S. The other is at rest in frame S' and lies along the positive y'-axis with one end at O', the origin of S'. Frame S' moves in the positive x-direction relative to frame S. Observers Stanley and Mavis, at rest in S and S' respectively, station themselves at the 50-cm mark of their sticks. At the instant the two origins coincide, the two sticks lie along the same line. At this instant, Mavis makes a mark on Stanley's stick at the point that coincides with her own 50-cm mark, and Stanley does the same to Mavis's stick.

Suppose for the sake of argument that Stanley observes Mavis's stick as longer than his own. Then the mark Stanley makes on her stick is *below* its center. In that case, Mavis will think Stanley's stick has become shorter, since half of its length coincides with *less* than half her stick's length. So Mavis observes moving sticks getting shorter and Stanley observes them getting longer. But this implies an asymmetry between the two frames that contradicts the basic postulate of relativity that tells us all inertial frames are equivalent. We conclude that consistency with the postulates of relativity requires that both observers measure the rulers as having the *same* length, even though to each observer one of them is stationary and the other is moving (**Fig. 37.12**). So *there is no length contraction perpendicular to the direction of relative motion of the coordinate systems.* We used this result in our derivation of Eq. (37.6) in assuming that the distance d is the same in both frames of reference.

For example, suppose a moving rod of length l_0 makes an angle θ_0 with the direction of relative motion (the x-axis) as measured in its rest frame. Its length component in that frame parallel to the motion, $l_0 \cos\theta_0$, is contracted to $(l_0 \cos\theta_0)/\gamma$. However, its length component perpendicular to the motion, $l_0 \sin\theta_0$, remains the same.

37.12 The meter sticks are perpendicular to the relative velocity. For any value of u, both Stanley and Mavis measure either meter stick to have a length of 1 meter.



PROBLEM-SOLVING STRATEGY 37.2 LENGTH CONTRACTION

IDENTIFY *the relevant concepts:* The concept of length contraction is used whenever we compare the length of an object as measured by observers in different inertial frames of reference.

SET UP *the problem* using the following steps:
1. Decide what defines the length in question. If the problem describes an object such as a ruler, it is just the distance between the ends of the object. If the problem is about a distance between two points in space, it helps to envision an object like a ruler that extends from one point to the other.
2. Identify the target variable.

EXECUTE *the solution* as follows:
1. Determine the reference frame in which the object in question is at rest. In this frame, the length of the object is its proper length l_0. In a second reference frame moving at speed u relative to the first frame, the object has contracted length l.
2. Keep in mind that length contraction occurs only for lengths parallel to the direction of relative motion of the two frames. Any length that is perpendicular to the relative motion is the same in both frames.
3. Use Eq. (37.16) to relate l and l_0, and then solve for the target variable.

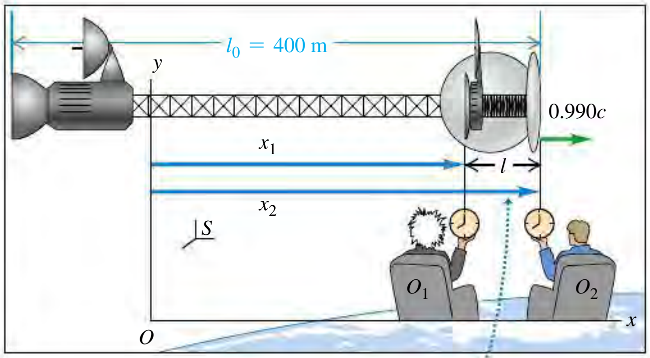
EVALUATE *your answer:* Check that your answers make sense: l is never larger than l_0, and u is never greater than c.

EXAMPLE 37.4 HOW LONG IS THE SPACESHIP?

SOLUTION

A spaceship flies past earth at a speed of $0.990c$. A crew member on board the spaceship measures its length, obtaining the value 400 m. What length do observers measure on earth?

37.13 Measuring the length of a moving spaceship.



The two observers on earth (S) must measure x_2 and x_1 simultaneously to obtain the correct length $l = x_2 - x_1$ in their frame of reference.

SOLUTION

IDENTIFY and SET UP: This problem is about the nose-to-tail length of the spaceship as measured on the spaceship and on earth. This length is along the direction of relative motion (**Fig. 37.13**), so there will be length contraction. The spaceship's 400-m length is the *proper* length l_0 because it is measured in the frame in which the spaceship is at rest. Our target variable is the length l measured in the earth frame, relative to which the spaceship is moving at $u = 0.990c$.

EXECUTE: From Eq. (37.16), the length in the earth frame is

$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = (400 \text{ m})\sqrt{1 - (0.990)^2} = 56.4 \text{ m}$$

EVALUATE: The spaceship is shorter in a frame in which it is in motion than in a frame in which it is at rest. To measure the length l, two earth observers with synchronized clocks could measure the positions of the two ends of the spaceship simultaneously in the earth's reference frame, as shown in Fig. 37.13. (These two measurements will *not* appear simultaneous to an observer in the spaceship.)

EXAMPLE 37.5 HOW FAR APART ARE THE OBSERVERS?

SOLUTION

Observers O_1 and O_2 in Fig. 37.13 are 56.4 m apart on the earth. How far apart does the spaceship crew measure them to be?

SOLUTION

IDENTIFY and SET UP: In this example the 56.4-m distance is the *proper* length l_0. It represents the length of a ruler that extends from O_1 to O_2 and is at rest in the earth frame in which the observers are at rest. Our target variable is the length l of this ruler measured in the spaceship frame, in which the earth and ruler are moving at $u = 0.990c$.

EXECUTE: As in Example 37.4, but with $l_0 = 56.4$ m,

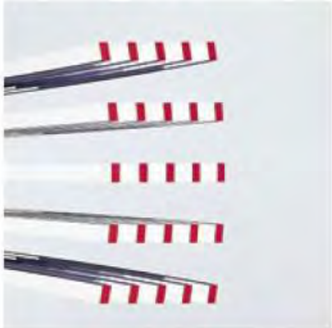
$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = (56.4 \text{ m})\sqrt{1 - (0.990)^2} = 7.96 \text{ m}$$

EVALUATE: This answer does *not* say that the crew measures their spaceship to be both 400 m long and 7.96 m long. As measured on earth, the tail of the spacecraft is at the position of O_1 at the same instant that the nose of the spacecraft is at the position of O_2. Hence the length of the spaceship measured on earth equals the 56.4-m distance between O_1 and O_2. But in the spaceship frame O_1 and O_2 are only 7.96 m apart, and the nose (which is 400 m in front of the tail) passes O_2 before the tail passes O_1.

37.14 Computer simulation of the appearance of an array of 25 rods with square cross section. The center rod is viewed end-on. The simulation ignores color changes in the array caused by the Doppler effect (see Section 37.6).

(a) Array at rest

(b) Array moving to the right at $0.2c$



(c) Array moving to the right at $0.9c$

How an Object Moving Near *c* Would Appear

Let's think a little about the visual appearance of a moving three-dimensional body. If we could see the positions of all points of the body simultaneously, it would appear to shrink only in the direction of motion. But we *don't* see all the points simultaneously; light from points farther from us takes longer to reach us than does light from points near to us, so we see the farther points at the positions they had at earlier times.

Suppose we have a rectangular rod with its faces parallel to the coordinate planes. When we look end-on at the center of the closest face of such a rod at rest, we see only that face. (See the center rod in computer-generated **Fig. 37.14a**.) But when that rod is moving past us toward the right at an appreciable fraction of the speed of light, we may also see its left side because of the earlier-time effect just described. That is, we can see some points that we couldn't see when the rod was at rest because the rod moves out of the way of the light rays from those points to us. Conversely, some light that can get to us when the rod is at rest is blocked by the moving rod. Because of all this, the rods in Figs. 37.14b and 37.14c appear rotated and distorted.

TEST YOUR UNDERSTANDING OF SECTION 37.4 A miniature spaceship is flying past you, moving horizontally at a substantial fraction of the speed of light. At a certain instant, you observe that the nose and tail of the spaceship align exactly with the two ends of a meter stick that you hold in your hands. Rank the following distances in order from longest to shortest: (i) The proper length of the meter stick; (ii) the proper length of the spaceship; (iii) the length of the spaceship measured in your frame of reference; (iv) the length of the meter stick measured in the spaceship's frame of reference. ❙

37.5 THE LORENTZ TRANSFORMATIONS

In Section 37.1 we discussed the Galilean coordinate transformation equations, Eqs. (37.1). They relate the coordinates (x, y, z) of a point in frame of reference S to the coordinates (x', y', z') of the point in a second frame S'. The second frame moves with constant speed u relative to S in the positive direction along the common x-x'-axis. This transformation also assumes that the time scale is the same in the two frames of reference, so $t = t'$. This Galilean transformation, as we have seen, is valid only in the limit when u approaches zero. We are now ready to derive more general transformations that are consistent with the principle of relativity. The more general relationships are called the **Lorentz transformations.**

The Lorentz Coordinate Transformation

Our first question is this: When an event occurs at point (x, y, z) at time t, as observed in a frame of reference S, what are the coordinates (x', y', z') and time t' of the event as observed in a second frame S' moving relative to S with constant speed u in the $+x$-direction?

To derive the coordinate transformation, we refer to **Fig. 37.15**, which is the same as Fig. 37.3. As before, we assume that the origins coincide at the initial time $t = 0 = t'$. Then in S the distance from O to O' at time t is still ut. The coordinate x' is a *proper length* in S', so in S it is contracted by the factor $1/\gamma = \sqrt{1 - u^2/c^2}$, as in Eq. (37.16). Thus the distance x from O to P, as measured in S, is not simply $x = ut + x'$, as in the Galilean coordinate transformation, but

$$x = ut + x'\sqrt{1 - \frac{u^2}{c^2}} \tag{37.17}$$

Solving this equation for x', we obtain

$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}} \tag{37.18}$$

Equation (37.18) is part of the Lorentz coordinate transformation; another part is the equation giving t' in terms of x and t. To obtain this, we note that the principle of relativity requires that the *form* of the transformation from S to S' be identical to that from S' to S. The only difference is a change in the sign of the relative velocity component u. Thus from Eq. (37.17) it must be true that

$$x' = -ut' + x\sqrt{1 - \frac{u^2}{c^2}} \qquad (37.19)$$

We now equate Eqs. (37.18) and (37.19) to eliminate x'. This gives us an equation for t' in terms of x and t. You can do the algebra to show that

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} \qquad (37.20)$$

As we discussed previously, lengths perpendicular to the direction of relative motion are not affected by the motion, so $y' = y$ and $z' = z$.

Collecting our results, we have the *Lorentz coordinate transformation:*

Lorentz coordinate transformation: Spacetime coordinates of an event are x, y, z, t in frame S and x', y', z', t' in frame S'.

Velocity of S' relative to S in positive direction along x-x'-axis

$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}} = \gamma(x - ut)$$

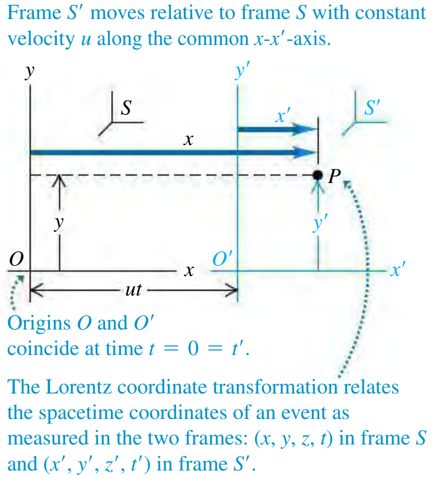
Lorentz factor relating the two frames

Speed of light in vacuum

$$y' = y \qquad (37.21)$$

$$z' = z$$

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} = \gamma(t - ux/c^2)$$



37.15 As measured in frame of reference S, x' is contracted to x'/γ, so $x = ut + (x'/\gamma)$ and $x' = \gamma(x - ut)$.

These equations are the relativistic generalization of the Galilean coordinate transformation, Eqs. (37.1) and $t = t'$. For values of u that approach zero, $\sqrt{1 - u^2/c^2}$ and γ approach 1, and the ux/c^2 term approaches zero. In this limit, Eqs. (37.21) become identical to Eqs. (37.1) along with $t = t'$. In general, though, both the coordinates and time of an event in one frame depend on its coordinates and time in another frame. *Space and time have become intertwined; we can no longer say that length and time have absolute meanings independent of the frame of reference*. For this reason, we refer to time and the three dimensions of space collectively as a four-dimensional entity called **spacetime,** and we call (x, y, z, t) together the **spacetime coordinates** of an event.

The Lorentz Velocity Transformation

We can use Eqs. (37.21) to derive the relativistic generalization of the Galilean velocity transformation, Eq. (37.2). We consider only one-dimensional motion along the x-axis and use the term "velocity" as being short for the "x-component of the velocity." Suppose that in a time dt a particle moves a distance dx, as measured in frame S. We obtain the corresponding distance dx' and time dt' in S' by taking differentials of Eqs. (37.21):

$$dx' = \gamma(dx - u\,dt)$$
$$dt' = \gamma(dt - u\,dx/c^2)$$

We divide the first equation by the second and then divide the numerator and denominator of the result by dt to obtain

$$\frac{dx'}{dt'} = \frac{\dfrac{dx}{dt} - u}{1 - \dfrac{u}{c^2}\dfrac{dx}{dt}}$$

DATA *SPEAKS*

The Lorentz Transformations

When students were given a problem involving the Lorentz transformation equations, more than 25% gave an incorrect response. Common errors:

- Confusion about which reference frame is which. It's essential in every problem to draw a diagram showing which reference frame goes with which observer.
- Confusion about length contraction. Remember that lengths perpendicular to the direction of relative motion do not contract.

Application Relative Velocity and Reference Frames A relay race illustrates the frames of reference used in Eqs. (37.22) and (37.23). The runner in purple is the particle, and the two frames of reference in which the particle's motion is observed are our rest frame S (we are spectators standing next to the track) and the rest frame S' of the runner in red, who has speed u relative to us. Since the runner in red is moving to the left relative to us, we must take the positive x-direction to the left. The runner in purple has positive velocity v_x relative to us (she is moving to the left); if the runner in red is moving faster ($u > v_x$), then from Eq. (37.22) the runner in purple has a *negative* velocity v'_x relative to the runner in red.

CAUTION Use the correct reference frame coordinates The Lorentz transformation equations given by Eqs. (37.21), (37.22), and (37.23) assume that frame S' is moving in the positive x-direction with velocity u relative to frame S. Always set up your coordinate system to follow this convention.

Now dx/dt is the velocity v_x in S, and dx'/dt' is the velocity v'_x in S', so

Lorentz velocity transformation (velocity in S' in terms of velocity in S):

$$\underbrace{v'_x}_{x\text{-velocity of object in frame } S'} = \frac{\overbrace{v_x}^{x\text{-velocity of object in frame } S} - \overbrace{u}^{\text{Velocity of } S' \text{ relative to } S \text{ in positive direction along } x\text{-}x'\text{-axis}}}{1 - uv_x/\underbrace{c^2}_{\text{Speed of light in vacuum}}} \tag{37.22}$$

When u and v_x are much smaller than c, the denominator in Eq. (37.22) approaches 1, and we approach the nonrelativistic result $v'_x = v_x - u$. The opposite extreme is the case $v_x = c$; then we find

$$v'_x = \frac{c - u}{1 - uc/c^2} = \frac{c(1 - u/c)}{1 - u/c} = c$$

This says that anything moving with velocity $v_x = c$ measured in S also has velocity $v'_x = c$ measured in S', despite the relative motion of the two frames. So Eq. (37.22) is consistent with Einstein's postulate that the speed of light in vacuum is the same in all inertial frames of reference.

The principle of relativity tells us there is no fundamental distinction between the two frames S and S'. Thus the expression for v_x in terms of v'_x must have the same form as Eq. (37.22), with v_x changed to v'_x, and vice versa, and the sign of u reversed. Carrying out these operations with Eq. (37.22), we find

Lorentz velocity transformation (velocity in S in terms of velocity in S'):

$$\underbrace{v_x}_{x\text{-velocity of object in frame } S} = \frac{\overbrace{v'_x}^{x\text{-velocity of object in frame } S'} + \overbrace{u}^{\text{Velocity of } S' \text{ relative to } S \text{ in positive direction along } x\text{-}x'\text{-axis}}}{1 + uv'_x/\underbrace{c^2}_{\text{Speed of light in vacuum}}} \tag{37.23}$$

You can also obtain this equation by solving Eq. (37.22) for v_x. Both Eqs. (37.22) and (37.23) are *Lorentz velocity transformations* for one-dimensional motion.

When u is less than c, the Lorentz velocity transformations show us that a body moving with a speed less than c in one frame of reference always has a speed less than c in *every other* frame of reference. This is one reason for concluding that no material body may travel with a speed equal to or greater than c relative to *any* inertial frame of reference. Later we'll see that the relativistic generalizations of energy and momentum give further support to this hypothesis.

PROBLEM-SOLVING STRATEGY 37.3 LORENTZ TRANSFORMATIONS

IDENTIFY *the relevant concepts:* The Lorentz *coordinate* transformation equations relate the spacetime coordinates of an event in one inertial reference frame to the coordinates of the same event in a second inertial frame. The Lorentz *velocity* transformation equations relate the velocity of an object in one inertial reference frame to its velocity in a second inertial frame.

SET UP *the problem* using the following steps:

1. Identify the target variable.
2. Define the two inertial frames S and S'. Remember that S' moves relative to S at a constant velocity u in the $+x$-direction.
3. If the coordinate transformation equations are needed, make a list of spacetime coordinates in the two frames, such as x_1, x'_1, t_1, t'_1, and so on. Label carefully which of these you know and which you don't.
4. In velocity-transformation problems, clearly identify u (the relative velocity of the two frames of reference), v_x (the velocity of the object relative to S), and v'_x (the velocity of the object relative to S').

EXECUTE *the solution* as follows:

1. In a coordinate-transformation problem, use Eqs. (37.21) to solve for the spacetime coordinates of the event as measured in S' in terms of the corresponding values in S. (If you need to solve for the spacetime coordinates in S in terms of the corresponding values in S', you can easily convert the expressions in Eqs. (37.21): Replace all of the primed quantities with unprimed ones, and vice versa, and replace u with $-u$.)
2. In a velocity-transformation problem, use either Eq. (37.22) or Eq. (37.23), as appropriate, to solve for the target variable.

EVALUATE *your answer:* Don't be discouraged if some of your results don't seem to make sense or if they disagree with "common sense." It takes time to develop intuition about relativity; you'll gain it with experience.

SOLUTION

EXAMPLE 37.6 WAS IT RECEIVED BEFORE IT WAS SENT?

Winning an interstellar race, Mavis pilots her spaceship across a finish line in space at a speed of $0.600c$ relative to that line. A "hooray" message is sent from the back of her ship (event 2) at the instant (in her frame of reference) that the front of her ship crosses the line (event 1). She measures the length of her ship to be 300 m. Stanley is at the finish line and is at rest relative to it. When and where does he measure events 1 and 2 to occur?

SOLUTION

IDENTIFY and SET UP: This example involves the Lorentz coordinate transformation. Our derivation of this transformation assumes that the origins of frames S and S' coincide at $t = 0 = t'$. Thus for simplicity we fix the origin of S at the finish line and the origin of S' at the front of the spaceship so that Stanley and Mavis measure event 1 to be at $x = 0 = x'$ and $t = 0 = t'$.

Mavis in S' measures her spaceship to be 300 m long, so she has the "hooray" sent from 300 m behind her spaceship's front at the instant she measures the front to cross the finish line. That is, she measures event 2 at $x' = -300$ m and $t' = 0$.

Our target variables are the coordinate x and time t of event 2 that Stanley measures in S.

EXECUTE: To solve for the target variables, we modify the first and last of Eqs. (37.21) to give x and t as functions of x' and t'. We do so in the same way that we obtained Eq. (37.23) from Eq. (37.22). We remove the primes from x' and t', add primes to x and t, and replace each u with $-u$. The results are

$$x = \gamma(x' + ut') \quad \text{and} \quad t = \gamma(t' + ux'/c^2)$$

From Eq. (37.7), $\gamma = 1.25$ for $u = 0.600c = 1.80 \times 10^8$ m/s. We also substitute $x' = -300$ m, $t' = 0$, $c = 3.00 \times 10^8$ m/s, and $u = 1.80 \times 10^8$ m/s in the equations for x and t to find $x = -375$ m at $t = -7.50 \times 10^{-7}$ s $= -0.750\ \mu$s for event 2.

EVALUATE: Mavis says that the events are simultaneous, but Stanley says that the "hooray" was sent *before* Mavis crossed the finish line. This does not mean that the effect preceded the cause. The fastest that Mavis can send a signal the length of her ship is $300\text{ m}/(3.00 \times 10^8\text{ m/s}) = 1.00\ \mu$s. She cannot send a signal from the front at the instant it crosses the finish line that would cause a "hooray" to be broadcast from the back at the same instant. She would have to send that signal from the front at least $1.00\ \mu$s before then, so she had to slightly anticipate her success.

SOLUTION

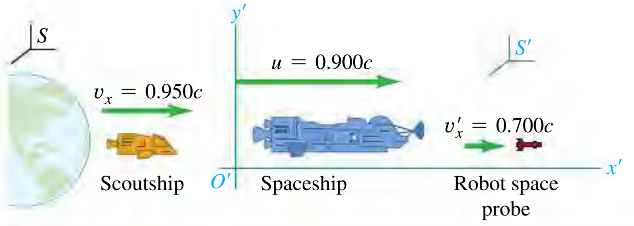
EXAMPLE 37.7 RELATIVE VELOCITIES

(a) A spaceship moving away from the earth at $0.900c$ fires a robot space probe in the same direction as its motion at $0.700c$ relative to the spaceship. What is the probe's velocity relative to the earth? (b) A scoutship is sent to catch up with the spaceship by traveling at $0.950c$ relative to the earth. What is the velocity of the scoutship relative to the spaceship?

SOLUTION

IDENTIFY and SET UP: This example uses the Lorentz velocity transformation. Let the earth and spaceship reference frames be S and S', respectively (**Fig. 37.16**); their relative velocity is $u = 0.900c$. In part (a) we are given the probe velocity $v'_x = 0.700c$ with respect to S', and the target variable is the velocity v_x of the probe relative to S. In part (b) we are given the velocity $v_x = 0.950c$ of the scoutship relative to S, and the target variable is its velocity v'_x relative to S'.

37.16 The spaceship, robot space probe, and scoutship.



EXECUTE: (a) We use Eq. (37.23) to find the probe velocity relative to the earth:

$$v_x = \frac{v'_x + u}{1 + uv'_x/c^2} = \frac{0.700c + 0.900c}{1 + (0.900c)(0.700c)/c^2} = 0.982c$$

(b) We use Eq. (37.22) to find the scoutship velocity relative to the spaceship:

$$v'_x = \frac{v_x - u}{1 - uv_x/c^2} = \frac{0.950c - 0.900c}{1 - (0.900c)(0.950c)/c^2} = 0.345c$$

EVALUATE: What would the Galilean velocity transformation formula, Eq. (37.2), say? In part (a) we would have found the probe's velocity relative to the earth to be $v_x = v'_x + u = 0.700c + 0.900c = 1.600c$, which is greater than c and hence impossible. In part (b), we would have found the scoutship's velocity relative to the spaceship to be $v'_x = v_x - u = 0.950c - 0.900c = 0.050c$; the relativistically correct value, $v'_x = 0.345c$, is almost seven times greater than the incorrect Galilean value.

TEST YOUR UNDERSTANDING OF SECTION 37.5 (a) In frame S events P_1 and P_2 occur at the same x-, y-, and z-coordinates, but event P_1 occurs before event P_2. In frame S', which event occurs first? (b) In frame S events P_3 and P_4 occur at the same time t and the same y- and z-coordinates, but event P_3 occurs at a less positive x-coordinate than event P_4. In frame S', which event occurs first?

37.6 THE DOPPLER EFFECT FOR ELECTROMAGNETIC WAVES

An additional important consequence of relativistic kinematics is the Doppler effect for electromagnetic waves. In Section 16.8 we quoted without proof the formula, Eq. (16.30), for the frequency shift that results from motion of a source of electromagnetic waves relative to an observer. We can now derive that result.

Here's a statement of the problem. A source of light is moving with constant speed u toward Stanley, who is stationary in an inertial frame (**Fig. 37.17**). As measured in its rest frame, the source emits light waves with frequency f_0 and period $T_0 = 1/f_0$. What is the frequency f of these waves as received by Stanley?

Let T be the time interval between *emission* of successive wave crests as observed in Stanley's reference frame. Note that this is *not* the interval between the *arrival* of successive crests at his position, because the crests are emitted at different points in Stanley's frame. In measuring only the frequency f he receives, he does not take into account the difference in transit times for successive crests. Therefore the frequency he receives is *not* $1/T$. What is the equation for f?

During a time T the crests ahead of the source move a distance cT, and the source moves a shorter distance uT in the same direction. The distance λ between successive crests—that is, the wavelength—is thus $\lambda = (c - u)T$, as measured in Stanley's frame. The frequency that he measures is c/λ. Therefore

$$f = \frac{c}{(c - u)T} \tag{37.24}$$

So far we have followed a pattern similar to that for the Doppler effect for sound from a moving source (see Section 16.8). In that discussion our next step was to equate T to the time T_0 between emissions of successive wave crests by the source. However, due to time dilation it is *not* relativistically correct to equate T to T_0. The time T_0 is measured in the rest frame of the source, so it is a proper time. From Eq. (37.6), T_0 and T are related by

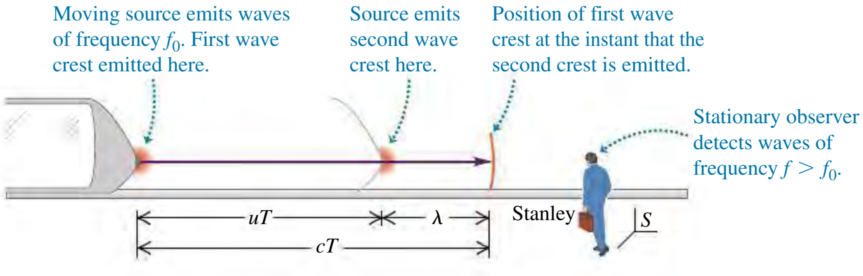
$$T = \frac{T_0}{\sqrt{1 - u^2/c^2}} = \frac{cT_0}{\sqrt{c^2 - u^2}}$$

or, since $T_0 = 1/f_0$,

$$\frac{1}{T} = \frac{\sqrt{c^2 - u^2}}{cT_0} = \frac{\sqrt{c^2 - u^2}}{c} f_0$$

Remember, $1/T$ is not equal to f. We must substitute this expression for $1/T$ into Eq. (37.24) to find f:

$$f = \frac{c}{c - u}\frac{\sqrt{c^2 - u^2}}{c} f_0$$



37.17 The Doppler effect for light. A light source moving at speed u relative to Stanley emits a wave crest, then travels a distance uT toward an observer and emits the next crest. In Stanley's reference frame S, the second crest is a distance λ behind the first crest.

Using $c^2 - u^2 = (c - u)(c + u)$ gives

Doppler effect, electromagnetic waves, source approaching observer:

Frequency measured by observer

$$f = \sqrt{\frac{c + u}{c - u}}\, f_0 \qquad (37.25)$$

Frequency measured in rest frame of source

Speed of light in vacuum

Speed of source relative to observer

This shows that when the source moves *toward* the observer, the observed frequency f is *greater* than the emitted frequency f_0. The difference $f - f_0 = \Delta f$ is called the Doppler frequency shift. When u/c is much smaller than 1, the fractional shift $\Delta f/f$ is also small and is approximately equal to u/c:

$$\frac{\Delta f}{f} = \frac{u}{c}$$

When the source moves *away from* the observer, we change the sign of u in Eq. (37.25) to get

$$f = \sqrt{\frac{c - u}{c + u}}\, f_0 \quad \text{(Doppler effect, electromagnetic waves, source moving away from observer)} \qquad (37.26)$$

This agrees with Eq. (16.30) with minor notation changes.

With light, unlike sound, there is no distinction between motion of source and motion of observer; only the *relative* velocity of the two is significant. The last four paragraphs of Section 16.8 discuss several practical applications of the Doppler effect with light and other electromagnetic radiation; we suggest you review those paragraphs now. **Figure 37.18** shows one common application.

37.18 This handheld radar gun emits a radio beam of frequency f_0, which in the frame of reference of an approaching car has a higher frequency f given by Eq. (37.25). The reflected beam also has frequency f in the car's frame, but has an even higher frequency f' in the police officer's frame. The radar gun calculates the car's speed by comparing the frequencies of the emitted beam and the doubly Doppler-shifted reflected beam. (Compare Example 16.18 in Section 16.8.)

EXAMPLE 37.8 A JET FROM A BLACK HOLE

SOLUTION

Many galaxies have supermassive black holes at their centers (see Section 13.8). As material swirls around such a black hole, it is heated, becomes ionized, and generates strong magnetic fields. The resulting magnetic forces steer some of the material into high-speed jets that blast out of the galaxy and into intergalactic space (**Fig. 37.19**). The light we observe from the jet in Fig. 37.19 has a frequency of 6.66×10^{14} Hz (in the far ultraviolet; see Fig. 32.4), but in the reference frame of the jet material the light has a frequency of 5.55×10^{13} Hz (in the infrared). What is the speed of the jet material with respect to us?

SOLUTION

IDENTIFY and SET UP: This problem involves the Doppler effect for electromagnetic waves. The frequency we observe is $f = 6.66 \times 10^{14}$ Hz, and the frequency in the frame of the source is $f_0 = 5.55 \times 10^{13}$ Hz. Since $f > f_0$, the jet is approaching us and we use Eq. (37.25) to find the target variable u.

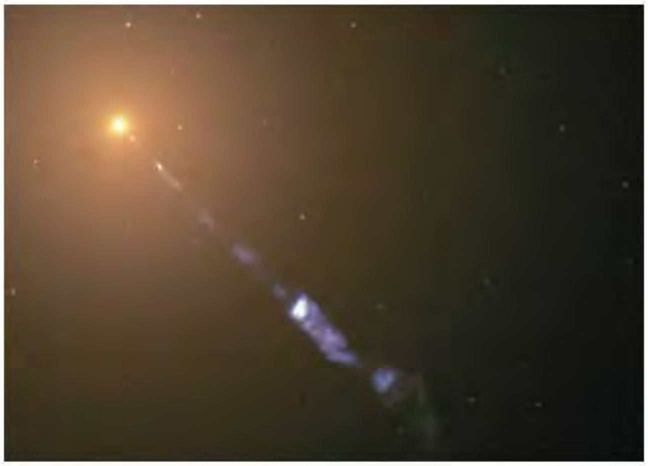
EXECUTE: We need to solve Eq. (37.25) for u. We'll leave it as an exercise for you to show that the result is

$$u = \frac{(f/f_0)^2 - 1}{(f/f_0)^2 + 1}c$$

We have $f/f_0 = (6.66 \times 10^{14}\ \text{Hz})/(5.55 \times 10^{13}\ \text{Hz}) = 12.0$, so

$$u = \frac{(12.0)^2 - 1}{(12.0)^2 + 1}c = 0.986c$$

37.19 This image shows a fast-moving jet 5000 light-years in length emanating from the center of the galaxy M87. The light from the jet is emitted by fast-moving electrons spiraling around magnetic field lines (see Fig. 27.18).



EVALUATE: Because the frequency shift is quite substantial, it would have been erroneous to use the approximate expression $\Delta f/f = u/c$. Had you done so, you would have found $u = c(\Delta f/f_0) = c(6.66 \times 10^{14}\ \text{Hz} - 5.55 \times 10^{13}\ \text{Hz})/(5.55 \times 10^{13}\ \text{Hz}) = 11.0c$. This result cannot be correct because the jet material cannot travel faster than light.

37.7 RELATIVISTIC MOMENTUM

Newton's laws of motion have the same form in all inertial frames of reference. When we use transformations to change from one inertial frame to another, the laws should be *invariant* (unchanging). But we have just learned that the principle of relativity forces us to replace the Galilean transformations with the more general Lorentz transformations. As we will see, this requires corresponding generalizations in the laws of motion and the definitions of momentum and energy.

The principle of conservation of momentum states that *when two bodies interact, the total momentum is constant,* provided that the net external force acting on the bodies in an inertial reference frame is zero (for example, if they form an isolated system, interacting only with each other). If conservation of momentum is a valid physical law, it must be valid in *all* inertial frames of reference. Now, here's the problem: Suppose we look at a collision in one inertial coordinate system S and find that momentum is conserved. Then we use the Lorentz transformation to obtain the velocities in a second inertial system S'. We find that if we use the Newtonian definition of momentum ($\vec{p} = m\vec{v}$), momentum is *not* conserved in the second system! The only way to make momentum conservation consistent with relativity is to generalize the *definition* of momentum.

We won't derive the correct relativistic generalization of momentum, but here is the result. Suppose we measure the mass of a particle to be m when it is at rest relative to us: We often call m the **rest mass.** We will use the term *material particle* for a particle that has a nonzero rest mass. When such a particle has a velocity $\vec{v}$, its **relativistic momentum** $\vec{p}$ is

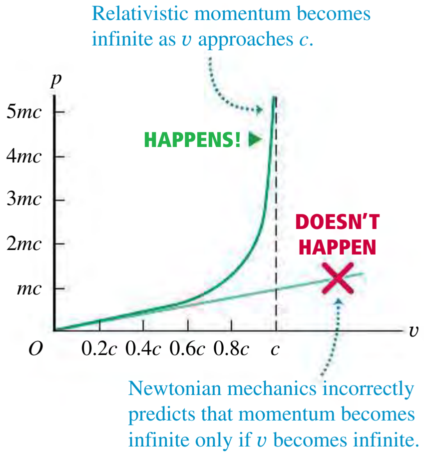
Rest mass of particle; Velocity of particle

Relativistic momentum $$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} \tag{37.27}$$

Speed of particle; Speed of light in vacuum

When the particle's speed v is much less than c, this is approximately equal to the Newtonian expression $\vec{p} = m\vec{v}$, but in general the momentum is greater in magnitude than mv (**Fig. 37.20**). In fact, as v approaches c, the momentum approaches infinity.

37.20 Graph of the magnitude of the momentum of a particle of rest mass m as a function of speed v. Also shown is the Newtonian prediction, which gives correct results only at speeds much less than c.



Relativity, Newton's Second Law, and Relativistic Mass

What about the relativistic generalization of Newton's second law? In Newtonian mechanics the most general form of the second law is

$$\vec{F} = \frac{d\vec{p}}{dt} \tag{37.28}$$

That is, the net force $\vec{F}$ on a particle equals the time rate of change of its momentum. Experiments show that this result is still valid in relativistic mechanics, provided that we use the relativistic momentum given by Eq. (37.27). That is, the relativistically correct generalization of Newton's second law is

$$\vec{F} = \frac{d}{dt}\frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} \tag{37.29}$$

Because momentum is no longer directly proportional to velocity, the rate of change of momentum is no longer directly proportional to the acceleration. As a result, *constant force does not cause constant acceleration*. For example, when the net force and the velocity are both along the x-axis, Eq. 37.29 gives

$$F = \frac{m}{(1 - v^2/c^2)^{3/2}}a \qquad (\vec{F} \text{ and } \vec{v} \text{ along the same line}) \tag{37.30}$$

where a is the acceleration, also along the x-axis. Solving Eq. (37.30) for the acceleration a gives

$$a = \frac{F}{m}\left(1 - \frac{v^2}{c^2}\right)^{3/2}$$

We see that as a particle's speed increases, the acceleration caused by a given force continuously *decreases*. As the speed approaches c, the acceleration approaches zero, no matter how great a force is applied. Thus it is impossible to accelerate a particle with nonzero rest mass to a speed equal to or greater than c. We again see that the speed of light in vacuum represents an ultimate speed limit.

Equation (37.27) for relativistic momentum is sometimes interpreted to mean that a rapidly moving particle undergoes an increase in mass. If the mass at zero velocity (the rest mass) is denoted by m, then the "relativistic mass" m_{rel} is

$$m_{\text{rel}} = \frac{m}{\sqrt{1 - v^2/c^2}}$$

Indeed, when we consider the motion of a system of particles (such as rapidly moving ideal-gas molecules in a stationary container), the total rest mass of the system is the sum of the relativistic masses of the particles, not the sum of their rest masses.

However, if blindly applied, the concept of relativistic mass has its pitfalls. As Eq. (37.29) shows, the relativistic generalization of Newton's second law is *not* $\vec{F} = m_{\text{rel}}\vec{a}$, and we will show in Section 37.8 that the relativistic kinetic energy of a particle is *not* $K = \frac{1}{2}m_{\text{rel}}v^2$. The use of relativistic mass has its supporters and detractors, some quite strong in their opinions. We will mostly deal with individual particles, so we will sidestep the controversy and use Eq. (37.27) as the generalized definition of momentum with m as a constant for each particle, independent of its state of motion.

The quantity $1/\sqrt{1 - v^2/c^2}$ in Eqs. (37.27) and (37.29) is the Lorentz factor γ from Eq. (37.7) (Section 37.3), but with a difference: We've replaced u, the relative speed of two coordinate systems, by v, the speed of a particle in a particular coordinate system—that is, the speed of the particle's *rest frame* with respect to that system. In terms of γ, Eqs. (37.27) and (37.30) become

Rest mass of particle Velocity of particle

Relativistic momentum $\vec{p} = \gamma m \vec{v}$ Lorentz factor relating rest frame of particle and frame of observer (37.31)

$$F = \gamma^3 ma \qquad (\vec{F} \text{ and } \vec{v} \text{ along the same line}) \tag{37.32}$$

In linear accelerators (used in medicine as well as nuclear and elementary-particle physics; see Fig. 37.11) the net force $\vec{F}$ and the velocity $\vec{v}$ of the accelerated particle are along the same straight line. But for much of the path in most *circular* accelerators the particle moves in uniform circular motion at constant speed v. Then the net force and velocity are perpendicular, so the force can do no work on the particle and the kinetic energy and speed remain constant. Thus the denominator in Eq. (37.29) is constant, and we obtain

$$F = \frac{m}{(1 - v^2/c^2)^{1/2}}a = \gamma ma \qquad (\vec{F} \text{ and } \vec{v} \text{ perpendicular}) \tag{37.33}$$

Recall from Section 3.4 that if the particle moves in a circle, the net force and acceleration are directed inward along the radius r, and $a = v^2/r$.

What about the general case in which $\vec{F}$ and $\vec{v}$ are neither along the same line nor perpendicular? Then we can resolve the net force $\vec{F}$ at any instant into components parallel to and perpendicular to $\vec{v}$. The resulting acceleration will have corresponding components obtained from Eqs. (37.32) and (37.33). Because of the different γ^3 and γ factors, the acceleration components will not be proportional to the net force components. That is, *unless the net force on a relativistic particle is either along the same line as the particle's velocity or perpendicular to it, the net force and acceleration vectors are not parallel.*

EXAMPLE 37.9 RELATIVISTIC DYNAMICS OF AN ELECTRON

SOLUTION

An electron (rest mass 9.11×10^{-31} kg, charge -1.60×10^{-19} C) is moving opposite to an electric field of magnitude $E = 5.00 \times 10^5$ N/C. All other forces are negligible in comparison to the electric-field force. (a) Find the magnitudes of momentum and of acceleration at the instants when $v = 0.010c$, $0.90c$, and $0.99c$. (b) Find the corresponding accelerations if a net force of the same magnitude is perpendicular to the velocity.

SOLUTION

IDENTIFY and SET UP: In addition to the expressions from this section for relativistic momentum and acceleration, we need the relationship between electric force and electric field from Chapter 21. In part (a) we use Eq. (37.31) to determine the magnitude of momentum; the force acts along the same line as the velocity, so we use Eq. (37.32) to determine the magnitude of acceleration. In part (b) the force is perpendicular to the velocity, so we use Eq. (37.33) rather than Eq. (37.32).

EXECUTE: (a) For $v = 0.010c$, $0.90c$, and $0.99c$ we have $\gamma = \sqrt{1 - v^2/c^2} = 1.00$, 2.29, and 7.09, respectively. The values of the momentum magnitude $p = \gamma m v$ are

$$p_1 = (1.00)(9.11 \times 10^{-31}\ \text{kg})(0.010)(3.00 \times 10^8\ \text{m/s})$$
$$= 2.7 \times 10^{-24}\ \text{kg} \cdot \text{m/s at } v_1 = 0.010c$$

$$p_2 = (2.29)(9.11 \times 10^{-31}\ \text{kg})(0.90)(3.00 \times 10^8\ \text{m/s})$$
$$= 5.6 \times 10^{-22}\ \text{kg} \cdot \text{m/s at } v_2 = 0.90c$$

$$p_3 = (7.09)(9.11 \times 10^{-31}\ \text{kg})(0.99)(3.00 \times 10^8\ \text{m/s})$$
$$= 1.9 \times 10^{-21}\ \text{kg} \cdot \text{m/s at } v_3 = 0.99c$$

From Eq. (21.4), the magnitude of the force on the electron is

$$F = |q|E = (1.60 \times 10^{-19}\ \text{C})(5.00 \times 10^5\ \text{N/C})$$
$$= 8.00 \times 10^{-14}\ \text{N}$$

From Eq. (37.32), $a = F/\gamma^3 m$. For $v = 0.010c$ and $\gamma = 1.00$,

$$a_1 = \frac{8.00 \times 10^{-14}\ \text{N}}{(1.00)^3(9.11 \times 10^{-31}\ \text{kg})} = 8.8 \times 10^{16}\ \text{m/s}^2$$

The accelerations at the two higher speeds are smaller than the nonrelativistic value by factors of $\gamma^3 = 12.0$ and 356, respectively:

$$a_2 = 7.3 \times 10^{15}\ \text{m/s}^2 \qquad a_3 = 2.5 \times 10^{14}\ \text{m/s}^2$$

(b) From Eq. (37.33), $a = F/\gamma m$ if $\vec{F}$ and $\vec{v}$ are perpendicular. When $v = 0.010c$ and $\gamma = 1.00$,

$$a_1 = \frac{8.00 \times 10^{-14}\ \text{N}}{(1.00)(9.11 \times 10^{-31}\ \text{kg})} = 8.8 \times 10^{16}\ \text{m/s}^2$$

Now the accelerations at the two higher speeds are smaller by factors of $\gamma = 2.29$ and 7.09, respectively:

$$a_2 = 3.8 \times 10^{16}\ \text{m/s}^2 \qquad a_3 = 1.2 \times 10^{16}\ \text{m/s}^2$$

These accelerations are larger than the corresponding ones in part (a) by factors of γ^2.

EVALUATE: Our results in part (a) show that at higher speeds, the relativistic values of momentum differ more and more from the nonrelativistic values calculated from $p = mv$. The momentum at $0.99c$ is more than three times as great as at $0.90c$ because of the increase in the factor γ. Our results also show that the acceleration drops off very quickly as v approaches c.

TEST YOUR UNDERSTANDING OF SECTION 37.7 According to relativistic mechanics, when you double the speed of a particle, the magnitude of its momentum increases by (i) a factor of 2; (ii) a factor greater than 2; (iii) a factor between 1 and 2 that depends on the mass of the particle. ❙

37.8 RELATIVISTIC WORK AND ENERGY

When we developed the relationship between work and kinetic energy in Chapter 6, we used Newton's laws of motion. Since we have generalized these laws according to the principle of relativity, we need a corresponding generalization of the equation for kinetic energy.

Relativistic Kinetic Energy

We use the work–energy theorem, beginning with the definition of work. When the net force and displacement are in the same direction, the work done by that force is $W = \int F\,dx$. We substitute the expression for F from Eq. (37.30), the relativistic version of Newton's second law for straight-line motion. In moving a particle of rest mass m from point x_1 to point x_2,

$$W = \int_{x_1}^{x_2} F\,dx = \int_{x_1}^{x_2} \frac{ma\,dx}{(1 - v_x{}^2/c^2)^{3/2}} \qquad (37.34)$$

We've replaced v in Eq. (37.34) with v_x because the motion is along the x-axis only. So v_x is the varying x-component of the particle's velocity as the net force accelerates it. To derive the generalized expression for kinetic energy K, first remember that the kinetic energy of a particle equals the net work done on it in moving it from rest to speed v: $K = W$. Thus we let the speeds be zero at point x_1 and v at point x_2. It's useful to convert Eq. (37.34) to an integral on v_x. To do this, note that dx and dv_x are the infinitesimal changes in x and v_x, respectively, in the time interval dt. Because $v_x = dx/dt$ and $a = dv_x/dt$, we can rewrite $a\,dx$ in Eq. (37.34) as

$$a\,dx = \frac{dv_x}{dt}dx = dx\frac{dv_x}{dt} = \frac{dx}{dt}dv_x = v_x dv_x$$

Making these substitutions gives us

$$K = W = \int_0^v \frac{mv_x dv_x}{(1 - v_x^2/c^2)^{3/2}} \tag{37.35}$$

We can evaluate this integral by a simple change of variable; the final result is

Rest mass of particle — Speed of light in vacuum

$$\text{Relativistic kinetic energy} \quad K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2 \tag{37.36}$$

Speed of particle — Lorentz factor relating rest frame of particle and frame of observer

As v approaches c, the kinetic energy approaches infinity. If Eq. (37.36) is correct, it must also approach the Newtonian expression $K = \frac{1}{2}mv^2$ when v is much smaller than c (**Fig. 37.21**). To verify this, we expand the radical, using the binomial theorem in the form

?

$$(1 + x)^n = 1 + nx + n(n - 1)x^2/2 + \cdots$$

In our case, $n = -\frac{1}{2}$ and $x = -v^2/c^2$, and we get

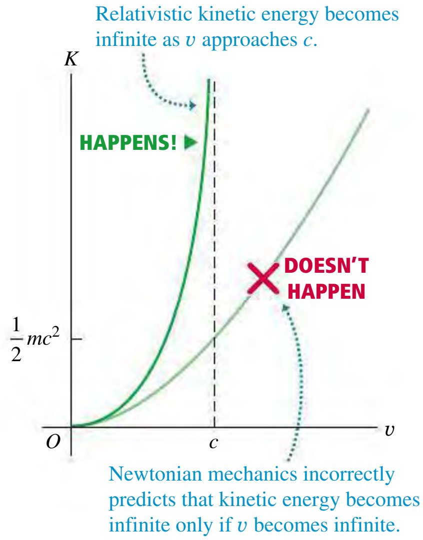
$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots$$

Combining this with $K = (\gamma - 1)mc^2$, we find

$$K = \left(1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots - 1\right)mc^2 = \frac{1}{2}mv^2 + \frac{3}{8}\frac{mv^4}{c^2} + \cdots \tag{37.37}$$

When v is much smaller than c, all the terms in the series in Eq. (37.37) except the first are negligibly small, and we obtain the Newtonian expression $\frac{1}{2}mv^2$.

37.21 Graph of the kinetic energy of a particle of rest mass m as a function of speed v. Also shown is the Newtonian prediction, which gives correct results only at speeds much less than c.



Rest Energy and $E = mc^2$

Equation (37.36) for the kinetic energy of a moving particle includes a term $mc^2/\sqrt{1 - v^2/c^2}$ that depends on the motion and a second energy term mc^2 that is independent of the motion. It seems that the kinetic energy of a particle is the difference between some **total energy** E and an energy mc^2 that it has even when it is at rest. Thus we can rewrite Eq. (37.36) as

Kinetic energy — Rest energy — Rest mass of particle — Speed of light in vacuum

$$\text{Total energy of a particle} \quad E = K + mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \gamma mc^2 \tag{37.38}$$

Speed of particle — Lorentz factor relating rest frame of particle and frame of observer

For a particle at rest ($K = 0$), we see that $E = mc^2$. The energy mc^2 associated with rest mass m rather than motion is called the **rest energy** of the particle.

Application Monitoring Mass-Energy Conversion Although the control room of a nuclear power plant is very complex, the physical principle on which such a plant operates is a simple one: Part of the rest energy of atomic nuclei is converted to thermal energy, which in turn is used to produce steam to drive electric generators.

There is abundant experimental evidence that rest energy really does exist. The simplest example is the decay of a neutral *pion*. This is an unstable subatomic particle of rest mass m_π; when it decays, it disappears and electromagnetic radiation appears. If a neutral pion has no kinetic energy before its decay, the total energy of the radiation after its decay is found to equal exactly $m_\pi c^2$. In many other fundamental particle transformations the sum of the rest masses of the particles changes. In every case there is a corresponding energy change, consistent with the assumption of a rest energy mc^2 associated with a rest mass m.

Historically, the principles of conservation of mass and of energy developed quite independently. The theory of relativity shows that they are actually two special cases of a single broader conservation principle, the *principle of conservation of mass and energy*. In some physical phenomena, neither the sum of the rest masses of the particles nor the total energy other than rest energy is separately conserved, but there is a more general conservation principle: In an isolated system, when the sum of the rest masses changes, there is always a change in $1/c^2$ times the total energy other than the rest energy. This change is equal in magnitude but opposite in sign to the change in the sum of the rest masses.

This more general mass-energy conservation law is the fundamental principle involved in the generation of nuclear power. When a uranium nucleus undergoes fission in a nuclear reactor, the sum of the rest masses of the resulting fragments is *less than* the rest mass of the parent nucleus. An amount of energy is released that equals the mass decrease multiplied by c^2. Most of this energy can be used to produce steam to operate turbines for electric power generators.

We can also relate the total energy E of a particle (kinetic energy plus rest energy) directly to its momentum by combining Eq. (37.27) for relativistic momentum and Eq. (37.38) for total energy to eliminate the particle's velocity. The simplest procedure is to rewrite these equations in the following forms:

$$\left(\frac{E}{mc^2}\right)^2 = \frac{1}{1 - v^2/c^2} \quad \text{and} \quad \left(\frac{p}{mc}\right)^2 = \frac{v^2/c^2}{1 - v^2/c^2}$$

Subtracting the second of these from the first and rearranging, we find

Total energy, rest energy, and momentum:

Total energy Rest energy Magnitude of momentum

$$E^2 = (mc^2)^2 + (pc)^2 \qquad (37.39)$$

Rest mass Speed of light in vacuum

Again we see that for a particle at rest $(p = 0)$, $E = mc^2$.

Equation (37.39) also suggests that a particle may have energy and momentum even when it has no rest mass. In such a case, $m = 0$ and

$$E = pc \qquad \text{(zero rest mass)} \qquad (37.40)$$

In fact, zero rest mass particles do exist. Such particles always travel at the speed of light in vacuum. One example is the *photon*, the quantum of electromagnetic radiation (to be discussed in Chapter 38). Photons are emitted and absorbed during changes of state of an atomic or nuclear system when the energy and momentum of the system change.

EXAMPLE 37.10 ENERGETIC ELECTRONS

SOLUTION

(a) Find the rest energy of an electron ($m = 9.109 \times 10^{-31}$ kg, $q = -e = -1.602 \times 10^{-19}$ C) in joules and in electron volts. (b) Find the speed of an electron that has been accelerated by an electric field, from rest, through a potential increase of 20.0 kV or of 5.00 MV (typical of a high-voltage x-ray machine).

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of rest energy, relativistic kinetic energy, and (from Chapter 23) electric potential energy. We use $E = mc^2$ to find the rest energy and Eqs. (37.7) and (37.38) to find the speed that gives the stated total energy.

EXECUTE: (a) The rest energy is

$$mc^2 = (9.109 \times 10^{-31}\ \text{kg})(2.998 \times 10^8\ \text{m/s})^2$$
$$= 8.187 \times 10^{-14}\ \text{J}$$

From the definition of the electron volt in Section 23.2, 1 eV = 1.602×10^{-19} J. Using this, we find

$$mc^2 = (8.187 \times 10^{-14}\ \text{J})\frac{1\ \text{eV}}{1.602 \times 10^{-19}\ \text{J}}$$
$$= 5.11 \times 10^5\ \text{eV} = 0.511\ \text{MeV}$$

(b) In calculations such as this, it is often convenient to work with the quantity $\gamma = 1/\sqrt{1 - v^2/c^2}$ from Eq. (37.38). Solving this for v, we find

$$v = c\sqrt{1 - (1/\gamma)^2}$$

The total energy E of the accelerated electron is the sum of its rest energy mc^2 and the kinetic energy eV_{ba} that it gains from the work done on it by the electric field in moving from point a to point b:

$$E = \gamma mc^2 = mc^2 + eV_{ba} \qquad \text{or}$$
$$\gamma = 1 + \frac{eV_{ba}}{mc^2}$$

An electron accelerated through a potential increase of V_{ba} = 20.0 kV gains 20.0 keV of energy, so for this electron

$$\gamma = 1 + \frac{20.0 \times 10^3\ \text{eV}}{0.511 \times 10^6\ \text{eV}} = 1.039$$

and

$$v = c\sqrt{1 - (1/1.039)^2} = 0.272c = 8.15 \times 10^7\ \text{m/s}$$

Repeating the calculation for $V_{ba} = 5.00$ MV, we find $eV_{ba}/mc^2 = 9.78$, $\gamma = 10.78$, and $v = 0.996c$.

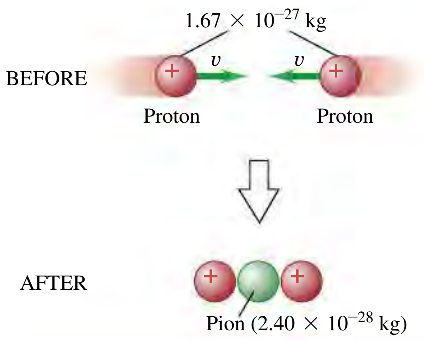
EVALUATE: With $V_{ba} = 20.0$ kV, the added kinetic energy of 20.0 keV is less than 4% of the rest energy of 0.511 MeV, and the final speed is about one-fourth the speed of light. With V_{ba} = 5.00 MV, the added kinetic energy of 5.00 MeV is much greater than the rest energy and the speed is close to c.

CAUTION **Three electron energies** All electrons have *rest* energy 0.511 MeV. An electron accelerated from rest through a 5.00-MeV potential increase has *kinetic* energy 5.00 MeV (we call it a "5.00-MeV electron") and *total* energy 5.51 MeV. Be careful to distinguish these energies from one another.

EXAMPLE 37.11 A RELATIVISTIC COLLISION

Two protons (each with mass $m_p = 1.67 \times 10^{-27}$ kg) are initially moving with equal speeds in opposite directions. They continue to exist after a head-on collision that also produces a neutral pion of mass $m_\pi = 2.40 \times 10^{-28}$ kg (**Fig. 37.22**). If all three particles are at rest after the collision, find the initial speed of the protons. Energy is conserved in the collision.

37.22 In this collision the kinetic energy of two protons is transformed into the rest energy of a new particle, a pion.



SOLUTION

IDENTIFY and SET UP: Relativistic total energy is conserved in the collision, so we can equate the (unknown) total energy of the two protons before the collision to the combined rest energies of the two protons and the pion after the collision. We then use Eq. (37.38) to find the speed of each proton.

EXECUTE: The total energy of each proton before the collision is $\gamma m_p c^2$. By conservation of energy,

$$2(\gamma m_p c^2) = 2(m_p c^2) + m_\pi c^2$$

$$\gamma = 1 + \frac{m_\pi}{2m_p} = 1 + \frac{2.40 \times 10^{-28}\ \text{kg}}{2(1.67 \times 10^{-27}\ \text{kg})} = 1.072$$

From Eq. (37.38), the initial proton speed is

$$v = c\sqrt{1 - (1/\gamma)^2} = 0.360c$$

EVALUATE: The proton rest energy is 938 MeV, so the initial kinetic energy of each proton is $(\gamma - 1)m_p c^2 = 0.072 m_p c^2 = (0.072)(938\ \text{MeV}) = 67.5$ MeV. You can verify that the rest energy $m_\pi c^2$ of the pion is twice this, or 135 MeV. All the kinetic energy "lost" in this completely inelastic collision is transformed into the rest energy of the pion.

TEST YOUR UNDERSTANDING OF SECTION 37.8 A proton is accelerated from rest by a constant force that always points in the direction of the particle's motion. Compared to the amount of kinetic energy that the proton gains during the first meter of its travel, how much kinetic energy does the proton gain during one meter of travel while it is moving at 99% of the speed of light? (i) The same amount; (ii) a greater amount; (iii) a smaller amount.

37.9 NEWTONIAN MECHANICS AND RELATIVITY

The sweeping changes required by the principle of relativity go to the very roots of Newtonian mechanics, including the concepts of length and time, the equations of motion, and the conservation principles. Thus it may appear that we have destroyed the foundations on which Newtonian mechanics is built. In one sense this is true, yet the Newtonian formulation is still accurate whenever speeds are small in comparison with the speed of light in vacuum. In such cases, time dilation, length contraction, and the modifications of the laws of motion are so small that they are unobservable. In fact, every one of the principles of Newtonian mechanics survives as a special case of the more general relativistic formulation.

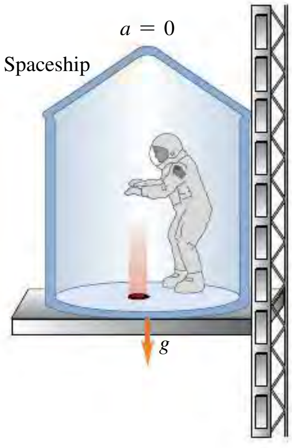
The laws of Newtonian mechanics are not *wrong;* they are *incomplete*. They are a limiting case of relativistic mechanics. They are *approximately* correct when all speeds are small in comparison to c, and they become exactly correct in the limit when all speeds approach zero. Thus relativity does not completely destroy the laws of Newtonian mechanics but *generalizes* them. This is a common pattern in the development of physical theory. Whenever a new theory is in partial conflict with an older, established theory, the new must yield the same predictions as the old in areas in which the old theory is supported by experimental evidence. Every new physical theory must pass this test, called the **correspondence principle.**

37.23 Without information from outside the spaceship, the astronaut cannot distinguish situation (b) from situation (c).

(a) An astronaut is about to drop her watch in a spaceship.

(b) In gravity-free space, the floor accelerates upward at $a = g$ and hits the watch.

(c) On the earth's surface, the watch accelerates downward at $a = g$ and hits the floor.



The General Theory of Relativity

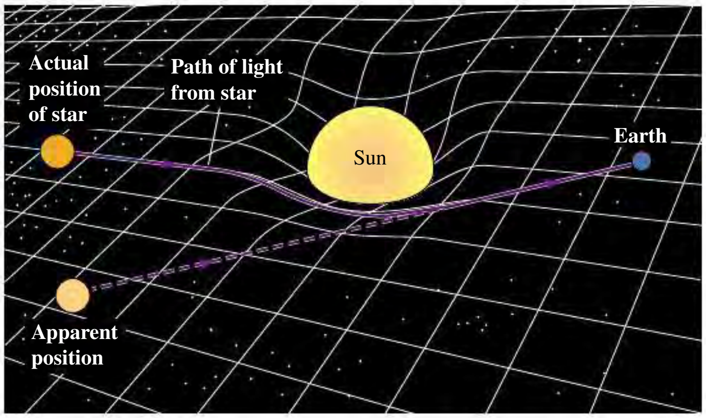
At this point we may ask whether the special theory of relativity gives the final word on mechanics or whether *further* generalizations are possible or necessary. For example, inertial frames have occupied a privileged position in our discussion. Can the principle of relativity be extended to noninertial frames as well?

Here's an example that illustrates some implications of this question. A student decides to go over Niagara Falls while enclosed in a large wooden box. During her free fall she doesn't fall to the floor of the box because both she and the box are in free fall with a downward acceleration of $9.8\ \mathrm{m/s^2}$. But an alternative interpretation, from her point of view, is that she doesn't fall to the floor because her gravitational interaction with the earth has suddenly been turned off. As long as she remains in the box and it remains in free fall, she cannot tell whether she is indeed in free fall or whether the gravitational interaction has vanished.

A similar problem occurs in a space station in orbit around the earth. Objects in the space station *seem* to be weightless, but without looking outside the station there is no way to determine whether gravity has been turned off or whether the station and all its contents are accelerating toward the center of the earth. **Figure 37.23** makes a similar point for a spaceship that is not in free fall but may be accelerating relative to an inertial frame or be at rest on the earth's surface.

These considerations form the basis of Einstein's **general theory of relativity.** If we cannot distinguish experimentally between a uniform gravitational field at a particular location and a uniformly accelerated reference frame, then there cannot be any real distinction between the two. Pursuing this concept, we may try to represent *any* gravitational field in terms of special characteristics of the coordinate system. This turns out to require even more sweeping revisions of our space-time concepts than did the special theory of relativity. In the general theory of relativity the geometric properties of space are affected by the presence of matter (**Fig. 37.24**).

The general theory of relativity has passed several experimental tests, including three proposed by Einstein. One test has to do with understanding the rotation of the axes of the planet Mercury's elliptical orbit, called the *precession of the perihelion*. (The perihelion is the point of closest approach to the sun.) A second test concerns the apparent bending of light rays from distant stars when they pass near the sun. The third test is the *gravitational red shift,* the increase in wavelength of light proceeding outward from a massive source. Some details of the general theory are more



37.24 A two-dimensional representation of curved space. We imagine the space (a plane) as being distorted as shown by a massive object (the sun). Light from a distant star (solid line) follows the distorted surface on its way to the earth. The dashed line shows the direction from which the light *appears* to be coming. The effect is greatly exaggerated; for the sun, the maximum deviation is only 0.00048°.

difficult to test, but this theory has played a central role in investigations of the formation and evolution of stars, black holes, and studies of the evolution of the universe.

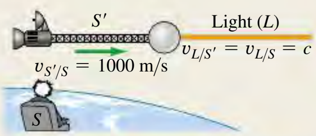
The general theory of relativity may seem to be an exotic bit of knowledge with little practical application. In fact, this theory plays an essential role in the global positioning system (GPS), which makes it possible to determine your position on the earth's surface to within a few meters using a handheld receiver (**Fig. 37.25**). The heart of the GPS system is a collection of more than two dozen satellites in very precise orbits. Each satellite emits carefully timed radio signals, and a GPS receiver simultaneously detects the signals from several satellites. The receiver then calculates the time delay between when each signal was emitted and when it was received, and uses this information to calculate the receiver's position. To ensure the proper timing of the signals, it's necessary to include corrections due to the special theory of relativity (because the satellites are moving relative to the receiver on earth) as well as the general theory (because the satellites are higher in the earth's gravitational field than the receiver). The corrections due to relativity are small—less than one part in 10^9—but are crucial to the superb precision of the GPS system.

37.25 A GPS receiver uses radio signals from the orbiting GPS satellites to determine its position. To account for the effects of relativity, the receiver must be tuned to a slightly higher frequency (10.23 MHz) than the frequency emitted by the satellites (10.22999999543 MHz).

CHAPTER 37 SUMMARY

SOLUTIONS TO ALL EXAMPLES

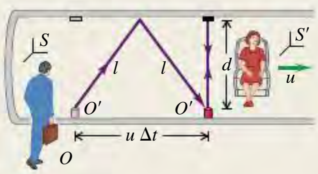
Invariance of physical laws, simultaneity: All of the fundamental laws of physics have the same form in all inertial frames of reference. The speed of light in vacuum is the same in all inertial frames and is independent of the motion of the source. Simultaneity is not an absolute concept; events that are simultaneous in one frame are not necessarily simultaneous in a second frame moving relative to the first.



Time dilation: If two events occur at the same space point in a particular frame of reference, the time interval Δt_0 between the events as measured in that frame is called a proper time interval. If this frame moves with constant velocity u relative to a second frame, the time interval Δt between the events as observed in the second frame is longer than Δt_0. (See Examples 37.1–37.3.)

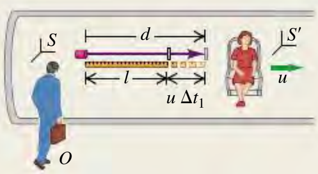
$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} = \gamma\, \Delta t_0 \qquad (37.6), (37.8)$$

$$\gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \qquad (37.7)$$



Length contraction: If two points are at rest in a particular frame of reference, the distance l_0 between the points as measured in that frame is called a proper length. If this frame moves with constant velocity u relative to a second frame and the distances are measured parallel to the motion, the distance l between the points as measured in the second frame is shorter than l_0. (See Examples 37.4 and 37.5.)

$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = \frac{l_0}{\gamma} \qquad (37.16)$$



The Lorentz transformations: The Lorentz coordinate transformations relate the coordinates and time of an event in an inertial frame S to the coordinates and time of the same event as observed in a second inertial frame S' moving at velocity u relative to the first. For one-dimensional motion, a particle's velocities v_x in S and v'_x in S' are related by the Lorentz velocity transformation. (See Examples 37.6 and 37.7.)

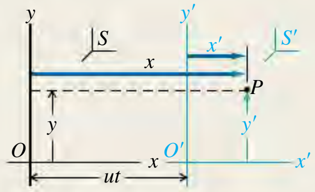
$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}} = \gamma(x - ut)$$

$$y' = y \qquad z' = z \tag{37.21}$$

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} = \gamma(t - ux/c^2)$$

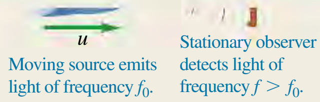
$$v'_x = \frac{v_x - u}{1 - uv_x/c^2} \tag{37.22}$$

$$v_x = \frac{v'_x + u}{1 + uv'_x/c^2} \tag{37.23}$$



The Doppler effect for electromagnetic waves: The Doppler effect is the frequency shift in light from a source due to the relative motion of source and observer. For a source moving toward the observer with speed u, Eq. (37.25) gives the received frequency f in terms of the emitted frequency f_0. (See Example 37.8.)

$$f = \sqrt{\frac{c + u}{c - u}}\,f_0 \tag{37.25}$$



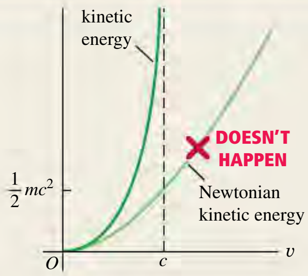
Relativistic momentum and energy: For a particle of rest mass m moving with velocity $\vec{v}$, the relativistic momentum $\vec{p}$ is given by Eq. (37.27) or (37.31) and the relativistic kinetic energy K is given by Eq. (37.36). The total energy E is the sum of the kinetic energy and the rest energy mc^2. The total energy can also be expressed in terms of the magnitude of momentum p and rest mass m. (See Examples 37.9–37.11.)

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} = \gamma m\vec{v} \qquad (37.27), (37.31)$$

$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2 \tag{37.36}$$

$$E = K + mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \gamma mc^2 \tag{37.38}$$

$$E^2 = (mc^2)^2 + (pc)^2 \tag{37.39}$$



BRIDGING PROBLEM COLLIDING PROTONS

SOLUTION

In an experiment, two protons are shot directly toward each other. Their speeds are such that in the frame of reference of each proton, the other proton is moving at $0.500c$. (a) What does an observer in the laboratory measure for the speed of each proton? (b) What is the kinetic energy of each proton as measured by an observer in the laboratory? (c) What is the kinetic energy of each proton as measured by the other proton?

SOLUTION GUIDE

IDENTIFY and SET UP

1. This problem uses the Lorentz velocity transformation, which allows us to relate the velocity v_x of a proton in one frame to its velocity v'_x in a different frame. It also uses the idea of relativistic kinetic energy.
2. Draw a sketch of the situation. Take the x-axis to be the line of motion of the protons, and take the $+x$-direction to be to the right. In the frame in which the left-hand proton is at rest, the right-hand proton has velocity $-0.500c$. In the laboratory frame the two protons have velocities $-\alpha c$ and $+\alpha c$, where α (each proton's laboratory speed as a fraction of c) is our first target variable. Given this we can find the laboratory kinetic energy of each proton.

EXECUTE

3. Write a Lorentz velocity-transformation equation that relates the velocity of the right-hand proton in the laboratory frame to its velocity in the frame of the left-hand proton. Solve this equation for α. (*Hint:* Remember that α cannot be greater than 1. Why?)
4. Use your result from step 3 to find the laboratory kinetic energy of each proton.
5. Find the kinetic energy of the right-hand proton as measured in the frame of the left-hand proton.

EVALUATE

6. How much total kinetic energy must be imparted to the protons by a scientist in the laboratory? If the experiment were to be repeated with one proton stationary, what kinetic energy would have to be given to the other proton for the collision to be equivalent?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

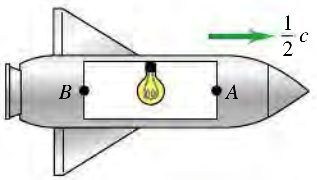
DISCUSSION QUESTIONS

Q37.1 You are standing on a train platform watching a high-speed train pass by. A light inside one of the train cars is turned on and then a little later it is turned off. (a) Who can measure the proper time interval for the duration of the light: you or a passenger on the train? (b) Who can measure the proper length of the train car: you or a passenger on the train? (c) Who can measure the proper length of a sign attached to a post on the train platform: you or a passenger on the train? In each case explain your answer.

Q37.2 If simultaneity is not an absolute concept, does that mean that we must discard the concept of causality? If event A is to *cause* event B, A must occur first. Is it possible that in some frames A appears to be the cause of B, and in others B appears to be the cause of A? Explain.

Q37.3 A rocket is moving to the right at $\frac{1}{2}$ the speed of light relative to the earth. A light bulb in the center of a room inside the rocket suddenly turns on. Call the light hitting the front end of the room event A and the light hitting the back of the room event B (**Fig. Q37.3**). Which event occurs first, A or B, or are they simultaneous, as viewed by (a) an astronaut riding in the rocket and (b) a person at rest on the earth?

Figure **Q37.3**



Q37.4 A spaceship is traveling toward the earth from the space colony on Asteroid 1040A. The ship is at the halfway point of the trip, passing Mars at a speed of $0.9c$ relative to the Mars frame of reference. At the same instant, a passenger on the spaceship receives a radio message from her boyfriend on 1040A and another from her sister on earth. According to the passenger on the ship, were these messages sent simultaneously or at different times? If at different times, which one was sent first? Explain your reasoning.

Q37.5 The average life span in the United States is about 70 years. Does this mean that it is impossible for an average person to travel a distance greater than 70 light-years away from the earth? (A light-year is the distance light travels in a year.) Explain.

Q37.6 You are holding an elliptical serving platter. How would you need to travel for the serving platter to appear round to another observer?

Q37.7 Two events occur at the same space point in a particular inertial frame of reference and are simultaneous in that frame. Is it possible that they may not be simultaneous in a different inertial frame? Explain.

Q37.8 A high-speed train passes a train platform. Larry is a passenger on the train, Adam is standing on the train platform, and David is riding a bicycle toward the platform in the same direction as the train is traveling. Compare the length of a train car as measured by Larry, Adam, and David.

Q37.9 The theory of relativity sets an upper limit on the speed that a particle can have. Are there also limits on the energy and momentum of a particle? Explain.

Q37.10 A student asserts that a material particle must always have a speed slower than that of light, and a massless particle must always move at exactly the speed of light. Is she correct? If so, how do massless particles such as photons and neutrinos acquire this speed? Can't they start from rest and accelerate? Explain.

Q37.11 The speed of light relative to still water is 2.25×10^8 m/s. If the water is moving past us, the speed of light we measure depends on the speed of the water. Do these facts violate Einstein's second postulate? Explain.

Q37.12 When a monochromatic light source moves toward an observer, its wavelength appears to be shorter than the value measured when the source is at rest. Does this contradict the hypothesis that the speed of light is the same for all observers? Explain.

Q37.13 In principle, does a hot gas have more mass than the same gas when it is cold? Explain. In practice, would this be a measurable effect? Explain.

Q37.14 Why do you think the development of Newtonian mechanics preceded the more refined relativistic mechanics by so many years?

Q37.15 What do you think would be different in everyday life if the speed of light were 10 m/s instead of 3.00×10^8 m/s?

EXERCISES

Section 37.2 Relativity of Simultaneity

37.1 • Suppose the two lightning bolts shown in Fig. 37.5a are simultaneous to an observer on the train. Show that they are *not* simultaneous to an observer on the ground. Which lightning strike does the ground observer measure to come first?

Section 37.3 Relativity of Time Intervals

37.2 • The positive muon (μ^+), an unstable particle, lives on average 2.20×10^{-6} s (measured in its own frame of reference) before decaying. (a) If such a particle is moving, with respect to the laboratory, with a speed of $0.900c$, what average lifetime is measured in the laboratory? (b) What average distance, measured in the laboratory, does the particle move before decaying?

37.3 • How fast must a rocket travel relative to the earth so that time in the rocket "slows down" to half its rate as measured by earth-based observers? Do present-day jet planes approach such speeds?

37.4 • A spaceship flies past Mars with a speed of $0.985c$ relative to the surface of the planet. When the spaceship is directly overhead, a signal light on the Martian surface blinks on and then off. An observer on Mars measures that the signal light was on for 75.0 μs. (a) Does the observer on Mars or the pilot on the spaceship measure the proper time? (b) What is the duration of the light pulse measured by the pilot of the spaceship?

37.5 • The negative pion (π^-) is an unstable particle with an average lifetime of 2.60×10^{-8} s (measured in the rest frame of the pion). (a) If the pion is made to travel at very high speed relative to a laboratory, its average lifetime is measured in the laboratory to be 4.20×10^{-7} s. Calculate the speed of the pion expressed as a fraction of c. (b) What distance, measured in the laboratory, does the pion travel during its average lifetime?

37.6 •• As you pilot your space utility vehicle at a constant speed toward the moon, a race pilot flies past you in her spaceracer at a constant speed of $0.800c$ relative to you. At the instant the spaceracer passes you, both of you start timers at zero. (a) At the instant when you measure that the spaceracer has traveled 1.20×10^8 m past you, what does the race pilot read on her timer? (b) When the race pilot reads the value calculated in part (a) on her timer, what does she measure to be your distance from her? (c) At the instant when the race pilot reads the value calculated in part (a) on her timer, what do you read on yours?

37.7 •• A spacecraft flies away from the earth with a speed of 4.80×10^6 m/s relative to the earth and then returns at the same speed. The spacecraft carries an atomic clock that has been carefully synchronized with an identical clock that remains at rest on earth. The spacecraft returns to its starting point 365 days (1 year) later, as measured by the clock that remained on earth. What is the difference in the elapsed times on the two clocks, measured in hours? Which clock, the one in the spacecraft or the one on earth, shows the shorter elapsed time?

37.8 • An alien spacecraft is flying overhead at a great distance as you stand in your backyard. You see its searchlight blink on for 0.150 s. The first officer on the spacecraft measures that the searchlight is on for 12.0 ms. (a) Which of these two measured times is the proper time? (b) What is the speed of the spacecraft relative to the earth, expressed as a fraction of the speed of light c?

Section 37.4 Relativity of Length

37.9 • A spacecraft of the Trade Federation flies past the planet Coruscant at a speed of $0.600c$. A scientist on Coruscant measures the length of the moving spacecraft to be 74.0 m. The spacecraft later lands on Coruscant, and the same scientist measures the length of the now stationary spacecraft. What value does she get?

37.10 • A meter stick moves past you at great speed. Its motion relative to you is parallel to its long axis. If you measure the length of the moving meter stick to be 1.00 ft $(1 \text{ ft} = 0.3048 \text{ m})$—for example, by comparing it to a 1-foot ruler that is at rest relative to you—at what speed is the meter stick moving relative to you?

37.11 •• **Why Are We Bombarded by Muons?** Muons are unstable subatomic particles that decay to electrons with a mean lifetime of 2.2 μs. They are produced when cosmic rays bombard the upper atmosphere about 10 km above the earth's surface, and they travel very close to the speed of light. The problem we want to address is why we see any of them at the earth's surface. (a) What is the greatest distance a muon could travel during its 2.2-μs lifetime? (b) According to your answer in part (a), it would seem that muons could never make it to the ground. But the 2.2-μs lifetime is measured in the frame of the muon, and muons are moving very fast. At a speed of $0.999c$, what is the mean lifetime of a muon as measured by an observer at rest on the earth? How far would the muon travel in this time? Does this result explain why we find muons in cosmic rays? (c) From the point of view of the muon, it still lives for only 2.2 μs, so how does it make it to the ground? What is the thickness of the 10 km of atmosphere through which the muon must travel, as measured by the muon? Is it now clear how the muon is able to reach the ground?

37.12 • An unstable particle is created in the upper atmosphere from a cosmic ray and travels straight down toward the surface of the earth with a speed of $0.99540c$ relative to the earth. A scientist at rest on the earth's surface measures that the particle is created at an altitude of 45.0 km. (a) As measured by the scientist, how much time does it take the particle to travel the 45.0 km to the surface of the earth? (b) Use the length-contraction formula to calculate the distance from where the particle is created to the surface of the earth as measured in the particle's frame. (c) In the particle's frame, how much time does it take the particle to travel from where it is created to the surface of the earth? Calculate this time both by the time dilation formula and from the distance calculated in part (b). Do the two results agree?

37.13 • As measured by an observer on the earth, a spacecraft runway on earth has a length of 3600 m. (a) What is the length of the runway as measured by a pilot of a spacecraft flying past at a speed of 4.00×10^7 m/s relative to the earth? (b) An observer on earth measures the time interval from when the spacecraft is directly over one end of the runway until it is directly over the other end. What result does she get? (c) The pilot of the spacecraft measures the time it takes him to travel from one end of the runway to the other end. What value does he get?

37.14 • A rocket ship flies past the earth at 91.0% of the speed of light. Inside, an astronaut who is undergoing a physical examination is having his height measured while he is lying down parallel to the direction in which the ship is moving. (a) If his height is measured to be 2.00 m by his doctor inside the ship, what height would a person watching this from the earth measure? (b) If the earth-based person had measured 2.00 m, what would the doctor in the spaceship have measured for the astronaut's height? Is this a reasonable height? (c) Suppose the astronaut in part (a) gets up after the examination and stands with his body perpendicular to the direction of motion. What would the doctor in the rocket and the observer on earth measure for his height now?

Section 37.5 The Lorentz Transformations

37.15 • An observer in frame S' is moving to the right ($+x$-direction) at speed $u = 0.600c$ away from a stationary observer in frame S. The observer in S' measures the speed v' of a particle moving to the right away from her. What speed v does the observer in S measure for the particle if (a) $v' = 0.400c$; (b) $v' = 0.900c$; (c) $v' = 0.990c$?

37.16 • Space pilot Mavis zips past Stanley at a constant speed relative to him of $0.800c$. Mavis and Stanley start timers at zero when the front of Mavis's ship is directly above Stanley. When Mavis reads 5.00 s on her timer, she turns on a bright light under the front of her spaceship. (a) Use the Lorentz coordinate transformation derived in Example 37.6 to calculate x and t as measured by Stanley for the event of turning on the light. (b) Use the time dilation formula, Eq. (37.6), to calculate the time interval between the two events (the front of the spaceship passing overhead and turning on the light) as measured by Stanley. Compare to the value of t you calculated in part (a). (c) Multiply the time interval by Mavis's speed, both as measured by Stanley, to calculate the distance she has traveled as measured by him when the light turns on. Compare to the value of x you calculated in part (a).

37.17 •• A pursuit spacecraft from the planet Tatooine is attempting to catch up with a Trade Federation cruiser. As measured by an observer on Tatooine, the cruiser is traveling away from the planet with a speed of $0.600c$. The pursuit ship is traveling at a speed of $0.800c$ relative to Tatooine, in the same direction as the cruiser. (a) For the pursuit ship to catch the cruiser, should the velocity of the cruiser relative to the pursuit ship be directed toward or away from the pursuit ship? (b) What is the speed of the cruiser relative to the pursuit ship?

37.18 •• An enemy spaceship is moving toward your starfighter with a speed, as measured in your frame, of $0.400c$. The enemy ship fires a missile toward you at a speed of $0.700c$ relative to the enemy ship (**Fig. E37.18**). (a) What is the speed of the missile

Figure **E37.18**

relative to you? Express your answer in terms of the speed of light. (b) If you measure that the enemy ship is 8.00×10^6 km away from you when the missile is fired, how much time, measured in your frame, will it take the missile to reach you?

37.19 •• Two particles are created in a high-energy accelerator and move off in opposite directions. The speed of one particle, as measured in the laboratory, is $0.650c$, and the speed of each particle relative to the other is $0.950c$. What is the speed of the second particle, as measured in the laboratory?

37.20 •• Two particles in a high-energy accelerator experiment are approaching each other head-on, each with a speed of $0.9380c$ as measured in the laboratory. What is the magnitude of the velocity of one particle relative to the other?

37.21 •• Two particles in a high-energy accelerator experiment approach each other head-on with a relative speed of $0.890c$. Both particles travel at the same speed as measured in the laboratory. What is the speed of each particle, as measured in the laboratory?

37.22 • An imperial spaceship, moving at high speed relative to the planet Arrakis, fires a rocket toward the planet with a speed of $0.920c$ relative to the spaceship. An observer on Arrakis measures that the rocket is approaching with a speed of $0.360c$. What is the speed of the spaceship relative to Arrakis? Is the spaceship moving toward or away from Arrakis?

Section 37.6 The Doppler Effect for Electromagnetic Waves

37.23 • **Tell It to the Judge.** (a) How fast must you be approaching a red traffic light ($\lambda = 675$ nm) for it to appear yellow ($\lambda = 575$ nm)? Express your answer in terms of the speed of light. (b) If you used this as a reason not to get a ticket for running a red light, how much of a fine would you get for speeding? Assume that the fine is \$1.00 for each kilometer per hour that your speed exceeds the posted limit of 90 km/h.

37.24 • Electromagnetic radiation from a star is observed with an earth-based telescope. The star is moving away from the earth at a speed of $0.520c$. If the radiation has a frequency of 8.64×10^{14} Hz in the rest frame of the star, what is the frequency measured by an observer on earth?

37.25 • A source of electromagnetic radiation is moving in a radial direction relative to you. The frequency you measure is 1.25 times the frequency measured in the rest frame of the source. What is the speed of the source relative to you? Is the source moving toward you or away from you?

Section 37.7 Relativistic Momentum

37.26 • **Relativistic Baseball.** Calculate the magnitude of the force required to give a 0.145-kg baseball an acceleration $a = 1.00\ \text{m/s}^2$ in the direction of the baseball's initial velocity when this velocity has a magnitude of (a) 10.0 m/s; (b) $0.900c$; (c) $0.990c$. (d) Repeat parts (a), (b), and (c) if the force and acceleration are perpendicular to the velocity.

37.27 • A proton has momentum with magnitude p_0 when its speed is $0.400c$. In terms of p_0, what is the magnitude of the proton's momentum when its speed is doubled to $0.800c$?

37.28 • **When Should You Use Relativity?** As you have seen, relativistic calculations usually involve the quantity γ. When γ is appreciably greater than 1, we must use relativistic formulas instead of Newtonian ones. For what speed v (in terms of c) is the value of γ (a) 1.0% greater than 1; (b) 10% greater than 1; (c) 100% greater than 1?

37.29 • (a) At what speed is the momentum of a particle twice as great as the result obtained from the nonrelativistic expression mv? Express your answer in terms of the speed of light. (b) A force is applied to a particle along its direction of motion. At what speed is the magnitude of force required to produce a given acceleration twice as great as the force required to produce the same acceleration when the particle is at rest? Express your answer in terms of the speed of light.

37.30 • An electron is acted upon by a force of 5.00×10^{-15} N due to an electric field. Find the acceleration this force produces in each case: (a) The electron's speed is 1.00 km/s. (b) The electron's speed is 2.50×10^8 m/s and the force is parallel to the velocity.

Section 37.8 Relativistic Work and Energy

37.31 •• What is the speed of a particle whose kinetic energy is equal to (a) its rest energy and (b) five times its rest energy?

37.32 • If a muon is traveling at $0.999c$, what are its momentum and kinetic energy? (The mass of such a muon at rest in the laboratory is 207 times the electron mass.)

37.33 • A proton (rest mass 1.67×10^{-27} kg) has total energy that is 4.00 times its rest energy. What are (a) the kinetic energy of the proton; (b) the magnitude of the momentum of the proton; (c) the speed of the proton?

37.34 •• (a) How much work must be done on a particle with mass m to accelerate it (a) from rest to a speed of $0.090c$ and (b) from a speed of $0.900c$ to a speed of $0.990c$? (Express the answers in terms of mc^2.) (c) How do your answers in parts (a) and (b) compare?

37.35 • **An Antimatter Reactor.** When a particle meets its antiparticle, they annihilate each other and their mass is converted to light energy. The United States uses approximately 1.0×10^{20} J of energy per year. (a) If all this energy came from a futuristic antimatter reactor, how much mass of matter and antimatter fuel would be consumed yearly? (b) If this fuel had the density of iron ($7.86\ \text{g/cm}^3$) and were stacked in bricks to form a cubical pile, how high would it be? (Before you get your hopes up, antimatter reactors are a *long* way in the future—if they ever will be feasible.)

37.36 •• Electrons are accelerated through a potential difference of 750 kV, so that their kinetic energy is 7.50×10^5 eV. (a) What is the ratio of the speed v of an electron having this energy to the speed of light, c? (b) What would the speed be if it were computed from the principles of classical mechanics?

37.37 • A particle has rest mass 6.64×10^{-27} kg and momentum $2.10 \times 10^{-18}\ \text{kg} \cdot \text{m/s}$. (a) What is the total energy (kinetic plus rest energy) of the particle? (b) What is the kinetic energy of the particle? (c) What is the ratio of the kinetic energy to the rest energy of the particle?

37.38 • **Creating a Particle.** Two protons (each with rest mass $M = 1.67 \times 10^{-27}$ kg) are initially moving with equal speeds in opposite directions. The protons continue to exist after a collision that also produces an η^0 particle (see Chapter 44). The rest mass of the η^0 is $m = 9.75 \times 10^{-28}$ kg. (a) If the two protons and the η^0 are all at rest after the collision, find the initial speed of the protons, expressed as a fraction of the speed of light. (b) What is the kinetic energy of each proton? Express your answer in MeV. (c) What is the rest energy of the η^0, expressed in MeV? (d) Discuss the relationship between the answers to parts (b) and (c).

37.39 • Compute the kinetic energy of a proton (mass 1.67×10^{-27} kg) using both the nonrelativistic and relativistic expressions, and compute the ratio of the two results (relativistic divided by nonrelativistic) for speeds of (a) 8.00×10^7 m/s and (b) 2.85×10^8 m/s.

37.40 • What is the kinetic energy of a proton moving at (a) $0.100c$; (b) $0.500c$; (c) $0.900c$? How much work must be done to (d) increase the proton's speed from $0.100c$ to $0.500c$ and (e) increase the proton's speed from $0.500c$ to $0.900c$? (f) How do the last two results compare to results obtained in the nonrelativistic limit?

37.41 • (a) Through what potential difference does an electron have to be accelerated, starting from rest, to achieve a speed of $0.980c$? (b) What is the kinetic energy of the electron at this speed? Express your answer in joules and in electron volts.

37.42 • The sun produces energy by nuclear fusion reactions, in which matter is converted into energy. By measuring the amount of energy we receive from the sun, we know that it is producing energy at a rate of 3.8×10^{26} W. (a) How many kilograms of matter does the sun lose each second? Approximately how many tons of matter is this (1 ton $=$ 2000 lb)? (b) At this rate, how long would it take the sun to use up all its mass?

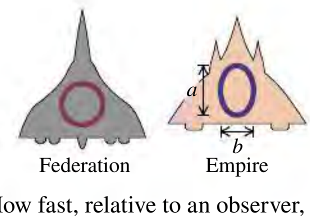
PROBLEMS

37.43 • After being produced in a collision between elementary particles, a positive pion (π^+) must travel down a 1.90-km-long tube to reach an experimental area. A π^+ particle has an average lifetime (measured in its rest frame) of 2.60×10^{-8} s; the π^+ we are considering has this lifetime. (a) How fast must the π^+ travel if it is not to decay before it reaches the end of the tube? (Since u will be very close to c, write $u = (1 - \Delta)c$ and give your answer in terms of Δ rather than u.) (b) The π^+ has a rest energy of 139.6 MeV. What is the total energy of the π^+ at the speed calculated in part (a)?

37.44 • Inside a spaceship flying past the earth at three-fourths the speed of light, a pendulum is swinging. (a) If each swing takes 1.80 s as measured by an astronaut performing an experiment inside the spaceship, how long will the swing take as measured by a person at mission control (on earth) who is watching the experiment? (b) If each swing takes 1.80 s as measured by a person at mission control, how long will it take as measured by the astronaut in the spaceship?

37.45 ••• The starships of the Solar Federation are marked with the symbol of the federation, a circle, while starships of the Denebian Empire are marked with the empire's symbol, an ellipse whose major axis is 1.40 times longer than its minor axis ($a = 1.40b$ in **Fig. P37.45**). How fast, relative to an observer, does an empire ship have to travel for its marking to be confused with the marking of a federation ship?

Figure **P37.45**



37.46 •• A cube of metal with sides of length a sits at rest in a frame S with one edge parallel to the x-axis. Therefore, in S the cube has volume a^3. Frame S' moves along the x-axis with a speed u. As measured by an observer in frame S', what is the volume of the metal cube?

37.47 •• A space probe is sent to the vicinity of the star Capella, which is 42.2 light-years from the earth. (A light-year is the distance light travels in a year.) The probe travels with a speed of $0.9930c$. An astronaut recruit on board is 19 years old when the probe leaves the earth. What is her biological age when the probe reaches Capella?

37.48 •• A muon is created 55.0 km above the surface of the earth (as measured in the earth's frame). The average lifetime of a muon, measured in its own rest frame, is 2.20 μs, and the muon we are considering has this lifetime. In the frame of the muon, the earth is moving toward the muon with a speed of $0.9860c$. (a) In the muon's frame, what is its initial height above the surface of the earth? (b) In the muon's frame, how much closer does the earth get during the lifetime of the muon? What fraction is this of the muon's original height, as measured in the muon's frame? (c) In the earth's frame, what is the lifetime of the muon? In the earth's frame, how far does the muon travel during its lifetime? What fraction is this of the muon's original height in the earth's frame?

37.49 • **The Large Hadron Collider (LHC).** Physicists and engineers from around the world came together to build the largest accelerator in the world, the Large Hadron Collider (LHC) at the CERN Laboratory in Geneva, Switzerland. The machine accelerates protons to high kinetic energies in an underground ring 27 km in circumference. (a) What is the speed v of a proton in the LHC if the proton's kinetic energy is 7.0 TeV? (Because v is very close to c, write $v = (1 - \Delta)c$ and give your answer in terms of Δ.) (b) Find the relativistic mass, m_{rel}, of the accelerated proton in terms of its rest mass.

37.50 •• The net force $\vec{F}$ on a particle of mass m is directed at 30.0° counterclockwise from the $+x$-axis. At one instant of time, the particle is traveling in the $+x$-direction with a speed (measured relative to the earth) of $0.700c$. At this instant, what is the direction of the particle's acceleration?

37.51 •• **Everyday Time Dilation.** Two atomic clocks are carefully synchronized. One remains in New York, and the other is loaded on an airliner that travels at an average speed of 250 m/s and then returns to New York. When the plane returns, the elapsed time on the clock that stayed behind is 4.00 h. By how much will the readings of the two clocks differ, and which clock will show the shorter elapsed time? (*Hint:* Since $u \ll c$, you can simplify $\sqrt{1 - u^2/c^2}$ by a binomial expansion.)

37.52 •• The distance to a particular star, as measured in the earth's frame of reference, is 7.11 light-years (1 light-year is the distance that light travels in 1 y). A spaceship leaves the earth and takes 3.35 y to arrive at the star, as measured by passengers on the ship. (a) How long does the trip take, according to observers on earth? (b) What distance for the trip do passengers on the spacecraft measure?

37.53 • CP **Čerenkov Radiation.** The Russian physicist P. A. Čerenkov discovered that a charged particle traveling in a solid with a speed exceeding the speed of light in that material radiates electromagnetic radiation. (This is analogous to the sonic boom produced by an aircraft moving faster than the speed of sound in air; see Section 16.9. Čerenkov shared the 1958 Nobel Prize for this discovery.) What is the minimum kinetic energy (in electron volts) that an electron must have while traveling inside a slab of crown glass ($n = 1.52$) in order to create this Čerenkov radiation?

37.54 •• Scientists working with a particle accelerator determine that an unknown particle has a speed of 1.35×10^8 m/s and a momentum of 2.52×10^{-19} kg·m/s. From the curvature of the particle's path in a magnetic field, they also deduce that it has a positive charge. Using this information, identify the particle.

37.55 • CP A nuclear bomb containing 12.0 kg of plutonium explodes. The sum of the rest masses of the products of the explosion is less than the original rest mass by one part in 10^4. (a) How much energy is released in the explosion? (b) If the explosion takes place in 4.00 μs, what is the average power developed by the bomb? (c) What mass of water could the released energy lift to a height of 1.00 km?

37.56 •• In the earth's rest frame, two protons are moving away from each other at equal speed. In the frame of each proton, the other proton has a speed of $0.700c$. What does an observer in the rest frame of the earth measure for the speed of each proton?

37.57 • In certain radioactive beta decay processes, the beta particle (an electron) leaves the atomic nucleus with a speed of 99.95% the speed of light relative to the decaying nucleus. If this nucleus is moving at 75.00% the speed of light in the laboratory reference frame, find the speed of the emitted electron relative to the laboratory reference frame if the electron is emitted (a) in the same direction that the nucleus is moving and (b) in the opposite direction from the nucleus's velocity. (c) In each case in parts (a) and (b), find the kinetic energy of the electron as measured in (i) the laboratory frame and (ii) the reference frame of the decaying nucleus.

37.58 •• Two events are observed in a frame of reference S to occur at the same space point, the second occurring 1.80 s after the first. In a frame S' moving relative to S, the second event is observed to occur 2.15 s after the first. What is the difference between the positions of the two events as measured in S'?

37.59 • One of the wavelengths of light emitted by hydrogen atoms under normal laboratory conditions is $\lambda = 656.3$ nm, in the red portion of the electromagnetic spectrum. In the light emitted from a distant galaxy this same spectral line is observed to be Doppler-shifted to $\lambda = 953.4$ nm, in the infrared portion of the spectrum. How fast are the emitting atoms moving relative to the earth? Are they approaching the earth or receding from it?

37.60 •• **Albert in Wonderland.** Einstein and Lorentz, being avid tennis players, play a fast-paced game on a court where they stand 20.0 m from each other. Being very skilled players, they play without a net. The tennis ball has mass 0.0580 kg. You can ignore gravity and assume that the ball travels parallel to the ground as it travels between the two players. Unless otherwise specified, all measurements are made by the two men. (a) Lorentz serves the ball at 80.0 m/s. What is the ball's kinetic energy? (b) Einstein slams a return at 1.80×10^8 m/s. What is the ball's kinetic energy? (c) During Einstein's return of the ball in part (a), a white rabbit runs beside the court in the direction from Einstein to Lorentz. The rabbit has a speed of 2.20×10^8 m/s relative to the two men. What is the speed of the rabbit relative to the ball? (d) What does the rabbit measure as the distance from Einstein to Lorentz? (e) How much time does it take for the rabbit to run 20.0 m, according to the players? (f) The white rabbit carries a pocket watch. He uses this watch to measure the time (as he sees it) for the distance from Einstein to Lorentz to pass by under him. What time does he measure?

37.61 •• **Measuring Speed by Radar.** A baseball coach uses a radar device to measure the speed of an approaching pitched baseball. This device sends out electromagnetic waves with frequency f_0 and then measures the shift in frequency Δf of the waves reflected from the moving baseball. If the fractional frequency shift produced by a baseball is $\Delta f/f_0 = 2.86 \times 10^{-7}$, what is the baseball's speed in km/h? (*Hint:* Are the waves Doppler-shifted a second time when reflected off the ball?)

37.62 •• A spaceship moving at constant speed u relative to us broadcasts a radio signal at constant frequency f_0. As the spaceship approaches us, we receive a higher frequency f; after it has passed, we receive a lower frequency. (a) As the spaceship passes by, so it is instantaneously moving neither toward nor away from us, show that the frequency we receive is not f_0, and derive an expression for the frequency we do receive. Is the frequency we receive higher or lower than f_0? (*Hint:* In this case, successive wave crests move the same distance to the observer and so they have the same transit time. Thus f equals $1/T$. Use the time dilation formula to relate the periods in the stationary and moving frames.) (b) A spaceship emits electromagnetic waves of frequency $f_0 = 345$ MHz as measured in a frame moving with the ship. The spaceship is moving at a constant speed $0.758c$ relative to us. What frequency f do we receive when the spaceship is approaching us? When it is moving away? In each case what is the shift in frequency, $f - f_0$? (c) Use the result of part (a) to calculate the frequency f and the frequency shift $(f - f_0)$ we receive at the instant that the ship passes by us. How does the shift in frequency calculated here compare to the shifts calculated in part (b)?

37.63 • **CP** In a particle accelerator a proton moves with constant speed $0.750c$ in a circle of radius 628 m. What is the net force on the proton?

37.64 •• **CP** The French physicist Armand Fizeau was the first to measure the speed of light accurately. He also found experimentally that the speed, relative to the lab frame, of light traveling in a tank of water that is itself moving at a speed V relative to the lab frame is

$$v = \frac{c}{n} + kV$$

where $n = 1.333$ is the index of refraction of water. Fizeau called k the dragging coefficient and obtained an experimental value of $k = 0.44$. What value of k do you calculate from relativistic transformations?

37.65 •• **DATA** As a research scientist at a linear accelerator, you are studying an unstable particle. You measure its mean lifetime Δt as a function of the particle's speed relative to your laboratory equipment. You record the speed of the particle u as a fraction of the speed of light in vacuum c. The table gives the results of your measurements.

u/c	0.70	0.80	0.85	0.88	0.90	0.92	0.94
Δt (10^{-8} s)	3.57	4.41	5.02	5.47	6.05	6.58	7.62

(a) Your team leader suggests that if you plot your data as $(\Delta t)^2$ versus $(1 - u^2/c^2)^{-1}$, the data points will be fit well by a straight line. Construct this graph and verify the team leader's prediction. Use the best-fit straight line to your data to calculate the mean lifetime of the particle in its rest frame. (b) What is the speed of the particle relative to your lab equipment (expressed as u/c) if the lifetime that you measure is four times its rest-frame lifetime?

37.66 •• **DATA** You are an astronomer investigating four astronomical sources of infrared radiation. You have identified the nature of each source, so you know the frequency f_0 of each when it is at rest relative to you. Your detector, which is at rest relative to the earth, measures the frequency f of the moving source. Your results are given in the table.

Source	A	B	C	D
f (THz)	7.1	5.4	6.1	8.1
f_0 (THz)	9.2	8.6	7.9	8.9

(a) Which source is moving at the highest speed relative to your detector? What is its speed? Is that source moving toward or away from the detector? (b) Which source is moving at the lowest speed relative to your detector? What is its speed? Is that source moving toward or away from the detector? (c) For source B, what frequency would your detector measure if the source were moving at the same speed relative to the detector but toward it rather than away from it?

37.67 •• DATA You are a scientist studying small aerosol particles that are contained in a vacuum chamber. The particles carry a net charge, and you use a uniform electric field to exert a constant force of 8.00×10^{-14} N on one of them. That particle moves in the direction of the exerted force. Your instruments measure the acceleration of the particle as a function of its speed v. The table gives the results of your measurements for this particular particle.

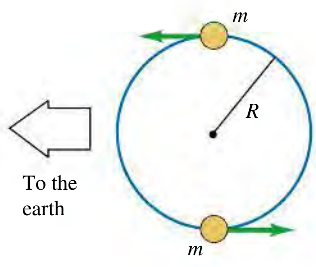
v/c	0.60	0.65	0.70	0.75	0.80	0.85
a (10^3 m/s^2)	20.3	17.9	14.8	11.2	8.5	5.9

(a) Graph your data so that the data points are well fit by a straight line. Use the slope of this line to calculate the mass m of the particle. (b) What magnitude of acceleration does the exerted force produce if the speed of the particle is 100 m/s?

CHALLENGE PROBLEMS

37.68 ••• CP Determining the Masses of Stars. Many of the stars in the sky are actually *binary stars,* in which two stars orbit about their common center of mass. If the orbital speeds of the stars are high enough, the motion of the stars can be detected by the Doppler shifts of the light they emit. Stars for which this is the case are called *spectroscopic binary stars.* **Figure P37.68** shows the simplest case of a spectroscopic binary star: two identical stars, each with mass m, orbiting their center of mass in a circle of radius R. The plane of the stars' orbits is edge-on to the line of sight of an observer on the earth. (a) The light produced by heated hydrogen gas in a laboratory on the earth has a frequency of 4.568110×10^{14} Hz. In the light received from the stars by a telescope on the earth, hydrogen light is observed to vary in frequency between 4.567710×10^{14} Hz and 4.568910×10^{14} Hz. Determine whether the binary star system as a whole is moving toward or away from the earth, the speed of this motion, and the orbital speeds of the stars. (*Hint:* The speeds involved are much less than c, so you may use the approximate result $\Delta f/f = u/c$ given in Section 37.6.) (b) The light from each star in the binary system varies from its maximum frequency to its minimum frequency and back again in 11.0 days. Determine the orbital radius R and the mass m of each star. Give your answer for m in kilograms and as a multiple of the mass of the sun, 1.99×10^{30} kg. Compare the value of R to the distance from the earth to the sun, 1.50×10^{11} m. (This technique is actually used in astronomy to determine the masses of stars. In practice, the problem is more complicated because the two stars in a binary system are usually not identical, the orbits are usually not circular, and the plane of the orbits is usually tilted with respect to the line of sight from the earth.)

Figure **P37.68**



37.69 ••• CP Kaon Production. In high-energy physics, new particles can be created by collisions of fast-moving projectile particles with stationary particles. Some of the kinetic energy of the incident particle is used to create the mass of the new particle. A proton–proton collision can result in the creation of a negative kaon (K^-) and a positive kaon (K^+):

$$p + p \rightarrow p + p + K^- + K^+$$

(a) Calculate the minimum kinetic energy of the incident proton that will allow this reaction to occur if the second (target) proton is initially at rest. The rest energy of each kaon is 493.7 MeV, and the rest energy of each proton is 938.3 MeV. (*Hint:* It is useful here to work in the frame in which the total momentum is zero. But note that the Lorentz transformation must be used to relate the velocities in the laboratory frame to those in the zero-total-momentum frame.) (b) How does this calculated minimum kinetic energy compare with the total rest mass energy of the created kaons? (c) Suppose that instead the two protons are both in motion with velocities of equal magnitude and opposite direction. Find the minimum combined kinetic energy of the two protons that will allow the reaction to occur. How does this calculated minimum kinetic energy compare with the total rest mass energy of the created kaons? (This example shows that when colliding beams of particles are used instead of a stationary target, the energy requirements for producing new particles are reduced substantially.)

37.70 ••• CP CALC Relativity and the Wave Equation. (a) Consider the Galilean transformation along the x-direction: $x' = x - vt$ and $t' = t$. In frame S the wave equation for electromagnetic waves in a vacuum is

$$\frac{\partial^2 E(x,t)}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 E(x,t)}{\partial t^2} = 0$$

where E represents the electric field in the wave. Show that by using the Galilean transformation the wave equation in frame S' is found to be

$$\left(1 - \frac{v^2}{c^2}\right)\frac{\partial^2 E(x',t')}{\partial x'^2} + \frac{2v}{c^2}\frac{\partial^2 E(x',t')}{\partial x'\partial t'} - \frac{1}{c^2}\frac{\partial^2 E(x',t')}{\partial t'^2} = 0$$

This has a different form than the wave equation in S. Hence the Galilean transformation *violates* the first relativity postulate that all physical laws have the same form in all inertial reference frames. (*Hint:* Express the derivatives $\partial/\partial x$ and $\partial/\partial t$ in terms of $\partial/\partial x'$ and $\partial/\partial t'$ by use of the chain rule.) (b) Repeat the analysis of part (a), but use the Lorentz coordinate transformations, Eqs. (37.21), and show that in frame S' the wave equation has the same form as in frame S:

$$\frac{\partial^2 E(x',t')}{\partial x'^2} - \frac{1}{c^2}\frac{\partial^2 E(x',t')}{\partial t'^2} = 0$$

Explain why this shows that the speed of light in vacuum is c in both frames S and S'.

PASSAGE PROBLEMS

SPEED OF LIGHT. Our universe has properties that are determined by the values of the fundamental physical constants, and it would be a much different place if the charge of the electron, the mass of the proton, or the speed of light was substantially different from its actual value. For instance, the speed of light is so great that the effects of relativity usually go unnoticed in everyday events. Let's imagine an alternate universe where the speed of light is 1,000,000 times less than it is in our universe to see what would happen.

37.71 An airplane has a length of 60 m when measured at rest. When the airplane is moving at 180 m/s (400 mph) in the alternate universe, how long would the plane appear to be to a stationary observer? (a) 24 m; (b) 36 m; (c) 48 m; (d) 60 m; (e) 75 m.

37.72 If the airplane of Passage Problem 37.71 has a rest mass of 20,000 kg, what is its relativistic mass when the plane is moving at 180 m/s? (a) 8000 kg; (b) 12,000 kg; (c) 16,000 kg; (d) 25,000 kg; (e) 33,300 kg.

37.73 In our universe, the rest energy of an electron is approximately 8.2×10^{-14} J. What would it be in the alternate universe? (a) 8.2×10^{-8} J; (b) 8.2×10^{-26} J; (c) 8.2×10^{-2} J; (d) 0.82 J.

37.74 In the alternate universe, how fast must an object be moving for it to have a kinetic energy equal to its rest mass? (a) 225 m/s; (b) 260 m/s; (c) 300 m/s; (d) The kinetic energy could not be equal to the rest mass.

Answers

Chapter Opening Question ?

(v) From Eq. (37.36), the relativistic expression for the kinetic energy of a particle of mass m moving at speed v is $K = (\gamma - 1)mc^2$, where $\gamma = 1/\sqrt{1 - v^2/c^2}$. If $v = 0.99000c$, $\gamma - 1 = 6.08881$; if $v = 0.99995c$, $\gamma - 1 = 99.001$, which is 16.260 times greater than the value at $v = 0.99000c$. As the speed approaches c, a relatively small increase in v corresponds to a large increase in kinetic energy (see Fig. 37.21).

Test Your Understanding Questions

37.1 (a) (i), (b) no You, too, will measure a spherical wave front that expands at the same speed c in all directions. This is a consequence of Einstein's second postulate. The wave front that you measure does *not* stay centered on the position of the moving spaceship; rather, it is centered on the point P where the spaceship was located at the instant that it emitted the light pulse. For example, suppose the spaceship is moving at speed $c/2$. When your watch shows that a time t has elapsed since the pulse of light was emitted, your measurements will show that the wave front is a sphere of radius ct centered on P and that the spaceship is a distance $ct/2$ from P.

37.2 (iii) In Mavis's frame of reference, the two events (the Ogdenville clock striking noon and the North Haverbrook clock striking noon) are not simultaneous. Figure 37.5 shows that the event toward the front of the rail car occurs first. Since the rail car is moving toward North Haverbrook, that clock struck noon before the one on Ogdenville. So, according to Mavis, it is after noon in North Haverbrook.

37.3 (a) (ii), (b) (ii) The statement that moving clocks run slow refers to any clock that is moving relative to an observer. Maria and her stopwatch are moving relative to Samir, so Samir measures Maria's stopwatch to be running slow and to have ticked off fewer seconds than his own stopwatch. Samir and his stopwatch are moving relative to Maria, so she likewise measures Samir's stopwatch to be running slow. Each observer's measurement is correct for his or her own frame of reference. *Both* observers conclude that a moving stopwatch runs slow. This is consistent with the principle of relativity (see Section 37.1), which states that the laws of physics are the same in all inertial frames of reference.

37.4 (ii), (i) and (iii) (tie), (iv) You measure both the rest length of the stationary meter stick and the contracted length of the moving spaceship to be 1 meter. The rest length of the spaceship is greater than the contracted length that you measure, and so must be greater than 1 meter. A miniature observer on board the spaceship would measure a contracted length for the meter stick of less than 1 meter. Note that in your frame of reference the nose and tail of the spaceship can simultaneously align with the two ends of the meter stick, since in your frame of reference they have the same length of 1 meter. In the spaceship's frame these two alignments cannot happen simultaneously because the meter stick is shorter than the spaceship. Section 37.2 tells us that this shouldn't be a surprise; two events that are simultaneous to one observer may not be simultaneous to a second observer moving relative to the first one.

37.5 (a) P_1, (b) P_4 (a) The last of Eqs. (37.21) tells us the times of the two events in S': $t_1' = \gamma(t_1 - ux_1/c^2)$ and $t_2' = \gamma(t_2 - ux_2/c^2)$. In frame S the two events occur at the same x-coordinate, so $x_1 = x_2$, and event P_1 occurs before event P_2, so $t_1 < t_2$. Hence you can see that $t_1' < t_2'$ and event P_1 happens before P_2 in frame S', too. This says that if event P_1 happens before P_2 in a frame of reference S where the two events occur at the same position, then P_1 happens before P_2 in any other frame moving relative to S. (b) In frame S the two events occur at different x-coordinates such that $x_3 < x_4$, and events P_3 and P_4 occur at the same time, so $t_3 = t_4$. Hence you can see that $t_3' = \gamma(t_3 - ux_3/c^2)$ is greater than $t_4' = \gamma(t_4 - ux_4/c^2)$, so event P_4 happens before P_3 in frame S'. This says that even though the two events are simultaneous in frame S, they need not be simultaneous in a frame moving relative to S.

37.7 (ii) Equation (37.27) tells us that the magnitude of momentum of a particle with mass m and speed v is $p = mv/\sqrt{1 - v^2/c^2}$. If v increases by a factor of 2, the numerator mv increases by a factor of 2 *and* the denominator $\sqrt{1 - v^2/c^2}$ decreases. Hence p increases by a factor greater than 2. (Note that in order to double the speed, the initial speed must be less than $c/2$. That's because the speed of light is the ultimate speed limit.)

37.8 (i) As the proton moves a distance s, the constant force of magnitude F does work $W = Fs$ and increases the kinetic energy by an amount $\Delta K = W = Fs$. This is true no matter what the speed of the proton before moving this distance. Thus the constant force increases the proton's kinetic energy by the same amount during the first meter of travel as during any subsequent meter of travel. (It's true that as the proton approaches the ultimate speed limit of c, the increase in the proton's *speed* is less and less with each subsequent meter of travel. That's not what the question is asking, however.)

Bridging Problem

(a) $0.268c$ **(b)** 35.6 MeV **(c)** 145 MeV

? This plastic surgeon is using two light sources: a headlamp that emits a beam of visible light and a handheld laser that emits infrared light. The light from both sources is emitted in the form of packets of energy called photons. For which source are the photons more energetic? (i) The headlamp; (ii) the laser; (iii) both are equally energetic; (iv) not enough information is given.

38 PHOTONS: LIGHT WAVES BEHAVING AS PARTICLES

LEARNING GOALS

Looking forward at ...

38.1 How Einstein's photon picture of light explains the photoelectric effect.

38.2 How experiments with x-ray production provided evidence that light is emitted in the form of photons.

38.3 How the scattering of gamma rays helped confirm the photon picture of light.

38.4 How the Heisenberg uncertainty principle imposes fundamental limits on what can be measured.

Looking back at ...

8.5 Center of mass.

16.7 Beats.

23.2 Electron volts.

32.1, 32.4 Light as an electromagnetic wave.

33.6 Light scattering.

36.2, 36.3, 36.6 Single-slit diffraction, x-ray diffraction.

37.8 Relativistic energy and momentum.

MP

PhET: Photoelectric Effect

In Chapter 32 we saw how Maxwell, Hertz, and others established firmly that light is an electromagnetic wave. Interference, diffraction, and polarization, discussed in Chapters 35 and 36, further demonstrate this *wave nature* of light.

When we look more closely at the emission, absorption, and scattering of electromagnetic radiation, however, we discover a completely different aspect of light. We find that the energy of an electromagnetic wave is *quantized;* it is emitted and absorbed in particle-like packages of definite energy, called *photons*. The energy of a single photon is proportional to the frequency of the radiation.

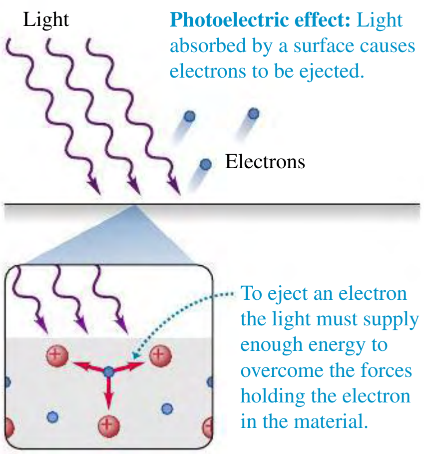
We'll find that light and other electromagnetic radiation exhibits *wave–particle duality:* Light acts sometimes like waves and sometimes like particles. Interference and diffraction demonstrate wave behavior, while emission and absorption of photons demonstrate the particle behavior. This radical reinterpretation of light will lead us in the next chapter to no less radical changes in our views of the nature of matter.

38.1 LIGHT ABSORBED AS PHOTONS: THE PHOTOELECTRIC EFFECT

A phenomenon that gives insight into the nature of light is the **photoelectric effect,** in which a material emits electrons from its surface when illuminated (**Fig. 38.1**). To escape from the surface, an electron must absorb enough energy from the incident light to overcome the attraction of positive ions in the material. These attractions constitute a potential-energy barrier; the light supplies the "kick" that enables the electron to escape.

The photoelectric effect has a number of applications. Digital cameras and night-vision scopes use it to convert light energy into an electric signal that is reconstructed into an image (**Fig. 38.2**). Sunlight striking the moon causes surface dust to eject electrons, leaving the dust particles with a positive charge. The mutual electric repulsion of these charged dust particles causes them to rise above the moon's surface, a phenomenon that was observed from lunar orbit by the *Apollo* astronauts.

38.1 The photoelectric effect.



Threshold Frequency and Stopping Potential

In Section 32.1 we explored the wave model of light, which Maxwell formulated two decades before the photoelectric effect was observed. Is the photoelectric effect consistent with this model? **Figure 38.3a** (next page) shows a modern version of one of the experiments that explored this question. Two conducting electrodes are enclosed in an evacuated glass tube and connected by a battery, and the cathode is illuminated. Depending on the potential difference V_{AC} between the two electrodes, electrons emitted by the illuminated cathode (called *photoelectrons*) may travel across to the anode, producing a *photocurrent* in the external circuit. (The tube is evacuated to a pressure of 0.01 Pa or less to minimize collisions between the electrons and gas molecules.)

The illuminated cathode emits photoelectrons with various kinetic energies. If the electric field points toward the cathode, as in Fig. 38.3a, all the electrons are accelerated toward the anode and contribute to the photocurrent. But by reversing the field and adjusting its strength as in Fig. 38.3b, we can prevent the less energetic electrons from reaching the anode. In fact, we can determine the *maximum* kinetic energy K_{max} of the emitted electrons by making the potential of the anode relative to the cathode, V_{AC}, just negative enough so that the current stops. This occurs for $V_{AC} = -V_0$, where V_0 is called the **stopping potential.** As an electron moves from the cathode to the anode, the potential decreases by V_0 and negative work $-eV_0$ is done on the (negatively charged) electron. The most energetic electron leaves the cathode with kinetic energy $K_{max} = \frac{1}{2}mv_{max}{}^2$ and has zero kinetic energy at the anode. Using the work–energy theorem, we have

$$\begin{aligned} W_{tot} &= -eV_0 = \Delta K = 0 - K_{max} \\ K_{max} &= \tfrac{1}{2}mv_{max}{}^2 = eV_0 \end{aligned} \quad \text{(maximum kinetic energy of photoelectrons)} \qquad (38.1)$$

Hence by measuring the stopping potential V_0, we can determine the maximum kinetic energy with which electrons leave the cathode. (We are ignoring any effects due to differences in the materials of the cathode and anode.)

In this experiment, how does the photocurrent depend on the voltage across the electrodes and on the frequency and intensity of the light? Based on Maxwell's picture of light as an electromagnetic wave, here is what we would predict:

Wave-Model Prediction 1: We saw in Section 32.4 that the intensity of an electromagnetic wave depends on its amplitude but not on its frequency. So the photoelectric effect should occur for light of any frequency, and *the magnitude of the photocurrent should not depend on the frequency of the light.*

Wave-Model Prediction 2: It takes a certain minimum amount of energy, called the **work function,** to eject a single electron from a particular surface (see Fig. 38.1). If the light falling on the surface is very faint, some time may elapse before the total energy absorbed by the surface equals the work function. Hence, for faint illumination, *we expect a time delay* between when we switch on the light and when photoelectrons appear.

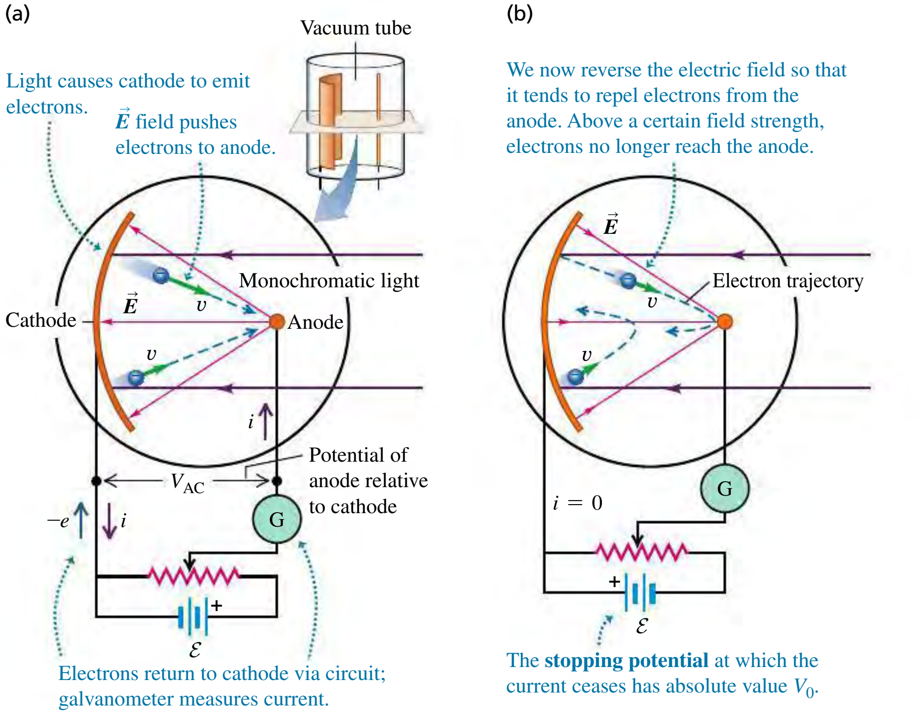
Wave-Model Prediction 3: Because the energy delivered to the cathode surface depends on the intensity of illumination, *we expect the stopping potential to increase with increasing light intensity.* Since intensity does not depend on frequency, we further expect that *the stopping potential should not depend on the frequency of the light.*

38.2 (a) A night-vision scope makes use of the photoelectric effect. Photons entering the scope strike a plate, ejecting electrons that pass through a thin disk in which there are millions of tiny channels. The current through each channel is amplified electronically and then directed toward a screen that glows when hit by electrons. (b) The image formed on the screen, which is a combination of these millions of glowing spots, is thousands of times brighter than the naked-eye view.

(a)

(b)

38.3 An experiment testing whether the photoelectric effect is consistent with the wave model of light.

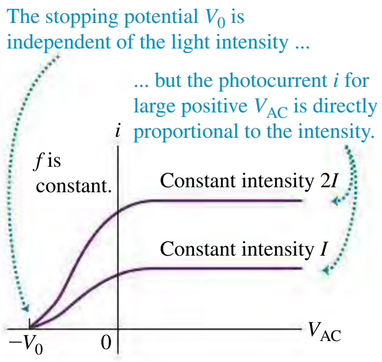


The experimental results proved to be *very* different from these predictions. Here is what was found in the years between 1877 and 1905:

Experimental Result 1: *The photocurrent depends on the light frequency.* For a given material, monochromatic light with a frequency below a minimum **threshold frequency** produces *no* photocurrent, regardless of intensity. For most metals the threshold frequency is in the ultraviolet (corresponding to wavelengths λ between 200 and 300 nm), but for other materials like potassium oxide and cesium oxide it is in the visible spectrum (λ between 380 and 750 nm).

Experimental Result 2: There is *no measurable time delay* between when the light is turned on and when the cathode emits photoelectrons (assuming the frequency of the light exceeds the threshold frequency). This is true no matter how faint the light is.

38.4 Photocurrent i for light frequency f as a function of the potential V_{AC} of the anode with respect to the cathode.



Experimental Result 3: *The stopping potential does not depend on intensity, but does depend on frequency.* **Figure 38.4** shows graphs of photocurrent as a function of potential difference V_{AC} for light of a given frequency and two different intensities. The reverse potential difference $-V_0$ needed to reduce the current to zero is the same for both intensities. The only effect of increasing the intensity is to increase the number of electrons per second and hence the photocurrent i. (The curves level off when V_{AC} is large and positive because at that point all the emitted electrons are being collected by the anode.) If the intensity is held constant but the frequency is increased, the stopping potential also increases. In other words, the greater the light frequency, the higher the energy of the ejected photoelectrons.

These results directly contradict Maxwell's description of light as an electromagnetic wave. A solution to this dilemma was provided by Albert Einstein in 1905. His proposal involved nothing less than a new picture of the nature of light.

Einstein's Photon Explanation

Einstein made the radical postulate that a beam of light consists of small packages of energy called **photons** or *quanta*. This postulate was an extension of an idea developed five years earlier by Max Planck to explain the properties of blackbody radiation, which we discussed in Section 17.7. (We'll explore

Planck's ideas in Section 39.5.) In Einstein's picture, the energy E of an individual photon is equal to a constant times the photon frequency f. From the relationship $f = c/\lambda$ for electromagnetic waves in vacuum, we have

Planck's constant; Speed of light in vacuum

Energy of a photon $$E = hf = \frac{hc}{\lambda} \qquad (38.2)$$

Frequency; Wavelength

Here **Planck's constant,** h, is a universal constant. Its numerical value, to the accuracy known at present, is

$$h = 6.62606957(29) \times 10^{-34}\ \mathrm{J \cdot s}$$

In Einstein's picture, an individual photon arriving at the surface in Fig. 38.1a or 38.2 is absorbed by a single electron. This energy transfer is an all-or-nothing process, in contrast to the continuous transfer of energy in the wave theory of light; the electron gets all of the photon's energy or none at all. The electron can escape from the surface only if the energy it acquires is greater than the work function ϕ. Thus photoelectrons will be ejected only if $hf > \phi$, or $f > \phi/h$. Einstein's postulate therefore explains why the photoelectric effect occurs only for frequencies greater than a minimum threshold frequency. This postulate is also consistent with the observation that greater intensity causes a greater photocurrent (Fig. 38.4). Greater intensity at a particular frequency means a greater number of photons per second absorbed, and thus a greater number of electrons emitted per second and a greater photocurrent.

Einstein's postulate also explains why there is no delay between illumination and the emission of photoelectrons. As soon as photons of sufficient energy strike the surface, electrons can absorb them and be ejected.

Finally, Einstein's postulate explains why the stopping potential for a given surface depends only on the light frequency. Recall that ϕ is the *minimum* energy needed to remove an electron from the surface. Einstein applied conservation of energy to find that the *maximum* kinetic energy $K_{max} = \frac{1}{2}mv_{max}{}^2$ for an emitted electron is the energy hf gained from a photon minus the work function ϕ:

$$K_{max} = \tfrac{1}{2}mv_{max}{}^2 = hf - \phi \qquad (38.3)$$

Substituting $K_{max} = eV_0$ from Eq. (38.1), we find

Maximum kinetic energy of photoelectron; Energy of absorbed photon; Work function; Light frequency

Photoelectric effect: $$eV_0 = hf - \phi \qquad (38.4)$$

Magnitude of electron charge; Stopping potential; Planck's constant

Equation (38.4) shows that the stopping potential V_0 increases with increasing frequency f. The intensity doesn't appear in Eq. (38.4), so V_0 is independent of intensity. As a check of Eq. (38.4), we can measure the stopping potential V_0 for each of several values of frequency f for a given cathode material (**Fig. 38.5**). A graph of V_0 as a function of f turns out to be a straight line, verifying Eq. (38.4), and from such a graph we can determine both the work function ϕ for the material and the value of the quantity h/e. After the electron charge $-e$ was measured by Robert Millikan in 1909, Planck's constant h could also be determined from these measurements.

Electron energies and work functions are usually expressed in electron volts (eV), defined in Section 23.2. To four significant figures,

$$1\ \mathrm{eV} = 1.602 \times 10^{-19}\ \mathrm{J}$$

CAUTION Photons are not "particles" in the usual sense It's common, but inaccurate, to envision photons as miniature billiard balls. Billiard balls have a rest mass and travel slower than the speed of light c, while photons travel at the speed of light and have *zero* rest mass. Furthermore, photons have wave aspects (frequency and wavelength) that are easy to observe. The photon concept is a very strange one, and the true nature of photons is difficult to visualize in a simple way. We'll discuss this in more detail in Section 38.4.

DATA *SPEAKS*

Photons

When students were given a problem involving photons and their properties, more than 20% gave an incorrect response. Common errors:

- Confusion about photon energy, frequency, and wavelength. The greater the frequency of a photon, the greater the photon energy and the shorter its wavelength; the longer the wavelength of a photon, the smaller the photon energy and the lower its frequency [see Eq. (38.2)].
- Confusion about the photoelectric effect. The *greater* the work function of the material, the *smaller* the kinetic energy of the electrons emitted when photons of a given frequency shine on the material [see Eq. (38.3)].

38.5 Stopping potential as a function of frequency for a particular cathode material.

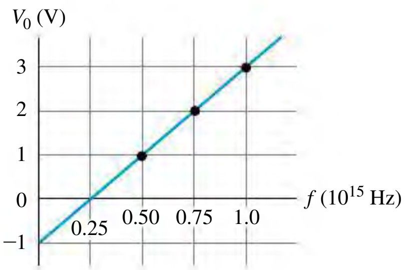
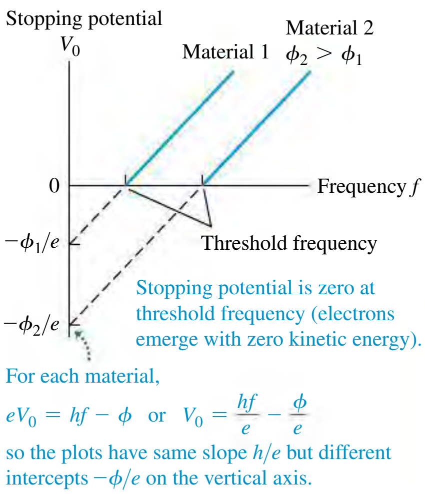


TABLE 38.1 Work Functions of Several Elements

Element	Work Function (eV)
Aluminum	4.3
Carbon	5.0
Copper	4.7
Gold	5.1
Nickel	5.1
Silicon	4.8
Silver	4.3
Sodium	2.7

38.6 Stopping potential as a function of frequency for two cathode materials having different work functions ϕ.



To this accuracy, Planck's constant is

$$h = 6.626 \times 10^{-34}\ \mathrm{J \cdot s} = 4.136 \times 10^{-15}\ \mathrm{eV \cdot s}$$

Table 38.1 lists the work functions of several elements. These values are approximate because they are very sensitive to surface impurities. The greater the work function, the higher the minimum frequency needed to emit photoelectrons (**Fig. 38.6**).

The photon picture also explains other phenomena in which light is absorbed. A *suntan* is caused when the energy in sunlight triggers a chemical reaction in skin cells that leads to increased production of the pigment melanin. This reaction can occur only if a specific molecule in the cell absorbs a certain minimum amount of energy. A short-wavelength ultraviolet photon has enough energy to trigger the reaction, but a longer-wavelength visible-light photon does not. Hence ultraviolet light causes tanning, while visible light cannot.

Photon Momentum

Einstein's photon concept applies to *all* regions of the electromagnetic spectrum, including radio waves, x rays, and so on. A photon of any frequency f and wavelength λ has energy E given by Eq. (38.2). Furthermore, according to the special theory of relativity, every particle that has energy must have momentum. Photons have zero rest mass, and a particle with zero rest mass and energy E has momentum with magnitude p given by $E = pc$ [Section 37.8; see (Eq. 37.40)]. Thus the magnitude p of the momentum of a photon is

Photon energy; Planck's constant

Momentum of a photon
$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda} \qquad (38.5)$$
Wavelength; Speed of light in vacuum; Frequency

The direction of the photon's momentum is simply the direction in which the electromagnetic wave is moving.

PROBLEM-SOLVING STRATEGY 38.1 PHOTONS

IDENTIFY *the relevant concepts:* The energy and momentum of an individual photon are proportional to the frequency and inversely proportional to the wavelength. Einstein's interpretation of the photoelectric effect is that energy is conserved as a photon ejects an electron from a material surface.

SET UP *the problem:* Identify the target variable. It could be the photon's wavelength λ, frequency f, energy E, or momentum p. If the problem involves the photoelectric effect, the target variable could be the maximum kinetic energy of photoelectrons K_{max}, the stopping potential V_0, or the work function ϕ.

EXECUTE *the solution* as follows:

1. Use Eqs. (38.2) and (38.5) to relate the energy and momentum of a photon to its wavelength and frequency. If the problem involves the photoelectric effect, use Eqs. (38.1), (38.3), and (38.4) to relate the photon frequency, stopping potential, work function, and maximum photoelectron kinetic energy.
2. The electron volt (eV), which we introduced in Section 23.2, is a convenient unit. It is the kinetic energy gained by an electron when it moves freely through an increase of potential of one volt: 1 eV $= 1.602 \times 10^{-19}$ J. If the photon energy E is given in electron volts, use $h = 4.136 \times 10^{-15}\ \mathrm{eV \cdot s}$; if E is in joules, use $h = 6.626 \times 10^{-34}\ \mathrm{J \cdot s}$.

EVALUATE *your answer:* In problems involving photons, at first the numbers will be unfamiliar to you and errors will not be obvious. It helps to remember that a visible-light photon with $\lambda = 600$ nm and $f = 5 \times 10^{14}$ Hz has an energy E of about 2 eV, or about 3×10^{-19} J.

SOLUTION

EXAMPLE 38.1 LASER-POINTER PHOTONS

A laser pointer with a power output of 5.00 mW emits red light ($\lambda = 650$ nm). (a) What is the magnitude of the momentum of each photon? (b) How many photons does the laser pointer emit each second?

SOLUTION

IDENTIFY and SET UP: This problem involves the ideas of (a) photon momentum and (b) photon energy. In part (a) we'll use Eq. (38.5) and the given wavelength to find the magnitude of each photon's momentum. In part (b), Eq. (38.5) gives the energy per photon, and the power output tells us the energy emitted per second. We can combine these quantities to calculate the number of photons emitted per second.

EXECUTE: (a) We have $\lambda = 650\text{ nm} = 6.50 \times 10^{-7}$ m, so from Eq. (38.5) the photon momentum is

$$p = \frac{h}{\lambda} = \frac{6.626 \times 10^{-34}\text{ J}\cdot\text{s}}{6.50 \times 10^{-7}\text{ m}}$$
$$= 1.02 \times 10^{-27}\text{ kg}\cdot\text{m/s}$$

(Recall that $1\text{ J} = 1\text{ kg}\cdot\text{m}^2/\text{s}^2$.)

(b) From Eq. (38.5), the energy of a single photon is

$$E = pc = (1.02 \times 10^{-27}\text{ kg}\cdot\text{m/s})(3.00 \times 10^8\text{ m/s})$$
$$= 3.06 \times 10^{-19}\text{ J} = 1.91\text{ eV}$$

The laser pointer emits energy at the rate of 5.00×10^{-3} J/s, so it emits photons at the rate of

$$\frac{5.00 \times 10^{-3}\text{ J/s}}{3.06 \times 10^{-19}\text{ J/photon}} = 1.63 \times 10^{16}\text{ photons/s}$$

EVALUATE: The result in part (a) is very small; a typical oxygen molecule in room-temperature air has 2500 times more momentum. As a check on part (b), we can calculate the photon energy from Eq. (38.2):

$$E = hf = \frac{hc}{\lambda} = \frac{(6.626 \times 10^{-34}\text{ J}\cdot\text{s})(3.00 \times 10^8\text{ m/s})}{6.50 \times 10^{-7}\text{ m}}$$
$$= 3.06 \times 10^{-19}\text{ J} = 1.91\text{ eV}$$

Our result in part (b) shows that a huge number of photons leave the laser pointer each second, each of which has an infinitesimal amount of energy. Hence the discreteness of the photons isn't noticed, and the radiated energy appears to be a continuous flow.

SOLUTION

EXAMPLE 38.2 A PHOTOELECTRIC-EFFECT EXPERIMENT

While conducting a photoelectric-effect experiment with light of a certain frequency, you find that a reverse potential difference of 1.25 V is required to reduce the current to zero. Find (a) the maximum kinetic energy and (b) the maximum speed of the emitted photoelectrons.

SOLUTION

IDENTIFY and SET UP: The value of 1.25 V is the stopping potential V_0 for this experiment. We'll use this in Eq. (38.1) to find the maximum photoelectron kinetic energy K_{max}, and from this we'll find the maximum photoelectron speed.

EXECUTE: (a) From Eq. (38.1),

$$K_{max} = eV_0 = (1.60 \times 10^{-19}\text{ C})(1.25\text{ V}) = 2.00 \times 10^{-19}\text{ J}$$

(Recall that $1\text{ V} = 1\text{ J/C}$.) In terms of electron volts,

$$K_{max} = eV_0 = e(1.25\text{ V}) = 1.25\text{ eV}$$

because the electron volt (eV) is the magnitude of the electron charge e times one volt (1 V).

(b) From $K_{max} = \frac{1}{2}mv_{max}^2$ we get

$$v_{max} = \sqrt{\frac{2K_{max}}{m}} = \sqrt{\frac{2(2.00 \times 10^{-19}\text{ J})}{9.11 \times 10^{-31}\text{ kg}}}$$
$$= 6.63 \times 10^5\text{ m/s}$$

EVALUATE: The value of v_{max} is about 0.2% of the speed of light, so we are justified in using the nonrelativistic expression for kinetic energy. (An equivalent justification is that the electron's 1.25-eV kinetic energy is much less than its rest energy $mc^2 = 0.511\text{ MeV} = 5.11 \times 10^5$ eV.)

SOLUTION

EXAMPLE 38.3 DETERMINING ϕ AND h EXPERIMENTALLY

For a particular cathode material in a photoelectric-effect experiment, you measure stopping potentials $V_0 = 1.0$ V for light of wavelength $\lambda = 600$ nm, 2.0 V for 400 nm, and 3.0 V for 300 nm. Determine the work function ϕ for this material and the implied value of Planck's constant h.

SOLUTION

IDENTIFY and SET UP: This example uses the relationship among stopping potential V_0, frequency f, and work function ϕ in the photoelectric effect. According to Eq. (38.4), a graph of V_0 versus f should be a straight line as in Fig. 38.5 or 38.6. Such a graph is completely determined by its slope and the value at which it intercepts the vertical axis; we will use these to determine the values of the target variables ϕ and h.

EXECUTE: We rewrite Eq. (38.4) as

$$V_0 = \frac{h}{e}f - \frac{\phi}{e}$$

In this form we see that the slope of the line is h/e and the vertical-axis intercept (corresponding to $f = 0$) is $-\phi/e$. The frequencies, obtained from $f = c/\lambda$ and $c = 3.00 \times 10^8$ m/s, are 0.50×10^{15} Hz, 0.75×10^{15} Hz, and 1.0×10^{15} Hz, respectively. From a graph of these data (see Fig. 38.6), we find

$$-\frac{\phi}{e} = \text{vertical intercept} = -1.0\ \text{V}$$

$$\phi = 1.0\ \text{eV} = 1.6 \times 10^{-19}\ \text{J}$$

and

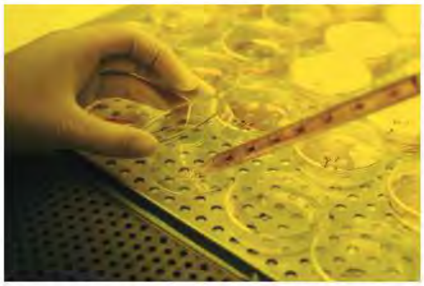
$$\text{Slope} = \frac{\Delta V_0}{\Delta f} = \frac{3.0\ \text{V} - (-1.0\ \text{V})}{1.00 \times 10^{15}\ \text{s}^{-1} - 0} = 4.0 \times 10^{-15}\ \text{J} \cdot \text{s/C}$$

$$h = \text{slope} \times e = (4.0 \times 10^{-15}\ \text{J} \cdot \text{s/C})(1.60 \times 10^{-19}\ \text{C})$$

$$= 6.4 \times 10^{-34}\ \text{J} \cdot \text{s}$$

EVALUATE: The value of Planck's constant h determined from your experiment differs from the accepted value by only about 3%. The small value $\phi = 1.0$ eV tells us that the cathode surface is not composed solely of one of the elements in Table 38.1.

BIO Application Sterilizing with High-Energy Photons One technique for killing harmful microorganisms is to illuminate them with ultraviolet light with a wavelength shorter than 254 nm. If a photon of such short wavelength strikes a DNA molecule within a microorganism, the energy of the photon is great enough to break the bonds within the molecule. This renders the microorganism unable to grow or reproduce. Such ultraviolet germicidal irradiation is used for medical sanitation, to keep laboratories sterile (as shown here), and to treat both drinking water and wastewater.



TEST YOUR UNDERSTANDING OF SECTION 38.1 Silicon films become better electrical conductors when illuminated by photons with energies of 1.14 eV or greater, an effect called *photoconductivity*. Which of the following wavelengths of electromagnetic radiation can cause photoconductivity in silicon films? (i) Ultraviolet light with $\lambda = 300$ nm; (ii) red light with $\lambda = 600$ nm; (iii) infrared light with $\lambda = 1200$ nm; (iv) both (i) and (ii); (v) all of (i), (ii), and (iii). ▌

38.2 LIGHT EMITTED AS PHOTONS: X-RAY PRODUCTION

The photoelectric effect provides convincing evidence that light is *absorbed* in the form of photons. For physicists to accept Einstein's radical photon concept, however, it was also necessary to show that light is *emitted* as photons. An experiment that demonstrates this convincingly is the inverse of the photoelectric effect: Instead of releasing electrons from a surface by shining electromagnetic radiation on it, we cause a surface to emit radiation—specifically, *x rays*—by bombarding it with fast-moving electrons.

X-Ray Photons

X rays were first produced in 1895 by the German physicist Wilhelm Röntgen, using an apparatus similar in principle to the setup shown in **Fig. 38.7**. When the cathode is heated to a very high temperature, it releases electrons in a process called *thermionic emission*. (As in the photoelectric effect, the minimum energy that an individual electron must be given to escape from the cathode's surface is equal to the work function for the surface. In this case the energy is provided to the electrons by heat rather than by light.) The electrons are then accelerated toward the anode by a potential difference V_{AC}. The bulb is evacuated (residual pressure 10^{-7} atm or less), so the electrons can travel from the cathode to the anode without colliding with air molecules. When V_{AC} is a few thousand volts or more, x rays are emitted from the anode surface.

The anode produces x rays in part simply by slowing the electrons abruptly. (Recall from Section 32.1 that accelerated charges emit electromagnetic waves.) This process is called *bremsstrahlung* (German for “braking radiation”). Because the electrons undergo accelerations of very great magnitude, they emit much of their radiation at short wavelengths in the x-ray range, about 10^{-9} to 10^{-12} m (1 nm to 1 pm). (X-ray wavelengths can be measured quite precisely by crystal diffraction techniques, which we discussed in Section 36.6.) Most electrons are braked by a series of collisions and interactions with anode atoms, so bremsstrahlung produces a continuous spectrum of electromagnetic radiation.

Just as we did for the photoelectric effect in Section 38.1, let’s compare what Maxwell’s wave theory of electromagnetic radiation would predict about this radiation to what is observed experimentally.

Wave-Model Prediction: The electromagnetic waves produced when an electron slams into the anode should be analogous to the sound waves produced by crashing cymbals together. These waves include sounds of all frequencies. By analogy, the x rays produced by bremsstrahlung should have a spectrum that includes *all* frequencies and hence *all* wavelengths.

Experimental Result: **Figure 38.8** shows bremsstrahlung spectra obtained when the same cathode and anode are used with four different accelerating voltages V_{AC}. *Not* all x-ray frequencies and wavelengths are emitted: Each spectrum has a maximum frequency f_{max} and a corresponding minimum wavelength λ_{min}. The greater the value of V_{AC}, the higher the maximum frequency and the shorter the minimum wavelength.

The wave model of electromagnetic radiation cannot explain these experimental results. But we can readily understand them by using the photon model. An electron has charge $-e$ and gains kinetic energy eV_{AC} when accelerated through a potential increase V_{AC}. The most energetic photon (highest frequency and shortest wavelength) is produced if the electron is braked to a stop all at once when it hits the anode, so that all of its kinetic energy goes to produce one photon; that is,

Kinetic energy lost by electron
Maximum energy of an emitted photon
Planck’s constant
Speed of light in vacuum

Bremsstrahlung: $$eV_{AC} = hf_{max} = \frac{hc}{\lambda_{min}} \qquad (38.6)$$

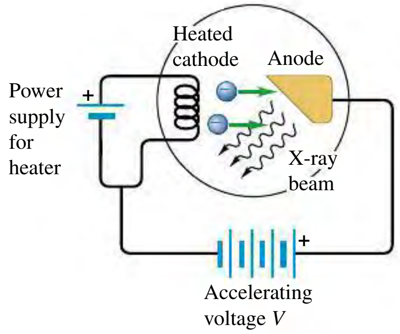
Magnitude of electron charge
Accelerating voltage
Maximum photon frequency
Minimum photon wavelength

(In this equation we ignore the work function of the target anode and the initial kinetic energy of the electrons “boiled off” from the cathode. These energies are very small compared to the kinetic energy eV_{AC} gained due to the potential difference.) If only a portion of an electron’s kinetic energy goes into producing a photon, the photon energy will be less than eV_{AC} and the wavelength will be greater than λ_{min}. Experiment shows that the measured values for λ_{min} for different values of eV_{AC} (see Fig. 38.8) agree with Eq. (38.6). Note that according to Eq. (38.6), the maximum frequency and minimum wavelength in the bremsstrahlung process do not depend on the target material; this also agrees with experiment. So we can conclude that the photon picture of electromagnetic radiation is valid for the *emission* as well as the absorption of radiation.

The apparatus shown in Fig. 38.7 can also produce x rays by a second process in which electrons transfer their kinetic energy partly or completely to individual atoms within the target. It turns out that this process not only is consistent with the photon model of electromagnetic radiation, but also provides insight into the structure of atoms. We’ll return to this process in Section 41.5.

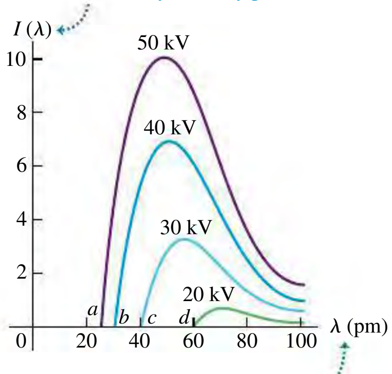
38.7 An apparatus used to produce x rays, similar to Röntgen’s 1895 apparatus.

Electrons are emitted thermionically from the heated cathode and are accelerated toward the anode; when they strike it, x rays are produced.



38.8 The continuous spectrum of x rays produced when a tungsten target is struck by electrons accelerated through a voltage V_{AC}. The curves represent different values of V_{AC}; points a, b, c, and d show the minimum wavelength for each voltage.

Vertical axis: x-ray intensity per unit wavelength



Horizontal axis: x-ray wavelength in picometers (1 pm $= 10^{-12}$ m)

EXAMPLE 38.4 PRODUCING X RAYS

Electrons in an x-ray tube accelerate through a potential difference of 10.0 kV before striking a target. If an electron produces one photon on impact with the target, what is the minimum wavelength of the resulting x rays? Find the answer by expressing energies in both SI units and electron volts.

SOLUTION

IDENTIFY and SET UP: To produce an x-ray photon with minimum wavelength and hence maximum energy, all of the electron's kinetic energy must go into producing a single x-ray photon. We'll use Eq. (38.6) to determine the wavelength.

EXECUTE: From Eq. (38.6), using SI units we have

$$\lambda_{\min} = \frac{hc}{eV_{AC}} = \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})(3.00 \times 10^{8}\ \mathrm{m/s})}{(1.602 \times 10^{-19}\ \mathrm{C})(10.0 \times 10^{3}\ \mathrm{V})}$$
$$= 1.24 \times 10^{-10}\ \mathrm{m} = 0.124\ \mathrm{nm}$$

Using electron volts, we have

$$\lambda_{\min} = \frac{hc}{eV_{AC}} = \frac{(4.136 \times 10^{-15}\ \mathrm{eV \cdot s})(3.00 \times 10^{8}\ \mathrm{m/s})}{e(10.0 \times 10^{3}\ \mathrm{V})}$$
$$= 1.24 \times 10^{-10}\ \mathrm{m} = 0.124\ \mathrm{nm}$$

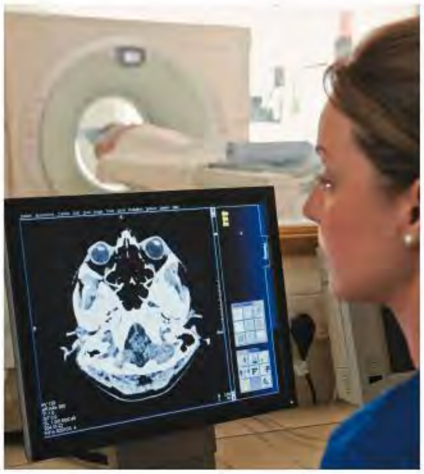
In the second calculation, the "e" for the magnitude of the electron charge cancels the "e" in the unit "eV," because the electron volt (eV) is the magnitude of the electron charge e times one volt (1 V).

EVALUATE: To check our result, recall from Example 38.1 that a 1.91-eV photon has a wavelength of 650 nm. Here the electron energy, and therefore the x-ray photon energy, is 10.0×10^3 eV = 10.0 keV, about 5000 times greater than in Example 38.1, and the wavelength is about $\frac{1}{5000}$ as great as in Example 38.1. This makes sense, since wavelength and photon energy are inversely proportional.

Applications of X Rays

X rays have many practical applications in medicine and industry. Because x-ray photons are of such high energy, they can penetrate several centimeters of solid matter. Hence they can be used to visualize the interiors of materials that are opaque to ordinary light, such as broken bones or defects in structural steel. The object to be visualized is placed between an x-ray source and an electronic detector (like that used in a digital camera). The darker an area in the image recorded by such a detector, the greater the radiation exposure. Bones are much more effective x-ray absorbers than soft tissue, so bones appear as light areas. A crack or air bubble allows greater transmission and shows as a dark area.

38.9 This radiologist is operating a CT scanner (seen through the window) from a separate room to avoid repeated exposure to x rays.



A widely used and vastly improved x-ray technique is *computed tomography*; the corresponding instrument is called a *CT scanner*. The x-ray source produces a thin, fan-shaped beam that is detected on the opposite side of the subject by an array of several hundred detectors in a line. Each detector measures absorption along a thin line through the subject. The entire apparatus is rotated around the subject in the plane of the beam, and the changing photon-counting rates of the detectors are recorded digitally. A computer processes this information and reconstructs a picture of absorption over an entire cross section of the subject (see **Fig. 38.9**). Differences in absorption as small as 1% or less can be detected with CT scans, and tumors and other anomalies that are much too small to be seen with older x-ray techniques can be detected.

X rays cause damage to living tissues. As x-ray photons are absorbed in tissues, their energy breaks molecular bonds and creates highly reactive free radicals (such as neutral H and OH), which in turn can disturb the molecular structure of proteins and especially genetic material. Young and rapidly growing cells are particularly susceptible, which is why x rays are useful for selective destruction of cancer cells. Conversely, however, a cell may be damaged by radiation but survive, continue dividing, and produce generations of defective cells; thus x rays can *cause* cancer.

Even when the organism itself shows no apparent damage, excessive exposure to x rays can cause changes in the organism's reproductive system that will affect its offspring. A careful assessment of the balance between risks and benefits of radiation exposure is essential in each individual case.

TEST YOUR UNDERSTANDING OF SECTION 38.2 In the apparatus shown in Fig. 38.7, suppose you increase the number of electrons that are emitted from the cathode per second while keeping the potential difference V_{AC} the same. How will this affect the intensity I and minimum wavelength λ_{min} of the emitted x rays? (i) I and λ_{min} will both increase; (ii) I will increase but λ_{min} will be unchanged; (iii) I will increase but λ_{min} will decrease; (iv) I will remain the same but λ_{min} will decrease; (v) none of these.

38.3 LIGHT SCATTERED AS PHOTONS: COMPTON SCATTERING AND PAIR PRODUCTION

The final aspect of light that we must test against Einstein's photon model is its behavior after the light is produced and before it is eventually absorbed. We can do this by considering the *scattering* of light. As we discussed in Section 33.6, scattering is what happens when light bounces off particles such as molecules in the air.

Compton Scattering

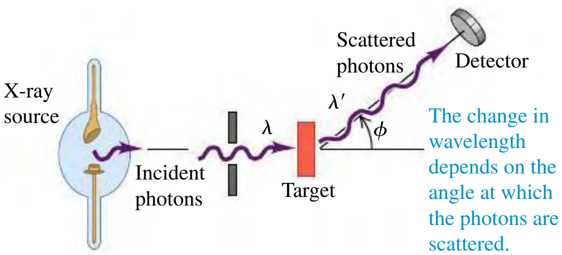
Let's see what Maxwell's wave model and Einstein's photon model predict for how light behaves when it undergoes scattering by a single electron, such as an individual electron within an atom.

Wave-Model Prediction: In the wave description, scattering would be a process of absorption and re-radiation. Part of the energy of the light wave would be absorbed by the electron, which would oscillate in response to the oscillating electric field of the wave. The oscillating electron would act like a miniature antenna (see Section 32.1), re-radiating its acquired energy as *scattered* waves in a variety of directions. The frequency at which the electron oscillates would be the same as the frequency of the incident light, and the re-radiated light would have the same frequency as the oscillations of the electron. So, *in the wave model, the scattered light and incident light have the same frequency and same wavelength*.

Photon-Model Prediction: In the photon model we imagine the scattering process as a collision of two *particles*, the incident photon and an electron that is initially at rest (**Fig. 38.10a**). The incident photon would give up part of its energy and momentum to the electron, which recoils as a result of this impact. The scattered photon that remains can fly off at a variety of angles ϕ with respect to the incident direction, but it has less energy and less momentum than the incident photon (Fig. 38.10b). The energy and momentum of a photon are given by $E = hf = hc/\lambda$ (Eq. 38.2) and $p = hf/c = h/\lambda$ (Eq. 38.5). Therefore, *in the photon model, the scattered light has a lower frequency f and longer wavelength λ than the incident light*.

The definitive experiment that tested these predictions was carried out in 1922 by the American physicist Arthur H. Compton. He aimed a beam of x rays at a solid target and measured the wavelength of the radiation scattered from the target (**Fig. 38.11**). Compton discovered that some of the scattered radiation has

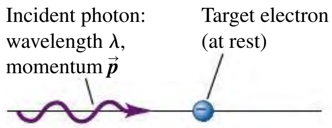
38.11 A Compton-effect experiment.



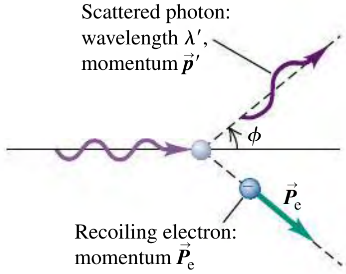
BIO Application X-Ray Absorption and Medical Imaging Atomic electrons can absorb x rays. Hence materials with many electrons per atom tend to be better x-ray absorbers than materials with few electrons. In this x-ray image the lighter areas show where x rays are absorbed as they pass through the body; the darker areas indicate regions that are relatively transparent to x rays. Bones contain large amounts of elements such as phosphorus and calcium, with 15 and 20 electrons per atom, respectively. In soft tissue, the predominant elements are hydrogen, carbon, and oxygen, with only 1, 6, and 8 electrons per atom, respectively. Hence x rays are absorbed by bone but pass relatively easily through soft tissue.

38.10 The photon model of light scattering by an electron.

(a) Before collision: The target electron is at rest.



(b) After collision: The angle between the directions of the scattered photon and the incident photon is ϕ.



smaller frequency (longer wavelength) than the incident radiation and that the change in wavelength depends on the angle through which the radiation is scattered. This is precisely what the photon model predicts for light scattered from electrons in the target, a process that is now called **Compton scattering.**

Specifically, Compton found that if the scattered radiation emerges at an angle ϕ with respect to the incident direction, as shown in Fig. 38.11, then

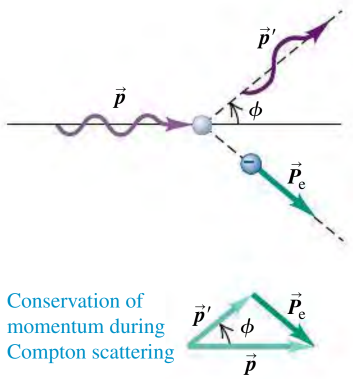
Compton scattering:
$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\phi) \tag{38.7}$$
Wavelength of scattered radiation (λ'); Wavelength of incident radiation (λ); Planck's constant (h); Scattering angle (ϕ); Electron rest mass (m); Speed of light in vacuum (c)

In other words, λ' is greater than λ. The quantity h/mc that appears in Eq. (38.7) has units of length. Its numerical value is

$$\frac{h}{mc} = \frac{6.626 \times 10^{-34}\ \mathrm{J \cdot s}}{(9.109 \times 10^{-31}\ \mathrm{kg})(2.998 \times 10^{8}\ \mathrm{m/s})} = 2.426 \times 10^{-12}\ \mathrm{m}$$

Compton showed that Einstein's photon theory, combined with the principles of conservation of energy and conservation of momentum, provides a beautifully clear explanation of his experimental results. We outline the derivation below. The electron recoil energy may be in the relativistic range, so we have to use the relativistic energy–momentum relationships, Eqs. (37.39) and (37.40). The incident photon has momentum $\vec{p}$, with magnitude p and energy pc. The scattered photon has momentum $\vec{p}\,'$, with magnitude p' and energy $p'c$. The electron is initially at rest, so its initial momentum is zero and its initial energy is its rest energy mc^2. The final electron momentum $\vec{P}_e$ has magnitude P_e, and the final electron energy is $E_e^2 = (mc^2)^2 + (P_e c)^2$. Then energy conservation gives us the relationship

$$pc + mc^2 = p'c + E_e$$



38.12 Vector diagram showing conservation of momentum in Compton scattering.

Rearranging, we find

$$(pc - p'c + mc^2)^2 = E_e^2 = (mc^2)^2 + (P_e c)^2 \tag{38.8}$$

We can eliminate the electron momentum $\vec{P}_e$ from Eq. (38.8) by using momentum conservation. From **Fig. 38.12** we see that $\vec{p} = \vec{p}\,' + \vec{P}_e$, or

$$\vec{P}_e = \vec{p} - \vec{p}\,' \tag{38.9}$$

By taking the scalar product of each side of Eq. (38.9) with itself, we find

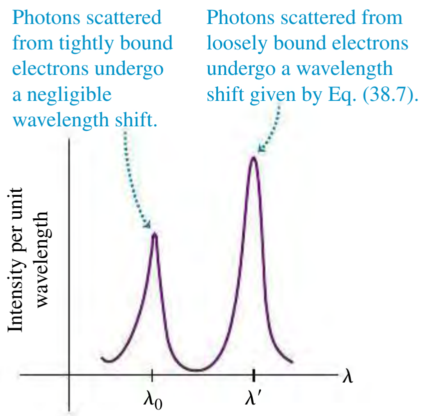
$$P_e^2 = p^2 + p'^2 - 2pp'\cos\phi \tag{38.10}$$

We now substitute this expression for P_e^2 into Eq. (38.8) and multiply out the left side. We divide out a common factor c^2; several terms cancel, and when the resulting equation is divided through by (pp'), the result is

$$\frac{mc}{p'} - \frac{mc}{p} = 1 - \cos\phi \tag{38.11}$$

Finally, we substitute $p' = h/\lambda'$ and $p = h/\lambda$, then multiply by h/mc to obtain Eq. (38.7).

38.13 Intensity as a function of wavelength for photons scattered at an angle of 135° in a Compton-scattering experiment.



When the wavelengths of x rays scattered at a certain angle are measured, the curve of intensity per unit wavelength as a function of wavelength has two peaks (**Fig. 38.13**). The longer-wavelength peak represents Compton scattering. The shorter-wavelength peak, λ_0, is at the wavelength of the incident x rays and corresponds to x-ray scattering from tightly bound electrons. In such scattering processes the entire atom must recoil, so m in Eq. (38.7) is the mass of the entire atom rather than of a single electron. The resulting wavelength shifts are negligible.

EXAMPLE 38.5 COMPTON SCATTERING

SOLUTION

You use 0.124-nm x-ray photons in a Compton-scattering experiment. (a) At what angle is the wavelength of the scattered x rays 1.0% longer than that of the incident x rays? (b) At what angle is it 0.050% longer?

SOLUTION

IDENTIFY and SET UP: We'll use the relationship between scattering angle and wavelength shift in the Compton effect. In each case our target variable is the angle ϕ (see Fig. 38.10b). We solve for ϕ by using Eq. (38.7).

EXECUTE: (a) In Eq. (38.7) we want $\Delta\lambda = \lambda' - \lambda$ to be 1.0% of 0.124 nm, so $\Delta\lambda = 0.00124\text{ nm} = 1.24 \times 10^{-12}$ m. Using the value $h/mc = 2.426 \times 10^{-12}$ m, we find

$$\Delta\lambda = \frac{h}{mc}(1 - \cos\phi)$$

$$\cos\phi = 1 - \frac{\Delta\lambda}{h/mc} = 1 - \frac{1.24 \times 10^{-12}\text{ m}}{2.426 \times 10^{-12}\text{ m}} = 0.4889$$

$$\phi = 60.7°$$

(b) For $\Delta\lambda$ to be 0.050% of 0.124 nm, or 6.2×10^{-14} m,

$$\cos\phi = 1 - \frac{6.2 \times 10^{-14}\text{ m}}{2.426 \times 10^{-12}\text{ m}} = 0.9744$$

$$\phi = 13.0°$$

EVALUATE: Our results show that smaller scattering angles give smaller wavelength shifts. Thus in a grazing collision the photon energy loss and the electron recoil energy are smaller than when the scattering angle is larger. This is just what we would expect for an elastic collision, whether between a photon and an electron or between two billiard balls.

Pair Production

Another effect that can be explained only with the photon picture involves *gamma rays,* the shortest-wavelength and highest-frequency variety of electromagnetic radiation. If a gamma-ray photon of sufficiently short wavelength is fired at a target, it may not scatter. Instead, as depicted in **Fig. 38.14**, it may disappear completely and be replaced by two new particles: an electron and a **positron** (a particle that has the same rest mass m as an electron but has a positive charge $+e$ rather than the negative charge $-e$ of the electron). This process, called **pair production,** was first observed by the physicists Patrick Blackett and Giuseppe Occhialini in 1933. The electron and positron have to be produced in pairs in order to conserve electric charge: The incident photon has zero charge, and the electron–positron pair has net charge $(-e) + (+e) = 0$. Enough energy must be available to account for the rest energy $2mc^2$ of the two particles. To four significant figures, this minimum energy is

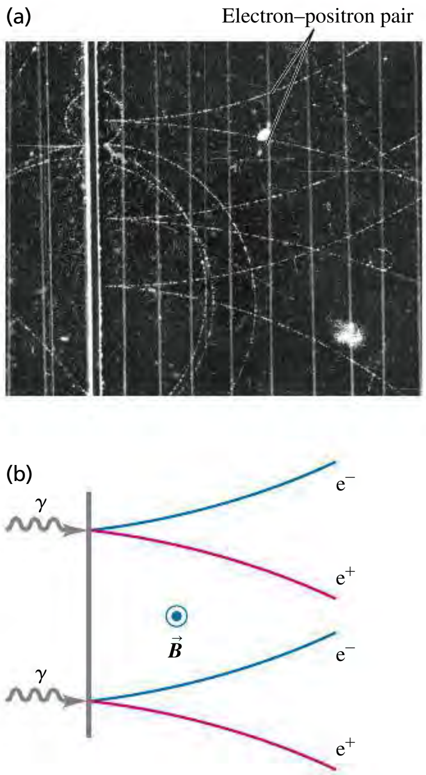
$$E_{\text{min}} = 2mc^2 = 2(9.109 \times 10^{-31}\text{ kg})(2.998 \times 10^8\text{ m/s})^2$$
$$= 1.637 \times 10^{-13}\text{ J} = 1.022\text{ MeV}$$

Thus the photon must have at least this much energy to produce an electron–positron pair. From Eq. (38.2), $E = hc/\lambda$, the photon wavelength has to be shorter than

$$\lambda_{\text{max}} = \frac{hc}{E_{\text{min}}} = \frac{(6.626 \times 10^{-34}\text{ J}\cdot\text{s})(2.998 \times 10^8\text{ m/s})}{1.637 \times 10^{-13}\text{ J}}$$
$$= 1.213 \times 10^{-12}\text{ m} = 1.213 \times 10^{-3}\text{ nm} = 1.213\text{ pm}$$

This is a very short wavelength, about $\frac{1}{1000}$ as large as the x-ray wavelengths that Compton used in his scattering experiments. (The requisite minimum photon energy is actually a bit higher than 1.022 MeV, so the photon wavelength must be a bit shorter than 1.213 pm. The reason is that when the incident photon encounters an atomic nucleus in the target, some of the photon energy goes into the kinetic energy of the recoiling nucleus.) Just as for the photoelectric effect, the wave model of electromagnetic radiation cannot explain why pair production occurs only when very short wavelengths are used.

38.14 **(a)** Photograph of bubble-chamber tracks of electron–positron pairs that are produced when 300-MeV photons strike a lead sheet. A magnetic field directed out of the photograph made the electrons (e^-) and positrons (e^+) curve in opposite directions. **(b)** Diagram showing the pair-production process for two of the gamma-ray photons (γ).



The inverse process, *electron–positron pair annihilation,* occurs when a positron and an electron collide. Both particles disappear, and two (or occasionally three) photons can appear, with total energy of at least $2m_e c^2 = 1.022$ MeV. Decay into a *single* photon is impossible because such a process could not conserve both energy and momentum. It's easiest to analyze this annihilation process in the frame of reference called the *center-of-momentum system,* in which the total momentum is zero. It is the relativistic generalization of the center-of-mass system that we discussed in Section 8.5.

EXAMPLE 38.6 PAIR ANNIHILATION

SOLUTION

An electron and a positron, initially far apart, move toward each other with the same speed. They collide head-on, annihilating each other and producing two photons. Find the energies, wavelengths, and frequencies of the photons if the initial kinetic energies of the electron and positron are (a) both negligible and (b) both 5.000 MeV. The electron rest energy is 0.511 MeV.

SOLUTION

IDENTIFY and SET UP: Just as in the elastic collisions we studied in Chapter 8, both momentum and energy are conserved in pair annihilation. The electron and positron are initially far apart, so the initial electric potential energy is zero and the initial energy is the sum of the particle kinetic and rest energies. The final energy is the sum of the photon energies. The total initial momentum is zero; the total momentum of the two photons must likewise be zero. We find the photon energy E by using conservation of energy, conservation of momentum, and the relationship $E = pc$ (see Section 38.1). We then calculate the wavelengths and frequencies from $E = hc/\lambda = hf$.

EXECUTE: If the total momentum of the two photons is to be zero, their momenta must have equal magnitudes p and opposite directions. From $E = pc = hc/\lambda = hf$, the two photons must also have the same energy E, wavelength λ, and frequency f.

Before the collision the energy of each electron is $K + mc^2$, where K is its kinetic energy and $mc^2 = 0.511$ MeV. Conservation of energy then gives

$$(K + mc^2) + (K + mc^2) = E + E$$

Hence the energy of each photon is $E = K + mc^2$.

(a) In this case the electron kinetic energy K is negligible compared to its rest energy mc^2, so each photon has energy $E = mc^2 = 0.511$ MeV. The corresponding photon wavelength and frequency are

$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(3.00 \times 10^{8}\ \text{m/s})}{0.511 \times 10^{6}\ \text{eV}}$$

$$= 2.43 \times 10^{-12}\ \text{m} = 2.43\ \text{pm}$$

$$f = \frac{E}{h} = \frac{0.511 \times 10^{6}\ \text{eV}}{4.136 \times 10^{-15}\ \text{eV}\cdot\text{s}} = 1.24 \times 10^{20}\ \text{Hz}$$

(b) In this case $K = 5.000$ MeV, so each photon has energy $E = 5.000\ \text{MeV} + 0.511\ \text{MeV} = 5.511$ MeV. Proceeding as in part (a), you can show that the photon wavelength is 0.2250 pm and the frequency is 1.333×10^{21} Hz.

EVALUATE: As a check, recall from Example 38.1 that a 650-nm visible-light photon has energy 1.91 eV and frequency 4.62×10^{14} Hz. The photon energy in part (a) is about 2.5×10^{5} times greater. As expected, the photon's wavelength is shorter and its frequency higher than those for a visible-light photon by the same factor. You can check the results for part (b) in the same way.

TEST YOUR UNDERSTANDING OF SECTION 38.3 If you used visible-light photons in the experiment shown in Fig. 38.11, would the photons undergo a wavelength shift due to the scattering? If so, is it possible to detect the shift with the human eye?

38.4 WAVE–PARTICLE DUALITY, PROBABILITY, AND UNCERTAINTY

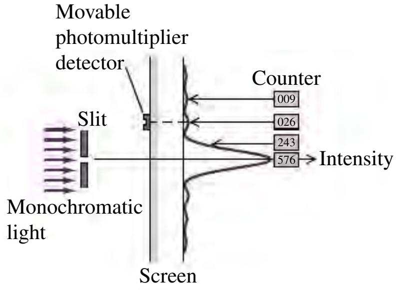
We have studied many examples of the behavior of light and other electromagnetic radiation. Some, including the interference and diffraction effects described in Chapters 35 and 36, demonstrate conclusively the *wave* nature of light. Others, the subject of the present chapter, point with equal force to the *particle* nature of light. This *wave–particle duality* means that light has two aspects that seem to be in direct conflict. How can light be a wave and a particle at the same time?

We can find the answer to this apparent wave–particle conflict in the **principle of complementarity,** first stated by the Danish physicist Niels Bohr in 1928. The wave descriptions and the particle descriptions are complementary. That is, we need both to complete our model of nature, but we will never need to use both at the same time to describe a single part of an occurrence.

Diffraction and Interference in the Photon Picture

Let's start by considering again the diffraction pattern for a single slit, which we analyzed in Sections 36.2 and 36.3. Instead of recording the pattern on a digital camera chip or photographic film, we use a detector called a *photomultiplier* that can actually detect individual photons. Using the setup shown in **Fig. 38.15**, we place the photomultiplier at various positions for equal time intervals, count the photons at each position, and plot out the intensity distribution.

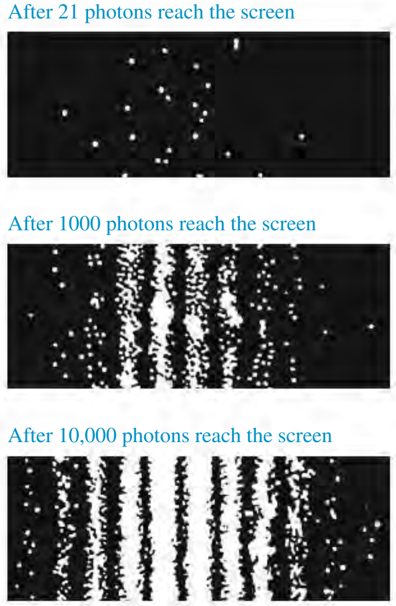
38.15 Single-slit diffraction pattern of light observed with a movable photomultiplier. The curve shows the intensity distribution predicted by the wave picture. The photon distribution is shown by the numbers of photons counted at various positions.



We find that, on average, the distribution of photons agrees with our predictions from Section 36.3. At points corresponding to the maxima of the pattern, we count many photons; at minimum points, we count almost none; and so on. The graph of the counts at various points gives the same diffraction pattern that we predicted with Eq. (36.7).

But suppose we now reduce the intensity to such a low level that only a few photons per second pass through the slit. We now record a series of discrete strikes, each representing a single photon. While we *cannot predict* where any given photon will strike, over time the accumulating strikes build up the familiar diffraction pattern we expect for a wave. To reconcile the wave and particle aspects of this pattern, we have to regard the pattern as a *statistical* distribution that tells us how many photons, on average, go to each spot. Equivalently, the pattern tells us the *probability* that any individual photon will land at a given spot. If we shine our faint light beam on a two-slit apparatus, we get an analogous result (**Fig. 38.16**). Again we can't predict exactly where an individual photon will go; the interference pattern is a statistical distribution.

38.16 These images record the positions where individual photons in a two-slit interference experiment strike the screen. As more photons reach the screen, a recognizable interference pattern appears.



How does the principle of complementarity apply to these diffraction and interference experiments? The wave description, not the particle description, explains the single- and double-slit patterns. But the particle description, not the wave description, explains why the photomultiplier records discrete packages of energy. The two descriptions complete our understanding of the results. For instance, suppose we consider an individual photon and ask how it knows "which way to go" when passing through the slit. This question seems like a conundrum, but that is because it is framed in terms of a *particle* description—whereas it is the *wave* nature of light that determines the distribution of photons. Conversely, the fact that the photomultiplier detects faint light as a sequence of individual "spots" can't be explained in wave terms.

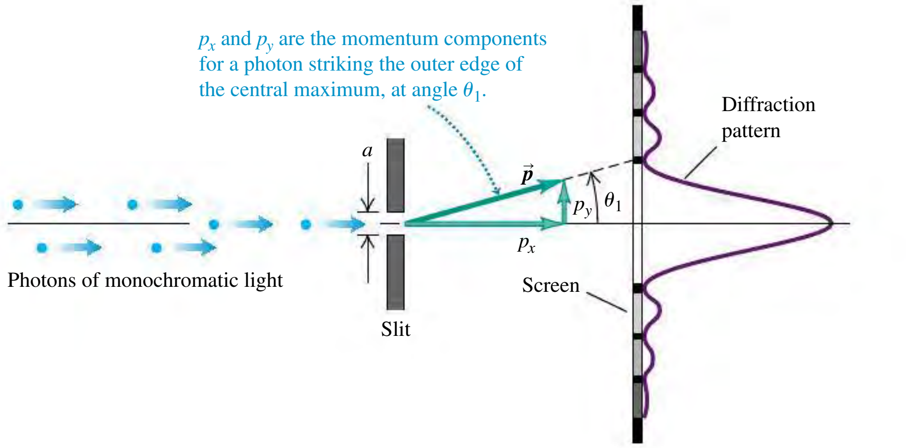
Probability and Uncertainty

Although photons have energy and momentum, they are nonetheless very different from the particle model we used for Newtonian mechanics in Chapters 4 through 8. The Newtonian particle model treats an object as a point mass. We can describe the location and state of motion of such a particle at any instant with three spatial coordinates and three components of momentum, and we can then predict the particle's future motion. This model doesn't work at all for photons, however: We *cannot* treat a photon as a point object. This is because there are fundamental limitations on the precision with which we can simultaneously determine the position and momentum of a photon. Many aspects of a photon's behavior can be stated only in terms of *probabilities.* (In Chapter 39 we will find that the non-Newtonian ideas we develop for photons in this section also apply to particles such as electrons.)

To get more insight into the problem of measuring a photon's position and momentum simultaneously, let's look again at the single-slit diffraction of light.

MP

PhET: Fourier: Making Waves
PhET: Quantum Wave Interference



38.17 Interpreting single-slit diffraction in terms of photon momentum.

Suppose the wavelength λ is much less than the slit width a (**Fig. 38.17**). Then most (85%) of the photons go into the central maximum of the diffraction pattern, and the remainder go into other parts of the pattern. We use θ_1 to denote the angle between the central maximum and the first minimum. Using Eq. (36.2) with $m = 1$, we find that θ_1 is given by $\sin\theta_1 = \lambda/a$. Since we assume $\lambda \ll a$, it follows that θ_1 is very small, $\sin\theta_1$ is very nearly equal to θ_1 (in radians), and

$$\theta_1 = \frac{\lambda}{a} \tag{38.12}$$

Even though the photons all have the same initial state of motion, they don't all follow the same path. We can't predict the exact trajectory of any individual photon from knowledge of its initial state; we can only describe the *probability* that an individual photon will strike a given spot on the screen. This fundamental indeterminacy has no counterpart in Newtonian mechanics.

Furthermore, there are fundamental *uncertainties* in both the position and the momentum of an individual particle, and these uncertainties are related inseparably. To clarify this point, let's go back to Fig. 38.17. A photon that strikes the screen at the outer edge of the central maximum, at angle θ_1, must have a component of momentum p_y in the y-direction, as well as a component p_x in the x-direction, despite the fact that initially the beam was directed along the x-axis. From the geometry of the situation the two components are related by $p_y/p_x = \tan\theta_1$. Since θ_1 is small, we may use the approximation $\tan\theta_1 = \theta_1$, and

$$p_y = p_x\theta_1 \tag{38.13}$$

Substituting Eq. (38.12), $\theta_1 = \lambda/a$, into Eq. (38.13) gives

$$p_y = p_x\frac{\lambda}{a} \tag{38.14}$$

Equation (38.14) says that for the 85% of the photons that strike the detector within the central maximum (that is, at angles between $-\lambda/a$ and $+\lambda/a$), the y-component of momentum is spread out over a range from $-p_x\lambda/a$ to $+p_x\lambda/a$. Now let's consider *all* the photons that pass through the slit and strike the screen. Again, they may hit above or below the center of the pattern, so their component p_y may be positive or negative. However the symmetry of the diffraction pattern shows us the average value $(p_y)_{\text{av}} = 0$. There will be an *uncertainty* Δp_y in the y-component of momentum at least as great as $p_x\lambda/a$. That is,

$$\Delta p_y \geq p_x\frac{\lambda}{a} \tag{38.15}$$

The narrower the slit width a, the broader is the diffraction pattern and the greater is the uncertainty in the y-component of momentum p_y.

The photon wavelength λ is related to the momentum p_x by Eq. (38.5), which we can rewrite as $\lambda = h/p_x$. Using this relationship in Eq. (38.15) and simplifying, we find

$$\Delta p_y \geq p_x \frac{h}{p_x a} = \frac{h}{a}$$

$$\Delta p_y a \geq h \tag{38.16}$$

What does Eq. (38.16) mean? The slit width a represents an uncertainty in the y-component of the *position* of a photon as it passes through the slit. We don't know exactly *where* in the slit each photon passes through. So both the y-position and the y-component of momentum have uncertainties, and the two uncertainties are related by Eq. (38.16). We can reduce the *momentum* uncertainty Δp_y only by reducing the width of the diffraction pattern. To do this, we have to increase the slit width a, which increases the *position* uncertainty. Conversely, when we *decrease* the position uncertainty by narrowing the slit, the diffraction pattern broadens and the corresponding momentum uncertainty *increases*.

You may protest that it doesn't seem to be consistent with common sense for a photon not to have a definite position and momentum. But what we call *common sense* is based on familiarity gained through experience. Our usual experience includes very little contact with the microscopic behavior of particles like photons. Sometimes we have to accept conclusions that violate our intuition when we are dealing with areas that are far removed from everyday experience.

The Uncertainty Principle

In more general discussions of uncertainty relationships, the uncertainty of a quantity is usually described in terms of the statistical concept of *standard deviation,* which is a measure of the spread or dispersion of a set of numbers around their average value. Suppose we now begin to describe uncertainties in this way [neither Δp_y nor a in Eq. (38.16) is a standard deviation]. If a coordinate x has an uncertainty Δx and if the corresponding momentum component p_x has an uncertainty Δp_x, then we find that in general

Heisenberg uncertainty principle for position and momentum:

Uncertainty in coordinate x

Planck's constant divided by 2π

$$\Delta x \Delta p_x \geq \hbar/2 \tag{38.17}$$

Uncertainty in corresponding momentum component p_x

CAUTION **h versus h-bar** It's common for students to plug in the value of h when what they wanted was $\hbar = h/2\pi$, or vice versa. Don't make the same mistake, or your answer will be off by a factor of 2π!

The quantity $\hbar$ (pronounced "h-bar") is Planck's constant divided by 2π:

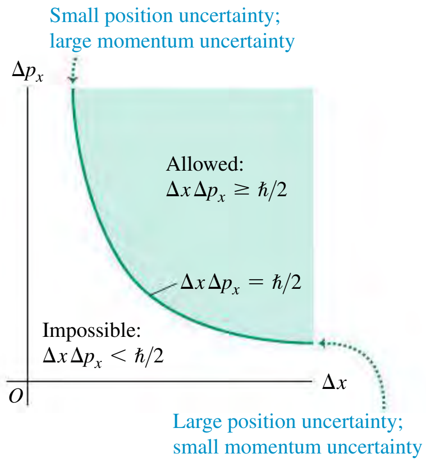
$$\hbar = \frac{h}{2\pi} = 1.054571628(53) \times 10^{-34}\ \mathrm{J \cdot s}$$

We will use $\hbar$ frequently to avoid writing a lot of factors of 2π in later equations.

Equation (38.17) is one form of the **Heisenberg uncertainty principle,** first discovered by the German physicist Werner Heisenberg (1901–1976). It states that, in general, it is impossible to simultaneously determine both the position and the momentum of a particle with arbitrarily great precision, as classical physics would predict. Instead, the uncertainties in the two quantities play complementary roles, as we have described. **Figure 38.18** shows the relationship between the two uncertainties. Our derivation of Eq. (38.16), a less refined form of the uncertainty principle given by Eq. (38.17), shows that this principle has its roots in the wave aspect of photons. We will see in Chapter 39 that electrons and other subatomic particles also have a wave aspect, and the same uncertainty principle applies to them as well.

It is tempting to suppose that we could get greater precision by using more sophisticated detectors of position and momentum. This turns out not to be possible. To detect a particle, the detector must *interact* with it, and this interaction

38.18 The Heisenberg uncertainty principle for position and momentum components. It is impossible for the product $\Delta x \Delta p_x$ to be less than $\hbar/2 = h/4\pi$.



Application Butterfly Hunting with Heisenberg Because $\hbar$ has such a small value, the Heisenberg uncertainty principle comes into play only for objects on the scale of atoms or smaller. To visualize what this principle means, imagine that we could make the value of $\hbar$ larger by a factor of 10^{34} so that $\hbar = 1.05 \text{ J} \cdot \text{s}$. If you trap a butterfly in a butterfly net, you know the butterfly's position to within the 0.25-m diameter of the net. Then the uncertainty in the butterfly's position is approximately $\Delta x = 0.25$ m. The minimum uncertainty in its momentum is then $\Delta p_x = (\hbar/2\,\Delta x) = (1.05 \text{ J} \cdot \text{s})/2(0.25 \text{ m}) = 2.1 \text{ kg} \cdot \text{m/s}$, so just by trapping the butterfly you could impart this much momentum to it. A typical butterfly has a mass of about 3×10^{-4} kg. With this much momentum, the butterfly's speed would be about 7000 m/s (about 20 times the speed of sound!) and its kinetic energy about 7000 J (that of a baseball traveling at about 300 m/s, just under the speed of sound). By confining the butterfly in the net, you could give it so much momentum and kinetic energy that it could burst out of the net!

unavoidably changes the state of motion of the particle, introducing uncertainty about its original state. For example, we could imagine placing an electron at a certain point in the middle of the slit in Fig. 38.17. If the photon passes through the middle, we would see the electron recoil. We would then know that the photon passed through that point in the slit, and we would be much more certain about the x-coordinate of the photon. However, the collision between the photon and the electron would change the photon momentum, giving us greater uncertainty in the value of that momentum. A more detailed analysis of such hypothetical experiments shows that the uncertainties we have described are fundamental and intrinsic. They *cannot* be circumvented *even in principle* by any experimental technique, no matter how sophisticated.

There is nothing special about the x-axis. In a three-dimensional situation with coordinates (x, y, z) there is an uncertainty relationship for each coordinate and its corresponding momentum component: $\Delta x \Delta p_x \geq \hbar/2$, $\Delta y \Delta p_y \geq \hbar/2$, and $\Delta z \Delta p_z \geq \hbar/2$. However, the uncertainty in one coordinate is *not* related to the uncertainty in a different component of momentum. For example, Δx is not related directly to Δp_y.

Waves and Uncertainty

Here's an alternative way to understand the Heisenberg uncertainty principle in terms of the properties of waves. Consider a sinusoidal electromagnetic wave propagating in the positive x-direction with its electric field polarized in the y-direction. If the wave has wavelength λ, frequency f, and amplitude A, we can write the wave function as

$$E_y(x, t) = A\sin(kx - \omega t) \tag{38.18}$$

In this expression the wave number is $k = 2\pi/\lambda$ and the angular frequency is $\omega = 2\pi f$. We can think of the wave function in Eq. (38.18) as a description of a photon with a definite wavelength and a definite frequency. In terms of k and ω we can express the momentum and energy of the photon as

$$p_x = \frac{h}{\lambda} = \frac{h}{2\pi}\frac{2\pi}{\lambda} = \hbar k \quad \text{(photon momentum in terms of wave number)} \tag{38.19a}$$

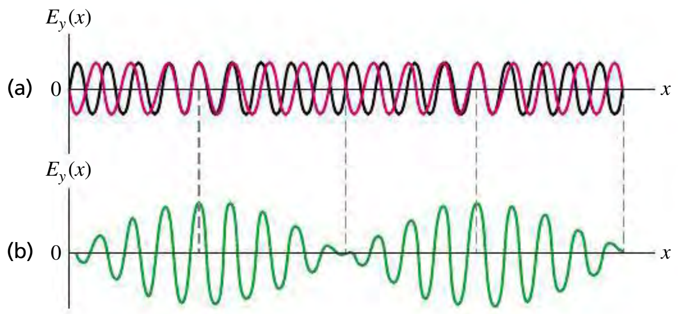
$$E = hf = \frac{h}{2\pi}2\pi f = \hbar\omega \quad \text{(photon energy in terms of angular frequency)} \tag{38.19b}$$

Using Eqs. (38.19) in Eq. (38.18), we can rewrite our photon wave equation as

$$E_y(x, t) = A\sin[(p_x x - Et)/\hbar] \quad \text{(wave function for a photon with } x\text{-momentum } p_x \text{ and energy } E) \tag{38.20}$$

Since this wave function has a definite value of x-momentum p_x, there is *no* uncertainty in the value of this quantity: $\Delta p_x = 0$. The Heisenberg uncertainty principle, Eq. (38.17), says that $\Delta x \Delta p_x \geq \hbar/2$. If Δp_x is zero, then Δx must be infinite. Indeed, the wave described by Eq. (38.20) extends along the entire x-axis and has the same amplitude everywhere. The price we pay for knowing the photon's momentum precisely is that we have no idea *where* the photon is!

In practical situations we always have *some* idea where a photon is. To describe this situation, we need a wave function that is more localized in space. We can create one by superimposing two or more sinusoidal functions. To keep things simple, we'll consider only waves propagating in the positive x-direction. For example, let's add together two sinusoidal wave functions like those in Eqs. (38.18) and (38.20), but with slightly different wavelengths and frequencies and hence slightly different values p_{x1} and p_{x2} of x-momentum and slightly



38.19 (a) Two sinusoidal waves with slightly different wave numbers k and hence slightly different values of momentum $p_x = \hbar k$ shown at one instant of time. (b) The superposition of these waves has a momentum equal to the average of the two individual values of momentum. The amplitude varies, giving the total wave a lumpy character not possessed by either individual wave.

different values E_1 and E_2 of energy. The total wave function is

$$E_y(x, t) = A_1 \sin[(p_{1x}x - E_1 t)/\hbar] + A_2 \sin[(p_{2x}x - E_2 t)/\hbar] \quad (38.21)$$

Consider what this wave function looks like at a particular instant of time, say, $t = 0$. At this instant Eq. (38.21) becomes

$$E_y(x, t = 0) = A_1 \sin(p_{1x}x/\hbar) + A_2 \sin(p_{2x}x/\hbar) \quad (38.22)$$

Figure 38.19a is a graph of the individual wave functions at $t = 0$ for the case $A_2 = -A_1$, and Fig. 38.19b graphs the combined wave function $E_y(x, t = 0)$ given by Eq. (38.22). We saw something very similar to Fig. 38.19b in our discussion of beats in Section 16.7: When we superimposed two sinusoidal waves with slightly different frequencies (see Fig. 16.25), the resulting wave exhibited amplitude variations not present in the original waves. In the same way, a photon represented by the wave function in Eq. (38.21) is most likely to be found in the regions where the wave function's amplitude is greatest. That is, the photon is *localized*. However, the photon's momentum no longer has a definite value because we began with two different x-momentum values, p_{x1} and p_{x2}. This agrees with the Heisenberg uncertainty principle: By decreasing the uncertainty in the photon's position, we have increased the uncertainty in its momentum.

Uncertainty in Energy

Our discussion of combining waves also shows that there is an uncertainty principle that involves *energy* and *time*. To see why this is so, imagine measuring the combined wave function described by Eq. (38.21) at a certain position, say $x = 0$, over a period of time. At $x = 0$, the wave function from Eq. (38.21) becomes

$$\begin{aligned} E_y(x, t) &= A_1 \sin(-E_1 t/\hbar) + A_2 \sin(-E_2 t/\hbar) \\ &= -A_1 \sin(E_1 t/\hbar) - A_2 \sin(E_2 t/\hbar) \end{aligned} \quad (38.23)$$

What we measure at $x = 0$ is a combination of two oscillating electric fields with slightly different angular frequencies $\omega_1 = E_1/\hbar$ and $\omega_2 = E_2/\hbar$. This is exactly the phenomenon of beats that we discussed in Section 16.7 (compare Fig. 16.25). The amplitude of the combined field rises and falls, so the photon described by this field is localized in *time* as well as in position. The photon is most likely to be found at the times when the amplitude is large. The price we pay for localizing the photon in time is that the wave does not have a definite energy. By contrast, if the photon is described by a sinusoidal wave like that in Eq. (38.20) that *does* have a definite energy E but that has the same amplitude at all times, we have no idea when the photon will appear at $x = 0$. So the better we know the photon's energy, the less certain we are of when we will observe the photon.

Just as for the momentum–position uncertainty principle, we can write a mathematical expression for the uncertainty principle that relates energy and time. In fact, except for an overall minus sign, Eq. (38.23) is identical to Eq. (38.22) if we replace the x-momentum p_x by energy E and the position x by time t.

This tells us that in the momentum–position uncertainty relation, Eq. (38.17), we can replace the momentum uncertainty Δp_x with the energy uncertainty ΔE and replace the position uncertainty Δx with the time uncertainty Δt. The result is

Heisenberg uncertainty principle for energy and time:

Time uncertainty of a phenomenon

Planck's constant divided by 2π

$$\Delta t \, \Delta E \geq \hbar/2 \tag{38.24}$$

Energy uncertainty of same phenomenon

In practice, any real photon has a limited spatial extent and hence passes any point in a limited amount of time. The following example illustrates how this affects the momentum and energy of the photon.

EXAMPLE 38.7 ULTRASHORT LASER PULSES AND THE UNCERTAINTY PRINCIPLE

SOLUTION

Many varieties of lasers emit light in the form of pulses rather than a steady beam. A tellurium–sapphire laser can produce light at a wavelength of 800 nm in ultrashort pulses that last only 4.00×10^{-15} s (4.00 femtoseconds, or 4.00 fs). The energy in a single pulse produced by one such laser is $2.00\,\mu\text{J} = 2.00 \times 10^{-6}$ J, and the pulses propagate in the positive x-direction. Find (a) the frequency of the light; (b) the energy and minimum energy uncertainty of a single photon in the pulse; (c) the minimum frequency uncertainty of the light in the pulse; (d) the spatial length of the pulse, in meters and as a multiple of the wavelength; (e) the momentum and minimum momentum uncertainty of a single photon in the pulse; and (f) the approximate number of photons in the pulse.

SOLUTION

IDENTIFY and SET UP: It's important to distinguish between the light pulse as a whole (which contains a very large number of photons) and an individual photon within the pulse. The 4.00-fs pulse duration represents the time it takes the pulse to emerge from the laser; it is also the time *uncertainty* for an individual photon within the pulse, since we don't know when during the pulse that photon emerges. Similarly, the position uncertainty of a photon is the spatial length of the pulse, since a given photon could be found anywhere within the pulse. To find our target variables, we'll use the relationships for photon energy and momentum from Section 38.1 and the two Heisenberg uncertainty principles, Eqs. (38.17) and (38.24).

EXECUTE: (a) From the relationship $c = \lambda f$, the frequency of 800-nm light is

$$f = \frac{c}{\lambda} = \frac{3.00 \times 10^{8}\ \text{m/s}}{8.00 \times 10^{-7}\ \text{m}} = 3.75 \times 10^{14}\ \text{Hz}$$

(b) From Eq. (38.2) the energy of a single 800-nm photon is

$$E = hf = (6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(3.75 \times 10^{14}\ \text{Hz})$$
$$= 2.48 \times 10^{-19}\ \text{J}$$

The time uncertainty equals the pulse duration, $\Delta t = 4.00 \times 10^{-15}$ s. From Eq. (38.24) the minimum uncertainty in energy corresponds to the case $\Delta t \, \Delta E = \hbar/2$, so

$$\Delta E = \frac{\hbar}{2\Delta t} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(4.00 \times 10^{-15}\ \text{s})} = 1.32 \times 10^{-20}\ \text{J}$$

This is 5.3% of the photon energy $E = 2.48 \times 10^{-19}$ J, so the energy of a given photon is uncertain by at least 5.3%. The uncertainty could be greater, depending on the shape of the pulse.

(c) From the relationship $f = E/h$, the minimum frequency uncertainty is

$$\Delta f = \frac{\Delta E}{h} = \frac{1.32 \times 10^{-20}\ \text{J}}{6.626 \times 10^{-34}\ \text{J}\cdot\text{s}} = 1.99 \times 10^{13}\ \text{Hz}$$

This is 5.3% of the frequency $f = 3.75 \times 10^{14}$ Hz we found in part (a). Hence these ultrashort pulses do not have a definite frequency; the average frequency of many such pulses will be 3.75×10^{14} Hz, but the frequency of any individual pulse can be anywhere from 5.3% higher to 5.3% lower.

(d) The spatial length Δx of the pulse is the distance that the front of the pulse travels during the time $\Delta t = 4.00 \times 10^{-15}$ s it takes the pulse to emerge from the laser:

$$\Delta x = c\Delta t = (3.00 \times 10^{8}\ \text{m/s})(4.00 \times 10^{-15}\ \text{s})$$
$$= 1.20 \times 10^{-6}\ \text{m}$$
$$\Delta x = \frac{1.20 \times 10^{-6}\ \text{m}}{8.00 \times 10^{-7}\ \text{m/wavelength}} = 1.50\ \text{wavelengths}$$

This justifies the term *ultrashort*. The pulse is less than two wavelengths long!

(e) From Eq. (38.5), the momentum of an average photon in the pulse is

$$p_x = \frac{E}{c} = \frac{2.48 \times 10^{-19}\ \text{J}}{3.00 \times 10^{8}\ \text{m/s}} = 8.28 \times 10^{-28}\ \text{kg}\cdot\text{m/s}$$

The spatial uncertainty is $\Delta x = 1.20 \times 10^{-6}$ m. From Eq. (38.17) minimum momentum uncertainty corresponds to $\Delta x \, \Delta p_x = \hbar/2$, so

$$\Delta p_x = \frac{\hbar}{2\Delta x} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(1.20 \times 10^{-6}\ \text{m})} = 4.40 \times 10^{-29}\ \text{kg}\cdot\text{m/s}$$

This is 5.3% of the average photon momentum p_x. An individual photon within the pulse can have a momentum that is 5.3% greater or less than the average.

(f) To estimate the number of photons in the pulse, we divide the total pulse energy by the average photon energy:

$$\frac{2.00 \times 10^{-6}\ \text{J/pulse}}{2.48 \times 10^{-19}\ \text{J/photon}} = 8.06 \times 10^{12}\ \text{photons/pulse}$$

The energy of an individual photon is uncertain, so this is the *average* number of photons per pulse.

EVALUATE: The percentage uncertainties in energy and momentum are large because this laser pulse is so short. If the pulse were longer, both Δt and Δx would be greater and the corresponding uncertainties in photon energy and photon momentum would be smaller.

Our calculation in part (f) shows an important distinction between photons and other kinds of particles. In principle it is possible to make an exact count of the number of electrons, protons, and neutrons in an object such as this book. If you repeated the count, you would get the same answer as the first time. By contrast, if you counted the number of photons in a laser pulse you would *not* necessarily get the same answer every time! The uncertainty in photon energy means that on each count there could be a different number of photons whose individual energies sum to 2.00×10^{-6} J. That's yet another of the many strange properties of photons.

TEST YOUR UNDERSTANDING OF SECTION 38.4 Through which of the following angles is a photon of wavelength λ most likely to be deflected after passing through a slit of width a? Assume that λ is much less than a. (i) $\theta = \lambda/a$; (ii) $\theta = 3\lambda/2a$; (iii) $\theta = 2\lambda/a$; (iv) $\theta = 3\lambda/a$; (v) not enough information given to decide.

CHAPTER 38 SUMMARY

SOLUTIONS TO ALL EXAMPLES

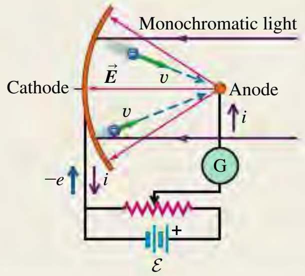
Photons: Electromagnetic radiation behaves as both waves and particles. The energy in an electromagnetic wave is carried in units called photons. The energy E of one photon is proportional to the wave frequency f and inversely proportional to the wavelength λ, and is proportional to a universal quantity h called Planck's constant. The momentum of a photon has magnitude E/c. (See Example 38.1.)

$$E = hf = \frac{hc}{\lambda} \quad (38.2)$$

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda} \quad (38.5)$$

The photoelectric effect: In the photoelectric effect, a surface can eject an electron by absorbing a photon whose energy hf is greater than or equal to the work function ϕ of the material. The stopping potential V_0 is the voltage required to stop a current of ejected electrons from reaching an anode. (See Examples 38.2 and 38.3.)

$$eV_0 = hf - \phi \quad (38.4)$$



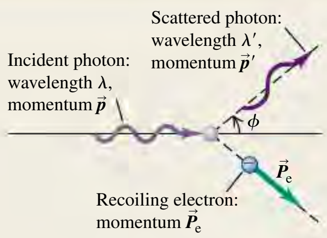
Photon production, photon scattering, and pair production: X rays can be produced when electrons accelerated to high kinetic energy across a potential increase V_{AC} strike a target. The photon model explains why the maximum frequency and minimum wavelength produced are given by Eq. (38.6). (See Example 38.4.) In Compton scattering a photon transfers some of its energy and momentum to an electron with which it collides. For free electrons (mass m), the wavelengths of incident and scattered photons are related to the photon scattering angle ϕ by Eq. (38.7). (See Example 38.5.) In pair production a photon of sufficient energy can disappear and be replaced by an electron–positron pair. In the inverse process, an electron and a positron can annihilate and be replaced by a pair of photons. (See Example 38.6.)

$$eV_{AC} = hf_{max} = \frac{hc}{\lambda_{min}} \quad (38.6)$$

(bremsstrahlung)

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\phi) \quad (38.7)$$

(Compton scattering)



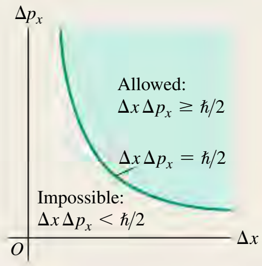
The Heisenberg uncertainty principle: It is impossible to determine both a photon's position and its momentum at the same time to arbitrarily high precision. The precision of such measurements for the x-components is limited by the Heisenberg uncertainty principle, Eq. (38.17); there are corresponding relationships for the y- and z-components. The uncertainty ΔE in the energy of a state that is occupied for a time Δt is given by Eq. (38.24). In these expressions, $\hbar = h/2\pi$. (See Example 38.7.)

$$\Delta x \Delta p_x \geq \hbar/2 \quad (38.17)$$

(Heisenberg uncertainty principle for position and momentum)

$$\Delta t \Delta E \geq \hbar/2 \quad (38.24)$$

(Heisenberg uncertainty principle for energy and time)



BRIDGING PROBLEM COMPTON SCATTERING AND ELECTRON RECOIL

SOLUTION

An incident x-ray photon is scattered from a free electron that is initially at rest. The photon is scattered straight back at an angle of 180° from its initial direction. The wavelength of the scattered photon is 0.0830 nm. (a) What is the wavelength of the incident photon? (b) What are the magnitude of the momentum and the speed of the electron after the collision? (c) What is the kinetic energy of the electron after the collision?

SOLUTION GUIDE

IDENTIFY and SET UP

1. In this problem a photon is scattered by an electron initially at rest. In Section 38.3 you learned how to relate the wavelengths of the incident and scattered photons; in this problem you must also find the momentum, speed, and kinetic energy of the recoiling electron. You can find these because momentum and energy are conserved in the collision. Draw a diagram showing the momentum vectors of the photon and electron before and after the scattering.
2. Which key equation can be used to find the incident photon wavelength? What is the photon scattering angle ϕ in this problem?

EXECUTE

3. Use the equation you selected in step 2 to find the wavelength of the incident photon.
4. Use momentum conservation and your result from step 3 to find the momentum of the recoiling electron. (*Hint:* All of the momentum vectors are along the same line, but not all point in the same direction. Be careful with signs.)
5. Find the speed of the recoiling electron from your result in step 4. (*Hint:* Assume that the electron is nonrelativistic, so you can use the relationship between momentum and speed from Chapter 8. This is acceptable if the speed of the electron is less than about $0.1c$. Is it?)
6. Use your result from step 4 or step 5 to find the electron kinetic energy.

EVALUATE

7. You can check your answer in step 6 by finding the difference between the energies of the incident and scattered photons. Is your result consistent with conservation of energy?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q38.1 In what ways do photons resemble other particles such as electrons? In what ways do they differ? Do photons have mass? Do they have electric charge? Can they be accelerated? What mechanical properties do they have?

Q38.2 There is a certain probability that a single electron may simultaneously absorb *two* identical photons from a high-intensity laser. How would such an occurrence affect the threshold frequency and the equations of Section 38.1? Explain.

Q38.3 According to the photon model, light carries its energy in packets called quanta or photons. Why then don't we see a series of flashes when we look at things?

Q38.4 Would you expect effects due to the photon nature of light to be generally more important at the low-frequency end of the electromagnetic spectrum (radio waves) or at the high-frequency end (x rays and gamma rays)? Why?

Q38.5 During the photoelectric effect, light knocks electrons out of metals. So why don't the metals in your home lose their electrons when you turn on the lights?

Q38.6 Most black-and-white photographic film (with the exception of some special-purpose films) is less sensitive to red light than blue light and has almost no sensitivity to infrared. How can these properties be understood on the basis of photons?

Q38.7 Human skin is relatively insensitive to visible light, but ultraviolet radiation can cause severe burns. Does this have anything to do with photon energies? Explain.

Q38.8 Explain why Fig. 38.4 shows that most photoelectrons have kinetic energies less than $hf - \phi$, and also explain how these smaller kinetic energies occur.

Q38.9 In a photoelectric-effect experiment, the photocurrent i for large positive values of V_{AC} has the same value no matter what the light frequency f (provided that f is higher than the threshold frequency f_0). Explain why.

Q38.10 In an experiment involving the photoelectric effect, if the intensity of the incident light (having frequency higher than the threshold frequency) is reduced by a factor of 10 without changing anything else, which (if any) of the following statements about this process will be true? (a) The number of photoelectrons will most likely be reduced by a factor of 10. (b) The maximum kinetic energy of the ejected photoelectrons will most likely be reduced by a factor of 10. (c) The maximum speed of the ejected photoelectrons will most likely be reduced by a factor of 10. (d) The maximum speed of the ejected photoelectrons will most likely be reduced by a factor of $\sqrt{10}$. (e) The time for the first photoelectron to be ejected will be increased by a factor of 10.

Q38.11 The materials called *phosphors* that coat the inside of a fluorescent lamp convert ultraviolet radiation (from the mercury-vapor discharge inside the tube) into visible light. Could one also make a phosphor that converts visible light to ultraviolet? Explain.

Q38.12 In a photoelectric-effect experiment, which of the following will increase the maximum kinetic energy of the photoelectrons? (a) Use light of greater intensity; (b) use light of higher frequency; (c) use light of longer wavelength; (d) use a metal surface with a larger work function. In each case justify your answer.

Q38.13 A photon of frequency f undergoes Compton scattering from an electron at rest and scatters through an angle ϕ. The frequency of the scattered photon is f'. How is f' related to f? Does your answer depend on ϕ? Explain.

Q38.14 Can Compton scattering occur with protons as well as electrons? For example, suppose a beam of x rays is directed at a target of liquid hydrogen. (Recall that the nucleus of hydrogen consists of a single proton.) Compared to Compton scattering with electrons, what similarities and differences would you expect? Explain.

Q38.15 Why must engineers and scientists shield against x-ray production in high-voltage equipment?

Q38.16 In attempting to reconcile the wave and particle models of light, some people have suggested that the photon rides up and down on the crests and troughs of the electromagnetic wave. What things are *wrong* with this description?

Q38.17 Some lasers emit light in pulses that are only 10^{-12} s in duration. The length of such a pulse is $(3 \times 10^8 \text{ m/s})(10^{-12} \text{ s}) = 3 \times 10^{-4} \text{ m} = 0.3 \text{ mm}$. Can pulsed laser light be as monochromatic as light from a laser that emits a steady, continuous beam? Explain.

EXERCISES

Section 38.1 Light Absorbed as Photons: The Photoelectric Effect

38.1 • A photon of green light has a wavelength of 520 nm. Find the photon's frequency, magnitude of momentum, and energy. Express the energy in both joules and electron volts.

38.2 • BIO **Response of the Eye.** The human eye is most sensitive to green light of wavelength 505 nm. Experiments have found that when people are kept in a dark room until their eyes adapt to the darkness, a *single* photon of green light will trigger receptor cells in the rods of the retina. (a) What is the frequency of this photon? (b) How much energy (in joules and electron volts) does it deliver to the receptor cells? (c) To appreciate what a small amount of energy this is, calculate how fast a typical bacterium of mass 9.5×10^{-12} g would move if it had that much energy.

38.3 • A 75-W light source consumes 75 W of electrical power. Assume all this energy goes into emitted light of wavelength 600 nm. (a) Calculate the frequency of the emitted light. (b) How many photons per second does the source emit? (c) Are the answers to parts (a) and (b) the same? Is the frequency of the light the same thing as the number of photons emitted per second? Explain.

38.4 • BIO A laser used to weld detached retinas emits light with a wavelength of 652 nm in pulses that are 20.0 ms in duration. The average power during each pulse is 0.600 W. (a) How much energy is in each pulse in joules? In electron volts? (b) What is the energy of one photon in joules? In electron volts? (c) How many photons are in each pulse?

38.5 • A photon has momentum of magnitude 8.24×10^{-28} kg · m/s. (a) What is the energy of this photon? Give your answer in joules and in electron volts. (b) What is the wavelength of this photon? In what region of the electromagnetic spectrum does it lie?

38.6 • The photoelectric threshold wavelength of a tungsten surface is 272 nm. Calculate the maximum kinetic energy of the electrons ejected from this tungsten surface by ultraviolet radiation of frequency 1.45×10^{15} Hz. Express the answer in electron volts.

38.7 •• A clean nickel surface is exposed to light of wavelength 235 nm. What is the maximum speed of the photoelectrons emitted from this surface? Use Table 38.1.

38.8 •• What would the minimum work function for a metal have to be for visible light (380–750 nm) to eject photoelectrons?

38.9 •• When ultraviolet light with a wavelength of 400.0 nm falls on a certain metal surface, the maximum kinetic energy of the emitted photoelectrons is measured to be 1.10 eV. What is the maximum kinetic energy of the photoelectrons when light of wavelength 300.0 nm falls on the same surface?

38.10 •• The photoelectric work function of potassium is 2.3 eV. If light that has a wavelength of 190 nm falls on potassium, find (a) the stopping potential in volts; (b) the kinetic energy, in electron volts, of the most energetic electrons ejected; (c) the speed of these electrons.

38.11 • When ultraviolet light with a wavelength of 254 nm falls on a clean copper surface, the stopping potential necessary to stop emission of photoelectrons is 0.181 V. (a) What is the photoelectric threshold wavelength for this copper surface? (b) What is the work function for this surface, and how does your calculated value compare with that given in Table 38.1?

Section 38.2 Light Emitted as Photons: X-Ray Production

38.12 • The cathode-ray tubes that generated the picture in early color televisions were sources of x rays. If the acceleration voltage in a television tube is 15.0 kV, what are the shortest-wavelength x rays produced by the television?

38.13 • Protons are accelerated from rest by a potential difference of 4.00 kV and strike a metal target. If a proton produces one photon on impact, what is the minimum wavelength of the resulting x rays? How does your answer compare to the minimum wavelength if 4.00-keV electrons are used instead? Why do x-ray tubes use electrons rather than protons to produce x rays?

38.14 •• (a) What is the minimum potential difference between the filament and the target of an x-ray tube if the tube is to produce x rays with a wavelength of 0.150 nm? (b) What is the shortest wavelength produced in an x-ray tube operated at 30.0 kV?

Section 38.3 Light Scattered as Photons: Compton Scattering and Pair Production

38.15 • An x ray with a wavelength of 0.100 nm collides with an electron that is initially at rest. The x ray's final wavelength is 0.110 nm. What is the final kinetic energy of the electron?

38.16 • X rays are produced in a tube operating at 24.0 kV. After emerging from the tube, x rays with the minimum wavelength produced strike a target and undergo Compton scattering through an angle of 45.0°. (a) What is the original x-ray wavelength? (b) What is the wavelength of the scattered x rays? (c) What is the energy of the scattered x rays (in electron volts)?

38.17 •• X rays with initial wavelength 0.0665 nm undergo Compton scattering. What is the longest wavelength found in the scattered x rays? At which scattering angle is this wavelength observed?

38.18 •• A photon with wavelength $\lambda = 0.1385$ nm scatters from an electron that is initially at rest. What must be the angle between the direction of propagation of the incident and scattered photons if the speed of the electron immediately after the collision is 8.90×10^6 m/s?

38.19 •• If a photon of wavelength 0.04250 nm strikes a free electron and is scattered at an angle of 35.0° from its original direction, find (a) the change in the wavelength of this photon; (b) the wavelength of the scattered light; (c) the change in energy of the photon (is it a loss or a gain?); (d) the energy gained by the electron.

38.20 •• A photon scatters in the backward direction ($\phi = 180°$) from a free proton that is initially at rest. What must the wavelength of the incident photon be if it is to undergo a 10.0% change in wavelength as a result of the scattering?

38.21 •• X rays with an initial wavelength of 0.900×10^{-10} m undergo Compton scattering. For what scattering angle is the wavelength of the scattered x rays greater by 1.0% than that of the incident x rays?

38.22 • An electron and a positron are moving toward each other and each has speed $0.500c$ in the lab frame. (a) What is the kinetic energy of each particle? (b) The e^+ and e^- meet head-on and annihilate. What is the energy of each photon that is produced? (c) What is the wavelength of each photon? How does the wavelength compare to the photon wavelength when the initial kinetic energy of the e^+ and e^- is negligibly small (see Example 38.6)?

Section 38.4 Wave-Particle Duality, Probability, and Uncertainty

38.23 • An ultrashort pulse has a duration of 9.00 fs and produces light at a wavelength of 556 nm. What are the momentum and momentum uncertainty of a single photon in the pulse?

38.24 • A horizontal beam of laser light of wavelength 585 nm passes through a narrow slit that has width 0.0620 mm. The intensity of the light is measured on a vertical screen that is 2.00 m from the slit. (a) What is the minimum uncertainty in the vertical component of the momentum of each photon in the beam after the photon has passed through the slit? (b) Use the result of part (a) to estimate the width of the central diffraction maximum that is observed on the screen.

38.25 • A laser produces light of wavelength 625 nm in an ultrashort pulse. What is the minimum duration of the pulse if the minimum uncertainty in the energy of the photons is 1.0%?

PROBLEMS

38.26 •• (a) If the average frequency emitted by a 120-W light bulb is 5.00×10^{14} Hz and 10.0% of the input power is emitted as visible light, approximately how many visible-light photons are emitted per second? (b) At what distance would this correspond to 1.00×10^{11} visible-light photons per cm^2 per second if the light is emitted uniformly in all directions?

38.27 •• CP BIO **Removing Vascular Lesions.** A pulsed dye laser emits light of wavelength 585 nm in 450-μs pulses. Because this wavelength is strongly absorbed by the hemoglobin in the blood, the method is especially effective for removing various types of blemishes due to blood, such as port-wine–colored birthmarks. To get a reasonable estimate of the power required for such laser surgery, we can model the blood as having the same specific heat and heat of vaporization as water ($4190\ \text{J/kg}\cdot\text{K}, 2.256 \times 10^6\ \text{J/kg}$). Suppose that each pulse must remove 2.0 μg of blood by evaporating it, starting at 33°C. (a) How much energy must each pulse deliver to the blemish? (b) What must be the power output of this laser? (c) How many photons does each pulse deliver to the blemish?

38.28 • A 2.50-W beam of light of wavelength 124 nm falls on a metal surface. You observe that the maximum kinetic energy of the ejected electrons is 4.16 eV. Assume that each photon in the beam ejects a photoelectron. (a) What is the work function (in electron volts) of this metal? (b) How many photoelectrons are ejected each second from this metal? (c) If the power of the light beam, but not its wavelength, were reduced by half, what would be the answer to part (b)? (d) If the wavelength of the beam, but not its power, were reduced by half, what would be the answer to part (b)?

38.29 •• An incident x-ray photon of wavelength 0.0900 nm is scattered in the backward direction from a free electron that is initially at rest. (a) What is the magnitude of the momentum of the scattered photon? (b) What is the kinetic energy of the electron after the photon is scattered?

38.30 •• CP A photon with wavelength $\lambda = 0.0980$ nm is incident on an electron that is initially at rest. If the photon scatters in the backward direction, what is the magnitude of the linear momentum of the electron just after the collision with the photon?

38.31 •• CP A photon with wavelength $\lambda = 0.1050$ nm is incident on an electron that is initially at rest. If the photon scatters at an angle of 60.0° from its original direction, what are the magnitude and direction of the linear momentum of the electron just after it collides with the photon?

38.32 •• CP A photon of wavelength 4.50 pm scatters from a free electron that is initially at rest. (a) For $\phi = 90.0°$, what is the kinetic energy of the electron immediately after the collision with the photon? What is the ratio of this kinetic energy to the rest energy of the electron? (b) What is the speed of the electron immediately after the collision? (c) What is the magnitude of the momentum of the electron immediately after the collision? What is the ratio of this momentum value to the nonrelativistic expression mv?

38.33 •• Nuclear fusion reactions at the center of the sun produce gamma-ray photons with energies of about 1 MeV (10^6 eV). By contrast, what we see emanating from the sun's surface are visible-light photons with wavelengths of about 500 nm. A simple model that explains this difference in wavelength is that a photon undergoes Compton scattering many times—in fact, about 10^{26} times, as suggested by models of the solar interior—as it travels from the center of the sun to its surface. (a) Estimate the increase in wavelength of a photon in an average Compton-scattering event. (b) Find the angle in degrees through which the photon is scattered in the scattering event described in part (a). (*Hint:* A useful approximation is $\cos\phi \approx 1 - \phi^2/2$, which is valid for $\phi \ll 1$. Note that ϕ is in radians in this expression.) (c) It is estimated that a photon takes about 10^6 years to travel from the core to the surface of the sun. Find the average distance that light can travel within the interior of the sun without being scattered. (This distance is roughly equivalent to how far you could see if you were inside the sun and could survive the extreme temperatures there. As your answer shows, the interior of the sun is *very* opaque.)

38.34 •• **CP** An x-ray tube is operating at voltage V and current I. (a) If only a fraction p of the electric power supplied is converted into x rays, at what rate is energy being delivered to the target? (b) If the target has mass m and specific heat c (in J/kg · K), at what average rate would its temperature rise if there were no thermal losses? (c) Evaluate your results from parts (a) and (b) for an x-ray tube operating at 18.0 kV and 60.0 mA that converts 1.0% of the electric power into x rays. Assume that the 0.250-kg target is made of lead ($c = 130$ J/kg · K). (d) What must the physical properties of a practical target material be? What would be some suitable target elements?

38.35 •• A photon with wavelength 0.1100 nm collides with a free electron that is initially at rest. After the collision the wavelength is 0.1132 nm. (a) What is the kinetic energy of the electron after the collision? What is its speed? (b) If the electron is suddenly stopped (for example, in a solid target), all of its kinetic energy is used to create a photon. What is the wavelength of this photon?

38.36 •• An x-ray photon is scattered from a free electron (mass m) at rest. The wavelength of the scattered photon is λ', and the final speed of the struck electron is v. (a) What was the initial wavelength λ of the photon? Express your answer in terms of λ', v, and m. (*Hint:* Use the relativistic expression for the electron kinetic energy.) (b) Through what angle ϕ is the photon scattered? Express your answer in terms of λ, λ', and m. (c) Evaluate your results in parts (a) and (b) for a wavelength of 5.10×10^{-3} nm for the scattered photon and a final electron speed of 1.80×10^8 m/s. Give ϕ in degrees.

38.37 •• **DATA** In developing night-vision equipment, you need to measure the work function for a metal surface, so you perform a photoelectric-effect experiment. You measure the stopping potential V_0 as a function of the wavelength λ of the light that is incident on the surface. You get the results in the table.

λ (nm)	100	120	140	160	180	200
V_0 (V)	7.53	5.59	3.98	2.92	2.06	1.43

In your analysis, you use $c = 2.998 \times 10^8$ m/s and $e = 1.602 \times 10^{-19}$ C, which are values obtained in other experiments. (a) Select a way to plot your results so that the data points fall close to a straight line. Using that plot, find the slope and y-intercept of the best-fit straight line to the data. (b) Use the results of part (a) to calculate Planck's constant h (as a test of your data) and the work function (in eV) of the surface. (c) What is the longest wavelength of light that will produce photoelectrons from this surface? (d) What wavelength of light is required to produce photoelectrons with kinetic energy 10.0 eV?

38.38 •• **DATA** While analyzing smoke detector designs that rely on the photoelectric effect, you are evaluating surfaces made from each of the materials listed in Table 38.1. One particular application uses ultraviolet light with wavelength 270 nm. (a) For which of the materials in Table 38.1 will this light produce photoelectrons? (b) Which material will result in photoelectrons of the greatest kinetic energy? What will be the maximum speed of the photoelectrons produced as they leave this material's surface? (c) What is the longest wavelength that will produce photoelectrons from a gold surface, if the surface has a work function equal to the value given for gold in Table 38.1? (d) For the wavelength calculated in part (c), what will be the maximum kinetic energy of the photoelectrons produced from a sodium surface that has a work function equal to the value given in Table 38.1 for sodium?

38.39 •• **DATA** To test the photon concept, you perform a Compton-scattering experiment in a research lab. Using photons of very short wavelength, you measure the wavelength λ' of scattered photons as a function of the scattering angle ϕ, the angle between the direction of a scattered photon and the incident photon. You obtain these results.

ϕ (deg)	30.6	58.7	90.2	119.2	151.3
λ' (pm)	5.52	6.40	7.60	8.84	9.69

Your analysis assumes that the target is a free electron at rest. (a) Graph your data as λ' versus $1 - \cos\phi$. What are the slope and y-intercept of the best-fit straight line to your data? (b) The Compton wavelength λ_C is defined as $\lambda_C = h/mc$, where m is the mass of an electron. Use the results of part (a) to calculate λ_C. (c) Use the results of part (a) to calculate the wavelength λ of the incident light.

CHALLENGE PROBLEM

38.40 ••• Consider Compton scattering of a photon by a *moving* electron. Before the collision the photon has wavelength λ and is moving in the $+x$-direction, and the electron is moving in the $-x$-direction with total energy E (including its rest energy mc^2). The photon and electron collide head-on. After the collision, both are moving in the $-x$-direction (that is, the photon has been scattered by 180°). (a) Derive an expression for the wavelength λ' of the scattered photon. Show that if $E \gg mc^2$, where m is the rest mass of the electron, your result reduces to

$$\lambda' = \frac{hc}{E}\left(1 + \frac{m^2c^4\lambda}{4hcE}\right)$$

(b) A beam of infrared radiation from a CO_2 laser ($\lambda = 10.6\ \mu$m) collides head-on with a beam of electrons, each of total energy $E = 10.0$ GeV (1 GeV $= 10^9$ eV). Calculate the wavelength λ' of the scattered photons, assuming a 180° scattering angle. (c) What kind of scattered photons are these (infrared, microwave, ultraviolet, etc.)? Can you think of an application of this effect?

PASSAGE PROBLEMS

BIO RADIATION THERAPY FOR TUMORS. Malignant tumors are commonly treated with targeted x-ray radiation therapy. To generate these medical x rays, a linear accelerator directs a high-energy beam of electrons toward a metal target—typically tungsten. As they near the tungsten nuclei, the electrons are deflected and accelerated, emitting high-energy photons via bremsstrahlung. The resulting x rays are collimated into a beam that is directed at the tumor. The photons can deposit energy in the tumor through Compton and photoelectric interactions. A typical tumor has 10^8 cells/cm^3, and in a full treatment, 4-MeV photons may produce a dose of 70 Gy in 35 fractional exposures on different days. The *gray* (Gy) is a measure of the absorbed energy dose of radiation per unit mass of tissue: 1 Gy = 1 J/kg.

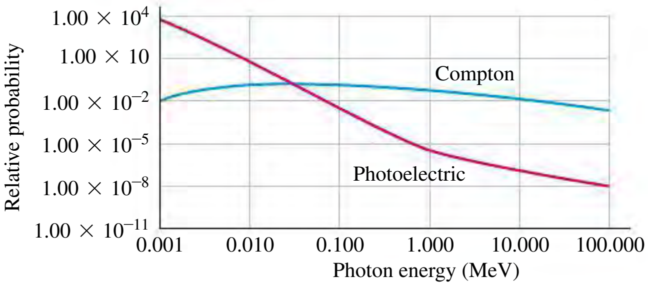
38.41 How much energy is imparted to one cell during one day's treatment? Assume that the specific gravity of the tumor is 1 and that 1 J = 6×10^{18} eV. (a) 120 keV; (b) 12 MeV; (c) 120 MeV; (d) 120×10^3 MeV.

38.42 While interacting with molecules (mainly water) in the tumor tissue, each Compton electron or photoelectron causes a series of ionizations, each of which takes about 40 eV. Estimate the maximum number of ionizations that one photon generated by this linear accelerator can produce in tissue. (a) 100; (b) 1000; (c) 10^4; (d) 10^5.

38.43 The high-energy photons can undergo Compton scattering off electrons in the tumor. The energy imparted by a photon is a maximum when the photon scatters straight back from the electron. In this process, what is the maximum energy that a photon with the energy described in the passage can give to an electron? (a) 3.8 MeV; (b) 2.0 MeV; (c) 0.40 MeV; (d) 0.23 MeV.

38.44 The probability of a photon interacting with tissue via the photoelectric effect or the Compton effect depends on the photon energy. Use **Fig. P38.44** to determine the best description of how the photons from the linear accelerator described in the passage interact with a tumor. (a) Via the Compton effect only; (b) mostly via the photoelectric effect until they have lost most of their energy, and then mostly via the Compton effect; (c) mostly via the Compton effect until they have lost most of their energy, and then mostly via the photoelectric effect; (d) via the Compton effect and the photoelectric effect equally.

Figure **P38.44**



38.45 Higher-energy photons might be desirable for the treatment of certain tumors. Which of these actions would generate higher-energy photons in this linear accelerator? (a) Increasing the number of electrons that hit the tungsten target; (b) accelerating the electrons through a higher potential difference; (c) both (a) and (b); (d) none of these.

Answers

Chapter Opening Question ?

(i) The energy of a photon E is inversely proportional to its wavelength λ: The shorter the wavelength, the more energetic is the photon. Since visible light has shorter wavelengths than infrared light, the headlamp emits photons of greater energy. However, the light from the infrared laser is far more *intense* (delivers much more energy per second per unit area to the patient's skin) because it emits many more photons per second than does the headlamp and concentrates them onto a very small spot.

Test Your Understanding Questions

38.1 (iv) From Eq. (38.2), a photon of energy $E = 1.14$ eV has wavelength

$$\begin{aligned}\lambda &= hc/E \\ &= (4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(3.00 \times 10^{8}\ \text{m/s})/(1.14\ \text{eV}) \\ &= 1.09 \times 10^{-6}\ \text{m} = 1090\ \text{nm}\end{aligned}$$

This is in the infrared part of the spectrum. Since wavelength is inversely proportional to photon energy, the *minimum* photon energy of 1.14 eV corresponds to the *maximum* wavelength that causes photoconductivity in silicon. Thus the wavelength must be 1090 nm or less.

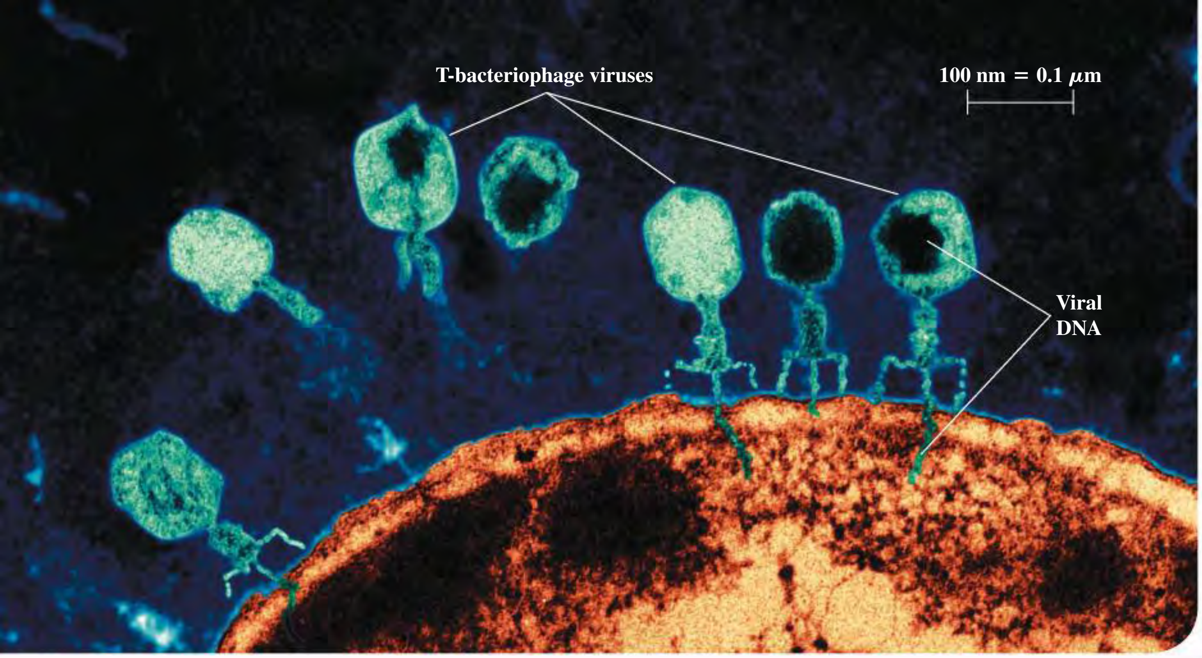
38.2 (ii) Equation (38.6) shows that the minimum wavelength of x rays produced by bremsstrahlung depends on the potential difference V_{AC} but does *not* depend on the rate at which electrons strike the anode. Increasing the number of electrons per second will only cause an increase in the number of x-ray photons emitted per second (that is, the x-ray intensity I).

38.3 yes, no Equation (38.7) shows that the wavelength shift $\Delta\lambda = \lambda' - \lambda$ depends only on the photon scattering angle ϕ, not on the wavelength of the incident photon. So a visible-light photon scattered through an angle ϕ undergoes the same wavelength shift as an x-ray photon. Equation (38.7) also shows that this shift is of the order of $h/mc = 2.426 \times 10^{-12}$ m $= 0.002426$ nm. This is a few percent of the wavelength of x rays (see Example 38.5), so the effect is noticeable in x-ray scattering. However, h/mc is a tiny fraction of the wavelength of visible light (between 380 and 750 nm). The human eye cannot distinguish such minuscule differences in wavelength (that is, differences in color).

38.4 (ii) There is *zero* probability that a photon will be deflected by one of the angles where the diffraction pattern has zero intensity. These angles are given by $a \sin\theta = m\lambda$ with $m = \pm 1, \pm 2, \pm 3, \ldots$. Since λ is much less than a, we can write these angles as $\theta = m\lambda/a = \pm\lambda/a, \pm 2\lambda/a, \pm 3\lambda/a, \ldots$. These values include answers (i), (iii), and (iv), so it is impossible for a photon to be deflected through any of these angles. The intensity is not zero at $\theta = 3\lambda/2a$ (located between two zeros in the diffraction pattern), so there is some probability that a photon will be deflected through this angle.

Bridging Problem

(a) 0.0781 nm **(b)** 1.65×10^{-23} kg·m/s, 1.81×10^{7} m/s
(c) 1.49×10^{-16} J



? Viruses (shown in blue) have landed on an *E. coli* bacterium and injected their DNA, converting the bacterium into a virus factory. This false-color image was made by using a beam of electrons rather than a light beam. Electrons are used for imaging such fine details because, compared to visible-light photons, (i) electrons can have much shorter wavelengths; (ii) electrons can have much longer wavelengths; (iii) electrons can have much less momentum; (iv) electrons have more total energy for the same momentum; (v) more than one of these.

39 PARTICLES BEHAVING AS WAVES

LEARNING GOALS

Looking forward at ...

39.1 De Broglie's proposal that electrons and other particles can behave like waves, and the experimental evidence for de Broglie's ideas.

39.2 How physicists discovered the atomic nucleus.

39.3 How Bohr's model of electron orbits explained the spectra of hydrogen and hydrogenlike atoms.

39.4 How a laser operates.

39.5 How the idea of energy levels, coupled with the photon model of light, explains the spectrum of light emitted by a hot, opaque object.

39.6 What the uncertainty principle tells us about the nature of the atom.

Looking back at ...

13.4 Satellites.

17.7 Stefan–Boltzmann law.

18.4, 18.5 Equipartition principle; Maxwell–Boltzmann distribution function.

32.1, 32.5 Radiation from an accelerating charge; electromagnetic standing waves.

36.5–36.7 Light diffraction, x-ray diffraction, resolution.

38.1, 38.4 Photoelectric effect, photons, interference.

In Chapter 38 we discovered one aspect of nature's wave–particle duality: Light and other electromagnetic radiation act sometimes like waves and sometimes like particles. Interference and diffraction demonstrate wave behavior, while emission and absorption of photons demonstrate particle behavior.

If light waves can behave like particles, can the particles of matter behave like waves? The answer is a resounding yes. Electrons can interfere and diffract just like other kinds of waves. The wave nature of electrons is not merely a laboratory curiosity: It is the fundamental reason why atoms, which according to classical physics should be unstable, are able to exist. In this chapter the wave nature of matter will help us understand the structure of atoms, the operating principles of a laser, and the curious properties of the light emitted by a heated, glowing object. Without the wave picture of matter, there would be no way to explain these phenomena.

In Chapter 40 we'll introduce an even more complete wave picture of matter called *quantum mechanics*. Through the remainder of this book we'll use the ideas of quantum mechanics to understand the nature of molecules, solids, atomic nuclei, and the fundamental particles that are the building blocks of our universe.

39.1 ELECTRON WAVES

In 1924 a French physicist, Louis de Broglie (pronounced "de broy"; **Fig. 39.1**, next page), made a remarkable proposal about the nature of matter. His reasoning, freely paraphrased, went like this: Nature loves symmetry. Light is dualistic in nature, behaving in some situations like waves and in others like particles. If nature is symmetric, this duality should also hold for matter. Electrons, which we usually think of as *particles,* may in some situations behave like *waves*.

If a particle acts like a wave, it should have a wavelength and a frequency. De Broglie postulated that a free particle with rest mass m, moving with nonrelativistic speed v, should have a wavelength λ related to its momentum

39.1 Louis-Victor de Broglie, the seventh Duke de Broglie (1892–1987), broke with family tradition by choosing a career in physics rather than as a diplomat. His revolutionary proposal that particles have wave characteristics—for which de Broglie won the 1929 Nobel Prize in physics—was published in his doctoral thesis.

$p = mv$ in exactly the same way as for a photon, as expressed by Eq. (38.5) from Section 38.1: $\lambda = h/p$. The **de Broglie wavelength** of a particle is then

$$\underbrace{\lambda}_{\text{De Broglie wavelength of a particle}} = \frac{\overbrace{h}^{\text{Planck's constant}}}{\underbrace{p}_{\text{Particle's momentum}}} = \frac{h}{\underbrace{m}_{\text{Particle's mass}}\underbrace{v}_{\text{Particle's speed}}} \qquad (39.1)$$

If the particle's speed is an appreciable fraction of the speed of light c, we replace mv in Eq. (39.1) with $\gamma mv = mv/\sqrt{1 - v^2/c^2}$ [Eq. (37.27) from Section 37.7]. The frequency f, according to de Broglie, is also related to the particle's energy E in exactly the same way as for a photon:

$$\underbrace{E}_{\text{Energy of a particle}} = \overbrace{h}^{\text{Planck's constant}}\underbrace{f}_{\text{Frequency}} \qquad (39.2)$$

CAUTION **Not all photon equations apply to particles with mass** Be careful when applying $E = hf$ to particles with nonzero rest mass, such as electrons. Unlike a photon, they do *not* travel at speed c, so the equations $f = c/\lambda$ and $E = pc$ do *not* apply to them!

Observing the Wave Nature of Electrons

De Broglie's proposal was a bold one, made at a time when there was no direct experimental evidence that particles have wave characteristics. But within a few years, his ideas were resoundingly verified by a diffraction experiment with electrons. This experiment was analogous to those we described in Section 36.6, in which atoms in a crystal act as a three-dimensional diffraction grating for x rays. An x-ray beam is strongly reflected when it strikes a crystal at an angle that gives constructive interference among the waves scattered from the various atoms in the crystal. These interference effects demonstrate the *wave* nature of x rays.

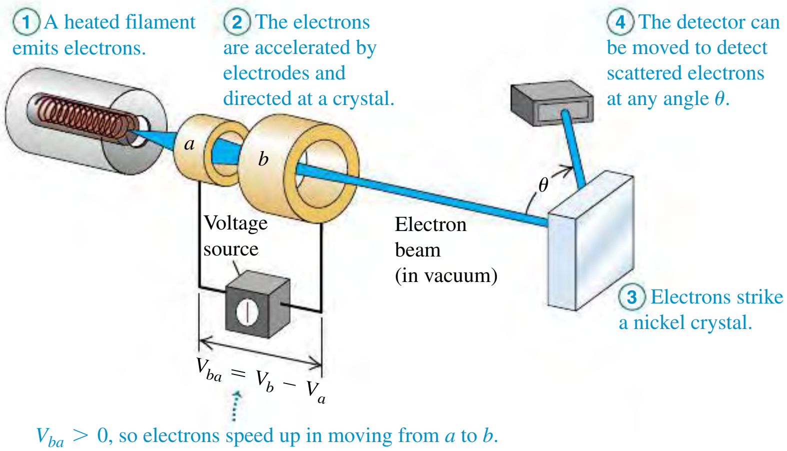
In 1927 the American physicists Clinton Davisson and Lester Germer, working at the Bell Telephone Laboratories, were studying the surface of a piece of nickel by directing a beam of *electrons* at the surface and observing how many electrons bounced off at various angles. **Figure 39.2** shows an experimental setup like theirs. Like many ordinary metals, the sample was *polycrystalline:* It consisted of many randomly oriented microscopic crystals bonded together. As a result, the electron beam reflected diffusely, like light bouncing off a rough surface (see Fig. 33.6b), with a smooth distribution of intensity as a function of the angle θ.

MP

PhET: Davisson-Germer: Electron Diffraction

During the experiment an accident occurred that permitted air to enter the vacuum chamber, and an oxide film formed on the metal surface. To remove this film, Davisson and Germer baked the sample in a high-temperature oven. Unknown to them, this had the effect of creating large regions within the nickel with crystal

39.2 An apparatus similar to that used by Davisson and Germer to discover electron diffraction.



planes that were continuous over the width of the electron beam. From the perspective of the electrons, the sample looked like a *single* crystal of nickel.

When the observations were repeated with this sample, the results were quite different. Now strong maxima in the intensity of the reflected electron beam occurred at specific angles (**Fig. 39.3a**), in contrast to the smooth variation of intensity with angle that Davisson and Germer had observed before the accident. The angular positions of the maxima depended on the accelerating voltage V_{ba} used to produce the electron beam. Davisson and Germer were familiar with de Broglie's hypothesis, and they noticed the similarity of this behavior to x-ray diffraction. This was not the effect they had been looking for, but they immediately recognized that the electron beam was being *diffracted*. They had discovered a very direct experimental confirmation of the wave hypothesis.

Davisson and Germer could determine the speeds of the electrons from the accelerating voltage, so they could compute the de Broglie wavelength from Eq. (39.1). If an electron is accelerated from rest at point a to point b through a potential increase $V_{ba} = V_b - V_a$ as shown in Fig. 39.2, the work done on the electron eV_{ba} equals its kinetic energy K. Using $K = \left(\frac{1}{2}\right)mv^2 = p^2/2m$ for a nonrelativistic particle, we have

$$eV_{ba} = \frac{p^2}{2m} \qquad p = \sqrt{2meV_{ba}}$$

We substitute this into Eq. (39.1) for the de Broglie wavelength of the electron:

De Broglie wavelength of an electron

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2meV_{ba}}} \tag{39.3}$$

Planck's constant; Electron momentum; Accelerating voltage; Magnitude of electron charge; Electron mass

The greater the accelerating voltage V_{ba}, the shorter the wavelength of the electron.

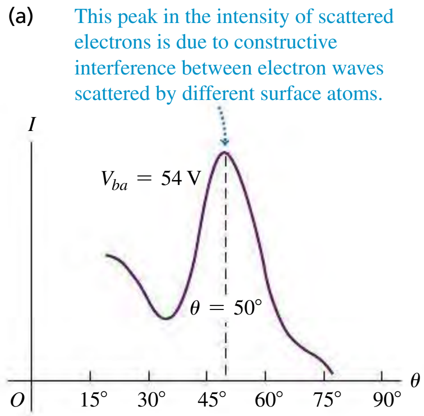
To predict the angles at which strong reflection occurs, note that the electrons were scattered primarily by the planes of atoms near the surface of the crystal. Atoms in a surface plane are arranged in rows, with a distance d that can be measured by x-ray diffraction techniques. These rows act like a reflecting diffraction grating; the angles at which strong reflection occurs are the same as for a grating with center-to-center distance d between its slits (Fig. 39.3b). From Eq. (36.13) the angles of maximum reflection are given by

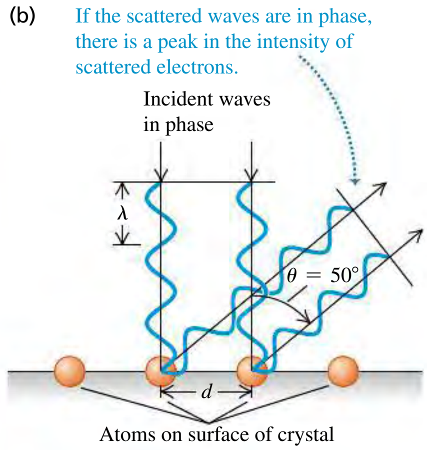
$$d\sin\theta = m\lambda \qquad (m = 1, 2, 3, \ldots) \tag{39.4}$$

where θ is the angle shown in Fig. 39.2. (Note that the geometry in Fig. 39.3b is different from that for Fig. 36.22, so Eq. (39.4) is different from Eq. (36.16).) Davisson and Germer found that the angles predicted by this equation, with the de Broglie wavelength given by Eq. (39.3), agreed with the observed values (Fig. 39.3a). Thus the accidental discovery of **electron diffraction** was the first direct evidence confirming de Broglie's hypothesis.

In 1928, just a year after the Davisson–Germer discovery, the English physicist G. P. Thomson carried out electron-diffraction experiments using a thin, polycrystalline, metallic foil as a target. Debye and Sherrer had used a similar technique several years earlier to study x-ray diffraction from polycrystalline specimens. In these experiments the beam passes *through* the target rather than being reflected from it. Because of the random orientations of the individual microscopic crystals in the foil, the diffraction pattern consists of intensity maxima forming rings around the direction of the incident beam. Thomson's results again confirmed the de Broglie relationship. **Figure 39.4** shows both x-ray and electron diffraction patterns for a polycrystalline aluminum foil. (G. P. Thomson was the son of J. J. Thomson, who 31 years earlier discovered the electron. Davisson and the younger Thomson shared the 1937 Nobel Prize in physics for their discoveries.)

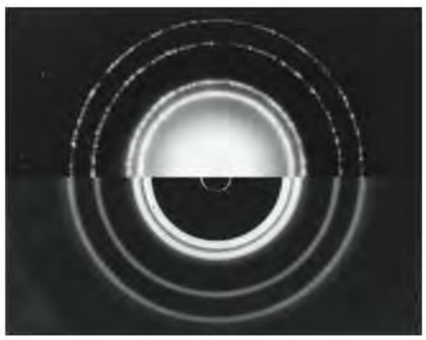
39.3 (a) Intensity of the scattered electron beam in Fig. 39.2 as a function of the scattering angle θ. (b) Electron waves scattered from two adjacent atoms interfere constructively when $d\sin\theta = m\lambda$. In the case shown here, $\theta = 50°$ and $m = 1$.





39.4 X-ray and electron diffraction. The upper half of the photo shows the diffraction pattern for 71-pm x rays passing through aluminum foil. The lower half, with a different scale, shows the diffraction pattern for 600-eV electrons from aluminum. The similarity shows that electrons undergo the same kind of diffraction as x rays.

Top: x-ray diffraction



Bottom: electron diffraction

Additional diffraction experiments were soon carried out in many laboratories using not only electrons but also various ions and low-energy neutrons. All of these are in agreement with de Broglie's bold predictions. Thus the wave nature of particles, so strange in 1924, became firmly established in the years that followed.

PROBLEM-SOLVING STRATEGY 39.1 WAVELIKE PROPERTIES OF PARTICLES

IDENTIFY *the relevant concepts:* Particles have wavelike properties. A particle's (de Broglie) wavelength is inversely proportional to its momentum, and its frequency is proportional to its energy.

SET UP *the problem:* Identify the target variables and decide which equations you will use to calculate them.

EXECUTE *the solution* as follows:

1. Use Eq. (39.1) to relate a particle's momentum p to its wavelength λ; use Eq. (39.2) to relate its energy E to its frequency f.
2. Nonrelativistic kinetic energy may be expressed as either $K = \frac{1}{2}mv^2$ or (because $p = mv$) $K = p^2/2m$. The latter form is useful in calculations involving the de Broglie wavelength.
3. You may express energies in either joules or electron volts, using $h = 6.626 \times 10^{-34}\ \text{J}\cdot\text{s}$ or $h = 4.136 \times 10^{-15}\ \text{eV}\cdot\text{s}$ as appropriate.

EVALUATE *your answer:* To check numerical results, it helps to remember some approximate orders of magnitude. Here's a partial list:

Size of an atom: $10^{-10}\ \text{m} = 0.1\ \text{nm}$

Mass of an atom: $10^{-26}\ \text{kg}$

Mass of an electron: $m = 10^{-30}\ \text{kg}$; $mc^2 = 0.511\ \text{MeV}$

Electron charge magnitude: $10^{-19}\ \text{C}$

kT at room temperature: $\frac{1}{40}\ \text{eV}$

Difference between energy levels of an atom (to be discussed in Section 39.3): 1 to 10 eV

Speed of an electron in the Bohr model of a hydrogen atom (to be discussed in Section 39.3): $10^6\ \text{m/s}$

EXAMPLE 39.1 AN ELECTRON-DIFFRACTION EXPERIMENT

SOLUTION

In an electron-diffraction experiment using an accelerating voltage of 54 V, an intensity maximum occurs for $\theta = 50°$ (see Fig. 39.3a). X-ray diffraction indicates that the atomic spacing in the target is $d = 2.18 \times 10^{-10}\ \text{m} = 0.218\ \text{nm}$. The electrons have negligible kinetic energy before being accelerated. Find the electron wavelength.

SOLUTION

IDENTIFY, SET UP, and EXECUTE: We'll determine λ from both de Broglie's equation, Eq. (39.3), and the diffraction equation, Eq. (39.4). From Eq. (39.3),

$$\lambda = \frac{6.626 \times 10^{-34}\ \text{J}\cdot\text{s}}{\sqrt{2(9.109 \times 10^{-31}\ \text{kg})(1.602 \times 10^{-19}\ \text{C})(54\ \text{V})}}$$
$$= 1.7 \times 10^{-10}\ \text{m} = 0.17\ \text{nm}$$

Alternatively, using Eq. (39.4) and assuming $m = 1$, we get

$$\lambda = d\sin\theta = (2.18 \times 10^{-10}\ \text{m})\sin 50° = 1.7 \times 10^{-10}\ \text{m}$$

EVALUATE: The two numbers agree within the accuracy of the experimental results, which gives us an excellent check on our calculations. Note that this electron wavelength is less than the spacing between the atoms.

EXAMPLE 39.2 ENERGY OF A THERMAL NEUTRON

SOLUTION

Find the speed and kinetic energy of a neutron ($m = 1.675 \times 10^{-27}\ \text{kg}$) with de Broglie wavelength $\lambda = 0.200\ \text{nm}$, a typical interatomic spacing in crystals. Compare this energy with the average translational kinetic energy of an ideal-gas molecule at room temperature ($T = 20°\text{C} = 293\ \text{K}$).

SOLUTION

IDENTIFY and SET UP: This problem uses the relationships between particle speed and wavelength, between particle speed and kinetic energy, and between gas temperature and the average kinetic energy of a gas molecule. We'll find the neutron speed v by using Eq. (39.1) and from that calculate the neutron kinetic energy $K = \frac{1}{2}mv^2$. We'll use Eq. (18.16) to find the average kinetic energy of a gas molecule.

EXECUTE: From Eq. (39.1), the neutron speed is

$$v = \frac{h}{\lambda m} = \frac{6.626 \times 10^{-34}\ \text{J}\cdot\text{s}}{(0.200 \times 10^{-9}\ \text{m})(1.675 \times 10^{-27}\ \text{kg})}$$
$$= 1.98 \times 10^3\ \text{m/s}$$

The neutron kinetic energy is

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}(1.675 \times 10^{-27}\ \text{kg})(1.98 \times 10^3\ \text{m/s})^2$$
$$= 3.28 \times 10^{-21}\ \text{J} = 0.0205\ \text{eV}$$

From Eq. (18.16), the average translational kinetic energy of an ideal-gas molecule at $T = 293$ K is

$$\tfrac{1}{2}m(v^2)_{\text{av}} = \tfrac{3}{2}kT = \tfrac{3}{2}(1.38 \times 10^{-23}\ \text{J/K})(293\ \text{K})$$
$$= 6.07 \times 10^{-21}\ \text{J} = 0.0379\ \text{eV}$$

The two energies are comparable in magnitude, which is why a neutron with kinetic energy in this range is called a *thermal neutron*. Diffraction of thermal neutrons is used to study crystal and molecular structure in the same way as x-ray diffraction. Neutron diffraction has proved to be especially useful in the study of large organic molecules.

EVALUATE: Note that the calculated neutron speed is much less than the speed of light. This justifies our use of the nonrelativistic form of Eq. (39.1).

De Broglie Waves and the Macroscopic World

If the de Broglie picture is correct and matter has wave aspects, you might wonder why we don't see these aspects in everyday life. As an example, we know that waves diffract when sent through a single slit. Yet when we walk through a doorway (a kind of single slit), we don't worry about our body diffracting!

The main reason we don't see these effects on human scales is that Planck's constant h has such a minuscule value. As a result, the de Broglie wavelengths of even the smallest ordinary objects that you can see are extremely small, and the wave effects are unimportant. For instance, what is the wavelength of a falling grain of sand? If the grain's mass is 5×10^{-10} kg and its diameter is $0.07\ \text{mm} = 7 \times 10^{-5}$ m, it will fall in air with a terminal speed of about 0.4 m/s. The magnitude of its momentum is $p = mv = (5 \times 10^{-10}\ \text{kg})(0.4\ \text{m/s}) = 2 \times 10^{-10}\ \text{kg} \cdot \text{m/s}$. The de Broglie wavelength of this falling sand grain is then

$$\lambda = \frac{h}{p} = \frac{6.626 \times 10^{-34}\ \text{J} \cdot \text{s}}{2 \times 10^{-10}\ \text{kg} \cdot \text{m/s}} = 3 \times 10^{-24}\ \text{m}$$

Not only is this wavelength far smaller than the diameter of the sand grain, but it's also far smaller than the size of a typical atom (about 10^{-10} m). A more massive, faster-moving object would have an even larger momentum and an even smaller de Broglie wavelength. The effects of such tiny wavelengths are so small that they are never noticed in daily life.

The Electron Microscope

The **electron microscope** offers an important and interesting example of the interplay of wave and particle properties of electrons. An electron beam can be used to form an image of an object in much the same way as a light beam. A ray of light can be bent by reflection or refraction, and an electron trajectory can be bent by an electric or magnetic field. Rays of light diverging from a point on an object can be brought to convergence by a converging lens or concave mirror, and electrons diverging from a small region can be brought to convergence by electric and/or magnetic fields.

The analogy between light rays and electrons goes deeper. The *ray* model of geometric optics is an approximate representation of the more general *wave* model. Geometric optics (ray optics) is valid whenever interference and diffraction effects can be ignored. Similarly, the model of an electron as a point particle following a line trajectory is an approximate description of the actual behavior of the electron; this model is useful when we can ignore effects associated with the wave nature of electrons.

How is an electron microscope superior to an optical microscope? The *resolution* of an optical microscope is limited by diffraction effects, as we discussed in Section 36.7. Since an optical microscope uses wavelengths around 500 nm, it can't resolve objects smaller than a few hundred nanometers, no matter how carefully its lenses are made. The resolution of an electron microscope is similarly limited by the wavelengths of the electrons, but these wavelengths may be many thousands of times smaller than wavelengths of visible light. As a result, the useful magnification of an electron microscope can be thousands of times greater than that of an optical microscope.

Note that the ability of the electron microscope to form a magnified image *does not* depend on the wave properties of electrons. Within the limitations of the Heisenberg uncertainty principle (which we'll discuss in Section 39.6), we can compute the electron trajectories by treating them as classical charged particles under the action of electric and magnetic forces. Only when we talk about *resolution* do the wave properties become important.

EXAMPLE 39.3 AN ELECTRON MICROSCOPE

SOLUTION

In an electron microscope, the nonrelativistic electron beam is formed by a setup similar to the electron gun used in the Davisson–Germer experiment (see Fig. 39.2). The electrons have negligible kinetic energy before they are accelerated. What accelerating voltage is needed to produce electrons with wavelength 10 pm = 0.010 nm (roughly 50,000 times smaller than typical visible-light wavelengths)?

SOLUTION

IDENTIFY, SET UP, and EXECUTE: We can use the same concepts we used to understand the Davisson–Germer experiment. The accelerating voltage is the quantity V_{ba} in Eq. (39.3). Rewrite this equation to solve for V_{ba}:

$$V_{ba} = \frac{h^2}{2me\lambda^2}$$

$$= \frac{(6.626 \times 10^{-34}\ \mathrm{J\cdot s})^2}{2(9.109 \times 10^{-31}\ \mathrm{kg})(1.602 \times 10^{-19}\ \mathrm{C})(10 \times 10^{-12}\ \mathrm{m})^2}$$

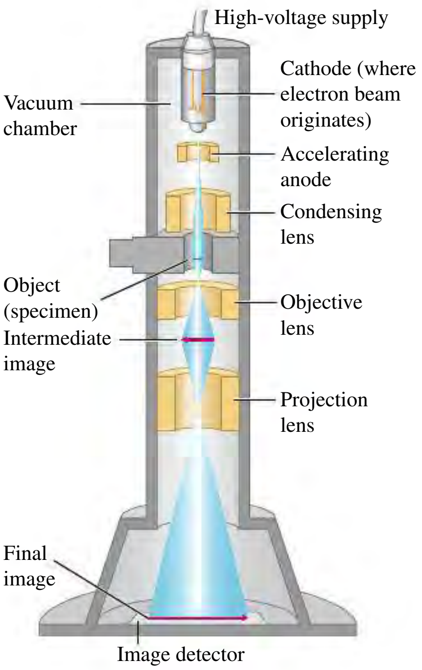
$$= 1.5 \times 10^4\ \mathrm{V} = 15{,}000\ \mathrm{V}$$

EVALUATE: It is easy to attain 15-kV accelerating voltages from 120-V or 240-V line voltage by using a step-up transformer (Section 31.6) and a rectifier (Section 31.1). The accelerated electrons have kinetic energy 15 keV; since the electron rest energy is 0.511 MeV = 511 keV, these electrons are indeed nonrelativistic.

Types of Electron Microscope

Figure 39.5 shows the design of a *transmission electron microscope,* in which electrons actually pass through the specimen being studied. The specimen to be viewed can be no more than 10 to 100 nm thick so the electrons are not slowed appreciably as they pass through. The electrons used in a transmission electron microscope are emitted from a hot cathode and accelerated by a potential difference, typically 40 to 400 kV. They then pass through a condensing "lens" that uses magnetic fields to focus the electrons into a parallel beam before they pass through the specimen. The beam then passes through two more magnetic lenses: an objective lens that forms an intermediate image of the specimen and a projection lens that produces a final real image of the intermediate image. The objective and projection lenses play the roles of the objective and eyepiece lenses, respectively, of a compound optical microscope (see Section 34.8). The final image is projected onto a fluorescent screen for viewing or photographing. The entire apparatus, including the specimen, must be enclosed in a vacuum container; otherwise, electrons would scatter off air molecules and muddle the image. The image that opens this chapter was made with a transmission electron microscope.

39.5 Schematic diagram of a transmission electron microscope (TEM).



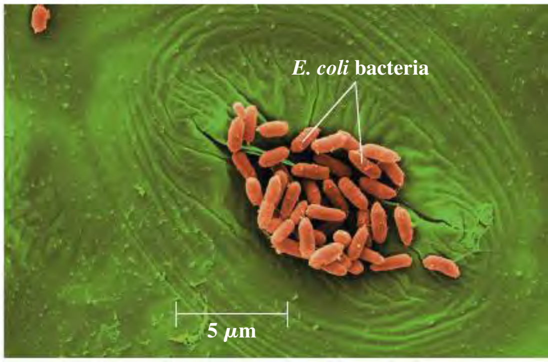
We might think that when the electron wavelength is 0.01 nm (as it is in Example 39.3), the resolution would also be about 0.01 nm. In fact, it is seldom better than 0.1 nm, in part because the focal length of a magnetic lens depends on the electron speed, which is never exactly the same for all electrons in the beam.

An important variation is the *scanning electron microscope*. The electron beam is focused to a very fine line and scanned across the specimen. The beam knocks additional electrons off the specimen wherever it hits. These ejected electrons are collected by an anode that is kept at a potential a few hundred volts positive with respect to the specimen. The current of ejected electrons flowing to the collecting anode varies as the microscope beam sweeps across the specimen. The varying strength of the current is then used to create a "map" of the scanned specimen, and this map forms a greatly magnified image of the specimen.

This scheme has several advantages. The specimen can be thick because the beam does not need to pass through it. Also, the knock-off electron production depends on the *angle* at which the beam strikes the surface. Thus scanning electron micrographs have an appearance that is much more three-dimensional than conventional visible-light micrographs (**Fig. 39.6**). The resolution is typically of the order of 10 nm, not as good as a transmission electron microscope but still much finer than the best optical microscopes.

TEST YOUR UNDERSTANDING OF SECTION 39.1 (a) A proton has a slightly smaller mass than a neutron. Compared to the neutron described in Example 39.2, would a proton of the same wavelength have (i) more kinetic energy; (ii) less kinetic energy; or (iii) the same kinetic energy? (b) Example 39.1 shows that to give electrons a wavelength of 1.7×10^{-10} m, they must be accelerated from rest through a voltage of 54 V and so acquire a kinetic energy of 54 eV. Does a photon of this same energy also have a wavelength of 1.7×10^{-10} m? ▌

39.6 This scanning electron microscope image shows *Escherichia coli* bacteria crowded into a stoma, or respiration opening, on the surface of a lettuce leaf. (False color has been added.) If not washed off before the lettuce is eaten, these bacteria can be a health hazard. The *transmission* electron micrograph that opens this chapter shows a greatly magnified view of the surface of an *E. coli* bacterium.

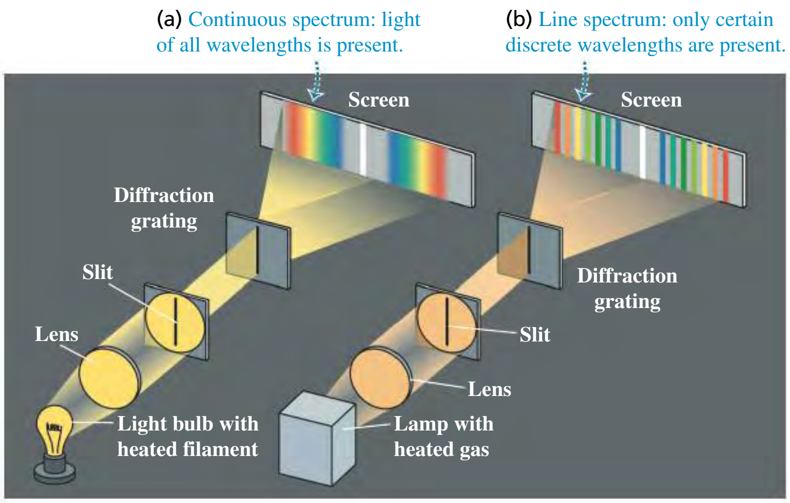


39.2 THE NUCLEAR ATOM AND ATOMIC SPECTRA

Every neutral atom contains at least one electron. How does the wave aspect of electrons affect atomic structure? As we will see, it is crucial for understanding not only the structure of atoms but also how they interact with light. Historically, the quest to understand the nature of the atom was intimately linked with both the idea that electrons have wave characteristics and the notion that light has particle characteristics. Before we explore how these ideas shaped atomic theory, it's useful to look at what was known about atoms—as well as what remained mysterious—by the first decade of the 20th century.

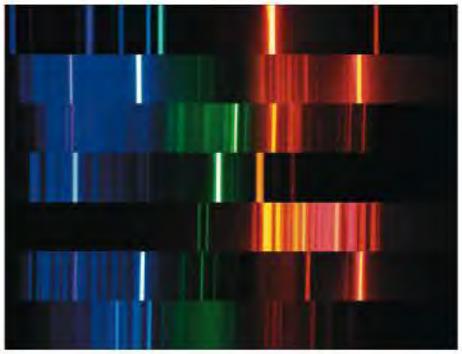
Line Spectra

Heated materials emit light, and different materials emit different kinds of light. The coils of a toaster glow red when in operation, the flame of a match has a characteristic yellow color, and the flame from a gas range is a distinct blue. To analyze these different types of light, we can use a prism or a diffraction grating to separate the various wavelengths in a beam of light into a spectrum. If the light source is a hot solid (such as the filament of an incandescent light bulb) or liquid, the spectrum is *continuous;* light of all wavelengths is present (**Fig. 39.7a**). But if the source is a heated *gas,* such as the neon in a sign or the sodium vapor formed when table salt is thrown into a campfire, the spectrum includes only a few colors in the form of isolated sharp parallel lines (Fig. 39.7b). (Each "line" is



39.7 (a) Continuous spectrum produced by a glowing light bulb filament.
(b) Emission line spectrum emitted by a lamp containing a heated gas.

Application Using Spectra to Analyze an Interstellar Gas Cloud The light from this glowing gas cloud—located in the Small Magellanic Cloud, a small satellite galaxy of the Milky Way some 200,000 light-years (1.9×10^{18} km) from earth—has an emission line spectrum. Despite its immense distance, astronomers can tell that this cloud is composed mostly of hydrogen because its spectrum is dominated by red light at a wavelength of 656.3 nm, a wavelength emitted by hydrogen and no other element.



39.8 The emission line spectra of several kinds of atoms and molecules. No two are alike. Note that the spectrum of water vapor (H_2O) is similar to that of hydrogen (H_2), but there are important differences that make it straightforward to distinguish these two spectra.

an image of the spectrograph slit, deviated through an angle that depends on the wavelength of the light forming that image; see Section 36.5.) A spectrum of this sort is called an **emission line spectrum,** and the lines are called **spectral lines.** Each spectral line corresponds to a definite wavelength and frequency.

It was discovered early in the 19th century that each element in its gaseous state has a unique set of wavelengths in its line spectrum. The spectrum of hydrogen always contains a certain set of wavelengths; mercury produces a different set, neon still another, and so on (**Fig. 39.8**). Scientists find the use of spectra to identify elements and compounds to be an invaluable tool. For instance, astronomers have detected the spectra from more than 100 different molecules in interstellar space, including some that are not found naturally on earth.

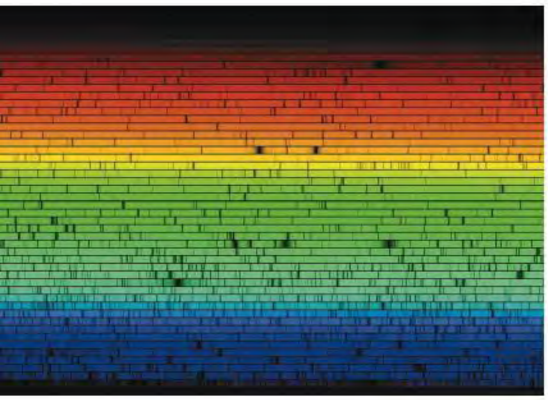
While a *heated* gas selectively *emits* only certain wavelengths, a *cool* gas selectively *absorbs* certain wavelengths. If we pass white (continuous-spectrum) light through a gas and look at the *transmitted* light with a spectrometer, we find a series of dark lines corresponding to the wavelengths that have been absorbed (**Fig. 39.9**). This is called an **absorption line spectrum.** What's more, a given kind of atom or molecule absorbs the *same* characteristic set of wavelengths when it's cool as it emits when heated. Hence scientists can use absorption line spectra to identify substances in the same manner that they use emission line spectra.

As useful as emission line spectra and absorption line spectra are, they presented a quandary to scientists: *Why* does a given kind of atom emit and absorb only certain very specific wavelengths? To answer this question, we need to have a better idea of what the inside of an atom is like. We know that atoms are much smaller than the wavelengths of visible light, so there is no hope of actually using that light to *see* an atom. But we can still describe how the mass and electric charge are distributed throughout the volume of the atom.

Here's where things stood in 1910. In 1897 the English physicist J. J. Thomson had discovered the electron and measured its charge-to-mass ratio e/m. By 1909, the American physicist Robert Millikan had made the first measurements of the electron charge $-e$. These and other experiments showed that almost all the mass of an atom had to be associated with the *positive* charge, not with the electrons. It was also known that the overall size of atoms is of the order of 10^{-10} m and that all atoms except hydrogen contain more than one electron.

In 1910 the best available model of atomic structure was one developed by Thomson. He envisioned the atom as a sphere of some as yet unidentified positively charged substance, within which the electrons were embedded like raisins in cake. This model offered an explanation for line spectra. If the atom collided with another atom, as in a heated gas, each electron would oscillate around its

39.9 The absorption line spectrum of the sun. (The spectrum "lines" read from left to right and from top to bottom, like text on a page.) The spectrum is produced by the sun's relatively cool atmosphere, which absorbs photons from deeper, hotter layers. The absorption lines thus indicate what kinds of atoms are present in the solar atmosphere.



equilibrium position with a characteristic frequency and emit electromagnetic radiation with that frequency. If the atom were illuminated with light of many frequencies, each electron would selectively absorb only light whose frequency matched the electron's natural oscillation frequency. (This is the phenomenon of resonance that we discussed in Section 14.8.)

Rutherford's Exploration of the Atom

The first experiments designed to test Thomson's model by probing the interior structure of the atom were carried out in 1910–1911 by Ernest Rutherford (**Fig. 39.10**) and two of his students, Hans Geiger and Ernest Marsden, at the University of Manchester in England. These experiments consisted of shooting a beam of charged particles at thin foils of various elements and observing how the foil deflected the particles.

The particle accelerators now in common use in laboratories had not yet been invented, and Rutherford's projectiles were *alpha particles* emitted from naturally radioactive elements. The nature of these alpha particles was not completely understood, but it was known that they are ejected from unstable nuclei with speeds of the order of 10^7 m/s, are positively charged, and can travel several centimeters through air or 0.1 mm or so through solid matter before they are brought to rest by collisions.

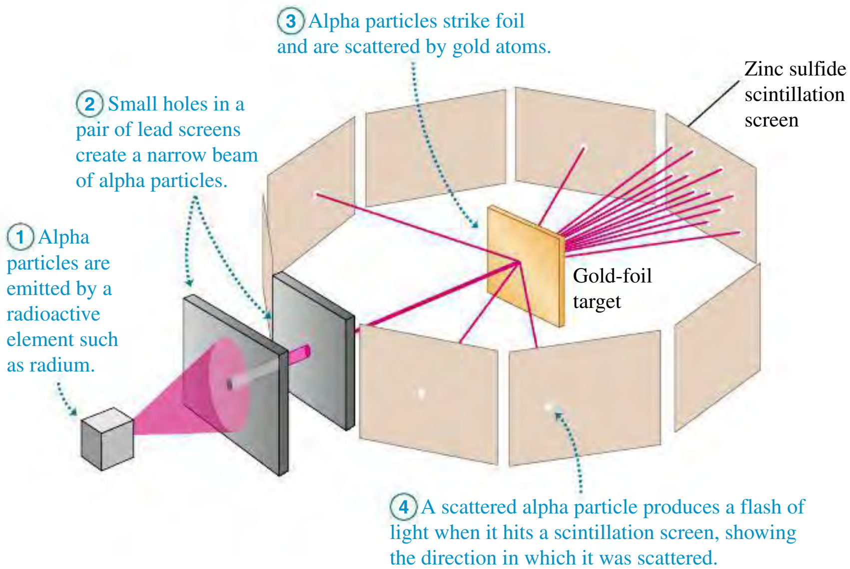
Figure 39.11 is a schematic view of Rutherford's experimental setup. A radioactive substance at the left emits alpha particles. Thick lead screens stop all particles except those in a narrow beam. The beam passes through the foil target (consisting of gold, silver, or copper) and strikes screens coated with zinc sulfide, creating a momentary flash, or *scintillation*. Rutherford and his students counted the numbers of particles deflected through various angles.

The atoms in a metal foil are packed together like marbles in a box (not spaced apart). Because the particle beam passes through the foil, the alpha particles must pass through the interior of atoms. Within an atom, the charged alpha particle will interact with the electrons and the positive charge. (Because the *total* charge of the atom is zero, alpha particles feel little electric force outside an atom.) An electron has about 7300 times less mass than an alpha particle, so momentum considerations indicate that the atom's electrons cannot appreciably deflect the alpha particle—any more than a swarm of gnats deflects a tossed pebble. Any deflection will be due to the positively charged material that makes up almost all of the atom's mass.

39.10 Born in New Zealand, Ernest Rutherford (1871–1937) spent his professional life in England and Canada. Before carrying out the experiments that established the existence of atomic nuclei, he shared (with Frederick Soddy) the 1908 Nobel Prize in chemistry for showing that radioactivity results from the disintegration of atoms.

MP

PhET: Rutherford Scattering

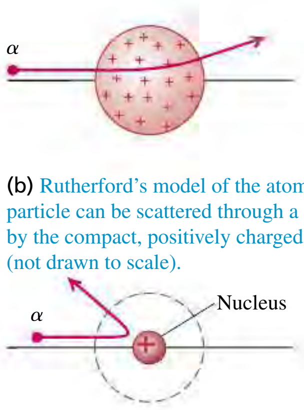


39.11 The Rutherford scattering experiments investigated what happens to alpha particles fired at a thin gold foil. The results of this experiment helped reveal the structure of atoms.

39.12 A comparison of Thomson's and Rutherford's models of the atom.

(a) Thomson's model of the atom: An alpha particle is scattered through only a small angle.

(b) Rutherford's model of the atom: An alpha particle can be scattered through a large angle by the compact, positively charged nucleus (not drawn to scale).



In the Thomson model, the positive charge and the negative electrons are distributed through the whole atom. Hence the electric field inside the atom should be quite small, and the electric force on an alpha particle that enters the atom should be quite weak. The maximum deflection to be expected is then only a few degrees (**Fig. 39.12a**). The results of the Rutherford experiments were *very* different from the Thomson prediction. Some alpha particles were scattered by nearly 180°—that is, almost straight backward (Fig. 39.12b). Rutherford later wrote:

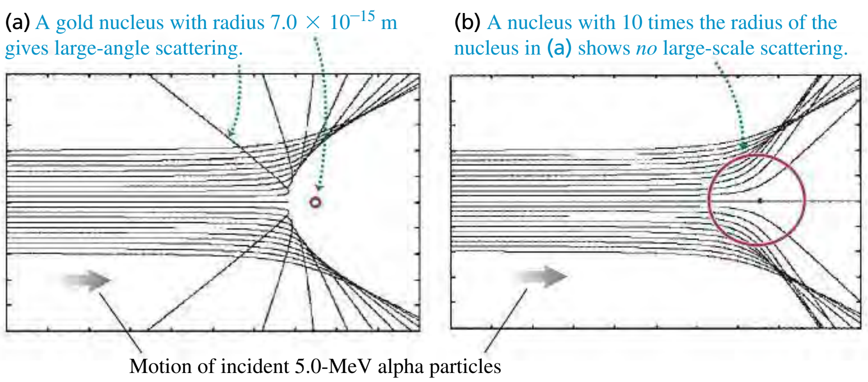
> **It was quite the most incredible event that ever happened to me in my life. It was almost as incredible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you.**

Clearly the Thomson model was wrong and a new model was needed. Suppose the positive charge, instead of being distributed through a sphere with atomic dimensions (of the order of 10^{-10} m), is all concentrated in a much *smaller* volume. Then it would act like a point charge down to much smaller distances. The maximum electric field repelling the alpha particle would be much larger, and the amazing large-angle scattering that Rutherford observed could occur. Rutherford developed this model and called the concentration of positive charge the **nucleus.** He again computed the numbers of particles expected to be scattered through various angles. Within the accuracy of his experiments, the computed and measured results agreed, down to distances of the order of 10^{-14} m. His experiments therefore established that the atom does have a nucleus—a very small, very dense structure, no larger than 10^{-14} m in diameter. The nucleus occupies only about 10^{-12} of the total volume of the atom or less, but it contains *all* the positive charge and at least 99.95% of the total mass of the atom.

Figure 39.13 shows a computer simulation of alpha particles with a kinetic energy of 5.0 MeV being scattered from a gold nucleus of radius 7.0×10^{-15} m (the actual value) and from a nucleus with a hypothetical radius ten times larger. In the second case there is *no* large-angle scattering. The presence of large-angle scattering in Rutherford's experiments thus attested to the small size of the nucleus.

Later experiments showed that all nuclei are composed of positively charged protons (discovered in 1918) and electrically neutral neutrons (discovered in 1930). For example, the gold atoms in Rutherford's experiments have 79 protons and 118 neutrons. In fact, an alpha particle is itself the nucleus of a helium atom, with two protons and two neutrons. It is much more massive than an electron but only about 2% as massive as a gold nucleus, which helps explain why alpha particles are scattered by gold nuclei but not by electrons.

39.13 Computer simulation of scattering of 5.0-MeV alpha particles from a gold nucleus. Each curve shows a possible alpha-particle trajectory. (a) The scattering curves match Rutherford's experimental data if a radius of 7.0×10^{-15} m is assumed for a gold nucleus. (b) A model with a much larger radius for the gold nucleus does not match the data.



EXAMPLE 39.4 A RUTHERFORD EXPERIMENT

SOLUTION

An alpha particle (charge $2e$) is aimed directly at a gold nucleus (charge $79e$). What minimum initial kinetic energy must the alpha particle have to approach within 5.0×10^{-14} m of the center of the gold nucleus before reversing direction? Assume that the gold nucleus, which has about 50 times the mass of an alpha particle, remains at rest.

SOLUTION

IDENTIFY: The repulsive electric force exerted by the gold nucleus makes the alpha particle slow to a halt as it approaches, then reverse direction. This force is conservative, so the total mechanical energy (kinetic energy of the alpha particle plus electric potential energy of the system) is conserved.

SET UP: Let point 1 be the initial position of the alpha particle, very far from the gold nucleus, and let point 2 be 5.0×10^{-14} m from the center of the gold nucleus. Our target variable is the kinetic energy K_1 of the alpha particle at point 1 that allows it to reach point 2 with $K_2 = 0$. To find this we'll use the law of conservation of energy and Eq. (23.9) for electric potential energy, $U = qq_0/4\pi\epsilon_0 r$.

EXECUTE: At point 1 the separation r of the alpha particle and gold nucleus is effectively infinite, so from Eq. (23.9) $U_1 = 0$. At point 2 the potential energy is

$$U_2 = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r}$$
$$= (9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2})\frac{(2)(79)(1.60 \times 10^{-19}\ \mathrm{C})^2}{5.0 \times 10^{-14}\ \mathrm{m}}$$
$$= 7.3 \times 10^{-13}\ \mathrm{J} = 4.6 \times 10^6\ \mathrm{eV} = 4.6\ \mathrm{MeV}$$

In accordance with energy conservation, $K_1 + U_1 = K_2 + U_2$, so $K_1 = K_2 + U_2 - U_1 = 0 + 4.6\ \mathrm{MeV} - 0 = 4.6\ \mathrm{MeV}$. Thus, to approach within 5.0×10^{-14} m, the alpha particle must have initial kinetic energy $K_1 = 4.6$ MeV.

EVALUATE: Alpha particles emitted from naturally occurring radioactive elements typically have energies in the range 4 to 6 MeV. For example, the common isotope of radium, $^{226}\mathrm{Ra}$, emits an alpha particle with energy 4.78 MeV.

Was it valid to assume that the gold nucleus remains at rest? To find out, note that when the alpha particle stops momentarily, all of its initial momentum has been transferred to the gold nucleus. An alpha particle has a mass $m_\alpha = 6.64 \times 10^{-27}$ kg; if its initial kinetic energy $K_1 = \frac{1}{2}mv_1^2$ is 7.3×10^{-13} J, you can show that its initial speed is $v_1 = 1.5 \times 10^7$ m/s and its initial momentum is $p_1 = m_\alpha v_1 = 9.8 \times 10^{-20}\ \mathrm{kg \cdot m/s}$. A gold nucleus (mass $m_{\mathrm{Au}} = 3.27 \times 10^{-25}$ kg) with this much momentum has a much slower speed $v_{\mathrm{Au}} = 3.0 \times 10^5$ m/s and kinetic energy $K_{\mathrm{Au}} = \frac{1}{2}mv_{\mathrm{Au}}{}^2 = 1.5 \times 10^{-14}\ \mathrm{J} = 0.092$ MeV. This *recoil kinetic energy* of the gold nucleus is only 2% of the total energy in this situation, so we are justified in ignoring it.

The Failure of Classical Physics

Rutherford's discovery of the atomic nucleus raised a serious question: What prevented the negatively charged electrons from falling into the positively charged nucleus due to the strong electrostatic attraction? Rutherford suggested that perhaps the electrons *revolve* in orbits about the nucleus, just as the planets revolve around the sun.

But according to classical electromagnetic theory, any accelerating electric charge (either oscillating or revolving) radiates electromagnetic waves. An example is the radiation from an oscillating point charge that we depicted in Fig. 32.3 (Section 32.1). An electron orbiting inside an atom would always have a centripetal acceleration toward the nucleus, and so should be emitting radiation *at all times.* The energy of an orbiting electron should therefore decrease continuously, its orbit should become smaller and smaller, and it should spiral into the nucleus within a fraction of a second (**Fig. 39.14**). Even worse, according to classical theory the *frequency* of the electromagnetic waves emitted should equal the frequency of revolution. As the electrons radiated energy, their angular speeds would change continuously, and they would emit a *continuous* spectrum (a mixture of all frequencies), not the *line* spectrum actually observed.

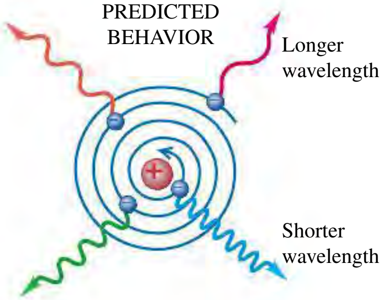
Thus Rutherford's model of electrons orbiting the nucleus, which is based on Newtonian mechanics and classical electromagnetic theory, makes three entirely *wrong* predictions about atoms: They should emit light continuously, they should be unstable, and the light they emit should have a continuous spectrum. Clearly a radical reappraisal of physics on the scale of the atom was needed. In the next section we will see the bold idea that led to a new understanding of the atom, and see how this idea meshes with de Broglie's no less bold notion that electrons have wave attributes.

39.14 Classical physics makes predictions about the behavior of atoms that do not match reality.

ACCORDING TO CLASSICAL PHYSICS:

- An orbiting electron is accelerating, so it should radiate electromagnetic waves.
- The waves would carry away energy, so the electron should lose energy and spiral inward.
- The electron's angular speed would increase as its orbit shrank, so the frequency of the radiated waves should increase.

Thus, classical physics says that atoms should collapse within a fraction of a second and should emit light with a continuous spectrum as they do so.



IN FACT:

- Atoms are stable.
- They emit light only when excited, and only at specific frequencies (as a line spectrum).

TEST YOUR UNDERSTANDING OF SECTION 39.2 Suppose you repeated Rutherford's scattering experiment with a thin sheet of solid hydrogen in place of the gold foil. (Hydrogen is a solid at temperatures below 14.0 K.) The nucleus of a hydrogen atom is a single proton, with about one-fourth the mass of an alpha particle. Compared to the original experiment with gold foil, would you expect the alpha particles in this experiment to undergo (i) more large-angle scattering; (ii) the same amount of large-angle scattering; or (iii) less large-angle scattering? ❙

39.3 ENERGY LEVELS AND THE BOHR MODEL OF THE ATOM

39.15 Niels Bohr (1885–1962) was a young postdoctoral researcher when he proposed the novel idea that the energy of an atom could have only certain discrete values. He won the 1922 Nobel Prize in physics for these ideas. Bohr went on to make seminal contributions to nuclear physics and to become a passionate advocate for the free exchange of scientific ideas among all nations.

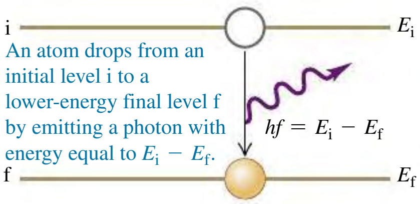
In 1913 a young Danish physicist working with Ernest Rutherford at the University of Manchester made a revolutionary proposal to explain both the stability of atoms and their emission and absorption line spectra. The physicist was Niels Bohr (**Fig. 39.15**), and his innovation was to combine the photon concept that we introduced in Chapter 38 with a fundamentally new idea: The energy of an atom can have only certain particular values. His hypothesis represented a clean break from 19th-century ideas.

Photon Emission and Absorption by Atoms

Bohr's reasoning went like this. The emission line spectrum of an element tells us that atoms of that element emit photons with only certain specific frequencies f and hence certain specific energies $E = hf$. During the emission of a photon, the internal energy of the atom changes by an amount equal to the energy of the photon. Therefore, said Bohr, each atom must be able to exist with only certain specific values of internal energy. Each atom has a set of possible **energy levels.** An atom can have an amount of internal energy equal to any one of these levels, but it *cannot* have an energy *intermediate* between two levels. All isolated atoms of a given element have the same set of energy levels, but atoms of different elements have different sets.

Suppose an atom is raised, or *excited*, to a high energy level. (In a hot gas this happens when fast-moving atoms undergo inelastic collisions with each other or with the walls of the gas container. In an electric discharge tube, such as those used in a neon light fixture, atoms are excited by collisions with fast-moving electrons.) According to Bohr, an excited atom can make a *transition* from one energy level to a lower level by emitting a photon with energy equal to the energy *difference* between the initial and final levels (**Fig. 39.16**):

39.16 An excited atom emitting a photon.



Energy of emitted photon; Planck's constant; Photon frequency; Speed of light in vacuum; Photon wavelength; Final energy of atom after transition; Initial energy of atom before transition

$$hf = \frac{hc}{\lambda} = E_i - E_f \tag{39.5}$$

For example, an excited lithium atom emits red light with wavelength $\lambda = 671$ nm. The corresponding photon energy is

$$E = \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34}\ \mathrm{J \cdot s})(3.00 \times 10^{8}\ \mathrm{m/s})}{671 \times 10^{-9}\ \mathrm{m}}$$
$$= 2.96 \times 10^{-19}\ \mathrm{J} = 1.85\ \mathrm{eV}$$

This photon is emitted during a transition like that shown in Fig. 39.16 between two levels of the atom that differ in energy by $E_i - E_f = 1.85$ eV.

Emission line spectra (Fig. 39.8) show that many different wavelengths are emitted by each atom. Hence each kind of atom must have a number of energy levels, with different spacings in energy between them. Each wavelength in the spectrum corresponds to a transition between two specific atomic energy levels.

MP

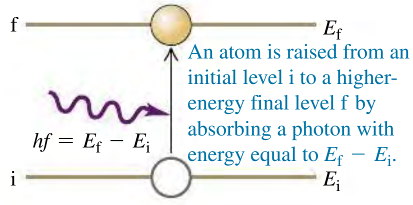
PhET: Models of the Hydrogen Atom

CAUTION **Producing a line spectrum** The lines of an emission line spectrum, such as the helium spectrum shown at the top of Fig. 39.8, are *not* all produced by a single atom. The sample of helium gas that produced the spectrum in Fig. 39.8 contained a large number of helium atoms; these were excited in an electric discharge tube to various energy levels. The spectrum of the gas shows the light emitted from all the different transitions that occurred in different atoms of the sample.

The observation that atoms are stable means that each atom has a *lowest* energy level, called the **ground level.** Levels with energies greater than the ground level are called **excited levels.** An atom in an excited level, called an *excited atom,* can make a transition into the ground level by emitting a photon as in Fig. 39.16. But since there are no levels below the ground level, an atom in the ground level cannot lose energy and so cannot emit a photon.

Collisions are not the only way that an atom's energy can be raised from one level to a higher level. If an atom initially in the lower energy level in Fig. 39.16 is struck by a photon with just the right amount of energy, the photon can be *absorbed* and the atom will end up in the higher level (**Fig. 39.17**). As an example, we previously mentioned two levels in the lithium atom with an energy difference of 1.85 eV. For a photon to be absorbed and excite the atom from the lower level to the higher one, the photon must have an energy of 1.85 eV and a wavelength of 671 nm. In other words, an atom *absorbs* the same wavelengths that it *emits*. This explains the correspondence between an element's emission line spectrum and its absorption line spectrum that we described in Section 39.2.

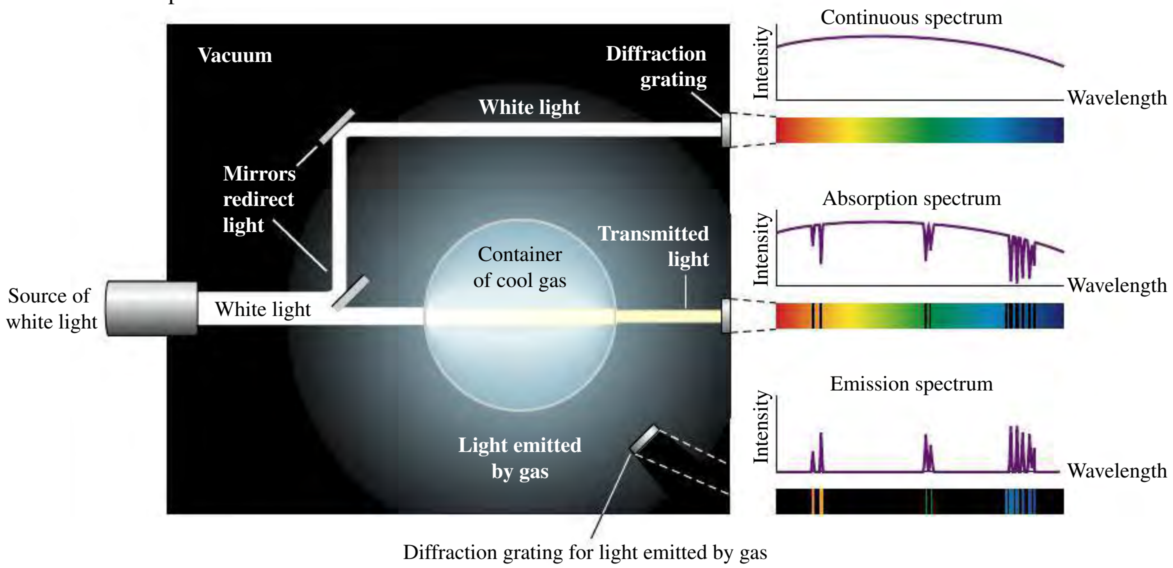
39.17 An atom absorbing a photon. (Compare with Fig. 39.16.)



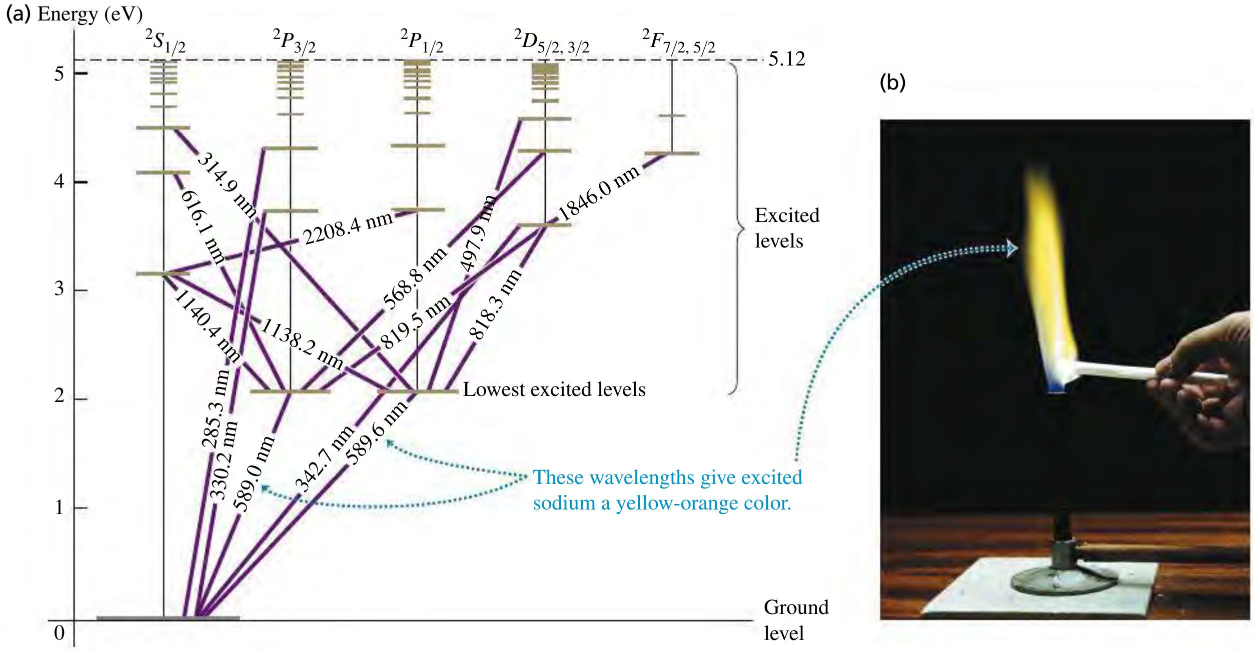
Note that a lithium atom *cannot* absorb a photon with a slightly longer wavelength (say, 672 nm) or one with a slightly shorter wavelength (say, 670 nm). That's because these photons have, respectively, slightly too little or slightly too much energy to raise the atom's energy from one level to the next, and an atom cannot have an energy that's intermediate between levels. This explains why absorption line spectra have distinct dark lines (see Fig. 39.9): Atoms can absorb only photons with specific wavelengths.

An atom that's been excited into a high energy level, either by photon absorption or by collisions, does not stay there for long. After a short time, called the *lifetime* of the level (typically around 10^{-8} s), the excited atom will emit a photon and make a transition into a lower excited level or the ground level. A cool gas that's illuminated by white light to make an *absorption* line spectrum thus also produces an *emission* line spectrum when viewed from the side, since when the atoms de-excite they emit photons in all directions (**Fig. 39.18**). To keep a gas of

39.18 When a beam of white light with a continuous spectrum passes through a cool gas, the transmitted light has an absorption spectrum. The absorbed light energy excites the gas and causes it to emit light of its own, which has an emission spectrum.



39.19 (a) Energy levels of the sodium atom relative to the ground level. Numbers on the lines between levels are wavelengths of the light emitted or absorbed during transitions between those levels. The column labels, such as $^2S_{1/2}$, refer to certain quantum states of the atom. (b) When a sodium compound is placed in a flame, sodium atoms are excited into the lowest excited levels. As they drop back to the ground level, the atoms emit photons of yellow-orange light with wavelengths 589.0 and 589.6 nm.



atoms glowing, you have to continually provide energy to the gas in order to re-excite atoms so that they can emit more photons. If you turn off the energy supply (for example, by turning off the electric current through a neon light fixture, or by shutting off the light source in Fig. 39.18), the atoms drop back into their ground levels and cease to emit light.

DEMO

By working backward from the observed emission line spectrum of an element, physicists can deduce the arrangement of energy levels in an atom of that element. As an example, **Fig. 39.19a** shows some of the energy levels for a sodium atom. You may have noticed the yellow-orange light emitted by sodium vapor street lights. Sodium atoms emit this characteristic yellow-orange light with wavelengths 589.0 and 589.6 nm when they make transitions from the two closely spaced levels labeled *lowest excited levels* to the ground level. A standard test for the presence of sodium compounds is to look for this yellow-orange light from a sample placed in a flame (Fig. 39.19b).

EXAMPLE 39.5 EMISSION AND ABSORPTION SPECTRA

SOLUTION

A hypothetical atom (**Fig. 39.20a**) has energy levels at 0.00 eV (the ground level), 1.00 eV, and 3.00 eV. (a) What are the frequencies and wavelengths of the spectral lines this atom can emit when excited? (b) What wavelengths can this atom absorb if it is in its ground level?

SOLUTION

IDENTIFY and SET UP: Energy is conserved when a photon is emitted or absorbed. In each transition the photon energy is equal to the difference between the energies of the levels involved in the transition.

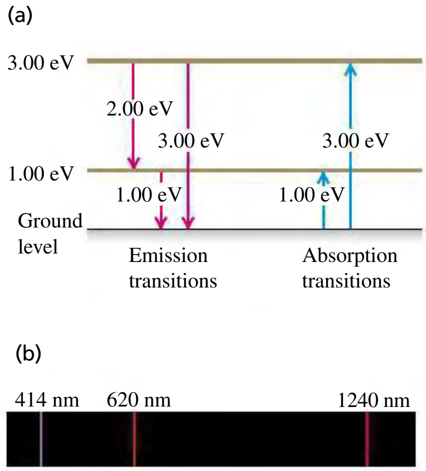
EXECUTE: (a) The possible energies of emitted photons are 1.00 eV, 2.00 eV, and 3.00 eV. For 1.00 eV, Eq. (39.2) gives

$$f = \frac{E}{h} = \frac{1.00\ \text{eV}}{4.136 \times 10^{-15}\ \text{eV} \cdot \text{s}} = 2.42 \times 10^{14}\ \text{Hz}$$

For 2.00 eV and 3.00 eV, $f = 4.84 \times 10^{14}$ Hz and 7.25×10^{14} Hz, respectively. For 1.00-eV photons,

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^{8}\ \text{m/s}}{2.42 \times 10^{14}\ \text{Hz}} = 1.24 \times 10^{-6}\ \text{m} = 1240\ \text{nm}$$

39.20 (a) Energy-level diagram for the hypothetical atom, showing the possible transitions for emission from excited levels and for absorption from the ground level. (b) Emission spectrum of this hypothetical atom.



This is in the infrared region of the spectrum (Fig. 39.20b). For 2.00 eV and 3.00 eV, the wavelengths are 620 nm (red) and 414 nm (violet), respectively.

(b) From the ground level, only a 1.00-eV or a 3.00-eV photon can be absorbed (Fig. 39.20a); a 2.00-eV photon cannot be absorbed because the atom has no energy level 2.00 eV above the ground level. Passing light from a hot solid through a gas of these hypothetical atoms (almost all of which would be in the ground level if the gas were cool) would yield a continuous spectrum with dark absorption lines at 1240 nm and 414 nm.

EVALUATE: Note that if a gas of these atoms were at a sufficiently high temperature, collisions would excite a number of atoms into the 1.00-eV energy level. Such excited atoms *can* absorb 2.00-eV photons, as Fig. 39.20a shows, and an absorption line at 620 nm would appear in the spectrum. Thus the observed spectrum of a given substance depends on its energy levels and its temperature.

Suppose we take a gas of the hypothetical atoms in Example 39.5 and illuminate it with violet light of wavelength 414 nm. Atoms in the ground level can absorb this photon and make a transition to the 3.00-eV level. Some of these atoms will make a transition back to the ground level by emitting a 414-nm photon. But other atoms will return to the ground level in two steps, first emitting a 620-nm photon to transition to the 1.00-eV level, then a 1240-nm photon to transition back to the ground level. Thus this gas will emit longer-wavelength radiation than it absorbs, a phenomenon called *fluorescence*. For example, the electric discharge in a fluorescent lamp causes the mercury vapor in the tube to emit ultraviolet radiation. This radiation is absorbed by the atoms of the coating on the inside of the tube. The coating atoms then re-emit light in the longer-wavelength, visible portion of the spectrum. Fluorescent lamps are more efficient than incandescent lamps in converting electrical energy to visible light because they do not waste as much energy producing (invisible) infrared photons.

Our discussion of energy levels and spectra has concentrated on *atoms,* but the same ideas apply to *molecules*. Figure 39.8 shows the emission line spectra of two molecules, hydrogen (H_2) and water (H_2O). Just as for sodium or other atoms, physicists can work backward from these molecular spectra and deduce the arrangement of energy levels for each kind of molecule. We'll return to molecules and molecular structure in Chapter 42.

BIO Application Fish Fluorescence
When illuminated by blue light, this tropical lizardfish (family Synodontidae) fluoresces and emits longer-wavelength green light. The fluorescence may be a sexual signal or a way for the fish to camouflage itself among coral (which also have a green fluorescence).

The Franck-Hertz Experiment: Are Energy Levels Real?

Are atomic energy levels real, or just a convenient fiction that helps us to explain spectra? In 1914, the German physicists James Franck and Gustav Hertz answered this question when they found direct experimental evidence for the existence of atomic energy levels.

Franck and Hertz studied the motion of electrons through mercury vapor under the action of an electric field. They found that when the electron kinetic energy was 4.9 eV or greater, the vapor emitted ultraviolet light of wavelength 250 nm. Suppose mercury atoms have an excited energy level 4.9 eV above the ground level. An atom can be raised to this level by collision with an electron;

it later decays back to the ground level by emitting a photon. From the photon formula $E = hc/\lambda$, the wavelength of the photon should be

$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\ \text{eV} \cdot \text{s})(3.00 \times 10^{8}\ \text{m/s})}{4.9\ \text{eV}}$$
$$= 2.5 \times 10^{-7}\ \text{m} = 250\ \text{nm}$$

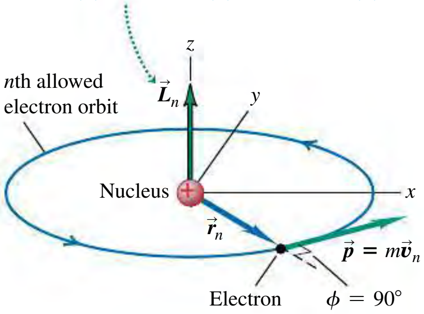
This is equal to the wavelength that Franck and Hertz measured, which demonstrates that this energy level actually exists in the mercury atom. Similar experiments with other atoms yield the same kind of evidence for atomic energy levels. Franck and Hertz shared the 1925 Nobel Prize in physics for their research.

Electron Waves and the Bohr Model of Hydrogen

Bohr's hypothesis established the relationship between atomic spectra and energy levels. By itself, however, it provided no general principles for *predicting* the energy levels of a particular atom. Bohr addressed this problem for the case of the simplest atom, hydrogen, which has just one electron.

In the **Bohr model,** Bohr postulated that each energy level of a hydrogen atom corresponds to a specific *stable* circular orbit of the electron around the nucleus. In a break with classical physics, Bohr further postulated that an electron in such an orbit does *not* radiate. Instead, an atom radiates energy only when an electron makes a transition from an orbit of energy E_i to a different orbit with lower energy E_f, emitting a photon of energy $hf = E_i - E_f$ in the process.

39.21 Calculating the angular momentum of an electron in a circular orbit around an atomic nucleus.



As a result of a rather complicated argument that related the angular frequency of the light emitted to the angular speed of the electron in highly excited energy levels, Bohr found that the magnitude of the electron's angular momentum is *quantized;* that is, this magnitude must be an integral multiple of $h/2\pi$. (Because $1\ \text{J} = 1\ \text{kg} \cdot \text{m}^2/\text{s}^2$, the SI units of Planck's constant h, $\text{J} \cdot \text{s}$, are the same as the SI units of angular momentum, usually written as $\text{kg} \cdot \text{m}^2/\text{s}$.) Let's number the orbits by an integer n, where $n = 1, 2, 3, \ldots$, and call the radius of orbit n r_n and the speed of the electron in that orbit v_n. The value of n for each orbit is called the **principal quantum number** for the orbit. From Section 10.5, Eq. (10.25), the magnitude of the angular momentum of an electron of mass m in such an orbit is $L_n = mv_nr_n$ (**Fig. 39.21**). So Bohr's argument led to

Quantization of angular momentum:

$$\overbrace{L_n}^{\text{Orbital angular momentum}} = \underbrace{m}_{\text{Electron mass}}\underbrace{v_n}_{\text{Electron speed}}\underbrace{r_n}_{\text{Electron orbital radius}} = \overbrace{n}^{\text{Principal quantum number } (n = 1, 2, 3, \ldots)}\frac{\overbrace{h}^{\text{Planck's constant}}}{2\pi} \tag{39.6}$$

Instead of going through Bohr's argument to justify Eq. (39.6), we can use de Broglie's picture of electron waves. Rather than visualizing the orbiting electron as a particle moving around the nucleus in a circular path, think of it as a sinusoidal *standing wave* with wavelength λ that extends around the circle. A standing wave on a string transmits no energy (see Section 15.7), and electrons in Bohr's orbits radiate no energy. For the wave to "come out even" and join onto itself smoothly, the circumference of this circle must include some *whole number* of wavelengths, as **Fig. 39.22** suggests. Hence for an orbit with radius r_n and circumference $2\pi r_n$, we must have $2\pi r_n = n\lambda_n$, where λ_n is the wavelength and $n = 1, 2, 3, \ldots$. According to the de Broglie relationship, Eq. (39.1), the wavelength of a particle with rest mass m moving with nonrelativistic speed v_n is $\lambda_n = h/mv_n$. Combining $2\pi r_n = n\lambda_n$ and $\lambda_n = h/mv_n$, we find $2\pi r_n = nh/mv_n$ or

$$mv_nr_n = n\frac{h}{2\pi}$$

This is the same as Bohr's result, Eq. (39.6). Thus a wave picture of the electron leads naturally to the quantization of the electron's angular momentum.

Now let's consider a model of the hydrogen atom that is Newtonian in spirit but incorporates this quantization assumption (**Fig. 39.23**). This atom consists of a single electron with mass m and charge $-e$ in a circular orbit around a single proton with charge $+e$. The proton is nearly 2000 times as massive as the electron, so we can assume that the proton does not move. We learned in Section 5.4 that when a particle with mass m moves with speed v_n in a circular orbit with radius r_n, its centripetal (inward) acceleration is v_n^2/r_n. According to Newton's second law, a radially inward net force with magnitude $F = mv_n^2/r_n$ is needed to cause this acceleration. We discussed in Section 13.4 how the gravitational attraction provides that inward force for satellite orbits. In hydrogen the force is provided by the electrical attraction between the proton (charge $+e$) and the electron (charge $-e$). From Coulomb's law, Eq. (21.2),

$$F = \frac{1}{4\pi\epsilon_0}\frac{|(+e)(-e)|}{r_n^2} = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n^2}$$

Hence Newton's second law states that

$$\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n^2} = \frac{mv_n^2}{r_n} \qquad (39.7)$$

When we solve Eqs. (39.6) and (39.7) simultaneously for r_n and v_n, we get

Radius of nth orbit in the Bohr model

$$r_n = \epsilon_0 \frac{n^2h^2}{\pi m e^2} \qquad (39.8)$$

Principal quantum number ($n = 1, 2, 3, \ldots$)
Planck's constant
Magnitude of electron charge
Electric constant
Electron mass

Orbital speed in nth orbit in the Bohr model

$$v_n = \frac{1}{\epsilon_0}\frac{e^2}{2nh} \qquad (39.9)$$

Magnitude of electron charge
Planck's constant
Electric constant
Principal quantum number ($n = 1, 2, 3, \ldots$)

Equation (39.8) shows that the orbit radius r_n is proportional to n^2, so the smallest orbit radius corresponds to $n = 1$. We'll denote this minimum radius, called the *Bohr radius,* as a_0:

$$a_0 = \epsilon_0 \frac{h^2}{\pi m e^2} \quad \text{(Bohr radius)} \qquad (39.10)$$

Then we can rewrite Eq. (39.8) as

Radius of nth orbit in the Bohr model

$$r_n = n^2 a_0 \qquad (39.11)$$

Bohr radius
Principal quantum number ($n = 1, 2, 3, \ldots$)

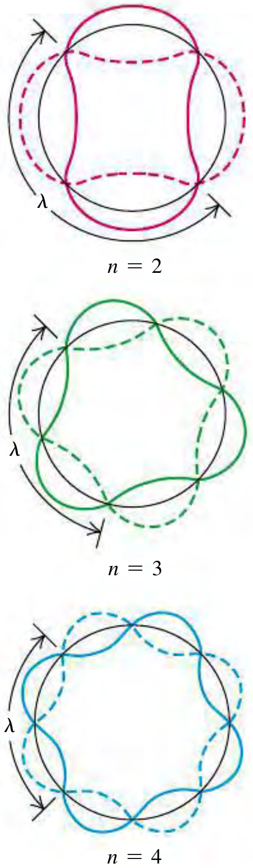
The permitted orbits have radii a_0, $4a_0$, $9a_0$, and so on.

You can find the numerical values of the quantities on the right-hand side of Eq. (39.10) in Appendix F. Using these values, we find that the radius a_0 of the smallest Bohr orbit is

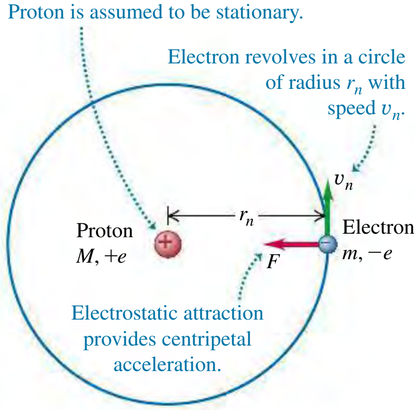
$$a_0 = \frac{(8.854 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)(6.626 \times 10^{-34}\,\text{J}\cdot\text{s})^2}{\pi(9.109 \times 10^{-31}\,\text{kg})(1.602 \times 10^{-19}\,\text{C})^2}$$
$$= 5.29 \times 10^{-11}\,\text{m}$$

This gives an atomic diameter of about 10^{-10} m $= 0.1$ nm, which is consistent with atomic dimensions estimated by other methods.

39.22 The idea of fitting a standing electron wave around a circular orbit. For the wave to join onto itself smoothly, the circumference of the orbit must be an integral number n of wavelengths.



39.23 The Bohr model of the hydrogen atom.



Equation (39.9) shows that the orbital speed v_n is proportional to $1/n$. Hence the greater the value of n, the larger the orbital radius of the electron and the slower its orbital speed. (We saw the same relationship between orbital radius and speed for satellite orbits in Section 13.4.) We leave it to you to calculate the speed in the $n = 1$ orbit, which is the greatest possible speed of the electron in the hydrogen atom (see Exercise 39.23); the result is $v_1 = 2.19 \times 10^6$ m/s. This is less than 1% of the speed of light, so relativistic considerations aren't significant.

Hydrogen Energy Levels in the Bohr Model

We can now use Eqs. (39.8) and (39.9) to find the kinetic and potential energies K_n and U_n when the electron is in the orbit with quantum number n:

$$K_n = \tfrac{1}{2}mv_n^2 = \frac{1}{\epsilon_0^2}\frac{me^4}{8n^2h^2} \quad \text{(kinetic energies in the Bohr model)} \tag{39.12}$$

$$U_n = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n} = -\frac{1}{\epsilon_0^2}\frac{me^4}{4n^2h^2} \quad \text{(potential energies in the Bohr model)} \tag{39.13}$$

The electric potential energy is negative because we have taken its value to be zero when the electron is infinitely far from the nucleus. We are interested only in the *differences* in energy between orbits, so the reference position doesn't matter. The total energy E_n is the sum of the kinetic and potential energies:

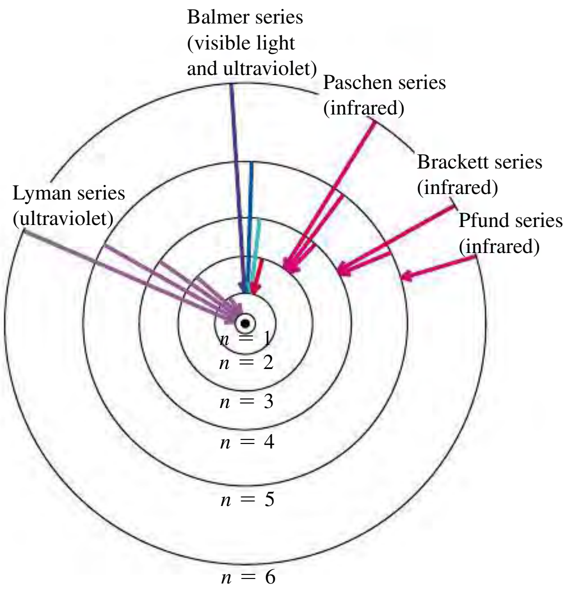
$$E_n = K_n + U_n = -\frac{1}{\epsilon_0^2}\frac{me^4}{8n^2h^2} \quad \text{(total energies in the Bohr model)} \tag{39.14}$$

Since E_n in Eq. (39.14) has a different value for each n, you can see that this equation gives the *energy levels* of the hydrogen atom in the Bohr model. Each distinct orbit corresponds to a distinct energy level.

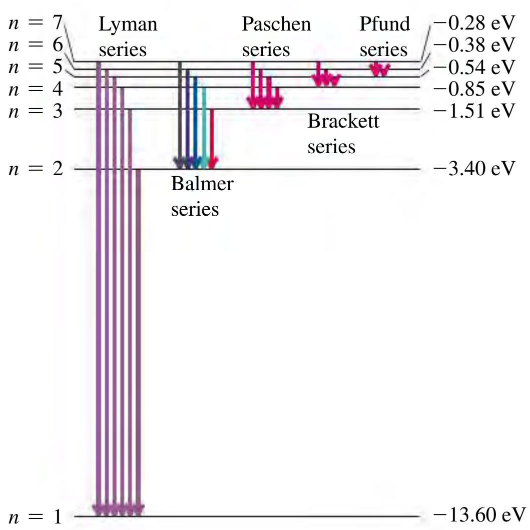
Figure 39.24 depicts the orbits and energy levels. We label the possible energy levels of the atom by values of the quantum number n. For each value of n there are corresponding values of orbit radius r_n, speed v_n, angular momentum

39.24 Two ways to represent the energy levels of the hydrogen atom and the transitions between them. Note that the radius of the nth permitted orbit is actually n^2 times the radius of the $n = 1$ orbit.

(a) Permitted orbits of an electron in the Bohr model of a hydrogen atom (not to scale). Arrows indicate the transitions responsible for some of the lines of various series.



(b) Energy-level diagram for hydrogen, showing some transitions corresponding to the various series



$L_n = nh/2\pi$, and total energy E_n. The energy of the atom is least when $n = 1$ and E_n has its most negative value. This is the *ground level* of the hydrogen atom; it is the level with the smallest orbit, of radius a_0. For $n = 2, 3, \ldots$, the absolute value of E_n is smaller and the energy is progressively larger (less negative).

Figure 39.24 also shows some of the possible transitions from one electron orbit to an orbit of lower energy. Consider a transition from orbit n_U (for "upper") to a smaller orbit n_L (for "lower"), with $n_L < n_U$—or, equivalently, from *level* n_U to a lower *level* n_L. Then the energy hc/λ of the emitted photon of wavelength λ is equal to $E_{n_U} - E_{n_L}$. Before we use this relationship to solve for λ, it's convenient to rewrite Eq. (39.14) for the energies as

Planck's constant; Speed of light in vacuum; Electron mass; Magnitude of electron charge

Total energy for *n*th orbit in the Bohr model

$$E_n = -\frac{hcR}{n^2}, \quad \text{where} \quad R = \frac{me^4}{8\epsilon_0^2 h^3 c} \tag{39.15}$$

Principal quantum number ($n = 1, 2, 3, \ldots$); Rydberg constant; Electric constant

The quantity R in Eq. (39.15) is called the **Rydberg constant** (named for the Swedish physicist Johannes Rydberg, who did pioneering work on the hydrogen spectrum). When we substitute the numerical values of the fundamental physical constants m, c, e, h, and ϵ_0, all of which can be determined quite independently of the Bohr theory, we find that $R = 1.097 \times 10^7\ \text{m}^{-1}$. Now we solve for the wavelength of the photon emitted in a transition from level n_U to level n_L:

$$\frac{hc}{\lambda} = E_{n_U} - E_{n_L} = \left(-\frac{hcR}{n_U^2}\right) - \left(-\frac{hcR}{n_L^2}\right) = hcR\left(\frac{1}{n_L^2} - \frac{1}{n_U^2}\right)$$

$$\frac{1}{\lambda} = R\left(\frac{1}{n_L^2} - \frac{1}{n_U^2}\right) \quad \text{(hydrogen wavelengths in the Bohr model, } n_L < n_U\text{)} \tag{39.16}$$

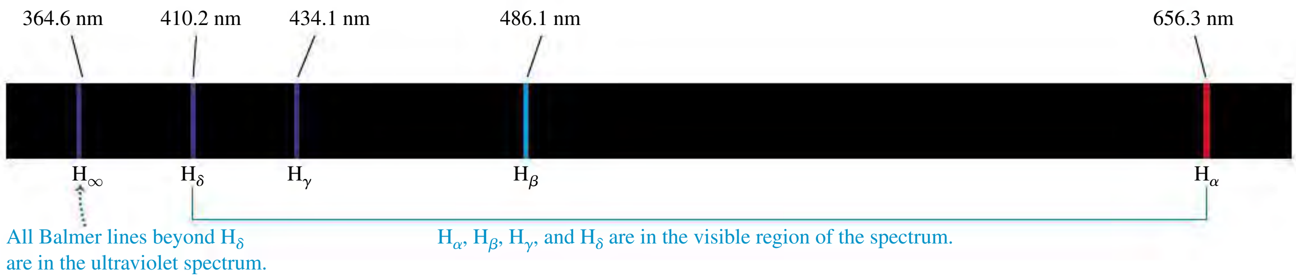
Equation (39.16) is a *theoretical prediction* of the wavelengths found in the *emission* line spectrum of hydrogen atoms. When a hydrogen atom *absorbs* a photon, an electron makes a transition from a level n_L to a *higher* level n_U. This can happen only if the photon energy hc/λ is equal to $E_{n_U} - E_{n_L}$, which is the same condition expressed by Eq. (39.16). So this equation also predicts the wavelengths found in the *absorption* line spectrum of hydrogen.

MP **PhET:** Neon Lights and Other Discharge Lamps

How does this prediction compare with experiment? If $n_L = 2$, corresponding to transitions to the second energy level in Fig. 39.24, the wavelengths predicted by Eq. (39.16)—collectively called the *Balmer series* (**Fig. 39.25**)—are all in the visible and ultraviolet parts of the electromagnetic spectrum. If we let $n_L = 2$ and $n_U = 3$ in Eq. (39.16) we obtain the wavelength of the H_α line:

$$\frac{1}{\lambda} = (1.097 \times 10^7\ \text{m}^{-1})\left(\tfrac{1}{4} - \tfrac{1}{9}\right) \qquad \text{or} \qquad \lambda = 656.3\ \text{nm}$$

39.25 The Balmer series of spectral lines for atomic hydrogen. You can see these same lines in the spectrum of *molecular* hydrogen (H_2) shown in Fig. 39.8, as well as additional lines that are present only when two hydrogen atoms are combined to make a molecule.



DATA *SPEAKS*

The Hydrogen Spectrum

When students were given a problem involving the spectrum of atomic hydrogen, more than 36% gave an incorrect response. Common errors:

- Confusion about energy levels, photon energy, and wavelength. The difference in energy between two energy levels of an atom equals the energy of a photon emitted or absorbed in a transition between these levels. Hence the *greater* the energy difference, the *shorter* the wavelength of the photon.
- Confusion about transitions between energy levels. A transition can "skip over" levels, so the quantum number n can change by more than 1 (for example, when an atom starts in the $n = 5$ level, emits a photon, and ends up in the $n = 2$ level).

With $n_L = 2$ and $n_U = 4$ we obtain the wavelength of the H_β line, and so on. With $n_L = 2$ and $n_U = \infty$ we obtain the shortest wavelength in the series, $\lambda = 364.6$ nm. These theoretical predictions are within 0.1% of the observed hydrogen wavelengths! This close agreement provides very strong and direct confirmation of Bohr's theory.

The Bohr model also predicts many other wavelengths in the hydrogen spectrum, as Fig. 39.24 shows. The observed wavelengths of all of these series, each of which is named for its discoverer, match the predicted values with the same percent accuracy as for the Balmer series. The *Lyman series* of spectral lines is caused by transitions between the ground level and the excited levels, corresponding to $n_L = 1$ and $n_U = 2, 3, 4, \ldots$ in Eq. (39.16). The energy difference between the ground level and any of the excited levels is large, so the emitted photons have wavelengths in the ultraviolet part of the electromagnetic spectrum. Transitions among the higher energy levels involve a much smaller energy difference, so the photons emitted in these transitions have little energy and long, infrared wavelengths. That's the case for both the *Brackett series* ($n_L = 3$ and $n_U = 4, 5, 6, \ldots$, corresponding to transitions between the third and higher energy levels) and the *Pfund series* ($n_L = 4$ and $n_U = 5, 6, 7, \ldots$, with transitions between the fourth and higher energy levels).

Figure 39.24 shows only transitions in which a hydrogen atom emits a photon. But as we discussed previously, the wavelengths of those photons that an atom can *absorb* are the same as those that it can emit. For example, a hydrogen atom in the $n = 2$ level can absorb a 656.3-nm photon and end up in the $n = 3$ level.

One additional test of the Bohr model is its predicted value of the *ionization energy* of the hydrogen atom. This is the energy required to remove the electron completely from the atom. Ionization corresponds to a transition from the ground level ($n = 1$) to an infinitely large orbit radius ($n = \infty$), so the energy that must be added to the atom is $E_\infty - E_1 = 0 - E_1 = -E_1$ (recall that E_1 is negative). Substituting the constants from Appendix F into Eq. (39.15) gives an ionization energy of 13.606 eV. The ionization energy can also be measured directly; the result is 13.60 eV. These two values agree within 0.1%.

EXAMPLE 39.6 EXPLORING THE BOHR MODEL

SOLUTION

Find the kinetic, potential, and total energies of the hydrogen atom in the first excited level, and find the wavelength of the photon emitted in a transition from that level to the ground level.

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of the Bohr model. We use simplified versions of Eqs. (39.12), (39.13), and (39.14) to find the energies of the atom, and Eq. (39.16), $hc/\lambda = E_{n_U} - E_{n_L}$, to find the photon wavelength λ in a transition from $n_U = 2$ (the first excited level) to $n_L = 1$ (the ground level).

EXECUTE: We could evaluate Eqs. (39.12), (39.13), and (39.14) for the nth level by substituting the values of m, e, ϵ_0, and h. But we can simplify the calculation by comparing with Eq. (39.15), which shows that the constant $me^4/8\epsilon_0^2h^2$ that appears in Eqs. (39.12), (39.13), and (39.14) is equal to hcR:

$$\begin{aligned}\frac{me^4}{8\epsilon_0^2h^2} &= hcR \\ &= (6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^8\ \text{m/s}) \\ &\quad \times (1.097 \times 10^7\ \text{m}^{-1}) \\ &= 2.179 \times 10^{-18}\ \text{J} = 13.60\ \text{eV}\end{aligned}$$

This allows us to rewrite Eqs. (39.12), (39.13), and (39.14) as

$$K_n = \frac{13.60\ \text{eV}}{n^2} \qquad U_n = \frac{-27.20\ \text{eV}}{n^2} \qquad E_n = \frac{-13.60\ \text{eV}}{n^2}$$

For the first excited level ($n = 2$), we have $K_2 = 3.40$ eV, $U_2 = -6.80$ eV, and $E_2 = -3.40$ eV. For the ground level ($n = 1$), $E_1 = -13.60$ eV. The energy of the emitted photon is then $E_2 - E_1 = -3.40\ \text{eV} - (-13.60\ \text{eV}) = 10.20$ eV, and

$$\begin{aligned}\lambda = \frac{hc}{E_2 - E_1} &= \frac{(4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(3.00 \times 10^8\ \text{m/s})}{10.20\ \text{eV}} \\ &= 1.22 \times 10^{-7}\ \text{m} = 122\ \text{nm}\end{aligned}$$

This is the wavelength of the Lyman-alpha (L_α) line, the longest-wavelength line in the Lyman series of ultraviolet lines in the hydrogen spectrum (see Fig. 39.24).

EVALUATE: The total mechanical energy for any level is negative and is equal to one-half the potential energy. We found the same energy relationship for Newtonian satellite orbits in Section 13.4. The situations are similar because both the electrostatic and gravitational forces are inversely proportional to $1/r^2$.

Nuclear Motion and the Reduced Mass of an Atom

The Bohr model is so successful that we can justifiably ask why its predictions for the wavelengths and ionization energy of hydrogen differ from the measured values by about 0.1%. The explanation is that we assumed that the nucleus (a proton) remains at rest. However, as **Fig. 39.26** shows, the proton and electron *both* orbit about their common center of mass (see Section 8.5). It turns out that we can take this motion into account by using in Bohr's equations not the electron rest mass m but a quantity called the **reduced mass** m_r of the system. For a system composed of two bodies of masses m_1 and m_2, the reduced mass is

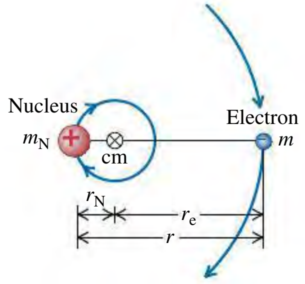
$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{39.17}$$

For ordinary hydrogen we let m_1 equal m and m_2 equal the proton mass, $m_p = 1836.2m$. Thus ordinary hydrogen has a reduced mass of

$$m_r = \frac{m(1836.2m)}{m + 1836.2m} = 0.99946m$$

When this value is used instead of the electron mass m in the Bohr equations, the predicted values agree very well with the measured values.

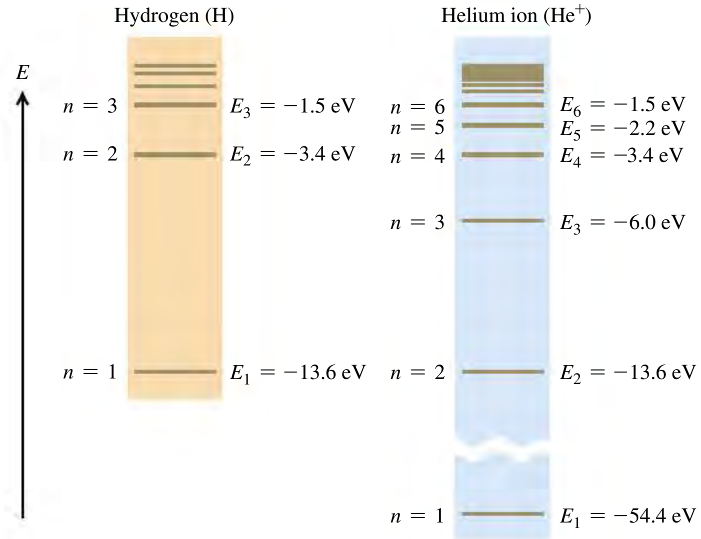
In an atom of deuterium, also called *heavy hydrogen,* the nucleus is not a single proton but a proton and a neutron bound together to form a composite body called the *deuteron*. The reduced mass of the deuterium atom turns out to be $0.99973m$. Equations (39.15) and (39.16) (with m replaced by m_r) show that all wavelengths are inversely proportional to m_r. Thus the wavelengths of the deuterium spectrum should be those of hydrogen divided by $(0.99973m)/(0.99946m) = 1.00027$. This is a small effect but well within the precision of modern spectrometers. This small wavelength shift led the American scientist Harold Urey to the discovery of deuterium in 1932, an achievement that earned him the 1934 Nobel Prize in chemistry.



39.26 The nucleus and the electron both orbit around their common center of mass. The distance r_N has been exaggerated for clarity; for ordinary hydrogen it actually equals $r_e/1836.2$.

Hydrogenlike Atoms

We can extend the Bohr model to other one-electron atoms, such as singly ionized helium (He^+), doubly ionized lithium (Li^{2+}), and so on. Such atoms are called *hydrogenlike* atoms. In such atoms, the nuclear charge is not e but Ze, where Z is the *atomic number,* equal to the number of protons in the nucleus. The effect in the previous analysis is to replace e^2 everywhere by Ze^2. You should verify that the orbital radii r_n given by Eq. (39.8) become smaller by a factor of Z, and the energy levels E_n given by Eq. (39.14) are multiplied by Z^2. The reduced-mass correction in these cases is even less than 0.1% because the nuclei are more massive than the single proton of ordinary hydrogen. **Figure 39.27** compares the energy levels for H and for He^+, which has $Z = 2$.



39.27 Energy levels of H and He^+. The energy expression, Eq. (39.14), is multiplied by $Z^2 = 4$ for He^+, so the energy of an He^+ ion with a given n is almost exactly four times that of an H atom with the same n. (There are small differences of the order of 0.05% because the reduced masses are slightly different.)

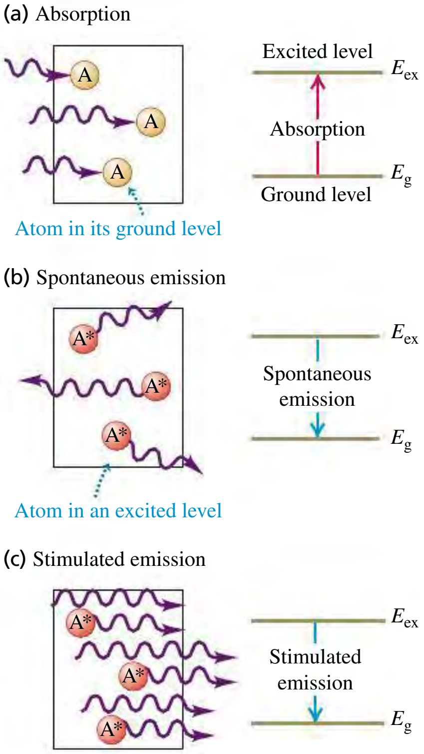
Atoms of the alkali metals (at the far left-hand side of the periodic table; see Appendix D) have one electron outside a core consisting of the nucleus and the inner electrons, with net core charge $+e$. These atoms are approximately hydrogenlike, especially in excited levels. Physicists have studied alkali atoms in which the outer electron has been excited into a very large orbit with $n = 1000$. From Eq. (39.8), the radius of such a *Rydberg atom* with $n = 1000$ is $n^2 = 10^6$ times the Bohr radius, or about 0.05 mm—about the size of a small grain of sand.

Although the Bohr model predicted the energy levels of the hydrogen atom correctly, it raised as many questions as it answered. It combined elements of classical physics with new postulates that were inconsistent with classical ideas. The model provided no insight into what happens during a transition from one orbit to another; the angular speeds of the electron motion were not in general the angular frequencies of the emitted radiation, a result that is contrary to classical electrodynamics. Attempts to extend the model to atoms with two or more electrons were not successful. An electron moving in one of Bohr's circular orbits forms a current loop and should produce a magnetic dipole moment (see Section 27.7). However, a hydrogen atom in its ground level has *no* magnetic moment due to orbital motion. In Chapters 40 and 41 we will find that an even more radical departure from classical concepts was needed before the understanding of atomic structure could progress further.

TEST YOUR UNDERSTANDING OF SECTION 39.3 Consider the possible transitions between energy levels in a He^+ ion. For which of these transitions in He^+ will the wavelength of the emitted photon be nearly the same as one of the wavelengths emitted by excited H atoms? (i) $n = 2$ to $n = 1$; (ii) $n = 3$ to $n = 2$; (iii) $n = 4$ to $n = 3$; (iv) $n = 4$ to $n = 2$; (v) more than one of these; (vi) none of these. ❙

PhET: Lasers

39.28 Three processes in which atoms interact with light.



39.4 THE LASER

The **laser** is a light source that produces a beam of highly coherent and very nearly monochromatic light as a result of cooperative emission from many atoms. The name "laser" is an acronym for "light amplification by stimulated emission of radiation." We can understand the principles of laser operation from what we have learned about atomic energy levels and photons. To do this we'll have to introduce two new concepts: *stimulated emission* and *population inversion.*

Spontaneous and Stimulated Emission

Consider a gas of atoms in a transparent container. Each atom is initially in its ground level of energy E_g and also has an excited level of energy E_{ex}. If we shine light of frequency f on the container, an atom can absorb one of the photons provided the photon energy $E = hf$ equals the energy difference $E_{ex} - E_g$ between the levels. **Figure 39.28a** shows this process, in which three atoms A each absorb a photon and go into the excited level. Some time later, the excited atoms (which we denote as A*) return to the ground level by each emitting a photon with the same frequency as the one originally absorbed (Fig. 39.28b). This process is called **spontaneous emission.** The direction and phase of each spontaneously emitted photon are random.

In **stimulated emission** (Fig. 39.28c), each incident photon encounters a previously excited atom. A kind of resonance effect induces each atom to emit a second photon with the same frequency, direction, phase, and polarization as the incident photon, which is not changed by the process. For each atom there is one photon before a stimulated emission and two photons after—thus the name *light amplification*. Because the two photons have the same phase, they emerge together as *coherent* radiation. The laser makes use of stimulated emission to produce a beam consisting of a large number of such coherent photons.

To discuss stimulated emission from atoms in excited levels, we need to know something about how many atoms are in each of the various energy levels. First, we need to make the distinction between the terms *energy level* and *state*. A system may have more than one way to attain a given energy level; each different way is a different **state.** For instance, there are two ways of putting an ideal unstretched spring in a given energy level. Remembering that the spring potential energy is $U = \frac{1}{2}kx^2$, we could compress the spring by $x = -b$ or we could stretch it by $x = +b$ to get the same $U = \frac{1}{2}kb^2$. The Bohr model had only one state in each energy level, but we will find in Chapter 41 that the hydrogen atom (Fig. 39.24b) actually has two *ground states* in its -13.60-eV ground level, eight *excited states* in its -3.40-eV first excited level, and so on.

The Maxwell–Boltzmann distribution function (see Section 18.5) determines the number of atoms in a given state in a gas. The function tells us that when the gas is in thermal equilibrium at absolute temperature T, the number n_i of atoms in a state with energy E_i equals $Ae^{-E_i/kT}$, where k is the Boltzmann constant and A is another constant determined by the total number of atoms in the gas. (In Section 18.5, E was the kinetic energy $\frac{1}{2}mv^2$ of a gas molecule; here we're talking about the internal energy of an atom.) Because of the negative exponent, fewer atoms are in higher-energy states. If E_g is a ground-state energy and E_{ex} is the energy of an excited state, then the ratio of numbers of atoms in the two states is

$$\frac{n_{ex}}{n_g} = \frac{Ae^{-E_{ex}/kT}}{Ae^{-E_g/kT}} = e^{-(E_{ex}-E_g)/kT} \tag{39.18}$$

For example, suppose $E_{ex} - E_g = 2.0\text{ eV} = 3.2 \times 10^{-19}\text{ J}$, the energy of a 620-nm visible-light photon. At $T = 3000\text{ K}$ (roughly the temperature of the filament in an incandescent light bulb or restaurant heat lamp),

$$\frac{E_{ex} - E_g}{kT} = \frac{3.2 \times 10^{-19}\text{ J}}{(1.38 \times 10^{-23}\text{ J/K})(3000\text{ K})} = 7.73$$

and

$$e^{-(E_{ex}-E_g)/kT} = e^{-7.73} = 0.00044$$

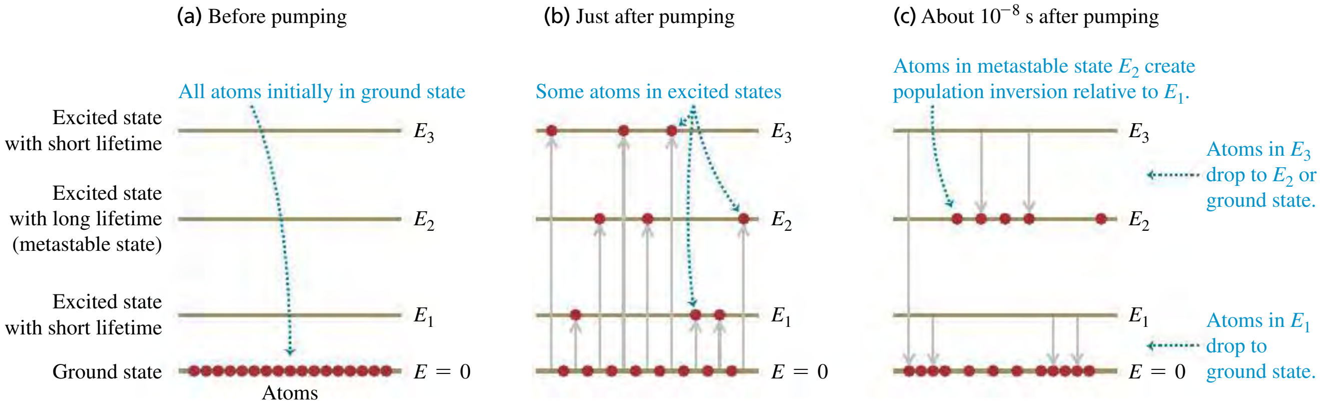
That is, the fraction of atoms in a state 2.0 eV above a ground state is extremely small, even at this high temperature. The point is that at any reasonable temperature there aren't enough atoms in excited states for any appreciable amount of stimulated emission from these states to occur. Rather, a photon emitted by one of the rare excited atoms will almost certainly be absorbed by an atom in the ground state rather than encountering another excited atom.

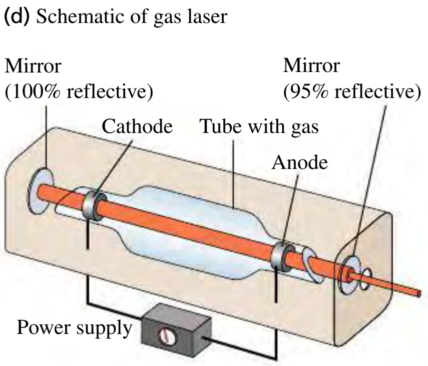
Enhancing Stimulated Emission: Population Inversions

To make a laser, we need to promote stimulated emission by increasing the number of atoms in excited states. Can we do that simply by illuminating the container with radiation of frequency $f = E/h$ corresponding to the energy difference $E = E_{ex} - E_g$, as in Fig. 39.28a? Some of the atoms absorb photons of energy E and are raised to the excited state, and the population ratio n_{ex}/n_g momentarily increases. But because n_g is originally so much larger than n_{ex}, an enormously intense beam of light would be required to momentarily increase n_{ex} to a value comparable to n_g. The rate at which energy is *absorbed* from the beam by the n_g ground-state atoms far exceeds the rate at which energy is added to the beam by stimulated emission from the relatively rare (n_{ex}) excited atoms.

We need to create a *nonequilibrium* situation in which there are more atoms in a higher-energy state than in a lower-energy state. Such a situation is called a **population inversion.** Then the rate of energy radiation by stimulated emission can *exceed* the rate of absorption, and the system will act as a net *source* of radiation with photon energy E. We can achieve a population inversion by

39.29 (a), (b), (c) Stages in the operation of a four-level laser. (d) The light emitted by atoms making spontaneous transitions from state E_2 to state E_1 is reflected between mirrors, so it continues to stimulate emission and gives rise to coherent light. One mirror is partially transmitting and allows the high-intensity light beam to escape.





starting with atoms that have the right kinds of excited states. **Figure 39.29a** shows an energy-level diagram for such an atom with a ground state and *three* excited states of energies E_1, E_2, and E_3. A laser that uses a material with energy levels like these is called a *four-level laser.* For the laser action to work, the states of energies E_1 and E_3 must have ordinary short lifetimes of about 10^{-8} s, while the state of energy E_2 must have an unusually long lifetime of 10^{-3} s or so. Such a long-lived **metastable state** can occur if, for instance, there are restrictions imposed by conservation of angular momentum that hinder photon emission from this state. (We'll discuss these restrictions in Chapter 41.) The metastable state is the one that we want to populate.

To produce a population inversion, we *pump* the material to excite the atoms out of the ground state into the states of energies E_1, E_2, and E_3 (Fig. 39.29b). If the atoms are in a gas, we can do this by inserting two electrodes into the gas container. When a burst of sufficiently high voltage is applied to the electrodes, an electric discharge occurs. Collisions between ionized atoms and electrons carrying the discharge current then excite the atoms to various energy states. Within about 10^{-8} s the atoms that are excited to states E_1 and E_3 undergo spontaneous photon emission, so these states end up depopulated. But atoms "pile up" in the metastable state with energy E_2. The number of atoms in the metastable state is *less* than the number in the ground state, but is *much greater* than in the nearly unoccupied state of energy E_1. Hence there is a population inversion of state E_2 relative to state E_1 (Fig. 39.29c). You can see why we need the two levels E_1 and E_3: Atoms that undergo spontaneous emission from the E_3 level help to populate the E_2 level, and the presence of the E_1 level makes a population inversion possible.

Over the next 10^{-3} s, some of the atoms in the long-lived metastable state E_2 transition to state E_1 by spontaneous emission. The emitted photons of energy $hf = E_2 - E_1$ are sent back and forth through the gas many times by a pair of parallel mirrors (Fig. 39.29d), so that they can *stimulate* emission from as many of the atoms in state E_2 as possible. The net result of all these processes is a beam of light of frequency f that can be quite intense, has parallel rays, is highly monochromatic, and is spatially *coherent* at all points within a given cross section—that is, a laser beam. One of the mirrors is partially transparent, so a portion of the beam emerges.

What we've described is a *pulsed* laser that produces a burst of coherent light every time the atoms are pumped. Pulsed lasers are used in LASIK eye surgery (an acronym for *laser-assisted in situ keratomileusis*) to reshape the cornea and correct for nearsightedness, farsightedness, or astigmatism. In a *continuous* laser, such as those found in the barcode scanners used at retail checkout counters, energy is supplied to the atoms continuously (for instance, by having the power supply in Fig. 39.29d provide a steady voltage to the electrodes) and a steady beam of light emerges from the laser. For such a laser the pumping must be intense enough to sustain the population inversion, so that the rate at which atoms are added to level E_2 through pumping equals the rate at which atoms in this level emit a photon and transition to level E_1.

Since a special arrangement of energy levels is needed for laser action, it's not surprising that only certain materials can be used to make a laser. Some types of laser use a solid, transparent material such as neodymium glass rather than a gas. The most common kind of laser—used in laser printers (Section 21.1), laser pointers, and to read the data on the disc in a DVD player or Blu-ray player—is a *semiconductor laser,* which doesn't use atomic energy levels at all. As we'll discuss in Chapter 42, these lasers instead use the energy levels of electrons that are free to roam throughout the volume of the semiconductors.

TEST YOUR UNDERSTANDING OF SECTION 39.4 An ordinary neon light fixture like those used in advertising signs emits red light of wavelength 632.8 nm. Neon atoms are also used in a helium–neon laser (a type of gas laser). The light emitted by a neon light fixture is (i) spontaneous emission; (ii) stimulated emission; (iii) both spontaneous and stimulated emission. ❙

39.5 CONTINUOUS SPECTRA

PhET: Blackbody Spectrum
PhET: The Greenhouse Effect

Emission line spectra come from matter in the gaseous state, in which the atoms are so far apart that interactions between them are negligible and each atom behaves as an isolated system. By contrast, a heated solid or liquid (in which atoms are close to each other) nearly always emits radiation with a *continuous* distribution of wavelengths rather than a line spectrum.

Here's an analogy that suggests why there is a difference. A tuning fork emits sound waves of a single definite frequency (a pure tone) when struck. But if you tightly pack a suitcase full of tuning forks and then shake the suitcase, the proximity of the tuning forks to each other affects the sound that they produce. What you hear is mostly noise, which is sound with a continuous distribution of all frequencies. In the same manner, isolated atoms in a gas emit light of certain distinct frequencies when excited, but if the same atoms are crowded together in a solid or liquid they produce a continuous spectrum of light.

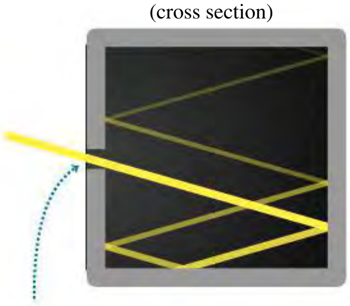
In this section we'll study an idealized case of continuous-spectrum radiation from a hot, dense object. Just as was the case for the emission line spectrum of light from an atom, we'll find that we can understand the continuous spectrum only if we use the ideas of energy levels and photons.

In the same way that an atom's emission spectrum has the same lines as its absorption spectrum, the ideal surface for *emitting* light with a continuous spectrum is one that also *absorbs* all wavelengths of electromagnetic radiation. Such an ideal surface is called a *blackbody* because it would appear perfectly black when illuminated; it would reflect no light at all. The continuous-spectrum radiation that a blackbody emits is called **blackbody radiation.** Like a perfectly frictionless incline or a massless rope, a perfect blackbody does not exist but is nonetheless a useful idealization.

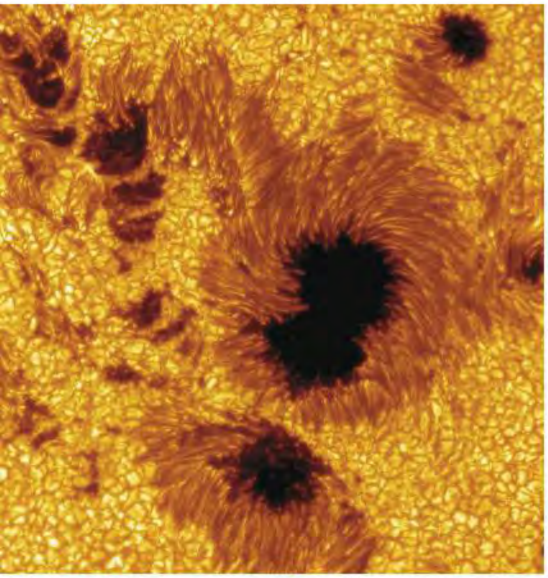
A good approximation to a blackbody is a hollow box with a small aperture in one wall (**Fig. 39.30**). Light that enters the aperture will eventually be absorbed by the walls of the box, so the box is a nearly perfect absorber. Conversely, when we heat the box, the light that emanates from the aperture is nearly ideal blackbody radiation with a continuous spectrum.

By 1900 blackbody radiation had been studied extensively, and three characteristics had been established. First, the total intensity I (the average rate of radiation of energy per unit surface area or average power per area) emitted from

39.30 A hollow box with a small aperture behaves like a blackbody. When the box is heated, the electromagnetic radiation that emerges from the aperture has a blackbody spectrum.



39.31 This close-up view of the sun's surface shows two dark sunspots. Their temperature is about 4000 K, while the surrounding solar material is at $T = 5800$ K. From the Stefan–Boltzmann law, the intensity from a given area of sunspot is only $(4000\text{ K}/5800\text{ K})^4 = 0.23$ as great as the intensity from the same area of the surrounding material—which is why sunspots appear dark.



the surface of an ideal radiator is proportional to the fourth power of the absolute temperature (**Fig. 39.31**). This is the **Stefan–Boltzmann law:**

Stefan–Boltzmann law for a blackbody:

Intensity of radiation from blackbody

$$I = \sigma T^4 \qquad (39.19)$$

Absolute temperature of blackbody

Stefan–Boltzmann constant

We encountered a version of this relationship in Section 17.7 during our study of heat transfer. In SI units, the value of the Stefan–Boltzmann constant σ is

$$\sigma = 5.670373(21) \times 10^{-8}\ \text{W/m}^2 \cdot \text{K}^4$$

Second, the intensity is not uniformly distributed over all wavelengths. Its distribution can be measured and described by the intensity per wavelength interval $I(\lambda)$, called the *spectral emittance*. Thus $I(\lambda)\,d\lambda$ is the intensity corresponding to wavelengths in the interval from λ to $\lambda + d\lambda$. The *total* intensity I, given by Eq. (39.19), is the *integral* of the distribution function $I(\lambda)$ over all wavelengths, which equals the area under the $I(\lambda)$-versus-λ curve:

$$I = \int_0^\infty I(\lambda)\,d\lambda \qquad (39.20)$$

CAUTION **Spectral emittance vs. intensity** Although we use the symbol $I(\lambda)$ for spectral emittance, keep in mind that spectral emittance is *not* the same thing as intensity I. Intensity is power per unit area, with units W/m^2. Spectral emittance is power per unit area *per unit wavelength interval*, with units W/m^3.

39.32 These graphs show the spectral emittance $I(\lambda)$ for radiation from a blackbody at three different temperatures.

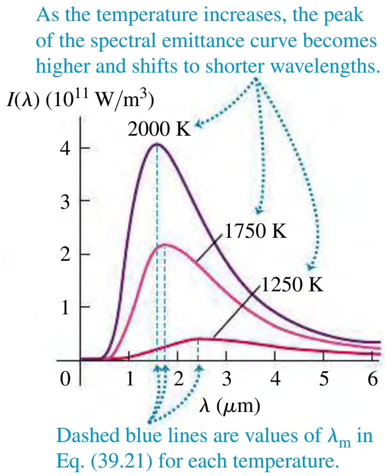


Figure 39.32 shows the measured spectral emittances $I(\lambda)$ for blackbody radiation at three different temperatures. Each has a peak wavelength λ_m at which the emitted intensity per wavelength interval is largest. Experiment shows that λ_m is inversely proportional to T, so their product is constant and equal to $2.90 \times 10^{-3}\ \text{m} \cdot \text{K}$. This observation is called the **Wien displacement law:**

Wien displacement law for a blackbody:

Peak wavelength in spectral emittance curve

$$\lambda_m T = 2.90 \times 10^{-3}\ \text{m} \cdot \text{K} \qquad (39.21)$$

Absolute temperature of blackbody

As the temperature rises, the peak of $I(\lambda)$ becomes higher and shifts to shorter wavelengths. Yellow light has shorter wavelengths than red light, so a body that glows yellow is hotter and brighter than one of the same size that glows red.

Third, experiments show that the *shape* of the distribution function is the same for all temperatures. We can make a curve for one temperature fit any other temperature by simply changing the scales on the graph.

Rayleigh and the "Ultraviolet Catastrophe"

During the last decade of the 19th century, many attempts were made to derive these empirical results about blackbody radiation from basic principles. In one attempt, the English physicist Lord Rayleigh considered the light enclosed within a rectangular box like that shown in Fig. 39.30. Such a box, he reasoned, has a series of possible *normal modes* for electromagnetic waves, as we discussed in Section 32.5. It also seemed reasonable to assume that the distribution of energy among the various modes would be given by the equipartition principle (see Section 18.4), which had been used successfully in the analysis of heat capacities.

Including both the electric- and magnetic-field energies, Rayleigh assumed that the total energy of each normal mode was equal to kT. Then by computing the *number* of normal modes corresponding to a wavelength interval $d\lambda$, Rayleigh calculated the expected distribution of wavelengths in the radiation within the box. Finally, he computed the predicted intensity distribution $I(\lambda)$ for the radiation emerging from the hole. His result was quite simple:

$$I(\lambda) = \frac{2\pi ckT}{\lambda^4} \qquad \text{(Rayleigh's calculation)} \qquad (39.22)$$

At large wavelengths this formula agrees quite well with the experimental results shown in Fig. 39.32, but there is serious disagreement at small wavelengths. The experimental curves in Fig. 39.32 fall toward zero at small λ. By contrast, Rayleigh's prediction in Eq. (39.22) goes in the opposite direction, approaching infinity as $1/\lambda^4$, a result that was called in Rayleigh's time the "ultraviolet catastrophe." Even worse, the integral of Eq. (39.22) over all λ is infinite, indicating an infinitely large *total* radiated intensity. Clearly, something is wrong.

Planck and the Quantum Hypothesis

Finally, in 1900, the German physicist Max Planck succeeded in deriving a function, now called the **Planck radiation law,** that agreed very well with experimental intensity distribution curves. In his derivation he made what seemed at the time to be a crazy assumption: that electromagnetic oscillators (electrons) in the walls of Rayleigh's box vibrating at a frequency f could have only certain values of energy equal to nhf, where $n = 0, 1, 2, 3, \ldots$ and h is the constant that now bears Planck's name. These oscillators were in equilibrium with the electromagnetic waves in the box, so they both emitted and absorbed light. His assumption gave quantized energy levels and said that the energy in each normal mode was also a multiple of hf. This was in sharp contrast to Rayleigh's point of view that each normal mode could have any amount of energy.

Planck was not comfortable with this quantum hypothesis; he regarded it as a calculational trick rather than a fundamental principle. In a letter to a friend, he called it "an act of desperation" into which he was forced because "a theoretical explanation had to be found at any cost, whatever the price." But five years later, Einstein identified the energy change hf between levels as the energy of a photon (see Section 38.1), and other evidence quickly mounted. By 1915 there was little doubt about the validity of the quantum concept and the existence of photons. By discussing atomic spectra *before* continuous spectra, we have departed from the historical order of things. The credit for inventing the concept of quantization of energy levels goes to Planck, even though he didn't believe it at first. He received the 1918 Nobel Prize in physics for his achievements.

Figure 39.33 shows energy-level diagrams for two of the oscillators that Planck envisioned in the walls of the rectangular box, one with a low frequency

BIO Application Blackbody Eyes

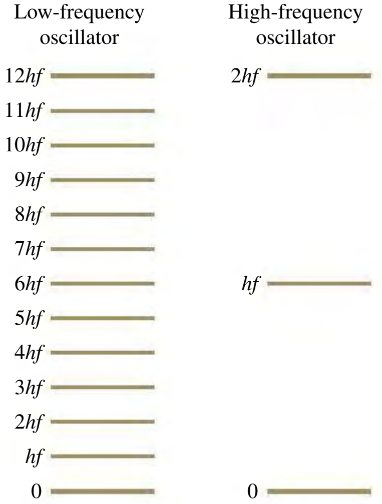
The interior of (a) a human eye, (b) a cat eye, or (c) a fish eye looks black, even though the tissue inside the eye is *not* black. That's because each eye acts like a blackbody, akin to the hollow box in Fig. 39.30: Light entering the eye is eventually absorbed after several reflections from the interior surfaces. Each eye also radiates like a blackbody, although the temperature is so low (around 300 K) that this radiation is principally at invisible infrared wavelengths.

(a)

(b)

(c)

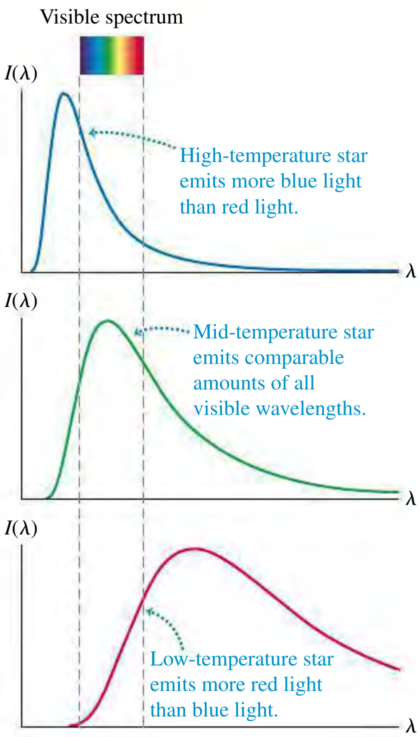
39.33 Energy levels for two of the oscillators that Planck envisioned in the walls of a blackbody like that shown in Fig. 39.30. The spacing between adjacent energy levels for each oscillator is hf, which is smaller for the low-frequency oscillator.



Application Star Colors and the Planck Radiation Law Stars (with radiation very similar to that of a blackbody) have a broad range of surface temperatures, from lower than 2500 K to higher than 30,000 K. The Wien displacement law and the shape of the Planck spectral emittance curve explain why these stars have different colors. From Eq. (39.21), a star with a high surface temperature of, say, 12,000 K has a short peak wavelength λ_m in the ultraviolet. Hence such a star emits more blue light than red light and appears blue to the eye. A star with a low surface temperature of, say, 3000 K has a long peak wavelength λ_m in the infrared, emits more red light than blue light, and appears red to the eye. For a star like the sun, which has a surface temperature of 5800 K, λ_m lies in the visible spectrum and the star appears white.

High-temperature stars appear blue.

Low-temperature stars appear red.



and the other with a high frequency. The spacing in energy between adjacent levels is hf. This spacing is small for the low-frequency oscillator that emits and absorbs photons of low frequency f and long wavelength $\lambda = c/f$. The energy spacing is greater for the high-frequency oscillator, which emits high-frequency photons of short wavelength.

According to Rayleigh's picture, both of these oscillators have the same amount of energy kT and are equally effective at emitting radiation. In Planck's model, however, the high-frequency oscillator is very ineffective as a source of light. To see why, we can use the ideas from Section 39.4 about the populations of various energy states. If we consider all the oscillators of a given frequency f in a box at temperature T, the number of oscillators that have energy nhf is $Ae^{-nhf/kT}$. The ratio of the number of oscillators in the first excited state ($n = 1$, energy hf) to the number of oscillators in the ground state ($n = 0$, energy zero) is

$$\frac{n_1}{n_0} = \frac{Ae^{-hf/kT}}{Ae^{-(0)/kT}} = e^{-hf/kT} \tag{39.23}$$

Let's evaluate Eq. (39.23) for $T = 2000$ K, one of the temperatures shown in Fig. 39.32. At this temperature $kT = 2.76 \times 10^{-20}$ J $= 0.172$ eV. For an oscillator that emits photons of wavelength $\lambda = 3.00\ \mu$m, $hf = hc/\lambda = 0.413$ eV; for a higher-frequency oscillator that emits photons of wavelength $\lambda = 0.500\ \mu$m, $hf = hc/\lambda = 2.48$ eV. For these two cases Eq. (39.23) gives

$$\frac{n_1}{n_0} = e^{-hf/kT} = 0.0909 \text{ for } \lambda = 3.00\ \mu\text{m}$$

$$\frac{n_1}{n_0} = e^{-hf/kT} = 5.64 \times 10^{-7} \text{ for } \lambda = 0.500\ \mu\text{m}$$

The value for $\lambda = 3.00\ \mu$m means that of all the oscillators that can emit light at this wavelength, 0.0909 of them—about one in 11—are in the first excited state. These excited oscillators can each emit a 3.00-μm photon and contribute it to the radiation inside the box. Hence we would expect that this radiation would be rather plentiful in the spectrum of radiation from a 2000 K blackbody. By contrast, the value for $\lambda = 0.500\ \mu$m means that only 5.64×10^{-7} (about one in two million) of the oscillators that can emit this wavelength are in the first excited state. An oscillator can't emit if it's in the ground state, so the amount of radiation in the box at this wavelength is *tremendously* suppressed compared to Rayleigh's prediction. That's why the spectral emittance curve for 2000 K in Fig. 39.32 has such a low value at $\lambda = 0.500\ \mu$m and shorter wavelengths. So Planck's quantum hypothesis provided a natural way to suppress the spectral emittance of a blackbody at short wavelengths, and hence averted the ultraviolet catastrophe that plagued Rayleigh's calculations.

We won't go into all the details of Planck's derivation of the spectral emittance. Here is his result:

Planck radiation law:

Spectral emittance of blackbody; Planck's constant; Speed of light in vacuum

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5\left(e^{hc/\lambda kT} - 1\right)} \tag{39.24}$$

Absolute temperature of blackbody; Wavelength; Boltzmann constant

This function turns out to agree well with experimental emittance curves such as those in Fig. 39.32.

The Planck radiation law also contains the Wien displacement law and the Stefan–Boltzmann law as consequences. To derive the Wien law, we find the value of λ at which $I(\lambda)$ is maximum by taking the derivative of Eq. (39.24) and setting it equal to zero. We leave it to you to fill in the details; the result is

$$\lambda_m = \frac{hc}{4.965kT} \qquad (39.25)$$

To obtain this result, you have to solve the equation

$$5 - x = 5e^{-x} \qquad (39.26)$$

The root of this equation, found by trial and error or more sophisticated means, is 4.965 to four significant figures. You should evaluate the constant $hc/4.965k$ and show that it agrees with the experimental value of $2.90 \times 10^{-3}\ \mathrm{m \cdot K}$ given in Eq. (39.21).

We can obtain the Stefan–Boltzmann law for a blackbody by integrating Eq. (39.24) over all λ to find the *total* radiated intensity (see Problem 39.61). This is not a simple integral; the result is

$$I = \int_0^\infty I(\lambda)\, d\lambda = \frac{2\pi^5 k^4}{15c^2h^3} T^4 = \sigma T^4 \qquad (39.27)$$

in agreement with Eq. (39.19). Our result in Eq. (39.27) also shows that the constant σ in that law can be expressed in terms of other fundamental constants:

$$\sigma = \frac{2\pi^5 k^4}{15c^2h^3} \qquad (39.28)$$

Substitute the values of k, c, and h from Appendix F and verify that you obtain the value $\sigma = 5.6704 \times 10^{-8}\ \mathrm{W/m^2 \cdot K^4}$ for the Stefan–Boltzmann constant.

The Planck radiation law, Eq. (39.24), looks so different from the unsuccessful Rayleigh expression, Eq. (39.22), that it may seem unlikely that they would agree for any value of λ. But when λ is large, the exponent in the denominator of Eq. (39.24) is very small. We can then use the approximation $e^x \approx 1 + x$ (for x much less than 1). You should verify that when this is done, the result approaches Eq. (39.22), showing that the two expressions do agree in the limit of very large λ. We also note that the Rayleigh expression does not contain h. At very long wavelengths (very small photon energies), quantum effects become unimportant.

EXAMPLE 39.7 LIGHT FROM THE SUN

SOLUTION

To a good approximation, the sun's surface is a blackbody with a surface temperature of 5800 K. (We are ignoring the absorption produced by the sun's atmosphere, shown in Fig. 39.9.) (a) At what wavelength does the sun emit most strongly? (b) What is the total radiated power per unit surface area?

SOLUTION

IDENTIFY and SET UP: Our target variables are the peak-intensity wavelength λ_m and the radiated power per area I. Hence we'll use the Wien displacement law, Eq. (39.21) (which relates λ_m to the blackbody temperature T), and the Stefan–Boltzmann law, Eq. (39.19) (which relates I to T).

EXECUTE: (a) From Eq. (39.21),

$$\lambda_m = \frac{2.90 \times 10^{-3}\ \mathrm{m \cdot K}}{T} = \frac{2.90 \times 10^{-3}\ \mathrm{m \cdot K}}{5800\ \mathrm{K}}$$
$$= 0.500 \times 10^{-6}\ \mathrm{m} = 500\ \mathrm{nm}$$

(b) From Eq. (39.19),

$$I = \sigma T^4 = (5.67 \times 10^{-8}\ \mathrm{W/m^2 \cdot K^4})(5800\ \mathrm{K})^4$$
$$= 6.42 \times 10^7\ \mathrm{W/m^2} = 64.2\ \mathrm{MW/m^2}$$

EVALUATE: The 500-nm wavelength found in part (a) is near the middle of the visible spectrum. This should not be a surprise: The human eye evolved to take maximum advantage of natural light.

The enormous value $I = 64.2\ \mathrm{MW/m^2}$ that we obtained in part (b) is the intensity at the *surface* of the sun, which is a sphere of radius 6.96×10^8 m. When this radiated energy reaches the earth, 1.50×10^{11} m away, the intensity has decreased by the factor $[(6.96 \times 10^8\ \mathrm{m})/(1.50 \times 10^{11}\ \mathrm{m})]^2 = 2.15 \times 10^{-5}$ to the still-impressive $1.4\ \mathrm{kW/m^2}$.

EXAMPLE 39.8 A SLICE OF SUNLIGHT

SOLUTION

Find the power per unit area radiated from the sun's surface in the wavelength range 600.0 to 605.0 nm.

SOLUTION

IDENTIFY and SET UP: This question concerns the power emitted by a blackbody over a narrow range of wavelengths, and so involves the spectral emittance $I(\lambda)$ given by the Planck radiation law, Eq. (39.24). This requires that we find the area under the $I(\lambda)$ curve between 600.0 and 605.0 nm. We'll *approximate* this area as the product of the height of the curve at the median wavelength $\lambda = 602.5$ nm and the width of the interval, $\Delta\lambda = 5.0$ nm. From Example 39.7, $T = 5800$ K.

EXECUTE: To obtain the height of the $I(\lambda)$ curve at $\lambda = 602.5\text{ nm} = 6.025 \times 10^{-7}$ m, we first evaluate the quantity $hc/\lambda kT$ in Eq. (39.24) and then substitute the result into Eq. (39.24):

$$\frac{hc}{\lambda kT} = \frac{(6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^{8}\ \text{m/s})}{(6.025 \times 10^{-7}\ \text{m})(1.381 \times 10^{-23}\ \text{J/K})(5800\ \text{K})} = 4.116$$

$$I(\lambda) = \frac{2\pi(6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^{8}\ \text{m/s})^2}{(6.025 \times 10^{-7}\ \text{m})^5(e^{4.116} - 1)}$$

$$= 7.81 \times 10^{13}\ \text{W/m}^3$$

The intensity in the 5.0-nm range from 600.0 to 605.0 nm is then approximately

$$I(\lambda)\Delta\lambda = (7.81 \times 10^{13}\ \text{W/m}^3)(5.0 \times 10^{-9}\ \text{m})$$

$$= 3.9 \times 10^{5}\ \text{W/m}^2 = 0.39\ \text{MW/m}^2$$

EVALUATE: In part (b) of Example 39.7, we found the power radiated per unit area by the sun at *all* wavelengths to be $I = 64.2\ \text{MW/m}^2$; here we have found that the power radiated per unit area in the wavelength range from 600 to 605 nm is $I(\lambda)\Delta\lambda = 0.39\ \text{MW/m}^2$, about 0.6% of the total.

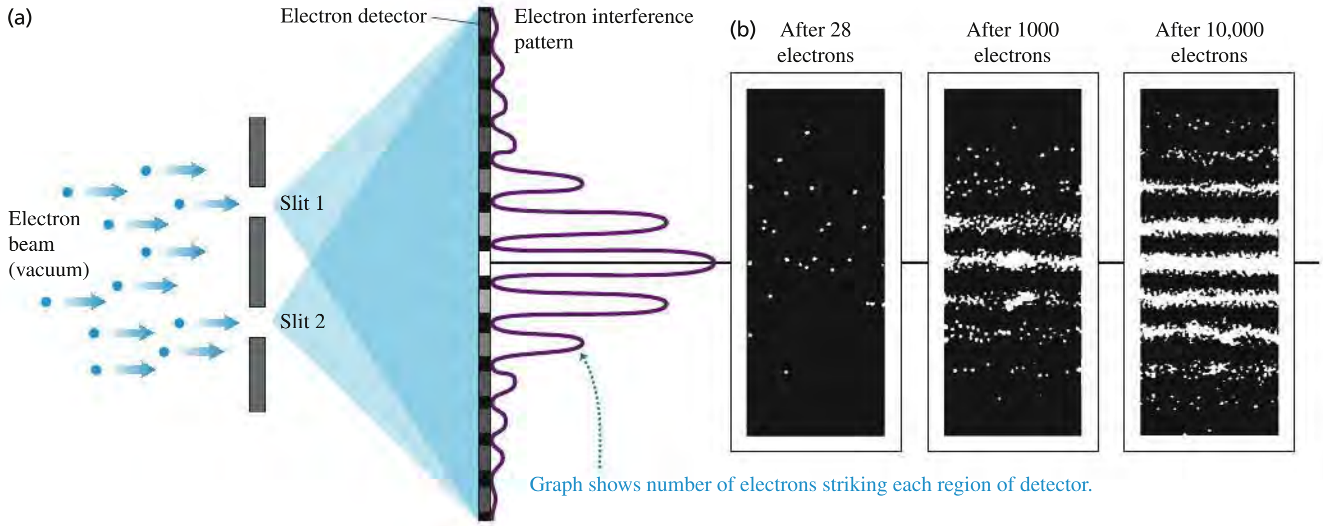
TEST YOUR UNDERSTANDING OF SECTION 39.5 (a) Does a blackbody at 2000 K emit x rays? (b) Does it emit radio waves?

39.6 THE UNCERTAINTY PRINCIPLE REVISITED

The discovery of the dual wave–particle nature of matter forces us to reevaluate the kinematic language we use to describe the position and motion of a particle. In classical Newtonian mechanics we think of a particle as a point. We can describe its location and state of motion at any instant with three spatial coordinates and three components of velocity. But because matter also has a wave aspect, when we look at the behavior on a small enough scale—comparable to the de Broglie wavelength of the particle—we can no longer use the Newtonian description. Certainly no Newtonian particle would undergo diffraction like electrons do (Section 39.1).

To demonstrate just how non-Newtonian the behavior of matter can be, let's look at an experiment involving the two-slit interference of electrons (**Fig. 39.34**).

39.34 (a) A two-slit interference experiment for electrons. (b) The interference pattern after 28, 1000, and 10,000 electrons.



We aim an electron beam at two parallel slits, as we did for light in Section 38.4. (The electron experiment has to be done in vacuum so that the electrons don't collide with air molecules.) What kind of pattern appears on the detector on the other side of the slits? The answer is: *exactly the same* kind of interference pattern we saw for photons in Section 38.4! Moreover, the principle of complementarity, which we introduced in Section 38.4, tells us that we cannot apply the wave and particle models simultaneously to describe any single element of this experiment. Thus we *cannot* predict exactly where in the pattern (a wave phenomenon) any individual electron (a particle) will land. We can't even ask which slit an individual electron passes through. If we tried to look at where the electrons were going by shining a light on them—that is, by scattering photons off them—the electrons would recoil, which would modify their motions so that the two-slit interference pattern would not appear.

CAUTION **Electron two-slit interference is not interference between two electrons** It's a common misconception that the pattern in Fig. 39.34b is due to the interference between *two* electron waves, each representing an electron passing through one slit. To show that this cannot be the case, we can send just one electron at a time through the apparatus. It makes no difference; we end up with the same interference pattern. In a sense, each electron wave interferes with itself.

The Heisenberg Uncertainty Principles for Matter

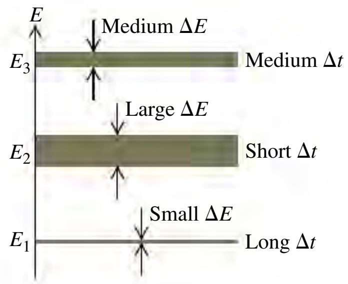
Just as electrons and photons show the same behavior in a two-slit interference experiment, electrons and other forms of matter obey the same Heisenberg uncertainty principles as photons do:

$$\begin{aligned}\Delta x \Delta p_x &\geq \hbar/2 \\ \Delta y \Delta p_y &\geq \hbar/2 \\ \Delta z \Delta p_z &\geq \hbar/2\end{aligned} \quad \text{(Heisenberg uncertainty principle for position and momentum)} \qquad (39.29)$$

$$\Delta t\, \Delta E \geq \hbar/2 \quad \text{(Heisenberg uncertainty principle for energy and time interval)} \qquad (39.30)$$

In these equations $\hbar = h/2\pi = 1.055 \times 10^{-34}\ \text{J}\cdot\text{s}$. The uncertainty principle for energy and time interval has a direct application to energy levels. We have assumed that each energy level in an atom has a very definite energy. However, Eq. (39.30) says that this is not true for all energy levels. A system that remains in a metastable state for a very long time (large Δt) can have a very well-defined energy (small ΔE), but if it remains in a state for only a short time (small Δt) the uncertainty in energy must be correspondingly greater (large ΔE). **Figure 39.35** illustrates this idea.

39.35 The longer the lifetime Δt of a state, the smaller is its spread in energy (shown by the width of the energy levels).



EXAMPLE 39.9 THE UNCERTAINTY PRINCIPLE: POSITION AND MOMENTUM

An electron is confined within a region of width 5.000×10^{-11} m (roughly the Bohr radius). (a) Estimate the minimum uncertainty in the x-component of the electron's momentum. (b) What is the kinetic energy of an electron with this magnitude of momentum? Express your answer in both joules and electron volts.

SOLUTION

IDENTIFY and SET UP: This problem uses the Heisenberg uncertainty principle for position and momentum and the relationship between a particle's momentum and its kinetic energy. The electron could be anywhere within the region, so we take $\Delta x = 5.000 \times 10^{-11}$ m as its position uncertainty. We then find the momentum uncertainty Δp_x from Eq. (39.29) and the kinetic energy from the relationships $p = mv$ and $K = \frac{1}{2}mv^2$.

EXECUTE: (a) From Eqs. (39.29), for a given value of Δx, the uncertainty in momentum is minimum when the product $\Delta x\, \Delta p_x$ equals $\hbar/2$. Hence

$$\Delta p_x = \frac{\hbar}{2\Delta x} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(5.000 \times 10^{-11}\ \text{m})} = 1.055 \times 10^{-24}\ \text{J}\cdot\text{s/m}$$
$$= 1.055 \times 10^{-24}\ \text{kg}\cdot\text{m/s}$$

(b) We can rewrite the nonrelativistic expression for kinetic energy as

$$K = \tfrac{1}{2}mv^2 = \frac{(mv)^2}{2m} = \frac{p^2}{2m}$$

Hence an electron with a magnitude of momentum equal to Δp_x from part (a) has kinetic energy

$$K = \frac{p^2}{2m} = \frac{(1.055 \times 10^{-24}\ \text{kg}\cdot\text{m/s})^2}{2(9.11 \times 10^{-31}\ \text{kg})}$$
$$= 6.11 \times 10^{-19}\ \text{J} = 3.81\ \text{eV}$$

EVALUATE: This energy is typical of electron energies in atoms. This agreement suggests that the uncertainty principle is deeply involved in atomic structure.

A similar calculation explains why electrons in atoms do not fall into the nucleus. If an electron were confined to the interior of a nucleus, its position uncertainty would be $\Delta x \approx 10^{-14}$ m. This would give the electron a momentum uncertainty about 5000 times greater than that of the electron in this example, and a kinetic energy so great that the electron would immediately be ejected from the nucleus.

SOLUTION

EXAMPLE 39.10 THE UNCERTAINTY PRINCIPLE: ENERGY AND TIME

A sodium atom in one of the states labeled "Lowest excited levels" in Fig. 39.19a remains in that state, on average, for 1.6×10^{-8} s before it makes a transition to the ground state, emitting a photon with wavelength 589.0 nm and energy 2.105 eV. What is the uncertainty in energy of that excited state? What is the wavelength spread of the corresponding spectral line?

SOLUTION

IDENTIFY and SET UP: We use the Heisenberg uncertainty principle for energy and time interval and the relationship between photon energy and wavelength. The average time that the atom spends in this excited state is equal to Δt in Eq. (39.30). We find the minimum uncertainty in the energy of the excited state by replacing the $\geq$ sign in Eq. (39.30) with an equals sign and solving for ΔE.

EXECUTE: From Eq. (39.30),

$$\Delta E = \frac{\hbar}{2\Delta t} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(1.6 \times 10^{-8}\ \text{s})}$$
$$= 3.3 \times 10^{-27}\ \text{J} = 2.1 \times 10^{-8}\ \text{eV}$$

The atom remains in the ground state indefinitely, so that state has *no* associated energy uncertainty. The fractional uncertainty of the *photon* energy is therefore

$$\frac{\Delta E}{E} = \frac{2.1 \times 10^{-8}\ \text{eV}}{2.105\ \text{eV}} = 1.0 \times 10^{-8}$$

You can use some simple calculus and the relationship $E = hc/\lambda$ to show that $\Delta\lambda/\lambda \approx \Delta E/E$, so that the corresponding spread in wavelength, or "width," of the spectral line is approximately

$$\Delta\lambda = \lambda\frac{\Delta E}{E} = (589.0\ \text{nm})(1.0 \times 10^{-8}) = 0.0000059\ \text{nm}$$

EVALUATE: This irreducible uncertainty $\Delta\lambda$ is called the *natural line width* of this particular spectral line. Though very small, it is within the limits of resolution of present-day spectrometers. Ordinarily, the natural line width is much smaller than the line width arising from other causes such as the Doppler effect and collisions among the rapidly moving atoms.

The Uncertainty Principle and the Limits of the Bohr Model

We saw in Section 39.3 that the Bohr model of the hydrogen atom was tremendously successful. However, the Heisenberg uncertainty principle for position and momentum shows that this model *cannot* be a correct description of how an electron in an atom behaves. Figure 39.22 shows that in the Bohr model as interpreted by de Broglie, an electron wave moves in a plane around the nucleus. Let's call this the xy-plane, so the z-axis is perpendicular to the plane. Hence the Bohr model says that an electron is always found at $z = 0$, and its z-momentum p_z is always zero (the electron does not move out of the xy-plane). But this implies that there are *no* uncertainties in either z or p_z, so $\Delta z = 0$ and $\Delta p_z = 0$. This directly contradicts Eq. (39.29), which says that the product $\Delta z \Delta p_z$ must be greater than or equal to $\hbar/2$.

This conclusion isn't too surprising, since the electron in the Bohr model is a mix of particle and wave ideas (the electron moves in an orbit like a miniature planet, but has a wavelength). To get an accurate picture of how electrons behave inside an atom and elsewhere, we need a description that is based *entirely* on the electron's wave properties. Our goal in Chapter 40 will be to develop this description, which we call *quantum mechanics*. To do this we'll introduce the *Schrödinger equation,* the fundamental equation that describes the dynamics of matter waves. This equation, as we will see, is as fundamental to quantum mechanics as Newton's laws are to classical mechanics or as Maxwell's equations are to electromagnetism.

TEST YOUR UNDERSTANDING OF SECTION 39.6 Rank the following situations according to the uncertainty in x-momentum, from largest to smallest. The mass of the proton is 1836 times the mass of the electron. (i) An electron whose x-coordinate is known to within 2×10^{-15} m; (ii) an electron whose x-coordinate is known to within 4×10^{-15} m; (iii) a proton whose x-coordinate is known to within 2×10^{-15} m; (iv) a proton whose x-coordinate is known to within 4×10^{-15} m. ❙

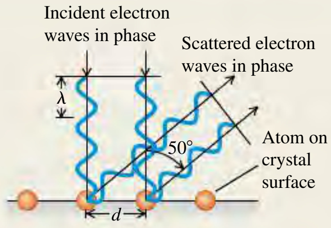
SOLUTIONS TO ALL EXAMPLES

De Broglie waves and electron diffraction: Electrons and other particles have wave properties. A particle's wavelength depends on its momentum in the same way as for photons. A nonrelativistic electron accelerated from rest through a potential difference V_{ba} has a wavelength given by Eq. (39.3). Electron microscopes use the very small wavelengths of fast-moving electrons to make images with resolution thousands of times finer than is possible with visible light. (See Examples 39.1–39.3.)

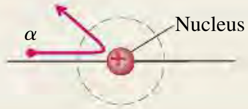
$$\lambda = \frac{h}{p} = \frac{h}{mv} \quad (39.1)$$

$$E = hf \quad (39.2)$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2meV_{ba}}} \quad (39.3)$$

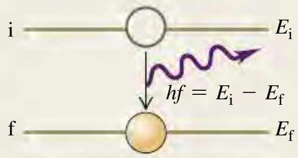


The nuclear atom: The Rutherford scattering experiments show that most of an atom's mass and all of its positive charge are concentrated in a tiny, dense nucleus at the center of the atom. (See Example 39.4.)



Atomic line spectra and energy levels: The energies of atoms are quantized: They can have only certain definite values, called energy levels. When an atom makes a transition from an energy level E_i to a lower level E_f, it emits a photon of energy $E_i - E_f$. The same photon can be absorbed by an atom in the lower energy level, which excites the atom to the upper level. (See Example 39.5.)

$$hf = \frac{hc}{\lambda} = E_i - E_f \quad (39.5)$$



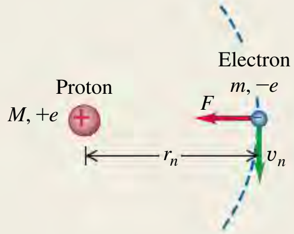
The Bohr model: In the Bohr model of the hydrogen atom, the permitted values of angular momentum are integral multiples of $h/2\pi$. The integer multiplier n is called the principal quantum number for the level. The orbital radii are proportional to n^2. The energy levels of the hydrogen atom are given by Eq. (39.15), where R is the Rydberg constant. (See Example 39.6.)

$$L_n = mv_n r_n = n\frac{h}{2\pi} \quad (39.6)$$
$$(n = 1, 2, 3, \ldots)$$

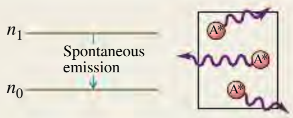
$$r_n = \epsilon_0 \frac{n^2 h^2}{\pi m e^2} = n^2 a_0 \quad (39.8), (39.11)$$

$$v_n = \frac{1}{\epsilon_0}\frac{e^2}{2nh} \quad (39.9)$$

$$E_n = -\frac{hcR}{n^2} = -\frac{13.60 \text{ eV}}{n^2} \quad (39.15)$$
$$(n = 1, 2, 3, \ldots)$$



The laser: The laser operates on the principle of stimulated emission, by which many photons with identical wavelength and phase are emitted. Laser operation requires a nonequilibrium condition called a population inversion, in which more atoms are in a higher-energy state than are in a lower-energy state.

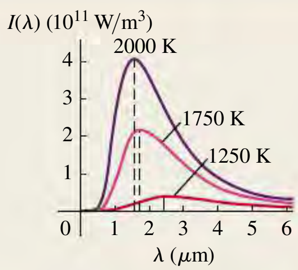


Blackbody radiation: The total radiated intensity (average power radiated per area) from a blackbody surface is proportional to the fourth power of the absolute temperature T. The quantity $\sigma = 5.67 \times 10^{-8}\ \text{W/m}^2 \cdot \text{K}^4$ is called the Stefan–Boltzmann constant. The wavelength λ_m at which a blackbody radiates most strongly is inversely proportional to T. The Planck radiation law gives the spectral emittance $I(\lambda)$ (intensity per wavelength interval in blackbody radiation). (See Examples 39.7 and 39.8.)

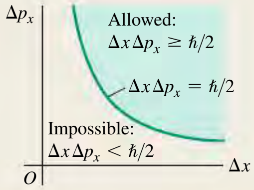
$$I = \sigma T^4$$
(Stefan–Boltzmann law) (39.19)

$$\lambda_m T = 2.90 \times 10^{-3}\ \text{m} \cdot \text{K}$$
(Wien displacement law) (39.21)

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5 (e^{hc/\lambda kT} - 1)}$$
(Planck radiation law) (39.24)



The Heisenberg uncertainty principle for particles: The same uncertainty considerations that apply to photons also apply to particles such as electrons. The uncertainty ΔE in the energy of a state that is occupied for a time Δt is given by Eq. (39.30), $\Delta t\, \Delta E \geq \hbar/2$. (See Examples 39.9 and 39.10.)



BRIDGING PROBLEM HOT STARS AND HYDROGEN CLOUDS

SOLUTION

Figure 39.36 shows a cloud, or *nebula,* of glowing hydrogen in interstellar space. The atoms in this cloud are excited by short-wavelength radiation emitted by the bright blue stars at the center of the nebula.

(a) The blue stars act as blackbodies and emit light with a continuous spectrum. What is the wavelength at which a star with a surface temperature of 15,100 K (about $2\frac{1}{2}$ times the surface temperature of the sun) has the maximum spectral emittance? In what region of the electromagnetic spectrum is this?

(b) Figure 39.32 shows that most of the energy radiated by a blackbody is at wavelengths between about one half and three times the wavelength of maximum emittance. If a hydrogen atom near the star in part (a) is initially in the ground level, what is the principal quantum number of the highest energy level to which it could be excited by a photon in this wavelength range?

(c) The red color of the nebula is primarily due to hydrogen atoms making a transition from $n = 3$ to $n = 2$ and emitting photons of wavelength 656.3 nm. In the Bohr model as interpreted by de Broglie, what are the *electron* wavelengths in the $n = 2$ and $n = 3$ levels?

39.36 The Rosette Nebula.

SOLUTION GUIDE

IDENTIFY and SET UP

1. To solve this problem you need to use your knowledge of both blackbody radiation (Section 39.5) and the Bohr model of the hydrogen atom (Section 39.3).
2. In part (a) the target variable is the wavelength at which the star emits most strongly; in part (b) the target variable is a principal quantum number, and in part (c) it is the de Broglie wavelength of an electron in the $n = 2$ and $n = 3$ Bohr orbits (see Fig. 39.24). Select the equations you will need to find the target variables. (*Hint:* In Section 39.5 you learned how to find the energy change involved in a transition between two given levels of a hydrogen atom. Part (b) is a variation on this: You are to find the final level in a transition that starts in the $n = 1$ level and involves the absorption of a photon of a given wavelength and hence a given energy.)

EXECUTE

3. Use the Wien displacement law to find the wavelength at which the star has maximum spectral emittance. In what part of the electromagnetic spectrum is this wavelength?
4. Use your result from step 3 to find the range of wavelengths in which the star radiates most of its energy. Which end of this range corresponds to a photon with the greatest energy?
5. Write an expression for the wavelength of a photon that must be absorbed to cause an electron transition from the ground level $(n = 1)$ to a higher level n. Solve for the value of n that corresponds to the highest-energy photon in the range you calculated in step 4. (*Hint:* Remember that n must be an integer.)
6. Find the electron wavelengths that correspond to the $n = 2$ and $n = 3$ orbits shown in Fig. 39.22.

EVALUATE

7. Check your result in step 5 by calculating the wavelength needed to excite a hydrogen atom from the ground level into the level *above* the highest-energy level that you found in step 5. Is it possible for light in the range of wavelengths you found in step 4 to excite hydrogen atoms from the ground level into this level?
8. How do the electron wavelengths you found in step 6 compare to the wavelength of a *photon* emitted in a transition from the $n = 3$ level to the $n = 2$ level?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q39.1 If a proton and an electron have the same speed, which has the longer de Broglie wavelength? Explain.

Q39.2 If a proton and an electron have the same kinetic energy, which has the longer de Broglie wavelength? Explain.

Q39.3 Does a photon have a de Broglie wavelength? If so, how is it related to the wavelength of the associated electromagnetic wave? Explain.

Q39.4 When an electron beam goes through a very small hole, it produces a diffraction pattern on a screen, just like that of light. Does this mean that an electron spreads out as it goes through the hole? What does this pattern mean?

Q39.5 Galaxies tend to be strong emitters of Lyman-α photons (from the $n = 2$ to $n = 1$ transition in atomic hydrogen). But the intergalactic medium—the very thin gas between the galaxies—tends to *absorb* Lyman-α photons. What can you infer from these observations about the temperature in these two environments? Explain.

Q39.6 A doubly ionized lithium atom (Li^{++}) is one that has had two of its three electrons removed. The energy levels of the remaining single-electron ion are closely related to those of the hydrogen atom. The nuclear charge for lithium is $+3e$ instead of just $+e$. How are the energy levels related to those of hydrogen? How is the *radius* of the ion in the ground level related to that of the hydrogen atom? Explain.

Q39.7 The emission of a photon by an isolated atom is a recoil process in which momentum is conserved. Thus Eq. (39.5) should include a recoil kinetic energy K_r for the atom. Why is this energy negligible in that equation?

Q39.8 How might the energy levels of an atom be measured directly—that is, without recourse to analysis of spectra?

Q39.9 Elements in the gaseous state emit line spectra with well-defined wavelengths. But hot solid bodies always emit a continuous spectrum—that is, a continuous smear of wavelengths. Can you account for this difference?

Q39.10 As a body is heated to a very high temperature and becomes self-luminous, the apparent color of the emitted radiation shifts from red to yellow and finally to blue as the temperature increases. Why does the color shift? What other changes in the character of the radiation occur?

Q39.11 Do the planets of the solar system obey a distance law ($r_n = n^2 r_1$) as the electrons of the Bohr atom do? Should they? Why (or why not)? (Consult Appendix F for the appropriate distances.)

Q39.12 You have been asked to design a magnet system to steer a beam of 54-eV electrons like those described in Example 39.1 (Section 39.1). The goal is to be able to direct the electron beam to a specific target location with an accuracy of ± 1.0 mm. In your design, do you need to take the wave nature of electrons into account? Explain.

Q39.13 Why go through the expense of building an electron microscope for studying very small objects such as organic molecules? Why not just use extremely short electromagnetic waves, which are much cheaper to generate?

Q39.14 Which has more total energy: a hydrogen atom with an electron in a high shell (large n) or in a low shell (small n)? Which is moving faster: the high-shell electron or the low-shell electron? Is there a contradiction here? Explain.

Q39.15 Does the uncertainty principle have anything to do with marksmanship? That is, is the accuracy with which a bullet can be aimed at a target limited by the uncertainty principle? Explain.

Q39.16 Suppose a two-slit interference experiment is carried out using an electron beam. Would the same interference pattern result if one slit at a time is uncovered instead of both at once? If not, why not? Doesn't each electron go through one slit or the other? Or does every electron go through both slits? Discuss the latter possibility in light of the principle of complementarity.

Q39.17 Equation (39.30) states that the energy of a system can have uncertainty. Does this mean that the principle of conservation of energy is no longer valid? Explain.

Q39.18 Laser light results from transitions from long-lived metastable states. Why is it more monochromatic than ordinary light?

Q39.19 Could an electron-diffraction experiment be carried out using three or four slits? Using a grating with many slits? What sort of results would you expect with a grating? Would the uncertainty principle be violated? Explain.

Q39.20 As the lower half of Fig. 39.4 shows, the diffraction pattern made by electrons that pass through aluminum foil is a series of concentric rings. But if the aluminum foil is replaced by a single crystal of aluminum, only certain points on these rings appear in the pattern. Explain.

Q39.21 Why can an electron microscope have greater magnification than an ordinary microscope?

Q39.22 When you check the air pressure in a tire, a little air always escapes; the process of making the measurement changes the quantity being measured. Think of other examples of measurements that change or disturb the quantity being measured.

EXERCISES

Section 39.1 Electron Waves

39.1 • (a) An electron moves with a speed of 4.70×10^6 m/s. What is its de Broglie wavelength? (b) A proton moves with the same speed. Determine its de Broglie wavelength.

39.2 •• For crystal diffraction experiments (discussed in Section 39.1), wavelengths on the order of 0.20 nm are often appropriate. Find the energy in electron volts for a particle with this wavelength if the particle is (a) a photon; (b) an electron; (c) an alpha particle ($m = 6.64 \times 10^{-27}$ kg).

39.3 • An electron has a de Broglie wavelength of 2.80×10^{-10} m. Determine (a) the magnitude of its momentum and (b) its kinetic energy (in joules and in electron volts).

39.4 •• **Wavelength of an Alpha Particle.** An alpha particle ($m = 6.64 \times 10^{-27}$ kg) emitted in the radioactive decay of uranium-238 has an energy of 4.20 MeV. What is its de Broglie wavelength?

39.5 • An electron is moving with a speed of 8.00×10^6 m/s. What is the speed of a proton that has the same de Broglie wavelength as this electron?

39.6 • (a) A nonrelativistic free particle with mass m has kinetic energy K. Derive an expression for the de Broglie wavelength of the particle in terms of m and K. (b) What is the de Broglie wavelength of an 800-eV electron?

39.7 • (a) If a photon and an electron each have the same energy of 20.0 eV, find the wavelength of each. (b) If a photon and an electron each have the same wavelength of 250 nm, find the energy of each. (c) You want to study an organic molecule that is about 250 nm long using either a photon or an electron microscope. Approximately what wavelength should you use, and which probe, the electron or the photon, is likely to damage the molecule the least?

39.8 •• What is the de Broglie wavelength for an electron with speed (a) $v = 0.480c$ and (b) $v = 0.960c$? (*Hint:* Use the correct relativistic expression for linear momentum if necessary.)

39.9 • **Wavelength of a Bullet.** Calculate the de Broglie wavelength of a 5.00-g bullet that is moving at 340 m/s. Will the bullet exhibit wavelike properties?

39.10 •• Through what potential difference must electrons be accelerated if they are to have (a) the same wavelength as an x ray of wavelength 0.220 nm and (b) the same energy as the x ray in part (a)?

39.11 •• (a) What accelerating potential is needed to produce electrons of wavelength 5.00 nm? (b) What would be the energy of photons having the same wavelength as these electrons? (c) What would be the wavelength of photons having the same energy as the electrons in part (a)?

39.12 •• CP A beam of electrons is accelerated from rest through a potential difference of 0.100 kV and then passes through a thin slit. When viewed far from the slit, the diffracted beam shows its first diffraction minima at $\pm 14.6°$ from the original direction of the beam. (a) Do we need to use relativity formulas? How do you know? (b) How wide is the slit?

39.13 •• A beam of neutrons that all have the same energy scatters from atoms that have a spacing of 0.0910 nm in the surface plane of a crystal. The $m = 1$ intensity maximum occurs when the angle θ in Fig. 39.2 is 28.6°. What is the kinetic energy (in electron volts) of each neutron in the beam?

39.14 • (a) In an electron microscope, what accelerating voltage is needed to produce electrons with wavelength 0.0600 nm? (b) If protons are used instead of electrons, what accelerating voltage is needed to produce protons with wavelength 0.0600 nm? (*Hint:* In each case the initial kinetic energy is negligible.)

39.15 • A CD-ROM is used instead of a crystal in an electron-diffraction experiment. The surface of the CD-ROM has tracks of tiny pits with a uniform spacing of 1.60 μm. (a) If the speed of the electrons is 1.26×10^4 m/s, at which values of θ will the $m = 1$ and $m = 2$ intensity maxima appear? (b) The scattered electrons in these maxima strike at normal incidence a piece of photographic film that is 50.0 cm from the CD-ROM. What is the spacing on the film between these maxima?

Section 39.2 The Nuclear Atom and Atomic Spectra

39.16 •• CP A 4.78-MeV alpha particle from a ^{226}Ra decay makes a head-on collision with a uranium nucleus. A uranium nucleus has 92 protons. (a) What is the distance of closest approach of the alpha particle to the center of the nucleus? Assume that the uranium nucleus remains at rest and that the distance of closest approach is much greater than the radius of the uranium nucleus. (b) What is the force on the alpha particle at the instant when it is at the distance of closest approach?

39.17 • A beam of alpha particles is incident on a target of lead. A particular alpha particle comes in "head-on" to a particular lead nucleus and stops 6.50×10^{-14} m away from the center of the nucleus. (This point is well outside the nucleus.) Assume that the lead nucleus, which has 82 protons, remains at rest. The mass of the alpha particle is 6.64×10^{-27} kg. (a) Calculate the electrostatic potential energy at the instant that the alpha particle stops. Express your result in joules and in MeV. (b) What initial kinetic energy (in joules and in MeV) did the alpha particle have? (c) What was the initial speed of the alpha particle?

Section 39.3 Energy Levels and the Bohr Model of the Atom

39.18 • The silicon–silicon single bond that forms the basis of the mythical silicon-based creature the Horta has a bond strength of 3.80 eV. What wavelength of photon would you need in a (mythical) phasor disintegration gun to destroy the Horta?

39.19 •• A hydrogen atom is in a state with energy -1.51 eV. In the Bohr model, what is the angular momentum of the electron in the atom, with respect to an axis at the nucleus?

39.20 • A hydrogen atom initially in its ground level absorbs a photon, which excites the atom to the $n = 3$ level. Determine the wavelength and frequency of the photon.

39.21 • A triply ionized beryllium ion, Be^{3+} (a beryllium atom with three electrons removed), behaves very much like a hydrogen atom except that the nuclear charge is four times as great. (a) What is the ground-level energy of Be^{3+}? How does this compare to the ground-level energy of the hydrogen atom? (b) What is the ionization energy of Be^{3+}? How does this compare to the ionization energy of the hydrogen atom? (c) For the hydrogen atom, the wavelength of the photon emitted in the $n = 2$ to $n = 1$ transition is 122 nm (see Example 39.6). What is the wavelength of the photon emitted when a Be^{3+} ion undergoes this transition? (d) For a given value of n, how does the radius of an orbit in Be^{3+} compare to that for hydrogen?

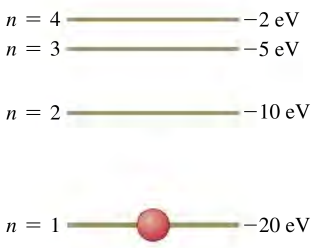
39.22 •• Consider the Bohr-model description of a hydrogen atom. (a) Calculate $E_2 - E_1$ and $E_{10} - E_9$. As n increases, does the energy separation between adjacent energy levels increase, decrease, or stay the same? (b) Show that $E_{n+1} - E_n$ approaches $(27.2\text{ eV})/n^3$ as n becomes large. (c) How does $r_{n+1} - r_n$ depend on n? Does the radial distance between adjacent orbits increase, decrease, or stay the same as n increases?

39.23 • (a) Using the Bohr model, calculate the speed of the electron in a hydrogen atom in the $n = 1$, 2, and 3 levels. (b) Calculate the orbital period in each of these levels. (c) The average lifetime of the first excited level of a hydrogen atom is 1.0×10^{-8} s. In the Bohr model, how many orbits does an electron in the $n = 2$ level complete before returning to the ground level?

39.24 • Consider the Bohr-model description of a hydrogen atom. (a) Calculate K_1, U_1, and E_1 for the $n = 1$ energy level. How are K_1 and U_1 related? (b) Show that for any value of n, both $U_n = -2K_n$ and $K_n = -E_n$.

39.25 • CP The energy-level scheme for the hypothetical one-electron element Searsium is shown in **Fig. E39.25**. The potential energy is taken to be zero for an electron at an infinite distance from the nucleus. (a) How much energy (in electron volts) does it take to ionize an electron from the ground level? (b) An 18-eV photon is absorbed by a Searsium atom in its ground level. As the

Figure **E39.25**

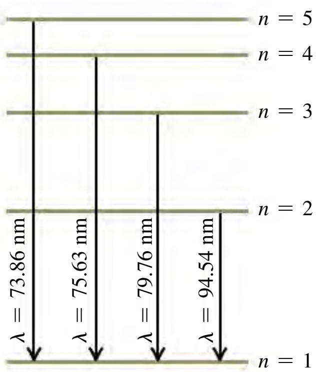


atom returns to its ground level, what possible energies can the emitted photons have? Assume that there can be transitions between all pairs of levels. (c) What will happen if a photon with an energy of 8 eV strikes a Searsium atom in its ground level? Why? (d) Photons emitted in the Searsium transitions $n = 3 \rightarrow n = 2$ and $n = 3 \rightarrow n = 1$ will eject photoelectrons from an unknown metal, but the photon emitted from the transition $n = 4 \rightarrow n = 3$ will not. What are the limits (maximum and minimum possible values) of the work function of the metal?

39.26 • (a) For one-electron ions with nuclear charge Z, what is the speed of the electron in a Bohr-model orbit labeled with n? Give your answer in terms of v_1, the orbital speed for the $n = 1$ Bohr orbit in hydrogen. (b) What is the largest value of Z for which the $n = 1$ orbital speed is less than 10% of the speed of light in vacuum?

39.27 • In a set of experiments on a hypothetical one-electron atom, you measure the wavelengths of the photons emitted from transitions ending in the ground level ($n = 1$), as shown in the energy-level diagram in **Fig. E39.27**. You also observe that it takes 17.50 eV to ionize this atom. (a) What is the energy of the atom in each of the levels ($n = 1$, $n = 2$, etc.) shown in the figure? (b) If an electron made a transition from the $n = 4$ to the $n = 2$ level, what wavelength of light would it emit?

Figure **E39.27**



39.28 • Find the longest and shortest wavelengths in the Lyman and Paschen series for hydrogen. In what region of the electromagnetic spectrum does each series lie?

39.29 • (a) An atom initially in an energy level with $E = -6.52$ eV absorbs a photon that has wavelength 860 nm. What is the internal energy of the atom after it absorbs the photon? (b) An atom initially in an energy level with $E = -2.68$ eV emits a photon that has wavelength 420 nm. What is the internal energy of the atom after it emits the photon?

39.30 •• Use Balmer's formula to calculate (a) the wavelength, (b) the frequency, and (c) the photon energy for the H_γ line of the Balmer series for hydrogen.

Section 39.4 The Laser

39.31 • BIO **Laser Surgery.** Using a mixture of CO_2, N_2, and sometimes He, CO_2 lasers emit a wavelength of 10.6 μm. At power outputs of 0.100 kW, such lasers are used for surgery. How many photons per second does a CO_2 laser deliver to the tissue during its use in an operation?

39.32 • BIO **Removing Birthmarks.** Pulsed dye lasers emit light of wavelength 585 nm in 0.45-ms pulses to remove skin blemishes such as birthmarks. The beam is usually focused onto a circular spot 5.0 mm in diameter. Suppose that the output of one such laser is 20.0 W. (a) What is the energy of each photon, in eV? (b) How many photons per square millimeter are delivered to the blemish during each pulse?

39.33 • How many photons per second are emitted by a 7.50-mW CO_2 laser that has a wavelength of 10.6 μm?

39.34 • BIO **PRK Surgery.** Photorefractive keratectomy (PRK) is a laser-based surgical procedure that corrects near- and farsightedness by removing part of the lens of the eye to change its curvature and hence focal length. This procedure can remove layers 0.25 μm thick using pulses lasting 12.0 ns from a laser beam of wavelength 193 nm. Low-intensity beams can be used because each individual photon has enough energy to break the covalent bonds of the tissue. (a) In what part of the electromagnetic spectrum does this light lie? (b) What is the energy of a single photon? (c) If a 1.50-mW beam is used, how many photons are delivered to the lens in each pulse?

39.35 • A large number of neon atoms are in thermal equilibrium. What is the ratio of the number of atoms in a 5s state to the number in a 3p state at (a) 300 K; (b) 600 K; (c) 1200 K? The energies of these states, relative to the ground state, are $E_{5s} = 20.66$ eV and $E_{3p} = 18.70$ eV. (d) At any of these temperatures, the rate at which a neon gas will spontaneously emit 632.8-nm radiation is quite low. Explain why.

39.36 • Figure 39.19a shows the energy levels of the sodium atom. The two lowest excited levels are shown in columns labeled $^2P_{3/2}$ and $^2P_{1/2}$. Find the ratio of the number of atoms in a $^2P_{3/2}$ state to the number in a $^2P_{1/2}$ state for a sodium gas in thermal equilibrium at 500 K. In which state are more atoms found?

Section 39.5 Continuous Spectra

39.37 •• A 100-W incandescent light bulb has a cylindrical tungsten filament 30.0 cm long, 0.40 mm in diameter, and with an emissivity of 0.26. (a) What is the temperature of the filament? (b) For what wavelength does the spectral emittance of the bulb peak? (c) Incandescent light bulbs are not very efficient sources of visible light. Explain why this is so.

39.38 • Determine λ_m, the wavelength at the peak of the Planck distribution, and the corresponding frequency f, at these temperatures: (a) 3.00 K; (b) 300 K; (c) 3000 K.

39.39 • Radiation has been detected from space that is characteristic of an ideal radiator at $T = 2.728$ K. (This radiation is a relic of the Big Bang at the beginning of the universe.) For this temperature, at what wavelength does the Planck distribution peak? In what part of the electromagnetic spectrum is this wavelength?

39.40 • The shortest visible wavelength is about 400 nm. What is the temperature of an ideal radiator whose spectral emittance peaks at this wavelength?

39.41 •• Two stars, both of which behave like ideal blackbodies, radiate the same total energy per second. The cooler one has a surface temperature T and a diameter 3.0 times that of the hotter star. (a) What is the temperature of the hotter star in terms of T? (b) What is the ratio of the peak-intensity wavelength of the hot star to the peak-intensity wavelength of the cool star?

39.42 • The wavelength 10.0 μm is in the infrared region of the electromagnetic spectrum, whereas 600 nm is in the visible region and 100 nm is in the ultraviolet. What is the temperature of an ideal blackbody for which the peak wavelength λ_m is equal to each of these wavelengths?

39.43 • **Sirius B.** The brightest star in the sky is Sirius, the Dog Star. It is actually a binary system of two stars, the smaller one (Sirius B) being a white dwarf. Spectral analysis of Sirius B indicates that its surface temperature is 24,000 K and that it radiates energy at a total rate of 1.0×10^{25} W. Assume that it behaves like an ideal blackbody. (a) What is the total radiated intensity of Sirius B? (b) What is the peak-intensity wavelength? Is this wavelength visible to humans? (c) What is the radius of Sirius B? Express your answer in kilometers and as a fraction of our sun's radius. (d) Which star radiates more *total* energy per second, the hot Sirius B or the (relatively) cool sun with a surface temperature of 5800 K? To find out, calculate the ratio of the total power radiated by our sun to the power radiated by Sirius B.

Section 39.6 The Uncertainty Principle Revisited

39.44 • A pesky 1.5-mg mosquito is annoying you as you attempt to study physics in your room, which is 5.0 m wide and 2.5 m high. You decide to swat the bothersome insect as it flies toward you, but you need to estimate its speed to make a successful hit. (a) What is the maximum uncertainty in the horizontal position of the mosquito? (b) What limit does the Heisenberg uncertainty principle place on your ability to know the horizontal velocity of this mosquito? Is this limitation a serious impediment to your attempt to swat it?

39.45 • (a) The uncertainty in the y-component of a proton's position is 2.0×10^{-12} m. What is the minimum uncertainty in a simultaneous measurement of the y-component of the proton's velocity? (b) The uncertainty in the z-component of an electron's velocity is 0.250 m/s. What is the minimum uncertainty in a simultaneous measurement of the z-coordinate of the electron?

39.46 • A 10.0-g marble is gently placed on a horizontal tabletop that is 1.75 m wide. (a) What is the maximum uncertainty in the horizontal position of the marble? (b) According to the Heisenberg uncertainty principle, what is the minimum uncertainty in the horizontal velocity of the marble? (c) In light of your answer to part (b), what is the longest time the marble could remain on the table? Compare this time to the age of the universe, which is approximately 14 billion years. (*Hint:* Can you know that the horizontal velocity of the marble is *exactly* zero?)

39.47 • A scientist has devised a new method of isolating individual particles. He claims that this method enables him to detect simultaneously the position of a particle along an axis with a standard deviation of 0.12 nm and its momentum component along this axis with a standard deviation of $3.0 \times 10^{-25}\ \text{kg} \cdot \text{m/s}$. Use the Heisenberg uncertainty principle to evaluate the validity of this claim.

39.48 • (a) The x-coordinate of an electron is measured with an uncertainty of 0.30 mm. What is the x-component of the electron's velocity, v_x, if the minimum percent uncertainty in a simultaneous measurement of v_x is 1.0%? (b) Repeat part (a) for a proton.

39.49 • An atom in a metastable state has a lifetime of 5.2 ms. What is the uncertainty in energy of the metastable state?

PROBLEMS

39.50 • An atom with mass m emits a photon of wavelength λ. (a) What is the recoil speed of the atom? (b) What is the kinetic energy K of the recoiling atom? (c) Find the ratio K/E, where E is the energy of the emitted photon. If this ratio is much less than unity, the recoil of the atom can be neglected in the emission process. Is the recoil of the atom more important for small or large atomic masses? For long or short wavelengths? (d) Calculate K (in electron volts) and K/E for a hydrogen atom (mass 1.67×10^{-27} kg) that emits an ultraviolet photon of energy 10.2 eV. Is recoil an important consideration in this emission process?

39.51 •• The negative muon has a charge equal to that of an electron but a mass that is 207 times as great. Consider a hydrogenlike atom consisting of a proton and a muon. (a) What is the reduced mass of the atom? (b) What is the ground-level energy (in electron volts)? (c) What is the wavelength of the radiation emitted in the transition from the $n = 2$ level to the $n = 1$ level?

39.52 • A large number of hydrogen atoms are in thermal equilibrium. Let n_2/n_1 be the ratio of the number of atoms in an $n = 2$ excited state to the number of atoms in an $n = 1$ ground state. At what temperature is n_2/n_1 equal to (a) 10^{-12}; (b) 10^{-8}; (c) 10^{-4}? (d) Like the sun, other stars have continuous spectra with dark absorption lines (see Fig. 39.9). The absorption takes place in the star's atmosphere, which in all stars is composed primarily of hydrogen. Explain why the Balmer absorption lines are relatively weak in stars with low atmospheric temperatures such as the sun (atmosphere temperature 5800 K) but strong in stars with higher atmospheric temperatures.

39.53 • (a) What is the smallest amount of energy in electron volts that must be given to a hydrogen atom initially in its ground level so that it can emit the H_α line in the Balmer series? (b) How many different possibilities of spectral-line emissions are there for this atom when the electron starts in the $n = 3$ level and eventually ends up in the ground level? Calculate the wavelength of the emitted photon in each case.

39.54 •• In the Bohr model of the hydrogen atom, what is the de Broglie wavelength of the electron when it is in (a) the $n = 1$ level and (b) the $n = 4$ level? In both cases, compare the de Broglie wavelength to the circumference $2\pi r_n$ of the orbit.

39.55 ••• A sample of hydrogen atoms is irradiated with light with wavelength 85.5 nm, and electrons are observed leaving the gas. (a) If each hydrogen atom were initially in its ground level, what would be the maximum kinetic energy in electron volts of these photoelectrons? (b) A few electrons are detected with energies as much as 10.2 eV greater than the maximum kinetic energy calculated in part (a). How can this be?

39.56 •• Take 380–750 nm to be the wavelength range of the visible spectrum. (a) What are the largest and smallest photon energies for visible light? (b) The lowest six energy levels of the one-electron He^+ ion are given in Fig. 39.27. For these levels, what transitions give absorption or emission of visible-light photons?

39.57 •• **The Red Supergiant Betelgeuse.** The star Betelgeuse has a surface temperature of 3000 K and is 600 times the diameter of our sun. (If our sun were that large, we would be inside it!) Assume that it radiates like an ideal blackbody. (a) If Betelgeuse were to radiate all of its energy at the peak-intensity wavelength, how many photons per second would it radiate? (b) Find the ratio of the power radiated by Betelgeuse to the power radiated by our sun (at 5800 K).

39.58 •• CP Light from an ideal spherical blackbody 15.0 cm in diameter is analyzed by using a diffraction grating that has 3850 lines/cm. When you shine this light through the grating, you observe that the peak-intensity wavelength forms a first-order bright fringe at $\pm 14.4°$ from the central bright fringe. (a) What is the temperature of the blackbody? (b) How long will it take this sphere to radiate 12.0 MJ of energy at constant temperature?

39.59 • What must be the temperature of an ideal blackbody so that photons of its radiated light having the peak-intensity wavelength can excite the electron in the Bohr-model hydrogen atom from the ground level to the $n = 4$ energy level?

39.60 •• **An Ideal Blackbody.** A large cavity that has a very small hole and is maintained at a temperature T is a good approximation to an ideal radiator or blackbody. Radiation can pass into or out of the cavity only through the hole. The cavity is a perfect absorber, since any radiation incident on the hole becomes trapped inside the cavity. Such a cavity at 400°C has a hole with area 4.00 mm^2. How long does it take for the cavity to radiate 100 J of energy through the hole?

39.61 •• **CALC** (a) Write the Planck distribution law in terms of the frequency f, rather than the wavelength λ, to obtain $I(f)$. (b) Show that

$$\int_0^\infty I(\lambda)\,d\lambda = \frac{2\pi^5 k^4}{15c^2h^3}T^4$$

where $I(\lambda)$ is the Planck distribution formula of Eq. (39.24). *Hint:* Change the integration variable from λ to f. You will need to use the following tabulated integral:

$$\int_0^\infty \frac{x^3}{e^{\alpha x}-1}dx = \frac{1}{240}\left(\frac{2\pi}{\alpha}\right)^4$$

(c) The result of part (b) is I and has the form of the Stefan–Boltzmann law, $I = \sigma T^4$ (Eq. 39.19). Evaluate the constants in part (b) to show that σ has the value given in Section 39.5.

39.62 •• **CP** A beam of 40-eV electrons traveling in the $+x$-direction passes through a slit that is parallel to the y-axis and 5.0 μm wide. The diffraction pattern is recorded on a screen 2.5 m from the slit. (a) What is the de Broglie wavelength of the electrons? (b) How much time does it take the electrons to travel from the slit to the screen? (c) Use the width of the central diffraction pattern to calculate the uncertainty in the y-component of momentum of an electron just after it has passed through the slit. (d) Use the result of part (c) and the Heisenberg uncertainty principle (Eq. 39.29 for y) to estimate the minimum uncertainty in the y-coordinate of an electron just after it has passed through the slit. Compare your result to the width of the slit.

39.63 • (a) What is the energy of a photon that has wavelength 0.10 μm? (b) Through approximately what potential difference must electrons be accelerated so that they will exhibit wave nature in passing through a pinhole 0.10 μm in diameter? What is the speed of these electrons? (c) If protons rather than electrons were used, through what potential difference would protons have to be accelerated so they would exhibit wave nature in passing through this pinhole? What would be the speed of these protons?

39.64 • **CP** Electrons go through a single slit 300 nm wide and strike a screen 24.0 cm away. At angles of $\pm 20.0°$ from the center of the diffraction pattern, no electrons hit the screen, but electrons hit at all points closer to the center. (a) How fast were these electrons moving when they went through the slit? (b) What will be the next pair of larger angles at which no electrons hit the screen?

39.65 •• **CP** A beam of electrons is accelerated from rest and then passes through a pair of identical thin slits that are 1.25 nm apart. You observe that the first double-slit interference dark fringe occurs at $\pm 18.0°$ from the original direction of the beam when viewed on a distant screen. (a) Are these electrons relativistic? How do you know? (b) Through what potential difference were the electrons accelerated?

39.66 • **CP** Coherent light is passed through two narrow slits whose separation is 20.0 μm. The second-order bright fringe in the interference pattern is located at an angle of 0.0300 rad. If electrons are used instead of light, what must the kinetic energy (in electron volts) of the electrons be if they are to produce an interference pattern for which the second-order maximum is also at 0.0300 rad?

39.67 •• **CP** An electron beam and a photon beam pass through identical slits. On a distant screen, the first dark fringe occurs at the same angle for both of the beams. The electron speeds are much slower than that of light. (a) Express the energy of a photon in terms of the kinetic energy K of one of the electrons. (b) Which is greater, the energy of a photon or the kinetic energy of an electron?

39.68 • **BIO** What is the de Broglie wavelength of a red blood cell, with mass 1.00×10^{-11} g, that is moving with a speed of 0.400 cm/s? Do we need to be concerned with the wave nature of the blood cells when we describe the flow of blood in the body?

39.69 • High-speed electrons are used to probe the interior structure of the atomic nucleus. For such electrons the expression $\lambda = h/p$ still holds, but we must use the relativistic expression for momentum, $p = mv/\sqrt{1 - v^2/c^2}$. (a) Show that the speed of an electron that has de Broglie wavelength λ is

$$v = \frac{c}{\sqrt{1 + (mc\lambda/h)^2}}$$

(b) The quantity h/mc equals 2.426×10^{-12} m. (As we saw in Section 38.3, this same quantity appears in Eq. (38.7), the expression for Compton scattering of photons by electrons.) If λ is small compared to h/mc, the denominator in the expression found in part (a) is close to unity and the speed v is very close to c. In this case it is convenient to write $v = (1 - \Delta)c$ and express the speed of the electron in terms of Δ rather than v. Find an expression for Δ valid when $\lambda \ll h/mc$. [*Hint:* Use the binomial expansion $(1 + z)^n = 1 + nz + [n(n-1)z^2/2] + \cdots$, valid for the case $|z| < 1$.] (c) How fast must an electron move for its de Broglie wavelength to be 1.00×10^{-15} m, comparable to the size of a proton? Express your answer in the form $v = (1 - \Delta)c$, and state the value of Δ.

39.70 • Suppose that the uncertainty of position of an electron is equal to the radius of the $n = 1$ Bohr orbit for hydrogen. Calculate the simultaneous minimum uncertainty of the corresponding momentum component, and compare this with the magnitude of the momentum of the electron in the $n = 1$ Bohr orbit. Discuss your results.

39.71 • **CP** (a) A particle with mass m has kinetic energy equal to three times its rest energy. What is the de Broglie wavelength of this particle? (*Hint:* You must use the relativistic expressions for momentum and kinetic energy: $E^2 = (pc)^2 + (mc^2)^2$ and $K = E - mc^2$.) (b) Determine the numerical value of the kinetic energy (in MeV) and the wavelength (in meters) if the particle in part (a) is (i) an electron and (ii) a proton.

39.72 • **Proton Energy in a Nucleus.** The radii of atomic nuclei are of the order of 5.0×10^{-15} m. (a) Estimate the minimum uncertainty in the momentum of a proton if it is confined within a nucleus. (b) Take this uncertainty in momentum to be an estimate of the magnitude of the momentum. Use the relativistic relationship between energy and momentum, Eq. (37.39), to obtain an estimate of the kinetic energy of a proton confined within a nucleus. (c) For a proton to remain bound within a nucleus, what must the magnitude of the (negative) potential energy for a proton be within the nucleus? Give your answer in eV and in MeV. Compare to the potential energy for an electron in a hydrogen atom, which has a magnitude of a few tens of eV. (This shows why the interaction that binds the nucleus together is called the "strong nuclear force.")

39.73 • **Electron Energy in a Nucleus.** The radii of atomic nuclei are of the order of 5.0×10^{-15} m. (a) Estimate the minimum uncertainty in the momentum of an electron if it is confined within a nucleus. (b) Take this uncertainty in momentum to be an estimate of the magnitude of the momentum. Use the relativistic relationship between energy and momentum, Eq. (37.39), to obtain an estimate of the kinetic energy of an electron confined within a nucleus. (c) Compare the energy calculated in part (b) to the magnitude of the Coulomb potential energy of a proton and an electron separated by 5.0×10^{-15} m. On the basis of your result, could there be electrons within the nucleus? (*Note:* It is interesting to compare this result to that of Problem 39.72.)

39.74 • The neutral pion (π^0) is an unstable particle produced in high-energy particle collisions. Its mass is about 264 times that of the electron, and it exists for an average lifetime of 8.4×10^{-17} s before decaying into two gamma-ray photons. Using the relationship $E = mc^2$ between rest mass and energy, find the uncertainty in the mass of the particle and express it as a fraction of the mass.

39.75 • **Doorway Diffraction.** If your wavelength were 1.0 m, you would undergo considerable diffraction in moving through a doorway. (a) What must your speed be for you to have this wavelength? (Assume that your mass is 60.0 kg.) (b) At the speed calculated in part (a), how many years would it take you to move 0.80 m (one step)? Will you notice diffraction effects as you walk through doorways?

39.76 • **Atomic Spectra Uncertainties.** A certain atom has an energy level 2.58 eV above the ground level. Once excited to this level, the atom remains in this level for 1.64×10^{-7} s (on average) before emitting a photon and returning to the ground level. (a) What is the energy of the photon (in electron volts)? What is its wavelength (in nanometers)? (b) What is the smallest possible uncertainty in energy of the photon? Give your answer in electron volts. (c) Show that $|\Delta E/E| = |\Delta\lambda/\lambda|$ if $|\Delta\lambda/\lambda| \ll 1$. Use this to calculate the magnitude of the smallest possible uncertainty in the wavelength of the photon. Give your answer in nanometers.

39.77 •• For x rays with wavelength 0.0300 nm, the $m = 1$ intensity maximum for a crystal occurs when the angle θ in Fig. 39.2 is 35.8°. At what angle θ does the $m = 1$ maximum occur when a beam of 4.50-keV electrons is used instead? Assume that the electrons also scatter from the atoms in the surface plane of this same crystal.

39.78 •• A certain atom has an energy state 3.50 eV above the ground state. When excited to this state, the atom remains for 2.0 μs, on average, before it emits a photon and returns to the ground state. (a) What are the energy and wavelength of the photon? (b) What is the smallest possible uncertainty in energy of the photon?

39.79 •• BIO **Structure of a Virus.** To investigate the structure of extremely small objects, such as viruses, the wavelength of the probing wave should be about one-tenth the size of the object for sharp images. But as the wavelength gets shorter, the energy of a photon of light gets greater and could damage or destroy the object being studied. One alternative is to use electron matter waves instead of light. Viruses vary considerably in size, but 50 nm is not unusual. Suppose you want to study such a virus, using a wave of wavelength 5.00 nm. (a) If you use light of this wavelength, what would be the energy (in eV) of a single photon? (b) If you use an electron of this wavelength, what would be its kinetic energy (in eV)? Is it now clear why matter waves (such as in the electron microscope) are often preferable to electromagnetic waves for studying microscopic objects?

39.80 •• CALC **Zero-Point Energy.** Consider a particle with mass m moving in a potential $U = \frac{1}{2}kx^2$, as in a mass–spring system. The total energy of the particle is $E = (p^2/2m) + \frac{1}{2}kx^2$. Assume that p and x are approximately related by the Heisenberg uncertainty principle, so $px \approx h$. (a) Calculate the minimum possible value of the energy E, and the value of x that gives this minimum E. This lowest possible energy, which is not zero, is called the *zero-point energy*. (b) For the x calculated in part (a), what is the ratio of the kinetic to the potential energy of the particle?

39.81 •• CALC A particle with mass m moves in a potential energy $U(x) = A|x|$, where A is a positive constant. In a simplified picture, quarks (the constituents of protons, neutrons, and other particles, as will be described in Chapter 44) have a potential energy of interaction of approximately this form, where x represents the separation between a pair of quarks. Because $U(x) \to \infty$ as $x \to \infty$, it's not possible to separate quarks from each other (a phenomenon called *quark confinement*). (a) Classically, what is the force acting on this particle as a function of x? (b) Using the uncertainty principle as in Problem 39.80, determine approximately the zero-point energy of the particle.

39.82 •• Imagine another universe in which the value of Planck's constant is 0.0663 J · s, but in which the physical laws and all other physical constants are the same as in our universe. In this universe, two physics students are playing catch. They are 12 m apart, and one throws a 0.25-kg ball directly toward the other with a speed of 6.0 m/s. (a) What is the uncertainty in the ball's horizontal momentum, in a direction perpendicular to that in which it is being thrown, if the student throwing the ball knows that it is located within a cube with volume 125 cm^3 at the time she throws it? (b) By what horizontal distance could the ball miss the second student?

39.83 •• DATA For your work in a mass spectrometry lab, you are investigating the absorption spectrum of one-electron ions. To maintain the atoms in an ionized state, you hold them at low density in an ion trap, a device that uses a configuration of electric fields to confine ions. The majority of the ions are in their ground state, so that is the initial state for the absorption transitions that you observe. (a) If the longest wavelength that you observe in the absorption spectrum is 13.56 nm, what is the atomic number Z for the ions? (b) What is the next shorter wavelength that the ions will absorb? (c) When one of the ions absorbs a photon of wavelength 6.78 nm, a free electron is produced. What is the kinetic energy (in electron volts) of the electron?

39.84 •• DATA In the crystallography lab where you work, you are given a single crystal of an unknown substance to identify. To obtain one piece of information about the substance, you repeat the Davisson–Germer experiment to determine the spacing of the atoms in the surface planes of the crystal. You start with electrons that are essentially stationary and accelerate them through a potential difference of magnitude V_{ac}. The electrons then scatter off the atoms on the surface of the crystal (as in Fig. 39.3b). Next you measure the angle θ that locates the first-order diffraction peak. Finally, you repeat the measurement for different values of V_{ac}. Your results are given in the table.

V_{ac} (V)	106.3	69.1	49.9	25.2	16.9	13.6
θ (°)	20.4	24.8	30.2	45.5	59.1	73.1

(a) Graph your data in the form $\sin\theta$ versus $1/\sqrt{V_{ac}}$. What is the slope of the straight line that best fits the data points when plotted in this way? (b) Use your results from part (a) to calculate the value of d for this crystal.

39.85 •• **DATA** As an amateur astronomer, you are studying the apparent brightness of stars. You know that a star's apparent brightness depends on its distance from the earth and also on the fraction of its radiated energy that is in the visible region of the electromagnetic spectrum. But, as a first step, you search the Internet for information on the surface temperatures and radii of some selected stars so that you can calculate their total radiated power. You find the data given in the table.

Star	Polaris	Vega	Antares	α Centauri B
Surface temperature (K)	6015	9602	3400	5260
Radius relative to that of the sun (R_{sun})	46	2.73	883	0.865

The radius is given in units of the radius of the sun, $R_{sun} = 6.96 \times 10^8$ m. The surface temperature is the effective temperature that gives the measured photon luminosity of the star if the star is assumed to radiate as an ideal blackbody. The photon luminosity is the power emitted in the form of photons. (a) Which star in the table has the greatest radiated power? (b) For which of these stars, if any, is the peak wavelength λ_m in the visible range (380–750 nm)? (c) The sun has a total radiated power of 3.85×10^{26} W. Which of these stars, if any, have a total radiated power less than that of our sun?

CHALLENGE PROBLEMS

39.86 ••• **CP CALC** You have entered a contest in which the contestants drop a marble with mass 20.0 g from the roof of a building onto a small target 25.0 m below. From uncertainty considerations, what is the typical distance by which you will miss the target, given that you aim with the highest possible precision? (*Hint:* The uncertainty Δx_f in the x-coordinate of the marble when it reaches the ground comes in part from the uncertainty Δx_i in the x-coordinate initially and in part from the initial uncertainty in v_x. The latter gives rise to an uncertainty Δv_x in the horizontal motion of the marble as it falls. The values of Δx_i and Δv_x are related by the uncertainty principle. A small Δx_i gives rise to a large Δv_x, and vice versa. Find the value of Δx_i that gives the smallest total uncertainty in x at the ground. Ignore any effects of air resistance.)

39.87 ••• (a) Show that in the Bohr model, the frequency of revolution of an electron in its circular orbit around a stationary hydrogen nucleus is $f = me^4/4\epsilon_0^2 n^3 h^3$. (b) In classical physics, the frequency of revolution of the electron is equal to the frequency of the radiation that it emits. Show that when n is very large, the frequency of revolution does indeed equal the radiated frequency calculated from Eq. (39.5) for a transition from $n_1 = n + 1$ to $n_2 = n$. (This illustrates Bohr's *correspondence principle,* which is often used as a check on quantum calculations. When n is small, quantum physics gives results that are very different from those of classical physics. When n is large, the differences are not significant, and the two methods then "correspond." In fact, when Bohr first tackled the hydrogen atom problem, he sought to determine f as a function of n such that it would correspond to classical results for large n.)

PASSAGE PROBLEMS

BIO ION MICROSCOPES. Whereas electron microscopes make use of the wave properties of electrons, ion microscopes make use of the wave properties of atomic ions, such as helium ions (He^+), to image materials. A helium ion has a mass 7300 times that of an electron. In a typical helium-ion microscope, helium ions are accelerated by a high voltage of 10–50 kV and focused into a beam onto the sample to be imaged. At these energies, the ions don't travel very far into the sample, so this type of microscope is used primarily for the surface imaging of biological structures. The use of helium ions with much greater energies (in the MeV range) has been proposed as a way to image the entire thickness of a sample, because these faster helium ions can pass all the way through biological samples such as cells. In this type of ion microscope, the energy lost as the ion beam passes through different parts of a cell can be measured and related to the distribution of material in the cell, with thicker parts of the cell causing greater energy loss. [Source: "Whole-Cell Imaging at Nanometer Resolutions Using Fast and Slow Focused Helium Ions," by Xiao Chen et al., *Biophysical Journal* 101(7): 1788–1793, Oct. 5, 2011.]

39.88 How does the wavelength of a helium ion compare to that of an electron accelerated through the same potential difference? (a) The helium ion has a longer wavelength, because it has greater mass. (b) The helium ion has a shorter wavelength, because it has greater mass. (c) The wavelengths are the same, because the kinetic energy is the same. (d) The wavelengths are the same, because the electric charge is the same.

39.89 Can the first type of helium-ion microscope, used for surface imaging, produce helium ions with a wavelength of 0.1 pm? (a) Yes; the voltage required is 21 kV. (b) Yes; the voltage required is 42 kV. (c) No; a voltage higher than 50 kV is required. (d) No; a voltage lower than 10 kV is required.

39.90 Why is it easier to use helium ions rather than neutral helium atoms in such a microscope? (a) Helium atoms are not electrically charged, and only electrically charged particles have wave properties. (b) Helium atoms form molecules, which are too large to have wave properties. (c) Neutral helium atoms are more difficult to focus with electric and magnetic fields. (d) Helium atoms have much larger mass than helium ions do and thus are more difficult to accelerate.

39.91 In the second type of helium-ion microscope, a 1.2-MeV ion passing through a cell loses 0.2 MeV per μm of cell thickness. If the energy of the ion can be measured to 6 keV, what is the smallest difference in thickness that can be discerned? (a) 0.03 μm; (b) 0.06 μm; (c) 3 μm; (d) 6 μm.

Answers

Chapter Opening Question ?

(i) The smallest detail visible in an image is comparable to the wavelength used to make the image. Electrons can easily be given a large momentum p and hence a short wavelength $\lambda = h/p$, and so can be used to resolve extremely fine details. (See Section 39.1.)

Test Your Understanding Questions

39.1 (a) (i), (b) no From Example 39.2, the speed of a particle is $v = h/\lambda m$ and the kinetic energy is $K = \frac{1}{2}mv^2 = (m/2)(h/\lambda m)^2 = h^2/2\lambda^2 m$. This shows that for a given wavelength, the kinetic energy is inversely proportional to the mass. Hence the proton, with a smaller mass, has more kinetic energy than the neutron. For part (b), the energy of a photon is $E = hf$, and the frequency of a photon is $f = c/\lambda$. Hence $E = hc/\lambda$ and $\lambda = hc/E = (4.136 \times 10^{-15}\ \text{eV} \cdot \text{s})(2.998 \times 10^8\ \text{m/s})/(54\ \text{eV}) = 2.3 \times 10^{-8}$ m. This is more than 100 times greater than the wavelength of an electron of the same energy. While both photons and electrons have wavelike properties, they have different relationships between their energy and momentum and hence between their frequency and wavelength.

39.2 (iii) Because the alpha particle is more massive, it won't bounce back in even a head-on collision with a proton that's initially at rest, any more than a bowling ball would when colliding with a Ping-Pong ball at rest (see Fig. 8.23b). Thus there would be *no* large-angle scattering in this case. Rutherford saw large-angle scattering in his experiment because gold nuclei are more massive than alpha particles (see Fig. 8.23a).

39.3 (iv) Figure 39.27 shows that many (though *not* all) of the energy levels of He^+ are the same as those of H. Hence photons emitted during transitions between corresponding pairs of levels in He^+ and H have the same energy E and the same wavelength $\lambda = hc/E$. An H atom that drops from the $n = 2$ level to the $n = 1$ level emits a photon of energy 10.20 eV and wavelength 122 nm (see Example 39.6); a He^+ ion emits a photon of the same energy and wavelength when it drops from the $n = 4$ level to the $n = 2$ level. Inspecting Fig. 39.27 will show you that every even-numbered level in He^+ matches a level in H, while none of the odd-numbered He^+ levels do. The first three He^+ transitions given in the question ($n = 2$ to $n = 1$, $n = 3$ to $n = 2$, and $n = 4$ to $n = 3$) all involve an odd-numbered level, so none of their wavelengths match a wavelength emitted by H atoms.

39.4 (i) In a neon light fixture, a large potential difference is applied between the ends of a neon-filled glass tube. This ionizes some of the neon atoms, allowing a current of electrons to flow through the gas. Some of the neon atoms are struck by fast-moving electrons, making them transition to an excited level. From this level the atoms undergo *spontaneous* emission, as depicted in Fig. 39.28b, and emit 632.8-nm photons in the process. No population inversion occurs and the photons are not trapped by mirrors as shown in Fig. 39.29d, so there is no stimulated emission. Hence there is no laser action.

39.5 (a) yes, (b) yes The Planck radiation law, Eq. (39.24), shows that an ideal blackbody emits radiation at *all* wavelengths: The spectral emittance $I(\lambda)$ is equal to zero only for $\lambda = 0$ and in the limit $\lambda \rightarrow \infty$. So a blackbody at 2000 K does indeed emit both x rays and radio waves. However, Fig. 39.32 shows that the spectral emittance for this temperature is very low for wavelengths much shorter than 1 μm (including x rays) and for wavelengths much longer than a few μm (including radio waves). Hence such a blackbody emits very little in the way of x rays or radio waves.

39.6 (i) and (iii) (tie), (ii) and (iv) (tie) According to the Heisenberg uncertainty principle, the smaller the uncertainty Δx in the x-coordinate, the greater the uncertainty Δp_x in the x-momentum. The relationship between Δx and Δp_x does not depend on the mass of the particle, and so is the same for a proton as for an electron.

Bridging Problem

(a) 192 nm; ultraviolet **(b)** $n = 4$
(c) $\lambda_2 = 0.665$ nm, $\lambda_3 = 0.997$ nm

? These containers hold solutions of microscopic semiconductor particles of different sizes. The particles glow when exposed to ultraviolet light; the smallest particles glow blue and the largest particles glow red. This is because the energy levels of electrons (i) are spaced farther apart in smaller particles; (ii) are spaced farther apart in larger particles; (iii) have the same spacing in all particles but have higher energies in smaller particles; (iv) have the same spacing in all particles but have higher energies in larger particles; (v) depend on the wavelength of ultraviolet light used.

40 QUANTUM MECHANICS I: WAVE FUNCTIONS

LEARNING GOALS

Looking forward at …

40.1 The wave function that describes the behavior of a particle and the Schrödinger equation that this function must satisfy.

40.2 How to calculate the wave functions and energy levels for a particle confined to a box.

40.3 How to analyze the quantum-mechanical behavior of a particle in a potential well.

40.4 How quantum mechanics makes it possible for particles to go where Newtonian mechanics says they cannot.

40.5 How to use quantum mechanics to analyze a harmonic oscillator.

40.6 How measuring a quantum-mechanical system can change that system's state.

Looking back at …

7.5 Potential wells.

14.2, 14.3 Harmonic oscillators.

15.3, 15.7, 15.8 Wave functions for waves on a string; standing waves.

32.3 Wave functions for electromagnetic waves.

38.1, 38.4 Work function; photon interpretation of interference and diffraction.

39.1, 39.3, 39.5, 39.6 De Broglie relationships; Bohr model for hydrogen; Planck radiation law; Heisenberg uncertainty principle.

In Chapter 39 we found that particles can behave like waves. In fact, it turns out that we can use the wave picture to completely describe the behavior of a particle. This approach, called *quantum mechanics,* is the key to understanding the behavior of matter on the molecular, atomic, and nuclear scales. In this chapter we'll see how to find the *wave function* of a particle by solving the *Schrödinger equation,* which is as fundamental to quantum mechanics as Newton's laws are to mechanics or as Maxwell's equations are to electromagnetism.

We'll begin with a quantum-mechanical analysis of a *free particle* that moves along a straight line without being acted on by forces of any kind. We'll then consider particles that are acted on by forces and are trapped in *bound states,* just as electrons are bound within an atom. We'll see that solving the Schrödinger equation automatically gives the possible energy levels for the system.

Besides energies, solving the Schrödinger equation gives us the probabilities of finding a particle in various regions. One surprising result is that there is a nonzero probability that microscopic particles will pass through thin barriers, even though such a process is forbidden by Newtonian mechanics.

In this chapter we'll consider the Schrödinger equation for one-dimensional motion only. In Chapter 41 we'll see how to extend this equation to three-dimensional problems such as the hydrogen atom. The hydrogen-atom wave functions will in turn form the foundation for our analysis of more complex atoms, of the periodic table of the elements, of x-ray energy levels, and of other properties of atoms.

40.1 WAVE FUNCTIONS AND THE ONE-DIMENSIONAL SCHRÖDINGER EQUATION

We have now seen compelling evidence that on an atomic or subatomic scale, an object such as an electron cannot be described simply as a classical, Newtonian point particle. Instead, we must take into account its *wave* characteristics. In the Bohr model of the hydrogen atom (Section 39.3) we tried to have it both ways: We

MP

PhET: Fourier: Making Waves

40.1 These children are talking over a cup-and-string "telephone." The displacement of the string is completely described by a wave function $y(x, t)$. In an analogous way, a particle is completely described by a quantum-mechanical wave function $\Psi(x, y, z, t)$.

CAUTION Particle waves vs. mechanical waves Unlike waves on a string or sound waves in air, the wave function for a particle is *not* a mechanical wave that needs a material medium in order to propagate. The wave function describes the particle, but we cannot define the function itself in terms of anything material. We can describe only how it is related to physically observable effects.

pictured the electron as a classical particle in a circular orbit around the nucleus, and used the de Broglie relationship between particle momentum and wavelength to explain why only orbits of certain radii are allowed. As we saw in Section 39.6, however, the Heisenberg uncertainty principle tells us that a hybrid description of this kind can't be wholly correct. In this section we'll explore how to describe the state of a particle by using *only* the language of waves. This new description, called **quantum mechanics,** replaces the classical scheme of describing the state of a particle by its coordinates and velocity components.

Our new quantum-mechanical scheme for describing a particle has a lot in common with the language of classical wave motion. In Section 15.3 of Chapter 15, we described transverse waves on a string by specifying the position of each point in the string at each instant of time by means of a *wave function* $y(x, t)$ that represents the displacement from equilibrium, at time t, of a point on the string at a distance x from the origin (**Fig. 40.1**). Once we know the wave function for a particular wave motion, we know everything there is to know about the motion. For example, we can find the velocity and acceleration of any point on the string at any time. We worked out specific forms for these functions for *sinusoidal* waves, in which each particle undergoes simple harmonic motion.

We followed a similar pattern for sound waves in Chapter 16. The wave function $p(x, t)$ for a wave traveling along the x-direction represented the pressure variation at any point x and any time t. In Section 32.3 we used *two* wave functions to describe the $\vec{E}$ and $\vec{B}$ fields in an electromagnetic wave.

Thus it's natural to use a wave function as the central element of our new language of quantum mechanics. The customary symbol for this wave function is the Greek letter psi, Ψ or ψ. In general, we'll use an uppercase Ψ to denote a function of all the space coordinates and time, and a lowercase ψ for a function of the space coordinates only—*not* of time. Just as the wave function $y(x, t)$ for mechanical waves on a string provides a complete description of the motion, so the wave function $\Psi(x, y, z, t)$ for a particle contains all the information that can be known about the particle.

Waves in One Dimension: Waves on a String

The wave function of a particle depends in general on all three dimensions of space. For simplicity, however, we'll begin our study of these functions by considering *one-dimensional* motion, in which a particle of mass m moves parallel to the x-axis and the wave function Ψ depends on the coordinate x and the time t only. (In the same way, we studied one-dimensional kinematics in Chapter 2 before going on to study two- and three-dimensional motion in Chapter 3.)

What does a one-dimensional quantum-mechanical wave look like, and what determines its properties? We can answer this question by first recalling the properties of a wave on a string. We saw in Section 15.3 that any wave function $y(x, t)$ that describes a wave on a string must satisfy the *wave equation:*

$$\frac{\partial^2 y(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x,t)}{\partial t^2} \quad \text{(wave equation for waves on a string)} \quad (40.1)$$

In Eq. (40.1) v is the speed of the wave, which is the same no matter what the wavelength. As an example, consider the following wave function for a wave of wavelength λ and frequency f moving in the positive x-direction along a string:

$$y(x,t) = A\cos(kx - \omega t) + B\sin(kx - \omega t) \quad \text{(sinusoidal wave on a string)} \quad (40.2)$$

Here $k = 2\pi/\lambda$ is the *wave number* and $\omega = 2\pi f$ is the *angular frequency.* (We used these same quantities for mechanical waves in Chapter 15 and electromagnetic waves in Chapter 32.) The quantities A and B are constants that determine the amplitude and phase of the wave. The expression in Eq. (40.2) is a valid wave function if and only if it satisfies the wave equation, Eq. (40.1). To check this, take the first and second derivatives of $y(x, t)$ with respect to x and take the first and second derivatives with respect to t:

$$\frac{\partial y(x, t)}{\partial x} = -kA\sin(kx - \omega t) + kB\cos(kx - \omega t) \tag{40.3a}$$

$$\frac{\partial^2 y(x, t)}{\partial x^2} = -k^2 A\cos(kx - \omega t) - k^2 B\sin(kx - \omega t) \tag{40.3b}$$

$$\frac{\partial y(x, t)}{\partial t} = \omega A\sin(kx - \omega t) - \omega B\cos(kx - \omega t) \tag{40.3c}$$

$$\frac{\partial^2 y(x, t)}{\partial t^2} = -\omega^2 A\cos(kx - \omega t) - \omega^2 B\sin(kx - \omega t) \tag{40.3d}$$

If we substitute Eqs. (40.3b) and (40.3d) into the wave equation, Eq. (40.1), we get

$$\begin{aligned} &-k^2 A\cos(kx - \omega t) - k^2 B\sin(kx - \omega t) \\ &= \frac{1}{v^2}\left[-\omega^2 A\cos(kx - \omega t) - \omega^2 B\sin(kx - \omega t)\right] \end{aligned} \tag{40.4}$$

For Eq. 40.4 to be satisfied at all coordinates x and all times t, the coefficients of $\cos(kx - \omega t)$ must be the same on both sides of the equation, and likewise for the coefficients of $\sin(kx - \omega t)$. Both of these conditions will be satisfied if

$$k^2 = \frac{\omega^2}{v^2} \quad \text{or} \quad \omega = vk \quad \text{(waves on a string)} \tag{40.5}$$

Since $\omega = 2\pi f$ and $k = 2\pi/\lambda$, Eq. (40.5) is equivalent to

$$2\pi f = v\frac{2\pi}{\lambda} \quad \text{or} \quad v = \lambda f \quad \text{(waves on a string)}$$

This equation is just the familiar relationship among wave speed, wavelength, and frequency for waves on a string. So our calculation shows that Eq. (40.2) is a valid wave function for waves on a string for any values of A and B, provided that ω and k are related by Eq. (40.5).

Waves in One Dimension: Particle Waves

What we need is a quantum-mechanical version of the wave equation, Eq. (40.1), valid for particle waves. We expect this equation to involve partial derivatives of the wave function $\Psi(x, t)$ with respect to x and with respect to t. However, this new equation *cannot* be the same as Eq. (40.1) for waves on a string because the relationship between ω and k is different. We can show this by considering a **free particle,** one that experiences no force at all as it moves along the x-axis. For such a particle the potential energy $U(x)$ has the same value for all x (recall from Chapter 7 that $F_x = -dU(x)/dx$, so zero force means the potential energy has zero derivative). For simplicity let $U = 0$ for all x. Then the energy of the free particle is equal to its kinetic energy, which we can express in terms of its momentum p:

$$E = \tfrac{1}{2}mv^2 = \frac{m^2v^2}{2m} = \frac{(mv)^2}{2m} = \frac{p^2}{2m} \quad \text{(energy of a free particle)} \tag{40.6}$$

The de Broglie relationships (Section 39.1) tell us that the energy E is proportional to the angular frequency ω and the momentum p is proportional to the wave number:

$$E = hf = \frac{h}{2\pi}2\pi f = \hbar\omega \tag{40.7a}$$

$$p = \frac{h}{\lambda} = \frac{h}{2\pi}\frac{2\pi}{\lambda} = \hbar k \tag{40.7b}$$

Remember that $\hbar = h/2\pi$. If we substitute Eqs. (40.7) into Eq. (40.6), we find that the relationship between ω and k for a free particle is

$$\hbar\omega = \frac{\hbar^2 k^2}{2m} \qquad \text{(free particle)} \tag{40.8}$$

Equation (40.8) is *very* different from the corresponding relationship for waves on a string, Eq. (40.5): The angular frequency ω for particle waves is proportional to the *square* of the wave number, while for waves on a string ω is directly proportional to k. Our task is therefore to construct a quantum-mechanical version of the wave equation whose free-particle solutions satisfy Eq. (40.8).

We'll attack this problem by assuming a sinusoidal wave function $\Psi(x, t)$ of the same form as Eq. (40.2) for a sinusoidal wave on a string. For a wave on a string, Eq. (40.2) represents a wave of wavelength $\lambda = 2\pi/k$ and frequency $f = \omega/2\pi$ propagating in the positive x-direction. By analogy, our sinsuoidal wave function $\Psi(x, t)$ represents a free particle of mass m, momentum $p = \hbar k$, and energy $E = \hbar\omega$ moving in the positive x-direction:

$$\Psi(x, t) = A\cos(kx - \omega t) + B\sin(kx - \omega t) \qquad \begin{array}{l}\text{(sinusoidal wave}\\ \text{function representing}\\ \text{a free particle)}\end{array} \tag{40.9}$$

The wave number k and angular frequency ω in Eq. (40.9) must satisfy Eq. (40.8). If you look at Eq. (40.3b), you'll see that taking the second derivative of $\Psi(x, t)$ in Eq. (40.9) with respect to x gives us $\Psi(x, t)$ multiplied by $-k^2$. Hence if we multiply $\partial^2\Psi(x, t)/\partial x^2$ by $-\hbar^2/2m$, we get

$$\begin{aligned} -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x, t)}{\partial x^2} &= -\frac{\hbar^2}{2m}\left[-k^2 A\cos(kx - \omega t) - k^2 B\sin(kx - \omega t)\right] \\ &= \frac{\hbar^2 k^2}{2m}\left[A\cos(kx - \omega t) + B\sin(kx - \omega t)\right] \\ &= \frac{\hbar^2 k^2}{2m}\Psi(x, t) \end{aligned} \tag{40.10}$$

Equation (40.10) suggests that $(-\hbar^2/2m)\partial^2\Psi(x, t)/\partial x^2$ should be one side of our quantum-mechanical wave equation, with the other side equal to $\hbar\omega\Psi(x, t)$ in order to satisfy Eq. (40.8). If you look at Eq. (40.3c), you'll see that taking the *first* time derivative of $\Psi(x, t)$ in Eq. (40.9) brings out a factor of ω. So we'll make the educated guess that the right-hand side of our quantum-mechanical wave equation involves $\hbar = h/2\pi$ times $\partial\Psi(x, t)/\partial t$. So our tentative equation is

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x, t)}{\partial x^2} = C\hbar\frac{\partial\Psi(x, t)}{\partial t} \tag{40.11}$$

At this point we include a constant C as a "fudge factor" to make sure that everything turns out right. Now let's substitute the wave function from Eq. (40.9) into Eq. (40.11). From Eq. (40.10) and Eq. (40.3c), we get

$$\begin{aligned} &\frac{\hbar^2 k^2}{2m}\left[A\cos(kx - \omega t) + B\sin(kx - \omega t)\right] \\ &\quad = C\hbar\omega\left[A\sin(kx - \omega t) - B\cos(kx - \omega t)\right] \end{aligned} \tag{40.12}$$

From Eq. (40.8), $\hbar\omega = \hbar^2 k^2/2m$, so we can cancel these factors on the two sides of Eq. (40.12). What remains is

$$\begin{aligned} &A\cos(kx - \omega t) + B\sin(kx - \omega t) \\ &\quad = CA\sin(kx - \omega t) - CB\cos(kx - \omega t) \end{aligned} \tag{40.13}$$

As in our discussion above of the wave equation for waves on a string, in order for Eq. (40.13) to be satisfied for all values of x and all values of t, the coefficients of $\cos(kx - \omega t)$ must be the same on both sides of the equation, and likewise for the coefficients of $\sin(kx - \omega t)$. Hence we have the following relationships among the coefficients A, B, and C in Eqs. (40.9) and (40.11):

$$A = -CB \tag{40.14a}$$

$$B = CA \tag{40.14b}$$

If we use Eq. (40.14b) to eliminate B from Eq. (40.14a), we get $A = -C^2A$, which means that $C^2 = -1$. Thus C is equal to the *imaginary* number $i = \sqrt{-1}$, and Eq. (40.11) becomes

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x, t)}{\partial x^2} = i\hbar\frac{\partial\Psi(x, t)}{\partial t} \quad \text{(one-dimensional Schrödinger equation for a free particle)} \tag{40.15}$$

Equation (40.15) is the one-dimensional **Schrödinger equation** for a free particle, developed in 1926 by the Austrian physicist Erwin Schrödinger (**Fig. 40.2**). The presence of the imaginary number i in Eq. (40.15) means that the solutions to the Schrödinger equation are complex quantities, with a real part and an imaginary part. (The imaginary part of $\Psi(x, t)$ is a real function multiplied by the imaginary number $i = \sqrt{-1}$.) An example is our free-particle wave function from Eq. (40.9). Since we found that $C = i$ in Eqs. (40.14), it follows from Eq. (40.14b) that $B = iA$. Then Eq. (40.9) becomes

$$\Psi(x, t) = A[\cos(kx - \omega t) + i\sin(kx - \omega t)] \quad \text{(sinusoidal wave function representing a free particle)} \tag{40.16}$$

The real part of $\Psi(x, t)$ is $\text{Re}\Psi(x, t) = A\cos(kx - \omega t)$ and the imaginary part is $\text{Im}\Psi(x, t) = A\sin(kx - \omega t)$. **Figure 40.3** graphs the real and imaginary parts of $\Psi(x, t)$ at $t = 0$, so $\Psi(x, 0) = A\cos kx + iA\sin kx$.

We can rewrite Eq. (40.16) with *Euler's formula,* which states that for any angle θ,

$$e^{i\theta} = \cos\theta + i\sin\theta$$
$$e^{-i\theta} = \cos(-\theta) + i\sin(-\theta) = \cos\theta - i\sin\theta \tag{40.17}$$

Thus our sinusoidal free-particle wave function becomes

$$\Psi(x, t) = Ae^{i(kx-\omega t)} = Ae^{ikx}e^{-i\omega t} \quad \text{(sinusoidal wave function representing a free particle)} \tag{40.18}$$

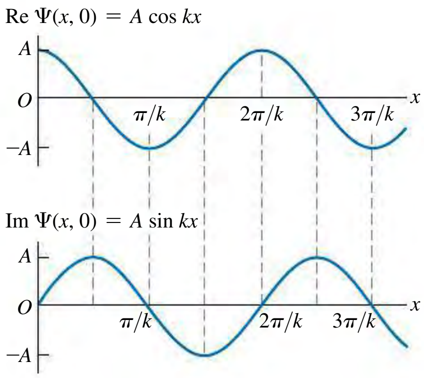
If k is positive in Eq. (40.16), the wave function represents a free particle moving in the positive x-direction with momentum $p = \hbar k$ and energy $E = \hbar\omega = \hbar^2k^2/2m$. If k is negative, the momentum and hence the motion are in the negative x-direction. (With a negative value of k, the wavelength is $\lambda = 2\pi/|k|$).

Interpreting the Wave Function

The complex nature of the wave function for a free particle makes this function challenging to interpret. (We certainly haven't needed imaginary numbers before this point to describe real physical phenomena.) Here's how to think about this function: $\Psi(x, t)$ describes the *distribution* of a particle in space, just as the wave functions for an electromagnetic wave describe the distribution of the electric and magnetic fields. When we worked out interference and diffraction patterns in Chapters 35 and 36, we found that the intensity I of the radiation at any point in a pattern is proportional to the square of the electric-field magnitude—that is, to E^2. In the photon interpretation of interference and diffraction (see Section 38.4), the intensity at each point is proportional to the number of photons striking around that point or, alternatively, to the *probability* that any individual photon will

40.2 Erwin Schrödinger (1887–1961) developed the equation that bears his name in 1926, an accomplishment for which he shared (with the British physicist P. A. M. Dirac) the 1933 Nobel Prize in physics. His grave marker is adorned with a version of Eq. (40.15).

40.3 The spatial wave function $\Psi(x, t) = Ae^{i(kx-\omega t)}$ for a free particle of definite momentum $p = \hbar k$ is a complex function: It has both a real part and an imaginary part. These are graphed here as functions of x for $t = 0$.



strike around the point. Thus the square of the electric-field magnitude at each point is proportional to the probability of finding a photon around that point.

In exactly the same way, the square of the wave function of a particle at each point tells us about the probability of finding the particle around that point. More precisely, we should say the square of the *absolute value* of the wave function, $|\Psi|^2$. This is necessary because, as we have seen, the wave function is a complex quantity with real and imaginary parts.

For a particle that can move only along the x-direction, the quantity $|\Psi(x, t)|^2 dx$ is the probability that the particle will be found at time t at a coordinate in the range from x to $x + dx$. The particle is most likely to be found in regions where $|\Psi|^2$ is large, and so on. This interpretation, first made by the German physicist Max Born (**Fig. 40.4**), requires that the wave function Ψ be *normalized*. That is, the integral of $|\Psi(x, t)|^2 dx$ over all possible values of x must equal exactly 1. In other words, the probability is exactly 1, or 100%, that the particle is *somewhere*.

40.4 In 1926, the German physicist Max Born (1882–1970) devised the interpretation that $|\Psi|^2$ is the probability distribution function for a particle that is described by the wave function Ψ. He also coined the term "quantum mechanics" (in the original German, *Quantenmechanik*). For his contributions, Born shared (with Walther Bothe) the 1954 Nobel Prize in physics.

CAUTION Interpreting $|\Psi|^2$ Note that $|\Psi(x, t)|^2$ itself is *not* a probability. Rather, $|\Psi(x, t)|^2 dx$ is the probability of finding the particle between position x and position $x + dx$ at time t. If the length dx is made smaller, it becomes less likely that the particle will be found within that length, so the probability decreases. A better name for $|\Psi(x, t)|^2$ is the **probability distribution function,** since it describes how the probability of finding the particle at different locations is distributed over space. Another common name for $|\Psi(x, t)|^2$ is the *probability density.*

We can use the probability interpretation of $|\Psi|^2$ to get a better understanding of Eq. (40.18), the wave function for a free particle. This function describes a particle that has a definite momentum $p = \hbar k$ in the x-direction and *no* uncertainty in momentum: $\Delta p_x = 0$. The Heisenberg uncertainty principle for position and momentum, Eqs. (39.29), says that $\Delta x \Delta p_x \geq \hbar/2$. If Δp_x is zero, then Δx must be infinite, and we have no idea where along the x-axis the particle can be found. (We saw a similar result for photons in Section 38.4.) We can show this by calculating the probability distribution function $|\Psi(x, t)|^2$: the product of Ψ and its *complex conjugate* Ψ^*. To find the complex conjugate of a complex number, we replace all i with $-i$. For example, the complex conjugate of $c = a + ib$, where a and b are real, is $c^* = a - ib$, so $|c|^2 = c^*c = (a + ib)(a - ib) = a^2 + b^2$ (recall that $i^2 = -1$). The complex conjugate of Eq. (40.18) is

$$\Psi^*(x, t) = A^* e^{-i(kx-\omega t)} = A^* e^{-ikx} e^{i\omega t}$$

(We have to allow for the possibility that the coefficient A is itself a complex number.) Hence the probability distribution function is

$$\begin{aligned}|\Psi(x, t)|^2 &= \Psi^*(x, t)\Psi(x, t) = (A^* e^{-ikx} e^{i\omega t})(A e^{ikx} e^{-i\omega t}) \\ &= A^* A e^0 = |A|^2\end{aligned}$$

The probability distribution function doesn't depend on position, which says that we are equally likely to find the particle *anywhere* along the x-axis! Mathematically, this is because the sinusoidal wave function $\Psi(x, t) = Ae^{i(kx-\omega t)} = A[\cos(kx - \omega t) + i\sin(kx - \omega t)]$ extends all the way from $x = -\infty$ to $x = +\infty$ with the same amplitude A. This also means that the wave function can't be normalized: The integral of $|\Psi(x, t)|^2$ over all space is infinite for any value of A.

Note also that the wave function in Eq. (40.18) describes a particle with a definite energy $E = \hbar\omega$, so there is zero uncertainty in energy: $\Delta E = 0$. The Heisenberg uncertainty principle for energy and time interval, $\Delta t \Delta E \geq \hbar/2$ [Eq. (39.30)], tells us that the time uncertainty Δt for this particle is infinite. In other words, we can have no idea *when* the particle will pass a given point on the x-axis. That also agrees with our result $|\Psi(x, t)|^2 = |A|^2$; the probability distribution function has the same value at all times.

Since we always have some idea of where a particle is, the wave function given in Eq. (40.18) isn't a realistic description. In our study of light in Section 38.4, we saw that we can make a wave function that's more *localized* in space by superimposing two or more sinusoidal functions. (This would be a good time to review that section.) As an illustration, let's calculate $|\Psi(x, t)|^2$ for a wave function of this kind.

EXAMPLE 40.1 A LOCALIZED FREE-PARTICLE WAVE FUNCTION

SOLUTION

The wave function $\Psi(x, t) = Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}$ is a superposition of *two* free-particle wave functions of the form given by Eq. (40.18). Both k_1 and k_2 are positive. (a) Show that this wave function satisfies the Schrödinger equation for a free particle of mass m. (b) Find the probability distribution function for $\Psi(x, t)$.

SOLUTION

IDENTIFY and SET UP: Both wave functions $Ae^{i(k_1x-\omega_1t)}$ and $Ae^{i(k_2x-\omega_2t)}$ represent a particle moving in the positive x-direction, but with different momenta and kinetic energies: $p_1 = \hbar k_1$ and $E_1 = \hbar\omega_1 = \hbar^2k_1^2/2m$ for the first function, $p_2 = \hbar k_2$ and $E_2 = \hbar\omega_2 = \hbar^2k_2^2/2m$ for the second function. To test whether a superposition of these is also a valid wave function for a free particle, we'll see whether our function $\Psi(x, t)$ satisfies the free-particle Schrödinger equation, Eq. (40.15). We'll use the derivatives of the exponential function: $(d/du)e^{au} = ae^{au}$ and $(d^2/du^2)e^{au} = a^2e^{au}$. The probability distribution function $|\Psi(x, t)|^2$ is the product of $\Psi(x, t)$ and its complex conjugate.

EXECUTE: (a) If we substitute $\Psi(x, t)$ into Eq. (40.15), the left-hand side of the equation is

$$
\begin{aligned}
-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} &= -\frac{\hbar^2}{2m}\frac{\partial^2(Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)})}{\partial x^2}\\
&= -\frac{\hbar^2}{2m}\left[(ik_1)^2Ae^{i(k_1x-\omega_1t)} + (ik_2)^2Ae^{i(k_2x-\omega_2t)}\right]\\
&= \frac{\hbar^2k_1^2}{2m}Ae^{i(k_1x-\omega_1t)} + \frac{\hbar^2k_2^2}{2m}Ae^{i(k_2x-\omega_2t)}
\end{aligned}
$$

The right-hand side is

$$
\begin{aligned}
i\hbar\frac{\partial\Psi(x,t)}{\partial t} &= i\hbar\frac{\partial(Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)})}{\partial t}\\
&= i\hbar\left[(-i\omega_1)Ae^{i(k_1x-\omega_1t)} + (-i\omega_2)Ae^{i(k_2x-\omega_2t)}\right]\\
&= \hbar\omega_1Ae^{i(k_1x-\omega_1t)} + \hbar\omega_2Ae^{i(k_2x-\omega_2t)}
\end{aligned}
$$

The two sides *are* equal, provided that $\hbar\omega_1 = \hbar^2k_1^2/2m$ and $\hbar\omega_2 = \hbar^2k_2^2/2m$. These are just the relationships that we noted above. So we conclude that $\Psi(x, t) = Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}$ is a valid free-particle wave function. In general, if we take any two wave functions that are solutions of the Schrödinger equation and then make a superposition of these to create a third wave function $\Psi(x, t)$, then $\Psi(x, t)$ is also a solution of the Schrödinger equation.

(b) The complex conjugate of $\Psi(x, t)$ is

$$\Psi^*(x, t) = A^*e^{-i(k_1x-\omega_1t)} + A^*e^{-i(k_2x-\omega_2t)}$$

Hence

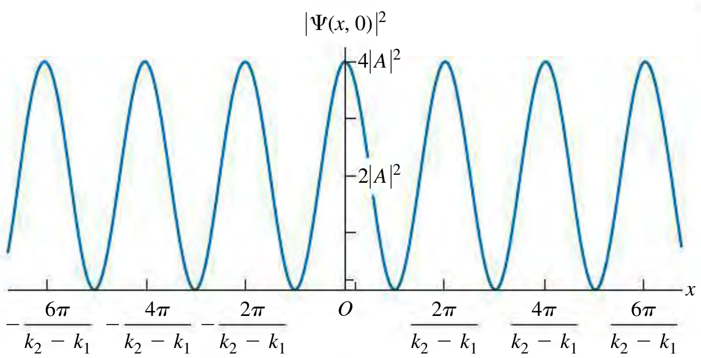
$$
\begin{aligned}
&|\Psi(x,t)|^2\\
&\quad = \Psi^*(x,t)\Psi(x,t)\\
&\quad = (A^*e^{-i(k_1x-\omega_1t)} + A^*e^{-i(k_2x-\omega_2t)})(Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)})\\
&\quad = A^*A\begin{bmatrix} e^{-i(k_1x-\omega_1t)}e^{i(k_1x-\omega_1t)} + e^{-i(k_2x-\omega_2t)}e^{i(k_2x-\omega_2t)}\\ + e^{-i(k_1x-\omega_1t)}e^{i(k_2x-\omega_2t)} + e^{-i(k_2x-\omega_2t)}e^{i(k_1x-\omega_1t)}\end{bmatrix}\\
&\quad = |A|^2\left[e^0 + e^0 + e^{i[(k_2-k_1)x-(\omega_2-\omega_1)t]} + e^{-i[(k_2-k_1)x-(\omega_2-\omega_1)t]}\right]
\end{aligned}
$$

To simplify this expression, recall that $e^0 = 1$. From Euler's formula, $e^{i\theta} = \cos\theta + i\sin\theta$ and $e^{-i\theta} = \cos\theta - i\sin\theta$, so $e^{i\theta} + e^{-i\theta} = 2\cos\theta$. Hence

$$
\begin{aligned}
|\Psi(x,t)|^2 &= |A|^2\{2 + 2\cos[(k_2-k_1)x - (\omega_2-\omega_1)t]\}\\
&= 2|A|^2\{1 + \cos[(k_2-k_1)x - (\omega_2-\omega_1)t]\}
\end{aligned}
$$

EVALUATE: **Figure 40.5** is a graph of the probability distribution function $|\Psi(x, t)|^2$ at $t = 0$. The value of $|\Psi(x, t)|^2$ varies between 0 and $4|A|^2$; probabilities can never be negative! The particle has become *somewhat* localized: The particle is most likely to be found near a point where $|\Psi(x, t)|^2$ is maximum (where the functions $Ae^{i(k_1x-\omega_1t)}$ and $Ae^{i(k_2x-\omega_2t)}$ interfere constructively) and is very unlikely to be found near a point where $|\Psi(x, t)|^2 = 0$ (where $Ae^{i(k_1x-\omega_1t)}$ and $Ae^{i(k_2x-\omega_2t)}$ interfere destructively).

40.5 The probability distribution function at $t = 0$ for $\Psi(x, t) = Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}$.



Continued

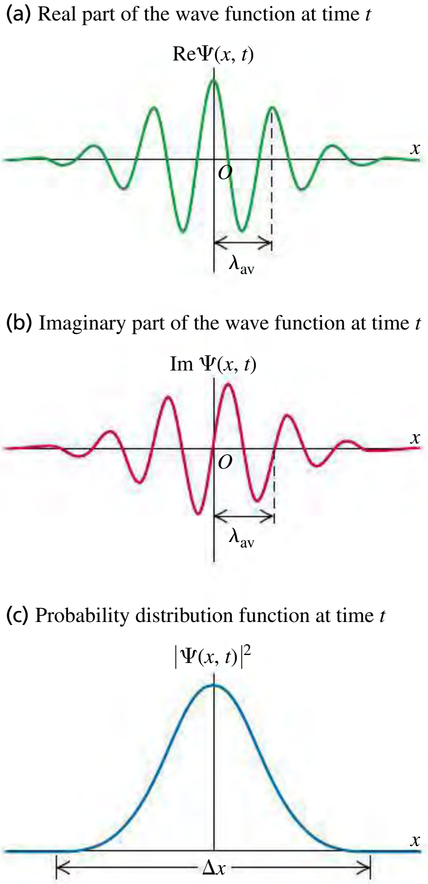
Note also that the probability distribution function is not stationary: It moves in the positive x-direction like the particle that it represents. To see this, recall from Section 15.3 that a sinusoidal wave given by $y(x, t) = A\cos(kx - \omega t)$ moves in the positive x-direction with speed $v = \omega/k$; since $|\Psi(x, t)|^2$ includes a term $\cos[(k_2 - k_1)x - (\omega_2 - \omega_1)t]$, the probability distribution moves at a speed $v_{av} = (\omega_2 - \omega_1)/(k_2 - k_1)$. The subscript "av" reminds us that v_{av} represents the *average* value of the particle's speed.

The price we pay for localizing the particle somewhat is that, unlike a particle represented by Eq. (40.18), it no longer has either a definite momentum or a definite energy. That's consistent with the Heisenberg uncertainty principles: If we decrease the uncertainties about where a particle is and when it passes a certain point, the uncertainties in its momentum and energy must increase.

The average momentum of the particle is $p_{av} = (\hbar k_2 + \hbar k_1)/2$, which is the average of the momenta associated with the free-particle wave functions we added to create $\Psi(x, t)$. This corresponds to the particle having an average speed $v_{av} = p_{av}/m = (\hbar k_2 + \hbar k_1)/2m$. Can you show that this is equal to the expression $v_{av} = (\omega_2 - \omega_1)/(k_2 - k_1)$ that we found above?

Wave Packets

40.6 Superposing a large number of sinusoidal waves with different wave numbers and appropriate amplitudes can produce a wave pulse that has a wavelength $\lambda_{av} = 2\pi/k_{av}$ and is localized within a region of space of length Δx. This localized pulse has aspects of both particle and wave.



The wave function that we examined in Example 40.1 is not very well localized: The probability distribution function still extends from $x = -\infty$ to $x = +\infty$. Hence this wave function can't be normalized. To make a wave function that's more highly localized, imagine superposing two additional sinusoidal waves with different wave numbers and amplitudes so as to reinforce alternate maxima of $|\Psi(x, t)|^2$ in Fig. 40.5 and cancel out the in-between ones. Finally, if we superpose waves with a very large number of different wave numbers, we can construct a wave with only *one* maximum of $|\Psi(x, t)|^2$ (**Fig. 40.6**). Then we have something that begins to look like both a particle and a wave. It is a particle in the sense that it is localized in space; if we look from a distance, it may look like a point. But it also has a periodic structure that is characteristic of a wave.

A localized wave pulse like that shown in Fig. 40.6 is called a **wave packet.** We can represent a wave packet by an expression such as

$$\Psi(x, t) = \int_{-\infty}^{\infty} A(k)e^{i(kx - \omega t)}\,dk \qquad (40.19)$$

This integral represents a superposition of a very large number of waves, each with a different wave number k and angular frequency $\omega = \hbar k^2/2m$, and each with an amplitude $A(k)$ that depends on k.

There is an important relationship between the two functions $\Psi(x, t)$ and $A(k)$, which we show qualitatively in **Fig. 40.7**. If the function $A(k)$ is sharply peaked, as in Fig. 40.7a, we are superposing only a narrow range of wave numbers. The resulting wave pulse is then relatively broad (Fig. 40.7b). But if we use a wider range of wave numbers, so that the function $A(k)$ is broader (Fig. 40.7c), then the wave pulse is more narrowly localized (Fig. 40.7d). This is simply the uncertainty principle in action. A narrow range of k means a narrow range of $p_x = \hbar k$ and thus a small Δp_x; the result is a relatively large Δx. A broad range of k corresponds to a large Δp_x, and the resulting Δx is smaller. You can see that the uncertainty principle for position and momentum, $\Delta x \Delta p_x \geq \hbar/2$, is really just a consequence of the properties of integrals like Eq. (40.19).

CAUTION Matter waves versus light waves in vacuum We can regard both a wave packet that represents a particle and a short pulse of light from a laser as superpositions of waves of different wave numbers and angular frequencies. An important difference is that the speed of light in vacuum is the same for all wavelengths λ and hence all wave numbers $k = 2\pi/\lambda$, but the speed of a matter wave is *different* for different wavelengths. You can see this from the formula for the speed of the wave crests in a periodic wave, $v = \lambda f = \omega/k$. For a matter wave, $\omega = \hbar k^2/2m$, so $v = \hbar k/2m = h/2m\lambda$. Hence matter waves with longer wavelengths and smaller wave numbers travel more slowly than those with short wavelengths and large wave numbers. (This shouldn't be too surprising. The de Broglie relationships that we learned in Section 39.1 tell us that shorter wavelength corresponds to greater momentum and hence a greater speed.) Since the individual sinusoidal waves that make up a wave packet travel at different speeds, the shape of the packet changes as it moves. That's why we've specified the time for which the wave packets in Figs. 40.6 and 40.7 are drawn; at later times, the packets become more spread out. By contrast, a pulse of light waves in vacuum retains the same shape at all times because all of its constituent sinusoidal waves travel together at the same speed.

The One-Dimensional Schrödinger Equation with Potential Energy

The one-dimensional Schrödinger equation that we presented in Eq. (40.15) is valid only for free particles, for which the potential energy function is zero: $U(x) = 0$. But for an electron within an atom, a proton within an atomic nucleus, and many other real situations, the potential energy plays an important role. To study the behavior of matter waves in these situations, we need a version of the Schrödinger equation that describes a particle moving in the presence of a nonzero potential energy function $U(x)$. This equation is

General one-dimensional Schrödinger equation:

Planck's constant divided by 2π Particle's wave function

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + U(x)\Psi(x,t) = i\hbar\frac{\partial\Psi(x,t)}{\partial t} \quad (40.20)$$

Particle's mass Potential-energy function

Note that if $U(x) = 0$, Eq. (40.20) reduces to the free-particle Schrödinger equation given in Eq. (40.15).

Here's the motivation behind Eq. (40.20). If $\Psi(x, t)$ is a sinusoidal wave function for a free particle, $\Psi(x,t) = Ae^{i(kx-\omega t)} = Ae^{ikx}e^{-i\omega t}$, the derivative terms in Eq. (40.20) become

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}(Ae^{ikx}e^{-i\omega t}) = -\frac{\hbar^2}{2m}(ik)^2(Ae^{ikx}e^{-i\omega t})$$
$$= \frac{\hbar^2k^2}{2m}\Psi(x,t)$$
$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = i\hbar\frac{\partial}{\partial t}(Ae^{ikx}e^{-i\omega t}) = i\hbar(-i\omega)(Ae^{ikx}e^{-i\omega t}) = \hbar\omega\Psi(x,t)$$

In these expressions $(\hbar^2k^2/2m)\Psi(x,t)$ is just the kinetic energy $K = p^2/2m = \hbar^2k^2/2m$ multiplied by the wave function, and $\hbar\omega\Psi(x,t)$ is the total energy $E = \hbar\omega$ multiplied by the wave function. So for a wave function of this kind, Eq. (40.20) says that kinetic energy times $\Psi(x,t)$ plus potential energy times $\Psi(x,t)$ equals total energy times $\Psi(x,t)$. That's equivalent to the statement in classical physics that the sum of kinetic energy and potential energy equals total mechanical energy: $K + U = E$.

The observations we've just made certainly aren't a *proof* that Eq. (40.20) is correct. The real reason we know this equation *is* correct is that it works: Predictions made with this equation agree with experimental results. Later in this chapter we'll apply Eq. (40.20) to several physical situations, each with a different form of the function $U(x)$.

Stationary States

We saw in our discussion of wave packets that any free-particle wave function can be built up as a superposition of sinusoidal wave functions of the form $\Psi(x,t) = Ae^{ikx}e^{-i\omega t}$. Each such sinusoidal wave function corresponds to a state of definite energy $E = \hbar\omega = \hbar^2k^2/2m$ and definite angular frequency $\omega = E/\hbar$, so we can rewrite these functions as $\Psi(x,t) = Ae^{ikx}e^{-iEt/\hbar}$. If the potential-energy function $U(x)$ is nonzero, these sinusoidal wave functions do not satisfy the Schrödinger equation, Eq. (40.20), and so these functions cannot be the basic "building blocks" of more complicated wave functions. However, we can still write the wave function for a state of definite energy E in the form

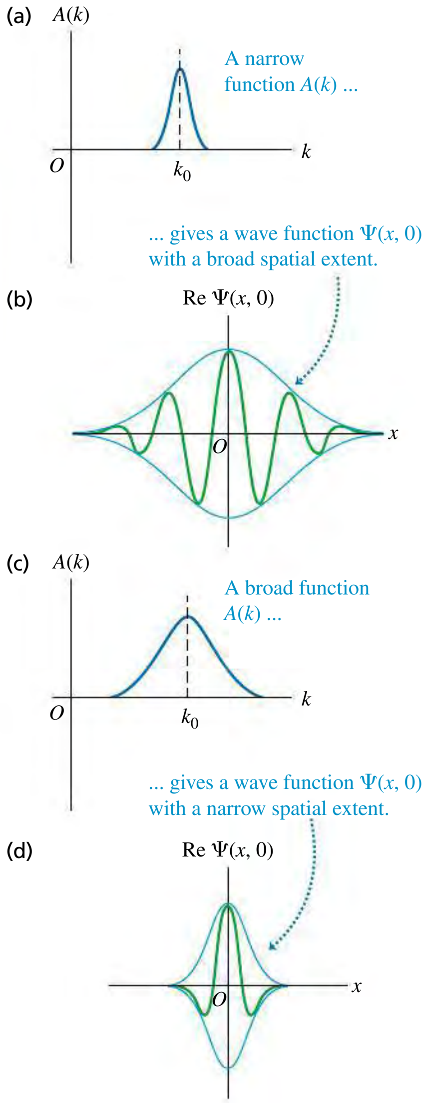
Time-dependent wave function for a state of definite energy

Time-independent wave function

$$\Psi(x,t) = \psi(x)e^{-iEt/\hbar} \quad (40.21)$$

Energy of state Planck's constant divided by 2π

40.7 How varying the function $A(k)$ in the wave-packet expression, Eq. (40.19), changes the character of the wave function $\Psi(x,t)$ (shown here at a specific time $t = 0$).



That is, the wave function $\Psi(x, t)$ for a state of definite energy is the product of a *time-independent* wave function $\psi(x)$ and a factor $e^{-iEt/\hbar}$. (For the free-particle sinusoidal wave function, $\psi(x) = Ae^{ikx}$.) States of definite energy are of tremendous importance in quantum mechanics. For example, for each energy level in a hydrogen atom (Section 39.3) there is a specific wave function. It is possible for an atom to be in a state that does not have a definite energy. The wave function for any such state can be written as a combination of definite-energy wave functions, in precisely the same way that a free-particle wave packet can be written as a superposition of sinusoidal wave functions of definite energy as in Eq. (40.19).

A state of definite energy is commonly called a **stationary state.** To see where this name comes from, let's multiply Eq. (40.21) by its complex conjugate to find the probability distribution function $|\Psi|^2$:

$$\begin{aligned}|\Psi(x, t)|^2 &= \Psi^*(x, t)\Psi(x, t) = [\psi^*(x)e^{+iEt/\hbar}][\psi(x)e^{-iEt/\hbar}] \\ &= \psi^*(x)\psi(x)e^{(+iEt/\hbar)+(-iEt/\hbar)} = |\psi(x)|^2 e^0 \\ &= |\psi(x)|^2\end{aligned} \tag{40.22}$$

PhET: Quantum Tunneling and Wave Packets

Since $|\psi(x)|^2$ does not depend on time, Eq. (40.22) shows that the same must be true for the probability distribution function $|\Psi(x, t)|^2$. This justifies the term "stationary state" for a state of definite energy.

CAUTION **A stationary state does not mean a stationary particle** Saying that a particle is in a stationary state does *not* mean that the particle is at rest. It's the *probability distribution* (that is, the relative likelihood of finding the particle at various positions), not the particle itself, that's stationary.

The Schrödinger equation, Eq. (40.20), becomes quite a bit simpler for stationary states. To see this, we substitute Eq. (40.21) into Eq. (40.20):

$$-\frac{\hbar^2}{2m}\frac{\partial^2[\psi(x)e^{-iEt/\hbar}]}{\partial x^2} + U(x)\psi(x)e^{-iEt/\hbar} = i\hbar\frac{\partial[\psi(x)e^{-iEt/\hbar}]}{\partial t}$$

The derivative in the first term on the left-hand side is with respect to x, so the factor $e^{-iEt/\hbar}$ comes outside of the derivative. Now we take the derivative with respect to t on the right-hand side of the equation:

$$\begin{aligned}-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2}e^{-iEt/\hbar} + U(x)\psi(x)e^{-iEt/\hbar} &= i\hbar\left(\frac{-iE}{\hbar}\right)[\psi(x)e^{-iEt/\hbar}] \\ &= E\psi(x)e^{-iEt/\hbar}\end{aligned}$$

If we divide both sides of this equation by $e^{-iEt/\hbar}$, we get

Time-independent one-dimensional Schrödinger equation:

Planck's constant divided by 2π — Time-independent wave function

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x) \tag{40.23}$$

Particle's mass — Potential-energy function — Energy of state

This is called the **time-independent one-dimensional Schrödinger equation.** The time-dependent factor $e^{-iEt/\hbar}$ does not appear, and Eq. (40.23) involves only the time-independent wave function $\psi(x)$. We'll devote much of this chapter to solving this equation to find the definite-energy, stationary-state wave functions $\psi(x)$ and the corresponding values of E—that is, the energies of the allowed levels—for different physical situations.

EXAMPLE 40.2 A STATIONARY STATE

Consider the wave function $\psi(x) = A_1 e^{ikx} + A_2 e^{-ikx}$, where k is positive. Is this a valid time-independent wave function for a free particle in a stationary state? What is the energy corresponding to this wave function?

SOLUTION

IDENTIFY and SET UP: A valid stationary-state wave function for a free particle must satisfy the time-independent Schrödinger equation, Eq. (40.23), with $U(x) = 0$. To test the given function $\psi(x)$, we simply substitute it into the left-hand side of the equation. If the result is a constant times $\psi(x)$, then the wave function is indeed a solution and the constant is equal to the particle energy E.

EXECUTE: Substituting $\psi(x) = A_1 e^{ikx} + A_2 e^{-ikx}$ and $U(x) = 0$ into Eq. (40.23), we obtain

$$
\begin{aligned}
-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} &= -\frac{\hbar^2}{2m}\frac{d^2(A_1 e^{ikx} + A_2 e^{-ikx})}{dx^2} \\
&= -\frac{\hbar^2}{2m}\left[(ik)^2 A_1 e^{ikx} + (-ik)^2 A_2 e^{-ikx}\right] \\
&= \frac{\hbar^2 k^2}{2m}(A_1 e^{ikx} + A_2 e^{-ikx}) = \frac{\hbar^2 k^2}{2m}\psi(x)
\end{aligned}
$$

The result is a constant times $\psi(x)$, so this $\psi(x)$ is indeed a valid stationary-state wave function for a free particle. Comparing with Eq. (40.23) shows that the constant on the right-hand side is the particle energy: $E = \hbar^2 k^2/2m$.

EVALUATE: Note that $\psi(x)$ is a *superposition* of two different wave functions: one function $(A_1 e^{ikx})$ that represents a particle with magnitude of momentum $p = \hbar k$ moving in the positive x-direction, and one function $(A_2 e^{-ikx})$ that represents a particle with the same magnitude of momentum moving in the negative x-direction. So while the combined wave function $\psi(x)$ represents a stationary state with a definite energy, this state does *not* have a definite momentum. We'll see in Section 40.2 that such a wave function can represent a *standing wave,* and we'll explore situations in which such standing matter waves can arise.

TEST YOUR UNDERSTANDING OF SECTION 40.1 Does a wave packet given by Eq. (40.19) represent a stationary state? ❙

40.2 PARTICLE IN A BOX

An important problem in quantum mechanics is how to use the time-independent Schrödinger equation, Eq. (40.23), to determine the possible energy levels and the corresponding wave functions for various systems. That is, for a given potential energy function $U(x)$, what are the possible stationary-state wave functions $\psi(x)$, and what are the corresponding energies E?

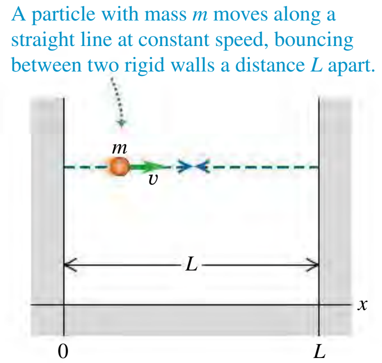
In Section 40.1 we solved this problem for the case $U(x) = 0$, corresponding to a *free* particle. The allowed wave functions and corresponding energies are

$$\psi(x) = Ae^{ikx} \qquad E = \frac{\hbar^2 k^2}{2m} \qquad \text{(free particle)} \tag{40.24}$$

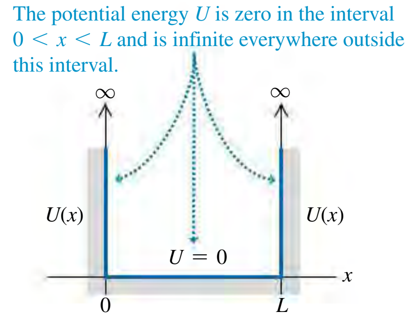
The wave number k is equal to $2\pi/\lambda$, where λ is the wavelength. We found that k can have any real value, so the energy E of a free particle can have any value from zero to infinity. Furthermore, the particle can be found with equal probability at any value of x from $-\infty$ to $+\infty$.

Now let's look at a simple model in which a particle is *bound* so that it cannot escape to infinity, but rather is confined to a restricted region of space. Our system consists of a particle confined between two rigid walls separated by a distance L (**Fig. 40.8**). The motion is purely one dimensional, with the particle moving along the x-axis only and the walls at $x = 0$ and $x = L$. The potential energy corresponding to the rigid walls is infinite, so the particle cannot escape; between the walls, the potential energy is zero (**Fig. 40.9**). This situation is often described as a **"particle in a box."** This model might represent an electron that is free to move within a long, straight molecule or along a very thin wire.

40.8 The Newtonian view of a particle in a box.



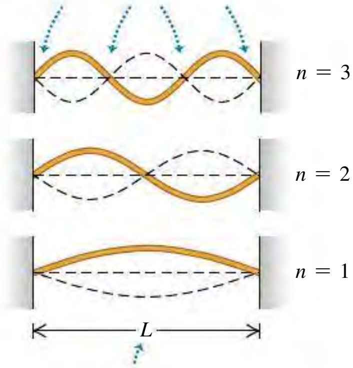
40.9 The potential-energy function for a particle in a box.



Wave Functions for a Particle in a Box

To solve the Schrödinger equation for this system, we begin with some restrictions on the particle's stationary-state wave function $\psi(x)$. Because the particle is confined to the region $0 \le x \le L$, we expect the probability distribution function $|\Psi(x, t)|^2 = |\psi(x)|^2$ and the wave function $\psi(x)$ to be zero outside that region. This agrees with the Schrödinger equation: If the term $U(x)\psi(x)$ in Eq. (40.23) is to be finite, then $\psi(x)$ must be zero where $U(x)$ is infinite.

Furthermore, $\psi(x)$ must be a *continuous* function to be a mathematically well-behaved solution to the Schrödinger equation. This implies that $\psi(x)$ must be zero at the region's boundary, $x = 0$ and $x = L$. These two conditions serve as *boundary conditions* for the problem. They should look familiar, because they are the same conditions that we used to find the normal modes of a vibrating string in Section 15.8 (**Fig. 40.10**); you should review that discussion.



40.10 Normal modes of vibration for a string with length L, held at both ends.

An additional condition is that to calculate the second derivative $d^2\psi(x)/dx^2$ in Eq. (40.23), the *first* derivative $d\psi(x)/dx$ must also be continuous except at points where the potential energy becomes infinite (as it does at the walls of the box). This is analogous to the requirement that a vibrating string, like those shown in Fig. 40.10, can't have any kinks in it (which would correspond to a discontinuity in the first derivative of the wave function) except at the ends of the string.

We now solve for the wave functions in the region $0 \le x \le L$ subject to the above conditions. In this region $U(x) = 0$, so $\psi(x)$ in this region must satisfy

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x) \qquad \text{(particle in a box)} \qquad (40.25)$$

Equation (40.25) is the *same* Schrödinger equation as for a free particle, so it is tempting to conclude that the wave functions and energies are given by Eq. (40.24). It is true that $\psi(x) = Ae^{ikx}$ satisfies the Schrödinger equation with $U(x) = 0$, is continuous, and has a continuous first derivative $d\psi(x)/dx = ikAe^{ikx}$. However, this wave function does *not* satisfy the boundary conditions that $\psi(x)$ must be zero at $x = 0$ and $x = L$: At $x = 0$ the wave function in Eq. (40.24) is equal to $Ae^0 = A$, and at $x = L$ it is equal to Ae^{ikL}. (These would be equal to zero if $A = 0$, but then the wave function would be zero and there would be no particle at all!)

The way out of this dilemma is to recall Example 40.2 (Section 40.1), in which we found that a more general stationary-state solution to the time-independent Schrödinger equation with $U(x) = 0$ is

$$\psi(x) = A_1e^{ikx} + A_2e^{-ikx} \qquad (40.26)$$

This wave function is a superposition of two waves: one traveling in the $+x$-direction of amplitude A_1, and one traveling in the $-x$-direction with the same wave number but amplitude A_2. This is analogous to a standing wave on a string (Fig. 40.10), which we can regard as the superposition of two sinusoidal waves propagating in opposite directions (see Section 15.7). The energy that corresponds to Eq. (40.26) is $E = \hbar^2k^2/2m$, just as for a single wave.

To see whether the wave function given by Eq. (40.26) can satisfy the boundary conditions, let's first rewrite it in terms of sines and cosines by using Euler's formula, Eq. (40.17):

$$\begin{aligned}\psi(x) &= A_1(\cos kx + i\sin kx) + A_2[\cos(-kx) + i\sin(-kx)]\\ &= A_1(\cos kx + i\sin kx) + A_2(\cos kx - i\sin kx)\\ &= (A_1 + A_2)\cos kx + i(A_1 - A_2)\sin kx\end{aligned} \qquad (40.27)$$

At $x = 0$ this is equal to $\psi(0) = A_1 + A_2$, which must equal zero to satisfy the boundary condition at that point. Hence $A_2 = -A_1$, and Eq. (40.27) becomes

$$\psi(x) = 2iA_1\sin kx = C\sin kx \qquad (40.28)$$

We have simplified the expression by introducing the constant $C = 2iA_1$. (We'll come back to this constant later.) We can also satisfy the second boundary condition that $\psi = 0$ at $x = L$ by choosing values of k such that $kL = n\pi$ $(n = 1, 2, 3, \dots)$. Hence Eq. (40.28) does indeed give the stationary-state wave functions for a particle in a box in the region $0 \le x \le L$. (Outside this region, $\psi(x) = 0$.) The possible values of k and the wavelength $\lambda = 2\pi/k$ are

$$k = \frac{n\pi}{L} \quad \text{and} \quad \lambda = \frac{2\pi}{k} = \frac{2L}{n} \quad (n = 1, 2, 3, \dots) \tag{40.29}$$

Just as for the string in Fig. 40.10, the length L of the region is an integral number of half-wavelengths.

Energy Levels for a Particle in a Box

The possible energy levels for a particle in a box are given by $E = \hbar^2k^2/2m = p^2/2m$, where $p = \hbar k = (h/2\pi)(2\pi/\lambda) = h/\lambda$ is the magnitude of momentum of a free particle with wave number k and wavelength λ. This makes sense, since inside the region $0 \le x \le L$ the potential energy is zero and the energy is all kinetic. For each value of n, there are corresponding values of p, λ, and E; let's call them p_n, λ_n, and E_n. Putting the pieces together, we get

$$p_n = \frac{h}{\lambda_n} = \frac{nh}{2L} \tag{40.30}$$

and so the energy levels for a particle in a box are

Magnitude of momentum; Planck's constant; Planck's constant divided by 2π

Energy levels for a particle in a box

$$E_n = \frac{p_n^2}{2m} = \frac{n^2h^2}{8mL^2} = \frac{n^2\pi^2\hbar^2}{2mL^2} \quad (n = 1, 2, 3, \dots) \tag{40.31}$$

Particle's mass; Width of box; Quantum number

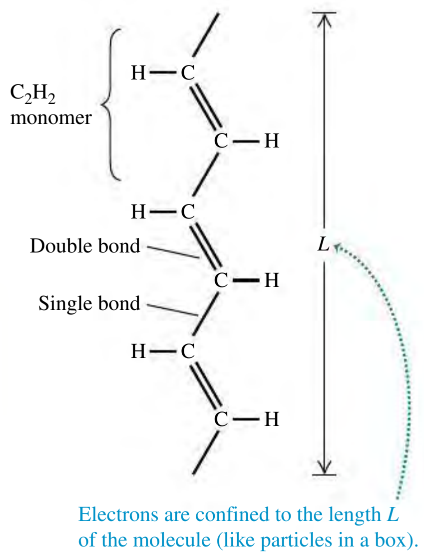
Each energy level has its own value of the quantum number n and a corresponding wave function, which we denote by ψ_n. When we replace k in Eq. (40.28) by $n\pi/L$ from Eq. (40.29), we find

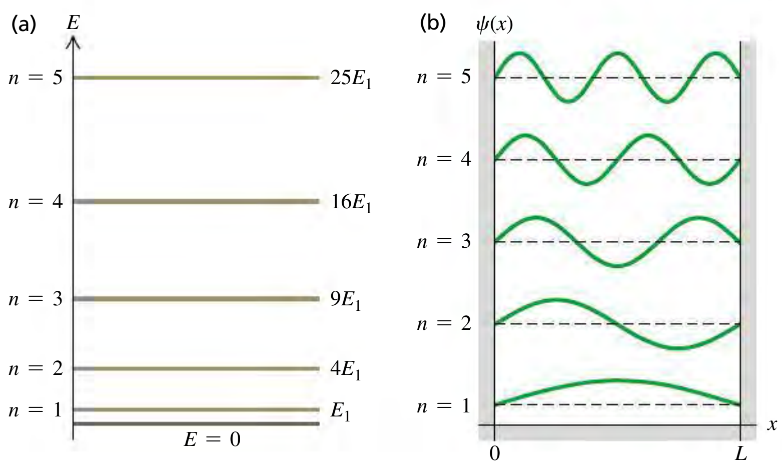
$$\psi_n(x) = C \sin\frac{n\pi x}{L} \quad (n = 1, 2, 3, \dots) \tag{40.32}$$

The energy-level diagram in **Fig. 40.11a** shows the five lowest levels for a particle in a box. The energy levels are proportional to n^2, so successively higher levels are spaced farther and farther apart. There are an infinite number of levels because the walls are perfectly rigid; even a particle of infinitely great kinetic

Application Particles in a Polymer "Box" Polyacetylene is one of a class of long-chain organic molecules that conduct electricity along their length. The molecule is made up of a large number of (C_2H_2) units, called *monomers* (only three monomers are shown here). Electrons can move freely along the length of the molecule but not perpendicular to the length, so the molecule is like a one-dimensional "box" for electrons. The length L of the molecule depends on the number of monomers. Experiment shows that the allowed energy levels agree well with Eq. (40.31): The greater the number of monomers and the greater the length L, the lower the energy levels and the smaller the spacing between these levels.

Polyacetylene





40.11 (a) Energy-level diagram for a particle in a box. Each energy is n^2E_1, where E_1 is the ground-level energy. (b) Wave functions for a particle in a box, with $n = 1, 2, 3, 4$, and 5. **CAUTION:** The five graphs have been displaced vertically for clarity, as in Fig. 40.10. Each of the horizontal dashed lines represents $\psi = 0$ for the respective wave function.

energy is confined within the box. Figure 40.11b shows graphs of the wave functions $\psi_n(x)$ for $n = 1, 2, 3, 4$, and 5. Note that these functions look identical to those for a standing wave on a string (see Fig. 40.10).

CAUTION **A particle in a box cannot have zero energy** The energy of a particle in a box *cannot* be zero. Equation (40.31) shows that $E = 0$ would require $n = 0$, but substituting $n = 0$ into Eq. (40.32) gives a zero wave function. Since a particle is described by a *nonzero* wave function, there cannot be a particle with $E = 0$. This is a consequence of the Heisenberg uncertainty principle: A particle in a zero-energy state would have a definite value of momentum (precisely zero), so its position uncertainty would be infinite and the particle could be found anywhere along the x-axis. But this is impossible, since a particle in a box can be found only between $x = 0$ and $x = L$. Hence $E = 0$ is not allowed. By contrast, the allowed stationary-state wave functions with $n = 1, 2, 3, \ldots$ do not represent states of definite momentum (each is an equal mixture of a state of x-momentum $+p_n = nh/2L$ and a state of x-momentum $-p_n = -nh/2L$). Hence each stationary state has a nonzero momentum uncertainty, consistent with having a finite position uncertainty.

EXAMPLE 40.3 ELECTRON IN AN ATOM-SIZE BOX

SOLUTION

Find the first two energy levels for an electron confined to a one-dimensional box 5.0×10^{-10} m across (about the diameter of an atom).

SOLUTION

IDENTIFY AND SET UP: This problem uses what we have learned in this section about a particle in a box. The first two energy levels correspond to $n = 1$ and $n = 2$ in Eq. (40.31).

EXECUTE: From Eq. (40.31),

$$E_1 = \frac{h^2}{8mL^2} = \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})^2}{8(9.109 \times 10^{-31}\ \mathrm{kg})(5.0 \times 10^{-10}\ \mathrm{m})^2}$$
$$= 2.4 \times 10^{-19}\ \mathrm{J} = 1.5\ \mathrm{eV}$$
$$E_2 = \frac{2^2h^2}{8mL^2} = 4E_1 = 9.6 \times 10^{-19}\ \mathrm{J} = 6.0\ \mathrm{eV}$$

EVALUATE: The difference between the first two energy levels is $E_2 - E_1 = 4.5$ eV. An electron confined to a box is different from an electron bound in an atom, but it is reassuring that this result is of the same order of magnitude as the difference between actual atomic energy levels.

You can show that for a proton or neutron ($m = 1.67 \times 10^{-27}$ kg) confined to a box 1.1×10^{-14} m across (the width of a medium-sized atomic nucleus), the energies of the first two levels are about a million times larger: $E_1 = 1.7 \times 10^6$ eV $= 1.7$ MeV, $E_2 = 4E_1 = 6.8$ MeV, $E_2 - E_1 = 5.1$ MeV. This suggests why nuclear reactions (which involve transitions between energy levels in nuclei) release so much more energy than chemical reactions (which involve transitions between energy levels of electrons in atoms).

Finally, you can show (see Exercise 40.9) that the energy levels of a billiard ball ($m = 0.2$ kg) confined to a box 1.3 m across—the width of a billiard table—are separated by about 5×10^{-67} J. Quantum effects won't disturb a game of billiards.

Probability and Normalization

Let's look a bit more closely at the wave functions for a particle in a box, keeping in mind the *probability* interpretation of the wave function ψ that we discussed in Section 40.1. In our one-dimensional situation the quantity $|\psi(x)|^2\,dx$ is proportional to the probability that the particle will be found within a small interval dx about x. For a particle in a box,

$$|\psi(x)|^2\,dx = C^2 \sin^2 \frac{n\pi x}{L}\,dx$$

Figure 40.12 shows graphs of both $\psi(x)$ and $|\psi(x)|^2$ for $n = 1$, 2, and 3. Note that not all positions are equally likely. By contrast, in classical mechanics the particle is equally likely to be found at any position between $x = 0$ and $x = L$. We see from Fig. 40.12b that $|\psi(x)|^2 = 0$ at some points, so there is zero probability of finding the particle at exactly these points. Don't let that bother you; the uncertainty principle has already shown us that we can't measure position exactly. The particle is localized only to be somewhere between $x = 0$ and $x = L$.

The particle must be *somewhere* on the x-axis—that is, somewhere between $x = -\infty$ and $x = +\infty$. So the *sum* of the probabilities for all the dx's everywhere (the *total* probability of finding the particle) must equal 1. That's the normalization condition that we discussed in Section 40.1:

Normalization condition, time-independent wave function:

Integral over all x

Time-independent wave function

$$\int_{-\infty}^{\infty} |\psi(x)|^2\, dx = 1 \qquad (40.33)$$

Probability distribution function

Probability that particle is *somewhere* on x-axis

A wave function is said to be *normalized* if it has a constant such as C in Eq. (40.32) that is calculated to make the total probability equal 1 in Eq. (40.33). For a normalized wave function, $|\psi(x)|^2\,dx$ is not merely proportional to, but *equals,* the probability of finding the particle between the coordinates x and $x + dx$. That's why we call $|\psi(x)|^2$ the probability distribution function. (In Section 40.1 we called $|\Psi(x, t)|^2$ the probability distribution function. For the case of a stationary-state wave function, however, $|\Psi(x, t)|^2$ is equal to $|\psi(x)|^2$.)

Let's normalize the particle-in-a-box wave functions $\psi_n(x)$ given by Eq. (40.32). Since $\psi_n(x)$ is zero except between $x = 0$ and $x = L$, Eq. (40.33) becomes

$$\int_0^L C^2 \sin^2 \frac{n\pi x}{L}\, dx = 1 \qquad (40.34)$$

You can evaluate this integral by using the trigonometric identity $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$; the result is $C^2L/2$. Thus our probability interpretation of the wave function demands that $C^2L/2 = 1$, or $C = (2/L)^{1/2}$; the constant C is *not* arbitrary. (This is in contrast to the classical vibrating string problem, in which C represents an amplitude that depends on initial conditions.) Thus the normalized stationary-state wave functions for a particle in a box are

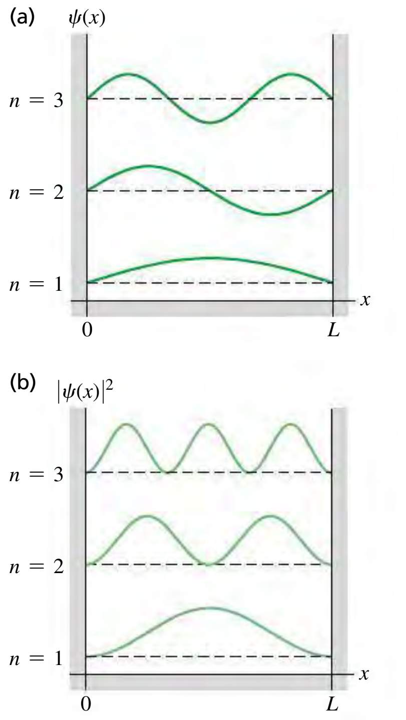
Stationary-state wave functions for a particle in a box

Quantum number

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \qquad (n = 1, 2, 3, \ldots) \qquad (40.35)$$

Width of box

40.12 Graphs of **(a)** $\psi(x)$ and **(b)** $|\psi(x)|^2$ for the first three wave functions ($n = 1, 2, 3$) for a particle in a box. The horizontal dashed lines represent $\psi(x) = 0$ and $|\psi(x)|^2 = 0$ for each of the three levels. The value of $|\psi(x)|^2 dx$ at each point is the probability of finding the particle in a small interval dx about the point. As in Fig. 40.11b, the three graphs in each part have been displaced vertically for clarity.



EXAMPLE 40.4 A NONSINUSOIDAL WAVE FUNCTION?

SOLUTION

(a) Show that $\psi(x) = Ax + B$, where A and B are constants, is a solution of the Schrödinger equation for an $E = 0$ energy level of a particle in a box. (b) What constraints do the boundary conditions at $x = 0$ and $x = L$ place on the constants A and B?

SOLUTION

IDENTIFY and SET UP: To be physically reasonable, a wave function must satisfy both the Schrödinger equation and the appropriate boundary conditions. In part (a) we'll substitute $\psi(x)$ into the Schrödinger equation for a particle in a box, Eq. (40.25), to determine whether it is a solution. In part (b) we'll see what restrictions on $\psi(x)$ arise from applying the boundary conditions that $\psi(x) = 0$ at $x = 0$ and $x = L$.

EXECUTE: (a) From Eq. (40.25), the Schrödinger equation for an $E = 0$ energy level of a particle in a box is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x) = 0$$

in the region $0 \le x \le L$. Differentiating $\psi(x) = Ax + B$ twice with respect to x gives $d^2\psi(x)/dx^2 = 0$, so the left side of the equation is zero, and so $\psi(x) = Ax + B$ is a solution of this Schrödinger equation for $E = 0$. (Note that both $\psi(x)$ and its derivative $d\psi(x)/dx = A$ are continuous functions, as they must be.)

(b) Applying the boundary condition at $x = 0$ gives $\psi(0) = B = 0$, and so $\psi(x) = Ax$. Applying the boundary condition at $x = L$ gives $\psi(L) = AL = 0$, so $A = 0$. Hence $\psi(x) = 0$ both inside the box ($0 \le x \le L$) *and outside:* There is *zero* probability of finding the particle anywhere with this wave function, and so $\psi(x) = Ax + B$ is *not* a physically valid wave function.

EVALUATE: The moral is that there are many functions that satisfy the Schrödinger equation for a given physical situation, but most of these—including the function considered here—have to be rejected because they don't satisfy the appropriate boundary conditions.

Time Dependence

The wave functions $\psi_n(x)$ in Eq. (40.35) depend only on the *spatial* coordinate x. Equation (40.21) shows that if $\psi(x)$ is the wave function for a state of definite energy E, the full time-dependent wave function is $\Psi(x, t) = \psi(x)e^{-iEt/\hbar}$. Hence the *time-dependent* stationary-state wave functions for a particle in a box are

$$\Psi_n(x, t) = \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x}{L}\right)e^{-iE_nt/\hbar} \qquad (n = 1, 2, 3, \dots) \qquad (40.36)$$

In this expression the energies E_n are given by Eq. (40.31). The higher the quantum number n, the greater the angular frequency $\omega_n = E_n/\hbar$ at which the wave function oscillates. Note that since $|e^{-iE_nt/\hbar}|^2 = e^{+iE_nt/\hbar}e^{-iE_nt/\hbar} = e^0 = 1$, the probability distribution function $|\Psi_n(x, t)|^2 = (2/L)\sin^2(n\pi x/L)$ is independent of time and does *not* oscillate. (Remember, this is why we say that these states of definite energy are *stationary*.)

TEST YOUR UNDERSTANDING OF SECTION 40.2 If a particle in a box is in the nth energy level, what is the average value of its x-component of momentum p_x? (i) $nh/2L$; (ii) $(\sqrt{2}/2)nh/L$; (iii) $(1/\sqrt{2})nh/L$; (iv) $[1/(2\sqrt{2})]nh/L$; (v) zero.

40.3 POTENTIAL WELLS

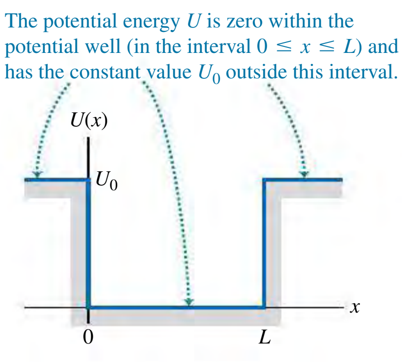
MP

PhET: Double Wells and Covalent Bonds
PhET: Quantum Bound States

A **potential well** is a potential-energy function $U(x)$ that has a minimum. We introduced this term in Section 7.5, and we also used it in our discussion of periodic motion in Chapter 14. In Newtonian mechanics a particle trapped in a potential well can vibrate back and forth with periodic motion. Our first application of the Schrödinger equation, the particle in a box, involved a rudimentary potential well with a function $U(x)$ that is zero within a certain interval and infinite everywhere else. As we mentioned in Section 40.2, this function corresponds to a few situations found in nature, but the correspondence is only approximate.

A better approximation to several physical situations is a **finite well,** which is a potential well that has straight sides but *finite* height. **Figure 40.13** shows a potential-energy function that is zero in the interval $0 \le x \le L$ and has the value U_0 outside this interval. This function is often called a **square-well potential.** It could serve as a simple model of an electron within a metallic sheet with thickness L, moving perpendicular to the surfaces of the sheet. The electron can move freely inside the metal but has to climb a potential-energy barrier with height U_0 to escape from either surface of the metal. The energy U_0 is related to the *work function* that we discussed in Section 38.1 in connection with the photoelectric effect. In three dimensions, a spherical version of a finite well gives an approximate description of the motions of protons and neutrons within a nucleus.

40.13 A square-well potential.



Bound States of a Square-Well Potential

In Newtonian mechanics, the particle is trapped (localized) in a well if the total mechanical energy E is less than U_0. In quantum mechanics, such a trapped state is often called a **bound state.** All states are bound for an infinitely deep well like the one we described in Section 40.2. For a finite well like that shown in Fig. 40.13, if E is greater than U_0, the particle is *not* bound.

Let's see how to solve the Schrödinger equation for the bound states of a square-well potential. Our goal is to find the energies and wave functions for which $E < U_0$. It's easiest to consider separately the regions where $U = 0$ and where $U = U_0$. Inside the square well $(0 \le x \le L)$, where $U = 0$, the time-independent Schrödinger equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x) \quad \text{or} \quad \frac{d^2\psi(x)}{dx^2} = -\frac{2mE}{\hbar^2}\psi(x) \qquad (40.37)$$

This is the same as Eq. (40.25) from Section 40.2, which describes a particle in a box. As in Section 40.2, we can express the solutions of this equation as combinations of cos kx and sin kx, where $E = \hbar^2k^2/2m$. We can rewrite the relationship between E and k as $k = \sqrt{2mE}/\hbar$. Hence inside the square well we have

$$\psi(x) = A\cos\left(\frac{\sqrt{2mE}}{\hbar}x\right) + B\sin\left(\frac{\sqrt{2mE}}{\hbar}x\right) \quad \text{(inside the well)} \qquad (40.38)$$

where A and B are constants. So far, this looks a lot like the particle-in-a-box analysis in Section 40.2. The difference is that for the square-well potential, the potential energy outside the well is not infinite, so the wave function $\psi(x)$ outside the well is *not* zero.

For the regions outside the well ($x < 0$ and $x > L$) the potential-energy function in the time-independent Schrödinger equation is $U = U_0$:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U_0\psi(x) = E\psi(x) \quad \text{or} \quad \frac{d^2\psi(x)}{dx^2} = \frac{2m(U_0 - E)}{\hbar^2}\psi(x) \qquad (40.39)$$

The quantity $U_0 - E$ is positive, so the solutions of this equation are exponential functions rather than sines or cosines. Using κ (the Greek letter kappa) to represent the quantity $[2m(U_0 - E)]^{1/2}/\hbar$ and taking κ as positive, we can write the solutions as

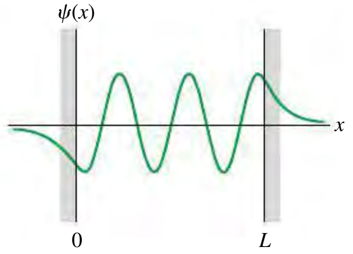
$$\psi(x) = Ce^{\kappa x} + De^{-\kappa x} \qquad \text{(outside the well)} \qquad (40.40)$$

where C and D are constants with different values in the two regions $x < 0$ and $x > L$. Note that ψ can't be allowed to approach infinity as $x \rightarrow +\infty$ or $x \rightarrow -\infty$. [If it did, we wouldn't be able to satisfy the normalization condition, Eq. (40.33).] This means that in Eq. (40.40), we must have $D = 0$ for $x < 0$ and $C = 0$ for $x > L$.

Our calculations so far show that the bound-state wave functions for a finite well are sinusoidal inside the well [Eq. (40.38)] and exponential outside it [Eq. (40.40)]. We have to *match* the wave functions inside and outside the well so that they satisfy the boundary conditions that we mentioned in Section 40.2: $\psi(x)$ and $d\psi(x)/dx$ must be continuous at the boundary points $x = 0$ and $x = L$. If the wave function $\psi(x)$ or the slope $d\psi(x)/dx$ were to change discontinuously at a point, the second derivative $d^2\psi(x)/dx^2$ would be *infinite* at that point. That would violate the time-independent Schrödinger equation, Eq. (40.23), which says that at every point $d^2\psi(x)/dx^2$ is proportional to $U - E$. For a finite well $U - E$ is finite everywhere, so $d^2\psi(x)/dx^2$ must also be finite everywhere.

Matching the sinusoidal and exponential functions at the boundary points so that they join smoothly is possible only for certain specific values of the total energy E, so this requirement determines the possible energy levels of the finite square well. There is no simple formula for the energy levels as there was for the infinitely deep well. Finding the levels is a fairly complex mathematical problem that requires solving a transcendental equation by numerical approximation; we won't go into the details. **Figure 40.14** shows the general shape of a possible wave function. The most striking features of this wave function are the "exponential tails" that extend outside the well into regions that are forbidden by Newtonian mechanics (because in those regions the particle would have negative kinetic energy). We see that there is some probability for finding the particle *outside* the potential well, which would be impossible in classical mechanics. In Section 40.4 we'll discuss an amazing result of this effect.

40.14 A possible wave function for a particle in a finite potential well. The function is sinusoidal inside the well $(0 \le x \le L)$ and exponential outside it. It approaches zero asymptotically at large $|x|$. The functions must join smoothly at $x = 0$ and $x = L$; the wave function and its derivative must be continuous.



EXAMPLE 40.5 OUTSIDE A FINITE WELL

SOLUTION

(a) Show that Eq. (40.40), $\psi(x) = Ce^{\kappa x} + De^{-\kappa x}$, is indeed a solution of the time-independent Schrödinger equation outside a finite well of height U_0. (b) What happens to $\psi(x)$ in the limit $U_0 \to \infty$?

SOLUTION

IDENTIFY and SET UP: In part (a), we try the given function $\psi(x)$ in the time-independent Schrödinger equation for $x < 0$ and for $x > L$, Eq. (40.39). In part (b), we note that in the limit $U_0 \to \infty$ the finite well becomes an *infinite* well, like that for a particle in a box (Section 40.2). So in this limit the wave functions outside a finite well must reduce to the wave functions outside the box.

EXECUTE: (a) We must show that $\psi(x) = Ce^{\kappa x} + De^{-\kappa x}$ satisfies $d^2\psi(x)/dx^2 = [2m(U_0 - E)/\hbar^2]\psi(x)$. We recall that $(d/du)e^{au} = ae^{au}$ and $(d^2/du^2)e^{au} = a^2e^{au}$; the left-hand side of the Schrödinger equation is then

$$\begin{aligned}\frac{d^2\psi(x)}{dx^2} &= \frac{d^2}{dx^2}(Ce^{\kappa x}) + \frac{d^2}{dx^2}(De^{-\kappa x})\\ &= C\kappa^2e^{\kappa x} + D(-\kappa)^2e^{-\kappa x}\\ &= \kappa^2(Ce^{\kappa x} + De^{-\kappa x}) = \kappa^2\psi(x)\end{aligned}$$

Since from Eq. (40.40) $\kappa^2 = 2m(U_0 - E)/\hbar^2$, this is equal to the right-hand side of the equation. The equation is satisfied, and $\psi(x)$ is a solution.

(b) As U_0 approaches infinity, κ also approaches infinity. In the region $x < 0$, $\psi(x) = Ce^{\kappa x}$; as $\kappa \to \infty$, $\kappa x \to -\infty$ (since x is negative) and $e^{\kappa x} \to 0$, so the wave function approaches zero for all $x < 0$. Likewise, we can show that the wave function also approaches zero for all $x > L$. This is just what we found in Section 40.2; the wave function for a particle in a box must be zero outside the box.

EVALUATE: Our result in part (b) shows that the infinite square well is a *limiting case* of the finite well. We've seen many cases in Newtonian mechanics where it's important to consider limiting cases (such as Examples 5.11 and 5.13 in Section 5.2). Limiting cases are no less important in quantum mechanics.

DATA *SPEAKS*

The Square-Well Potential

When students were given a problem involving a particle in a square-well potential, more than 41% gave an incorrect response. Common errors:

- Confusion about energy levels. The energy of a particle in a well is given relative to the *bottom* of the well (taken to be $E = 0$), not the *top* of the well (see Fig. 40.13). If the depth of the well is U_0, then for a bound state, $E < U_0$.
- Confusion about wave functions. The narrower the width L of the well, the farther outside the well the "exponential tails" of the wave function of a bound state extend.

Comparing Finite and Infinite Square Wells

Let's continue the comparison of the finite-depth potential well with the infinitely deep well, which we began in Example 40.5. First, because the wave functions for the finite well don't go to zero at $x = 0$ and $x = L$, the wavelength of the sinusoidal part of each wave function is *longer* than it would be with an infinite well. This increase in λ corresponds to a reduced magnitude of momentum $p = h/\lambda$ and therefore a reduced energy. Thus each energy level, including the ground level, is *lower* for a finite well than for an infinitely deep well with the same width.

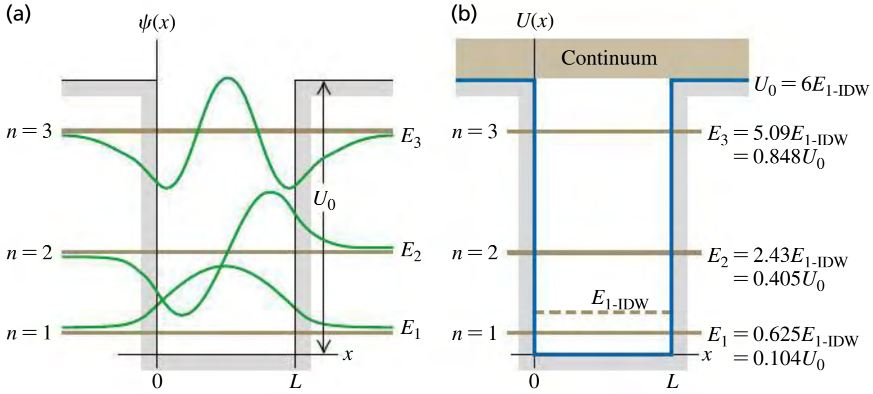
Second, a well with finite depth U_0 has only a *finite* number of bound states and corresponding energy levels, compared to the *infinite* number for an infinitely deep well. How many levels there are depends on the magnitude of U_0 in comparison with the ground-level energy for the infinitely deep well (IDW), which we call $E_{1-\mathrm{IDW}}$. From Eq. (40.31),

$$E_{1-\mathrm{IDW}} = \frac{\pi^2\hbar^2}{2mL^2} \qquad \text{(ground-level energy, infinitely deep well)} \qquad (40.41)$$

When the well is very deep so U_0 is much larger than $E_{1-\mathrm{IDW}}$, there are many bound states and the energies of the lowest few are nearly the same as the energies for the infinitely deep well. When U_0 is only a few times as large as $E_{1-\mathrm{IDW}}$ there are only a few bound states. (There is always at least *one* bound state, no matter how shallow the well.) As with the infinitely deep well, there is no state with $E = 0$; such a state would violate the uncertainty principle.

Figure 40.15 shows the case $U_0 = 6E_{1-\mathrm{IDW}}$, for which there are three bound states. In the figure, we express the energy levels both as fractions of the well depth U_0 and as multiples of $E_{1-\mathrm{IDW}}$. If the well were infinitely deep, the lowest three levels, as given by Eq. (40.31), would be $E_{1-\mathrm{IDW}}$, $4E_{1-\mathrm{IDW}}$, and $9E_{1-\mathrm{IDW}}$. Figure 40.15 also shows the wave functions for the three bound states.

40.15 (a) Wave functions for the three bound states for a particle in a finite potential well of depth $U_0 = 6E_{1-\text{IDW}}$. (Here $E_{1-\text{IDW}}$ is the ground-level energy for an infinite well of the same width.) The horizontal brown line for each wave function corresponds to $\psi = 0$; the vertical placement of these lines indicates the energy of each bound state (compare Fig. 40.11). (b) Energy-level diagram for this system. All energies greater than U_0 are possible; states with $E > U_0$ form a continuum.



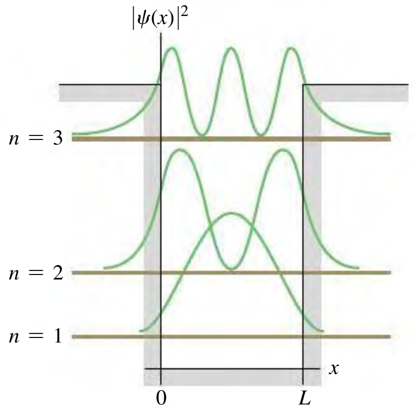
It turns out that when U_0 is less than $E_{1-\text{IDW}}$, there is only one bound state. In the limit when U_0 is *much smaller* than $E_{1-\text{IDW}}$ (a very shallow well), the energy of this single state is approximately $E = 0.68U_0$.

Figure 40.16 shows graphs of the probability distributions—that is, the values of $|\psi|^2$—for the wave functions shown in Fig. 40.15a. As with the infinite well, not all positions are equally likely. Unlike the infinite well, there is some probability of finding the particle outside the well in the classically forbidden regions.

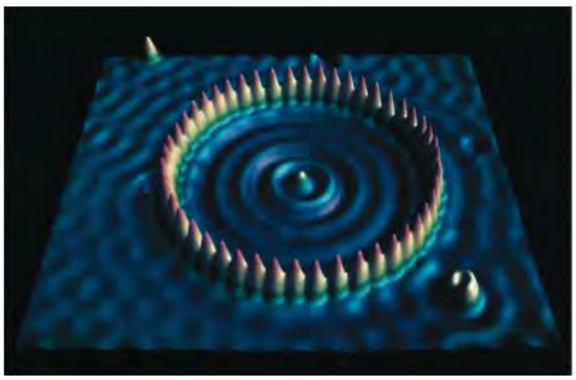
There are also states for which E is *greater* than U_0. In these *free-particle states* the particle is not bound but is free to move through all values of x. *Any* energy E greater than U_0 is possible, so the free-particle states form a *continuum* rather than a discrete set of states with definite energy levels. The free-particle wave functions are sinusoidal both inside and outside the well. The wavelength is shorter inside the well than outside, corresponding to greater kinetic energy inside the well than outside it.

Figure 40.17 shows particles in a *two*-dimensional finite potential well. Example 40.6 describes another application of the square-well potential.

40.16 Probability distribution functions $|\psi(x)|^2$ for the square-well wave functions shown in Fig. 40.15. The horizontal brown line for each wave function corresponds to $|\psi|^2 = 0$.



40.17 To make this image, 48 iron atoms (shown as yellow peaks) were placed in a circle on a copper surface. The "elevation" at each point inside the circle indicates the electron density within the circle. The standing-wave pattern is very similar to the probability distribution function for a particle in a one-dimensional finite potential well. (This image was made with a scanning tunneling microscope, discussed in Section 40.4.)



SOLUTION

EXAMPLE 40.6 AN ELECTRON IN A FINITE WELL

An electron is trapped in a square well 0.50 nm across (roughly five times a typical atomic diameter). (a) Find the ground-level energy $E_{1-\text{IDW}}$ if the well is infinitely deep. (b) Find the energy levels if the actual well depth U_0 is six times the ground-level energy found in part (a). (c) Find the wavelength of the photon emitted when the electron makes a transition from the $n = 2$ level to the $n = 1$ level. In what region of the electromagnetic spectrum does the photon wavelength lie? (d) If the electron is in the $n = 1$ (ground) level and absorbs a photon, what is the minimum photon energy that will free the electron from the well? In what region of the spectrum does the wavelength of this photon lie?

SOLUTION

IDENTIFY and SET UP: Equation (40.41) gives the ground-level energy $E_{1-\text{IDW}}$ for an infinitely deep well, and Fig. 40.15b shows the energies for a square well with $U_0 = 6E_{1-\text{IDW}}$. The energy of the photon emitted or absorbed in a transition is equal to the difference in energy between two levels involved in the transition; the photon wavelength is given by $E = hc/\lambda$ (see Chapter 38).

EXECUTE: (a) From Eq. (40.41),

$$E_{1-\text{IDW}} = \frac{\pi^2\hbar^2}{2mL^2} = \frac{\pi^2(1.055\times10^{-34}\ \text{J}\cdot\text{s})^2}{2(9.11\times10^{-31}\ \text{kg})(0.50\times10^{-9}\ \text{m})^2}$$
$$= 2.4\times10^{-19}\ \text{J} = 1.5\ \text{eV}$$

(b) We have $U_0 = 6E_{1-\text{IDW}} = 6(1.5\ \text{eV}) = 9.0\ \text{eV}$. We can read off the energy levels from Fig. 40.15b:

$$E_1 = 0.625E_{1-\text{IDW}} = 0.625(1.5\ \text{eV}) = 0.94\ \text{eV}$$
$$E_2 = 2.43E_{1-\text{IDW}} = 2.43(1.5\ \text{eV}) = 3.6\ \text{eV}$$
$$E_3 = 5.09E_{1-\text{IDW}} = 5.09(1.5\ \text{eV}) = 7.6\ \text{eV}$$

(c) The photon energy and wavelength for the $n = 2$ to $n = 1$ transition are

$$E_2 - E_1 = 3.6\ \text{eV} - 0.94\ \text{eV} = 2.7\ \text{eV}$$
$$\lambda = \frac{hc}{E} = \frac{(4.136\times10^{-15}\ \text{eV}\cdot\text{s})(3.00\times10^8\ \text{m/s})}{2.7\ \text{eV}}$$
$$= 460\ \text{nm}$$

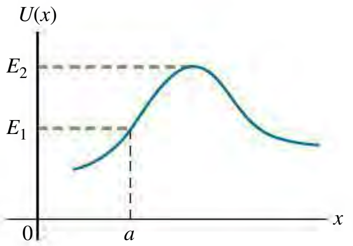
in the blue region of the visible spectrum.

(d) We see from Fig. 40.15b that the minimum energy needed to free the electron from the well from the $n = 1$ level is $U_0 - E_1 = 9.0\ \text{eV} - 0.94\ \text{eV} = 8.1\ \text{eV}$, which is three times the 2.7-eV photon energy found in part (c). Hence the corresponding photon wavelength is one-third of 460 nm, or (to two significant figures) 150 nm, which is in the ultraviolet region of the spectrum.

EVALUATE: As a check, you can calculate the bound-state energies by using the formulas $E_1 = 0.104U_0$, $E_2 = 0.405U_0$, and $E_3 = 0.848U_0$ given in Fig. 40.15b. As an additional check, note that the first three energy levels of an infinitely deep well of the same width are $E_{1-\text{IDW}} = 1.5\ \text{eV}$, $E_{2-\text{IDW}} = 4E_{1-\text{IDW}} = 6.0\ \text{eV}$, and $E_{3-\text{IDW}} = 9E_{1-\text{IDW}} = 13.5\ \text{eV}$. The energies we found in part (b) are less than these values: As we mentioned earlier, the finite depth of the well lowers the energy levels compared to the levels for an infinitely deep well.

One application of these ideas is *quantum dots,* which are nanometer-sized particles of a semiconductor such as cadmium selenide (CdSe). An electron within a quantum dot behaves much like a particle in a finite potential well of width L equal to the size of the dot. When quantum dots are illuminated with ultraviolet light, the electrons absorb the ultraviolet photons and are excited into high energy levels, such as the $n = 3$ level described in this example. If the electron returns to the ground level ($n = 1$) in two or more steps (for example, from $n = 3$ to $n = 2$ and from $n = 2$ to $n = 1$), one step will involve emitting a visible-light photon, as we have calculated here. (We described this process of *fluorescence* in Section 39.3.) Increasing the value of L decreases the energies of the levels and hence the spacing between them, and thus decreases the energy and increases the wavelength of the emitted photons. The photograph that opens this chapter shows quantum dots of different sizes in solution: Each emits a characteristic wavelength that depends on the dot size. Quantum dots can be injected into living tissue and their fluorescent glow used as a tracer for biological research and for medicine. They may also be the key to a new generation of lasers and ultrafast computers.

40.18 A potential-energy barrier. According to Newtonian mechanics, if the total energy of the system is E_1, a particle to the left of the barrier can go no farther than $x = a$. If the total energy is greater than E_2, the particle can pass over the barrier.



TEST YOUR UNDERSTANDING OF SECTION 40.3 Suppose that the width of the finite potential well shown in Fig. 40.15 is reduced by one-half. How must the value of U_0 change so that there are still just three bound energy levels whose energies are the fractions of U_0 shown in Fig. 40.15b? U_0 must (i) increase by a factor of four; (ii) increase by a factor of two; (iii) remain the same; (iv) decrease by a factor of one-half; (v) decrease by a factor of one-fourth.

40.4 POTENTIAL BARRIERS AND TUNNELING

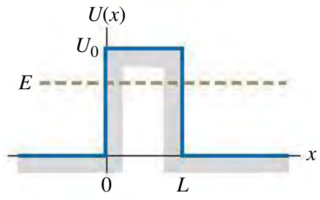
A **potential barrier** is the opposite of a potential well; it is a potential-energy function with a *maximum*. **Figure 40.18** shows an example. In classical Newtonian mechanics, if a particle (such as a roller coaster) is located to the left of the barrier (which might be a hill), and if the total mechanical energy of the system is E_1, the particle cannot move farther to the right than $x = a$. If it did, the potential energy U would be greater than the total energy E and the kinetic energy $K = E - U$ would be negative. This is impossible in classical mechanics since $K = \frac{1}{2}mv^2$ can never be negative.

A quantum-mechanical particle behaves differently: If it encounters a barrier like the one in Fig. 40.18 and has energy less than E_2, it *may* appear on the other side. This phenomenon is called *tunneling.* In quantum-mechanical tunneling, unlike macroscopic, mechanical tunneling, the particle does not actually push through the barrier and loses no energy in the process.

Tunneling Through a Rectangular Barrier

To understand how tunneling can occur, let's look at the potential-energy function $U(x)$ shown in **Fig. 40.19**. It's like Fig. 40.13 turned upside-down; the potential energy is zero everywhere except in the range $0 \le x \le L$, where it has the value U_0. This might represent a simple model for the potential energy of an electron in the presence of two slabs of metal separated by an air gap of thickness L. The potential energy is lower within either slab than in the gap between them.

40.19 A rectangular potential-energy barrier with width L and height U_0. According to Newtonian mechanics, if the total energy E is less than U_0, a particle cannot pass over this barrier but is confined to the side where it starts.



Let's consider solutions of the Schrödinger equation for this potential-energy function for the case in which E is less than U_0. We can use our results from Section 40.3. In the regions $x < 0$ and $x > L$, where $U = 0$, the solution is sinusoidal and is given by Eq. (40.38). Within the barrier $(0 \le x \le L)$, $U = U_0$ and the solution is exponential as in Eq. (40.40). Just as with the finite potential well, the functions have to join smoothly at the boundary points $x = 0$ and $x = L$, which means that both $\psi(x)$ and $d\psi(x)/dx$ have to be continuous at these points.

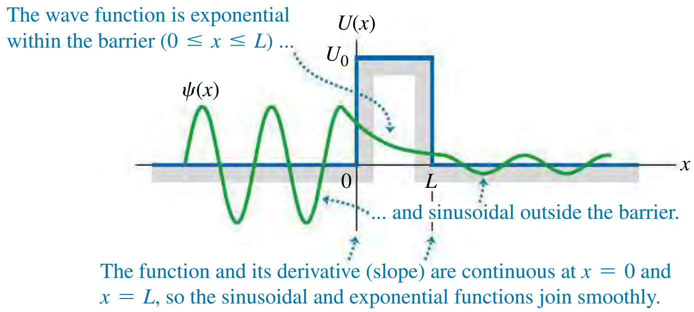
These requirements lead to a wave function like the one shown in **Fig. 40.20**. The function is *not* zero inside the barrier (the region forbidden by Newtonian mechanics). Even more remarkable, a particle that is initially to the *left* of the barrier has some probability of being found to the *right* of the barrier. How great this probability is depends on the width L of the barrier and the particle's energy E in comparison with the barrier height U_0. The **tunneling probability** T that the particle gets through the barrier is proportional to the square of the ratio of the amplitudes of the sinusoidal wave functions on the two sides of the barrier. These amplitudes are determined by matching wave functions and their derivatives at the boundary points, a fairly involved mathematical problem. When T is much smaller than unity, it is given approximately by

PhET: Quantum Tunneling and Wave Packets

$$T = Ge^{-2\kappa L} \text{ where } G = 16\frac{E}{U_0}\left(1 - \frac{E}{U_0}\right) \text{ and } \kappa = \frac{\sqrt{2m(U_0 - E)}}{\hbar} \qquad (40.42)$$

(probability of tunneling)

The probability decreases rapidly with increasing barrier width L. It also depends critically on the energy difference $U_0 - E$, which in Newtonian physics is the additional kinetic energy the particle would need to be able to climb over the barrier.



40.20 A possible wave function for a particle tunneling through the potential-energy barrier shown in Fig. 40.19.

EXAMPLE 40.7 TUNNELING THROUGH A BARRIER

A 2.0-eV electron encounters a barrier 5.0 eV high. What is the probability that it will tunnel through the barrier if the barrier width is (a) 1.00 nm and (b) 0.50 nm?

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of tunneling through a rectangular barrier, as in Figs. 40.19 and 40.20. Our target variable is the tunneling probability T in Eq. (40.42), which we evaluate for the given values $E = 2.0\text{ eV}$ (electron energy), $U = 5.0\text{ eV}$ (barrier height), $m = 9.11 \times 10^{-31}\text{ kg}$ (mass of the electron), and $L = 1.00\text{ nm}$ or 0.50 nm (barrier width).

EXECUTE: First we evaluate G and κ in Eq. (40.42), using $E = 2.0\text{ eV}$:

$$G = 16\left(\frac{2.0\text{ eV}}{5.0\text{ eV}}\right)\left(1 - \frac{2.0\text{ eV}}{5.0\text{ eV}}\right) = 3.8$$

$$U_0 - E = 5.0\text{ eV} - 2.0\text{ eV} = 3.0\text{ eV} = 4.8 \times 10^{-19}\text{ J}$$

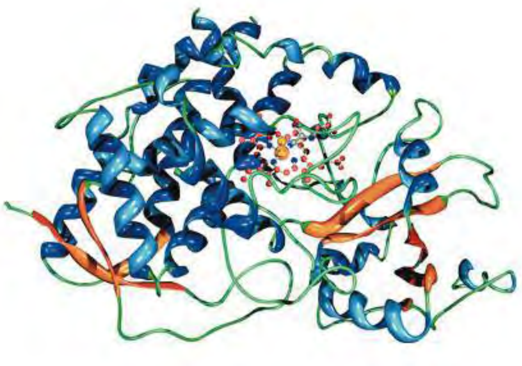
$$\kappa = \frac{\sqrt{2(9.11 \times 10^{-31}\text{ kg})(4.8 \times 10^{-19}\text{ J})}}{1.055 \times 10^{-34}\text{ J}\cdot\text{s}} = 8.9 \times 10^{9}\text{ m}^{-1}$$

(a) When $L = 1.00\text{ nm} = 1.00 \times 10^{-9}\text{ m}$, we have $2\kappa L = 2(8.9 \times 10^{9}\text{ m}^{-1})(1.00 \times 10^{-9}\text{ m}) = 17.8$ and $T = Ge^{-2\kappa L} = 3.8e^{-17.8} = 7.1 \times 10^{-8}$.

(b) When $L = 0.50\text{ nm}$, one-half of 1.00 nm, $2\kappa L$ is one-half of 17.8, or 8.9. Hence $T = 3.8e^{-8.9} = 5.2 \times 10^{-4}$.

EVALUATE: Halving the width of this barrier increases the tunneling probability T by a factor of $(5.2 \times 10^{-4})/(7.1 \times 10^{-8}) = 7.3 \times 10^{3}$, or nearly ten thousand. The tunneling probability is an *extremely* sensitive function of the barrier width.

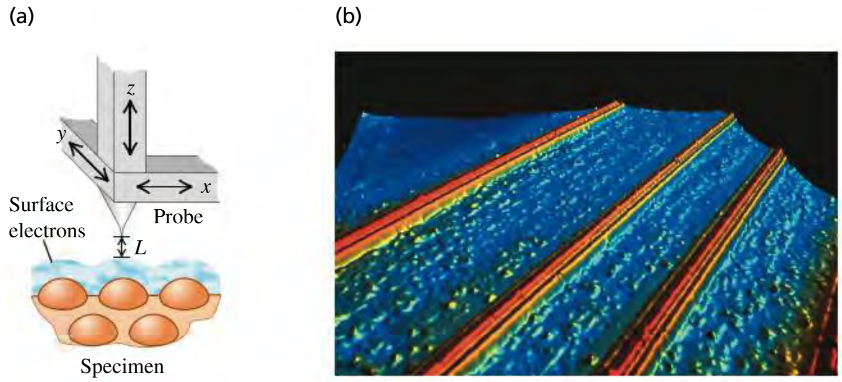
BIO Application Electron Tunneling in Enzymes Protein molecules play essential roles as enzymes in living organisms. Enzymes like the one shown here are large molecules, and in many cases their function depends on the ability of electrons to tunnel across the space that separates one part of the molecule from another. Without tunneling, life as we know it would be impossible!



Applications of Tunneling

Tunneling has a number of practical applications, some of considerable importance. When you twist two copper wires together or close the contacts of a switch, current passes from one conductor to the other despite a thin layer of nonconducting copper oxide between them. The electrons tunnel through this thin insulating layer. A *tunnel diode* is a semiconductor device in which electrons tunnel through a potential barrier. The current can be switched on and off very quickly (within a few picoseconds) by varying the height of the barrier. A *Josephson junction* consists of two superconductors separated by an oxide layer a few atoms (1 to 2 nm) thick. Electron pairs in the superconductors can tunnel through the barrier layer, giving such a device unusual circuit properties. Josephson junctions are useful for establishing precise voltage standards and measuring tiny magnetic fields, and they play a crucial role in the developing field of quantum computing.

The *scanning tunneling microscope* (STM) uses electron tunneling to create images of surfaces down to the scale of individual atoms. An extremely sharp conducting needle is brought very close to the surface, within 1 nm or so (**Fig. 40.21a**). When the needle is at a positive potential with respect to the surface, electrons can tunnel through the surface potential-energy barrier and reach the needle. As Example 40.7 shows, the tunneling probability and hence the tunneling current are very sensitive to changes in the width L of the barrier (the distance between the surface and the needle tip). In one mode of operation the needle is scanned across the surface while being moved perpendicular to the surface to

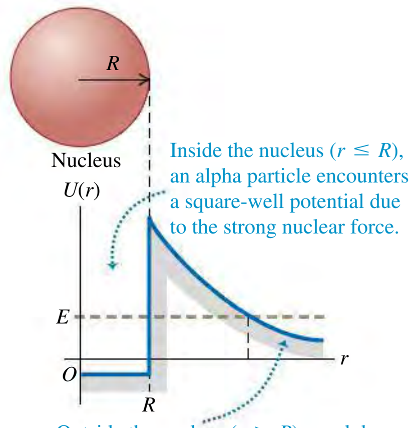


40.21 (a) Schematic diagram of the probe of a scanning tunneling microscope (STM). As the sharp conducting probe is scanned across the surface in the x- and y-directions, it is also moved in the z-direction to maintain a constant tunneling current. The changing position of the probe is recorded and used to construct an image of the surface. (b) This colored STM image shows "quantum wires": thin strips, just 10 atoms wide, of a conductive rare-earth silicide atop a silicon surface. Such quantum wires may one day be the basis of ultraminiaturized circuits.

maintain a constant tunneling current. The needle motion is recorded; after many parallel scans, an image of the surface can be reconstructed. Extremely precise control of needle motion, including isolation from vibration, is essential. Figure 40.21b shows an STM image. (Figure 40.17 is also an STM image.)

Tunneling is also of great importance in nuclear physics. A fusion reaction can occur when two nuclei tunnel through the barrier caused by their electrical repulsion and approach each other closely enough for the attractive nuclear force to cause them to fuse. Fusion reactions occur in the cores of stars, including the sun; without tunneling, the sun wouldn't shine. The emission of alpha particles from unstable nuclei such as radium also involves tunneling. An alpha particle is a cluster of two protons and two neutrons (the same as a nucleus of the most common form of helium). Such clusters form naturally within larger atomic nuclei. An alpha particle trying to escape from a nucleus encounters a potential barrier that results from the combined effect of the attractive nuclear force and the electrical repulsion of the remaining part of the nucleus (**Fig. 40.22**). To escape, the alpha particle must tunnel through this barrier. Depending on the barrier height and width for a given kind of alpha-emitting nucleus, the tunneling probability can be low or high, and the alpha-emitting material will have low or high radioactivity. Recall from Section 39.2 that Ernest Rutherford used alpha particles from a radioactive source to discover the atomic nucleus. Although Rutherford did not know it, tunneling allowed these alpha particles to escape from their parent nuclei, which made his experiments possible! We'll learn more about alpha decay in Chapter 43.

40.22 Approximate potential-energy function for an alpha particle interacting with a nucleus of radius R. If an alpha particle inside the nucleus has energy E greater than zero, it can tunnel through the barrier and escape from the nucleus.



TEST YOUR UNDERSTANDING OF SECTION 40.4 Is it possible for a particle undergoing tunneling to be found *within* the barrier rather than on either side of it?

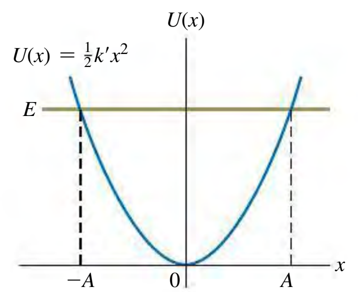
40.5 THE HARMONIC OSCILLATOR

Systems that *oscillate* are of tremendous importance in the physical world, from the oscillations of your eardrums in response to a sound wave to the vibrations of the ground caused by an earthquake. Oscillations are equally important on the microscopic scale where quantum effects dominate. The molecules of the air around you can be set into vibration when they collide with each other, the protons and neutrons in an excited atomic nucleus can oscillate in opposite directions, and a microwave oven transfers energy to food by making water molecules in the food flip back and forth. In this section we'll look at the solutions of the Schrödinger equation for the simplest kind of vibrating system, the quantum-mechanical harmonic oscillator.

As we learned in Section 14.2, a **harmonic oscillator** is a particle with mass m that moves along the x-axis under the influence of a conservative force $F_x = -k'x$. The constant k' is called the *force constant*. (In Section 14.2 we used the symbol k for the force constant. In this section we'll use the symbol k' instead to minimize confusion with the wave number $k = 2\pi/\lambda$.) The force is proportional to the particle's displacement x from its equilibrium position, $x = 0$. The corresponding potential-energy function is $U = \frac{1}{2}k'x^2$ (**Fig. 40.23**). In Newtonian mechanics, when the particle is displaced from equilibrium, it undergoes sinusoidal motion with frequency $f = (1/2\pi)(k'/m)^{1/2}$ and angular frequency $\omega = 2\pi f = (k'/m)^{1/2}$. The amplitude (that is, the maximum displacement from equilibrium) of these Newtonian oscillations is A, which is related to the energy E of the oscillator by $E = \frac{1}{2}k'A^2$.

Let's make an enlightened guess about the energy levels of a quantum-mechanical harmonic oscillator. In classical physics an electron oscillating with angular frequency ω emits electromagnetic radiation with that same angular frequency. It's reasonable to guess that when an excited quantum-mechanical harmonic oscillator with angular frequency $\omega = (k'/m)^{1/2}$ (according to Newtonian mechanics, at least) makes a transition from one energy level to a

40.23 Potential-energy function for the harmonic oscillator. In Newtonian mechanics the amplitude A is related to the total energy E by $E = \frac{1}{2}k'A^2$, and the particle is restricted to the range from $x = -A$ to $x = A$. In quantum mechanics the particle can be found at $x > A$ or $x < -A$.



lower level, it would emit a photon with this same angular frequency ω. The energy of such a photon is $hf = (2\pi\hbar)(\omega/2\pi) = \hbar\omega$. So we would expect that the spacing between adjacent energy levels of the harmonic oscillator would be

$$hf = \hbar\omega = \hbar\sqrt{\frac{k'}{m}} \tag{40.43}$$

That's the same spacing between energy levels that Planck assumed in deriving his radiation law (see Section 39.5). It was a good assumption; as we'll see, the energy levels are in fact half-integer $\left(\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots\right)$ multiples of $\hbar\omega$.

Wave Functions, Boundary Conditions, and Energy Levels

We'll begin our quantum-mechanical analysis of the harmonic oscillator by writing down the one-dimensional time-independent Schrödinger equation, Eq. (40.23), with $\frac{1}{2}k'x^2$ in place of U:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \tfrac{1}{2}k'x^2\psi(x) = E\psi(x) \quad \text{(Schrödinger equation for the harmonic oscillator)} \tag{40.44}$$

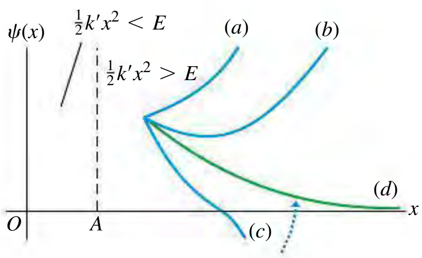
The solutions of this equation are wave functions for the physically possible states of the system.

In the discussion of square-well potentials in Section 40.2 we found that the energy levels are determined by boundary conditions at the walls of the well. However, the harmonic-oscillator potential has no walls as such; what, then, are the appropriate boundary conditions? Classically, $|x|$ cannot be greater than the amplitude A given by $E = \frac{1}{2}k'A^2$. Quantum mechanics does allow some penetration into classically forbidden regions, but the probability decreases as that penetration increases. Thus the wave functions must approach zero as $|x|$ grows large.

Satisfying the requirement that $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$ is not as trivial as it may seem. To see why this is, let's rewrite Eq. (40.44) in the form

$$\frac{d^2\psi(x)}{dx^2} = \frac{2m}{\hbar^2}\left(\tfrac{1}{2}k'x^2 - E\right)\psi(x) \tag{40.45}$$

40.24 Possible behaviors of harmonic-oscillator wave functions in the region $\frac{1}{2}k'x^2 > E$. In this region, $\psi(x)$ and $d^2\psi(x)/dx^2$ have the same sign. The curve is concave upward when $d^2\psi(x)/dx^2$ is positive and concave downward when $d^2\psi(x)/dx^2$ is negative.



Equation (40.45) shows that when x is large enough (either positive or negative) to make the quantity $\left(\frac{1}{2}k'x^2 - E\right)$ positive, the function $\psi(x)$ and its second derivative $d^2\psi(x)/dx^2$ have the same sign. **Figure 40.24** shows four possible kinds of behavior of $\psi(x)$ beginning at a point where x is greater than the classical amplitude A, so that $\frac{1}{2}k'x^2 - \frac{1}{2}k'A^2 = \frac{1}{2}k'x^2 - E > 0$. Let's look at these four cases more closely. If $\psi(x)$ is positive as shown in Fig. 40.24, Eq. (40.45) tells us that $d^2\psi(x)/dx^2$ is also positive and the function is *concave upward*. Note also that $d^2\psi(x)/dx^2$ is the rate of change of the *slope* of $\psi(x)$; this will help us understand how our four possible wave functions behave.

- *Curve a:* The slope of $\psi(x)$ is positive at point x. Since $d^2\psi(x)/dx^2 > 0$, the function curves upward increasingly steeply and goes to infinity. This violates the boundary condition that $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$, so this isn't a viable wave function.
- *Curve b:* The slope of $\psi(x)$ is negative at point x, and $d^2\psi(x)/dx^2$ has a large positive value. Hence the slope changes rapidly from negative to positive and keeps on increasing—so, again, the wave function goes to infinity. This wave function isn't viable either.
- *Curve c:* As for curve b, the slope is negative at point x. However, $d^2\psi(x)/dx^2$ now has a *small* positive value, so the slope increases only gradually as $\psi(x)$ decreases to zero and crosses over to negative values. Equation (40.45) tells us that once $\psi(x)$ becomes negative, $d^2\psi(x)/dx^2$ also becomes negative. Hence the curve becomes concave *downward* and heads for *negative* infinity. This wave function, too, fails to satisfy the requirement that $\psi(x) \rightarrow 0$ as $|x| \rightarrow \infty$ and thus isn't viable.

- *Curve d:* If the slope of $\psi(x)$ at point x is negative, and the positive value of $d^2\psi(x)/dx^2$ at this point is neither too large nor too small, the curve bends just enough to glide in asymptotically to the x-axis. In this case $\psi(x)$, $d\psi(x)/dx$, and $d^2\psi(x)/dx^2$ all approach zero at large x. This case offers the only hope of satisfying the boundary condition that $\psi(x) \to 0$ as $|x| \to \infty$, and it occurs only for certain very special values of the energy E.

This qualitative discussion suggests how the boundary conditions as $|x| \to \infty$ determine the possible energy levels for the quantum-mechanical harmonic oscillator. It turns out that these boundary conditions are satisfied only if the energy E is equal to one of the values E_n:

Energy levels for a harmonic oscillator

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m}} = \left(n + \tfrac{1}{2}\right)\hbar\omega \quad (n = 0, 1, 2, \ldots) \tag{40.46}$$

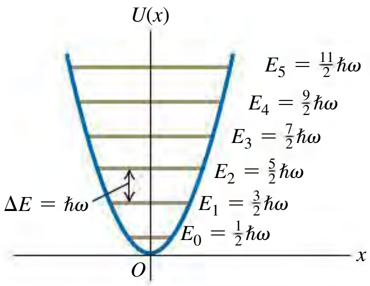
Quantum number; Planck's constant divided by 2π; Particle's mass; Force constant; Oscillation angular frequency

Note that the ground level of energy $E_0 = \frac{1}{2}\hbar\omega$ is denoted by the quantum number $n = 0$, *not* $n = 1$.

Equation (40.46) confirms our guess [(Eq. 40.43)] that adjacent energy levels are separated by a constant interval of $\hbar\omega = hf$, as Planck assumed in 1900. There are infinitely many levels; this shouldn't be surprising because we are dealing with an infinitely deep potential well. As $|x|$ increases, $U = \frac{1}{2}k'x^2$ increases without bound.

Figure 40.25 shows the lowest six energy levels and the potential-energy function $U(x)$. For each level n, the value of $|x|$ at which the horizontal line representing the total energy E_n intersects $U(x)$ gives the amplitude A_n of the corresponding Newtonian oscillator.

40.25 Energy levels for the harmonic oscillator. The spacing between any two adjacent levels is $\Delta E = \hbar\omega$. The energy of the ground level is $E_0 = \frac{1}{2}\hbar\omega$.



EXAMPLE 40.8 VIBRATION IN A CRYSTAL

SOLUTION

A sodium atom of mass 3.82×10^{-26} kg vibrates within a crystal. The potential energy increases by 0.0075 eV when the atom is displaced 0.014 nm from its equilibrium position. Treat the atom as a harmonic oscillator. (a) Find the angular frequency of the oscillations according to Newtonian mechanics. (b) Find the spacing (in electron volts) of adjacent vibrational energy levels according to quantum mechanics. (c) What is the wavelength of a photon emitted as the result of a transition from one level to the next lower level? In what region of the electromagnetic spectrum does this lie?

SOLUTION

IDENTIFY and SET UP: We'll find the force constant k' from the expression $U = \frac{1}{2}k'x^2$ for potential energy. We'll then find the angular frequency $\omega = (k'/m)^{1/2}$ and use this in Eq. (40.46) to find the spacing between adjacent energy levels. We'll calculate the wavelength of the emitted photon as in Example 40.6.

EXECUTE: We are given that $U = 0.0075\text{ eV} = 1.2 \times 10^{-21}$ J when $x = 0.014 \times 10^{-9}$ m, so we can solve $U = \frac{1}{2}k'x^2$ for k':

$$k' = \frac{2U}{x^2} = \frac{2(1.2 \times 10^{-21}\text{ J})}{(0.014 \times 10^{-9}\text{ m})^2} = 12.2\text{ N/m}$$

(a) The Newtonian angular frequency is

$$\omega = \sqrt{\frac{k'}{m}} = \sqrt{\frac{12.2\text{ N/m}}{3.82 \times 10^{-26}\text{ kg}}} = 1.79 \times 10^{13}\text{ rad/s}$$

(b) From Eq. (40.46) and Fig. 40.25, the spacing between adjacent energy levels is

$$\begin{aligned}\hbar\omega &= (1.055 \times 10^{-34}\text{ J}\cdot\text{s})(1.79 \times 10^{13}\text{ s}^{-1})\\ &= 1.89 \times 10^{-21}\text{ J}\left(\frac{1\text{ eV}}{1.602 \times 10^{-19}\text{ J}}\right) = 0.0118\text{ eV}\end{aligned}$$

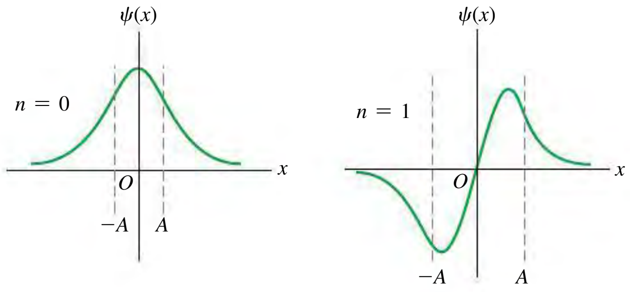
(c) The energy E of the emitted photon is equal to the energy lost by the oscillator in the transition, 0.0118 eV. Then

$$\begin{aligned}\lambda &= \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\text{ eV}\cdot\text{s})(3.00 \times 10^{8}\text{ m/s})}{0.0118\text{ eV}}\\ &= 1.05 \times 10^{-4}\text{ m} = 105\ \mu\text{m}\end{aligned}$$

This photon wavelength is in the infrared region of the spectrum.

EVALUATE: This example shows us that interatomic force constants are a few newtons per meter, about the same as those of household springs or spring-based toys such as the Slinky®. It also suggests that we can learn about the vibrations of molecules by measuring the radiation that they emit in transitioning to a lower vibrational state. We will explore this idea further in Chapter 42.

40.26 The first four wave functions for the harmonic oscillator. The amplitude A of a Newtonian oscillator with the same total energy is shown for each. Each wave function penetrates somewhat into the classically forbidden regions $|x| > A$. The total number of finite maxima and minima for each function is $n + 1$, one more than the quantum number.



Comparing Quantum and Newtonian Oscillators

The wave functions for the levels $n = 0, 1, 2, \ldots$ of the harmonic oscillator are called *Hermite functions;* they aren't encountered in elementary calculus courses but are well known to mathematicians. Each Hermite function is an exponential function multiplied by a polynomial in x. The harmonic-oscillator wave function corresponding to $n = 0$ and $E = E_0$ (the ground level) is

$$\psi(x) = Ce^{-\sqrt{mk'}\,x^2/2\hbar} \tag{40.47}$$

The constant C is chosen to normalize the function—that is, to make $\int_{-\infty}^{\infty}|\psi|^2\,dx = 1$. (We're using C rather than A as a normalization constant in this section, since we've already appropriated the symbol A to denote the Newtonian amplitude of a harmonic oscillator.) You can find C by using the following result from integral tables:

$$\int_{-\infty}^{\infty} e^{-a^2x^2}dx = \frac{\sqrt{\pi}}{a}$$

To confirm that $\psi(x)$ as given by Eq. (40.47) really *is* a solution of the Schrödinger equation for the harmonic oscillator, you can calculate the second derivative of this wave function, substitute it into Eq. (40.44), and verify that the equation is satisfied if the energy E is equal to $E_0 = \frac{1}{2}\hbar\omega$ (see Exercise 40.34). It's a little messy, but the result is satisfying and worth the effort.

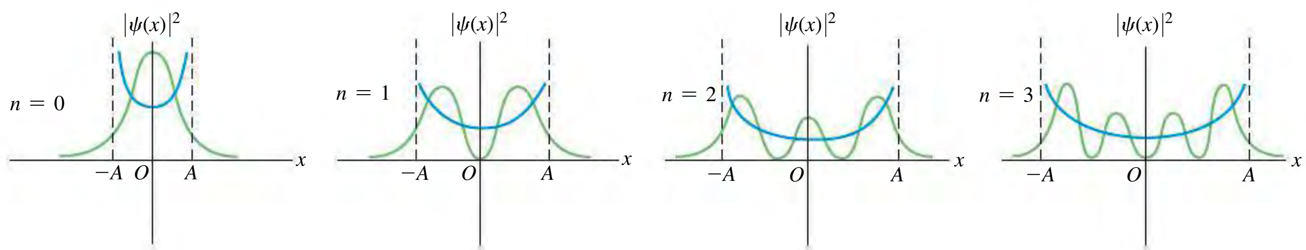
Figure 40.26 shows the the first four harmonic-oscillator wave functions. Each graph also shows the amplitude A of a Newtonian harmonic oscillator with the same energy—that is, the value of A determined from

$$\tfrac{1}{2}k'A^2 = \left(n + \tfrac{1}{2}\right)\hbar\omega \tag{40.48}$$

In each case there is some penetration of the wave function into the regions $|x| > A$ that are forbidden by Newtonian mechanics. This is similar to the effect that we noted in Section 40.3 for a particle in a finite square well.

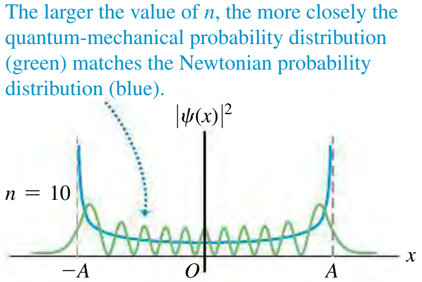
Figure 40.27 shows the probability distributions $|\psi(x)|^2$ for these states. Each graph also shows the probability distribution determined from Newtonian physics, in which the probability of finding the particle near a randomly chosen

40.27 Probability distribution functions $|\psi(x)|^2$ for the harmonic-oscillator wave functions shown in Fig. 40.26. The amplitude A of the Newtonian motion with the same energy is shown for each. The blue lines show the corresponding probability distributions for the Newtonian motion. As n increases, the averaged-out quantum-mechanical functions resemble the Newtonian curves more and more.



point is inversely proportional to the particle's speed at that point. If we average out the wiggles in the quantum-mechanical probability curves, the results for $n > 0$ resemble the Newtonian predictions. This agreement improves with increasing n; **Fig. 40.28** shows the classical and quantum-mechanical probability functions for $n = 10$. Notice that the spacing between zeros of $|\psi(x)|^2$ in Fig. 40.28 increases with increasing distance from $x = 0$. This makes sense from the Newtonian perspective: As a particle moves away from $x = 0$, its kinetic energy K and the magnitude p of its momentum both decrease. Thinking quantum-mechanically, this means that the wavelength $\lambda = h/p$ increases, so the spacing between zeros of $\psi(x)$ (and hence of $|\psi(x)|^2$) also increases.

40.28 Newtonian and quantum-mechanical probability distribution functions for a harmonic oscillator for the state $n = 10$. The Newtonian amplitude A is also shown.



In the Newtonian analysis of the harmonic oscillator the minimum energy is zero, with the particle at rest at its equilibrium position $x = 0$. This is not possible in quantum mechanics; no solution of the Schrödinger equation has $E = 0$ and satisfies the boundary conditions. Furthermore, such a state would violate the Heisenberg uncertainty principle because there would be no uncertainty in either position or momentum. The energy must be at least $\frac{1}{2}\hbar\omega$ for the system to conform to the uncertainty principle. To see qualitatively why this is so, consider a Newtonian oscillator with total energy $\frac{1}{2}\hbar\omega$. We can find the amplitude A and the maximum velocity just as we did in Section 14.3. When the particle is at its maximum displacement $(x = \pm A)$ and instantaneously at rest, $K = 0$ and $E = U = \frac{1}{2}k'A^2$. When the particle is at equilibrium $(x = 0)$ and moving at its maximum speed, $U = 0$ and $E = K = \frac{1}{2}mv_{\text{max}}{}^2$. Setting $E = \frac{1}{2}\hbar\omega$, we find

$$E = \tfrac{1}{2}k'A^2 = \tfrac{1}{2}\hbar\omega = \tfrac{1}{2}\hbar\left(\frac{k'}{m}\right)^{1/2} \qquad \text{so} \qquad A = \frac{\hbar^{1/2}}{k'^{1/4}m^{1/4}}$$

$$E = \tfrac{1}{2}mv_{\text{max}}{}^2 = \tfrac{1}{2}k'A^2 \qquad \text{so} \qquad v_{\text{max}} = A\left(\frac{k'}{m}\right)^{1/2} = \frac{\hbar^{1/2}k'^{1/4}}{m^{3/4}}$$

The maximum *momentum* of the particle is

$$p_{\text{max}} = mv_{\text{max}} = \hbar^{1/2}k'^{1/4}m^{1/4}$$

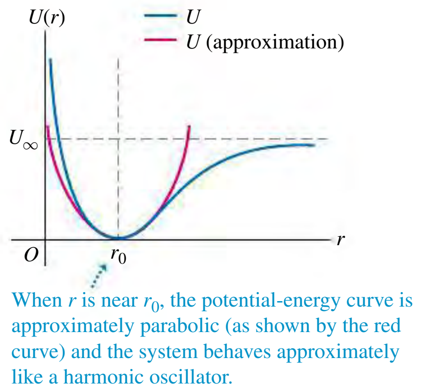
Here's where the Heisenberg uncertainty principle comes in. It turns out that the uncertainties in the particle's position and momentum (calculated as standard deviations) are, respectively, $\Delta x = A/\sqrt{2} = A/2^{1/2}$ and $\Delta p_x = p_{\text{max}}/\sqrt{2} = p_{\text{max}}/2^{1/2}$. Then the product of the two uncertainties is

$$\Delta x\,\Delta p_x = \left(\frac{\hbar^{1/2}}{2^{1/2}\,k'^{1/4}\,m^{1/4}}\right)\left(\frac{\hbar^{1/2}\,k'^{1/4}\,m^{1/4}}{2^{1/2}}\right) = \frac{\hbar}{2}$$

This product equals the minimum value allowed by Eq. (39.29), $\Delta x\,\Delta p_x \geq \hbar/2$, and thus satisfies the uncertainty principle. If the energy had been less than $\frac{1}{2}\hbar\omega$, the product $\Delta x\,\Delta p_x$ would have been less than $\hbar/2$, and the uncertainty principle would have been violated.

Even when a potential-energy function isn't precisely parabolic in shape, we may be able to approximate it by the harmonic-oscillator potential for sufficiently small displacements from equilibrium. **Figure 40.29** shows a typical potential-energy function for an interatomic force in a molecule. At large separations the curve of $U(r)$ versus r levels off, corresponding to the absence of force at great distances. But the curve is approximately parabolic near the minimum of $U(r)$ (the equilibrium separation of the atoms). Near equilibrium the molecular vibration is approximately simple harmonic with energy levels given by Eq. (40.46), as we assumed in Example 40.8.

40.29 A potential-energy function describing the interaction of two atoms in a diatomic molecule. The distance r is the separation between the centers of the atoms, and the equilibrium separation is $r = r_0$. The energy needed to dissociate the molecule is U_∞.



TEST YOUR UNDERSTANDING OF SECTION 40.5 A quantum-mechanical system initially in its ground level absorbs a photon and ends up in the first excited state. The system then absorbs a second photon and ends up in the second excited state. For which of the following systems does the second photon have a longer wavelength than the first one? (i) A harmonic oscillator; (ii) a hydrogen atom; (iii) a particle in a box. ❙

40.6 MEASUREMENT IN QUANTUM MECHANICS

We've seen how to use the Schrödinger equation to calculate the stationary-state wave functions and energy levels for various potential-energy functions $U(x)$. We've also seen how to interpret the wave function $\Psi(x, t)$ of a particle in terms of the probability distribution function $|\Psi(x, t)|^2$. We'll conclude with a brief discussion of what happens when we try to *measure* the properties of a quantum-mechanical particle. As we will see, the consequences of such a measurement can be startlingly different from what happens when we measure the properties of a familiar Newtonian particle, such as a marble or billiard ball.

Let's consider a "particle in a box"—that is, a particle in an infinite square well of width L, as described in Section 40.2. This particle of mass m is free to move along the x-axis in the region $0 \le x \le L$ but cannot move beyond this region. Let's suppose the particle is in a stationary state with definite energy E, equal to one of the energy levels E_n given by Eq. (40.31). If we measure the x-component of momentum of this particle, what is the result?

First let's consider the answer to that question for a Newtonian particle in a box (see Fig. 40.8). This could be a hockey puck sliding on frictionless ice and bouncing back and forth between two parallel walls. The energy E of the puck is equal to its kinetic energy $p^2/2m$, so the magnitude of its momentum is $p = \sqrt{2mE}$. The x-component of its momentum p_x is therefore

$$p_x = +\sqrt{2mE} \quad \text{or} \quad p_x = -\sqrt{2mE} \tag{40.49}$$

Whether p_x is positive or negative depends on whether the hockey puck is moving in the $+x$-direction (then $p_x = +\sqrt{2mE}$) or the $-x$-direction (then $p_x = -\sqrt{2mE}$). To determine which value of p_x is correct at a given time, we need only look at the puck to see in which direction it's moving.

We can't make such an observation in the dark; we need to shine some light on the hockey puck. We know from Section 38.1 that light comes in the form of photons and that a photon of wavelength λ has momentum $p = h/\lambda$. When we shine light on the puck to observe it, the photons collide with the puck and *change* its momentum. The mere act of measuring the puck's momentum can affect the quantity that we're trying to measure! The good news is that this change is minuscule: A hockey puck of mass $m = 0.165$ kg moving at speed $v = 1.00$ m/s has momentum $p = mv = 0.165$ kg · m/s, while a visible-light photon of wavelength 500 nm has momentum $p = h/\lambda = 1.33 \times 10^{-27}$ kg · m/s. Even if we directed all of the photons from a 100-W light source onto the puck for a 1.00-s burst of light, the total momentum in this burst would be only 3.33×10^{-7} kg · m/s, and the resulting change in the momentum of the puck would be negligible. In general, we can measure any of the properties of a Newtonian particle—its momentum, position, energy, and so on—without appreciably changing the quantity that we are measuring.

The situation is very different for a quantum-mechanical particle in a box. From Eq. (40.21) the state of such a particle with energy $E = E_n$ is described by the wave function

$$\Psi(x, t) = \psi_n(x)e^{-iE_nt/\hbar} = \psi_n(x)e^{-i\omega_nt} \tag{40.50}$$

In Eq. (40.50) the angular frequency is $\omega_n = E_n/\hbar$ and the time-independent, stationary-state wave function $\psi_n(x)$ is given by Eq. (40.35):

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin\frac{n\pi x}{L} \qquad (n = 1, 2, 3, \ldots) \tag{40.51}$$

This is a state of definite *energy,* but it is *not* a state of definite momentum: It represents a standing wave with equal amounts of momentum in the $+x$-direction and the $-x$-direction. To make this more explicit, recall from

Eqs. (40.30) and (40.31) that the magnitude of momentum in a state of energy E_n is $p_n = \sqrt{2mE_n} = nh/2L = n\pi\hbar/L$, and the corresponding wave number is $k_n = p_n/\hbar = n\pi/L$. So we can replace $n\pi/L$ in Eq. (40.51) with k_n:

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin k_n x$$

Recall also Euler's formula from Eq. (40.17): $e^{i\theta} = \cos\theta + i\sin\theta$ and $e^{-i\theta} = \cos\theta - i\sin\theta$. Hence $\sin\theta = (e^{i\theta} - e^{-i\theta})/2i$, and we can write

$$\psi_n(x) = \sqrt{\frac{2}{L}}\left(\frac{e^{ik_nx} - e^{-ik_nx}}{2i}\right) = \frac{1}{i\sqrt{2L}}(e^{ik_nx} - e^{-ik_nx}) \tag{40.52}$$

Now we substitute Eq. (40.52) into Eq. (40.50) and distribute the factors $1/i\sqrt{2L}$ and $e^{-i\omega_n t}$:

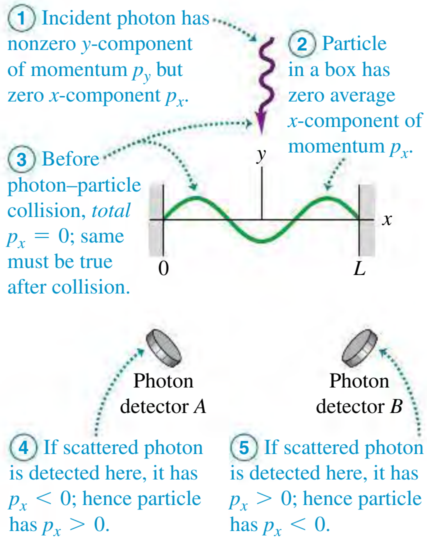
$$\begin{aligned}\Psi(x,t) &= \frac{1}{i\sqrt{2L}}(e^{ik_nx} - e^{-ik_nx})e^{-i\omega_n t} \\ &= \frac{1}{i\sqrt{2L}}e^{ik_nx}e^{-i\omega_n t} - \frac{1}{i\sqrt{2L}}e^{-ik_nx}e^{-i\omega_n t}\end{aligned} \tag{40.53}$$

In Eq. (40.53) the $e^{ik_nx}e^{-i\omega_n t}$ term is a wave function for a free particle with energy $E_n = \hbar\omega_n$ and a *positive* x-component of momentum $p_x = p_n = \hbar k_n$. In the $e^{-ik_nx}e^{-i\omega_n t}$ term, k_n is replaced by $-k_n$, so this term is a wave function for a free particle with the same energy $E_n = \hbar\omega_n$ but a *negative* x-component of momentum $p_x = -p_n = -\hbar k_n$. These two possible values of p_x are the same as for a Newtonian particle in a box, Eq. (40.49). The difference is that as the Newtonian particle bounces back and forth between the walls of the box, it has positive p_x half of the time and negative p_x half of the time. Only its time-averaged value of p_x is zero. But because both terms for positive p_x and negative p_x are present in Eq. (40.53), the quantum-mechanical particle has *both* signs of the x-component of momentum present at *all* times. As we stated earlier, this stationary state for a quantum-mechanical particle in a box has a definite energy [both terms in Eq. (40.53) have the same value of ω_n and hence the same value of $E_n = \hbar\omega_n$] but does not have a definite momentum. Because the $e^{ik_nx}e^{-i\omega_n t}$ and $e^{-ik_nx}e^{-i\omega_n t}$ terms in Eq. (40.53) have coefficients of the same magnitude, $1/\sqrt{2L}$, the *instantaneous* average value of p_x for the quantum-mechanical particle is zero (the average of $\hbar k_n$ and $-\hbar k_n$) at *all* times.

What value do we get if we *measure* the momentum of the quantum-mechanical particle in a box? As for the Newtonian particle, we can measure the momentum by shining a light on it. Let's fire a single photon, moving in the $-y$-direction, at the particle and allow the photon and particle to collide (**Fig. 40.30**). Before the collision the total x-component of momentum of the system of photon and particle is zero. Momentum is conserved in the collision, so the same is true after the collision. After the collision, whichever sign of p_x the photon has, the x-component of momentum of the particle will have the opposite sign. If the photon is detected in detector A, we conclude that the particle has $p_x = \hbar k_n$; if instead the photon is detected in detector B, we conclude that the particle has $p_x = -\hbar k_n$.

In this experiment, we need to be even more concerned about how the photon changes the momentum of the particle than in the Newtonian case. For an electron in a box of width $L = 1.00 \times 10^{-6}$ m $= 1.00\ \mu$m, the electron momentum has a minimum magnitude of $p = 3.31 \times 10^{-28}$ kg·m/s (corresponding to the $n = 1$ energy level), which is only about one-quarter that of a visible-light photon of wavelength 500 nm. To minimize the change in the electron's magnitude of momentum due to the collision, we should use a photon of much longer wavelength (say, a radio-wave photon) and hence much smaller momentum.

40.30 Using photon scattering to measure the x-component of momentum of a particle in a box.



Even when we use a photon with the lowest possible momentum, however, we find that the state of the particle in the box *must* change as a result of the experiment. Here's a summary of the results:

1. If the measurement shows that the particle has positive $p_x = \hbar k_n$, the wave function *changes* from that given in Eq. (40.53) to one with an $e^{ik_n x}e^{-i\omega_n t}$ term *only*. The other term, which corresponds to $p_x = -\hbar k_n$, disappears. We say that the wave function, which was a combination of two terms with different values of p_x, has undergone *wave-function collapse*—it has collapsed to one term with $p_x = \hbar k_n$ as a consequence of measuring the value of p_x. To test this result, we fire a second photon immediately after the first. The second photon scatters from the particle as we would expect if the particle had the value $p_x = \hbar k_n$.
2. If the measurement shows that the particle has negative $p_x = -\hbar k_n$, the wave function collapses in the opposite way: It changes to one with an $e^{-ik_n x}e^{-i\omega_n t}$ term only. The $p_x = \hbar k_n$ term disappears.
3. If we repeat the experiment many times, each time starting with the particle described by the wave function in Eq. (40.53), 50% of the time we measure the particle to have $p_x = \hbar k_n$ and 50% of the time we measure the particle to have $p_x = -\hbar k_n$. For any given time that we try the experiment, there is no way to predict which outcome will occur. We can state only that there is equal probability of either outcome.

These results reveal a fact of quantum-mechanical life: *Measuring a physical property of a system can change the wave function of that system.* By measuring the value of p_x for a particle in a box, we changed the wave function from one that was a combination of two wave functions, one for $p_x = \hbar k_n$ and one for $p_x = -\hbar k_n$, to one with a definite value of p_x. This change in the wave function is not described by the time-dependent Schrödinger equation [Eq. (40.20)] but is a consequence of the measurement process. It is also independent of how the measurement is carried out: No matter how small the momentum of the incident photon shown in Fig. 40.30, the same collapse of the wave function takes place. Indeed, *any* experiment to measure p_x for a particle in a box in a steady state, no matter how the experiment is designed, will have the results that we described earlier.

(*After* the measurement, the wave function will undergo further change that *is* described by the Schrödinger equation. Neither $e^{ik_n x}e^{-i\omega_n t}$ nor $e^{-ik_n x}e^{-i\omega_n t}$ by itself satisfies the boundary conditions for a particle in a box—namely, that the wave function vanishes at $x = 0$ and $x = L$. The wave function must evolve to satisfy these conditions.)

CAUTION Quantum measurement misconceptions If we measure the particle to have $p_x = \hbar k_n$, does that mean it had $p_x = \hbar k_n$ before the measurement? No; the particle acquired that value as a result of the measurement. If we measure the particle to have $p_x = \hbar k_n$ instead of $p_x = -\hbar k_n$, does that mean there was some bias in the way we did the measurement? Again, no; the result of any given experiment is random. All quantum mechanics can do is predict the probability that this experiment will give us a certain result.

Note that not every measurement of a quantum-mechanical system causes a change in the wave function. If we perform an experiment that measures only the *energy* of a particle given by the wave function in Eq. (40.53), the wave function does *not* change. That's because the wave function already corresponds to a state of definite energy $E_n = \hbar\omega_n$, so there is a 100% probability that we will measure that value of energy.

You may ask, Does the wave function really collapse? Most physicists would answer yes, but some theorists have devised alternative models of what happens in a quantum-mechanical measurement. One model, called the *many-worlds interpretation,* asserts that there is a *universal* wave function that describes all particles in the universe. Whenever a measurement of any sort takes place, whether of human origin (like our experiment) or natural origin (for example, a photon of sunlight scattering from an electron in an atom in the atmosphere), this universal wave function does not collapse. Instead, every measurement causes the universe to branch into alternative timelines. So, when we carry out the experiment depicted in Fig. 40.30, the universe splits into one timeline in which the photon goes into detector A and a second timeline in which the photon goes into detector B. These two timelines then no longer communicate.

As weird as these aspects of quantum mechanics are, others are far weirder. We will investigate these in Chapter 41 after we have learned more about the nature of the electron.

TEST YOUR UNDERSTANDING OF SECTION 40.6 A particle in a box is described by a wave function that is a combination of the $n = 1$ and $n = 2$ stationary states: $\Psi(x, t) = C\psi_1(x)e^{-iE_1t/\hbar} + D\psi_2(x)e^{-iE_2t/\hbar}$, where $\psi_1(x)$ and $\psi_2(x)$ are given by Eq. (40.35), E_1 and E_2 are given by Eq. (40.31), and C and D are nonzero constants. If you carry out an experiment to measure the energy of this particle, the result is *guaranteed* to be (i) E_1; (ii) E_2; (iii) $(E_1 + E_2)/2$; (iv) intermediate between E_1 and E_2, with a value that depends on the values of C and D; (v) none of these.

CHAPTER 40 SUMMARY

SOLUTIONS TO ALL EXAMPLES

Wave functions: The wave function for a particle contains all of the information about that particle. If the particle moves in one dimension in the presence of a potential energy function $U(x)$, the wave function $\Psi(x, t)$ obeys the one-dimensional Schrödinger equation. (For a *free* particle on which no forces act, $U(x) = 0$.) The quantity $|\Psi(x, t)|^2$, called the probability distribution function, determines the relative probability of finding a particle near a given position at a given time. If the particle is in a state of definite energy, called a stationary state, $\Psi(x, t)$ is a product of a function $\psi(x)$ that depends on only spatial coordinates and a function $e^{-iEt/\hbar}$ that depends on only time. For a stationary state, the probability distribution function is independent of time.

A spatial stationary-state wave function $\psi(x)$ for a particle that moves in one dimension in the presence of a potential-energy function $U(x)$ satisfies the time-independent Schrödinger equation. More complex wave functions can be constructed by superposing stationary-state wave functions. These can represent particles that are localized in a certain region, thus representing both particle and wave aspects. (See Examples 40.1 and 40.2.)

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + U(x)\Psi(x,t) = i\hbar\frac{\partial\Psi(x,t)}{\partial t} \quad (40.20)$$

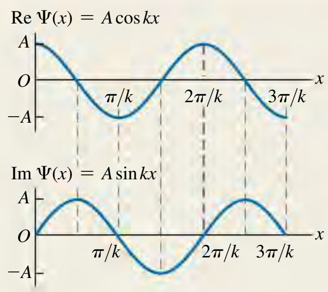
(general 1-D Schrödinger equation)

$$\Psi(x,t) = \psi(x)e^{-iEt/\hbar} \quad (40.21)$$

(time-dependent wave function for a state of definite energy)

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x)$$

(time-independent 1-D Schrödinger equation) (40.23)



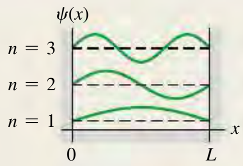
Particle in a box: The energy levels for a particle of mass m in a box (an infinitely deep square potential well) with width L are given by Eq. (40.31). The corresponding normalized stationary-state wave functions of the particle are given by Eq. (40.35). (See Examples 40.3 and 40.4.)

$$E_n = \frac{p_n^2}{2m} = \frac{n^2h^2}{8mL^2} = \frac{n^2\pi^2\hbar^2}{2mL^2}$$

$(n = 1, 2, 3, \ldots)$ (40.31)

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin\frac{n\pi x}{L}$$

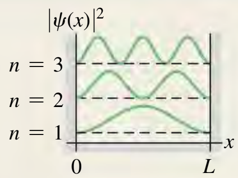
$(n = 1, 2, 3, \ldots)$ (40.35)



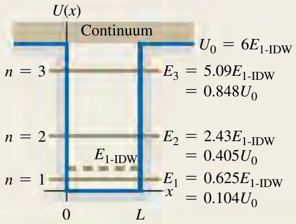
Wave functions and normalization: To be a solution of the Schrödinger equation, the wave function $\psi(x)$ and its derivative $d\psi(x)/dx$ must be continuous everywhere except where the potential-energy function $U(x)$ has an infinite discontinuity. Wave functions are usually normalized so that the total probability of finding the particle somewhere is unity.

$$\int_{-\infty}^{\infty}|\psi(x)|^2\,dx = 1 \quad (40.33)$$

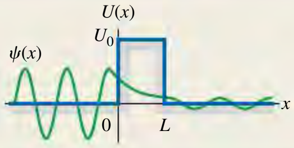
(normalization condition)



Finite potential well: In a potential well with finite depth U_0, the energy levels are lower than those for an infinitely deep well with the same width, and the number of energy levels corresponding to bound states is finite. The levels are obtained by matching wave functions at the well walls to satisfy the continuity of $\psi(x)$ and $d\psi(x)/dx$. (See Examples 40.5 and 40.6.)



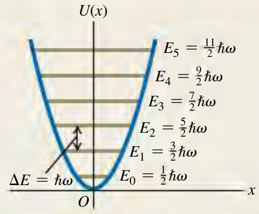
Potential barriers and tunneling: There is a certain probability that a particle will penetrate a potential-energy barrier even though its initial energy is less than the barrier height. This process is called tunneling. (See Example 40.7.)



Quantum harmonic oscillator: The energy levels for the harmonic oscillator (for which $U(x) = \frac{1}{2}k'x^2$) are given by Eq. (40.46). The spacing between any two adjacent levels is $\hbar\omega$, where $\omega = \sqrt{k'/m}$ is the oscillation angular frequency of the corresponding Newtonian harmonic oscillator. (See Example 40.8.)

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m}} = \left(n + \tfrac{1}{2}\right)\hbar\omega$$

$$(n = 0, 1, 2, 3, \ldots) \qquad (40.46)$$



Measurement in quantum mechanics: If the wave function of a particle does not correspond to a definite value of a certain physical property (such as momentum or energy), the wave function changes when we measure that property. This phenomenon is called wave-function collapse.

BRIDGING PROBLEM A PACKET IN A BOX

SOLUTION

A particle of mass m in an infinitely deep well (see Fig. 40.9) has the following wave function in the region from $x = 0$ to $x = L$:

$$\Psi(x, t) = \frac{1}{\sqrt{2}}\psi_1(x)e^{-iE_1t/\hbar} + \frac{1}{\sqrt{2}}\psi_2(x)e^{-iE_2t/\hbar}$$

Here $\psi_1(x)$ and $\psi_2(x)$ are the normalized stationary-state wave functions for the first two levels ($n = 1$ and $n = 2$), given by Eq. (40.35). E_1 and E_2, given by Eq. (40.31), are the energies of these levels. The wave function is zero for $x < 0$ and for $x > L$. (a) Find the probability distribution function for this wave function. (b) Does $\Psi(x, t)$ represent a stationary state of definite energy? How can you tell? (c) Show that the wave function $\Psi(x, t)$ is normalized. (d) Find the angular frequency of oscillation of the probability distribution function. What is the interpretation of this oscillation? (e) Suppose instead that $\Psi(x, t)$ is a combination of the wave functions of the two lowest levels of a finite well of length L and height U_0 equal to six times the energy of the lowest-energy bound state of an infinite well of length L. What would be the angular frequency of the probability distribution function in this case?

SOLUTION GUIDE

IDENTIFY and SET UP

1. In Section 40.1 we saw how to interpret a combination of two free-particle wave functions of different energies. In this problem you need to apply these same ideas to a combination of wave functions for the infinite well (Section 40.2) and the finite well (Section 40.3).

EXECUTE

2. Write down the full time-dependent wave function $\Psi(x, t)$ and its complex conjugate $\Psi^*(x, t)$ by using the functions $\psi_1(x)$ and $\psi_2(x)$ from Eq. (40.35). Use these to calculate the probability distribution function, and decide whether or not this function depends on time.
3. To check for normalization, you'll need to verify that when you integrate the probability distribution function from step 2 over all values of x, the integral is equal to 1. [*Hint:* The trigonometric identities $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$ and $\sin\theta\sin\phi = \cos(\theta - \phi) - \cos(\theta + \phi)$ may be helpful.]
4. To find the answer to part (d) you'll need to identify the oscillation angular frequency ω_{osc} in your expression from step 2 for the probability distribution function. To interpret the oscillations, draw graphs of the probability distribution functions at times $t = 0$, $t = T/4$, $t = T/2$, and $t = 3T/4$, where $T = 2\pi/\omega_{osc}$ is the oscillation period of the probability distribution function.
5. For the finite well you do not have simple expressions for the first two stationary-state wave functions $\psi_1(x)$ and $\psi_2(x)$. However, you can still find the oscillation angular frequency ω_{osc}, which is related to the energies E_1 and E_2 in the same way as for the infinite-well case. (Can you see why?)

EVALUATE

6. Why are the factors of $1/\sqrt{2}$ in the wave function $\Psi(x, t)$ important?
7. Why do you suppose the oscillation angular frequency for a finite well is lower than for an infinite well of the same width?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q40.1 If quantum mechanics replaces the language of Newtonian mechanics, why don't we have to use wave functions to describe the motion of macroscopic bodies such as baseballs and cars?

Q40.2 A student remarks that the relationship of ray optics to the more general wave picture is analogous to the relationship of Newtonian mechanics, with well-defined particle trajectories, to quantum mechanics. Comment on this remark.

Q40.3 As Eq. (40.21) indicates, the time-dependent wave function for a stationary state is a complex number having a real part and an imaginary part. How can this function have any physical meaning, since part of it is *imaginary*?

Q40.4 Why must the wave function of a particle be normalized?

Q40.5 If a particle is in a stationary state, does that mean that the particle is not moving? If a particle moves in empty space with constant momentum $\vec{p}$ and hence constant energy $E = p^2/2m$, is it in a stationary state? Explain your answers.

Q40.6 For the particle in a box, we chose $k = n\pi/L$ with $n = 1, 2, 3, \ldots$ to fit the boundary condition that $\psi = 0$ at $x = L$. However, $n = 0, -1, -2, -3, \ldots$ also satisfy that boundary condition. Why didn't we also choose those values of n?

Q40.7 If ψ is normalized, what is the physical significance of the area under a graph of $|\psi|^2$ versus x between x_1 and x_2? What is the total area under the graph of $|\psi|^2$ when all x are included? Explain.

Q40.8 For a particle in a box, what would the probability distribution function $|\psi|^2$ look like if the particle behaved like a classical (Newtonian) particle? Do the actual probability distributions approach this classical form when n is very large? Explain.

Q40.9 In Chapter 15 we represented a standing wave as a superposition of two waves traveling in opposite directions. Can the wave functions for a particle in a box also be thought of as a combination of two traveling waves? Why or why not? What physical interpretation does this representation have? Explain.

Q40.10 A particle in a box is in the ground level. What is the probability of finding the particle in the right half of the box? (Refer to Fig. 40.12, but don't evaluate an integral.) Is the answer the same if the particle is in an excited level? Explain.

Q40.11 The wave functions for a particle in a box (see Fig. 40.12a) are zero at certain points. Does this mean that the particle can't move past one of these points? Explain.

Q40.12 For a particle confined to an infinite square well, is it correct to say that each state of definite energy is also a state of definite wavelength? Is it also a state of definite momentum? Explain. (*Hint:* Remember that momentum is a vector.)

Q40.13 For a particle in a finite potential well, is it correct to say that each bound state of definite energy is also a state of definite wavelength? Is it a state of definite momentum? Explain.

Q40.14 In Fig. 40.12b, the probability function is zero at the points $x = 0$ and $x = L$, the "walls" of the box. Does this mean that the particle never strikes the walls? Explain.

Q40.15 A particle is confined to a finite potential well in the region $0 < x < L$. How does the area under the graph of $|\psi|^2$ in the region $0 < x < L$ compare to the total area under the graph of $|\psi|^2$ when including all possible x?

Q40.16 Compare the wave functions for the first three energy levels for a particle in a box of width L (see Fig. 40.12a) to the corresponding wave functions for a finite potential well of the same width (see Fig. 40.15a). How does the wavelength in the interval $0 \le x \le L$ for the $n = 1$ level of the particle in a box compare to the corresponding wavelength for the $n = 1$ level of the finite potential well? Use this to explain why E_1 is less than $E_{1-\mathrm{IDW}}$ in the situation depicted in Fig. 40.15b.

Q40.17 It is stated in Section 40.3 that a finite potential well always has at least one bound level, no matter how shallow the well. Does this mean that as $U_0 \to 0$, $E_1 \to 0$? Does this violate the Heisenberg uncertainty principle? Explain.

Q40.18 Figure 40.15a shows that the higher the energy of a bound state for a finite potential well, the more the wave function extends outside the well (into the intervals $x < 0$ and $x > L$). Explain why this happens.

Q40.19 In classical (Newtonian) mechanics, the total energy E of a particle can never be less than the potential energy U because the kinetic energy K cannot be negative. Yet in barrier tunneling (see Section 40.4) a particle passes through regions where E is less than U. Is this a contradiction? Explain.

Q40.20 Figure 40.17 shows the scanning tunneling microscope image of 48 iron atoms placed on a copper surface, the pattern indicating the density of electrons on the copper surface. What can you infer about the potential-energy function inside the circle of iron atoms?

Q40.21 Qualitatively, how would you expect the probability for a particle to tunnel through a potential barrier to depend on the height of the barrier? Explain.

Q40.22 The wave function shown in Fig. 40.20 is nonzero for both $x < 0$ and $x > L$. Does this mean that the particle splits into two parts when it strikes the barrier, with one part tunneling through the barrier and the other part bouncing off the barrier? Explain.

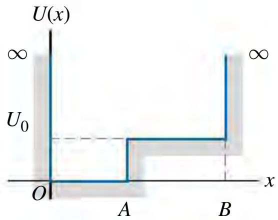
Q40.23 The probability distributions for the harmonic-oscillator wave functions (see Figs. 40.27 and 40.28) begin to resemble the classical (Newtonian) probability distribution when the quantum number n becomes large. Would the distributions become the same as in the classical case in the limit of very large n? Explain.

Q40.24 In Fig. 40.28, how does the probability of finding a particle in the center half of the region $-A < x < A$ compare to the probability of finding the particle in the outer half of the region? Is this consistent with the physical interpretation of the situation?

Q40.25 Compare the allowed energy levels for the hydrogen atom, the particle in a box, and the harmonic oscillator. What are the values of the quantum number n for the ground level and the second excited level of each system?

Q40.26 Sketch the wave function for the potential-energy well shown in **Fig. Q40.26** when E_1 is less than U_0 and when E_3 is greater than U_0.

Figure **Q40.26**



Q40.27 (a) A particle in a box has wave function $\Psi(x,t) = \psi_2(x)e^{-iE_2t/\hbar}$, where ψ_n and E_n are given by Eqs. (40.35) and (40.31), respectively. If the energy of the particle is measured, what is the result? (b) If instead the particle has wave function $\Psi(x,t) = (1/\sqrt{2})(\psi_1(x)e^{-iE_1t/\hbar} + \psi_2(x)e^{-iE_2t/\hbar})$ and the energy of the particle is measured, what is the result? (c) If we had many identical particles with the wave function of part (b) and measured the energy of each, what would be the average value of all of the measurements? Can we say that, before the measurement was made, each particle had this average energy? Explain.

EXERCISES

Section 40.1 Wave Functions and the One-Dimensional Schrödinger Equation

40.1 • An electron is moving as a free particle in the $-x$-direction with momentum that has magnitude 4.50×10^{-24} kg·m/s. What is the one-dimensional time-dependent wave function of the electron?

40.2 • A free particle moving in one dimension has wave function

$$\Psi(x,t) = A\left[e^{i(kx-\omega t)} - e^{i(2kx-4\omega t)}\right]$$

where k and ω are positive real constants. (a) At $t = 0$ what are the two smallest positive values of x for which the probability function $|\Psi(x,t)|^2$ is a maximum? (b) Repeat part (a) for time $t = 2\pi/\omega$. (c) Calculate v_{av} as the distance the maxima have moved divided by the elapsed time. Compare your result to the expression $v_{av} = (\omega_2 - \omega_1)/(k_2 - k_1)$ from Example 40.1.

40.3 • Consider the free-particle wave function of Example 40.1. Let $k_2 = 3k_1 = 3k$. At $t = 0$ the probability distribution function $|\Psi(x,t)|^2$ has a maximum at $x = 0$. (a) What is the smallest positive value of x for which the probability distribution function has a maximum at time $t = 2\pi/\omega$, where $\omega = \hbar k^2/2m$? (b) From your result in part (a), what is the average speed with which the probability distribution is moving in the $+x$-direction? Compare your result to the expression $v_{av} = (\omega_2 - \omega_1)/(k_2 - k_1)$ from Example 40.1.

40.4 • A particle is described by a wave function $\psi(x) = Ae^{-\alpha x^2}$, where A and α are real, positive constants. If the value of α is increased, what effect does this have on (a) the particle's uncertainty in position and (b) the particle's uncertainty in momentum? Explain your answers.

40.5 • Consider a wave function given by $\psi(x) = A\sin kx$, where $k = 2\pi/\lambda$ and A is a real constant. (a) For what values of x is there the highest probability of finding the particle described by this wave function? Explain. (b) For which values of x is the probability *zero?* Explain.

40.6 •• Compute $|\Psi|^2$ for $\Psi = \psi\sin\omega t$, where ψ is time independent and ω is a real constant. Is this a wave function for a stationary state? Why or why not?

40.7 • CALC Let ψ_1 and ψ_2 be two solutions of Eq. (40.23) with energies E_1 and E_2, respectively, where $E_1 \neq E_2$. Is $\psi = A\psi_1 + B\psi_2$, where A and B are nonzero constants, a solution to Eq. (40.23)? Explain your answer.

Section 40.2 Particle in a Box

40.8 •• CALC A particle moving in one dimension (the x-axis) is described by the wave function

$$\psi(x) = \begin{cases} Ae^{-bx}, & \text{for } x \geq 0 \\ Ae^{bx}, & \text{for } x < 0 \end{cases}$$

where $b = 2.00\ \text{m}^{-1}$, $A > 0$, and the $+x$-axis points toward the right. (a) Determine A so that the wave function is normalized. (b) Sketch the graph of the wave function. (c) Find the probability of finding this particle in each of the following regions: (i) within 50.0 cm of the origin, (ii) on the left side of the origin (can you first guess the answer by looking at the graph of the wave function?), (iii) between $x = 0.500$ m and $x = 1.00$ m.

40.9 • **Ground-Level Billiards.** (a) Find the lowest energy level for a particle in a box if the particle is a billiard ball ($m = 0.20$ kg) and the box has a width of 1.3 m, the size of a billiard table. (Assume that the billiard ball slides without friction rather than rolls; that is, ignore the *rotational* kinetic energy.) (b) Since the energy in part (a) is all kinetic, to what speed does this correspond? How much time would it take at this speed for the ball to move from one side of the table to the other? (c) What is the difference in energy between the $n = 2$ and $n = 1$ levels? (d) Are quantum-mechanical effects important for the game of billiards?

40.10 • A proton is in a box of width L. What must the width of the box be for the ground-level energy to be 5.0 MeV, a typical value for the energy with which the particles in a nucleus are bound? Compare your result to the size of a nucleus—that is, on the order of 10^{-14} m.

40.11 •• Find the width L of a one-dimensional box for which the ground-state energy of an electron in the box equals the absolute value of the ground state of a hydrogen atom.

40.12 •• When a hydrogen atom undergoes a transition from the $n = 2$ to the $n = 1$ level, a photon with $\lambda = 122$ nm is emitted. (a) If the atom is modeled as an electron in a one-dimensional box, what is the width of the box in order for the $n = 2$ to $n = 1$ transition to correspond to emission of a photon of this energy? (b) For a box with the width calculated in part (a), what is the ground-state energy? How does this correspond to the ground-state energy of a hydrogen atom? (c) Do you think a one-dimensional box is a good model for a hydrogen atom? Explain. (*Hint:* Compare the spacing between adjacent energy levels as a function of n.)

40.13 •• A certain atom requires 3.0 eV of energy to excite an electron from the ground level to the first excited level. Model the atom as an electron in a box and find the width L of the box.

40.14 • An electron in a one-dimensional box has ground-state energy 2.00 eV. What is the wavelength of the photon absorbed when the electron makes a transition to the second excited state?

40.15 •• CALC **Normalization of the Wave Function.** Consider a particle moving in one dimension, which we shall call the x-axis. (a) What does it mean for the wave function of this particle to be *normalized*? (b) Is the wave function $\psi(x) = e^{ax}$, where a is a positive real number, normalized? Could this be a valid wave function? (c) If the particle described by the wave function $\psi(x) = Ae^{-bx}$, where A and b are positive real numbers, is confined to the range $x \geq 0$, determine A (including its units) so that the wave function is normalized.

40.16 • Recall that $|\psi|^2dx$ is the probability of finding the particle that has normalized wave function $\psi(x)$ in the interval x to $x + dx$. Consider a particle in a box with rigid walls at $x = 0$ and $x = L$. Let the particle be in the ground level and use ψ_n as given in Eq. (40.35). (a) For which values of x, if any, in the range from 0 to L is the probability of finding the particle zero? (b) For which values of x is the probability highest? (c) In parts (a) and (b) are your answers consistent with Fig. 40.12? Explain.

40.17 • Repeat Exercise 40.16 for the particle in the first excited level.

40.18 • (a) Find the excitation energy from the ground level to the third excited level for an electron confined to a box of width 0.360 nm. (b) The electron makes a transition from the $n = 1$ to $n = 4$ level by absorbing a photon. Calculate the wavelength of this photon.

40.19 • An electron is in a box of width 3.0×10^{-10} m. What are the de Broglie wavelength and the magnitude of the momentum of the electron if it is in (a) the $n = 1$ level; (b) the $n = 2$ level; (c) the $n = 3$ level? In each case how does the wavelength compare to the width of the box?

40.20 •• When an electron in a one-dimensional box makes a transition from the $n = 1$ energy level to the $n = 2$ level, it absorbs a photon of wavelength 426 nm. What is the wavelength of that photon when the electron undergoes a transition (a) from the $n = 2$ to the $n = 3$ energy level and (b) from the $n = 1$ to the $n = 3$ energy level? (c) What is the width L of the box?

Section 40.3 Potential Wells

40.21 • An electron is bound in a square well of depth $U_0 = 6E_{1-\text{IDW}}$. What is the width of the well if its ground-state energy is 2.00 eV?

40.22 •• An electron is moving past the square well shown in Fig. 40.13. The electron has energy $E = 3U_0$. What is the ratio of the de Broglie wavelength of the electron in the region $x > L$ to the wavelength for $0 < x < L$?

40.23 •• An electron is bound in a square well of width 1.50 nm and depth $U_0 = 6E_{1-\text{IDW}}$. If the electron is initially in the ground level and absorbs a photon, what maximum wavelength can the photon have and still liberate the electron from the well?

40.24 •• An electron is in the ground state of a square well of width $L = 4.00 \times 10^{-10}$ m. The depth of the well is six times the ground-state energy of an electron in an infinite well of the same width. What is the kinetic energy of this electron after it has absorbed a photon of wavelength 72 nm and moved away from the well?

40.25 •• A proton is bound in a square well of width 4.0 fm $= 4.0 \times 10^{-15}$ m. The depth of the well is six times the ground-level energy $E_{1-\text{IDW}}$ of the corresponding infinite well. If the proton makes a transition from the level with energy E_1 to the level with energy E_3 by absorbing a photon, find the wavelength of the photon.

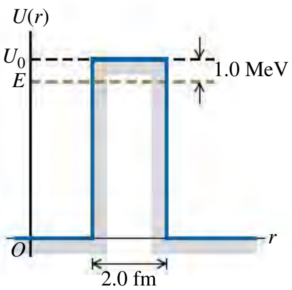
40.26 •• An electron is bound in a square well that has a depth equal to six times the ground-level energy $E_{1-\text{IDW}}$ of an infinite well of the same width. The longest-wavelength photon that is absorbed by this electron has a wavelength of 582 nm. Determine the width of the well.

Section 40.4 Potential Barriers and Tunneling

40.27 •• (a) An electron with initial kinetic energy 32 eV encounters a square barrier with height 41 eV and width 0.25 nm. What is the probability that the electron will tunnel through the barrier? (b) A proton with the same kinetic energy encounters the same barrier. What is the probability that the proton will tunnel through the barrier?

40.28 •• **Alpha Decay.** In a simple model for a radioactive nucleus, an alpha particle ($m = 6.64 \times 10^{-27}$ kg) is trapped by a square barrier that has width 2.0 fm and height 30.0 MeV. (a) What is the tunneling probability when the alpha particle encounters the barrier if its kinetic energy is 1.0 MeV below the top of the barrier (**Fig. E40.28**)? (b) What is the tunneling probability if the energy of the alpha particle is 10.0 MeV below the top of the barrier?

Figure **E40.28**



40.29 • An electron with initial kinetic energy 6.0 eV encounters a barrier with height 11.0 eV. What is the probability of tunneling if the width of the barrier is (a) 0.80 nm and (b) 0.40 nm?

40.30 • An electron with initial kinetic energy 5.0 eV encounters a barrier with height U_0 and width 0.60 nm. What is the transmission coefficient if (a) $U_0 = 7.0$ eV; (b) $U_0 = 9.0$ eV; (c) $U_0 = 13.0$ eV?

40.31 •• An electron is moving past the square barrier shown in Fig. 40.19, but the energy of the electron is *greater* than the barrier height. If $E = 2U_0$, what is the ratio of the de Broglie wavelength of the electron in the region $x > L$ to the wavelength for $0 < x < L$?

40.32 • A proton with initial kinetic energy 50.0 eV encounters a barrier of height 70.0 eV. What is the width of the barrier if the probability of tunneling is 8.0×10^{-3}? How does this compare with the barrier width for an electron with the same energy tunneling through a barrier of the same height with the same probability?

Section 40.5 The Harmonic Oscillator

40.33 • A wooden block with mass 0.250 kg is oscillating on the end of a spring that has force constant 110 N/m. Calculate the ground-level energy and the energy separation between adjacent levels. Express your results in joules and in electron volts. Are quantum effects important?

40.34 • **CALC** Show that $\psi(x)$ given by Eq. (40.47) is a solution to Eq. (40.44) with energy $E_0 = \hbar\omega/2$.

40.35 • Chemists use infrared absorption spectra to identify chemicals in a sample. In one sample, a chemist finds that light of wavelength 5.8 μm is absorbed when a molecule makes a transition from its ground harmonic oscillator level to its first excited level. (a) Find the energy of this transition. (b) If the molecule can be treated as a harmonic oscillator with mass 5.6×10^{-26} kg, find the force constant.

40.36 • A harmonic oscillator absorbs a photon of wavelength 6.35 μm when it undergoes a transition from the ground state to the first excited state. What is the ground-state energy, in electron volts, of the oscillator?

40.37 •• The ground-state energy of a harmonic oscillator is 5.60 eV. If the oscillator undergoes a transition from its $n = 3$ to $n = 2$ level by emitting a photon, what is the wavelength of the photon?

40.38 •• While undergoing a transition from the $n = 1$ to the $n = 2$ energy level, a harmonic oscillator absorbs a photon of wavelength 6.50 μm. What is the wavelength of the absorbed photon when this oscillator undergoes a transition (a) from the $n = 2$ to the $n = 3$ energy level and (b) from the $n = 1$ to the $n = 3$ energy level? (c) What is the value of $\sqrt{k'/m}$, the angular oscillation frequency of the corresponding Newtonian oscillator?

40.39 • In Section 40.5 it is shown that for the ground level of a harmonic oscillator, $\Delta x \Delta p_x = \hbar/2$. Do a similar analysis for an excited level that has quantum number n. How does the uncertainty product $\Delta x \, \Delta p_x$ depend on n?

40.40 •• For the ground-level harmonic oscillator wave function $\psi(x)$ given in Eq. (40.47), $|\psi|^2$ has a maximum at $x = 0$. (a) Compute the ratio of $|\psi|^2$ at $x = +A$ to $|\psi|^2$ at $x = 0$, where A is given by Eq. (40.48) with $n = 0$ for the ground level. (b) Compute the ratio of $|\psi|^2$ at $x = +2A$ to $|\psi|^2$ at $x = 0$. In each case is your result consistent with what is shown in Fig. 40.27?

40.41 •• For the sodium atom of Example 40.8, find (a) the ground-state energy; (b) the wavelength of a photon emitted when the $n = 4$ to $n = 3$ transition occurs; (c) the energy difference for any $\Delta n = 1$ transition.

PROBLEMS

40.42 •• CALC Consider the wave packet defined by

$$\psi(x) = \int_0^\infty B(k)\cos kx\,dk$$

Let $B(k) = e^{-\alpha^2k^2}$. (a) The function $B(k)$ has its maximum value at $k = 0$. Let k_h be the value of k at which $B(k)$ has fallen to half its maximum value, and define the width of $B(k)$ as $w_k = k_h$. In terms of α, what is w_k? (b) Use integral tables to evaluate the integral that gives $\psi(x)$. For what value of x is $\psi(x)$ maximum? (c) Define the width of $\psi(x)$ as $w_x = x_h$, where x_h is the positive value of x at which $\psi(x)$ has fallen to half its maximum value. Calculate w_x in terms of α. (d) The momentum p is equal to $hk/2\pi$, so the width of B in momentum is $w_p = hw_k/2\pi$. Calculate the product w_pw_x and compare to the Heisenberg uncertainty principle.

40.43 •• A particle of mass m in a one-dimensional box has the following wave function in the region $x = 0$ to $x = L$:

$$\Psi(x, t) = \frac{1}{\sqrt{2}}\psi_1(x)e^{-iE_1t/\hbar} + \frac{1}{\sqrt{2}}\psi_3(x)e^{-iE_3t/\hbar}$$

Here $\psi_1(x)$ and $\psi_3(x)$ are the normalized stationary-state wave functions for the $n = 1$ and $n = 3$ levels, and E_1 and E_3 are the energies of these levels. The wave function is zero for $x < 0$ and for $x > L$. (a) Find the value of the probability distribution function at $x = L/2$ as a function of time. (b) Find the angular frequency at which the probability distribution function oscillates.

40.44 •• CALC (a) Using the integral in Problem 40.42, determine the wave function $\psi(x)$ for a function $B(k)$ given by

$$B(k) = \begin{cases} 0 & k < 0 \\ 1/k_0, & 0 \le k \le k_0 \\ 0, & k > k_0 \end{cases}$$

This represents an equal combination of all wave numbers between 0 and k_0. Thus $\psi(x)$ represents a particle with average wave number $k_0/2$, with a total spread or uncertainty in wave number of k_0. We will call this spread the *width* w_k of $B(k)$, so $w_k = k_0$. (b) Graph $B(k)$ versus k and $\psi(x)$ versus x for the case $k_0 = 2\pi/L$, where L is a length. Locate the point where $\psi(x)$ has its maximum value and label this point on your graph. Locate the two points closest to this maximum (one on each side of it) where $\psi(x) = 0$, and define the distance along the x-axis between these two points as w_x, the width of $\psi(x)$. Indicate the distance w_x on your graph. What is the value of w_x if $k_0 = 2\pi/L$? (c) Repeat part (b) for the case $k_0 = \pi/L$. (d) The momentum p is equal to $hk/2\pi$, so the width of B in momentum is $w_p = hw_k/2\pi$. Calculate the product w_pw_x for each of the cases $k_0 = 2\pi/L$ and $k_0 = \pi/L$. Discuss your results in light of the Heisenberg uncertainty principle.

40.45 •• CALC Consider a beam of free particles that move with velocity $v = p/m$ in the x-direction and are incident on a potential-energy step $U(x) = 0$, for $x < 0$, and $U(x) = U_0 < E$, for $x > 0$. The wave function for $x < 0$ is $\psi(x) = Ae^{ik_1x} + Be^{-ik_1x}$, representing incident and reflected particles, and for $x > 0$ is $\psi(x) = Ce^{ik_2x}$, representing transmitted particles. Use the conditions that both ψ and its first derivative must be continuous at $x = 0$ to find the constants B and C in terms of k_1, k_2, and A.

40.46 • CALC A particle is in the ground level of a box that extends from $x = 0$ to $x = L$. (a) What is the probability of finding the particle in the region between 0 and $L/4$? Calculate this by integrating $|\psi(x)|^2\,dx$, where ψ is normalized, from $x = 0$ to $x = L/4$. (b) What is the probability of finding the particle in the region $x = L/4$ to $x = L/2$? (c) How do the results of parts (a) and (b) compare? Explain. (d) Add the probabilities calculated in parts (a) and (b). (e) Are your results in parts (a), (b), and (d) consistent with Fig. 40.12b? Explain.

40.47 • **Photon in a Dye Laser.** An electron in a long, organic molecule used in a dye laser behaves approximately like a particle in a box with width 4.18 nm. What is the wavelength of the photon emitted when the electron undergoes a transition (a) from the first excited level to the ground level and (b) from the second excited level to the first excited level?

40.48 •• Consider a particle in a box with rigid walls at $x = 0$ and $x = L$. Let the particle be in the ground level. Calculate the probability $|\psi|^2dx$ that the particle will be found in the interval x to $x + dx$ for (a) $x = L/4$; (b) $x = L/2$; (c) $x = 3L/4$.

40.49 •• Repeat Problem 40.48 for a particle in the first excited level.

40.50 •• CP A particle is confined within a box with perfectly rigid walls at $x = 0$ and $x = L$. Although the magnitude of the instantaneous force exerted on the particle by the walls is infinite and the time over which it acts is zero, the impulse (that involves a product of force and time) is both finite and quantized. Show that the impulse exerted by the wall at $x = 0$ is $(nh/L)\hat{\imath}$ and that the impulse exerted by the wall at $x = L$ is $-(nh/L)\hat{\imath}$. (*Hint:* You may wish to review Section 8.1.)

40.51 •• CALC What is the probability of finding a particle in a box of length L in the region between $x = L/4$ and $x = 3L/4$ when the particle is in (a) the ground level and (b) the first excited level? (*Hint:* Integrate $|\psi(x)|^2\,dx$, where ψ is normalized, between $L/4$ and $3L/4$.) (c) Are your results in parts (a) and (b) consistent with Fig. 40.12b? Explain.

40.52 •• The *penetration distance* η in a finite potential well is the distance at which the wave function has decreased to $1/e$ of the wave function at the classical turning point:

$$\psi(x = L + \eta) = \frac{1}{e}\psi(L)$$

The penetration distance can be shown to be

$$\eta = \frac{\hbar}{\sqrt{2m(U_0 - E)}}$$

The probability of finding the particle beyond the penetration distance is nearly zero. (a) Find η for an electron having a kinetic energy of 13 eV in a potential well with $U_0 = 20$ eV. (b) Find η for a 20.0-MeV proton trapped in a 30.0-MeV-deep potential well.

40.53 •• CALC A fellow student proposes that a possible wave function for a free particle with mass m (one for which the potential-energy function $U(x)$ is zero) is

$$\psi(x) = \begin{cases} e^{+\kappa x}, & x < 0 \\ e^{-\kappa x}, & x \ge 0 \end{cases}$$

where κ is a positive constant. (a) Graph this proposed wave function. (b) Show that the proposed wave function satisfies the Schrödinger equation for $x < 0$ if the energy is $E = -\hbar^2\kappa^2/2m$—

that is, if the energy of the particle is *negative*. (c) Show that the proposed wave function also satisfies the Schrödinger equation for $x \geq 0$ with the same energy as in part (b). (d) Explain why the proposed wave function is nonetheless *not* an acceptable solution of the Schrödinger equation for a free particle. (*Hint:* What is the behavior of the function at $x = 0$?) It is in fact impossible for a free particle (one for which $U(x) = 0$) to have an energy less than zero.

40.54 • An electron with initial kinetic energy 5.5 eV encounters a square potential barrier of height 10.0 eV. What is the width of the barrier if the electron has a 0.50% probability of tunneling through the barrier?

40.55 • **CALC** (a) For the finite potential well of Fig. 40.13, what relationships among the constants A and B of Eq. (40.38) and C and D of Eq. (40.40) are obtained by applying the boundary condition that ψ be continuous at $x = 0$ and at $x = L$? (b) What relationships among A, B, C, and D are obtained by applying the boundary condition that $d\psi/dx$ be continuous at $x = 0$ and at $x = L$?

40.56 • **CP** A harmonic oscillator consists of a 0.020-kg mass on a spring. The oscillation frequency is 1.50 Hz, and the mass has a speed of 0.480 m/s as it passes the equilibrium position. (a) What is the value of the quantum number n for its energy level? (b) What is the difference in energy between the levels E_n and E_{n+1}? Is this difference detectable?

40.57 • For small amplitudes of oscillation the motion of a pendulum is simple harmonic. For a pendulum with a period of 0.500 s, find the ground-level energy and the energy difference between adjacent energy levels. Express your results in joules and in electron volts. Are these values detectable?

40.58 •• **CALC** (a) Show by direct substitution in the Schrödinger equation for the one-dimensional harmonic oscillator that the wave function $\psi_1(x) = A_1 x e^{-\alpha^2 x^2/2}$, where $\alpha^2 = m\omega/\hbar$, is a solution with energy corresponding to $n = 1$ in Eq. (40.46). (b) Find the normalization constant A_1. (c) Show that the probability density has a minimum at $x = 0$ and maxima at $x = \pm 1/\alpha$, corresponding to the classical turning points for the ground state $n = 0$.

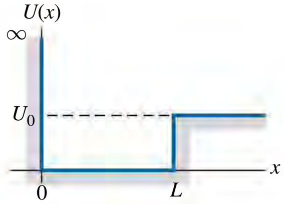
40.59 •• **CP** (a) The wave nature of particles results in the quantum-mechanical situation that a particle confined in a box can assume only wavelengths that result in standing waves in the box, with nodes at the box walls. Use this to show that an electron confined in a one-dimensional box of length L will have energy levels given by

$$E_n = \frac{n^2 h^2}{8mL^2}$$

(*Hint:* Recall that the relationship between the de Broglie wavelength and the speed of a nonrelativistic particle is $mv = h/\lambda$. The energy of the particle is $\frac{1}{2}mv^2$.) (b) If a hydrogen atom is modeled as a one-dimensional box with length equal to the Bohr radius, what is the energy (in electron volts) of the lowest energy level of the electron?

40.60 ••• Consider a potential well defined as $U(x) = \infty$ for $x < 0$, $U(x) = 0$ for $0 < x < L$, and $U(x) = U_0 > 0$ for $x > L$ (**Fig. P40.60**). Consider a particle with mass m and kinetic energy $E < U_0$ that is trapped in the well. (a) The boundary condition at the infinite wall ($x = 0$) is $\psi(0) = 0$. What must the form of the function $\psi(x)$ for $0 < x < L$ be in order to satisfy both the Schrödinger equation and this boundary condition? (b) The wave function must remain finite as $x \to \infty$. What must the form of the function $\psi(x)$ for $x > L$ be in order to satisfy both the Schrödinger equation and this boundary condition at infinity? (c) Impose the boundary conditions that ψ and $d\psi/dx$ are continuous at $x = L$. Show that the energies of the allowed levels are obtained from solutions of the equation $k \cot kL = -\kappa$, where $k = \sqrt{2mE}/\hbar$ and $\kappa = \sqrt{2m(U_0 - E)}/\hbar$.

Figure **P40.60**



40.61 •• **DATA** In your research on new solid-state devices, you are studying a solid-state structure that can be modeled accurately as an electron in a one-dimensional infinite potential well (box) of width L. In one of your experiments, electromagnetic radiation is absorbed in transitions in which the initial state is the $n = 1$ ground state. You measure that light of frequency $f = 9.0 \times 10^{14}$ Hz is absorbed and that the next higher absorbed frequency is 16.9×10^{14} Hz. (a) What is quantum number n for the final state in each of the transitions that leads to the absorption of photons of these frequencies? (b) What is the width L of the potential well? (c) What is the longest wavelength in air of light that can be absorbed by an electron if it is initially in the $n = 1$ state?

40.62 •• **DATA** As an intern at a research lab, you study the transmission of electrons through a potential barrier. You know the height of the barrier, 8.0 eV, but must measure the width L of the barrier. When you measure the tunneling probability T as a function of the energy E of the electron, you get the results shown in the table.

***E* (eV)**	4.0	5.0	6.0	7.0	7.6
T	2.4×10^{-6}	1.5×10^{-5}	1.2×10^{-4}	1.3×10^{-3}	8.1×10^{-3}

(a) For each value of E, calculate the quantities G and κ that appear in Eq. (40.42). Graph $\ln(T/G)$ versus κ. Explain why your data points, when plotted this way, fall close to a straight line. (b) Use the slope of the best-fit straight line to the data in part (a) to calculate L.

40.63 •• **DATA** When low-energy electrons pass through an ionized gas, electrons of certain energies pass through the gas as if the gas atoms weren't there and thus have transmission coefficients (tunneling probabilities) T equal to unity. The gas ions can be modeled approximately as a rectangular barrier. The value of $T = 1$ occurs when an integral or half-integral number of de Broglie wavelengths of the electron as it passes over the barrier equal the width L of the barrier. You are planning an experiment to measure this effect. To assist you in designing the necessary apparatus, you estimate the electron energies E that will result in $T = 1$. You assume a barrier height of 10 eV and a width of 1.8×10^{-10} m. Calculate the three lowest values of E for which $T = 1$.

CHALLENGE PROBLEMS

40.64 ••• **CALC** **The WKB Approximation.** It can be a challenge to solve the Schrödinger equation for the bound-state energy levels of an arbitrary potential well. An alternative approach that can yield good approximate results for the energy levels is the *WKB approximation* (named for the physicists Gregor Wentzel, Hendrik Kramers, and Léon Brillouin, who pioneered its application to quantum mechanics). The WKB approximation begins from three physical statements: (i) According to de Broglie, the magnitude of momentum p of a quantum-mechanical particle is $p = h/\lambda$. (ii) The magnitude of momentum is related to the kinetic energy K by the relationship $K = p^2/2m$. (iii) If there are no nonconservative forces, then in Newtonian mechanics the energy E for a particle is constant and equal at each point to the sum of the kinetic and potential energies at that point: $E = K + U(x)$, where x is the coordinate. (a) Combine these three relationships to show that the wavelength of the particle at a coordinate x can be written as

$$\lambda(x) = \frac{h}{\sqrt{2m[E - U(x)]}}$$

Thus we envision a quantum-mechanical particle in a potential well $U(x)$ as being like a free particle, but with a wavelength $\lambda(x)$ that is a function of position. (b) When the particle moves into a region of increasing potential energy, what happens to its wavelength? (c) At a point where $E = U(x)$, Newtonian mechanics says that the particle has zero kinetic energy and must be instantaneously at rest. Such a point is called a *classical turning point,* since this is where a Newtonian particle must stop its motion and reverse direction. As an example, an object oscillating in simple harmonic motion with amplitude A moves back and forth between the points $x = -A$ and $x = +A$; each of these is a classical turning point, since there the potential energy $\frac{1}{2}k'x^2$ equals the total energy $\frac{1}{2}k'A^2$. In the WKB expression for $\lambda(x)$, what is the wavelength at a classical turning point? (d) For a particle in a box with length L, the walls of the box are classical turning points (see Fig. 40.8). Furthermore, the number of wavelengths that fit within the box must be a half-integer (see Fig. 40.10), so that $L = (n/2)\lambda$ and hence $L/\lambda = n/2$, where $n = 1, 2, 3, \ldots$. [Note that this is a restatement of Eq. (40.29).] The WKB scheme for finding the allowed bound-state energy levels of an *arbitrary* potential well is an extension of these observations. It demands that for an allowed energy E, there must be a half-integer number of wavelengths between the classical turning points for that energy. Since the wavelength in the WKB approximation is not a constant but depends on x, the number of wavelengths between the classical turning points a and b for a given value of the energy is the integral of $1/\lambda(x)$ between those points:

$$\int_a^b \frac{dx}{\lambda(x)} = \frac{n}{2} \quad (n = 1, 2, 3, \ldots)$$

Using the expression for $\lambda(x)$ you found in part (a), show that the *WKB condition for an allowed bound-state energy* can be written as

$$\int_a^b \sqrt{2m[E - U(x)]}\,dx = \frac{nh}{2} \quad (n = 1, 2, 3, \ldots)$$

(e) As a check on the expression in part (d), apply it to a particle in a box with walls at $x = 0$ and $x = L$. Evaluate the integral and show that the allowed energy levels according to the WKB approximation are the same as those given by Eq. (40.31). (*Hint:* Since the walls of the box are infinitely high, the points $x = 0$ and $x = L$ are classical turning points for *any* energy E. Inside the box, the potential energy is zero.) (f) For the finite square well shown in Fig. 40.13, show that the WKB expression given in part (d) predicts the *same* bound-state energies as for an infinite square well of the same width. (*Hint:* Assume $E < U_0$. Then the classical turning points are at $x = 0$ and $x = L$.) This shows that the WKB approximation does a poor job when the potential-energy function changes discontinuously, as for a finite potential well. In the next two problems we consider situations in which the potential-energy function changes gradually and the WKB approximation is much more useful.

40.65 ••• CALC The WKB approximation (see Challenge Problem 40.64) can be used to calculate the energy levels for a harmonic oscillator. In this approximation, the energy levels are the solutions to the equation

$$\int_a^b \sqrt{2m[E - U(x)]}\,dx = \frac{nh}{2} \quad n = 1, 2, 3, \ldots$$

Here E is the energy, $U(x)$ is the potential-energy function, and $x = a$ and $x = b$ are the classical turning points (the points at which E is equal to the potential energy, so the Newtonian kinetic energy would be zero). (a) Determine the classical turning points for a harmonic oscillator with energy E and force constant k'. (b) Carry out the integral in the WKB approximation and show that the energy levels in this approximation are $E_n = \hbar\omega$, where $\omega = \sqrt{k'/m}$ and $n = 1, 2, 3, \ldots$. (*Hint:* Recall that $\hbar = h/2\pi$. A useful standard integral is

$$\int \sqrt{A^2 - x^2}\,dx = \frac{1}{2}\left[x\sqrt{A^2 - x^2} + A^2 \arcsin\left(\frac{x}{|A|}\right)\right]$$

where arcsin denotes the inverse sine function. Note that the integrand is even, so the integral from $-x$ to x is equal to twice the integral from 0 to x.) (c) How do the approximate energy levels found in part (b) compare with the true energy levels given by Eq. (40.46)? Does the WKB approximation give an underestimate or an overestimate of the energy levels?

40.66 ••• CALC Protons, neutrons, and many other particles are made of more fundamental particles called *quarks* and *antiquarks* (the antimatter equivalent of quarks). A quark and an antiquark can form a bound state with a variety of different energy levels, each of which corresponds to a different particle observed in the laboratory. As an example, the ψ particle is a low-energy bound state of a so-called charm quark and its antiquark, with a rest energy of 3097 MeV; the $\psi(2S)$ particle is an excited state of this same quark–antiquark combination, with a rest energy of 3686 MeV. A simplified representation of the potential energy of interaction between a quark and an antiquark is $U(x) = A|x|$, where A is a positive constant and x represents the distance between the quark and the antiquark. You can use the WKB approximation (see Challenge Problem 40.64) to determine the bound-state energy levels for this potential-energy function. In the WKB approximation, the energy levels are the solutions to the equation

$$\int_a^b \sqrt{2m[E - U(x)]}\,dx = \frac{nh}{2} \quad (n = 1, 2, 3, \ldots)$$

Here E is the energy, $U(x)$ is the potential-energy function, and $x = a$ and $x = b$ are the classical turning points (the points at which E is equal to the potential energy, so the Newtonian kinetic energy would be zero). (a) Determine the classical turning points for the potential $U(x) = A|x|$ and for an energy E. (b) Carry out the above integral and show that the allowed energy levels in the WKB approximation are given by

$$E_n = \frac{1}{2m}\left(\frac{3mAh}{4}\right)^{2/3} n^{2/3} \quad (n = 1, 2, 3, \ldots)$$

(*Hint:* The integrand is even, so the integral from $-x$ to x is equal to twice the integral from 0 to x.) (c) Does the difference in energy between successive levels increase, decrease, or remain the same as n increases? How does this compare to the behavior of the energy levels for the harmonic oscillator? For the particle in a box? Can you suggest a simple rule that relates the difference in energy between successive levels to the shape of the potential-energy function?

PASSAGE PROBLEMS

QUANTUM DOTS. A *quantum dot* is a type of crystal so small that quantum effects are significant. One application of quantum dots is in fluorescence imaging, in which a quantum dot is bound to a molecule or structure of interest. When the quantum dot is illuminated with light, it absorbs photons and then re-emits photons at a different wavelength. This phenomenon is called *fluorescence*. The wavelength that a quantum dot emits when stimulated with light depends on the dot's size, so the synthesis of quantum dots with different photon absorption and emission properties may be possible. We can understand many quantum-dot properties via a model in which a particle of mass M (roughly the mass of the electron) is confined to a two-dimensional rigid square box of sides L. In this model, the quantum-dot energy levels are given by $E_{m,n} = (m^2 + n^2)(\pi^2\hbar^2)/2ML^2$, where m and n are integers 1, 2, 3, ...

40.67 According to this model, which statement is true about the energy-level spacing of dots of different sizes? (a) Smaller dots have equally spaced levels, but larger dots have energy levels that get farther apart as the energy increases. (b) Larger dots have greater spacing between energy levels than do smaller dots. (c) Smaller dots have greater spacing between energy levels than do larger dots. (d) The spacing between energy levels is independent of the dot size.

40.68 When a given dot with side length L makes a transition from its first excited state to its ground state, the dot emits green (550 nm) light. If a dot with side length $1.1L$ is used instead, what wavelength is emitted in the same transition, according to this model? (a) 600 nm; (b) 670 nm; (c) 500 nm; (d) 460 nm.

40.69 Dots that are the same size but made from different materials are compared. In the same transition, a dot of material 1 emits a photon of longer wavelength than the dot of material 2 does. Based on this model, what is a possible explanation? (a) The mass of the confined particle in material 1 is greater. (b) The mass of the confined particle in material 2 is greater. (c) The confined particles make more transitions per second in material 1. (d) The confined particles make more transitions per second in material 2.

40.70 One advantage of the quantum dot is that, compared to many other fluorescent materials, excited states have relatively long lifetimes (10 ns). What does this mean for the spread in the energy of the photons emitted by quantum dots? (a) Quantum dots emit photons of more well-defined energies than do other fluorescent materials. (b) Quantum dots emit photons of less well-defined energies than do other fluorescent materials. (c) The spread in the energy is affected by the size of the dot, not by the lifetime. (d) There is no spread in the energy of the emitted photons, regardless of the lifetime.

Answers

Chapter Opening Question ?

(i) When an electron in one of these particles—called *quantum dots*—makes a transition from an excited level to a lower level, it emits a photon whose energy is equal to the difference in energy between the levels. The smaller the quantum dot, the larger the energy spacing between levels and hence the shorter (bluer) the wavelength of the emitted photons. See Example 40.6 (Section 40.3) for more details.

Test Your Understanding Questions

40.1 no Equation (40.19) represents a superposition of wave functions with different values of wave number k and hence different values of energy $E = \hbar^2k^2/2m$. The state that this combined wave function represents is not a state of definite energy, and therefore not a stationary state. Another way to see this is to note that there is a factor $e^{-iEt/\hbar}$ inside the integral in Eq. (40.19), with a different value of E for each value of k. This wave function therefore has a very complicated time dependence, and the probability distribution function $|\Psi(x,t)|^2$ does depend on time.

40.2 (v) Our derivation of the stationary-state wave functions for a particle in a box shows that they are superpositions of waves propagating in opposite directions, just like a standing wave on a string. One wave has momentum in the positive x-direction, while the other wave has an equal magnitude of momentum in the negative x-direction. The *total* x-component of momentum is zero.

40.3 (i) The energy levels are arranged as shown in Fig. 40.15b if $U_0 = 6E_{1-\text{IDW}}$, where $E_{1-\text{IDW}} = \pi^2\hbar^2/2mL^2$ is the ground-level energy of an infinite well. If the well width L is reduced to one-half of its initial value, $E_{1-\text{IDW}}$ increases by a factor of four and so U_0 must also increase by a factor of four. The energies E_1, E_2, and E_3 shown in Fig. 40.15b are all specific fractions of U_0, so they will also increase by a factor of four.

40.4 yes Figure 40.20 shows a possible wave function $\psi(x)$ for tunneling. Since $\psi(x)$ is not zero within the barrier ($0 \le x \le L$), there is some probability that the particle can be found there.

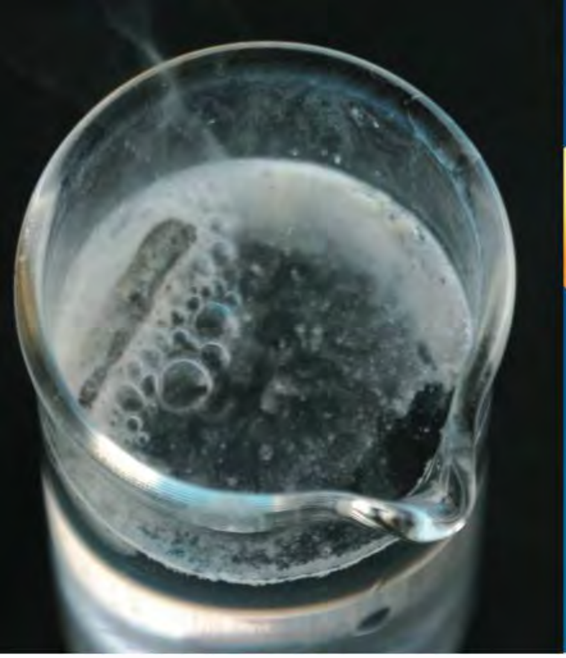
40.5 (ii) If the second photon has a longer wavelength and hence lower energy than the first photon, the difference in energy between the first and second excited levels must be less than the difference between the ground level and the first excited level. This is the case for the hydrogen atom, for which the energy difference between levels decreases as the energy increases (see Fig. 39.24). By contrast, the energy difference between successive levels increases for a particle in a box (see Fig. 40.11b) and is constant for a harmonic oscillator (see Fig. 40.25).

40.6 (v) The value of the energy of a particle in a box must be equal to one of the allowed energy levels, so the measured value will be either E_1 or E_2. *Neither* of these results is guaranteed. If $|C| = |D|$, then E_1 and E_2 are of equal probability. E_1 is the more likely result if $|C| > |D|$, and E_2 is the more likely result if $|C| < |D|$.

Bridging Problem

(a) $$|\Psi(x,t)|^2 = \frac{1}{L}\left[\sin^2\frac{\pi x}{L} + \sin^2\frac{2\pi x}{L} + 2\sin\frac{\pi x}{L}\sin\frac{2\pi x}{L}\cos\left(\frac{(E_2 - E_1)t}{\hbar}\right)\right]$$

(b) no **(d)** $\dfrac{3\pi^2\hbar}{2mL^2}$ **(e)** $\dfrac{0.903\pi^2\hbar}{mL^2}$



? Lithium (with three electrons per atom) is a metal that burns spontaneously in water, while helium (with two electrons per atom) is a gas that undergoes almost no chemical reactions. The additional electron makes lithium behave very differently from helium primarily because (i) the third electron is strongly repelled by electric forces from the other two electrons; (ii) the third electron and larger nucleus make the lithium atom more massive than the helium atom; (iii) there is a limit on the number of electrons that can occupy a given quantum-mechanical state; (iv) the lithium nucleus has more positive charge than a helium nucleus has.

41 QUANTUM MECHANICS II: ATOMIC STRUCTURE

LEARNING GOALS

Looking forward at ...

41.1 How to extend quantum-mechanical calculations to three-dimensional problems.

41.2 How to solve the Schrödinger equation for a particle trapped in a cubical box.

41.3 How to describe the states of a hydrogen atom in terms of quantum numbers.

41.4 How magnetic fields affect the orbital motion of atomic electrons.

41.5 How we know that electrons have their own intrinsic angular momentum.

41.6 How to analyze the structure of many-electron atoms.

41.7 How x rays emitted by atoms reveal their inner structure.

41.8 What happens when the quantum-mechanical states of two particles become entangled.

Looking back at ...

22.3 Gauss's law.

27.7 Magnetic dipole moment.

32.5 Standing electromagnetic waves.

38.2 X-ray production.

39.2, 39.3 Atoms and the Bohr model.

40.1, 40.2, 40.5, 40.6 One-dimensional Schrödinger equation; particle in a box; harmonic-oscillator wave functions; measuring a quantum-mechanical system.

Some physicists claim that all of chemistry is contained in the Schrödinger equation. This is somewhat of an exaggeration, but this equation can teach us a great deal about the chemical behavior of elements, the periodic table, and the nature of chemical bonds.

In order to learn about the quantum-mechanical structure of atoms, we'll first construct a three-dimensional version of the Schrödinger equation. We'll try this equation out by looking at a three-dimensional version of a particle in a box: a particle confined to a cubical volume.

We'll then see that we can learn a great deal about the structure and properties of *all* atoms from the solutions to the Schrödinger equation for the hydrogen atom. These solutions have quantized values of orbital angular momentum; we don't need to impose quantization as we did with the Bohr model. We label the states with a set of quantum numbers, which we'll use later with many-electron atoms as well. We'll find that the electron also has an intrinsic *spin* angular momentum with its own set of quantized values.

We'll also encounter the exclusion principle, a kind of microscopic zoning ordinance that is the key to understanding many-electron atoms. This principle says that no two electrons in an atom can have the same quantum-mechanical state. We'll then use the principles of this chapter to explain the characteristic x-ray spectra of atoms. Finally, we'll end our discussion of quantum mechanics with a look at the curious concept of quantum entanglement and its application to the new science of quantum computing.

41.1 THE SCHRÖDINGER EQUATION IN THREE DIMENSIONS

We have discussed the Schrödinger equation and its applications only for *one-dimensional* problems, the analog of a Newtonian particle moving along a straight line. The straight-line model is adequate for some applications, but to understand atomic structure, we need a three-dimensional generalization.

It's not difficult to guess what the three-dimensional Schrödinger equation should look like. First, the wave function Ψ is a function of time and all three space coordinates (x, y, z). In general, the potential-energy function also depends on all three coordinates and can be written as $U(x, y, z)$. Next, recall from Section 40.1 that the term $-(\hbar^2/2m)\partial^2\Psi/\partial x^2$ in the one-dimensional Schrödinger equation, Eq. (40.20), is related to the kinetic energy of the particle in the state described by the wave function Ψ. For example, if we insert into this term the wave function $\Psi(x, t) = Ae^{ikx}e^{-i\omega t}$ for a free particle with magnitude of momentum $p = \hbar k$ and kinetic energy $K = p^2/2m$, we obtain $-(\hbar^2/2m)(ik)^2Ae^{ikx}e^{-i\omega t} = (\hbar^2k^2/2m)Ae^{ikx}e^{-i\omega t} = (p^2/2m)\Psi(x, t) = K\Psi(x, t)$. If the particle can move in three dimensions, its momentum has three components (p_x, p_y, p_z) and its kinetic energy is

$$K = \frac{p_x{}^2}{2m} + \frac{p_y{}^2}{2m} + \frac{p_z{}^2}{2m} \tag{41.1}$$

These observations, taken together, suggest that the correct generalization of the Schrödinger equation to three dimensions is

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi(x, y, z, t)}{\partial x^2} + \frac{\partial^2\Psi(x, y, z, t)}{\partial y^2} + \frac{\partial^2\Psi(x, y, z, t)}{\partial z^2}\right)$$
$$+ U(x, y, z)\Psi(x, y, z, t) = i\hbar\frac{\partial\Psi(x, y, z, t)}{\partial t} \tag{41.2}$$

(general three-dimensional Schrödinger equation)

The three-dimensional wave function $\Psi(x, y, z, t)$ has a similar interpretation as in one dimension. The wave function itself is a complex quantity with both a real part and an imaginary part, but $|\Psi(x, y, z, t)|^2$—the square of its absolute value, equal to the product of $\Psi(x, y, z, t)$ and its complex conjugate $\Psi^*(x, y, z, t)$—is real and either positive or zero at every point in space. We interpret $|\Psi(x, y, z, t)|^2\,dV$ as the *probability* of finding the particle within a small volume dV centered on the point (x, y, z) at time t, so $|\Psi(x, y, z, t)|^2$ is the *probability distribution function* in three dimensions. The *normalization condition* on the wave function is that the probability that the particle is *somewhere* in space is exactly 1. Hence the integral of $|\Psi(x, y, z, t)|^2$ over all space must equal 1:

$$\int |\Psi(x, y, z, t)|^2\,dV = 1 \qquad \text{(normalization condition in three dimensions)} \tag{41.3}$$

If the wave function $\Psi(x, y, z, t)$ represents a state of a definite energy E—that is, a stationary state—we can write it as the product of a spatial wave function $\psi(x, y, z)$ and a function of time $e^{-iEt/\hbar}$:

$$\Psi(x, y, z, t) = \psi(x, y, z)e^{-iEt/\hbar} \qquad \text{(time-dependent wave function for a state of definite energy)} \tag{41.4}$$

(Compare this to Eq. (40.21) for a one-dimensional state of definite energy.) If we substitute Eq. (41.4) into Eq. (41.2), the right-hand side of the equation becomes $i\hbar\psi(x, y, z)(-iE/\hbar)e^{-iEt/\hbar} = E\psi(x, y, z)e^{-iEt/\hbar}$. We can then divide both sides by the factor $e^{-iEt/\hbar}$, leaving the *time-independent* Schrödinger equation in three dimensions for a stationary state:

Time-independent three-dimensional Schrödinger equation:

Planck's constant divided by 2π — Time-independent wave function

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi(x, y, z)}{\partial x^2} + \frac{\partial^2\psi(x, y, z)}{\partial y^2} + \frac{\partial^2\psi(x, y, z)}{\partial z^2}\right)$$
$$+ U(x, y, z)\psi(x, y, z) = E\psi(x, y, z) \tag{41.5}$$

Particle's mass — Potential-energy function — Energy of state

The probability distribution function for a stationary state is just the square of the absolute value of the spatial wave function: $|\psi(x, y, z)e^{-iEt/\hbar}|^2 = \psi^*(x, y, z)e^{+iEt/\hbar}\psi(x, y, z)e^{-iEt/\hbar} = |\psi(x, y, z)|^2$. Note that this doesn't depend on time. (As we discussed in Section 40.1, that's why we call these states *stationary*.) Hence for a stationary state the wave function normalization condition, Eq. (41.3), becomes

$$\int |\psi(x, y, z)|^2\, dV = 1 \qquad \text{(normalization condition for a stationary state in three dimensions)} \qquad (41.6)$$

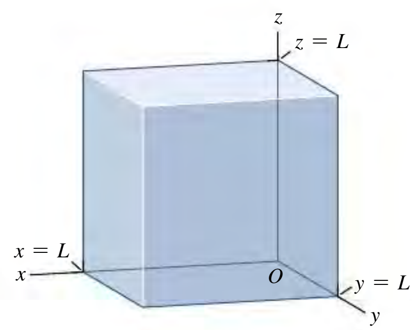
We won't pretend that we have *derived* Eqs. (41.2) and (41.5). Like their one-dimensional versions, these equations have to be tested by comparison of their predictions with experimental results. Happily, Eqs. (41.2) and (41.5) both pass this test with flying colors, so we are confident that they *are* the correct equations.

An important topic that we will address in this chapter is the solutions for Eq. (41.5) for the stationary states of the hydrogen atom. The potential-energy function for an electron in a hydrogen atom is *spherically symmetric*; it depends only on the distance $r = (x^2 + y^2 + z^2)^{1/2}$ from the origin of coordinates. To take advantage of this symmetry, it's best to use *spherical coordinates* rather than the Cartesian coordinates (x, y, z) to solve the Schrödinger equation for the hydrogen atom. Before introducing these new coordinates and investigating the hydrogen atom, it's useful to look at the three-dimensional version of the particle in a box that we considered in Section 40.2. Solving this simpler problem will give us insight into the more complicated stationary states found in atomic physics.

TEST YOUR UNDERSTANDING OF SECTION 41.1 In a certain region of space the potential-energy function for a quantum-mechanical particle is zero. In this region the wave function $\psi(x, y, z)$ for a certain stationary state is real and satisfies $\partial^2\psi/\partial x^2 > 0$, $\partial^2\psi/\partial y^2 > 0$, and $\partial^2\psi/\partial z^2 > 0$. The particle has a definite energy E that is positive. What can you conclude about $\psi(x, y, z)$ in this region? (i) It must be positive; (ii) it must be negative; (iii) it must be zero; (iv) not enough information given to decide. ❙

41.2 PARTICLE IN A THREE-DIMENSIONAL BOX

41.1 A particle is confined in a cubical box with walls at $x = 0$, $x = L$, $y = 0$, $y = L$, $z = 0$, and $z = L$.



Consider a particle enclosed within a cubical box of side L. This could represent an electron that's free to move anywhere within the interior of a solid metal cube but cannot escape the cube. We'll choose the origin to be at one corner of the box, with the x-, y-, and z-axes along edges of the box. Then the particle is confined to the region $0 \le x \le L$, $0 \le y \le L$, $0 \le z \le L$ (**Fig. 41.1**). What are the stationary states of this system?

As for the model of a particle in a one-dimensional box that we considered in Section 40.2, we'll say that the potential energy is zero inside the box but infinite outside. Hence the spatial wave function $\psi(x, y, z)$ must be zero outside the box in order that the term $U(x, y, z)\psi(x, y, z)$ in the time-independent Schrödinger equation, Eq. (41.5), not be infinite. Consequently the probability distribution function $|\psi(x, y, z)|^2$ is zero outside the box, and the probability that the particle will be found there is zero. Inside the box, the spatial wave function for a stationary state obeys the time-independent Schrödinger equation, Eq. (41.5), with $U(x, y, z) = 0$:

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi(x, y, z)}{\partial x^2} + \frac{\partial^2\psi(x, y, z)}{\partial y^2} + \frac{\partial^2\psi(x, y, z)}{\partial z^2}\right) = E\psi(x, y, z)$$

(particle in a three-dimensional box) (41.7)

In order for the wave function to be continuous from the inside to the outside of the box, $\psi(x, y, z)$ must equal zero on the walls. Hence our boundary conditions are that $\psi(x, y, z) = 0$ at $x = 0$, $x = L$, $y = 0$, $y = L$, $z = 0$, and $z = L$.

Guessing a solution to a complicated partial differential equation like Eq. (41.7) seems like quite a challenge. To make progress, recall that we wrote the time-*dependent* wave function for a stationary state as the product of one function that depends on only the spatial coordinates x, y, and z and a second function that depends on only the time t: $\Psi(x, y, z, t) = \psi(x, y, z)e^{-iEt/\hbar}$. In the same way, let's try a technique called *separation of variables:* We'll write the spatial wave function $\psi(x, y, z)$ as a product of one function X that depends on only x, a second function Y that depends on only y, and a third function Z that depends on only z:

$$\psi(x, y, z) = X(x)Y(y)Z(z) \tag{41.8}$$

If we substitute Eq. (41.8) into Eq. (41.7), we get

$$\begin{aligned} &-\frac{\hbar^2}{2m}\left(Y(y)Z(z)\frac{d^2X(x)}{dx^2} + X(x)Z(z)\frac{d^2Y(y)}{dy^2} + X(x)Y(y)\frac{d^2Z(z)}{dz^2}\right) \\ &= EX(x)Y(y)Z(z) \end{aligned} \tag{41.9}$$

The partial derivatives in Eq. (41.7) have become ordinary derivatives since they act on functions of a single variable. Now we divide both sides of Eq. (41.9) by the product $X(x)Y(y)Z(z)$:

$$\left(-\frac{\hbar^2}{2m}\frac{1}{X(x)}\frac{d^2X(x)}{dx^2}\right) + \left(-\frac{\hbar^2}{2m}\frac{1}{Y(y)}\frac{d^2Y(y)}{dy^2}\right) + \left(-\frac{\hbar^2}{2m}\frac{1}{Z(z)}\frac{d^2Z(z)}{dz^2}\right) = E \tag{41.10}$$

The right-hand side of Eq. (41.10) is the energy of the stationary state. Since E is a constant that does not depend on the values of x, y, and z, the left-hand side of the equation must also be independent of the values of x, y, and z. Hence the first term in parentheses on the left-hand side of Eq. (41.10) must equal a constant that doesn't depend on x, the second term in parentheses must equal another constant that doesn't depend on y, and the third term in parentheses must equal a third constant that doesn't depend on z. Let's call these constants E_X, E_Y, and E_Z, respectively. We then have a separate equation for each of the three functions $X(x)$, $Y(y)$, and $Z(z)$:

$$-\frac{\hbar^2}{2m}\frac{d^2X(x)}{dx^2} = E_X X(x) \tag{41.11a}$$

$$-\frac{\hbar^2}{2m}\frac{d^2Y(y)}{dy^2} = E_Y Y(y) \tag{41.11b}$$

$$-\frac{\hbar^2}{2m}\frac{d^2Z(z)}{dz^2} = E_Z Z(z) \tag{41.11c}$$

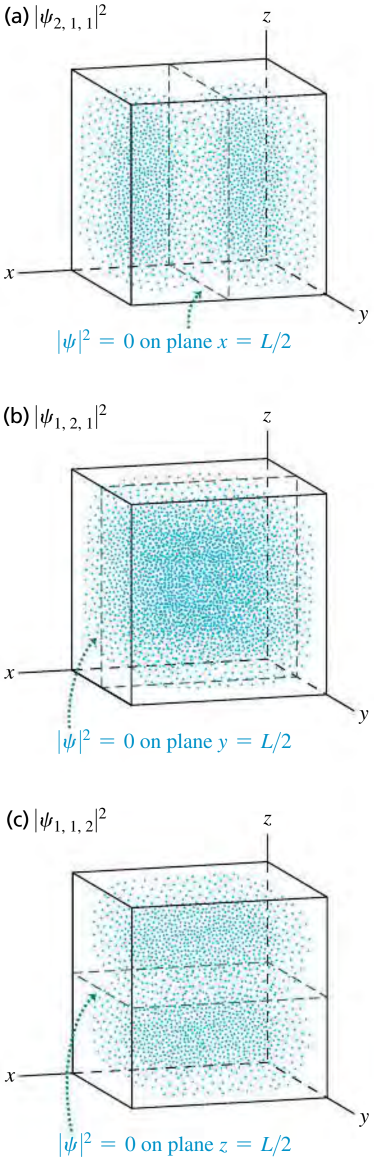
To satisfy the boundary conditions that $\psi(x, y, z) = X(x)Y(y)Z(z)$ be equal to zero on the walls of the box, we demand that $X(x) = 0$ at $x = 0$ and $x = L$, $Y(y) = 0$ at $y = 0$ and $y = L$, and $Z(z) = 0$ at $z = 0$ and $z = L$.

How can we interpret the three constants E_X, E_Y, and E_Z in Eqs. (41.11)? From Eq. (41.10), they are related to the energy E by

$$E_X + E_Y + E_Z = E \tag{41.12}$$

Equation (41.12) should remind you of Eq. (41.1) in Section 41.1, which states that the kinetic energy of a particle is the sum of contributions coming from its x-, y-, and z-components of momentum. Hence the constants E_X, E_Y, and E_Z tell us how much of the particle's energy is due to motion along each of the three coordinate axes. (Inside the box the potential energy is zero, so the particle's energy is purely kinetic.)

41.2 Probability distribution function $|\psi_{n_X,n_Y,n_Z}(x, y, z)|^2$ for (n_X, n_Y, n_Z) equal to (a) (2, 1, 1), (b) (1, 2, 1), and (c) (1, 1, 2). The value of $|\psi|^2$ is proportional to the density of dots. The wave function is zero on the walls of the box and on a midplane of the box, so $|\psi|^2 = 0$ at these locations.



Equations (41.11) represent an enormous simplification; we've reduced the problem of solving a fairly complex *partial* differential equation with three independent variables to the much simpler problem of solving three separate *ordinary* differential equations with one independent variable each. What's more, each of these ordinary differential equations is the same as the time-independent Schrödinger equation for a particle in a *one-dimensional* box, Eq. (40.25), and with exactly the same boundary conditions at 0 and L. (The only differences are that some of the quantities are labeled by different symbols.) By comparing with our work in Section 40.2, you can see that the solutions to Eqs. (41.11) are

$$X_{n_X}(x) = C_X \sin\frac{n_X \pi x}{L} \quad (n_X = 1, 2, 3, \ldots) \tag{41.13a}$$

$$Y_{n_Y}(y) = C_Y \sin\frac{n_Y \pi y}{L} \quad (n_Y = 1, 2, 3, \ldots) \tag{41.13b}$$

$$Z_{n_Z}(z) = C_Z \sin\frac{n_Z \pi z}{L} \quad (n_Z = 1, 2, 3, \ldots) \tag{41.13c}$$

where C_X, C_Y, and C_Z are constants. The corresponding values of E_X, E_Y, and E_Z are

$$E_X = \frac{{n_X}^2 \pi^2 \hbar^2}{2mL^2} \quad (n_X = 1, 2, 3, \ldots) \tag{41.14a}$$

$$E_Y = \frac{{n_Y}^2 \pi^2 \hbar^2}{2mL^2} \quad (n_Y = 1, 2, 3, \ldots) \tag{41.14b}$$

$$E_Z = \frac{{n_Z}^2 \pi^2 \hbar^2}{2mL^2} \quad (n_Z = 1, 2, 3, \ldots) \tag{41.14c}$$

There is only one quantum number n for the one-dimensional particle in a box, but *three* quantum numbers n_X, n_Y, and n_Z for the three-dimensional box. If we substitute Eqs. (41.13) back into Eq. (41.8) for the total spatial wave function, $\psi(x, y, z) = X(x)Y(y)Z(z)$, we get the following stationary-state wave functions for a particle in a three-dimensional cubical box:

$$\psi_{n_X,n_Y,n_Z}(x, y, z) = C \sin\frac{n_X \pi x}{L} \sin\frac{n_Y \pi y}{L} \sin\frac{n_Z \pi z}{L}$$
$$(n_X = 1, 2, 3, \ldots;\ n_Y = 1, 2, 3, \ldots;\ n_Z = 1, 2, 3, \ldots) \tag{41.15}$$

where $C = C_X C_Y C_Z$. The value of the constant C is determined by the normalization condition, Eq. (41.6).

In Section 40.2 we saw that the stationary-state wave functions for a particle in a one-dimensional box were analogous to standing waves on a string. In a similar way, the *three*-dimensional wave functions given by Eq. (41.15) are analogous to standing electromagnetic waves in a cubical cavity like the interior of a microwave oven (see Section 32.5). In a microwave oven there are "dead spots" where the wave intensity is zero, corresponding to the nodes of the standing wave. (The moving platform in a microwave oven ensures even cooking by making sure that no part of the food sits at any "dead spot.") In a similar fashion, the probability distribution function corresponding to Eq. (41.15) can have "dead spots" where there is zero probability of finding the particle. As an example, consider the case $(n_X, n_Y, n_Z) = (2, 1, 1)$. From Eq. (41.15), the probability distribution function for this case is

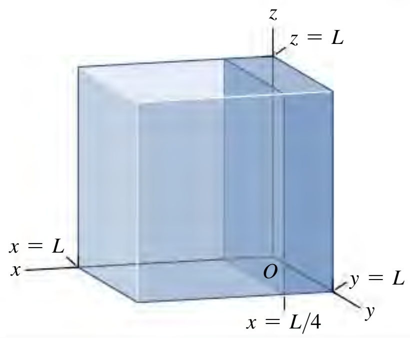
$$|\psi_{2,1,1}(x, y, z)|^2 = |C|^2 \sin^2\frac{2\pi x}{L} \sin^2\frac{\pi y}{L} \sin^2\frac{\pi z}{L}$$

As **Fig. 41.2a** shows, this probability distribution function is zero on the plane $x = L/2$, where $\sin^2(2\pi x/L) = \sin^2\pi = 0$. The particle is most likely to be found near where all three of the sine-squared functions are greatest, at $(x, y, z) = (L/4, L/2, L/2)$ or $(x, y, z) = (3L/4, L/2, L/2)$. Figures 41.2b and 41.2c show the similar cases $(n_X, n_Y, n_Z) = (1, 2, 1)$ and $(n_X, n_Y, n_Z) = (1, 1, 2)$. For higher values of the quantum numbers n_X, n_Y, and n_Z there are additional planes on which the probability distribution function equals zero, just as the probability distribution function $|\psi(x)|^2$ for a one-dimensional box has more zeros for higher values of n (see Fig. 40.12).

EXAMPLE 41.1 PROBABILITY IN A THREE-DIMENSIONAL BOX

(a) Find the value of the constant C that normalizes the wave function of Eq. (41.15). (b) Find the probability that the particle will be found somewhere in the region $0 \le x \le L/4$ (**Fig. 41.3**) for the cases (i) $(n_X, n_Y, n_Z) = (1, 2, 1)$, (ii) $(n_X, n_Y, n_Z) = (2, 1, 1)$, and (iii) $(n_X, n_Y, n_Z) = (3, 1, 1)$.

41.3 What is the probability that the particle is in the dark-colored quarter of the box?



SOLUTION

IDENTIFY and SET UP: Equation (41.6) tells us that to normalize the wave function, we have to choose the value of C so that the integral of the probability distribution function $|\psi_{n_X, n_Y, n_Z}(x, y, z)|^2$ over the volume within the box equals 1. (The integral is actually over *all* space, but the particle-in-a-box wave functions are zero outside the box.)

The probability of finding the particle within a certain volume within the box equals the integral of the probability distribution function over that volume. Hence in part (b) we'll integrate $|\psi_{n_X, n_Y, n_Z}(x, y, z)|^2$ for the given values of (n_X, n_Y, n_Z) over the volume $0 \le x \le L/4$, $0 \le y \le L$, $0 \le z \le L$.

EXECUTE: (a) From Eq. (41.15),

$$|\psi_{n_X,n_Y,n_Z}(x, y, z)|^2 = |C|^2 \sin^2 \frac{n_X \pi x}{L} \sin^2 \frac{n_Y \pi y}{L} \sin^2 \frac{n_Z \pi z}{L}$$

Hence the normalization condition is

$$\begin{aligned}
&\int |\psi_{n_X,n_Y,n_Z}(x, y, z)|^2 \, dV \\
&= |C|^2 \int_{x=0}^{x=L} \int_{y=0}^{y=L} \int_{z=0}^{z=L} \sin^2 \frac{n_X \pi x}{L} \sin^2 \frac{n_Y \pi y}{L} \sin^2 \frac{n_Z \pi z}{L} \, dx\, dy\, dz \\
&= |C|^2 \left(\int_{x=0}^{x=L} \sin^2 \frac{n_X \pi x}{L} dx \right) \left(\int_{y=0}^{y=L} \sin^2 \frac{n_Y \pi y}{L} dy \right) \\
&\quad \times \left(\int_{z=0}^{z=L} \sin^2 \frac{n_Z \pi z}{L} dz \right) \\
&= 1
\end{aligned}$$

We can use the identity $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$ and the variable substitution $\theta = n_X \pi x/L$ to show that

$$\begin{aligned}
\int \sin^2 \frac{n_X \pi x}{L} dx &= \frac{L}{2n_X\pi} \left[\frac{n_X \pi x}{L} - \tfrac{1}{2} \sin\left(\frac{2n_X \pi x}{L} \right) \right] \\
&= \frac{x}{2} - \frac{L}{4n_X\pi} \sin\left(\frac{2n_X \pi x}{L} \right)
\end{aligned}$$

If we evaluate this integral between $x = 0$ and $x = L$, the result is $L/2$ (recall that $\sin 0 = 0$ and $\sin 2n_X\pi = 0$ for any integer n_X).

The y- and z-integrals each yield the same result, so the normalization condition is

$$|C|^2 \left(\frac{L}{2} \right) \left(\frac{L}{2} \right) \left(\frac{L}{2} \right) = |C|^2 \left(\frac{L}{2} \right)^3 = 1$$

or $|C|^2 = (2/L)^3$. If we choose C to be real and positive, then $C = (2/L)^{3/2}$.

(b) We have the same y- and z-integrals as in part (a), but now the limits of integration on the x-integral are $x = 0$ and $x = L/4$:

$$\begin{aligned}
P &= \int_{0 \le x \le L/4} |\psi_{n_X,n_Y,n_Z}|^2 \, dV \\
&= |C|^2 \left(\int_{x=0}^{x=L/4} \sin^2 \frac{n_X \pi x}{L} dx \right) \left(\int_{y=0}^{y=L} \sin^2 \frac{n_Y \pi y}{L} dy \right) \\
&\quad \times \left(\int_{z=0}^{z=L} \sin^2 \frac{n_Z \pi z}{L} dz \right)
\end{aligned}$$

The x-integral is

$$\begin{aligned}
\int_{x=0}^{x=L/4} \sin^2 \frac{n_X \pi x}{L} dx &= \left[\frac{x}{2} - \frac{L}{4n_X\pi} \sin\left(\frac{2n_X \pi x}{L} \right) \right] \Bigg|_{x=0}^{x=L/4} \\
&= \frac{L}{8} - \frac{L}{4n_X\pi} \sin\left(\frac{n_X \pi}{2} \right)
\end{aligned}$$

Hence the probability of finding the particle somewhere in the region $0 \le x \le L/4$ is

$$\begin{aligned}
P &= \left(\frac{2}{L} \right)^3 \left[\frac{L}{8} - \frac{L}{4n_X\pi} \sin\left(\frac{n_X \pi}{2} \right) \right] \left(\frac{L}{2} \right) \left(\frac{L}{2} \right) \\
&= \tfrac{1}{4} - \frac{1}{2n_X\pi} \sin\left(\frac{n_X \pi}{2} \right)
\end{aligned}$$

Continued

This depends only on the value of n_X, not on n_Y or n_Z. Hence for the three cases we have

$$\text{(i) } n_X = 1\text{: } P = \tfrac{1}{4} - \frac{1}{2(1)\pi}\sin\left(\frac{\pi}{2}\right) = \tfrac{1}{4} - \frac{1}{2\pi}(1)$$

$$= \tfrac{1}{4} - \frac{1}{2\pi} = 0.091$$

$$\text{(ii) } n_X = 2\text{: } P = \tfrac{1}{4} - \frac{1}{2(2)\pi}\sin\left(\frac{2\pi}{2}\right) = \tfrac{1}{4} - \frac{1}{4\pi}\sin\pi$$

$$= \tfrac{1}{4} - 0 = 0.250$$

$$\text{(iii) } n_X = 3\text{: } P = \tfrac{1}{4} - \frac{1}{2(3)\pi}\sin\left(\frac{3\pi}{2}\right) = \tfrac{1}{4} - \frac{1}{6\pi}(-1)$$

$$= \tfrac{1}{4} + \frac{1}{6\pi} = 0.303$$

EVALUATE: You can see why the probabilities in part (b) are different by looking at part (b) of Fig. 40.12, which shows $\sin^2 n_X\pi x/L$ for $n_X = 1, 2,$ and 3. For $n_X = 2$ the area under the curve between $x = 0$ and $x = L/4$ (equal to the integral between these two points) is exactly $\frac{1}{4}$ of the total area between $x = 0$ and $x = L$. For $n_X = 1$ the area between $x = 0$ and $x = L/4$ is less than $\frac{1}{4}$ of the total area, and for $n_X = 3$ it is greater than $\frac{1}{4}$ of the total area.

Energy Levels, Degeneracy, and Symmetry

From Eqs. (41.12) and (41.14), the allowed energies for a particle of mass m in a cubical box of side L are

Quantum numbers n_X, n_Y, n_Z can each equal 1, 2, 3, ...

Planck's constant divided by 2π

Energy levels, particle in a three-dimensional cubical box

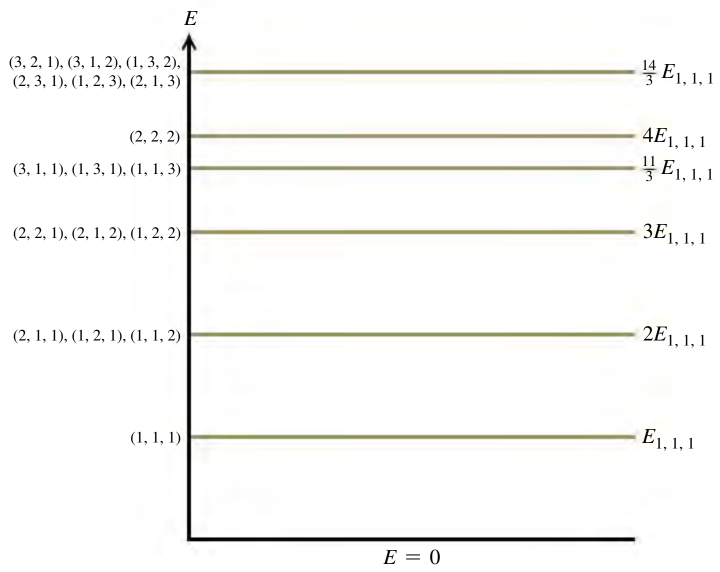
$$E_{n_X,n_Y,n_Z} = \frac{(n_X{}^2 + n_Y{}^2 + n_Z{}^2)\pi^2\hbar^2}{2mL^2} \qquad (41.16)$$

Particle's mass

Length of each side of box

Figure 41.4 shows the six lowest energy levels given by Eq. (41.16). Note that most energy levels correspond to more than one set of quantum numbers (n_X, n_Y, n_Z) and hence to more than one quantum state. Having two or more distinct quantum states with the same energy is called **degeneracy,** and states with the same energy are said to be **degenerate.** For example, Fig. 41.4 shows that the states $(n_X, n_Y, n_Z) = (2, 1, 1)$, $(1, 2, 1)$, and $(1, 1, 2)$ are degenerate. By comparison, for a particle in a one-dimensional box there is just one state for each energy level (see Fig. 40.11a) and no degeneracy.

The reason the cubical box exhibits degeneracy is that it is *symmetric:* All sides of the box have the same dimensions. As an illustration, Fig. 41.2 shows the probability distribution functions for the three states $(n_X, n_Y, n_Z) = (2, 1, 1)$,



41.4 Energy-level diagram for a particle in a three-dimensional cubical box. We label each level with the quantum numbers of the states (n_X, n_Y, n_Z) with that energy. Several of the levels are degenerate (more than one state has the same energy). The lowest (ground) level, $(n_X, n_Y, n_Z) = (1, 1, 1)$, has energy $E_{1,1,1} = (1^2 + 1^2 + 1^2)\pi^2\hbar^2/2mL^2 = 3\pi^2\hbar^2/2mL^2$; we show the energies of the other levels as multiples of $E_{1,1,1}$.

$(1, 2, 1)$, and $(1, 1, 2)$. You can transform any one of these three states into a different one by simply rotating the cubical box by 90°. This rotation doesn't change the energy, so the three states are degenerate.

Since degeneracy is a consequence of symmetry, we can remove the degeneracy by making the box asymmetric. We do this by giving the three sides of the box different lengths L_X, L_Y, and L_Z. If we repeat the steps that we followed to solve the time-independent Schrödinger equation, we find that the energy levels are given by

$$E_{n_X,n_Y,n_Z} = \left(\frac{n_X^2}{L_X^2} + \frac{n_Y^2}{L_Y^2} + \frac{n_Z^2}{L_Z^2}\right)\frac{\pi^2\hbar^2}{2m} \quad \begin{pmatrix} n_X = 1, 2, 3, \ldots; \\ n_Y = 1, 2, 3, \ldots; \\ n_Z = 1, 2, 3, \ldots \end{pmatrix} \tag{41.17}$$

(energy levels, particle in a three-dimensional box with sides of length L_X, L_Y, and L_Z)

If L_X, L_Y, and L_Z are all different, the states $(n_X, n_Y, n_Z) = (2, 1, 1)$, $(1, 2, 1)$, and $(1, 1, 2)$ have different energies and hence are no longer degenerate. Note that Eq. (41.17) reduces to Eq. (41.16) if $L_X = L_Y = L_Z = L$.

Returning to a particle in a three-dimensional cubical box, let's summarize the differences from the one-dimensional case that we examined in Section 40.2:

- We can write the wave function for a three-dimensional stationary state as a product of three functions, one for each spatial coordinate. Only a single function of the coordinate x is needed in one dimension.
- In the three-dimensional case, three quantum numbers are needed to describe each stationary state. Only one quantum number is needed in the one-dimensional case.
- Most of the energy levels for the three-dimensional case are degenerate: More than one stationary state has this energy. There is no degeneracy in the one-dimensional case.
- For a stationary state of the three-dimensional case, there are surfaces on which the probability distribution function $|\psi|^2$ is zero. In the one-dimensional case there are positions on the x-axis where $|\psi|^2$ is zero.

We'll see these same features in the following section for a three-dimensional situation that's more realistic than a particle in a cubical box: a hydrogen atom in which a negatively charged electron orbits a positively charged nucleus.

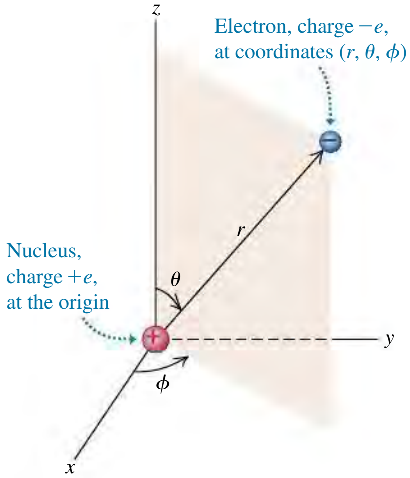
TEST YOUR UNDERSTANDING OF SECTION 41.2 Rank the following states of a particle in a cubical box of side L in order from highest to lowest energy: (i) $(n_X, n_Y, n_Z) = (2, 3, 2)$; (ii) $(n_X, n_Y, n_Z) = (4, 1, 1)$; (iii) $(n_X, n_Y, n_Z) = (2, 2, 3)$; (iv) $(n_X, n_Y, n_Z) = (1, 3, 3)$. ❙

41.3 THE HYDROGEN ATOM

Let's continue the discussion of the hydrogen atom that we began in Chapter 39. In the Bohr model, electrons move in circular orbits like Newtonian particles, but with quantized values of angular momentum. While this model gave the correct energy levels of the hydrogen atom, as deduced from spectra, it had many conceptual difficulties. It mixed classical physics with new and seemingly contradictory concepts. It provided no insight into the process by which photons are emitted and absorbed. It could not be generalized to atoms with more than one electron. It predicted the wrong magnetic properties for the hydrogen atom. And perhaps most important, its picture of the electron as a localized point particle was inconsistent with the more general view we developed in Chapters 39 and 40. To go beyond the Bohr model, let's apply the Schrödinger equation to find the wave functions for stationary states (states of definite energy) of the hydrogen atom. As in Section 39.3, we include the motion of the nucleus by simply replacing the electron mass m with the reduced mass m_r.

The Schrödinger Equation for the Hydrogen Atom

41.5 The Schrödinger equation for the hydrogen atom can be solved most readily by using spherical coordinates.



We discussed the three-dimensional version of the Schrödinger equation in Section 41.1. The potential-energy function is *spherically symmetric:* It depends only on the distance $r = (x^2 + y^2 + z^2)^{1/2}$ from the origin of coordinates:

$$U(r) = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r} \qquad (41.18)$$

The hydrogen-atom problem is best formulated in spherical coordinates (r, θ, ϕ), shown in **Fig. 41.5**; the spherically symmetric potential-energy function depends only on r, not on θ or ϕ. The Schrödinger equation with this potential-energy function can be solved exactly; the solutions are combinations of familiar functions. Without going into a lot of detail, we can describe the most important features of the procedure and the results.

First, we find the solutions by using the same method of separation of variables that we employed for a particle in a cubical box in Section 41.2. We express the wave function $\psi(r, \theta, \phi)$ as a product of three functions, each one a function of only one of the three coordinates:

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi) \qquad (41.19)$$

That is, the function $R(r)$ depends on only r, $\Theta(\theta)$ depends on only θ, and $\Phi(\phi)$ depends on only ϕ. Just as for a particle in a three-dimensional box, when we substitute Eq. (41.19) into the Schrödinger equation, we get three separate ordinary differential equations. One equation involves only r and $R(r)$, a second involves only θ and $\Theta(\theta)$, and a third involves only ϕ and $\Phi(\phi)$:

$$-\frac{\hbar^2}{2m_r r^2}\frac{d}{dr}\left(r^2\frac{dR(r)}{dr}\right) + \left(\frac{\hbar^2 l(l+1)}{2m_r r^2} + U(r)\right)R(r) = ER(r) \qquad (41.20a)$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta(\theta)}{d\theta}\right) + \left(l(l+1) - \frac{m_l^{\,2}}{\sin^2\theta}\right)\Theta(\theta) = 0 \qquad (41.20b)$$

$$\frac{d^2\Phi(\phi)}{d\phi^2} + m_l^{\,2}\Phi(\phi) = 0 \qquad (41.20c)$$

CAUTION Two uses of the symbol m Don't confuse the constant m_l in Eqs. (41.20b) and (41.20c) with the similar symbol m_r for the reduced mass of the electron and nucleus (see Section 39.3). The constant m_l is a dimensionless number; the reduced mass m_r has units of kilograms.

In Eqs. (41.20) E is the energy of the stationary state and l and m_l are constants that we'll discuss later.

We won't attempt to solve this set of three equations, but we can describe how it's done. As for the particle in a cubical box, the physically acceptable solutions of these three equations are determined by boundary conditions. The radial function $R(r)$ in Eq. (41.20a) must approach zero at large r, because we are describing *bound states* of the electron that are localized near the nucleus. This is analogous to the requirement that the harmonic-oscillator wave functions (see Section 40.5) must approach zero at large x. The angular functions $\Theta(\theta)$ and $\Phi(\phi)$ in Eqs. (41.20b) and (41.20c) must be *finite* for all relevant values of the angles. For example, there are solutions of the Θ equation that become infinite at $\theta = 0$ and $\theta = \pi$; these are unacceptable, since $\psi(r, \theta, \phi)$ must be normalizable. Furthermore, the angular function $\Phi(\phi)$ in Eq. (41.20c) must be *periodic*. For example, (r, θ, ϕ) and $(r, \theta, \phi + 2\pi)$ describe the same point, so $\Phi(\phi + 2\pi)$ must equal $\Phi(\phi)$.

The allowed radial functions $R(r)$ turn out to be an exponential function $e^{-\alpha r}$ (where α is positive) multiplied by a polynomial in r. The functions $\Theta(\theta)$ are polynomials containing various powers of $\sin\theta$ and $\cos\theta$, and the functions $\Phi(\phi)$ are simply proportional to $e^{im_l\phi}$, where $i = \sqrt{-1}$ and m_l is an integer that may be positive, zero, or negative.

In the process of finding solutions that satisfy the boundary conditions, we also find the corresponding energy levels. We denote the energies of these levels [E in

Eq. (41.20a)] by E_n $(n = 1, 2, 3, \ldots)$. These turn out to be *identical* to those from the Bohr model, as given by Eq. (39.15), with the electron rest mass m replaced by the reduced mass m_r. Rewriting that equation with $\hbar = h/2\pi$, we have

Reduced mass; Magnitude of electron charge

Energy levels of hydrogen

$$E_n = -\frac{1}{(4\pi\epsilon_0)^2}\frac{m_r e^4}{2n^2\hbar^2} = -\frac{13.60\ \text{eV}}{n^2} \qquad (41.21)$$

Electric constant; Principal quantum number $(n = 1, 2, 3, \ldots)$; Planck's constant divided by 2π

As in Section 39.3, we call n the **principal quantum number.**

Equation (41.21) is an important validation of our Schrödinger-equation analysis of the hydrogen atom. The Schrödinger analysis is quite different from the Bohr model, both mathematically and conceptually, yet both yield the same energy-level scheme—a scheme that agrees with the energies determined from spectra. As we will see, the Schrödinger analysis can explain many more aspects of the hydrogen atom than can the Bohr model.

Quantization of Orbital Angular Momentum

The solutions to Eqs. (41.20) that satisfy the boundary conditions mentioned above also have quantized values of *orbital angular momentum*. That is, only certain discrete values of the magnitude and components of orbital angular momentum are permitted. In discussing the Bohr model in Section 39.3, we mentioned that quantization of angular momentum was a result with little fundamental justification. With the Schrödinger equation it appears automatically.

The possible values of the magnitude L of orbital angular momentum $\vec{L}$ are determined by the requirement that the $\Theta(\theta)$ function in Eq. (41.20b) must be finite at $\theta = 0$ and $\theta = \pi$. In a level with energy E_n and principal quantum number n, the possible values of L are

Orbital quantum number

Magnitude of orbital angular momentum, hydrogen atom

$$L = \sqrt{l(l+1)}\,\hbar \quad (l = 0, 1, 2, \ldots, n-1) \qquad (41.22)$$

Planck's constant divided by 2π; Principal quantum number $(n = 1, 2, 3, \ldots)$

The **orbital quantum number** l in Eq. (41.22) is the same l that appears in Eqs. (41.20a) and (41.20b). In the Bohr model, each energy level corresponded to a single value of angular momentum. Equation (41.22) shows that in fact there are n different possible values of L for the nth energy level.

An interesting feature of Eq. (41.22) is that the orbital angular momentum is *zero* for $l = 0$ states. This result disagrees with the Bohr model, in which the electron always moved in a circle of definite radius and L was never zero. The $l = 0$ wave functions ψ depend only on r; for these states, the functions $\Theta(\theta)$ and $\Phi(\phi)$ are constants. Thus the wave functions for $l = 0$ states are spherically symmetric. There is nothing in their probability distribution $|\psi|^2$ to favor one direction over any other, and there is no orbital angular momentum.

The permitted values of the *component* of $\vec{L}$ in a given direction, say the z-component L_z, are determined by the requirement that the $\Phi(\phi)$ function must equal $\Phi(\phi + 2\pi)$. The possible values of L_z are

Orbital magnetic quantum number

***z*-component of orbital angular momentum, hydrogen atom**

$$L_z = m_l\hbar \qquad (m_l = 0, \pm 1, \pm 2, \ldots, \pm l) \qquad (41.23)$$

Planck's constant divided by 2π; Orbital quantum number

The quantum number m_l is the same as that in Eqs. (41.20b) and (41.20c). We see that m_l can be zero or a positive or negative integer up to, but no larger in magnitude than, l. That is, $|m_l| \leq l$. For example, if $l = 1$, m_l can equal 1, 0, or -1. For reasons that will emerge later, we call m_l the *orbital magnetic quantum number,* or **magnetic quantum number** for short.

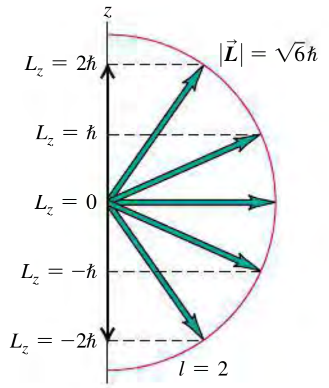
The component L_z can never be quite as large as L (unless both are zero). For example, when $l = 2$, the largest possible value of m_l is also 2; then Eqs. (41.22) and (41.23) give

$$L = \sqrt{2(2+1)}\hbar = \sqrt{6}\hbar = 2.45\hbar$$

$$L_z = 2\hbar$$

41.6 (a) When $l = 2$, the magnitude of the angular momentum vector $\vec{L}$ is $\sqrt{6}\hbar = 2.45\hbar$, but $\vec{L}$ does not have a definite direction. In this semiclassical vector picture, $\vec{L}$ makes an angle of 35.3° with the z-axis when the z-component has its maximum value of $2\hbar$. (b) These cones show the possible directions of $\vec{L}$ for different values of L_z.

(a)



(b)

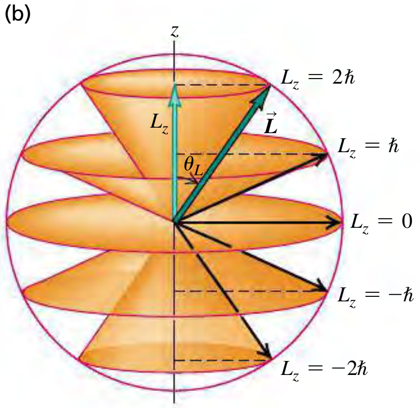


Figure 41.6 shows the situation. The minimum value of the angle θ_L between the vector $\vec{L}$ and the z-axis is

$$\theta_L = \arccos\frac{L_z}{L}$$

$$= \arccos\frac{2}{2.45} = 35.3°$$

That $|L_z|$ is always less than L is also required by the uncertainty principle. Suppose we could know the precise *direction* of the orbital angular momentum vector. Then we could let that be the direction of the z-axis, and L_z would equal L. This corresponds to a particle moving in the xy-plane only, in which case the z-component of the linear momentum $\vec{p}$ would be zero with no uncertainty Δp_z. Then the uncertainty principle $\Delta z \Delta p_z \geq \hbar$ requires infinite uncertainty Δz in the coordinate z. This is impossible for a localized state; we conclude that we can't know the direction of $\vec{L}$ precisely. Thus, as we've already stated, the component of $\vec{L}$ in a given direction can never be quite as large as its magnitude L. Also, if we can't know the direction of $\vec{L}$ precisely, we can't determine the components L_x and L_y precisely. Thus we show *cones* of possible directions for $\vec{L}$ in Fig. 41.6b.

You may wonder why we have singled out the z-axis. We can't determine all three components of orbital angular momentum with certainty, so we arbitrarily pick one as the component we want to measure. When we discuss interactions of the atom with a magnetic field, we will consistently choose the positive z-axis to be in the direction of $\vec{B}$.

Quantum Number Notation

The wave functions for the hydrogen atom are determined by the values of three quantum numbers n, l, and m_l. (Compare this to the particle in a three-dimensional box that we considered in Section 41.2. There, too, three quantum numbers were needed to describe each stationary state.) The energy E_n is determined by the principal quantum number n according to Eq. (41.21). The magnitude of orbital angular momentum is determined by the orbital quantum number l, as in Eq. (41.22). The component of orbital angular momentum in a specified axis direction (customarily the z-axis) is determined by the magnetic quantum number m_l, as in Eq. (41.23). The energy does not depend on the values of l or m_l (**Fig. 41.7**), so for each energy level E_n given by Eq. (41.21), there is more than one distinct state having the same energy but different quantum numbers. That is, these states are *degenerate,* just like most of the states of a particle in a three-dimensional box. As for the three-dimensional box, degeneracy arises because the hydrogen atom is symmetric: If you rotate the atom through any angle, the potential-energy function at a distance r from the nucleus has the same value.

States with various values of the orbital quantum number l are often labeled with letters, according to the following scheme:

$l = 0$: s states $\quad$ $l = 3$: f states
$l = 1$: p states $\quad$ $l = 4$: g states
$l = 2$: d states $\quad$ $l = 5$: h states

and so on alphabetically. This seemingly irrational choice of the letters s, p, d, and f originated in the early days of spectroscopy and has no fundamental significance. In an important form of *spectroscopic notation* that we'll use often, a state with $n = 2$ and $l = 1$ is called a $2p$ state; a state with $n = 4$ and $l = 0$ is a $4s$ state; and so on. Only s states $(l = 0)$ are spherically symmetric.

Here's another bit of notation. The radial extent of the wave functions increases with the principal quantum number n, and we can speak of a region of space associated with a particular value of n as a **shell**. Especially in discussions of many-electron atoms, these shells are denoted by capital letters:

$n = 1$: K shell
$n = 2$: L shell
$n = 3$: M shell
$n = 4$: N shell

and so on alphabetically. For each n, different values of l correspond to different *subshells*. For example, the L shell $(n = 2)$ contains the $2s$ and $2p$ subshells.

Table 41.1 shows some of the possible combinations of the quantum numbers n, l, and m_l for hydrogen-atom wave functions. The spectroscopic notation and the shell notation for each are also shown.

41.7 The energy for an orbiting satellite such as the Hubble Space Telescope depends on the average distance between the satellite and the center of the earth. It does *not* depend on whether the orbit is circular (with a large orbital angular momentum L) or elliptical (in which case L is smaller). In the same way, the energy of a hydrogen atom does not depend on the orbital angular momentum.

TABLE 41.1 Quantum States of the Hydrogen Atom

n	l	m_l	Spectroscopic Notation	Shell
1	0	0	$1s$	K
2	0	0	$2s$	L
2	1	$-1, 0, 1$	$2p$	
3	0	0	$3s$	M
3	1	$-1, 0, 1$	$3p$	
3	2	$-2, -1, 0, 1, 2$	$3d$	
4	0	0	$4s$	N

and so on

PROBLEM-SOLVING STRATEGY 41.1 ATOMIC STRUCTURE

IDENTIFY *the relevant concepts:* Many problems in atomic structure can be solved simply by reference to the quantum numbers n, l, and m_l that describe the total energy E, the magnitude of the orbital angular momentum $\vec{L}$, the z-component of $\vec{L}$, and other properties of an atom.

SET UP *the problem:* Identify the target variables and choose the appropriate equations, which may include Eqs. (41.21), (41.22), and (41.23).

EXECUTE *the solution* as follows:

1. Be sure you understand the possible values of the quantum numbers n, l, and m_l for the hydrogen atom. They are all integers; n is always greater than zero, l can be zero or positive up to $n - 1$, and m_l can range from $-l$ to l. You should know how to count the number of (n, l, m_l) states in each shell (K, L, M, and so on) and subshell ($3s$, $3p$, $3d$, and so on). Be able to *construct* Table 41.1, not just to write it from memory.
2. Solve for the target variables.

EVALUATE *your answer:* It helps to be familiar with typical magnitudes in atomic physics. For example, the electric potential energy of a proton and electron 0.10 nm apart (typical of atomic dimensions) is about -15 eV. Visible light has wavelengths around 500 nm and frequencies around 5×10^{14} Hz. Problem-Solving Strategy 39.1 (Section 39.1) gives other typical magnitudes.

SOLUTION

EXAMPLE 41.2 COUNTING HYDROGEN STATES

How many distinct (n, l, m_l) states of the hydrogen atom with $n = 3$ are there? What are their energies?

SOLUTION

IDENTIFY and SET UP: This problem uses the relationships among the principal quantum number n, orbital quantum number l, magnetic quantum number m_l, and energy of a state for the hydrogen atom. We use the rule that l can have n integer values, from 0 to $n - 1$, and that m_l can have $2l + 1$ values, from $-l$ to l. Equation (41.21) gives the energy of any particular state.

EXECUTE: When $n = 3$, l can be 0, 1, or 2. When $l = 0$, m_l can be only 0 (1 state). When $l = 1$, m_l can be -1, 0, or 1 (3 states). When $l = 2$, m_l can be -2, -1, 0, 1, or 2 (5 states). The total number of (n, l, m_l) states with $n = 3$ is therefore $1 + 3 + 5 = 9$. (In Section 41.5 we'll find that the total number of $n = 3$ states is in fact twice this, or 18, because of electron spin.)

The energy of a hydrogen-atom state depends only on n, so all 9 of these states have the same energy. From Eq. (41.21),

$$E_3 = \frac{-13.60\ \text{eV}}{3^2} = -1.51\ \text{eV}$$

EVALUATE: For a given value of n, the total number of (n, l, m_l) states turns out to be n^2. In this case $n = 3$ and there are $3^2 = 9$ states. Remember that the ground level of hydrogen has $n = 1$ and $E_1 = -13.6$ eV; the $n = 3$ excited states have a higher (less negative) energy.

SOLUTION

EXAMPLE 41.3 ANGULAR MOMENTUM IN AN EXCITED LEVEL OF HYDROGEN

Consider the $n = 4$ states of hydrogen. (a) What is the maximum magnitude L of the orbital angular momentum? (b) What is the maximum value of L_z? (c) What is the minimum angle between $\vec{L}$ and the z-axis? Give your answers to parts (a) and (b) in terms of $\hbar$.

SOLUTION

IDENTIFY and SET UP: We again need to relate the principal quantum number n and the orbital quantum number l for a hydrogen atom. We also need to relate the value of l and the magnitude and possible directions of the orbital angular momentum vector. We'll use Eq. (41.22) in part (a) to determine the maximum value of L; then we'll use Eq. (41.23) in part (b) to determine the maximum value of L_z. The angle between $\vec{L}$ and the z-axis is minimum when L_z is maximum (so that $\vec{L}$ is most nearly aligned with the positive z-axis).

EXECUTE: (a) When $n = 4$, the maximum value of the orbital quantum number l is $(n - 1) = (4 - 1) = 3$; from Eq. (41.22),

$$L_{\text{max}} = \sqrt{3(3 + 1)}\,\hbar = \sqrt{12}\,\hbar = 3.464\hbar$$

(b) For $l = 3$ the maximum value of m_l is 3. From Eq. (41.23),

$$(L_z)_{\text{max}} = 3\hbar$$

(c) The *minimum* allowed angle between $\vec{L}$ and the z-axis corresponds to the *maximum* allowed values of L_z and m_l (Fig. 41.6b shows an $l = 2$ example). For the state with $l = 3$ and $m_l = 3$,

$$\theta_{\text{min}} = \arccos\frac{(L_z)_{\text{max}}}{L} = \arccos\frac{3\hbar}{3.464\hbar} = 30.0°$$

EVALUATE: As a check, you can verify that θ is greater than 30.0° for all states with smaller values of l.

Electron Probability Distributions

Rather than picturing the electron as a point particle moving in a precise circle, the Schrödinger equation gives an electron *probability distribution* surrounding the nucleus. The hydrogen-atom probability distributions are three-dimensional, so they are harder to visualize than the two-dimensional circular orbits of the Bohr model. It's helpful to look at the *radial probability distribution function* $P(r)$—that is, the probability per radial length for the electron to be found at various distances from the proton. From Section 41.1 the probability for finding the electron in a small volume element dV is $|\psi|^2\,dV$. (We assume that ψ is normalized in accordance with Eq. (41.6)—that is, that the integral of $|\psi|^2\,dV$ over all space equals unity so that there is 100% probability of finding the electron somewhere in the universe.) Let's take as our volume element a thin spherical shell with inner radius r and outer radius $r + dr$. The volume dV of this shell is approximately its area $4\pi r^2$ multiplied by its thickness dr:

$$dV = 4\pi r^2\,dr \qquad (41.24)$$

We denote by $P(r)\,dr$ the probability of finding the particle within the radial range dr; then, using Eq. (41.24), we get

Probability that the electron is between r and $r + dr$

Radial probability distribution function; Probability distribution function

$$\underbrace{P(r)}_{\text{Radial probability distribution function}}\,dr = \underbrace{|\psi|^2}_{\text{Probability distribution function}}\,\underbrace{dV}_{\text{Volume of spherical shell with inner radius } r\text{, outer radius } r + dr} = |\psi|^2 \underbrace{4\pi r^2\,dr}_{} \tag{41.25}$$

Wave function; Volume of spherical shell with inner radius r, outer radius $r + dr$

For wave functions that depend on θ and ϕ as well as r, we use the value of $|\psi|^2$ averaged over all angles in Eq. (41.25).

Figure 41.8 shows graphs of $P(r)$ for several hydrogen-atom wave functions. The r scales are labeled in multiples of a, the smallest distance between the electron and the nucleus in the Bohr model:

Radius of smallest orbit in Bohr model

Electric constant; Planck's constant; Planck's constant divided by 2π

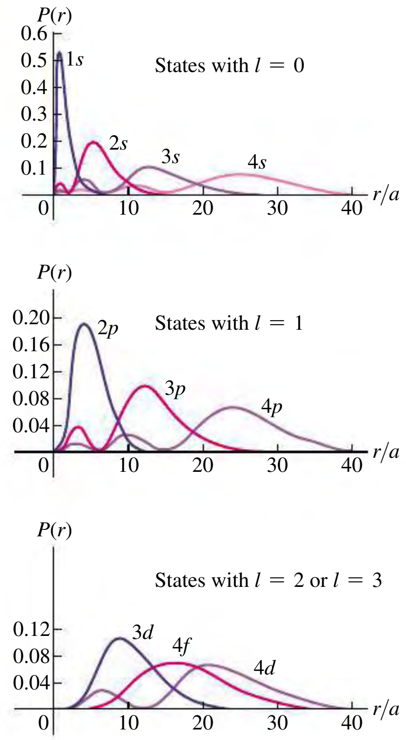
$$a = \frac{\epsilon_0 h^2}{\pi m_r e^2} = \frac{4\pi\epsilon_0\hbar^2}{m_r e^2} = 5.29 \times 10^{-11}\text{ m} \tag{41.26}$$

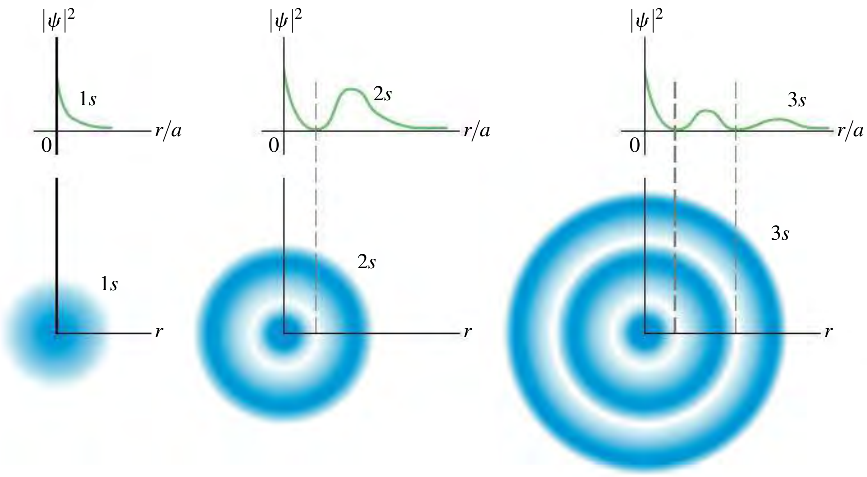
Reduced mass; Magnitude of electron charge

As for a particle in a cubical box (see Section 41.2), there are some locations where the probability is zero. These surfaces are planes for a particle in a box; for a hydrogen atom these are spherical surfaces (that is, surfaces of constant r). Note that for the states that have the largest possible l for each n (such as 1s, 2p, 3d, and 4f states), $P(r)$ has a single maximum at n^2a. For these states, the electron is most likely to be found at the distance from the nucleus that is predicted by the Bohr model, $r = n^2a$.

Figure 41.8 shows *radial* probability distribution functions $P(r) = 4\pi r^2|\psi|^2$, which indicate the relative probability of finding the electron within a thin spherical shell of radius r. By contrast, **Figs. 41.9** and **41.10** (next page) show the *three-dimensional* probability distribution functions $|\psi|^2$, which indicate the relative probability of finding the electron within a small box at a given position. The darker the blue "cloud," the greater the value of $|\psi|^2$. (These are similar to the "clouds" shown in Fig. 41.2.) Figure 41.9 shows cross sections of the spherically symmetric probability clouds for the lowest three s subshells, for which $|\psi|^2$ depends only on the radial coordinate r. Figure 41.10 shows cross sections of the clouds for other electron states for which $|\psi|^2$ depends on both r and θ. For these states the probability distribution function is zero for certain values of θ as well as for certain values of r. In *any* stationary state of the hydrogen atom, $|\psi|^2$ is independent of ϕ.

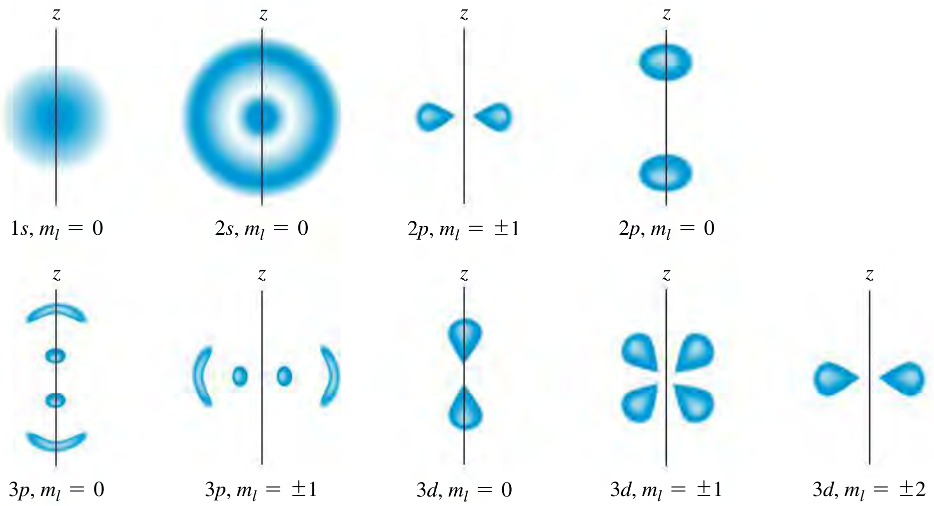
41.8 Radial probability distribution functions $P(r)$ for several hydrogen-atom wave functions, plotted as functions of the ratio r/a [see Eq. (41.26)]. For each function, the number of maxima is $(n - l)$. The curves for which $l = n - 1$ $(1s, 2p, 3d, \ldots)$ have only one maximum, located at $r = n^2a$.





41.9 Three-dimensional probability distribution functions $|\psi|^2$ for the spherically symmetric 1s, 2s, and 3s hydrogen-atom wave functions.

41.10 Cross sections of three-dimensional probability distributions for a few quantum states of the hydrogen atom. They are not to the same scale. Mentally rotate each drawing about the z-axis to obtain the three-dimensional representation of $|\psi|^2$. For example, the $2p$, $m_l = \pm 1$ probability distribution looks like a fuzzy donut.



EXAMPLE 41.4 A HYDROGEN WAVE FUNCTION

The ground-state wave function for hydrogen (a $1s$ state) is

$$\psi_{1s}(r) = \frac{1}{\sqrt{\pi a^3}} e^{-r/a}$$

(a) Verify that this function is normalized. (b) What is the probability that the electron will be found at a distance less than a from the nucleus?

SOLUTION

IDENTIFY and SET UP: This example is similar to Example 41.1 in Section 41.2. We need to show that this wave function satisfies the condition that the probability of finding the electron *somewhere* is 1. We then need to find the probability that it will be found in the region $r < a$. In part (a) we'll carry out the integral $\int |\psi|^2\,dV$ over all space; if it is equal to 1, the wave function is normalized. In part (b) we'll carry out the same integral over a spherical volume that extends from the origin (the nucleus) out to a distance a from the nucleus.

EXECUTE: (a) Since the wave function depends only on the radial coordinate r, we can choose our volume elements to be spherical shells of radius r, thickness dr, and volume dV given by Eq. (41.24). We then have

$$\int_{\text{all space}} |\psi_{1s}|^2\,dV = \int_0^\infty \frac{1}{\pi a^3} e^{-2r/a}(4\pi r^2\,dr)$$

$$= \frac{4}{a^3}\int_0^\infty r^2 e^{-2r/a}\,dr$$

You can find the following indefinite integral in a table of integrals or by integrating by parts:

$$\int r^2 e^{-2r/a}\,dr = \left(-\frac{ar^2}{2} - \frac{a^2 r}{2} - \frac{a^3}{4}\right)e^{-2r/a}$$

Evaluating this between the limits $r = 0$ and $r = \infty$ is simple; it is zero at $r = \infty$ because of the exponential factor, and at $r = 0$ only the last term in the parentheses survives. Thus the value of the definite integral is $a^3/4$. Putting it all together, we find

$$\int_0^\infty |\psi_{1s}|^2\,dV = \frac{4}{a^3}\int_0^\infty r^2 e^{-2r/a}\,dr = \frac{4}{a^3}\frac{a^3}{4} = 1$$

The wave function *is* normalized.

(b) To find the probability P that the electron is found within $r < a$, we carry out the same integration but with the limits 0 and a. We'll leave the details to you. From the upper limit we get $-5e^{-2}a^3/4$; the final result is

$$P = \int_0^a |\psi_{1s}|^2\,4\pi r^2\,dr = \frac{4}{a^3}\left(-\frac{5a^3 e^{-2}}{4} + \frac{a^3}{4}\right)$$

$$= (-5e^{-2} + 1) = 1 - 5e^{-2} = 0.323$$

EVALUATE: Our results tell us that in a ground state we expect to find the electron at a distance from the nucleus less than a about $\frac{1}{3}$ of the time and at a greater distance about $\frac{2}{3}$ of the time. It's hard to tell, but in Fig. 41.8, about $\frac{2}{3}$ of the area under the $1s$ curve is at distances greater than a (that is, $r/a > 1$).

Hydrogenlike Atoms

Two generalizations that we discussed with the Bohr model in Section 39.3 are equally valid in the Schrödinger analysis. First, if the "atom" is not composed of a single proton and a single electron, using the reduced mass m_r of the system in Eqs. (41.21) and (41.26) will lead to changes that are substantial for some exotic systems. One example is *positronium,* in which a positron and an electron orbit each other; another is a *muonic atom,* in which the electron is replaced by an unstable particle called a muon that has the same charge as an electron but is 207 times more massive. Second, our analysis is applicable to single-electron ions, such as He^+, Li^{2+}, and so on. For such ions we replace e^2 by Ze^2 in Eqs. (41.21) and (41.26), where Z is the number of protons (the **atomic number**).

TEST YOUR UNDERSTANDING OF SECTION 41.3 Rank the following states of the hydrogen atom in order from highest to lowest probability of finding the electron in the vicinity of $r = 5a$: (i) $n = 1, l = 0, m_l = 0$; (ii) $n = 2, l = 1, m_l = +1$; (iii) $n = 2, l = 1, m_l = 0$. ▌

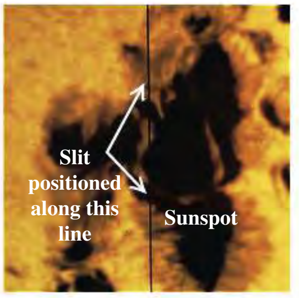
41.4 THE ZEEMAN EFFECT

The **Zeeman effect** is the splitting of atomic energy levels and the associated spectral lines when the atoms are placed in a magnetic field (**Fig. 41.11**). This effect confirms experimentally the quantization of angular momentum. In this section we'll assume that the only angular momentum is the *orbital* angular momentum of a single electron and learn why we call m_l the magnetic quantum number.

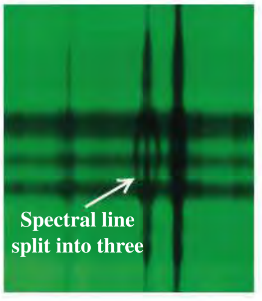
Atoms contain charges in motion, so it should not be surprising that magnetic forces cause changes in that motion and in the energy levels. In 1896 the Dutch physicist Pieter Zeeman was the first to show that in the presence of a magnetic field, some spectral lines were split into groups of closely spaced lines (**Fig. 41.12**). This effect now bears his name.

41.11 Magnetic effects on the spectrum of sunlight. (a) The slit of a spectrograph is positioned along the black line crossing a portion of a sunspot. (b) The 0.4-T magnetic field in the sunspot (a thousand times greater than the earth's field) splits the middle spectral line into three lines.

(a)



(b)



Magnetic Moment of an Orbiting Electron

Let's begin our analysis of the Zeeman effect by reviewing the concept of *magnetic dipole moment* or *magnetic moment,* introduced in Section 27.7. A plane current loop with vector area $\vec{A}$ carrying current I has a magnetic moment $\vec{\mu}$ given by

$$\vec{\mu} = I\vec{A} \tag{41.27}$$

When a magnetic dipole of moment $\vec{\mu}$ is placed in a magnetic field $\vec{B}$, the field exerts a torque $\vec{\tau} = \vec{\mu} \times \vec{B}$ on the dipole. The potential energy U associated with this interaction is given by Eq. (27.27):

$$U = -\vec{\mu} \cdot \vec{B} \tag{41.28}$$

Now let's use Eqs. (41.27) and (41.28) and the Bohr model to look at the interaction of a hydrogen atom with a magnetic field. The orbiting electron with speed v is equivalent to a current loop with radius r and area πr^2. The average current I is the average charge per unit time that passes a given point of the orbit. This is equal to the charge magnitude e divided by the time T for one revolution, given by $T = 2\pi r/v$. Thus $I = ev/2\pi r$, and from Eq. (41.27) the magnitude μ of the magnetic moment is

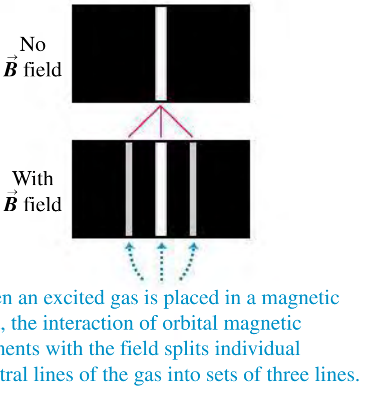
$$\mu = IA = \frac{ev}{2\pi r}\pi r^2 = \frac{evr}{2} \tag{41.29}$$

We can also express this in terms of the magnitude L of the orbital angular momentum. From Eq. (10.28) the angular momentum of a particle in a circular orbit is $L = mvr$, so Eq. (41.29) becomes

$$\mu = \frac{e}{2m}L \tag{41.30}$$

The ratio of the magnitude of $\vec{\mu}$ to the magnitude of $\vec{L}$ is $\mu/L = e/2m$ and is called the *gyromagnetic ratio.*

41.12 The normal Zeeman effect. Compare this to the magnetic splitting in the solar spectrum shown in Fig. 41.11b.



In the Bohr model, $L = nh/2\pi = n\hbar$, where $n = 1, 2, \ldots$. For an $n = 1$ state (a ground state), Eq. (41.30) becomes $\mu = (e/2m)\hbar$. This quantity is a natural unit for magnetic moment; it is called one **Bohr magneton,** denoted by μ_B:

$$\mu_B = \frac{e\hbar}{2m} \quad \text{(definition of the Bohr magneton)} \tag{41.31}$$

(We defined this quantity in Section 28.8.) Evaluating Eq. (41.31) gives

$$\mu_B = 5.788 \times 10^{-5}\ \text{eV/T} = 9.274 \times 10^{-24}\ \text{J/T or A} \cdot \text{m}^2$$

Note that the units J/T and A · m^2 are equivalent.

While the Bohr model suggests that the orbital motion of an atomic electron gives rise to a magnetic moment, this model does *not* give correct predictions about magnetic interactions. As an example, the Bohr model predicts that an electron in a hydrogen-atom ground state has an orbital magnetic moment of magnitude μ_B. But the Schrödinger picture tells us that such a ground-state electron is in an s state with zero angular momentum, so the orbital magnetic moment must be *zero!* To get the correct results, we must describe the states by using Schrödinger wave functions.

It turns out that in the Schrödinger formulation, electrons have the same ratio of μ to L (gyromagnetic ratio) as in the Bohr model—namely, $e/2m$. Suppose the magnetic field $\vec{B}$ is directed along the $+z$-axis. From Eq. (41.28) the interaction energy U of the atom's magnetic moment with the field is

$$U = -\mu_z B \tag{41.32}$$

where μ_z is the z-component of the vector $\vec{\mu}$.

Now we use Eq. (41.30) to find μ_z, recalling that e is the *magnitude* of the electron charge and that the actual charge is $-e$. Because the electron charge is negative, the orbital angular momentum and magnetic moment vectors are opposite. We find

$$\mu_z = -\frac{e}{2m}L_z \tag{41.33}$$

For the Schrödinger wave functions, $L_z = m_l\hbar$, with $m_l = 0, \pm 1, \pm 2, \ldots, \pm l$, so Eq. (41.33) becomes

$$\mu_z = -\frac{e}{2m}L_z = -m_l\frac{e\hbar}{2m} \tag{41.34}$$

CAUTION Again, two uses of the symbol *m* As in Section 41.3, the symbol m is used in two ways in Eq. (41.34). Don't confuse the electron mass m with the magnetic quantum number m_l.

Finally, using Eq. (41.31) for the Bohr magneton, we can express the interaction energy from Eq. (41.32) as

Orbital magnetic interaction energy

$$U = -\mu_z B = m_l\frac{e\hbar}{2m}B = m_l\mu_B B \tag{41.35}$$

Magnetic dipole component in direction of $\vec{B}$

Magnitude of electron charge

Planck's constant divided by 2π

Bohr magneton

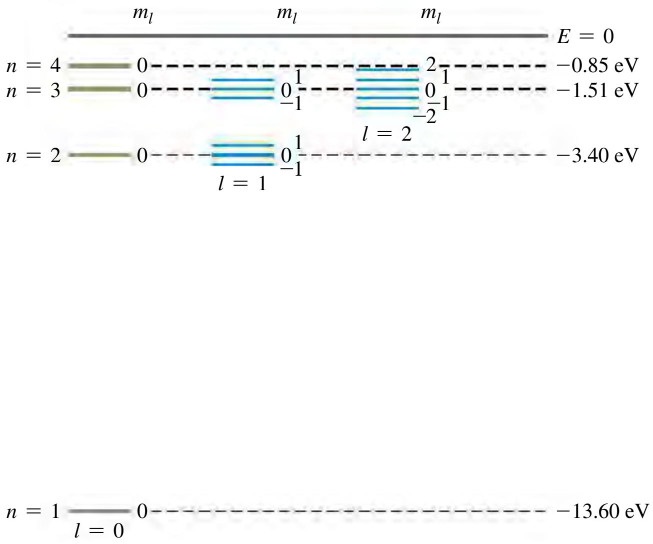
Magnetic-field magnitude

Electron mass

Orbital magnetic quantum number $= 0, \pm 1, \pm 2, \ldots, \pm l$

The magnetic field shifts the energy of each orbital state by an amount U. The interaction energy U depends on the value of m_l because m_l determines the orientation of the orbital magnetic moment relative to the magnetic field. This dependence is the reason m_l is called the magnetic quantum number.

The values of m_l range from $-l$ to $+l$ in steps of one, so an energy level with a particular value of the orbital quantum number l contains $(2l + 1)$ different orbital states. Without a magnetic field these states all have the same energy; that is, they are degenerate. The magnetic field removes this degeneracy. In the presence of a magnetic field they are split into $2l + 1$ distinct energy levels;



41.13 This energy-level diagram for hydrogen shows how the levels are split when the electron's orbital magnetic moment interacts with an external magnetic field. The values of m_l are shown adjacent to the various levels. The relative magnitudes of the level splittings are exaggerated for clarity. The $n = 4$ splittings are not shown; can you draw them in?

adjacent levels differ in energy by $(e\hbar/2m)B = \mu_B B$. We can understand this in terms of the connection between degeneracy and symmetry. With a magnetic field applied along the z-axis, the atom is no longer completely symmetric under rotation: There is a preferred direction in space. By removing the symmetry, we remove the degeneracy of states.

Figure 41.13 shows the effect on the energy levels of hydrogen. Spectral lines corresponding to transitions from one set of levels to another set are correspondingly split and appear as a series of three closely spaced spectral lines replacing a single line. As the following example shows, the splitting of spectral lines is quite small because the value of $\mu_B B$ is small even for substantial magnetic fields.

EXAMPLE 41.5 AN ATOM IN A MAGNETIC FIELD

SOLUTION

An atom in a state with $l = 1$ emits a photon with wavelength 600.000 nm as it decays to a state with $l = 0$. If the atom is placed in a magnetic field with magnitude $B = 2.00$ T, what are the shifts in the energy levels and in the wavelength that result from the interaction between the atom's orbital magnetic moment and the magnetic field?

SOLUTION

IDENTIFY and SET UP: This problem concerns the splitting of atomic energy levels by a magnetic field (the Zeeman effect). We use Eq. (41.35) to determine the energy-level shifts. The relationship $E = hc/\lambda$ between the energy and wavelength of a photon then lets us calculate the wavelengths emitted during transitions from the $l = 1$ states to the $l = 0$ state.

EXECUTE: The energy of a 600-nm photon is

$$E = \frac{hc}{\lambda} = \frac{(4.14 \times 10^{-15}\ \text{eV}\cdot\text{s})(3.00 \times 10^{8}\ \text{m/s})}{600 \times 10^{-9}\ \text{m}}$$
$$= 2.07\ \text{eV}$$

If there is no external magnetic field, that is the difference in energy between the $l = 0$ and $l = 1$ levels.

With a 2.00-T field present, Eq. (41.35) shows that there is no shift of the $l = 0$ state (which has $m_l = 0$). For the $l = 1$ states, the splitting of levels is given by

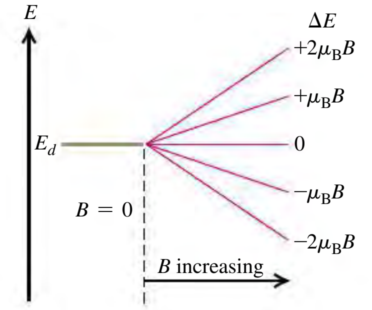
$$U = m_l \mu_B B = m_l(5.788 \times 10^{-5}\ \text{eV/T})(2.00\ \text{T})$$
$$= m_l(1.16 \times 10^{-4}\ \text{eV}) = m_l(1.85 \times 10^{-23}\ \text{J})$$

The possible values of m_l for $l = 1$ are -1, 0, and $+1$, and the three corresponding levels are separated by equal intervals of 1.16×10^{-4} eV. This is a small fraction of the 2.07-eV photon energy:

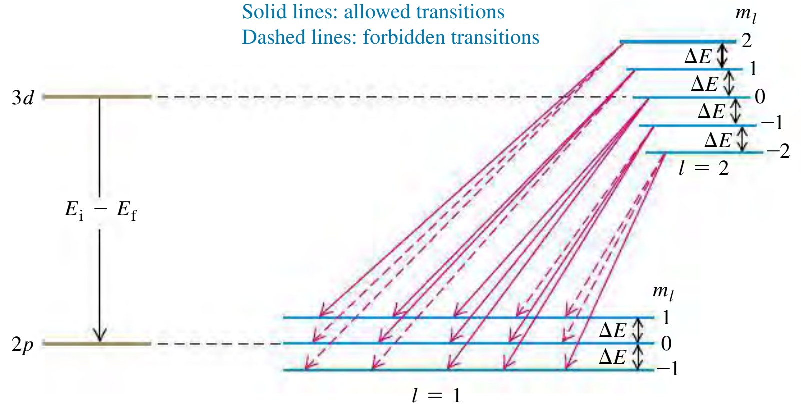
$$\frac{\Delta E}{E} = \frac{1.16 \times 10^{-4}\ \text{eV}}{2.07\ \text{eV}} = 5.60 \times 10^{-5}$$

The corresponding *wavelength* shifts are about $(5.60 \times 10^{-5}) \times (600\ \text{nm}) = 0.034$ nm. The original 600.000-nm line is split into a triplet with wavelengths 599.966, 600.000, and 600.034 nm.

EVALUATE: Even though 2.00 T would be a strong field in most laboratories, the wavelength splittings are extremely small. Nonetheless, modern spectrographs have more than enough chromatic resolving power to measure these splittings (see Section 36.5).



41.14 This figure shows how the splitting of the energy levels of a d state $(l = 2)$ depends on the magnitude B of an external magnetic field, assuming only an orbital magnetic moment.



41.15 The cause of the normal Zeeman effect. The magnetic field splits the levels, but selection rules allow transitions with only three different energy changes, giving three different photon frequencies and wavelengths.

Selection Rules

Figure 41.14 shows what happens to a set of d states $(l = 2)$ as the magnetic field increases. With zero field the five states $m_l = -2, -1, 0, 1,$ and 2 are degenerate (have the same energy), but the applied field spreads the states out. **Figure 41.15** shows the splittings of both the $3d$ and $2p$ states. Equal energy differences $(e\hbar/2m)B = \mu_B B$ separate adjacent levels. In the absence of a magnetic field, a transition from a $3d$ to a $2p$ state would yield a single spectral line with photon energy $E_i - E_f$. With the levels split as shown, it might seem that there are five possible photon energies.

In fact, there are only three possibilities. Not all combinations of initial and final levels are possible because of a restriction associated with conservation of angular momentum. The photon ordinarily carries off one unit $(\hbar)$ of angular momentum, which leads to the requirements that in a transition l must change by 1 and m_l must change by 0 or ± 1. These requirements are called **selection rules.** Transitions that obey these rules are called *allowed transitions*; those that don't are *forbidden transitions*. In Fig. 41.15 we show the allowed transitions by solid arrows. You should count the possible transition energies to convince yourself that the nine solid arrows give only three possible energies; the zero-field value $E_i - E_f$, and that value plus or minus $\Delta E = (e\hbar/2m)B = \mu_B B$. Figure 41.12 shows the corresponding spectral lines.

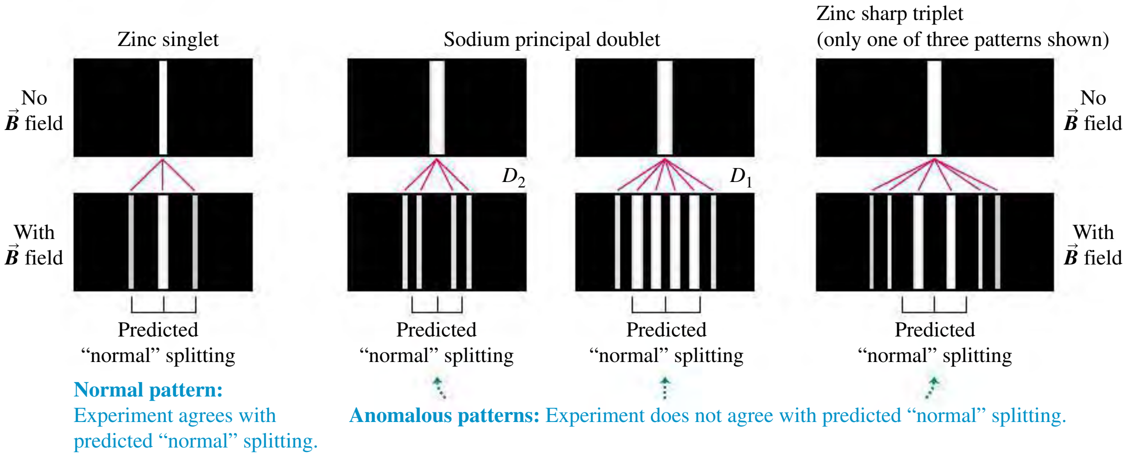
What we have described is called the *normal* Zeeman effect. It is based entirely on the orbital angular momentum of the electron. However, it leaves out a very important consideration: the electron *spin* angular momentum, the subject of the next section.

TEST YOUR UNDERSTANDING OF SECTION 41.4 In this section we assumed that the magnetic field points in the positive z-direction. Would the results be different if the magnetic field pointed in the positive x-direction?

41.5 ELECTRON SPIN

Despite the success of the Schrödinger equation in predicting the energy levels of the hydrogen atom, experimental observations indicate that it doesn't tell the whole story of the behavior of electrons in atoms. First, spectroscopists have found magnetic-field splitting into other than the three equally spaced lines that we explained in Section 41.4 (see Fig. 41.12). Before this effect was understood, it was called the *anomalous* Zeeman effect to distinguish it from the "normal"

41.16 Illustrations of the normal and anomalous Zeeman effects for two elements, zinc and sodium. The brackets under each illustration show the "normal" splitting predicted by ignoring the effect of electron spin.



effect discussed in the preceding section. **Figure 41.16** shows both kinds of splittings.

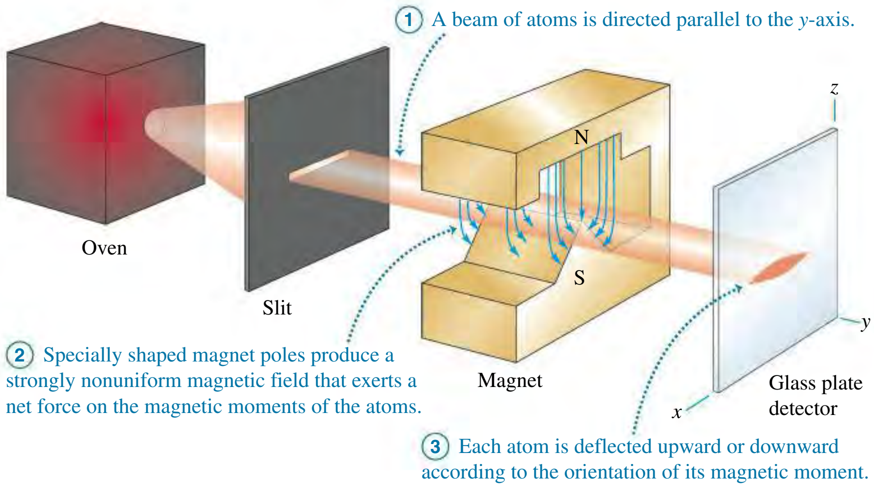
Second, some energy levels show splittings that resemble the Zeeman effect even when there is *no* external magnetic field. For example, when the lines in the hydrogen spectrum are examined with a high-resolution spectrograph, some lines are found to consist of sets of closely spaced lines called *multiplets*. Similarly, the orange-yellow line of sodium, corresponding to the transition $4p \longrightarrow 3s$ of the outer electron, is found to be a doublet $(\lambda = 589.0, 589.6\ \text{nm})$, suggesting that the $4p$ level might in fact be two closely spaced levels. The Schrödinger equation in its original form didn't predict any of this.

The Stern–Gerlach Experiment

Similar anomalies appeared in 1922 in atomic-beam experiments performed in Germany by Otto Stern and Walter Gerlach. When they passed a beam of neutral atoms through a nonuniform magnetic field (**Fig. 41.17**), atoms were deflected according to the orientation of their magnetic moments with respect to the field. These experiments demonstrated the quantization of angular momentum in a very direct way. If there were only orbital angular momentum, the deflections would split the beam into an odd number $(2l + 1)$ of different components. However, some atomic beams were split into an *even* number of components. If we use a different symbol j for an angular momentum quantum number,

MP

PhET: Stern–Gerlach Experiment



41.17 The Stern–Gerlach experiment.

setting $2j + 1$ equal to an even number gives $j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$, suggesting a half-integer angular momentum. This can't be understood on the basis of the Bohr model and similar pictures of atomic structure.

In 1925 two graduate students in the Netherlands, Samuel Goudsmidt and George Uhlenbeck, proposed that the electron might have some additional motion. Using a semiclassical model, they suggested that the electron might behave like a spinning sphere of charge instead of a particle. If so, it would have an additional *spin* angular momentum and magnetic moment. If these were quantized in much the same way as *orbital* angular momentum and magnetic moment, they might help explain the observed energy-level anomalies.

An Analogy for Electron Spin

To introduce the concept of **electron spin,** let's start with an analogy. The earth travels in a nearly circular orbit around the sun, and at the same time it *rotates* on its axis. Each motion has its associated angular momentum, which we call the *orbital* and *spin* angular momentum, respectively. The total angular momentum of the earth is the vector sum of the two. If we were to model the earth as a single point, it would have no moment of inertia about its spin axis and thus no spin angular momentum. But when our model includes the finite size of the earth, spin angular momentum becomes possible.

In the Bohr model, suppose the electron is not just a point charge but a small spinning sphere that orbits the nucleus. Then the electron has not only orbital angular momentum but also spin angular momentum associated with the rotation of its mass about its axis. The sphere carries an electric charge, so the spinning motion leads to current loops and to a magnetic moment, as we discussed in Section 27.7. In a magnetic field, the *spin* magnetic moment has an interaction energy in addition to that of the *orbital* magnetic moment (the normal Zeeman-effect interaction that we discussed in Section 41.4). We should see additional Zeeman shifts due to the spin magnetic moment.

As we mentioned, such shifts *are* indeed observed in precise spectroscopic analysis. This and a variety of other experimental evidence have shown conclusively that the electron *does* have a spin angular momentum and a spin magnetic moment that do not depend on its orbital motion but are intrinsic to the electron itself. The origin of this spin angular momentum is fundamentally quantum-mechanical, so it's not correct to model the electron as a spinning charged sphere. But just as the Bohr model can be a useful conceptual picture for the motion of an electron in an atom, the spinning-sphere analogy can help you visualize the intrinsic spin angular momentum of an electron.

Spin Quantum Numbers

Like orbital angular momentum, the spin angular momentum of an electron (denoted by $\vec{S}$) is found to be quantized. Suppose we have an apparatus that measures a particular component of $\vec{S}$, say the z-component S_z. We find that the only possible values are

$$\underbrace{S_z}_{z\text{-component of spin angular momentum of electron}} = \overbrace{m_s}^{\text{Spin magnetic quantum number } = \pm\frac{1}{2}} \underbrace{\hbar}_{\text{Planck's constant divided by } 2\pi} \qquad (41.36)$$

This relationship is reminiscent of the expression $L_z = m_l\hbar$ for the z-component of orbital angular momentum, except that $|S_z|$ is *one-half* of $\hbar$ instead of an *integer* multiple. In analogy to the orbital magnetic quantum number m_l, we call the quantum number m_s the **spin magnetic quantum number.** Since m_s has only two possible values, $+\frac{1}{2}$ and $-\frac{1}{2}$, it follows that the spin angular momentum vector $\vec{S}$ can have only two orientations in space relative to the z-axis: "*spin up*" with a z-component of $+\frac{1}{2}\hbar$ and "*spin down*" with a z-component of $-\frac{1}{2}\hbar$.

Equation (41.36) also suggests that the magnitude S of the spin angular momentum is given by an expression analogous to Eq. (41.22) with the orbital quantum number l replaced by the **spin quantum number** $s = \frac{1}{2}$:

Magnitude of spin angular momentum of electron; Maximum value of spin magnetic quantum number $= \frac{1}{2}$; Planck's constant divided by 2π

$$S = \sqrt{\tfrac{1}{2}\left(\tfrac{1}{2}+1\right)}\hbar = \sqrt{\tfrac{3}{4}}\hbar \qquad (41.37)$$

The electron is often called a "spin-one-half particle" or "spin-$\frac{1}{2}$ particle."

We see that to label the state of the electron in a hydrogen atom completely, we need *four* quantum numbers: n, l, and m_l (described in Section 41.3) to specify the electron's motion relative to the nucleus, plus the spin magnetic quantum number m_s to specify the electron spin orientation.

To visualize the quantized spin of an electron in a hydrogen atom, think of the electron probability distribution function $|\psi|^2$ as a cloud surrounding the nucleus like those shown in Figs. 41.9 and 41.10. Then imagine many tiny spin arrows distributed throughout the cloud, either all with components in the $+z$-direction or all with components in the $-z$-direction. But don't take this picture too seriously.

Just as the orbital magnetic moment of the electron is proportional to its orbital angular momentum $\vec{L}$ (see Section 41.4), the electron's spin magnetic moment is proportional to its spin angular momentum $\vec{S}$. The z-component of the spin magnetic moment (μ_z) turns out to be related to S_z by

$$\mu_z = -(2.00232)\frac{e}{2m}S_z \qquad (41.38)$$

where $-e$ and m are (as usual) the charge and mass of the electron. When the atom is placed in a magnetic field, the interaction energy $-\vec{\mu}\cdot\vec{B}$ of the spin magnetic dipole moment with the field causes further splittings in energy levels and in the corresponding spectral lines.

Equation (41.38) shows that the gyromagnetic ratio for electron spin is approximately *twice* as great as the value $e/2m$ for *orbital* angular momentum and magnetic dipole moment. This result has no classical analog. But in 1928 Paul Dirac developed a relativistic generalization of the Schrödinger equation for electrons. His equation gave a spin gyromagnetic ratio of exactly $2(e/2m)$. It took another two decades to develop the area of physics called *quantum electrodynamics,* abbreviated QED, which predicts the value we've given to "only" six significant figures as 2.00232. QED now predicts a value that agrees with the currently accepted experimental value of 2.00231930436153(53), making QED the most precise theory in all science.

BIO Application Electron Spins and Dating Human Origins In many atoms, the net spin of all of the electrons is zero (as many electrons are "spin up" as are "spin down"). If these atoms are ionized and lose an electron, however, the net spin of the ion that remains is nonzero. This happens naturally in tooth enamel, where ionization is caused by radioactivity in the environment. The longer a tooth is exposed, the more ions are present. To find the age of fossil teeth, such as those in this skull of *Homo neanderthalensis*, a sample of the enamel is placed in a strong magnetic field. The ion spins align opposite to this field (become "spin down"). The sample is then illuminated with microwave photons of just the right energy to flip the spins to the higher-energy configuration aligned with the field ("spin up"). The amount of microwave energy absorbed in this process (called *electron spin resonance*) indicates the number of ions present and hence the age of the enamel.

EXAMPLE 41.6 ENERGY OF ELECTRON SPIN IN A MAGNETIC FIELD

SOLUTION

Calculate the interaction energy for an electron in an $l = 0$ state in a magnetic field with magnitude 2.00 T.

SOLUTION

IDENTIFY and SET UP: For $l = 0$ the electron has zero orbital angular momentum and zero orbital magnetic moment. Hence the only magnetic interaction is that between the $\vec{B}$ field and the spin magnetic moment $\vec{\mu}$. From Eq. (41.28), the interaction energy is $U = -\vec{\mu}\cdot\vec{B}$. As in Section 41.4, we take $\vec{B}$ to be in the positive z-direction so that $U = -\mu_z B$ [Eq. (41.32)]. Equation (41.38) gives μ_z in terms of S_z, and Eq. (41.36) gives S_z.

EXECUTE: Combining Eqs. (41.36) and (41.38), we have

$$\begin{aligned}\mu_z &= -(2.00232)\left(\frac{e}{2m}\right)\left(\pm\tfrac{1}{2}\hbar\right)\\ &= \mp\tfrac{1}{2}(2.00232)\left(\frac{e\hbar}{2m}\right) = \mp(1.00116)\mu_B\\ &= \mp(1.00116)(9.274\times10^{-24}\ \text{J/T})\\ &= \mp9.285\times10^{-24}\ \text{J/T} = \mp5.795\times10^{-5}\ \text{eV/T}\end{aligned}$$

Continued

Then from Eq. (41.32),

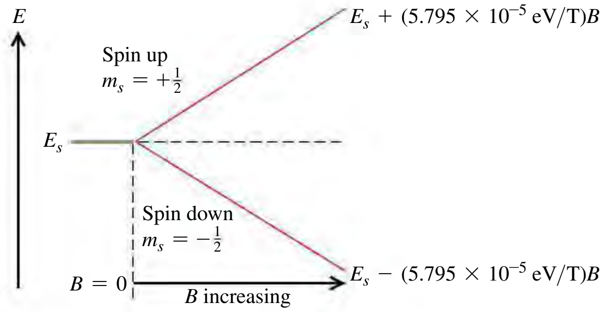
$$U = -\mu_z B = \pm(9.285 \times 10^{-24}\ \text{J/T})(2.00\ \text{T})$$
$$= \pm 1.86 \times 10^{-23}\ \text{J} = \pm 1.16 \times 10^{-4}\ \text{eV}$$

The positive value of U and the negative value of μ_z correspond to $S_z = +\frac{1}{2}\hbar$ (spin up); the negative value of U and the positive value of μ_z correspond to $S_z = -\frac{1}{2}\hbar$ (spin down).

EVALUATE: Let's check the *signs* of our results. If the electron is spin down, $\vec{S}$ points generally opposite to $\vec{B}$. Then the magnetic moment $\vec{\mu}$ (which is opposite to $\vec{S}$ because the electron charge is negative) points generally parallel to $\vec{B}$, and μ_z is positive. From Eq. (41.28), $U = -\vec{\mu} \cdot \vec{B}$, the interaction energy is negative if $\vec{\mu}$ and $\vec{B}$ are parallel. Our results show that U is indeed negative in this case. We can similarly confirm that U must be positive and μ_z negative for a spin-up electron.

The red lines in **Fig. 41.18** show how the interaction energies for the two spin states vary with the magnetic-field magnitude B. The graphs are straight lines because, from Eq. (41.32), U is proportional to B.

41.18 An $l = 0$ level of a single electron is split by interaction of the spin magnetic moment with an external magnetic field. The greater the magnitude B of the magnetic field, the greater the splitting. The quantity 5.795×10^{-5} eV/T is just $(1.00116)\mu_B$.



Spin-Orbit Coupling

We mentioned earlier that the spin magnetic dipole moment also gives splitting of energy levels even when there is *no* external field. One cause involves the orbital motion of the electron. In the Bohr model, observers moving with the electron would see the positively charged nucleus revolving around them (just as to earthbound observers the sun seems to be orbiting the earth). This apparent motion of charge causes a magnetic field at the location of the electron, as measured in the electron's moving frame of reference. The resulting interaction with the spin magnetic moment causes a twofold splitting of this level, corresponding to the two possible orientations of electron spin.

Discussions based on the Bohr model can't be taken too seriously, but a similar result can be derived from the Schrödinger equation. The interaction energy U can be expressed in terms of the scalar product of the angular momentum vectors $\vec{L}$ and $\vec{S}$. This effect is called **spin-orbit coupling;** it is responsible for the small energy difference between the two closely spaced, lowest excited levels of sodium shown in Fig. 39.19a and for the corresponding doublet (589.0, 589.6 nm) in the spectrum of sodium.

EXAMPLE 41.7 AN EFFECTIVE MAGNETIC FIELD

SOLUTION

To six significant figures, the wavelengths of the two spectral lines that make up the sodium doublet are $\lambda_1 = 588.995$ nm and $\lambda_2 = 589.592$ nm. Calculate the effective magnetic field experienced by the electron in the $3p$ levels of the sodium atom.

SOLUTION

IDENTIFY and SET UP: The two lines in the sodium doublet result from transitions from the two $3p$ levels, which are split by spin-orbit coupling, to the $3s$ level, which is *not* split because it has $L = 0$. We picture the spin-orbit coupling as an interaction between the electron spin magnetic moment and an effective magnetic field due to the nucleus. This example is like Example 41.6 in reverse: There we were given B and found the difference between the energies of the two spin states, while here we use the energy difference to find the target variable B. The difference in energy between the two $3p$ levels is equal to the difference in energy between the two photons of the sodium doublet. We use this relationship and the results of Example 41.6 to determine B.

EXECUTE: The energies of the two photons are $E_1 = hc/\lambda_1$ and $E_2 = hc/\lambda_2$. Here $E_1 > E_2$ because $\lambda_1 < \lambda_2$, so the difference in their energies is

$$\Delta E = \frac{hc}{\lambda_1} - \frac{hc}{\lambda_2} = hc\left(\frac{\lambda_2 - \lambda_1}{\lambda_2\lambda_1}\right)$$
$$= (4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(2.998 \times 10^{8}\ \text{m/s})$$
$$\times \frac{(589.592 \times 10^{-9}\ \text{m}) - (588.995 \times 10^{-9}\ \text{m})}{(589.592 \times 10^{-9}\ \text{m})(588.995 \times 10^{-9}\ \text{m})}$$
$$= 0.00213\ \text{eV} = 3.41 \times 10^{-22}\ \text{J}$$

This equals the energy difference between the two $3p$ levels. The spin-orbit interaction raises one level by 1.70×10^{-22} J (one-half of 3.41×10^{-22} J) and lowers the other by 1.70×10^{-22} J. From Example 41.6, the amount each state is raised or lowered is $|U| = (1.00116)\mu_B B$, so

$$B = \left|\frac{U}{(1.00116)\mu_B}\right| = \frac{1.70 \times 10^{-22}\ \text{J}}{9.28 \times 10^{-24}\ \text{J/T}} = 18.0\ \text{T}$$

EVALUATE: The electron experiences a *very* strong effective magnetic field. To produce a steady, macroscopic field of this magnitude in the laboratory requires state-of-the-art electromagnets.

Combining Orbital and Spin Angular Momenta

The orbital and spin angular momenta ($\vec{L}$ and $\vec{S}$, respectively) can combine in various ways. The vector sum of $\vec{L}$ and $\vec{S}$ is the *total* angular momentum $\vec{J}$:

$$\vec{J} = \vec{L} + \vec{S} \tag{41.39}$$

The possible values of the magnitude J are given in terms of a quantum number j, called the **total angular momentum quantum number:**

$$J = \sqrt{j(j+1)}\hbar \tag{41.40}$$

We can then have states in which $j = |l \pm \frac{1}{2}|$. The $l + \frac{1}{2}$ states correspond to the case in which the vectors $\vec{L}$ and $\vec{S}$ have parallel z-components; for the $l - \frac{1}{2}$ states, $\vec{L}$ and $\vec{S}$ have antiparallel z-components. For example, when $l = 1$, j can be $\frac{1}{2}$ or $\frac{3}{2}$. In another spectroscopic notation these p states are labeled $^2P_{1/2}$ and $^2P_{3/2}$, respectively. The superscript is the number of possible spin orientations, the letter P (now capitalized) indicates states with $l = 1$, and the subscript is the value of j. We used this scheme to label the energy levels of the sodium atom in Fig. 39.19a.

In addition to shifts in energy levels due to magnetic effects within the atom, there are shifts of the same magnitude due to relativistic corrections to the kinetic energy of the electron. (In the Bohr model, an electron in the $n = 1$ orbit of hydrogen moves at about 1% of the speed of light.) The term "fine structure" refers to the energy-level shifts caused by magnetic and relativistic effects together, as well as to the line splittings that result from these shifts. Including these effects, the energy levels of the hydrogen atom are

Energy levels of hydrogen, including fine structure

$$E_{n,j} = -\underbrace{\frac{13.60\ \text{eV}}{n^2}}_{\text{Principal quantum number } (n = 1, 2, 3, \ldots)}\left[1 + \frac{\overbrace{\alpha^2}^{\text{Fine-structure constant}}}{n^2}\left(\frac{n}{\underbrace{j + \frac{1}{2}}_{\text{Total angular momentum quantum number}}} - \frac{3}{4}\right)\right] \tag{41.41}$$

The *fine-structure constant* α that appears in Eq. (41.41) is a dimensionless number:

$$\alpha = \frac{1}{4\pi\epsilon_0}\frac{e^2}{\hbar c} = 7.2973525698(24) \times 10^{-3} \quad \text{(fine-structure constant)} \tag{41.42}$$

To five significant figures, $\alpha = 7.2974 \times 10^{-3} = 1/137.04$.

In Section 41.3 we found that the energy levels of the hydrogen atom are degenerate: All states that have the same principal quantum number n have the same energy. Our more complete treatment including fine structure shows that this degeneracy is removed: States with the same n but different values of the total angular momentum quantum number j have different energies. Example 41.8 illustrates this for the $n = 2$ levels of hydrogen.

EXAMPLE 41.8 FINE STRUCTURE AND SPECTRAL-LINE SPLITTING

SOLUTION

For an electron with orbital quantum number $l = 0$, the angular momentum is due to spin alone and the only possible value of the total angular momentum quantum number is $j = \frac{1}{2}$. If $l = 1$, two values are possible: $j = \frac{3}{2}$ (the spin and orbital angular momentum vectors are in roughly the same direction and so add together) and $j = \frac{1}{2}$ (the spin and orbital angular momentum vectors are in roughly opposite directions and so partially cancel). (a) Find the energies of a state of the electron in a hydrogen atom with $n = 2$, $l = 1$, $j = \frac{3}{2}$ (a $^2P_{3/2}$ state) and a state with $n = 2$, $l = 1$, $j = \frac{1}{2}$ (a $^2P_{1/2}$ state), and calculate the difference between the two energies. Which state has the higher energy? (b) Find the difference in wavelengths between (i) a photon emitted in a transition from a state with $n = 2, l = 1, j = \frac{3}{2}$ to a state with $n = 1, l = 0, j = \frac{1}{2}$ and (ii) a photon emitted in a transition from a state with $n = 2$, $l = 1$, $j = \frac{1}{2}$ to a state with $n = 1$, $l = 0$, $j = \frac{1}{2}$. Which photon has the longer wavelength?

SOLUTION

IDENTIFY and SET UP: In part (a) we use Eq. (41.41) to find the difference in energy between these two states, which have the same n value but different j values. The difference between the two energies is due to fine structure, so we expect this difference to be small. In part (b) both transitions end in the same state with $n = 1$, so we recognize from Section 39.3 that both are members of the Lyman series. If there were no fine structure, the two initial states would have the same energies and both photons would have the same energy E and hence the same wavelength $\lambda = hc/E$. But because the two initial states differ slightly in energy, the photons in the two transitions will have slightly different wavelengths.

EXECUTE: (a) From Eq. (41.41), the energies of the two states are

$$E_{n=2,\,j=3/2} = -\frac{13.60\text{ eV}}{2^2}\left[1 + \frac{\alpha^2}{2^2}\left(\frac{2}{\frac{3}{2}+\frac{1}{2}} - \frac{3}{4}\right)\right]$$

$$= -3.40\text{ eV}\left(1 + \frac{\alpha^2}{16}\right)$$

$$E_{n=2,\,j=1/2} = -\frac{13.60\text{ eV}}{2^2}\left[1 + \frac{\alpha^2}{2^2}\left(\frac{2}{\frac{1}{2}+\frac{1}{2}} - \frac{3}{4}\right)\right]$$

$$= -3.40\text{ eV}\left(1 + \frac{5\alpha^2}{16}\right)$$

The fine-structure terms involving α^2 cause both states to have lower (more negative) energies than in the Bohr model, in which both states would have energy $E_2 = -3.40$ eV. The fine-structure term is five times greater for the $j = \frac{1}{2}$ state, so the $j = \frac{3}{2}$ state has the higher (less negative) energy. Using the value of the fine-structure constant α from Eq. (41.42), we get the difference in energy between the two states:

$$E_{n=2,\,j=3/2} - E_{n=2,\,j=1/2}$$

$$= \left[-3.40\text{ eV}\left(1 + \frac{\alpha^2}{16}\right)\right] - \left[-3.40\text{ eV}\left(1 + \frac{5\alpha^2}{16}\right)\right]$$

$$= 3.40\text{ eV}\left(\frac{4\alpha^2}{16}\right) = (3.40\text{ eV})\left(\frac{4}{16}\right)\left(\frac{1}{137.04}\right)^2$$

$$= 4.53 \times 10^{-5}\text{ eV}$$

(b) The photon energy in each case equals the difference between the energies of the initial and final states of the electron. The final electron state for both transitions has $n = 1$ and $j = \frac{1}{2}$, which from Eq. (41.41) has energy

$$E_{n=1,\,j=1/2} = -\frac{13.60\text{ eV}}{1^2}\left[1 + \frac{\alpha^2}{1^2}\left(\frac{1}{\frac{1}{2}+\frac{1}{2}} - \frac{3}{4}\right)\right]$$

$$= -13.60\text{ eV}\left(1 + \frac{\alpha^2}{4}\right)$$

Note that as for the two $n = 2$ states in part (a), the fine-structure correction to the $n = 1$ state makes the energy more negative. The photon energies for the two transitions are then

$$E_{\text{photon}}\left(n = 2, l = 1, j = \tfrac{3}{2} \text{ to } n = 1, l = 0, j = \tfrac{1}{2}\right)$$

$$= E_{n=2,\,j=3/2} - E_{n=1,\,j=1/2}$$

$$= \left[-3.40\text{ eV}\left(1 + \frac{\alpha^2}{16}\right)\right] - \left[-13.60\text{ eV}\left(1 + \frac{\alpha^2}{4}\right)\right]$$

$$= 10.20\text{ eV} + (3.40\text{ eV})\left(\frac{15\alpha^2}{16}\right)$$

$$= 10.20\text{ eV} + 1.70 \times 10^{-4}\text{ eV}$$

$$E_{\text{photon}}\left(n = 2, l = 1, j = \tfrac{1}{2} \text{ to } n = 1, l = 0, j = \tfrac{1}{2}\right)$$

$$= E_{n=2,\,j=1/2} - E_{n=1,\,j=1/2}$$

$$= \left[-3.40\text{ eV}\left(1 + \frac{5\alpha^2}{16}\right)\right] - \left[-13.60\text{ eV}\left(1 + \frac{\alpha^2}{4}\right)\right]$$

$$= 10.20\text{ eV} + (3.40\text{ eV})\left(\frac{11\alpha^2}{16}\right)$$

$$= 10.20\text{ eV} + 1.24 \times 10^{-4}\text{ eV}$$

The photon emitted when the initial state is $n = 2$, $l = 1$, $j = \frac{1}{2}$ has a lower energy E_{photon} and hence will have a longer wavelength, as given by the equation $\lambda = hc/E_{\text{photon}}$. If you plug the two photon energies into this equation, your calculator will tell you that $\lambda = 1.216 \times 10^{-7}$ m $= 121.6$ nm in both cases because the energy difference is so small. To find the wavelength difference $\Delta\lambda$, we instead take the differential of both sides of $\lambda = hc/E_{\text{photon}}$:

$$d\lambda = d\left(\frac{hc}{E_{\text{photon}}}\right) = -\frac{hc}{(E_{\text{photon}})^2}dE_{\text{photon}}$$

$$= -\left(\frac{hc}{E_{\text{photon}}}\right)\left(\frac{1}{E_{\text{photon}}}\right)dE_{\text{photon}} = -\frac{\lambda}{E_{\text{photon}}}dE_{\text{photon}}$$

The minus sign means that a *decrease* in photon energy corresponds to an *increase* in photon wavelength. Replacing $d\lambda$ with $\Delta\lambda$ (the wavelength difference that we seek) and dE_{photon} with ΔE_{photon}, we get the difference between the two photon wavelengths:

$$\Delta\lambda = -\frac{\lambda}{E_{\text{photon}}}\Delta E_{\text{photon}}$$

To four significant digits, we have $\lambda = 121.6$ nm and $E_{\text{photon}} = 10.20$ eV. We find the photon energy difference ΔE_{photon} from the two expressions above, subtracting the larger energy from the smaller one so that $\Delta\lambda$ is positive:

$$\Delta\lambda = -\frac{121.6\text{ nm}}{10.20\text{ eV}}\left\{\left[10.20\text{ eV} + (3.40\text{ eV})\left(\frac{11\alpha^2}{16}\right)\right] - \left[10.20\text{ eV} + (3.40\text{ eV})\left(\frac{15\alpha^2}{16}\right)\right]\right\}$$

$$= -\frac{121.6\text{ nm}}{10.20\text{ eV}}(3.40\text{ eV})\left(-\frac{4\alpha^2}{16}\right)$$

$$= \frac{121.6\text{ nm}}{10.20\text{ eV}}(3.40\text{ eV})\left(\frac{4}{16}\right)\left(\frac{1}{137.04}\right)^2$$

$$= 5.40 \times 10^{-4}\text{ nm}$$

EVALUATE: This line splitting is very small, as predicted. Fine structure is fine indeed! It is nonetheless observable with a diffraction grating that has a sufficient number of lines (see Section 36.5). The measured wavelengths are 121.567364 nm for the transition that begins in the $j = \frac{1}{2}$ state and 121.566824 nm for the transition that begins in the $j = \frac{3}{2}$ state. These are ultraviolet wavelengths.

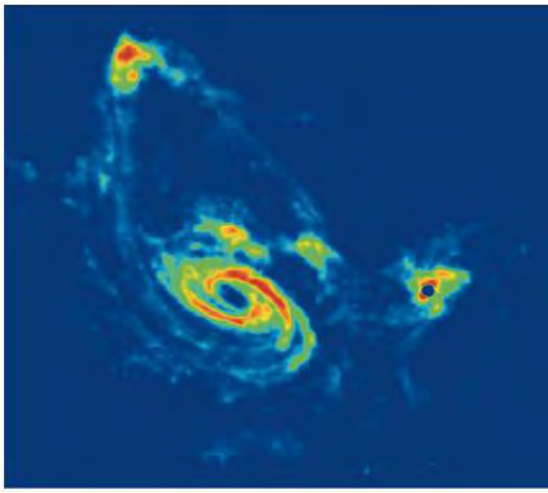
There are also states of the hydrogen atom with $n = 2$, $l = 0$, $j = \frac{1}{2}$. [From Eq. (41.41), these states have the same energy as those with $n = 2$, $l = 1$, $j = \frac{1}{2}$; the energy $E_{n,j}$ depends on n and j but not on l.] However, an electron in an $n = 2$, $l = 0$, $j = \frac{1}{2}$ state *cannot* emit a photon and transition to an $n = 1$, $l = 0$, $j = \frac{1}{2}$ state. Such a transition is forbidden by the selection rule that l must change by 1 when a photon is emitted (see Section 41.4).

Additional, much smaller splittings are associated with the fact that the *nucleus* of the atom has a magnetic dipole moment that interacts with the orbital and/or spin magnetic dipole moments of the electrons. These effects are called *hyperfine structure*. For example, the ground level of hydrogen is split into two states, separated by only 5.9×10^{-6} eV. The photon that is emitted in the transitions between these states has a wavelength of 21 cm. Radio astronomers use this wavelength to map clouds of interstellar hydrogen gas that are too cold to emit visible light (**Fig. 41.19**).

41.19 (a) In a visible-light image, these three distant galaxies appear to be unrelated. But in fact these galaxies are connected by immense streamers of hydrogen gas. This is revealed by (b) the false-color image made with a radio telescope tuned to the 21-cm wavelength emitted by hydrogen atoms.

(a) Galaxies in visible light (negative image; galaxies appear dark)

(b) Radio image of the same galaxies at wavelength 21 cm



TEST YOUR UNDERSTANDING OF SECTION 41.5 In which of the following situations is the magnetic moment of an electron perfectly aligned with a magnetic field that points in the positive z-direction? (i) $m_s = +\frac{1}{2}$; (ii) $m_s = -\frac{1}{2}$; (iii) both (i) and (ii); (iv) neither (i) nor (ii).

41.6 MANY-ELECTRON ATOMS AND THE EXCLUSION PRINCIPLE

So far our analysis of atomic structure has concentrated on the hydrogen atom. That's natural; neutral hydrogen, with only one electron, is the simplest atom. Let's now take what we've learned about the hydrogen atom and apply that knowledge to the more complicated case of many-electron atoms.

An atom in its normal (electrically neutral) state has Z electrons and Z protons. Recall from Section 41.3 that we call Z the *atomic number*. The total electric charge of such an atom is exactly zero because the neutron has no charge while the proton and electron charges have the same magnitude but opposite sign.

A complete understanding of such a general atom requires that we know the wave function that describes the behavior of all Z of its electrons. This wave function depends on $3Z$ coordinates (three for each electron), so its complexity increases very rapidly with increasing Z. What's more, each of the Z electrons interacts not only with the nucleus but also with every other electron. The potential energy is therefore a complicated function of all $3Z$ coordinates, and the Schrödinger equation contains second derivatives with respect to all of them. Finding exact solutions to such equations is such a complex task that it has not been successfully achieved even for the neutral helium atom, which has only two electrons.

Fortunately, various approximation schemes are available. The simplest approximation is to ignore all interactions between electrons and consider each electron as moving under the action only of the nucleus (considered to be a point charge). In this approximation, we write a separate wave function for each *individual* electron. Each such function is like that for the hydrogen atom, specified by four quantum numbers (n, l, m_l, m_s). The nuclear charge is Ze

instead of e, so we replace every factor of e^2 in the wave functions and the energy levels by Ze^2. In particular, the energy levels are given by Eq. (41.21) with e^4 replaced by Z^2e^4:

$$E_n = -\frac{1}{(4\pi\epsilon_0)^2}\frac{m_r Z^2 e^4}{2n^2\hbar^2} = -\frac{Z^2}{n^2}(13.6\text{ eV}) \qquad (41.43)$$

This approximation is fairly drastic; when there are many electrons, their interactions with each other are as important as the interaction of each with the nucleus. So this model isn't very useful for quantitative predictions.

The Central-Field Approximation

A less drastic and more useful approximation is to think of all the electrons together as making up a charge cloud that is, on average, *spherically symmetric*. We can then think of each individual electron as moving in the total electric field due to the nucleus and this averaged-out cloud of all the other electrons. There is a corresponding spherically symmetric potential-energy function $U(r)$. This picture is called the **central-field approximation;** it provides a useful starting point for understanding atomic structure.

In the central-field approximation we can again deal with one-electron wave functions. The Schrödinger equation for these functions differs from the equation for hydrogen, which we discussed in Section 41.3, only in that the $1/r$ potential-energy function is replaced by a different function $U(r)$. Now, Eqs. (41.20) show that $U(r)$ does not appear in the differential equations for $\Theta(\theta)$ and $\Phi(\phi)$. So those angular functions are exactly the same as for hydrogen, and the orbital angular momentum *states* are also the same as before. The quantum numbers l, m_l, and m_s have the same meanings as before, and Eqs. (41.22) and (41.23) again give the magnitude and z-component of the orbital angular momentum.

The radial wave functions and probabilities are different than for hydrogen because of the change in $U(r)$, so the energy levels are no longer given by Eq. (41.21). We can still label a state by using the four quantum numbers (n, l, m_l, m_s). In general, the energy of a state now depends on both n and l, rather than just on n as with hydrogen. (Due to fine-structure effects, the energy can also depend on the total angular momentum quantum number j. These effects are generally small, however, so we ignore them for this discussion.) The restrictions on the values of the quantum numbers are the same as before:

Allowed values of quantum numbers for one-electron wave functions:

Principal quantum number; Orbital magnetic quantum number; Spin magnetic quantum number

$$n \geq 1 \qquad 0 \leq l \leq n-1 \qquad |m_l| \leq l \qquad m_s = \pm\tfrac{1}{2} \qquad (41.44)$$

Orbital quantum number

The Exclusion Principle

To understand the structure of many-electron atoms, we need an additional principle, the *exclusion principle*. To see why this principle is needed, let's consider the lowest-energy state, or *ground state,* of a many-electron atom. In the one-electron states of the central-field model, there is a lowest-energy state (corresponding to an $n = 1$ state of hydrogen). We might expect that in the ground state of a complex atom, *all* the electrons should be in this lowest state. If so, then we should see only gradual changes in physical and chemical properties when we look at the behavior of atoms with increasing numbers of electrons (Z).

Such gradual changes are *not* what is observed. Instead, properties of elements vary widely from one to the next, with each element having its own distinct personality. For example, the elements fluorine, neon, and sodium have 9, 10, and 11 electrons, respectively, per atom. Fluorine $(Z = 9)$ is a *halogen;* it tends strongly to form compounds in which each fluorine atom acquires an extra electron.

Sodium $(Z = 11)$ is an *alkali metal;* it forms compounds in which each sodium atom *loses* an electron. Neon $(Z = 10)$ is a *noble gas,* forming no compounds at all. Such observations show that in the ground state of a complex atom the electrons *cannot* all be in the lowest-energy states. But why not?

The key to this puzzle, discovered by the Austrian physicist Wolfgang Pauli (**Fig. 41.20**) in 1925, is called the **exclusion principle.** This principle states that **no two electrons can occupy the same quantum-mechanical state** in a given system. That is, **no two electrons in an atom can have the same values of all four quantum numbers** (n, l, m_l, m_s). Each quantum state corresponds to a certain distribution of the electron "cloud" in space. Therefore the principle also says, in effect, that no more than two electrons with opposite values of the quantum number m_s can occupy the same region of space. We shouldn't take this last statement too seriously because the electron probability functions don't have sharp, definite boundaries. But the exclusion principle limits the amount by which electron wave functions can overlap. Think of it as the quantum-mechanical analog of a university rule that allows only one student per desk. This same exclusion principle applies to all spin-$\frac{1}{2}$ particles, not just electrons. (We'll see in Chapter 43 that protons and neutrons are also spin-$\frac{1}{2}$ particles. As a result, the exclusion principle plays an important role in the structure of atomic nuclei.)

41.20 The key to understanding the periodic table of the elements was the discovery by Wolfgang Pauli (1900–1958) of the exclusion principle. Pauli received the 1945 Nobel Prize in physics for his accomplishment. This photo shows Pauli (on the left) and Niels Bohr watching the physics of a toy top spinning on the floor—a macroscopic analog of a microscopic electron with spin.

CAUTION **The meaning of the exclusion principle** Don't confuse the exclusion principle with the electrical repulsion between electrons. While both effects tend to keep electrons within an atom separated from each other, they are very different in character. Two electrons can always be pushed closer together by adding energy to combat electrical repulsion, but *nothing* can overcome the exclusion principle and force two electrons into the same quantum-mechanical state.

Table 41.2 lists some of the sets of quantum numbers for electron states in an atom. It's similar to Table 41.1 (Section 41.3), but we've added the number of states in each subshell and shell. Because of the exclusion principle, the "number of states" is the same as the *maximum* number of electrons that can be found in those states. For each state, m_s can be either $+\frac{1}{2}$ or $-\frac{1}{2}$.

As with the hydrogen wave functions, different states correspond to different spatial distributions; electrons with larger values of n are concentrated at larger distances from the nucleus. Figure 41.8 (Section 41.3) shows this effect. When an atom has more than two electrons, they can't all huddle down in the low-energy $n = 1$ states nearest to the nucleus because there are only two of these states; the exclusion principle forbids multiple occupancy of a state. Some electrons are forced into states farther away, with higher energies. Each value of n corresponds roughly to a region of space around the nucleus in the form of a spherical *shell.* Hence we speak of the K shell as the region that is occupied by the electrons in the $n = 1$ states, the L shell as the region of the $n = 2$ states, and so on. States with the same n but different l form *subshells,* such as the $3p$ subshell.

TABLE 41.2 **Quantum States of Electrons in the First Four Shells**

n	l	m_l	Spectroscopic Notation	Number of States		Shell
1	0	0	$1s$	2		K
2	0	0	$2s$	2	8	L
2	1	$-1, 0, 1$	$2p$	6		
3	0	0	$3s$	2	18	M
3	1	$-1, 0, 1$	$3p$	6		
3	2	$-2, -1, 0, 1, 2$	$3d$	10		
4	0	0	$4s$	2	32	N
4	1	$-1, 0, 1$	$4p$	6		
4	2	$-2, -1, 0, 1, 2$	$4d$	10		
4	3	$-3, -2, -1, 0, 1, 2, 3$	$4f$	14		

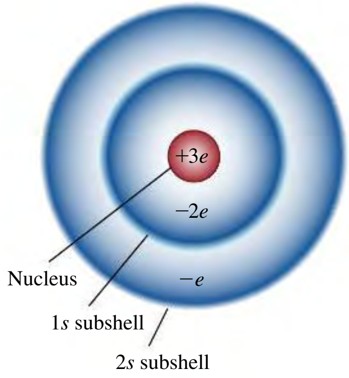
The Periodic Table

We can use the exclusion principle to derive the most important features of the structure and chemical behavior of multielectron atoms, including the periodic table of the elements. Let's imagine constructing a neutral atom by starting with a bare nucleus with Z protons and then adding Z electrons, one by one. To obtain the ground state of the atom as a whole, we fill the lowest-energy electron states (those closest to the nucleus, with the smallest values of n and l) first, and we use successively higher states until all the electrons are in place. The chemical properties of an atom are determined principally by interactions involving the outermost, or *valence,* electrons, so we particularly want to learn how these electrons are arranged.

Let's look at the ground-state electron configurations for the first few atoms (in order of increasing Z). For hydrogen the ground state is $1s$; the single electron is in a state $n = 1$, $l = 0$, $m_l = 0$, and $m_s = \pm\frac{1}{2}$. In the helium atom $(Z = 2)$, *both* electrons are in $1s$ states, with opposite spins; one has $m_s = -\frac{1}{2}$ and the other has $m_s = +\frac{1}{2}$. We denote the helium ground state as $1s^2$. (The superscript 2 is not an exponent; the notation $1s^2$ tells us that there are two electrons in the $1s$ subshell. Also, the superscript 1 is understood, as in $2s$.) For helium the K shell is completely filled, and all others are empty. Helium is a noble gas; it has no tendency to gain or lose an electron, and it forms no compounds.

41.21 Schematic representation of the charge distribution in a lithium atom. The nucleus has a charge of $+3e$.

On average, the $2s$ electron is considerably farther from the nucleus than the $1s$ electrons. Therefore, it experiences a net nuclear charge of approximately $+3e - 2e = +e$ (rather than $+3e$).



Lithium $(Z = 3)$ has three electrons. In its ground state, two are in $1s$ states and one is in a $2s$ state, so we denote the lithium ground state as $1s^2 2s$. ? On average, the $2s$ electron is considerably farther from the nucleus than are the $1s$ electrons (**Fig. 41.21**). According to Gauss's law, the *net* charge Q_{encl} attracting the $2s$ electron is nearer to $+e$ than to the value $+3e$ it would have without the two $1s$ electrons present. As a result, the $2s$ electron is loosely bound; only 5.4 eV is required to remove it, compared with the 30.6 eV given by Eq. (41.43) with $Z = 3$ and $n = 2$. Chemically, lithium is an *alkali metal*. It forms ionic compounds in which each lithium atom loses an electron and has a valence of $+1$.

Next is beryllium $(Z = 4)$; its ground-state configuration is $1s^2 2s^2$, with its two valence electrons filling the s subshell of the L shell. Beryllium is the first of the *alkaline earth* elements, forming ionic compounds in which the valence of the atoms is $+2$.

Table 41.3 shows the ground-state electron configurations of the first 30 elements. The L shell can hold eight electrons. At $Z = 10$, both the K and L shells are filled, and there are no electrons in the M shell. We expect this to be a particularly stable configuration, with little tendency to gain or lose electrons. This element is neon, a noble gas with no known compounds. The next element after neon is sodium $(Z = 11)$, with filled K and L shells and one electron in the M shell. Its "noble-gas-plus-one-electron" structure resembles that of lithium; both are alkali metals. The element *before* neon is fluorine, with $Z = 9$. It has a vacancy in the L shell and has an affinity for an extra electron to fill the shell. Fluorine forms ionic compounds in which it has a valence of -1. This behavior is characteristic of the *halogens* (fluorine, chlorine, bromine, iodine, and astatine), all of which have "noble-gas-minus-one" configurations (**Fig. 41.22**).

41.22 Salt (sodium chloride, NaCl) dissolves readily in water, making seawater salty. This is due to the electron configurations of sodium and chlorine: Sodium can easily lose an electron to form an Na^+ ion, and chlorine can easily gain an electron to form a Cl^- ion. These ions are held in solution because they are attracted to the polar ends of water molecules (see Fig. 21.30a).

Proceeding down the list, we can understand the regularities in chemical behavior displayed by the **periodic table of the elements** (Appendix D) on the basis of electron configurations. The similarity of elements in each *group* (vertical column) of the periodic table is the result of similarity in outer-electron configuration. All the noble gases (helium, neon, argon, krypton, xenon, and radon) have filled-shell or filled-shell plus filled p subshell configurations. All the alkali metals (lithium, sodium, potassium, rubidium, cesium, and francium) have "noble-gas-plus-one" configurations. All the alkaline earth metals (beryllium, magnesium, calcium, strontium, barium, and radium) have "noble-gas-plus-two" configurations, and, as we just mentioned, all the halogens (fluorine, chlorine, bromine, iodine, and astatine) have "noble-gas-minus-one" structures.

TABLE 41.3 Ground-State Electron Configurations

Element	Symbol	Atomic Number (Z)	Electron Configuration
Hydrogen	H	1	$1s$
Helium	He	2	$1s^2$
Lithium	Li	3	$1s^22s$
Beryllium	Be	4	$1s^22s^2$
Boron	B	5	$1s^22s^22p$
Carbon	C	6	$1s^22s^22p^2$
Nitrogen	N	7	$1s^22s^22p^3$
Oxygen	O	8	$1s^22s^22p^4$
Fluorine	F	9	$1s^22s^22p^5$
Neon	Ne	10	$1s^22s^22p^6$
Sodium	Na	11	$1s^22s^22p^63s$
Magnesium	Mg	12	$1s^22s^22p^63s^2$
Aluminum	Al	13	$1s^22s^22p^63s^23p$
Silicon	Si	14	$1s^22s^22p^63s^23p^2$
Phosphorus	P	15	$1s^22s^22p^63s^23p^3$
Sulfur	S	16	$1s^22s^22p^63s^23p^4$
Chlorine	Cl	17	$1s^22s^22p^63s^23p^5$
Argon	Ar	18	$1s^22s^22p^63s^23p^6$
Potassium	K	19	$1s^22s^22p^63s^23p^64s$
Calcium	Ca	20	$1s^22s^22p^63s^23p^64s^2$
Scandium	Sc	21	$1s^22s^22p^63s^23p^64s^23d$
Titanium	Ti	22	$1s^22s^22p^63s^23p^64s^23d^2$
Vanadium	V	23	$1s^22s^22p^63s^23p^64s^23d^3$
Chromium	Cr	24	$1s^22s^22p^63s^23p^64s3d^5$
Manganese	Mn	25	$1s^22s^22p^63s^23p^64s^23d^5$
Iron	Fe	26	$1s^22s^22p^63s^23p^64s^23d^6$
Cobalt	Co	27	$1s^22s^22p^63s^23p^64s^23d^7$
Nickel	Ni	28	$1s^22s^22p^63s^23p^64s^23d^8$
Copper	Cu	29	$1s^22s^22p^63s^23p^64s3d^{10}$
Zinc	Zn	30	$1s^22s^22p^63s^23p^64s^23d^{10}$

A slight complication occurs with the M and N shells because the $3d$ and $4s$ subshell levels ($n = 3$, $l = 2$, and $n = 4$, $l = 0$, respectively) have similar energies. (We'll discuss in the next subsection why this happens.) Argon $(Z = 18)$ has all the $1s$, $2s$, $2p$, $3s$, and $3p$ subshells filled, but in potassium $(Z = 19)$ the additional electron goes into a $4s$ energy state rather than a $3d$ state (because the $4s$ state has slightly lower energy).

The next several elements have one or two electrons in the $4s$ subshell and increasing numbers in the $3d$ subshell. These elements are all metals with rather similar chemical and physical properties; they form the first *transition series,* starting with scandium $(Z = 21)$ and ending with zinc $(Z = 30)$, for which all the $3d$ and $4s$ subshells are filled.

Something similar happens with $Z = 57$ through $Z = 71$, which have one or two electrons in the $6s$ subshell but only partially filled $4f$ and $5d$ subshells. These *rare earth* elements all have very similar physical and chemical properties. Another such series, called the *actinide* series, starts with $Z = 91$.

Screening

We have mentioned that in the central-field model, the energy levels depend on l as well as n. Let's take sodium $(Z = 11)$ as an example. If 10 of its electrons fill its K and L shells, the energies of some of the states for the remaining electron are found experimentally to be

$3s$ states: -5.138 eV

$3p$ states: -3.035 eV

$3d$ states: -1.521 eV

$4s$ states: -1.947 eV

BIO Application Electron Configurations and Bone Cancer Radiotherapy The orange spots in this colored x-ray image are bone cancer tumors. One method of treating bone cancer is to inject a radioactive isotope of strontium (^{89}Sr) into a patient's vein. Strontium is chemically similar to calcium because in both atoms the two outer electrons are in an s state (the structures are $1s^22s^22p^63s^23p^64s^23d^{10}4p^65s^2$ for strontium and $1s^22s^22p^63s^23p^64s^2$ for calcium). Hence the strontium is readily taken up by the tumors, where calcium turnover is more rapid than in healthy bone. Radiation from the strontium helps destroy the tumors.

The $3s$ states are the lowest (most negative); one is the ground state for the 11th electron in sodium. The energy of the $3d$ states is quite close to the energy of the $n = 3$ state in hydrogen. The surprise is that the $4s$ state energy is 0.426 eV *below* the $3d$ state, even though the $4s$ state has larger n.

We can understand these results by using Gauss's law (Section 22.3). For any spherically symmetric charge distribution, the electric-field magnitude at a distance r from the center is $Q_{\text{encl}}/4\pi\epsilon_0 r^2$, where Q_{encl} is the total charge enclosed within a sphere with radius r. Mentally remove the outer (valence) electron from a sodium atom. What you have left is a spherically symmetric collection of 10 electrons (filling the K and L shells) and 11 protons, so $Q_{\text{encl}} = -10e + 11e = +e$. If the 11th electron is completely outside this collection of charges, it is attracted by an effective charge of $+e$, not $+11e$. This is a more extreme example of the effect depicted in Fig. 41.21.

This effect is called **screening;** the 10 electrons *screen* 10 of the 11 protons in the sodium nucleus, leaving an effective net charge of $+e$. From the viewpoint of the 11th electron, this is equivalent to reducing the number of protons in the nucleus from $Z = 11$ to a smaller *effective atomic number* Z_{eff}. If the 11th electron is *completely* outside the charge distribution of the other 10 electrons, then $Z_{\text{eff}} = 1$. Since the probability distribution of the 11th electron does extend somewhat into those of the other electrons, in fact Z_{eff} is greater than 1 (but still much less than 11). In general, an electron that spends all its time completely outside a positive charge $Z_{\text{eff}}e$ has energy levels given by the hydrogen expression with e^2 replaced by $Z_{\text{eff}}e^2$. From Eq. (41.43) this is

Effective (screened) atomic number

Energy levels of an electron with screening

$$E_n = -\frac{Z_{\text{eff}}^{\,2}}{n^2}(13.6\ \text{eV}) \qquad (41.45)$$

Principal quantum number

CAUTION Different equations for different atoms Equations (41.21), (41.43), and (41.45) all give values of E_n in terms of $(13.6\ \text{eV})/n^2$, but they don't apply in general to the same atoms. Equation (41.21) is for hydrogen *only*. Equation (41.43) is for only the case in which there is no interaction with any other electron (and is thus accurate only when the atom has just one electron). Equation (41.45) is useful when one electron is screened from the nucleus by other electrons.

Now let's use the radial probability functions shown in Fig. 41.8 to explain why the energy of a sodium $3d$ state is approximately the same as the $n = 3$ value of hydrogen, -1.51 eV. The distribution for the $3d$ state (for which l has the maximum value $n - 1$) has one peak, and its most probable radius is *outside* the positions of the electrons with $n = 1$ or 2. (Those electrons also are pulled closer to the nucleus than in hydrogen because they are less effectively screened from the positive charge $11e$ of the nucleus.) Thus in sodium a $3d$ electron spends most of its time well outside the $n = 1$ and $n = 2$ states (the K and L shells). The 10 electrons in these shells screen about ten-elevenths of the charge of the 11 protons, leaving a net charge of about $Z_{\text{eff}}e = (1)e$. Then, from Eq. (41.45), the corresponding energy is approximately $-(1)^2(13.6\ \text{eV})/3^2 = -1.51$ eV. This approximation is very close to the experimental value of -1.521 eV.

Looking again at Fig. 41.8, we see that the radial probability density for the $3p$ state (for which $l = n - 2$) has two peaks and that for the $3s$ state $(l = n - 3)$ has three peaks. For sodium the first small peak in the $3p$ distribution gives a $3p$ electron a higher probability (compared to the $3d$ state) of being *inside* the charge distributions for the electrons in the $n = 2$ states. That is, a $3p$ electron is less completely screened from the nucleus than is a $3d$ electron because it spends some of its time within the filled K and L shells. Thus for the $3p$ electrons, Z_{eff} is greater than unity. From Eq. (41.45) the $3p$ energy is lower (more negative)

DATA *SPEAKS*

Many-Electron Atoms and Electron States

When students were given a problem involving quantum-mechanical states in many-electron atoms, more than 32% gave an incorrect response. Common errors:

- **Confusion about quantum numbers.** There are limits on the values of the four quantum numbers n, l, m_l, and m_s. For a given n value, l can be no greater than $n - 1$; for a given l value, m_l can be no greater than l and no less than $-l$; and m_s has only two possible values, $+\frac{1}{2}$ and $-\frac{1}{2}$.
- **Confusion about electron subshells.** A *subshell* corresponds to a given value of n and l. The total number of electrons that can be present in a given subshell is $2(2l + 1)$ (that is, two possible values of m_s multiplied by $2l + 1$ possible values of m_l, from l through 0 to $-l$).

than the $3d$ energy of -1.521 eV. The actual value is -3.035 eV. A $3s$ electron spends even more time within the inner electron shells than a $3p$ electron does, giving an even larger Z_{eff} and an even more negative energy.

This discussion shows that the energy levels given by Eq. (41.45) depend on both the principal quantum number n and the orbital quantum number l. That's because the value of Z_{eff} is different for the $3s$ state $(n = 3, l = 0)$, the $3p$ state $(n = 3, l = 1)$, and the $3d$ state $(n = 3, l = 2)$.

EXAMPLE 41.9 DETERMINING Z_{eff} EXPERIMENTALLY

SOLUTION

The measured energy of a $3s$ state of sodium is -5.138 eV. Calculate the value of Z_{eff}.

SOLUTION

IDENTIFY and SET UP: Sodium has a single electron in the M shell outside filled K and L shells. The ten K and L electrons partially screen the single M electron from the $+11e$ charge of the nucleus; our goal is to determine the extent of this screening. We are given $n = 3$ and $E_n = -5.138$ eV, so we can use Eq. (41.45) to determine Z_{eff}.

EXECUTE: Solving Eq. (41.45) for Z_{eff}, we have

$$Z_{eff}^2 = -\frac{n^2E_n}{13.6\text{ eV}} = -\frac{3^2(-5.138\text{ eV})}{13.6\text{ eV}} = 3.40$$

$$Z_{eff} = 1.84$$

EVALUATE: The effective charge attracting a $3s$ electron is $1.84e$. Sodium's 11 protons are screened by an average of $11 - 1.84 = 9.16$ electrons instead of 10 electrons because the $3s$ electron spends some time within the inner (K and L) shells.

Each alkali metal (lithium, sodium, potassium, rubidium, and cesium) has one more electron than the corresponding noble gas (helium, neon, argon, krypton, and xenon). This extra electron is mostly outside the other electrons in the filled shells and subshells. Therefore all the alkali metals behave similarly to sodium.

EXAMPLE 41.10 ENERGIES FOR A VALENCE ELECTRON

SOLUTION

The valence electron in potassium has a $4s$ ground state. Calculate the approximate energy of the $n = 4$ state having the smallest Z_{eff}, and discuss the relative energies of the $4s$, $4p$, $4d$, and $4f$ states.

SOLUTION

IDENTIFY and SET UP: The state with the smallest Z_{eff} is the one in which the valence electron spends the most time outside the inner filled shells and subshells, so that it is most effectively screened from the charge of the nucleus. Once we have determined which state has the smallest Z_{eff}, we can use Eq. (41.45) to determine the energy of this state.

EXECUTE: A $4f$ state has $n = 4$ and $l = 3 = 4 - 1$. Thus it is the state of greatest orbital angular momentum for $n = 4$, and thus the state in which the electron spends the most time outside the electron charge clouds of the inner filled shells and subshells. This makes Z_{eff} for a $4f$ state close to unity. Equation (41.45) then gives

$$E_4 = -\frac{Z_{eff}^2}{n^2}(13.6\text{ eV}) = -\frac{1}{4^2}(13.6\text{ eV}) = -0.85\text{ eV}$$

This approximation agrees with the measured energy of the sodium $4f$ state to the precision given.

An electron in a $4d$ state spends a bit more time within the inner shells, and its energy is therefore a bit more negative (measured to be -0.94 eV). For the same reason, a $4p$ state has an even lower energy (measured to be -2.73 eV) and a $4s$ state has the lowest energy (measured to be -4.339 eV).

EVALUATE: We can extend this analysis to the *singly ionized alkaline earth elements:* Be^+, Mg^+, Ca^+, Sr^+, and Ba^+. For any allowed value of n, the highest-l state $(l = n - 1)$ of the one remaining outer electron sees an effective charge of almost $+2e$, so for these states, $Z_{eff} = 2$. A $3d$ state for Mg^+, for example, has an energy of about $-2^2(13.6\text{ eV})/3^2 = -6.0$ eV.

TEST YOUR UNDERSTANDING OF SECTION 41.6 If electrons did *not* obey the exclusion principle, would it be easier or more difficult to remove the first electron from sodium? ▮

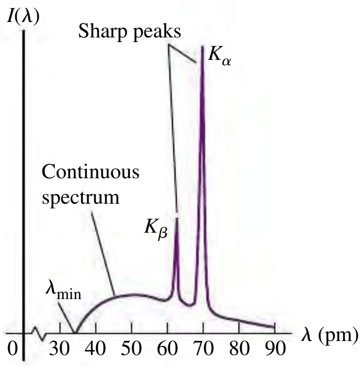
41.7 X-RAY SPECTRA

X-ray spectra provide an example of the richness and power of the model of atomic structure that we derived in the preceding section. In Section 38.2 we discussed how x-ray photons are produced when electrons strike a metal target (see Fig. 38.7). In this section we'll see how the spectrum of x rays produced in this way depends on the type of metal used in the target and how the ideas of atomic energy levels and screening help us understand this dependence.

Characteristic X Rays and Atomic Energy Levels

X-ray diffraction techniques (see Section 36.6) make it possible to measure x-ray wavelengths quite precisely (to within 0.1% or less). **Figure 41.23** shows the spectrum of x rays produced when fast-moving electrons strike a target of the metal molybdenum. This spectrum has two features:

41.23 Graph of intensity per unit wavelength as a function of wavelength for x rays produced with an accelerating voltage of 35 kV and a molybdenum target. The curve is a smooth function similar to the bremsstrahlung spectra in Fig. 38.8 (Section 38.2), but with two sharp spikes corresponding to part of the characteristic x-ray spectrum for molybdenum.



1. There is a *continuous* spectrum of wavelengths (see Fig. 38.8 in Section 38.2), with a minimum wavelength (corresponding to a maximum frequency and a maximum photon energy) that is determined by the voltage V_{AC} used to accelerate the electrons. As we saw in Section 38.2, this continuous spectrum is due to *bremsstrahlung*, in which the electrons slow down as they interact with the metal atoms in the target and convert their kinetic energy into photons. The minimum wavelength λ_{min} corresponds to all of the kinetic energy eV_{AC} of the electron being converted into the energy of a single photon of energy hc/λ_{min}, so

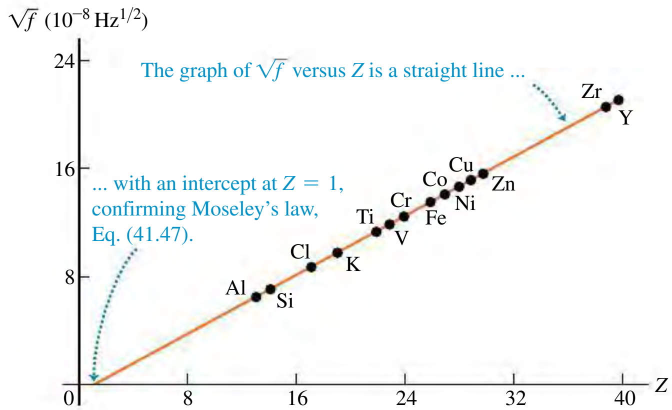
$$\lambda_{min} = \frac{hc}{eV_{AC}} \tag{41.46}$$

   This continuous-spectrum radiation is nearly independent of the target material in the x-ray tube.

2. Depending on the accelerating voltage, sharp peaks may be superimposed on the continuous bremsstrahlung spectrum, as in Fig. 41.23. These peaks are caused when the target atoms are struck by high-energy electrons and emit x rays of very definite wavelengths. Unlike the continuous spectrum, the wavelengths of the peaks are *different* for different target elements; they form what is called a *characteristic x-ray spectrum* for each target element.

In 1913 the English physicist Henry G. J. Moseley undertook a careful experimental study of characteristic x-ray spectra. He found that the most intense short-wavelength line in the characteristic x-ray spectrum from a particular target element, called the K_α line, varied smoothly with that element's atomic number Z (**Fig. 41.24**). This is in sharp contrast to optical spectra, in which elements with adjacent Z-values have spectra that often bear no resemblance to each other.

41.24 The square root of Moseley's measured frequencies of the K_α line for 14 elements.



Moseley found that the relationship could be expressed in terms of x-ray frequencies f by a simple formula called *Moseley's law:*

Moseley's law: Frequency of K_α line in characteristic x-ray spectrum of an element

$$f = (2.48 \times 10^{15}\ \text{Hz})(Z - 1)^2 \qquad (41.47)$$

Atomic number of element

Moseley went far beyond this empirical relationship; he showed how characteristic x-ray spectra could be understood on the basis of energy levels of atoms in the target. His analysis was based on the Bohr model, published in the same watershed year of 1913. We will recast it somewhat, using the ideas of atomic structure that we discussed in Section 41.6. First recall that the *outer* electrons of an atom are responsible for optical spectra. Their excited states are usually only a few electron volts above their ground state. In transitions from excited states to the ground state, they usually emit photons in or near the visible region.

Characteristic x rays, by contrast, are emitted in transitions involving the *inner* shells of a complex atom. In an x-ray tube the electrons may strike the target with enough energy to knock electrons out of the inner shells of the target atoms. These inner electrons are much closer to the nucleus than are the electrons in the outer shells; they are much more tightly bound, and hundreds or thousands of electron volts may be required to remove them.

Suppose one electron is knocked out of the K shell. This process leaves a vacancy, which we'll call a *hole*. (One electron remains in the K shell.) The hole can then be filled by an electron falling in from one of the outer shells, such as the $L, M, N, \ldots$ shell. This transition is accompanied by a decrease in the energy of the atom (because *less* energy would be needed to remove an electron from an $L, M, N, \ldots$ shell), and an x-ray photon is emitted with energy equal to this decrease. Each state has definite energy, so the emitted x rays have definite wavelengths; the emitted spectrum is a *line* spectrum.

We can estimate the energy and frequency of K_α x-ray photons by using the concept of screening from Section 41.6. A K_α x-ray photon is emitted when an electron in the L shell $(n = 2)$ drops down to fill a hole in the K shell $(n = 1)$. As the electron drops down, it is attracted by the Z protons in the nucleus screened by the one remaining electron in the K shell. We therefore approximate the energy by Eq. (41.45), with $Z_{\text{eff}} = Z - 1$, $n_\text{i} = 2$, and n_f. The energy before the transition is

$$E_i \approx -\frac{(Z-1)^2}{2^2}(13.6\ \text{eV}) = -(Z-1)^2(3.4\ \text{eV})$$

and the energy after the transition is

$$E_\text{f} \approx -\frac{(Z-1)^2}{1^2}(13.6\ \text{eV}) = -(Z-1)^2(13.6\ \text{eV})$$

$E_{K_\alpha} = E_\text{i} - E_\text{f} \approx (Z-1)^2(-3.4\ \text{eV} + 13.6\ \text{eV})$ is the energy of the K_α x-ray photon. That is,

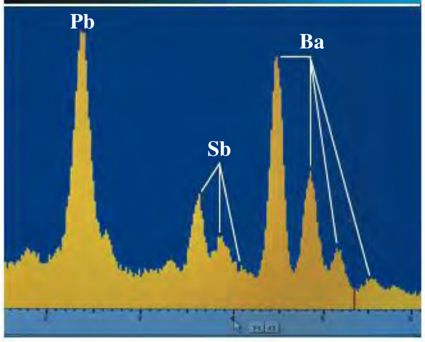
$$E_{K_\alpha} \approx (Z-1)^2(10.2\ \text{eV}) \qquad (41.48)$$

The frequency of the photon is its energy divided by Planck's constant:

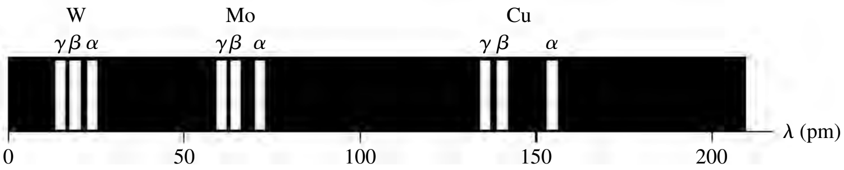
$$f = \frac{E}{h} \approx \frac{(Z-1)^2(10.2\ \text{eV})}{4.136 \times 10^{-15}\ \text{eV}\cdot\text{s}} = (2.47 \times 10^{15}\ \text{Hz})(Z-1)^2$$

This relationship agrees almost exactly with Moseley's experimental law, Eq. (41.47). Indeed, considering the approximations we have made, the agreement is better than we have a right to expect. But our calculation does show how Moseley's law can be understood on the bases of screening and transitions between energy levels.

Application X Rays in Forensic Science When a handgun is fired, a cloud of gunshot residue (GSR) is ejected from the barrel. The x-ray emission spectrum of GSR includes characteristic peaks from lead (Pb), antimony (Sb), and barium (Ba). If a sample taken from a suspect's skin or clothing has an x-ray emission spectrum with these characteristics, it indicates that the suspect recently fired a gun.



41.25 Wavelengths of the K_α, K_β, and K_γ lines of tungsten (W), molybdenum (Mo), and copper (Cu).



The three lines in each series are called the K_α, K_β, and K_γ lines. The K_α line is produced by the transition of an L electron to the vacancy in the K shell, the K_β line by an M electron, and the K_γ line by an N electron.

The hole in the K shell may also be filled by an electron falling from the M or N shell, assuming that these are occupied. If so, the x-ray spectrum of a large group of atoms of a single element shows a series, named the K series, of three lines, called the K_α, K_β, and K_γ lines. These three lines result from transitions in which the K-shell hole is filled by an L, M, or N electron, respectively. **Figure 41.25** shows the K series for tungsten $(Z = 74)$, molybdenum $(Z = 42)$, and copper $(Z = 29)$.

There are other series of x-ray lines, called the L, M, and N series, that are produced after the ejection of electrons from the L, M, and N shells rather than the K shell. Electrons in these outer shells are farther away from the nucleus and are not held as tightly as are those in the K shell, so removing these outer electrons requires less energy. Hence the x-ray photons that are emitted when these vacancies are filled have lower energy than those in the K series.

EXAMPLE 41.11 CHEMICAL ANALYSIS BY X-RAY EMISSION

You measure the K_α wavelength for an unknown element, obtaining the value 0.0709 nm. What is the element?

SOLUTION

IDENTIFY and SET UP: To determine which element this is, we need to know its atomic number Z. We can find this by using Moseley's law, which relates the frequency of an element's K_α x-ray emission line to that element's atomic number Z. We'll use the relationship $f = c/\lambda$ to calculate the frequency for the K_α line, and then use Eq. (41.47) to find the corresponding value of the atomic number Z. We'll then consult the periodic table (Appendix D) to determine which element has this atomic number.

EXECUTE: The frequency is

$$f = \frac{c}{\lambda} = \frac{3.00 \times 10^{8}\ \text{m/s}}{0.0709 \times 10^{-9}\ \text{m}} = 4.23 \times 10^{18}\ \text{Hz}$$

Solving Moseley's law for Z, we get

$$Z = 1 + \sqrt{\frac{f}{2.48 \times 10^{15}\ \text{Hz}}} = 1 + \sqrt{\frac{4.23 \times 10^{18}\ \text{Hz}}{2.48 \times 10^{15}\ \text{Hz}}} = 42.3$$

We know that Z has to be an integer; we conclude that $Z = 42$, corresponding to the element molybdenum.

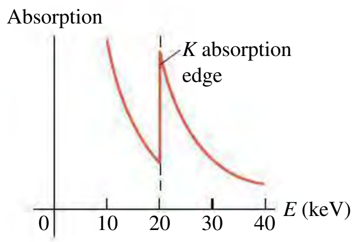
EVALUATE: If you're worried that our calculation did not give an integer for Z, remember that Moseley's law is an empirical relationship. There are slight variations from one atom to another due to differences in the structure of the electron shells. Nonetheless, this example suggests the power of Moseley's law.

Niels Bohr commented that it was Moseley's observations, not the alpha-particle scattering experiments of Rutherford, Geiger, and Marsden (see Section 39.2), that truly convinced physicists that the atom consists of a positive nucleus surrounded by electrons in motion. Unlike Bohr or Rutherford, Moseley did not receive a Nobel Prize for his important work; these awards are given to living scientists only, and Moseley (at age 27) was killed in combat during the First World War.

X-Ray Absorption Spectra

We can also observe x-ray *absorption* spectra. Unlike optical spectra, the absorption wavelengths are usually not the same as those for emission, especially in many-electron atoms, and do not give simple line spectra. For example, the K_α emission line results from a transition from the L shell to a hole in the K shell. The reverse transition doesn't occur in atoms with $Z \geq 10$ because in the atom's ground state, there is no vacancy in the L shell. To be absorbed, a photon must have enough energy to move an electron to an empty state. Since empty states are only a few electron volts in energy below the free-electron continuum, the minimum absorption energies in many-electron atoms are about the same as the minimum energies that are needed to remove an electron from its shell. Experimentally, if we gradually increase the accelerating voltage and hence the maximum photon energy, we observe sudden increases in absorption when we reach these minimum energies. These sudden jumps of absorption are called *absorption edges* (**Fig. 41.26**).

Characteristic x-ray spectra provide a very useful analytical tool. Satellite-borne x-ray spectrometers are used to study x-ray emission lines from highly excited atoms in distant astronomical sources. X-ray spectra are also used in air-pollution monitoring and in studies of the abundance of various elements in rocks.



41.26 When a beam of x rays is passed through a slab of molybdenum, the extent to which the beam is absorbed depends on the energy E of the x-ray photons. A sharp increase in absorption occurs at the K absorption edge at 20 keV. The increase occurs because photons with energies above this value can excite an electron from the K shell of a molybdenum atom into an empty state.

TEST YOUR UNDERSTANDING OF SECTION 41.7 A beam of photons is passed through a sample of high-temperature atomic hydrogen. At what photon energy would you expect there to be an absorption edge like that shown in Fig. 41.26? (i) 13.60 eV; (ii) 3.40 eV; (iii) 1.51 eV; (iv) all of these; (v) none of these. ❙

41.8 QUANTUM ENTANGLEMENT

We've seen that quantum mechanics is very successful at correctly predicting the results of experiments. As we will see in the remaining chapters, quantum mechanics is the basis of all electronic devices and is essential to our understanding of atomic nuclei and subatomic particles. But even though the central ideas of quantum mechanics have been established for decades, some aspects of the theory continue to baffle physicists and remain topics of ongoing research. We close this chapter with a discussion of one of these topics, *quantum entanglement*.

The Wave Function for Two Identical Particles

To understand what is meant by "quantum entanglement," let's consider how to write the wave function for two identical particles, such as two electrons (the electrons in the neutral helium atom, for instance). We'll use the subscripts 1 and 2 to refer to these particles.

As we discussed in Section 41.6, a wave function that describes both particles is a function of both the coordinates (x_1, y_1, z_1) of particle 1 and the coordinates (x_2, y_2, z_2) of particle 2. As a shorthand, we'll use the position vectors $\vec{r}_1 = x_1\hat{\imath} + y_1\hat{\jmath} + z_1\hat{k}$ and $\vec{r}_2 = x_2\hat{\imath} + y_2\hat{\jmath} + z_2\hat{k}$. In terms of these vectors, we can write the time-dependent two-particle wave function as $\Psi(\vec{r}_1, \vec{r}_2, t)$. Just as for a single particle, if the system of two particles has total energy E, we can write this time-dependent wave function as the product of a time-*independent* two-particle wave function $\psi(\vec{r}_1, \vec{r}_2)$ and a factor that depends on time only:

$$\Psi(\vec{r}_1, \vec{r}_2, t) = \psi(\vec{r}_1, \vec{r}_2)e^{-iEt/\hbar} \tag{41.49}$$

We saw in Section 41.2 that a technique called *separation of variables* is useful for expressing a wave function that depends on several variables as a product of functions of the individual variables. Let's see whether we can express the time-independent two-particle wave function $\psi(\vec{r}_1, \vec{r}_2)$ in Eq. (41.49) as a product of a function of $\vec{r}_1$ and a function of $\vec{r}_2$. We interpret the function of $\vec{r}_1$ to be the *single*-particle wave function for particle 1 and the function of $\vec{r}_2$ to be the single-particle wave function for particle 2. Suppose particle 1 is in a state A, for which the single-particle wave function is ψ_A, and particle 2 is in a state B, for which the single-particle wave function is ψ_B. [In the hydrogen atom, with two electrons, one electron could be in the spin-up state ($n = 1$, $l = 0$, $m_l = 0$, and $m_s = +\frac{1}{2}$) and the other could be in the spin-down state ($n = 1$, $l = 0$, $m_l = 0$, and $m_s = -\frac{1}{2}$).] Using separation of variables, we would write

$$\psi(\vec{r}_1, \vec{r}_2) = \psi_A(\vec{r}_1)\psi_B(\vec{r}_2) \tag{41.50}$$

(first guess at the two-particle wave function)

However, Eq. (41.50) *cannot* be correct, because particles 1 and 2 are *identical* and *indistinguishable*. We may be able to state with confidence that one particle is in state A and the other is in state B, but it's impossible to specify which particle is in which state. (There's no way even in principle to "tag" the particles.)

To account for this, let's make an improved guess for the two-particle wave function: a combination of two terms like Eq. (41.50)—one term for which particle 1 is in state A and particle 2 is in state B, and one term for which particle 1 is in state B and particle 2 is in state A. Our improved guess is then

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_A(\vec{r}_1)\psi_B(\vec{r}_2) \pm \psi_B(\vec{r}_1)\psi_A(\vec{r}_2)] \qquad (41.51)$$

(second guess at the two-particle wave function)

The factor $1/\sqrt{2}$ in Eq. (41.51) ensures that if ψ_A and ψ_B are normalized, then $\psi(\vec{r}_1, \vec{r}_2)$ will be normalized as well. Note that the terms $\psi_A(\vec{r}_1)\psi_B(\vec{r}_2)$ and $\psi_B(\vec{r}_1)\psi_A(\vec{r}_2)$ appear with equal magnitudes, so that the two possibilities (particle 1 in A and particle 2 in B, or particle 1 in B and particle 2 in A) are equally probable.

How can we decide whether the $\pm$ sign in Eq. (41.51) should be a plus or a minus? If the particles are two electrons, or two of any other type of spin-$\frac{1}{2}$ particle, the Pauli exclusion principle (Section 41.6) tells us that we must use the minus sign:

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_A(\vec{r}_1)\psi_B(\vec{r}_2) - \psi_B(\vec{r}_1)\psi_A(\vec{r}_2)] \qquad (41.52)$$

(two-particle wave function, spin-$\frac{1}{2}$ particles)

To check this, suppose we demand that *both* particles be in the same state A. Then we would replace ψ_B in Eq. (41.52) with ψ_A:

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_A(\vec{r}_1)\psi_A(\vec{r}_2) - \psi_A(\vec{r}_1)\psi_A(\vec{r}_2)] = 0$$

The zero value of the wave function says that our demand cannot be met. This is in agreement with the Pauli exclusion principle: No two electrons, and indeed no two identical spin-$\frac{1}{2}$ particles of any kind, can occupy the same quantum-mechanical state.

Measurement and "Spooky Action at a Distance"

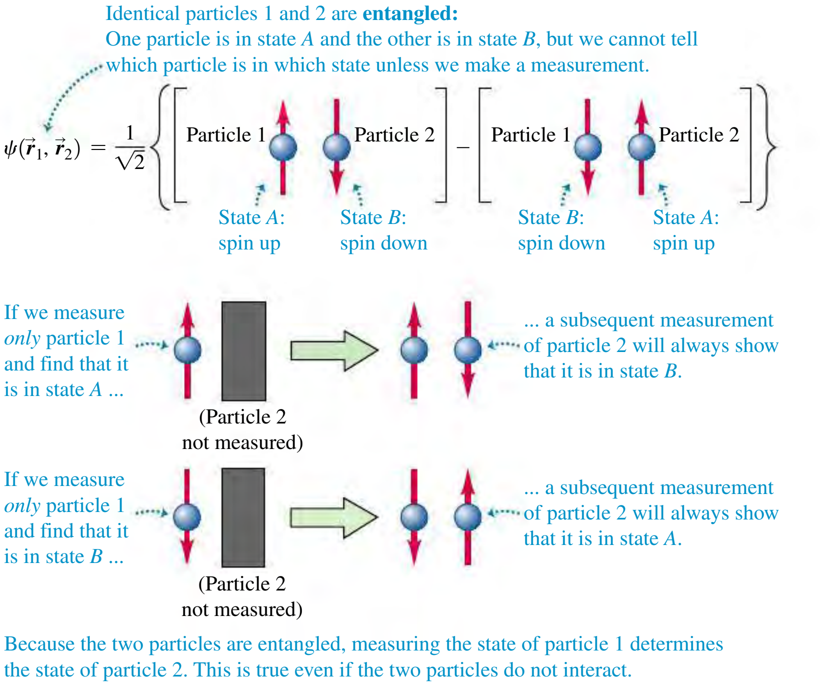
The wave function in Eq. (41.52) shares features with that of the quantum-mechanical particle in a box that we discussed in Section 40.6. That particle's wave function is also a combination of two terms of equal magnitude representing different situations: one in which the momentum of the particle is in the $+x$-direction, so $p_x > 0$, and one in which its momentum is in the $-x$-direction, so $p_x < 0$. We saw in Section 40.6 that if we *measured* the momentum of the particle, we would get either $p_x > 0$ or $p_x < 0$. Making such a measurement causes the wave function to *collapse,* and only the term corresponding to the measured value of p_x survives. The other term disappears from the wave function.

The same sort of wave-function collapse happens in our two-particle system. Suppose we make a measurement of one particle—call it particle 1—and determine that it is in state A. The measurement causes the wave function in Eq. (41.52) to collapse to $\psi(\vec{r}_1, \vec{r}_2) = \psi_A(\vec{r}_1)\psi_B(\vec{r}_2)$. This wave equation corresponds to particle 1 being in state A but also corresponds to particle 2 being in state B. It follows that, after the measurement, particle 2 *must* be in state B, even though we have not directly measured the state of particle 2. In other words, making a measurement on *one* particle affects the state of the *other* particle. Schrödinger described this situation by saying that the two particles are *entangled*.

If the two entangled particles are the two electrons within a helium atom, the idea that their states are entangled may not seem troubling. After all, these electrons are in very close proximity (a helium atom is only about 0.1 nm in diameter) and exert substantial electric forces on each other. You might imagine that when we measure electron 1 to be in state A (say, spin up with $m_s = +\frac{1}{2}$), it exerts forces on electron 2 that require electron 2 to be in state B (say, spin down with $m_s = -\frac{1}{2}$).

But suppose we arrange for two identical particles to be in an entangled state in which the particles are *not* close to each other, so they cannot exert forces on each other. When the same kind of measurement experiment is done on such a distant pair of entangled particles, the result is the same as if they are close together: If we measure particle 1 to be in state A and subsequently make a measurement on particle 2, we always find that particle 2 is in state B. If instead we measure particle 1 to be in state B and then make a measurement on particle 2, we always find that particle 2 is in state A. So measuring the state of one particle affects the state of the other particle, even when the two particles *cannot* exert forces on each other (**Fig. 41.27**). This finding has been confirmed with entangled particles that are *more than 300 km apart!* (These experiments with very large distances are done with photons rather than electrons. Like electrons, photons have spin, and the "spin-up" and "spin-down" states correspond to left and right circular polarization. The only difference is that photons are spin-1 particles, not spin-$\frac{1}{2}$, and do not obey the Pauli exclusion principle. As a result, we must use the plus sign rather than the minus sign in Eq. (41.51) to describe two entangled photons. The rest of the physics is identical, however.)

These results contradict the idea of *locality*—the notion that a particle responds to forces or fields that act at its position only, not at some other point in space. We used locality in Chapters 4 and 13 when we expressed the gravitational force on a particle of mass m as $\vec{F}_g = m\vec{g}$, where $\vec{g}$ is the acceleration due to gravity at the point in space where the particle is located. We used locality again in Chapters 21 and 27 when we wrote the electric force $\vec{F}_E$ and the magnetic force $\vec{F}_B$ on a particle of charge q moving with velocity $\vec{v}$ as $\vec{F}_E = q\vec{E}$ and $\vec{F}_B = q\vec{v} \times \vec{B}$, where $\vec{E}$ and $\vec{B}$, respectively, are the electric and magnetic fields at the position of the particle. But the interaction between two entangled, widely separated particles seems *not* to obey locality. For this reason, Albert Einstein



41.27 If two particles are in an entangled state, making a measurement of one particle determines the result of a subsequent measurement of the other particle.

referred to the results of an experiment like that shown in Fig. 41.27 as "spooky action at a distance." Spooky or not, quantum mechanics appears to be intrinsically nonlocal.

What makes these results even more striking is that no matter how far apart the two entangled particles are, there appears to be *zero delay* between the time that we make a measurement on one particle and the time that the state of the other particle changes as a result. At first glance this seems to violate a key idea of Einstein's special theory of relativity: that signals of any kind—radio waves, light signals, or beams of particles—cannot travel faster than the speed of light in vacuum, c. If measuring the state of particle 1 in Fig. 41.27 causes the state of particle 2 to change instantaneously, couldn't we make a "quantum radio" that sends signals faster than c, with particle 1 as the transmitter and particle 2 as the receiver?

The answer is no. The "message" in our quantum radio would be the result of a measurement of particle 1 by a physicist (call her Primo) at that particle's position. Primo's measurement collapses the wave function of the two particles, and her result would be that particle 1 is in either state A or state B. Another physicist (call him Secondo) at the position of particle 2 would measure particle 2 to be in state B if Primo measured particle 1 to be in state A, and to be in state A if Primo measured particle 1 to be in state B. But Secondo would have no way of knowing whether his result was caused by Primo making a measurement first or by *Secondo* himself making an independent measurement of particle 2 *without* Primo having made any measurement. (Secondo could determine this later by, for instance, sending text messages back and forth with Primo. But that method of communication involves signals that travel at the speed of light, not instantaneously.) So our quantum radio would transmit no information at all and would not allow us to communicate at speeds faster than c.

A remarkable practical application of quantum entanglement is *quantum computing*. In a conventional ("classical") computer, the memory is made up of *bits*. Each bit has only two possible values (say, 0 or 1), so a computer memory with N bits can have any of 2^N different configurations. (This is analogous to coins that can be either heads up or tails up. Figure 20.21 in Section 20.8 shows the possible configurations of four coins; the number of possibilities is $2^4 = 16$.) In a quantum computer, bits are replaced with *qubits* (short for "quantum bits"). An example is a spin-$\frac{1}{2}$ electron that can be in a spin-up state $\left(m_s = +\frac{1}{2}\right)$ or a spin-down state $\left(m_s = -\frac{1}{2}\right)$, as usual, but can also be in *any combination* of these states. The wave function of N entangled qubits can correspond to any of 2^N configurations (like ordinary bits or coins that can be heads up or tails up) or to an entangled state in which the qubits are in any combination of these configurations. So, unlike a classical computer memory, which can be in only one of its 2^N configurations at a time, a quantum computer memory can essentially be in *all* of these configurations simultaneously. This holds the promise of the ability to do certain types of computations, such as those involved in breaking codes in cryptography, much more rapidly than a classical computer could. As of this writing, the quest to build a fully quantum computer is still in its early stages, but intensive research is under way and rapid progress is being made.

TEST YOUR UNDERSTANDING OF SECTION 41.8 Particle 1 is an electron that can be in state C or D. Particle 2 is a proton that can be in state E or F. Is $\psi(\vec{r}_1, \vec{r}_2) = (1/\sqrt{2})\left[\psi_C(\vec{r}_1)\psi_E(\vec{r}_2) + \psi_C(\vec{r}_1)\psi_F(\vec{r}_2)\right]$ a possible wave function for this two-particle system? If so, does it represent an entangled state? ▌

SOLUTIONS TO ALL EXAMPLES

Three-dimensional problems: The time-independent Schrödinger equation for three-dimensional problems is given by Eq. (41.5).

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi(x,y,z)}{\partial x^2}+\frac{\partial^2\psi(x,y,z)}{\partial y^2}+\frac{\partial^2\psi(x,y,z)}{\partial z^2}\right)+U(x,y,z)\psi(x,y,z)$$
$$=E\psi(x,y,z)$$

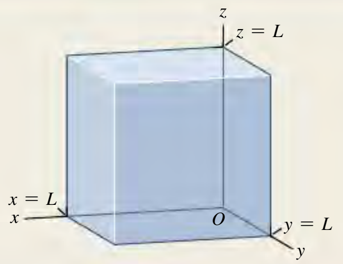
(three-dimensional time-independent Schrödinger equation) (41.5)

Particle in a three-dimensional box: The wave function for a particle in a cubical box is the product of a function of x only, a function of y only, and a function of z only. Each stationary state is described by three quantum numbers (n_X, n_Y, n_Z). Most of the energy levels given by Eq. (41.16) exhibit degeneracy: More than one quantum state has the same energy. (See Example 41.1.)

$$E_{n_X,n_Y,n_Z}=\frac{(n_X{}^2+n_Y{}^2+n_Z{}^2)\pi^2\hbar^2}{2mL^2}$$

$(n_X = 1, 2, 3, \ldots;$
$n_Y = 1, 2, 3, \ldots;$
$n_Z = 1, 2, 3, \ldots)$

(energy levels, particle in a three-dimensional cubical box) (41.16)



The hydrogen atom: The Schrödinger equation for the hydrogen atom gives the same energy levels as the Bohr model. If the nucleus has charge Ze, there is an additional factor of Z^2 in the numerator of Eq. (41.21). The possible magnitudes L of orbital angular momentum are given by Eq. (41.22), and the possible values of the z-component of orbital angular momentum are given by Eq. (41.23). (See Examples 41.2 and 41.3.)

The probability that an atomic electron is between r and $r + dr$ from the nucleus is $P(r)\,dr$, given by Eq. (41.25). Atomic distances are often measured in units of a, the smallest distance between the electron and the nucleus in the Bohr model. (See Example 41.4.)

$$E_n=-\frac{1}{(4\pi\epsilon_0)^2}\frac{m_r e^4}{2n^2\hbar^2}=-\frac{13.60\text{ eV}}{n^2}$$

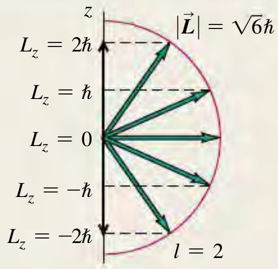
(energy levels of hydrogen) (41.21)

$$L=\sqrt{l(l+1)}\hbar$$
$$(l=0,1,2,\ldots,n-1) \qquad (41.22)$$

$$L_z=m_l\hbar$$
$$(m_l=0,\pm1,\pm2,\ldots,\pm l) \qquad (41.23)$$

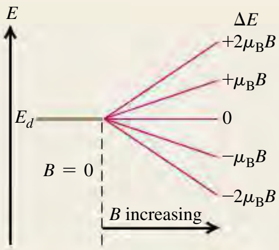
$$P(r)\,dr=|\psi|^2\,dV=|\psi|^2\,4\pi r^2\,dr \qquad (41.25)$$

$$a=\frac{\epsilon_0 h^2}{\pi m_r e^2}=\frac{4\pi\epsilon_0\hbar^2}{m_r e^2}$$
$$=5.29\times10^{-11}\text{ m} \qquad (41.26)$$



The Zeeman effect: The interaction energy of an electron (mass m) with magnetic quantum number m_l in a magnetic field $\vec{B}$ along the $+z$-direction is given by Eq. (41.35), where $\mu_B = e\hbar/2m$ is called the Bohr magneton. (See Example 41.5.)

$$U=-\mu_z B=m_l\frac{e\hbar}{2m}B=m_l\mu_B B$$
$$(m_l=0,\pm1,\pm2,\ldots,\pm l) \qquad (41.35)$$



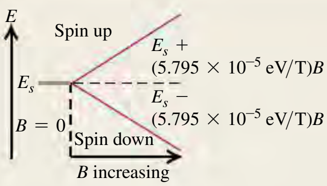
Electron spin: An electron has an intrinsic spin angular momentum of magnitude S, given by Eq. (41.37). The possible values of the z-component of the spin angular momentum are $S_z = m_s\hbar$, where $m_s = \pm\frac{1}{2}$. (See Examples 41.6 and 41.7.)

An orbiting electron experiences an interaction between its spin and the effective magnetic field produced by the relative motions of electron and nucleus. This spin-orbit coupling, along with relativistic effects, splits the energy levels according to their total angular momentum quantum number j. (See Example 41.8.)

$$S=\sqrt{\tfrac{1}{2}\left(\tfrac{1}{2}+1\right)}\hbar=\sqrt{\tfrac{3}{4}}\hbar \qquad (41.37)$$

$$S_z=m_s\hbar \quad \left(m_s=\pm\tfrac{1}{2}\right) \qquad (41.36)$$

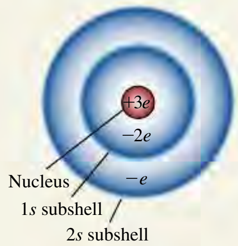
$$E_{n,j}=-\frac{13.60\text{ eV}}{n^2}\left[1+\frac{\alpha^2}{n^2}\left(\frac{n}{j+\frac{1}{2}}-\frac{3}{4}\right)\right] \qquad (41.41)$$



Many-electron atoms: In a hydrogen atom, the quantum numbers n, l, m_l, and m_s of the electron have certain allowed values given by Eq. (41.44). In a many-electron atom, the allowed quantum numbers for each electron are the same as in hydrogen, but the energy levels depend on both n and l because of screening, the partial cancellation of the field of the nucleus by the inner electrons. If the effective (screened) charge attracting an electron is $Z_{\text{eff}}e$, the energies of the levels are given approximately by Eq. (41.45). (See Examples 41.9 and 41.10.)

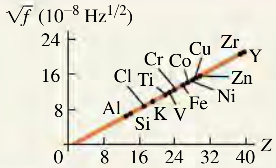
$$n \geq 1 \quad 0 \leq l \leq n - 1$$
$$|m_l| \leq l \quad m_s = \pm\tfrac{1}{2} \tag{41.44}$$

$$E_n = -\frac{Z_{\text{eff}}^2}{n^2}(13.6\ \text{eV}) \tag{41.45}$$

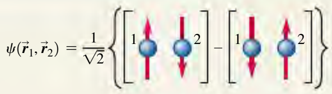


X-ray spectra: Moseley's law states that the frequency of a K_α x ray from a target with atomic number Z is given by Eq. (41.47). Characteristic x-ray spectra result from transitions to a hole in an inner energy level of an atom. (See Example 41.11.)

$$f = (2.48 \times 10^{15}\ \text{Hz})(Z - 1)^2 \tag{41.47}$$



Quantum entanglement: The wave function of two identical particles can be such that neither particle is itself in a definite state. For example, the wave function could be a combination of one term with particle 1 in state A and particle 2 in state B and one term with particle 1 in state B and particle 2 in state A. The two particles are said to be entangled, since measuring the state of one particle automatically determines the results of subsequent measurements of the other particle.



BRIDGING PROBLEM A MANY-ELECTRON ATOM IN A BOX

SOLUTION

An atom of titanium (Ti) has 22 electrons and has a radius of 1.47×10^{-10} m. As a simple model of this atom, imagine putting 22 electrons into a cubical box that has the same volume as a titanium atom. (a) What is the length of each side of the box? (b) What will be the configuration of the 22 electrons? (c) Find the energies of each of the levels occupied by the electrons. (Ignore the electric forces that the electrons exert on each other.) (d) You remove one of the electrons from the lowest level. As a result, one of the electrons from the highest occupied level drops into the lowest level to fill the hole, emitting a photon in the process. What is the energy of this photon? How does this compare to the energy of the K_α photon for titanium as predicted by Moseley's law?

SOLUTION GUIDE

IDENTIFY and SET UP

1. In this problem you'll use ideas from Section 41.2 about a particle in a cubical box. You'll also apply the exclusion principle from Section 41.6 to find the electron configuration of this cubical "atom." The ideas about x-ray spectra from Section 41.7 are also important.
2. The target variables are (a) the dimensions of the box, (b) the electron configurations (like those given in Table 41.3 for real atoms), (c) the occupied energy levels of the cubical box, and (d) the energy of the emitted photon.

EXECUTE

3. Use your knowledge of geometry to find the length of each side of the box.
4. Each electron state is described by four quantum numbers: n_X, n_Y, and n_Z as described in Section 41.2 and the spin magnetic quantum number m_s described in Section 41.5. Use the exclusion principle to determine the quantum numbers of each of the 22 electrons in the "atom." (*Hint:* Figure 41.4 in Section 41.2 shows the first several energy levels of a cubical box relative to the ground level $E_{1,1,1}$.)
5. Use your results from steps 3 and 4 to find the energies of each of the occupied levels.
6. Use your result from step 5 to find the energy of the photon emitted when an electron makes a transition from the highest occupied level to the ground level. Compare this to the energy that we calculated for titanium by using Moseley's law.

EVALUATE

7. Is this cubical "atom" a useful model for titanium? Why or why not?
8. In this problem you ignored the electrical interactions between electrons. To estimate how large these are, find the electrostatic potential energy of two electrons separated by half the length of the box. How does this compare to the energy levels you calculated in step 5? Is it a good approximation to ignore these interactions?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q41.1 Particle A is described by the wave function $\psi(x, y, z)$. Particle B is described by the wave function $\psi(x, y, z)e^{i\phi}$, where ϕ is a real constant. How does the probability of finding particle A within a volume dV around a certain point in space compare with the probability of finding particle B within this same volume?

Q41.2 What are the most significant differences between the Bohr model of the hydrogen atom and the Schrödinger analysis? What are the similarities?

Q41.3 For a body orbiting the sun, such as a planet, comet, or asteroid, is there any restriction on the z-component of its orbital angular momentum such as there is with the z-component of the electron's orbital angular momentum in hydrogen? Explain.

Q41.4 Why is the analysis of the helium atom much more complex than that of the hydrogen atom, either in a Bohr type of model or using the Schrödinger equation?

Q41.5 The Stern–Gerlach experiment is always performed with beams of *neutral* atoms. Wouldn't it be easier to form beams using *ionized* atoms? Why won't this work?

Q41.6 (a) If two electrons in hydrogen atoms have the same principal quantum number, can they have different orbital angular momenta? How? (b) If two electrons in hydrogen atoms have the same orbital quantum number, can they have different principal quantum numbers? How?

Q41.7 In the Stern–Gerlach experiment, why is it essential for the magnetic field to be *inhomogeneous* (that is, nonuniform)?

Q41.8 In the ground state of the helium atom one electron must have "spin down" and the other "spin up." Why?

Q41.9 An electron in a hydrogen atom is in an s level, and the atom is in a magnetic field $\vec{B} = B\hat{k}$. Explain why the "spin up" state $\left(m_s = +\frac{1}{2}\right)$ has a higher energy than the "spin down" state $\left(m_s = -\frac{1}{2}\right)$.

Q41.10 The central-field approximation is more accurate for alkali metals than for transition metals such as iron, nickel, or copper. Why?

Q41.11 Table 41.3 shows that for the ground state of the potassium atom, the outermost electron is in a 4s state. What does this tell you about the relative energies of the 3d and 4s levels for this atom? Explain.

Q41.12 Do gravitational forces play a significant role in atomic structure? Explain.

Q41.13 Why do the transition elements $(Z = 21 \text{ to } 30)$ all have similar chemical properties?

Q41.14 Use Table 41.3 to help determine the ground-state electron configuration of the neutral gallium atom (Ga) as well as the ions Ga^+ and Ga^-. Gallium has an atomic number of 31.

Q41.15 On the basis of the Pauli exclusion principle, the structure of the periodic table of the elements shows that there must be a fourth quantum number in addition to n, l, and m_l. Explain.

Q41.16 A small amount of magnetic-field splitting of spectral lines occurs even when the atoms are not in a magnetic field. What causes this?

Q41.17 The ionization energies of the alkali metals (that is, the lowest energy required to remove one outer electron when the atom is in its ground state) are about 4 or 5 eV, while those of the noble gases are in the range from 11 to 25 eV. Why is there a difference?

Q41.18 For magnesium, the first ionization potential is 7.6 eV. The second ionization potential (additional energy required to remove a second electron) is almost twice this, 15 eV, and the third ionization potential is much larger, about 80 eV. How can these numbers be understood?

Q41.19 What is the "central-field approximation" and why is it only an approximation?

Q41.20 The nucleus of a gold atom contains 79 protons. How does the energy required to remove a 1s electron completely from a gold atom compare with the energy required to remove the electron from the ground level in a hydrogen atom? In what region of the electromagnetic spectrum would a photon with this energy for each of these two atoms lie?

Q41.21 (a) Can you show that the orbital angular momentum of an electron in any given direction (e.g., along the z-axis) is *always* less than or equal to its total orbital angular momentum? In which cases would the two be equal to each other? (b) Is the result in part (a) true for a classical object, such as a spinning top or planet?

Q41.22 An atom in its ground level absorbs a photon with energy equal to the K absorption edge. Does absorbing this photon ionize this atom? Explain.

Q41.23 Can a hydrogen atom emit x rays? If so, how? If not, why not?

Q41.24 A system of two electrons has the wave function $\psi(\vec{r}_1, \vec{r}_2) = (1/\sqrt{2})[\psi_\alpha(\vec{r}_1)\psi_\beta(\vec{r}_2) - \psi_\beta(\vec{r}_1)\psi_\alpha(\vec{r}_2)]$, where ψ_α is a normalized wave function for a state with $S_z = +\frac{1}{2}\hbar$ and ψ_β is a normalized wave function for a state with $S_z = -\frac{1}{2}\hbar$. (a) If S_z for electron 1 is measured, what are the possible results? What is the probability of each result? (b) If S_z for electron 2 is measured, what are the possible results? What is the probability of each result? (c) If measurement of S_z for electron 1 yields the value $\frac{1}{2}\hbar$, what are the possible results of a subsequent measurement of S_z for electron 2? What is the probability of each result being obtained? Explain.

Q41.25 Repeat Discussion Question Q41.24 for the wave function $\psi(\vec{r}_1, \vec{r}_2) = \psi_\alpha(\vec{r}_1)\psi_\alpha(\vec{r}_2)$.

EXERCISES

Section 41.2 Particle in a Three-Dimensional Box

41.1 • For a particle in a three-dimensional cubical box, what is the degeneracy (number of different quantum states with the same energy) of the energy levels (a) $3\pi^2\hbar^2/2mL^2$ and (b) $9\pi^2\hbar^2/2mL^2$?

41.2 • **CP** Model a hydrogen atom as an electron in a cubical box with side length L. Set the value of L so that the volume of the box equals the volume of a sphere of radius $a = 5.29 \times 10^{-11}$ m, the Bohr radius. Calculate the energy separation between the ground and first excited levels, and compare the result to this energy separation calculated from the Bohr model.

41.3 • **CP** A photon is emitted when an electron in a three-dimensional cubical box of side length 8.00×10^{-11} m makes a transition from the $n_X = 2$, $n_Y = 2$, $n_Z = 1$ state to the $n_X = 1$, $n_Y = 1$, $n_Z = 1$ state. What is the wavelength of this photon?

41.4 • For each of the following states of a particle in a three-dimensional cubical box, at what points is the probability distribution function a maximum: (a) $n_X = 1$, $n_Y = 1$, $n_Z = 1$ and (b) $n_X = 2$, $n_Y = 2$, $n_Z = 1$?

41.5 •• A particle is in the three-dimensional cubical box of Section 41.1. For the state $n_X = 2$, $n_Y = 2$, $n_Z = 1$, for what planes (in addition to the walls of the box) is the probability distribution function zero? Compare this number of planes to the corresponding number of planes where $|\psi|^2$ is zero for the lower-energy state $n_X = 2$, $n_Y = 1$, $n_Z = 1$ and for the ground state $n_X = 1$, $n_Y = 1$, $n_Z = 1$.

41.6 • What is the energy difference between the two lowest energy levels for a proton in a cubical box with side length 1.00×10^{-14} m, the approximate diameter of a nucleus?

Section 41.3 The Hydrogen Atom

41.7 •• Consider an electron in the N shell. (a) What is the smallest orbital angular momentum it could have? (b) What is the largest orbital angular momentum it could have? Express your answers in terms of $\hbar$ and in SI units. (c) What is the largest orbital angular momentum this electron could have in any chosen direction? Express your answers in terms of $\hbar$ and in SI units. (d) What is the largest spin angular momentum this electron could have in any chosen direction? Express your answers in terms of $\hbar$ and in SI units. (e) For the electron in part (c), what is the ratio of its spin angular momentum in the z-direction to its orbital angular momentum in the z-direction?

41.8 • An electron is in the hydrogen atom with $n = 5$. (a) Find the possible values of L and L_z for this electron, in units of $\hbar$. (b) For each value of L, find all the possible angles between $\vec{L}$ and the z-axis. (c) What are the maximum and minimum values of the magnitude of the angle between $\vec{L}$ and the z-axis?

41.9 • The orbital angular momentum of an electron has a magnitude of $4.716 \times 10^{-34}\ \mathrm{kg \cdot m^2/s}$. What is the angular momentum quantum number l for this electron?

41.10 • Consider states with angular momentum quantum number $l = 2$. (a) In units of $\hbar$, what is the largest possible value of L_z? (b) In units of $\hbar$, what is the value of L? Which is larger: L or the maximum possible L_z? (c) For each allowed value of L_z, what angle does the vector $\vec{L}$ make with the $+z$-axis? How does the minimum angle for $l = 2$ compare to the minimum angle for $l = 3$ calculated in Example 41.3?

41.11 •• In a particular state of the hydrogen atom, the angle between the angular momentum vector $\vec{L}$ and the z-axis is $\theta = 26.6°$. If this is the smallest angle for this particular value of the orbital quantum number l, what is l?

41.12 •• A hydrogen atom is in a state that has $L_z = 2\hbar$. In the semiclassical vector model, the angular momentum vector $\vec{L}$ for this state makes an angle $\theta_L = 63.4°$ with the $+z$-axis. (a) What is the l quantum number for this state? (b) What is the smallest possible n quantum number for this state?

41.13 • Calculate, in units of $\hbar$, the magnitude of the maximum orbital angular momentum for an electron in a hydrogen atom for states with a principal quantum number of 2, 20, and 200. Compare each with the value of $n\hbar$ postulated in the Bohr model. What trend do you see?

41.14 • (a) Make a chart showing all possible sets of quantum numbers l and m_l for the states of the electron in the hydrogen atom when $n = 4$. How many combinations are there? (b) What are the energies of these states?

41.15 •• (a) How many different $5g$ states does hydrogen have? (b) Which of the states in part (a) has the largest angle between $\vec{L}$ and the z-axis, and what is that angle? (c) Which of the states in part (a) has the smallest angle between $\vec{L}$ and the z-axis, and what is that angle?

41.16 •• CALC (a) What is the probability that an electron in the $1s$ state of a hydrogen atom will be found at a distance less than $a/2$ from the nucleus? (b) Use the results of part (a) and of Example 41.4 to calculate the probability that the electron will be found at distances between $a/2$ and a from the nucleus.

41.17 • Show that $\Phi(\phi) = e^{im_l\phi} = \Phi(\phi + 2\pi)$ (that is, show that $\Phi(\phi)$ is periodic with period 2π) if and only if m_l is restricted to the values $0, \pm 1, \pm 2, \ldots$. (*Hint:* Euler's formula states that $e^{i\phi} = \cos\phi + i\sin\phi$.)

Section 41.4 The Zeeman Effect

41.18 • A hydrogen atom is in a d state. In the absence of an external magnetic field, the states with different m_l values have (approximately) the same energy. Consider the interaction of the magnetic field with the atom's orbital magnetic dipole moment. (a) Calculate the splitting (in electron volts) of the m_l levels when the atom is put in a 0.800-T magnetic field that is in the $+z$-direction. (b) Which m_l level will have the lowest energy? (c) Draw an energy-level diagram that shows the d levels with and without the external magnetic field.

41.19 • A hydrogen atom in a $3p$ state is placed in a uniform external magnetic field $\vec{B}$. Consider the interaction of the magnetic field with the atom's orbital magnetic dipole moment. (a) What field magnitude B is required to split the $3p$ state into multiple levels with an energy difference of 2.71×10^{-5} eV between adjacent levels? (b) How many levels will there be?

41.20 •• CP A hydrogen atom undergoes a transition from a $2p$ state to the $1s$ ground state. In the absence of a magnetic field, the energy of the photon emitted is 122 nm. The atom is then placed in a strong magnetic field in the z-direction. Ignore spin effects; consider only the interaction of the magnetic field with the atom's orbital magnetic moment. (a) How many different photon wavelengths are observed for the $2p \rightarrow 1s$ transition? What are the m_l values for the initial and final states for the transition that leads to each photon wavelength? (b) One observed wavelength is exactly the same with the magnetic field as without. What are the initial and final m_l values for the transition that produces a photon of this wavelength? (c) One observed wavelength with the field is longer than the wavelength without the field. What are the initial and final m_l values for the transition that produces a photon of this wavelength? (d) Repeat part (c) for the wavelength that is shorter than the wavelength in the absence of the field.

41.21 • A hydrogen atom in the $5g$ state is placed in a magnetic field of 0.600 T that is in the z-direction. (a) Into how many levels is this state split by the interaction of the atom's orbital magnetic dipole moment with the magnetic field? (b) What is the energy separation between adjacent levels? (c) What is the energy separation between the level of lowest energy and the level of highest energy?

Section 41.5 Electron Spin

41.22 •• A hydrogen atom in the $n = 1$, $m_s = -\frac{1}{2}$ state is placed in a magnetic field with a magnitude of 1.60 T in the $+z$-direction. (a) Find the magnetic interaction energy (in electron volts) of the electron with the field. (b) Is there any orbital magnetic dipole moment interaction for this state? Explain. Can there be an orbital magnetic dipole moment interaction for $n \neq 1$?

41.23 •• CP **Classical Electron Spin.** (a) If you treat an electron as a classical spherical object with a radius of 1.0×10^{-17} m, what angular speed is necessary to produce a spin angular momentum of magnitude $\sqrt{\frac{3}{4}}\hbar$? (b) Use $v = r\omega$ and the result of part (a) to calculate the speed v of a point at the electron's equator. What does your result suggest about the validity of this model?

41.24 • CP The hyperfine interaction in a hydrogen atom between the magnetic dipole moment of the proton and the spin magnetic dipole moment of the electron splits the ground level into two levels separated by 5.9×10^{-6} eV. (a) Calculate the wavelength and frequency of the photon emitted when the atom makes a transition between these states, and compare your answer to the value given at the end of Section 41.5. In what part of the electromagnetic spectrum does this lie? Such photons are emitted by cold hydrogen clouds in interstellar space; by detecting these photons, astronomers can learn about the number and density of such clouds. (b) Calculate the effective magnetic field experienced by the electron in these states (see Fig. 41.18). Compare your result to the effective magnetic field due to the spin-orbit coupling calculated in Example 41.7.

41.25 • Calculate the energy difference between the $m_s = \frac{1}{2}$ ("spin up") and $m_s = -\frac{1}{2}$ ("spin down") levels of a hydrogen atom in the 1s state when it is placed in a 1.45-T magnetic field in the *negative* z-direction. Which level, $m_s = \frac{1}{2}$ or $m_s = -\frac{1}{2}$, has the lower energy?

41.26 • A hydrogen atom in a particular orbital angular momentum state is found to have j quantum numbers $\frac{7}{2}$ and $\frac{9}{2}$. (a) What is the letter that labels the value of l for the state? (b) If $n = 5$, what is the energy difference between the $j = \frac{7}{2}$ and $j = \frac{9}{2}$ levels?

Section 41.6 Many-Electron Atoms and the Exclusion Principle

41.27 • Make a list of the four quantum numbers n, l, m_l, and m_s for each of the 10 electrons in the ground state of the neon atom. Do *not* refer to Table 41.2 or 41.3.

41.28 • For germanium (Ge, $Z = 32$), make a list of the number of electrons in each subshell (1s, 2s, 2p, . . .). Use the allowed values of the quantum numbers along with the exclusion principle; do *not* refer to Table 41.3.

41.29 •• (a) Write out the ground-state electron configuration ($1s^2, 2s^2, \ldots$) for the beryllium atom. (b) What element of next-larger Z has chemical properties similar to those of beryllium? Give the ground-state electron configuration of this element. (c) Use the procedure of part (b) to predict what element of next-larger Z than in (b) will have chemical properties similar to those of the element you found in part (b), and give its ground-state electron configuration.

41.30 •• (a) Write out the ground-state electron configuration ($1s^2, 2s^2, \ldots$) for the carbon atom. (b) What element of next-larger Z has chemical properties similar to those of carbon? Give the ground-state electron configuration for this element.

41.31 • The 5s electron in rubidium (Rb) sees an effective charge of 2.771e. Calculate the ionization energy of this electron.

41.32 • The energies of the 4s, 4p, and 4d states of potassium are given in Example 41.10. Calculate Z_{eff} for each state. What trend do your results show? How can you explain this trend?

41.33 • (a) The doubly charged ion N^{2+} is formed by removing two electrons from a nitrogen atom. What is the ground-state electron configuration for the N^{2+} ion? (b) Estimate the energy of the least strongly bound level in the L shell of N^{2+}. (c) The doubly charged ion P^{2+} is formed by removing two electrons from a phosphorus atom. What is the ground-state electron configuration for the P^{2+} ion? (d) Estimate the energy of the least strongly bound level in the M shell of P^{2+}.

41.34 • (a) The energy of the 2s state of lithium is -5.391 eV. Calculate the value of Z_{eff} for this state. (b) The energy of the 4s state of potassium is -4.339 eV. Calculate the value of Z_{eff} for this state. (c) Compare Z_{eff} for the 2s state of lithium, the 3s state of sodium (see Example 41.9), and the 4s state of potassium. What trend do you see? How can you explain this trend?

41.35 • Estimate the energy of the highest-l state for (a) the L shell of Be^{+} and (b) the N shell of Ca^{+}.

Section 41.7 X-Ray Spectra

41.36 • A K_α x ray emitted from a sample has an energy of 7.46 keV. Of which element is the sample made?

41.37 • Calculate the frequency, energy (in keV), and wavelength of the K_α x ray for the elements (a) calcium (Ca, $Z = 20$); (b) cobalt (Co, $Z = 27$); (c) cadmium (Cd, $Z = 48$).

41.38 •• The energies for an electron in the K, L, and M shells of the tungsten atom are $-69{,}500$ eV, $-12{,}000$ eV, and -2200 eV, respectively. Calculate the wavelengths of the K_α and K_β x rays of tungsten.

PROBLEMS

41.39 • In terms of the ground-state energy $E_{1,1,1}$, what is the energy of the highest level occupied by an electron when 10 electrons are placed into a cubical box?

41.40 •• An electron is in a three-dimensional box with side lengths $L_X = 0.600$ nm and $L_Y = L_Z = 2L_X$. What are the quantum numbers n_X, n_Y, and n_Z and the energies, in eV, for the four lowest energy levels? What is the degeneracy of each (including the degeneracy due to spin)?

41.41 •• CALC A particle is in the three-dimensional cubical box of Section 41.2. (a) Consider the cubical volume defined by $0 \le x \le L/4$, $0 \le y \le L/4$, and $0 \le z \le L/4$. What fraction of the total volume of the box is this cubical volume? (b) If the particle is in the ground state ($n_X = 1, n_Y = 1, n_Z = 1$), calculate the probability that the particle will be found in the cubical volume defined in part (a). (c) Repeat the calculation of part (b) when the particle is in the state $n_X = 2$, $n_Y = 1$, $n_Z = 1$.

41.42 ••• An electron is in a three-dimensional box. The x- and z-sides of the box have the same length, but the y-side has a different length. The two lowest energy levels are 2.24 eV and 3.47 eV, and the degeneracy of each of these levels (including the degeneracy due to the electron spin) is two. (a) What are the n_X, n_Y, and n_Z quantum numbers for each of these two levels? (b) What are the lengths L_X, L_Y, and L_Z for each side of the box? (c) What are the energy, the quantum numbers, and the degeneracy (including the spin degeneracy) for the next higher energy state?

41.43 •• CALC A particle in the three-dimensional cubical box of Section 41.2 is in the ground state, where $n_X = n_Y = n_Z = 1$. (a) Calculate the probability that the particle will be found somewhere between $x = 0$ and $x = L/2$. (b) Calculate the probability that the particle will be found somewhere between $x = L/4$ and $x = L/2$. Compare your results to the result of Example 41.1 for the probability of finding the particle in the region $x = 0$ to $x = L/4$.

41.44 •• CP CALC **A Three-Dimensional Isotropic Harmonic Oscillator.** An isotropic harmonic oscillator has the potential-energy function $U(x, y, z) = \frac{1}{2}k'(x^2 + y^2 + z^2)$. (*Isotropic* means that the force constant k' is the same in all three coordinate directions.) (a) Show that for this potential, a solution to Eq. (41.5) is given by $\psi = \psi_{n_x}(x)\psi_{n_y}(y)\psi_{n_z}(z)$. In this expression, $\psi_{n_x}(x)$ is a solution to the one-dimensional harmonic-oscillator Schrödinger equation, Eq. (40.44), with energy $E_{n_x} = \left(n_x + \frac{1}{2}\right)\hbar\omega$. The functions $\psi_{n_y}(y)$ and $\psi_{n_z}(z)$ are analogous one-dimensional wave functions for oscillations in the y- and z-directions. Find the energy associated with this ψ. (b) From your results in part (a) what are the ground-level and first-excited-level energies of the three-dimensional isotropic oscillator? (c) Show that there is only one state (one set of quantum numbers n_x, n_y, and n_z) for the ground level but three states for the first excited level.

41.45 •• CP CALC **Three-Dimensional Anisotropic Harmonic Oscillator.** An oscillator has the potential-energy function $U(x, y, z) = \frac{1}{2}k_1'(x^2 + y^2) + \frac{1}{2}k_2'z^2$, where $k_1' > k_2'$. This oscillator is called *anisotropic* because the force constant is not the same in all three coordinate directions. (a) Find a general expression for the energy levels of the oscillator (see Problem 41.44). (b) From your results in part (a), what are the ground-level and first-excited-level energies of this oscillator? (c) How many states (different sets of quantum numbers n_x, n_y, and n_z) are there for the ground level and for the first excited level? Compare to part (c) of Problem 41.44.

41.46 •• CALC A particle is described by the normalized wave function $\psi(x, y, z) = Axe^{-\alpha x^2}e^{-\beta y^2}e^{-\gamma z^2}$, where A, α, β, and γ are all real, positive constants. The probability that the particle will be found in the infinitesimal volume $dx\,dy\,dz$ centered at the point (x_0, y_0, z_0) is $|\psi(x_0, y_0, z_0)|^2\,dx\,dy\,dz$. (a) At what value of x_0 is the particle most likely to be found? (b) Are there values of x_0 for which the probability of the particle being found is zero? If so, at what x_0?

41.47 •• (a) Show that the total number of atomic states (including different spin states) in a shell of principal quantum number n is $2n^2$. [*Hint:* The sum of the first N integers $1 + 2 + 3 + \cdots + N$ is equal to $N(N + 1)/2$.] (b) Which shell has 50 states?

41.48 •• (a) What is the lowest possible energy (in electron volts) of an electron in hydrogen if its orbital angular momentum is $\sqrt{20}\,\hbar$? (b) What are the largest and smallest values of the z-component of the orbital angular momentum (in terms of $\hbar$) for the electron in part (a)? (c) What are the largest and smallest values of the spin angular momentum (in terms of $\hbar$) for the electron in part (a)? (d) What are the largest and smallest values of the orbital angular momentum (in terms of $\hbar$) for an electron in the M shell of hydrogen?

41.49 •• CALC Consider a hydrogen atom in the $1s$ state. (a) For what value of r is the potential energy $U(r)$ equal to the total energy E? Express your answer in terms of a. This value of r is called the *classical turning point,* since this is where a Newtonian particle would stop its motion and reverse direction. (b) For r greater than the classical turning point, $U(r) > E$. Classically, the particle cannot be in this region, since the kinetic energy cannot be negative. Calculate the probability of the electron being found in this classically forbidden region.

41.50 • CALC For a hydrogen atom, the probability $P(r)$ of finding the electron within a spherical shell with inner radius r and outer radius $r + dr$ is given by Eq. (41.25). For a hydrogen atom in the $1s$ ground state, at what value of r does $P(r)$ have its maximum value? How does your result compare to the distance between the electron and the nucleus for the $n = 1$ state in the Bohr model, Eq. (41.26)?

41.51 •• CALC The normalized radial wave function for the $2p$ state of the hydrogen atom is $R_{2p} = (1/\sqrt{24a^5})re^{-r/2a}$. After we average over the angular variables, the radial probability function becomes $P(r)\,dr = (R_{2p})^2r^2\,dr$. At what value of r is $P(r)$ for the $2p$ state a maximum? Compare your results to the radius of the $n = 2$ state in the Bohr model.

41.52 • CP **Rydberg Atoms.** *Rydberg atoms* are atoms whose outermost electron is in an excited state with a *very* large principal quantum number. Rydberg atoms have been produced in the laboratory and detected in interstellar space. (a) Why do all neutral Rydberg atoms with the same n value have essentially the same ionization energy, independent of the total number of electrons in the atom? (b) What is the ionization energy for a Rydberg atom with a principal quantum number of 300? By the Bohr model, what is the radius of the Rydberg electron's orbit? (c) Repeat part (b) for $n = 600$.

41.53 •• (a) For an excited state of hydrogen, show that the smallest angle that the orbital angular momentum vector $\vec{L}$ can have with the z-axis is

$$(\theta_L)_{\min} = \arccos\left(\frac{n-1}{\sqrt{n(n-1)}}\right)$$

(b) What is the corresponding expression for $(\theta_L)_{\max}$, the largest possible angle between $\vec{L}$ and the z-axis?

41.54 •• An atom in a $3d$ state emits a photon of wavelength 475.082 nm when it decays to a $2p$ state. (a) What is the energy (in electron volts) of the photon emitted in this transition? (b) Use the selection rules described in Section 41.4 to find the allowed transitions if the atom is now in an external magnetic field of 3.500 T. Ignore the effects of the electron's spin. (c) For the case in part (b), if the energy of the $3d$ state was originally -8.50000 eV with no magnetic field present, what will be the energies of the states into which it splits in the magnetic field? (d) What are the allowed wavelengths of the light emitted during transition in part (b)?

41.55 •• CALC **Spectral Analysis.** While studying the spectrum of a gas cloud in space, an astronomer magnifies a spectral line that results from a transition from a p state to an s state. She finds that the line at 575.050 nm has actually split into three lines, with adjacent lines 0.0462 nm apart, indicating that the gas is in an external magnetic field. (Ignore effects due to electron spin.) What is the strength of the external magnetic field?

41.56 •• CP **Stern–Gerlach Experiment.** In a Stern–Gerlach experiment, the deflecting force on the atom is $F_z = -\mu_z(dB_z/dz)$, where μ_z is given by Eq. (41.38) and dB_z/dz is the magnetic-field gradient. In a particular experiment, the magnetic-field region is 50.0 cm long; assume the magnetic-field gradient is constant in that region. A beam of silver atoms enters the magnetic field with a speed of 375 m/s. What value of dB_z/dz is required to give a separation of 1.0 mm between the two spin components as they exit the field? (*Note:* The magnetic dipole moment of silver is the same as that for hydrogen, since its valence electron is in an $l = 0$ state.)

41.57 • CP A large number of hydrogen atoms in $1s$ states are placed in an external magnetic field that is in the $+z$-direction. Assume that the atoms are in thermal equilibrium at room temperature, $T = 300$ K. According to the Maxwell–Boltzmann distribution (see Section 39.4), what is the ratio of the number of atoms in the $m_s = \frac{1}{2}$ state to the number in the $m_s = -\frac{1}{2}$ state when the magnetic-field magnitude is (a) 5.00×10^{-5} T (approximately the earth's field); (b) 0.500 T; (c) 5.00 T?

41.58 •• **Effective Magnetic Field.** An electron in a hydrogen atom is in the $2p$ state. In a simple model of the atom, assume that the electron circles the proton in an orbit with radius r equal to the Bohr-model radius for $n = 2$. Assume that the speed v of the orbiting electron can be calculated by setting $L = mvr$ and taking L to have the quantum-mechanical value for a $2p$ state. In the frame of the electron, the proton orbits with radius r and speed v. Model the orbiting proton as a circular current loop, and calculate the magnetic field it produces at the location of the electron.

41.59 •• Weird Universe. In another universe, the electron is a spin-$\frac{3}{2}$ rather than a spin-$\frac{1}{2}$ particle, but all other physics are the same as in our universe. In this universe, (a) what are the atomic numbers of the lightest two inert gases? (b) What is the ground-state electron configuration of sodium?

41.60 •• A lithium atom has three electrons, and the $^2S_{1/2}$ ground-state electron configuration is $1s^2 2s$. The $1s^2 2p$ excited state is split into two closely spaced levels, $^2P_{3/2}$ and $^2P_{1/2}$, by the spin-orbit interaction (see Example 41.7 in Section 41.5). A photon with wavelength 67.09608 μm is emitted in the $^2P_{3/2} \rightarrow {}^2S_{1/2}$ transition, and a photon with wavelength 67.09761 μm is emitted in the $^2P_{1/2} \rightarrow {}^2S_{1/2}$ transition. Calculate the effective magnetic field seen by the electron in the $1s^2 2p$ state of the lithium atom. How does your result compare to that for the $3p$ level of sodium found in Example 41.7?

41.61 •• A hydrogen atom in an $n = 2$, $l = 1$, $m_l = -1$ state emits a photon when it decays to an $n = 1$, $l = 0$, $m_l = 0$ ground state. (a) In the absence of an external magnetic field, what is the wavelength of this photon? (b) If the atom is in a magnetic field in the $+z$-direction and with a magnitude of 2.20 T, what is the shift in the wavelength of the photon from the zero-field value? Does the magnetic field increase or decrease the wavelength? Disregard the effect of electron spin. [*Hint:* Use the result of Problem 39.76(c).]

41.62 •• CP Electron Spin Resonance. Electrons in the lower of two spin states in a magnetic field can absorb a photon of the right frequency and move to the higher state. (a) Find the magnetic-field magnitude B required for this transition in a hydrogen atom with $n = 1$ and $l = 0$ to be induced by microwaves with wavelength λ. (b) Calculate the value of B for a wavelength of 4.20 cm.

41.63 • Estimate the minimum and maximum wavelengths of the characteristic x rays emitted by (a) vanadium $(Z = 23)$ and (b) rhenium $(Z = 45)$. Discuss any approximations that you make.

41.64 •• A hydrogen atom initially in an $n = 3$, $l = 1$ state makes a transition to the $n = 2$, $l = 0$, $j = \frac{1}{2}$ state. Find the difference in wavelength between the following two photons: one emitted in a transition that starts in the $n = 3$, $l = 1$, $j = \frac{3}{2}$ state and one that starts instead in the $n = 3$, $l = 1$, $j = \frac{1}{2}$ state. Which photon has the longer wavelength?

41.65 •• DATA In studying electron screening in multielectron atoms, you begin with the alkali metals. You look up experimental data and find the results given in the table.

Element	Li	Na	K	Rb	Cs	Fr
Ionization energy (kJ/mol)	520.2	495.8	418.8	403.0	375.7	380

The ionization energy is the minimum energy required to remove the least-bound electron from a ground-state atom. (a) The units kJ/mol given in the table are the minimum energy in kJ required to ionize 1 mol of atoms. Convert the given values for ionization energy to the energy in eV required to ionize one atom. (b) What is the value of the nuclear charge Z for each element in the table? What is the n quantum number for the least-bound electron in the ground state? (c) Calculate Z_{eff} for this electron in each alkali-metal atom. (d) The ionization energies decrease as Z increases. Does Z_{eff} increase or decrease as Z increases? Why does Z_{eff} have this behavior?

41.66 •• DATA You are studying the absorption of electromagnetic radiation by electrons in a crystal structure. The situation is well described by an electron in a cubical box of side length L. The electron is initially in the ground state. (a) You observe that the longest-wavelength photon that is absorbed has a wavelength in air of $\lambda = 624$ nm. What is L? (b) You find that $\lambda = 234$ nm is also absorbed when the initial state is still the ground state. What is the value of n^2 for the final state in the transition for which this wavelength is absorbed, where $n^2 = n_X^2 + n_Y^2 + n_Z^2$? What is the degeneracy of this energy level (including the degeneracy due to electron spin)?

41.67 •• DATA While working in a magnetics lab, you conduct an experiment in which a hydrogen atom in the $n = 1$ state is in a magnetic field of magnitude B. A photon of wavelength λ (in air) is absorbed in a transition from the $m_s = -\frac{1}{2}$ to the $m_s = +\frac{1}{2}$ state. The wavelengths λ as a function of B are given in the table.

B (T)	0.51	0.74	1.03	1.52	2.02	2.48	2.97
λ (mm)	21.4	14.3	10.7	7.14	5.35	4.28	3.57

(a) Graph the data in the table as photon frequency f versus B, where $f = c/\lambda$. Find the slope of the straight line that gives the best fit to the data. (b) Use your results of part (a) to calculate $|\mu_z|$, the magnitude of the spin magnetic moment. (c) Let $\gamma = |\mu_z|/|S_z|$ denote the gyromagnetic ratio for electron spin. Use your result of part (b) to calculate γ. What is the value of $\gamma/(e/2m)$ given by your experimental data?

CHALLENGE PROBLEMS

41.68 ••• Each of $2N$ electrons (mass m) is free to move along the x-axis. The potential-energy function for each electron is $U(x) = \frac{1}{2}k'x^2$, where k' is a positive constant. The electric and magnetic interactions between electrons can be ignored. Use the exclusion principle to show that the minimum energy of the system of $2N$ electrons is $\hbar N^2\sqrt{k'/m}$. (*Hint:* See Section 40.5 and the hint given in Problem 41.47.)

41.69 ••• CP Consider a simple model of the helium atom in which two electrons, each with mass m, move around the nucleus (charge $+2e$) in the same circular orbit. Each electron has orbital angular momentum $\hbar$ (that is, the orbit is the smallest-radius Bohr orbit), and the two electrons are always on opposite sides of the nucleus. Ignore the effects of spin. (a) Determine the radius of the orbit and the orbital speed of each electron. [*Hint:* Follow the procedure used in Section 39.3 to derive Eqs. (39.8) and (39.9). Each electron experiences an attractive force from the nucleus and a repulsive force from the other electron.] (b) What is the total kinetic energy of the electrons? (c) What is the potential energy of the system (the nucleus and the two electrons)? (d) In this model, how much energy is required to remove both electrons to infinity? How does this compare to the experimental value of 79.0 eV?

PASSAGE PROBLEMS

BIO ATOMS OF UNUSUAL SIZE. In photosynthesis in plants, light is absorbed in light-harvesting complexes that consist of protein and pigment molecules. The absorbed energy is then transported to a specialized complex called the *reaction center.* Quantum-mechanical effects may play an important role in this energy transfer. In a recent experiment, researchers cooled rubidium atoms to a very low temperature to study a similar energy-transfer process in the lab. Laser light was used to excite an electron in each atom to a state with large n. This highly excited electron behaves much like the single electron in a hydrogen atom, with an effective (screened) atomic number $Z_{\text{eff}} = 1$. Because n is so large, though, the excited electron is quite far from the atomic nucleus, with an orbital radius of approximately 1 μm, and is weakly bound. Using these so-called *Rydberg atoms,* the researchers were able to study the way energy is transported from one atom to the next. This process may be a model for understanding energy transport in photosynthesis. (Source: "Observing the Dynamics of Dipole-Mediated Energy Transport by Interaction Enhanced Imaging," by G. Günter et al., *Science* 342(6161): 954–956, Nov. 2013.)

41.70 In the Bohr model, what is the principal quantum number n at which the excited electron is at a radius of 1 μm? (a) 140; (b) 400; (c) 20; (d) 81.

41.71 Take the size of a Rydberg atom to be the diameter of the orbit of the excited electron. If the researchers want to perform this experiment with the rubidium atoms in a gas, with atoms separated by a distance 10 times their size, the density of atoms per cubic centimeter should be about (a) 10^5 atoms/cm^3; (b) 10^8 atoms/cm^3; (c) 10^{11} atoms/cm^3; (d) 10^{21} atoms/cm^3.

41.72 Assume that the researchers place an atom in a state with $n = 100$, $l = 2$. What is the magnitude of the orbital angular momentum $\vec{L}$ associated with this state? (a) $\sqrt{2}\,\hbar$; (b) $\sqrt{6}\,\hbar$; (c) $\sqrt{200}\,\hbar$; (d) $\sqrt{10{,}100}\,\hbar$.

41.73 How many different possible electron states are there in the $n = 100$, $l = 2$ subshell? (a) 2; (b) 100; (c) 10,000; (d) 10.

Answers

Chapter Opening Question ?

(iii) The Pauli exclusion principle is responsible. Helium is inert because its two electrons fill the K shell; lithium is very reactive because its third electron must go into the L shell and is loosely bound. See Section 41.6 for more details.

Test Your Understanding Questions

41.1 (iv) If $U(x, y, z) = 0$ in a certain region of space, we can rewrite the time-independent Schrödinger equation [Eq. (41.5)] for that region as $\partial^2\psi/\partial x^2 + \partial^2\psi/\partial y^2 + \partial^2\psi/\partial z^2 = (-2mE/\hbar^2)\psi$. We are told that all of the second derivatives of $\psi(x, y, z)$ are positive in this region, so the left-hand side of this equation is positive. Hence the right-hand side $(-2mE/\hbar^2)\psi$ must also be positive. Since $E > 0$, the quantity $-2mE/\hbar^2$ is negative, and so $\psi(x, y, z)$ must be negative.

41.2 (iv), (ii), (i) and (iii) (tie) Equation (41.16) shows that the energy levels for a cubical box are proportional to the quantity $n_X^2 + n_Y^2 + n_Z^2$. Hence ranking in order of this quantity is the same as ranking in order of energy. For the four cases we are given, we have (i) $n_X^2 + n_Y^2 + n_Z^2 = 2^2 + 3^2 + 2^2 = 17$; (ii) $n_X^2 + n_Y^2 + n_Z^2 = 4^2 + 1^2 + 1^2 = 18$; (iii) $n_X^2 + n_Y^2 + n_Z^2 = 2^2 + 2^2 + 3^2 = 17$; and (iv) $n_X^2 + n_Y^2 + n_Z^2 = 1^2 + 3^2 + 3^2 = 19$. The states $(n_X, n_Y, n_Z) = (2, 3, 2)$ and $(n_X, n_Y, n_Z) = (2, 2, 3)$ have the same energy (they are degenerate).

41.3 (ii) and (iii) (tie), (i) An electron in a state with principal quantum number n is most likely to be found at $r = n^2a$. This result is independent of the values of the quantum numbers l and m_l. Hence an electron with $n = 2$ (most likely to be found at $r = 4a$) is more likely to be found near $r = 5a$ than an electron with $n = 1$ (most likely to be found at $r = a$).

41.4 no All that matters is the component of the electron's orbital magnetic moment along the direction of $\vec{B}$. We called this quantity μ_z in Eq. (41.32) because we *defined* the positive z-axis to be in the direction of $\vec{B}$. In reality, the names of the axes are arbitrary.

41.5 (iv) For the magnetic moment to be perfectly aligned with the z-direction, the z-component of the spin vector $\vec{S}$ would have to have the same absolute value as $\vec{S}$. However, the possible values of S_z are $\pm\frac{1}{2}\hbar$ [Eq. (41.36)], while the magnitude of the spin vector is $S = \sqrt{\frac{3}{4}}\,\hbar$ [Eq. (41.37)]. Hence $\vec{S}$ can never be perfectly aligned with any one direction in space.

41.6 more difficult If there were no exclusion principle, all 11 electrons in the sodium atom would be in the level of lowest energy (the 1s level) and the configuration would be $1s^{11}$. Consequently, it would be more difficult to remove the first electron. (In a real sodium atom the valence electron is in a screened 3s state, which has a comparatively high energy.)

41.7 (iv) An absorption edge appears if the photon energy is just high enough to remove an electron in a given energy level from the atom. In a sample of high-temperature hydrogen we expect to find atoms whose electron is in the ground level $(n = 1)$, the first excited level $(n = 2)$, and the second excited level $(n = 3)$. From Eq. (41.21) these levels have energies $E_n = (-13.60\text{ eV})/n^2 = -13.60\text{ eV}, -3.40\text{ eV},$ and -1.51 eV (see Fig. 39.24b).

41.8 yes; no This wave function says that it is equally possible that the electron (particle 1) is in state C and the proton (particle 2) is in state E or that particle 1 is in state C and particle 2 is in state F. Since particles 1 and 2 are not identical and are distinguishable, there is no reason we can't know that particle 1 is in state C independent of which state particle 2 is in. So this is a valid wave function for the system, and the two particles are not entangled. If we measure the state of the electron alone, we are guaranteed to get C as a result; a subsequent measurement of the proton's state will give either E or F with equal probability, the same as if we had not first measured the state of the electron. Similarly, if we first measure the state of the proton, the results of that measurement will not affect a subsequent measurement of the state of the electron (for which the result is guaranteed to be C).

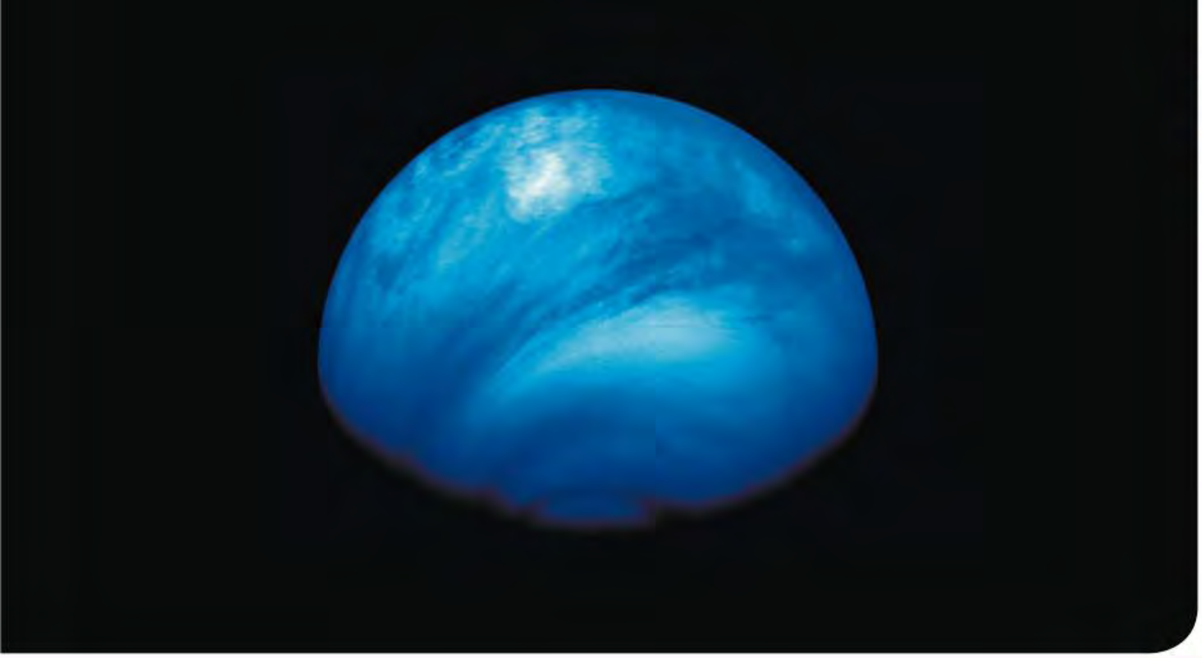
Bridging Problem

(a) 2.37×10^{-10} m

(b) Values of (n_X, n_Y, n_Z, m_s) for the 22 electrons: $(1, 1, 1, +\frac{1}{2})$, $(1, 1, 1, -\frac{1}{2})$, $(2, 1, 1, +\frac{1}{2})$, $(2, 1, 1, -\frac{1}{2})$, $(1, 2, 1, +\frac{1}{2})$, $(1, 2, 1, -\frac{1}{2})$, $(1, 1, 2, +\frac{1}{2})$, $(1, 1, 2, -\frac{1}{2})$, $(2, 2, 1, +\frac{1}{2})$, $(2, 2, 1, -\frac{1}{2})$, $(2, 1, 2, +\frac{1}{2})$, $(2, 1, 2, -\frac{1}{2})$, $(1, 2, 2, +\frac{1}{2})$, $(1, 2, 2, -\frac{1}{2})$, $(3, 1, 1, +\frac{1}{2})$, $(3, 1, 1, -\frac{1}{2})$, $(1, 3, 1, +\frac{1}{2})$, $(1, 3, 1, -\frac{1}{2})$, $(1, 1, 3, +\frac{1}{2})$, $(1, 1, 3, -\frac{1}{2})$, $(2, 2, 2, +\frac{1}{2})$, $(2, 2, 2, -\frac{1}{2})$

(c) 20.1 eV, 40.2 eV, 60.3 eV, 73.7 eV, and 80.4 eV

(d) 60.3 eV versus 4.52×10^3 eV



? Although Venus is almost twice as far as Mercury is from the sun, it has a higher surface temperature: 735 K (462°C = 863°F). The reason is that Venus has a thick, cloud-shrouded atmosphere (shown here in false color) that is 96.5% carbon dioxide (CO_2). Molecules of CO_2 are a potent agent for raising Venus's temperature because (i) they absorb infrared radiation in vibrational transitions; (ii) they absorb infrared radiation in electronic transitions; (iii) they absorb ultraviolet radiation in vibrational transitions; (iv) they absorb ultraviolet radiation in electronic transitions; (v) more than one of these.

42 MOLECULES AND CONDENSED MATTER

LEARNING GOALS

Looking forward at ...

42.1 The various types of bonds that hold atoms together.

42.2 How molecular spectra reveal the rotational and vibrational dynamics of molecules.

42.3 How and why atoms form into crystalline structures.

42.4 How to use the energy-band concept to explain the electrical properties of solids.

42.5 A model that explains many of the physical properties of metals.

42.6 How a small amount of an impurity can radically affect the character of a semiconductor.

42.7 Some of the technological applications of semiconductor devices.

42.8 Why certain materials become superconductors at low temperature.

Looking back at ...

10.5 Rotational kinetic energy.

17.7 Greenhouse effect.

18.3–18.5 Ideal gases; equipartition of energy; Maxwell–Boltzmann distribution.

23.1 Electric potential energy.

24.4, 24.5 Dielectrics.

25.2 Resistivity.

39.3 Reduced mass.

40.5 Quantum-mechanical harmonic oscillators.

41.2–41.4, 41.6 Particle in a 3-D box; rotational energy levels; selection rules; exclusion principle.

Isolated atoms, which we studied in Chapter 41, are the exception; usually we find atoms combined to form molecules or more extended structures we call condensed matter (liquid or solid). In this chapter we'll study the attractive forces, called molecular bonds, that cause atoms to combine into molecules. We will see that just as atoms have quantized energies determined by the quantum-mechanical state of their electrons, so molecules have quantized energies determined by their rotational and vibrational states.

The same physical principles behind molecular bonds also apply to the study of condensed matter, in which various types of bonding occur. We'll explore the concept of energy bands and see how it helps us understand the properties of solids. Then we'll look more closely at the properties of a special class of solids called semiconductors. Devices using semiconductors are found in every mobile phone, TV, and computer used today.

42.1 TYPES OF MOLECULAR BONDS

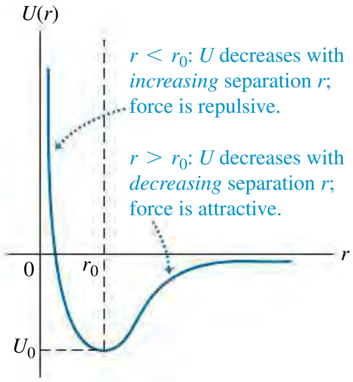
We can use our discussion of atomic structure in Chapter 41 as a basis for exploring the nature of *molecular bonds*, the interactions that hold atoms together to form stable structures such as molecules and solids.

Ionic Bonds

The **ionic bond** is an interaction between oppositely charged *ionized* atoms. The most familiar example is sodium chloride (NaCl), in which the sodium (Na) atom loses its one 3*s* electron, which fills the vacancy in the 3*p* subshell of the chlorine (Cl) atom.

Let's look at the energy balance in this transaction. Removing the 3*s* electron from a neutral Na atom requires 5.138 eV of energy; this is called the *ionization energy* of Na. The neutral Cl atom can attract an extra electron into the vacancy

42.1 When the separation r between two oppositely charged ions is large, the potential energy $U(r)$ is proportional to $1/r$ as for point charges and the force is attractive. As r decreases, the charge clouds of the two atoms overlap and the force becomes less attractive. If r is less than the equilibrium separation r_0, the force is repulsive.



in the $3p$ subshell, where it is incompletely screened by the other electrons and therefore is attracted to the nucleus. This state has 3.613 eV lower energy than a state with a neutral Cl atom and a distant free electron; 3.613 eV is the magnitude of the *electron affinity* of chlorine. Thus creating the well-separated Na^+ and Cl^- ions requires a net investment of only $5.138 \text{ eV} - 3.613 \text{ eV} = 1.525 \text{ eV}$.

When their mutual attraction brings the Na^+ and Cl^- ions together, the magnitude of their negative potential energy is determined by their separation r (**Fig. 42.1**). The exclusion principle (Section 41.6), which states that only one electron can occupy a given quantum-mechanical state, limits how small this separation can be. As r decreases, the exclusion principle distorts the charge clouds, so the ions no longer interact like point charges and the interaction eventually becomes repulsive.

The minimum electric potential energy for NaCl turns out to be -5.7 eV at a separation of 0.24 nm. The net energy released in creating the ions and letting them come together to the equilibrium separation of 0.24 nm is $5.7 \text{ eV} - 1.525 \text{ eV} = 4.2 \text{ eV}$. Thus, if the kinetic energy of the ions is ignored, 4.2 eV is the *binding energy* of the NaCl molecule, the energy that is needed to dissociate the molecule into separate neutral atoms.

Ionic bonds can involve more than one electron per atom. For instance, alkaline earth elements form ionic compounds in which an atom loses *two* electrons; an example is magnesium chloride, or $Mg^{2+}(Cl^-)_2$. Ionic bonds that involve a loss of more than two electrons are relatively rare. Instead, a different kind of bond, the *covalent* bond, comes into operation. We'll discuss this type of bond below.

EXAMPLE 42.1 ELECTRIC POTENTIAL ENERGY OF THE NaCl MOLECULE

SOLUTION

Find the electric potential energy of an Na^+ ion and a Cl^- ion separated by 0.24 nm. Consider the ions as point charges.

SOLUTION

IDENTIFY and SET UP: Equation (23.9) in Section 23.1 tells us that the electric potential energy of two point charges q and q_0 separated by a distance r is $U = qq_0/4\pi\epsilon_0 r$.

EXECUTE: We have $q = +e$ (for Na^+), $q_0 = -e$ (for Cl^-), and $r = 0.24 \text{ nm} = 0.24 \times 10^{-9}$ m. From Eq. (23.9),

$$U = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_0} = -(9.0 \times 10^9 \text{ N}\cdot\text{m}^2/\text{C}^2)\frac{(1.6 \times 10^{-19}\text{ C})^2}{0.24 \times 10^{-9}\text{ m}}$$
$$= -9.6 \times 10^{-19}\text{J} = -6.0 \text{ eV}$$

EVALUATE: This result agrees fairly well with the observed value of -5.7 eV. The reason for the difference is that when the two ions are at their equilibrium separation of 0.24 nm, the outer regions of their electron clouds overlap. Hence the two ions don't behave exactly like point charges.

Covalent Bonds

MP

PhET: Double Wells and Covalent Bonds

Unlike the transaction that occurs in an ionic bond, in a **covalent bond** there is no net transfer of electrons from one atom to another. The simplest covalent bond is found in the hydrogen molecule, a structure containing two protons and two electrons. As the separate atoms (**Fig. 42.2a**) come together, the electron wave functions are distorted and become more concentrated in the region between the two protons (Fig. 42.2b). The net attraction of the electrons for each proton more than balances the repulsion of the two protons and of the two electrons.

The attractive interaction is then supplied by a *pair* of electrons, one contributed by each atom, with charge clouds that are concentrated primarily in the region between the two atoms. The energy of the covalent bond in the hydrogen molecule H_2 is -4.48 eV.

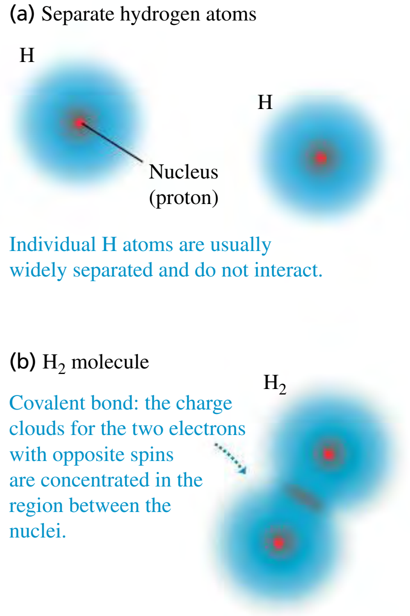
As we saw in Section 41.6, the exclusion principle permits two electrons to occupy the same region of space (that is, to be in the same spatial quantum state) only when they have opposite spins. Hence the two electrons in the H_2 covalent bond (Fig. 42.2b) must have opposite spins, since both occupy the same region between the two nuclei. Opposite spins are an essential requirement for a covalent bond, and no more than two electrons can participate in such a bond.

However, an atom with several electrons in its outermost shell can form several covalent bonds. The bonding of carbon and hydrogen atoms, of central importance in organic chemistry, is an example. In the *methane* molecule (CH_4) the carbon atom is at the center of a regular tetrahedron, with a hydrogen atom at each corner. The carbon atom has four electrons in its *L* shell, and each of these four electrons forms a covalent bond with one of the four hydrogen atoms (**Fig. 42.3**). Similar patterns occur in more complex organic molecules.

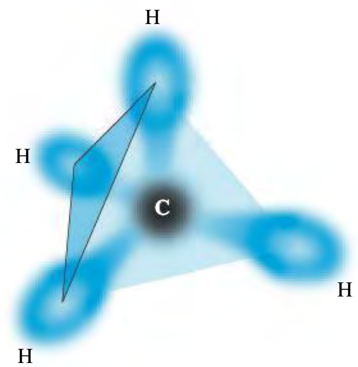
Covalent bonds are highly directional. In the methane molecule the wave function for each of carbon's four valence electrons is a combination of the $2s$ and $2p$ wave functions called a *hybrid wave function*. The probability distribution for each one has a lobe protruding toward a corner of a tetrahedron. This symmetric arrangement minimizes the overlap of wave functions for the electron pairs, which in turn minimizes the positive potential energy associated with repulsion between the pairs.

Ionic and covalent bonds represent two extremes in molecular bonding, but there is no sharp division between the two types. Often there is a *partial* transfer of one or more electrons from one atom to another. As a result, many molecules that have dissimilar atoms have electric dipole moments—that is, a preponderance of positive charge at one end and of negative charge at the other. Such molecules are called *polar* molecules. Water molecules have large electric dipole moments; these are responsible for the exceptionally large dielectric constant of liquid water (see Sections 24.4 and 24.5).

42.2 Covalent bond in a hydrogen molecule.



42.3 Schematic diagram of the methane (CH_4) molecule. The carbon atom is at the center of a regular tetrahedron and forms four covalent bonds with the hydrogen atoms at the corners. Each covalent bond includes two electrons with opposite spins, forming a charge cloud that is concentrated between the carbon atom and a hydrogen atom.



Van der Waals Bonds

Ionic and covalent bonds, with typical bond energies of 1 to 5 eV, are called *strong bonds*. There are also two types of weaker bonds. One of these, the **van der Waals bond**, is an interaction between the electric dipole moments of atoms or molecules; typical energies are 0.1 eV or less. The bonding of water molecules in the liquid and solid states results partly from dipole–dipole interactions.

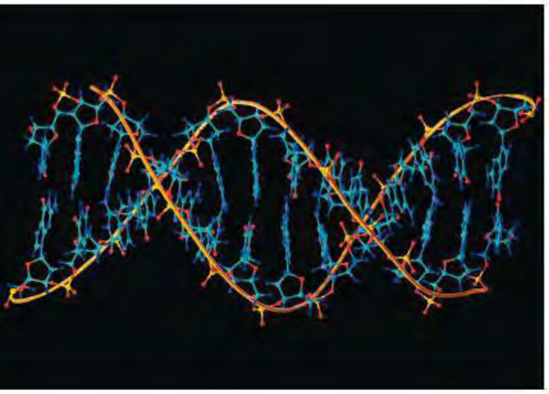
No atom has a permanent electric dipole moment, nor do many molecules. However, fluctuating charge distributions can lead to fluctuating dipole moments; these in turn can induce dipole moments in neighboring structures. Overall, the resulting dipole–dipole interaction is attractive, giving a weak bonding of atoms or molecules. The interaction potential energy drops off very quickly with distance r between molecules, usually as $1/r^6$. The liquefaction and solidification of the inert gases and of molecules such as H_2, O_2, and N_2 are due to induced-dipole van der Waals interactions. Not much thermal-agitation energy is needed to break these weak bonds, so such substances usually exist in the liquid and solid states only at very low temperatures.

Hydrogen Bonds

In the other type of weak bond, the **hydrogen bond**, a proton (H^+ ion) gets between two atoms, polarizing them and attracting them by means of the induced dipoles. This bond is unique to hydrogen-containing compounds because only hydrogen has a singly ionized state with no remaining electron cloud; the hydrogen ion is a bare proton, much smaller than any other singly ionized atom. The bond energy is usually less than 0.5 eV. The hydrogen bond is responsible for the cross-linking of long-chain organic molecules such as polyethylene (used in plastic bags). Hydrogen bonding also plays a role in the structure of ice.

All these bond types hold the atoms together in *solids* as well as in molecules. Indeed, a solid is in many respects a giant molecule. Still another type of

BIO Application Molecular Zipper
A DNA molecule functions like a twisted zipper. Each of the two strands of the "zipper" consists of an outer backbone and inward-facing nucleotide "teeth"; hydrogen bonds between facing teeth "zip" the strands together. The covalent bonds that hold together the atoms of each strand are strong, whereas the hydrogen bonds are relatively weak, so that the cell's biochemical machinery can easily separate the strands for reading or copying.



bonding, the *metallic bond*, comes into play in the structure of metallic solids. We'll return to this subject in Section 42.3.

TEST YOUR UNDERSTANDING OF SECTION 42.1 If electrons obeyed the exclusion principle but did *not* have spin, how many electrons could participate in a covalent bond? (i) One; (ii) two; (iii) three; (iv) more than three.

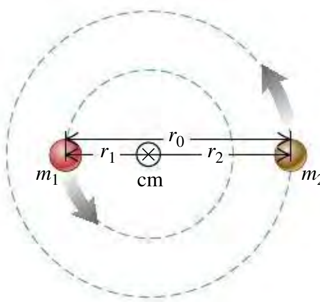
42.2 MOLECULAR SPECTRA

Molecules have energy levels that are associated with rotation of a molecule as a whole and with vibration of the atoms relative to each other. Just as transitions between energy levels in atoms lead to atomic spectra, transitions between rotational and vibrational levels in molecules lead to *molecular spectra*.

Rotational Energy Levels

In this discussion we'll concentrate mostly on *diatomic* molecules, to keep things as simple as possible. In **Fig. 42.4** we picture a diatomic molecule as a rigid dumbbell (two point masses m_1 and m_2 separated by a constant distance r_0) that can *rotate* about axes through its center of mass, perpendicular to the line joining them. What are the energy levels associated with this motion?

42.4 A diatomic molecule modeled as two point masses m_1 and m_2 separated by a distance r_0. The distances of the masses from the center of mass are r_1 and r_2, where $r_1 + r_2 = r_0$.



We showed in Section 10.5 that when a rigid body rotates with angular speed ω about a perpendicular axis through its center of mass, the magnitude L of its angular momentum is given by Eq. (10.28), $L = I\omega$, where I is its moment of inertia for that axis. Its kinetic energy is given by Eq. (9.17), $K = \frac{1}{2}I\omega^2$. Combining these two equations, we find $K = L^2/2I$. There is no potential energy U, so the kinetic energy K is equal to the total mechanical energy E:

$$E = \frac{L^2}{2I} \tag{42.1}$$

Zero potential energy means that U does not depend on the angular coordinate of the molecule. But the potential-energy function U for the hydrogen atom (see Section 41.3) also has no dependence on angular coordinates. Thus the angular solutions to the Schrödinger equation for rigid-body rotation are the same as for the hydrogen atom, and the angular momentum is quantized in the same way. As in Eq. (41.22),

$$L = \sqrt{l(l+1)}\hbar \qquad (l = 0, 1, 2, \ldots) \tag{42.2}$$

Combining Eqs. (42.1) and (42.2), we obtain the *rotational energy levels:*

Rotational energy levels of a diatomic molecule: Rotational quantum number ($l = 0, 1, 2, \ldots$); Planck's constant divided by 2π; Moment of inertia for axis through molecule's cm

$$E_l = l(l+1)\frac{\hbar^2}{2I} \tag{42.3}$$

42.5 The ground level and first four excited rotational energy levels for a diatomic molecule. The levels are not equally spaced.

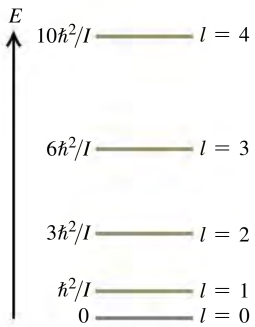


Figure 42.5 is an energy-level diagram showing these rotational levels. The $l = 0$ ground level has zero angular momentum (no rotation and zero rotational energy E). The spacing of adjacent levels increases with increasing l.

We can express the moment of inertia I in Eqs. (42.1) and (42.3) in terms of the *reduced mass* m_r of the molecule:

Reduced mass of a diatomic molecule: Mass of atom 1; Mass of atom 2

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{42.4}$$

We introduced the reduced mass in Section 39.3 to accommodate the finite nuclear mass of the hydrogen atom. In Fig. 42.4 the distances r_1 and r_2 are the distances from the center of mass to the centers of the atoms. By the definition of the center of mass, $m_1 r_1 = m_2 r_2$, and the figure also shows that $r_0 = r_1 + r_2$. Solving these equations for r_1 and r_2, we find

$$r_1 = \frac{m_2}{m_1 + m_2} r_0 \qquad r_2 = \frac{m_1}{m_1 + m_2} r_0 \tag{42.5}$$

The moment of inertia is $I = m_1 r_1^2 + m_2 r_2^2$; substituting Eq. (42.5), we find

$$I = m_1 \frac{m_2^2}{(m_1 + m_2)^2} r_0^2 + m_2 \frac{m_1^2}{(m_1 + m_2)^2} r_0^2 = \frac{m_1 m_2}{m_1 + m_2} r_0^2 \quad \text{or}$$

Moment of inertia of a diatomic molecule, axis through molecule's cm $\quad I = \underbrace{m_r}_{\text{Reduced mass}} \underbrace{r_0^2}_{\text{Distance between centers of molecule's two atoms}} \quad (42.6)$

The moment of inertia is the same as that of an equivalent *single* point mass m_r that orbits the axis in a circle of radius r_0.

To conserve angular momentum and account for the angular momentum of the emitted or absorbed photon, the allowed transitions between rotational states must satisfy the same selection rule that we discussed in Section 41.4 for allowed transitions between the states of an atom: l must change by exactly one unit; that is, $\Delta l = \pm 1$.

MP

PhET: The Greenhouse Effect

EXAMPLE 42.2 ROTATIONAL SPECTRUM OF CARBON MONOXIDE

SOLUTION

The two nuclei in the carbon monoxide (CO) molecule are 0.1128 nm apart. The mass of the most common carbon atom is 1.993×10^{-26} kg; that of the most common oxygen atom is 2.656×10^{-26} kg. (a) Find the energies of the lowest three rotational energy levels of CO. Express your results in meV ($1 \text{ meV} = 10^{-3} \text{ eV}$). (b) Find the wavelength of the photon emitted in the transition from the $l = 2$ to the $l = 1$ level.

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas developed in this section about the rotational energy levels of molecules. We are given the distance r_0 between the atoms and their masses m_1 and m_2. We find the reduced mass m_r from Eq. (42.4), the moment of inertia I from Eq. (42.6), and the energies E_l from Eq. (42.3). The energy E of the emitted photon is equal to the difference in energy between the $l = 2$ and $l = 1$ levels. (This transition obeys the $\Delta l = \pm 1$ selection rule, since $\Delta l = 1 - 2 = -1$.) We determine the photon wavelength by using $E = hc/\lambda$.

EXECUTE: (a) From Eqs. (42.4) and (42.6), the reduced mass and moment of inertia of the CO molecule are:

$$\begin{aligned} m_r &= \frac{m_1 m_2}{m_1 + m_2} \\ &= \frac{(1.993 \times 10^{-26}\text{ kg})(2.656 \times 10^{-26}\text{ kg})}{(1.993 \times 10^{-26}\text{ kg}) + (2.656 \times 10^{-26}\text{ kg})} \\ &= 1.139 \times 10^{-26}\text{ kg} \\ I &= m_r r_0^2 \\ &= (1.139 \times 10^{-26}\text{ kg})(0.1128 \times 10^{-9}\text{ m})^2 \\ &= 1.449 \times 10^{-46}\text{ kg} \cdot \text{m}^2 \end{aligned}$$

The rotational levels are given by Eq. (42.3):

$$\begin{aligned} E_l &= l(l + 1)\frac{\hbar^2}{2I} = l(l + 1)\frac{(1.0546 \times 10^{-34}\text{ J}\cdot\text{s})^2}{2(1.449 \times 10^{-46}\text{ kg}\cdot\text{m}^2)} \\ &= l(l + 1)(3.838 \times 10^{-23}\text{ J}) = l(l + 1)0.2395\text{ meV} \end{aligned}$$

($1 \text{ meV} = 10^{-3} \text{ eV}$.) Substituting $l = 0, 1, 2$, we find

$$E_0 = 0 \qquad E_1 = 0.479\text{ meV} \qquad E_2 = 1.437\text{ meV}$$

(b) The photon energy and wavelength are

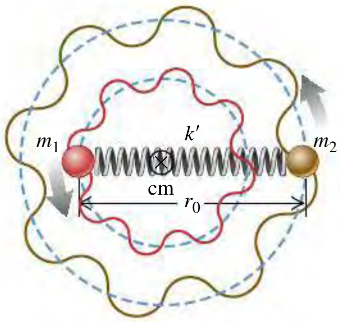
$$E = E_2 - E_1 = 0.958\text{ meV}$$

$$\begin{aligned} \lambda &= \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\text{ eV}\cdot\text{s})(3.00 \times 10^8\text{ m/s})}{0.958 \times 10^{-3}\text{ eV}} \\ &= 1.29 \times 10^{-3}\text{ m} = 1.29\text{ mm} \end{aligned}$$

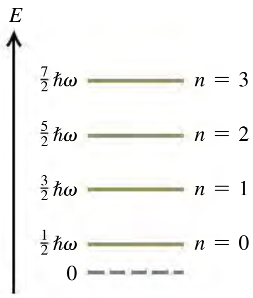
EVALUATE: The differences between the first few rotational energy levels of CO are very small (about $1 \text{ meV} = 10^{-3} \text{ eV}$) compared to the differences between atomic energy levels (typically a few eV). Hence a photon emitted by a CO molecule in a transition from the $l = 2$ to the $l = 1$ level has very low energy and a very long wavelength compared to the visible light emitted by excited atoms. Photon wavelengths for rotational transitions in molecules are typically in the microwave and far infrared regions of the spectrum.

In this example we were given the equilibrium separation between the atoms, also called the *bond length*, and we used it to calculate one of the wavelengths emitted by excited CO molecules. In experiments, scientists work this problem backward: By measuring the long-wavelength emissions of a sample of diatomic molecules, they determine the moment of inertia of the molecule and hence the bond length.

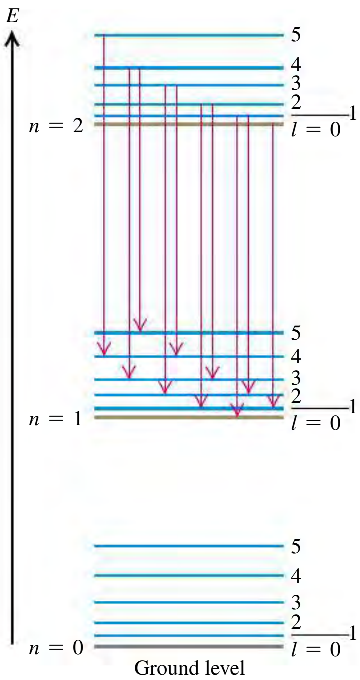
42.6 A diatomic molecule modeled as two point masses m_1 and m_2 connected by a spring with force constant k'.



42.7 The ground level and first three excited vibrational levels for a diatomic molecule, assuming small displacements from equilibrium so we can treat the oscillations as simple harmonic. (Compare Fig. 40.25.)



42.8 Energy-level diagram for vibrational and rotational energy levels of a diatomic molecule. For each vibrational level (n) there is a series of more closely spaced rotational levels (l). Several transitions corresponding to a single band in a band spectrum are shown. These transitions obey the selection rule $\Delta l = \pm 1$.



Vibrational Energy Levels

Molecules are never completely rigid. In a more realistic model of a diatomic molecule we represent the connection between atoms not as a rigid rod but as a spring (**Fig. 42.6**). Then, in addition to rotating, the atoms of the molecule can *vibrate* about their equilibrium positions along the line joining them. For small oscillations the restoring force can be taken as proportional to the displacement from the equilibrium separation r_0 (like a spring that obeys Hooke's law with a force constant k'), and the system is a harmonic oscillator. We discussed the quantum-mechanical harmonic oscillator in Section 40.5. The energy levels are given by Eq. (40.46) with the mass m replaced by the reduced mass m_r:

Vibrational energy levels of a diatomic molecule

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\omega = \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m_r}} \tag{42.7}$$

Vibrational quantum number ($n = 0, 1, 2, \ldots$)
Planck's constant divided by 2π
Oscillation angular frequency
Reduced mass
Force constant

The spacing in energy between any two adjacent vibrational levels is

$$\Delta E = \hbar\omega = \hbar\sqrt{\frac{k'}{m_r}} \tag{42.8}$$

Figure 42.7 is an energy-level diagram showing these vibrational levels. As an example, for the CO molecule of Example 42.2 the spacing $\hbar\omega$ between levels is 0.2690 eV. From Eq. (42.8) this corresponds to a force constant of 1.90×10^3 N/m, which is a fairly loose spring. (To stretch a macroscopic spring with this value of k' by 1.0 cm would require a force of only 19 N, or about 4 lb.) Force constants for diatomic molecules are typically about 100 to 2000 N/m.

CAUTION **Watch out for k, k', and K** As in Section 40.5 we're using k' for the force constant, this time to minimize confusion with Boltzmann's constant k, the gas constant per molecule (introduced in Section 18.3). Besides the quantities k and k', we also use the absolute temperature unit $1\text{ K} = 1$ kelvin.

Rotation and Vibration Combined

Visible-light photons have energies between 1.65 eV and 3.26 eV. The 0.2690-eV energy difference between vibrational levels for carbon monoxide (CO) corresponds to a photon of wavelength 4.613 μm, in the infrared region of the spectrum. This is much closer to the visible region than is the photon in the rotational transition in Example 42.2. In general the energy differences for molecular *vibration* are much smaller than those that produce atomic spectra, but much larger than the energy differences for molecular *rotation*.

When we include *both* rotational and vibrational energies, the energy levels for our diatomic molecule are

$$E_{nl} = l(l+1)\frac{\hbar^2}{2I} + \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m_r}} \tag{42.9}$$

Figure 42.8 shows the energy-level diagram. For each value of n there are many values of l, forming a series of closely spaced levels.

The red arrows in Fig. 42.8 show several possible transitions in which a molecule goes from a level with $n = 2$ to a level with $n = 1$ by emitting a photon. As we mentioned, these transitions must obey the selection rule $\Delta l = \pm 1$ to conserve angular momentum. Another selection rule states that if the vibrational level changes, the vibrational quantum number n in Eq. (42.9) must increase by 1 $(\Delta n = 1)$ if a photon is absorbed or decrease by 1 $(\Delta n = -1)$ if a photon is emitted.

42.9 A typical molecular band spectrum.

Illustrating these selection rules, Fig. 42.8 shows that a molecule in the $n = 2$, $l = 4$ level can emit a photon and drop into the $n = 1$, $l = 5$ level $(\Delta n = -1, \Delta l = +1)$ or the $n = 1$, $l = 3$ level $(\Delta n = -1, \Delta l = -1)$, but is forbidden from making a $\Delta n = -1$, $\Delta l = 0$ transition into the $n = 1$, $l = 4$ level.

Transitions between states with various pairs of n-values give different series of spectrum lines, and the resulting spectrum has a series of *bands*. Each band corresponds to a particular vibrational transition, and each individual line in a band represents a particular rotational transition, with the selection rule $\Delta l = \pm 1$. **Figure 42.9** shows a typical *band spectrum*.

All molecules can have excited states of the *electrons* in addition to the rotational and vibrational states that we have described. In general, these lie at higher energies than the rotational and vibrational states, and there is no simple rule relating them. When there is a transition between electronic states, the $\Delta n = \pm 1$ selection rule for the vibrational levels no longer holds.

SOLUTION

EXAMPLE 42.3 VIBRATION-ROTATION SPECTRUM OF CARBON MONOXIDE

Consider again the CO molecule of Example 42.2. Find the wavelength of the photon emitted by a CO molecule when its vibrational energy changes and its rotational energy is (a) initially zero and (b) finally zero.

SOLUTION

IDENTIFY and SET UP: We need to use the selection rules for the vibrational and rotational transitions of a diatomic molecule. Since a photon is emitted as the vibrational energy changes, the selection rule $\Delta n = -1$ tells us that the vibrational quantum number n decreases by 1 in both parts (a) and (b). In part (a) the initial value of l is zero; the selection rule $\Delta l = \pm 1$ tells us that the *final* value of l is 1, so the rotational energy increases in this case. In part (b) the *final* value of l is zero; $\Delta l = \pm 1$ then tells us that the *initial* value of l is 1, and the rotational energy decreases.

The energy E of the emitted photon is the difference between the initial and final energies of the molecule, accounting for the change in both vibrational and rotational energies. In part (a) E equals the difference $\hbar\omega$ between adjacent vibrational energy levels *minus* the rotational energy that the molecule *gains*; in part (b) E equals $\hbar\omega$ *plus* the rotational energy that the molecule *loses*. Example 42.2 tells us that the difference between the $l = 0$ and $l = 1$ rotational energy levels is $0.479 \text{ meV} = 0.000479 \text{ eV}$, and we learned above that the vibrational energy-level separation for CO is $\hbar\omega = 0.2690 \text{ eV}$. We use $E = hc/\lambda$ to determine the corresponding wavelengths (our target variables).

EXECUTE: (a) The CO molecule loses $\hbar\omega = 0.2690 \text{ eV}$ of vibrational energy and gains 0.000479 eV of rotational energy. Hence the energy E that goes into the emitted photon equals 0.2690 eV *less* 0.000479 eV, or 0.2685 eV. The photon wavelength is

$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15} \text{ eV}\cdot\text{s})(2.998 \times 10^{8} \text{ m/s})}{0.2685 \text{ eV}}$$
$$= 4.618 \times 10^{-6} \text{ m} = 4.618 \ \mu\text{m}$$

(b) Now the CO molecule loses $\hbar\omega = 0.2690 \text{ eV}$ of vibrational energy and also loses 0.000479 eV of rotational energy, so the energy that goes into the photon is $E = 0.2690 \text{ eV} + 0.000479 \text{ eV} = 0.2695 \text{ eV}$. The wavelength is

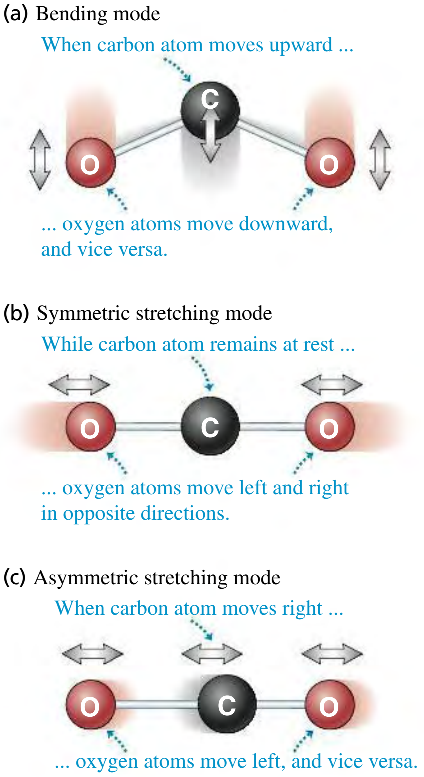
$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15} \text{ eV}\cdot\text{s})(2.998 \times 10^{8} \text{ m/s})}{0.2695 \text{ eV}}$$
$$= 4.601 \times 10^{-6} \text{ m} = 4.601 \ \mu\text{m}$$

EVALUATE: In part (b) the molecule loses more energy than it does in part (a), so the emitted photon must have greater energy and a shorter wavelength. That's just what our results show.

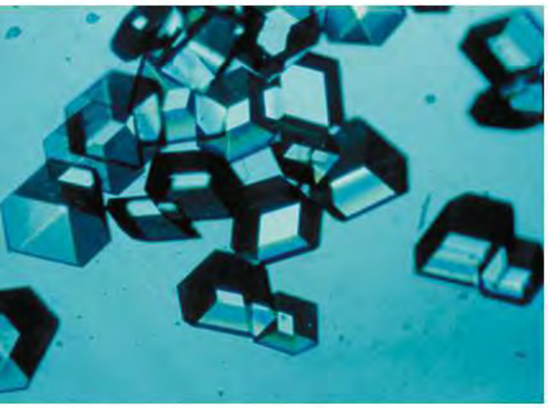
Complex Molecules

We can apply these same principles to more complex molecules. A molecule with three or more atoms has several different kinds or *modes* of vibratory motion. Each mode has its own set of energy levels, related to its frequency by Eq. (42.7). In nearly all cases the associated radiation lies in the infrared region of the electromagnetic spectrum.

42.10 The carbon dioxide molecule can vibrate in three different modes. For clarity, the atoms are not shown to scale: The separation between atoms is actually comparable to their diameters.



BIO Application Using Crystals to Determine Protein Structure Protein molecules can form crystals, such as these crystals of insulin (a protein composed of 51 amino acids). All of the molecules within a single crystal of a protein have the same orientation; how the crystal diffracts x rays or neutrons depends on the shape and size of the molecules. By analyzing these diffraction patterns, scientists have deduced the molecular structures of more than 100,000 types of proteins.



Infrared spectroscopy has proved to be an extremely valuable analytical tool. It provides information about the strength, rigidity, and length of molecular bonds and the structure of complex molecules. Also, because every molecule (like every atom) has its characteristic spectrum, infrared spectroscopy can be used to identify unknown compounds.

One molecule that can readily absorb and emit infrared radiation is carbon dioxide (CO_2). **Figure 42.10** shows the three possible modes of vibration of a CO_2 molecule. A number of transitions are possible between excited levels of the same vibrational mode as well as between levels of different vibrational modes. The energy differences are less than 1 eV in all of these transitions, and so involve infrared photons of wavelength longer than 1 μm. Hence a gas of CO_2 can readily absorb light at a number of different infrared wavelengths. This makes CO_2 primarily responsible for the greenhouse effect (Section 17.7) on the earth, even though CO_2 is only 0.04% of our atmosphere by volume. On Venus, however, the atmosphere has more than 90 times the total mass of our atmosphere and is almost entirely CO_2. The resulting greenhouse effect is tremendous: The surface temperature on Venus is more than 400 kelvins higher than what it would be if the planet had no atmosphere at all.

TEST YOUR UNDERSTANDING OF SECTION 42.2 A rotating diatomic molecule emits a photon when it makes a transition from level (n, l) to level $(n - 1, l - 1)$. If the value of l increases but n is unchanged, does the wavelength of the emitted photon (i) increase, (ii) decrease, or (iii) remain unchanged?

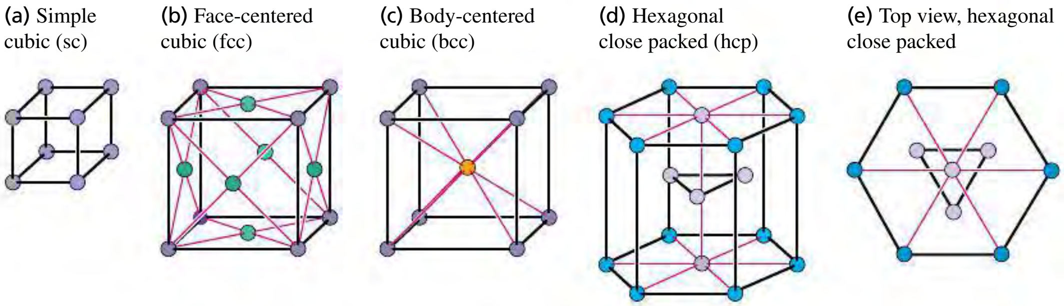
42.3 STRUCTURE OF SOLIDS

The term *condensed matter* includes both solids and liquids. In both states, the interactions between atoms or molecules are strong enough to give the material a definite volume that changes relatively little with applied stress. In condensed matter, adjacent atoms attract one another until their outer electron charge clouds begin to overlap significantly. Thus the distances between adjacent atoms in condensed matter are about the same as the diameters of the atoms themselves, typically 0.1 to 0.5 nm. Also, when we speak of the distances between atoms, we mean the center-to-center (nucleus-to-nucleus) distances.

Ordinarily, we think of a liquid as a material that can flow and of a solid as a material with a definite shape. However, if you heat a horizontal glass rod in the flame of a burner, you'll find that the rod begins to sag (flow) more and more easily as its temperature rises. Glass has no definite transition from solid to liquid, and no definite melting point. On this basis, we can consider glass at room temperature as being an extremely viscous liquid. Tar and butter show similar behavior.

What is the microscopic difference between materials like glass or butter and solids like ice or copper, which do have definite melting points? Ice and copper are examples of *crystalline solids* in which the atoms have *long-range order*, a recurring pattern of atomic positions that extends over many atoms. This pattern is called the *crystal structure*. In contrast, glass at room temperature is an example of an *amorphous* solid, one that has no long-range order, but only *short-range order* (correlations between neighboring atoms or molecules). Liquids also have only short-range order. The boundaries between crystalline solid, amorphous solid, and liquid may be sometimes blurred. Some solids, crystalline when perfect, can form with so many imperfections in their structure that they have almost no long-range order. (Yet another kind of order is found in a *liquid crystal*, which is made up of rod-shaped or disc-shaped molecules of an organic compound. The positions of the molecules in the liquid are not fixed, but there is *orientational* order; the axes of the molecules tend to align with each other. This ordering can extend over a distance of many molecules.)

42.11 Portions of some common types of crystal lattices.



Nearly everything we know about crystal structure was learned from diffraction experiments with x rays, electrons, or neutrons. A typical distance between atoms is of the order of 0.1 nm. You can show that 12.4-keV x rays, 150-eV electrons, and 0.0818-eV neutrons all have wavelengths $\lambda = 0.1$ nm.

Crystal Lattices and Structures

An essential part of understanding crystals is the idea of a *crystal lattice*, which is a repeating pattern of mathematical points that extends throughout space. There are 14 general types of such patterns; **Fig. 42.11** shows small portions of some common examples. The *simple cubic lattice* (sc) has a lattice point at each corner of a cubic array (Fig. 42.11a). The *face-centered cubic lattice* (fcc) is like the simple cubic but with an additional lattice point at the center of each cube face (Fig. 42.11b). The *body-centered cubic lattice* (bcc) is like the simple cubic but with an additional point at the center of each cube (Fig. 42.11c). The *hexagonal close-packed lattice* has layers of lattice points in hexagonal patterns, each hexagon made up of six equilateral triangles (Figs. 42.11d and 42.11e).

CAUTION **A perfect crystal lattice is infinitely large** Figure 42.11 shows just enough lattice points so that you can easily visualize the pattern; the lattice, a mathematical abstraction, extends throughout space. Thus the lattice points shown repeat endlessly in all directions.

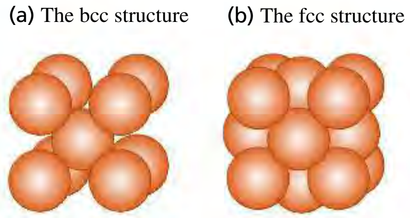
In a crystal structure, a single atom or a group of atoms is associated with each lattice point. The group may contain the same or different kinds of atoms. This atom or group of atoms is called a *basis*. Thus a complete description of a crystal structure includes both the lattice and the basis. We initially consider *perfect crystals*, or *ideal single crystals*, in which the crystal structure extends uninterrupted throughout space.

The bcc and fcc structures are two common simple crystal structures. The alkali metals have a bcc structure—that is, a bcc lattice with a basis of one atom at each lattice point. Each atom in a bcc structure has eight nearest neighbors (**Fig. 42.12a**). The elements Al, Ca, Cu, Ag, and Au have an fcc structure—that is, an fcc lattice with a basis of one atom at each lattice point. Each atom in an fcc structure has 12 nearest neighbors (Fig. 42.12b).

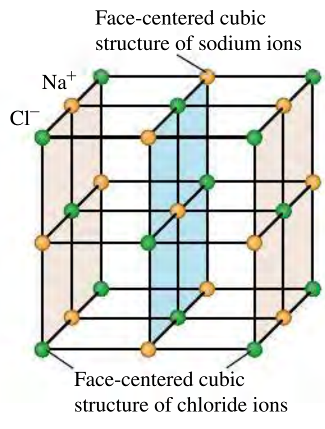
Figure 42.13 shows a representation of the structure of sodium chloride (NaCl, ordinary salt). It may look like a simple cubic structure, but it isn't. The sodium and chloride ions each form an fcc structure, so we can think roughly of the sodium chloride structure as being composed of two interpenetrating fcc structures. More correctly, the sodium chloride crystal structure of Fig. 42.13 has an fcc lattice with one chloride ion at each lattice point and one sodium ion half a cube length above it. That is, its basis consists of one chloride and one sodium ion.

Another example is the *diamond structure*; it's called that because it is the crystal structure of carbon in the diamond form. It's also the crystal structure of

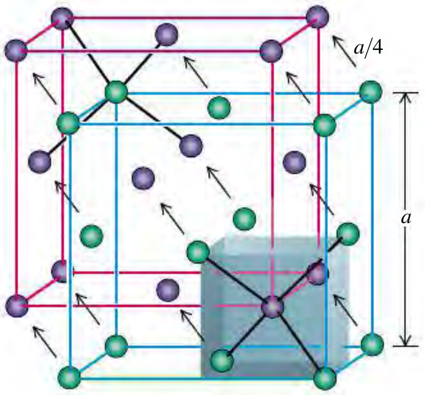
42.12 (a) The bcc *structure* is composed of a bcc *lattice* with a basis of one atom for each lattice point. (b) The fcc *structure* is composed of an fcc *lattice* with a basis of one atom for each lattice point. These structures repeat precisely to make up perfect crystals.



42.13 Representation of part of the sodium chloride crystal structure. The distances between ions are exaggerated.



42.14 The diamond structure, shown as two interpenetrating face-centered cubic structures with distances between atoms exaggerated. Relative to the corresponding green atom, each purple atom is shifted up, back, and to the left by a distance $a/4$.



silicon, germanium, and gray tin. The diamond lattice is fcc; the basis consists of one atom at each lattice point and a second *identical* atom displaced a quarter of a cube length in each of the three cube-edge directions. **Figure 42.14** will help you visualize this. The shaded volume in Fig. 42.14 shows the bottom right front eighth of the basic cube; the four atoms at alternate corners of this cube are at the corners of a regular tetrahedron, and there is an additional atom at the center. Thus each atom in the diamond structure is at the center of a regular tetrahedron with four nearest-neighbor atoms at the corners.

In the diamond structure, both the purple and green spheres in Fig. 42.14 represent *identical* atoms—for example, both carbon or both silicon. In the cubic zinc sulfide structure, the purple spheres represent one type of atom and the green spheres represent a *different* type. For example, in zinc sulfide (ZnS) each zinc atom (purple in Fig. 42.14) is at the center of a regular tetrahedron with four sulfur atoms (green in Fig. 42.14) at its corners, and vice versa. Gallium arsenide (GaAs) and similar compounds have this same structure.

Bonding in Solids

The forces that are responsible for the regular arrangement of atoms in a crystal are the same as those involved in molecular bonds, plus one additional type. Not surprisingly, *ionic* and *covalent* molecular bonds are found in ionic and covalent crystals, respectively. The most familiar *ionic crystals* are the alkali halides, such as ordinary salt (NaCl). The positive sodium ions and the negative chloride ions occupy adjacent positions in the crystal (see Fig. 42.13). The attractive forces are the familiar Coulomb's-law forces between charged particles. These forces have no preferred direction, and the arrangement in which the material crystallizes is partly determined by the relative sizes of the two ions. Such a structure is *stable* in the sense that it has lower total energy than the separated ions (see the following example). The negative potential energies of pairs of opposite charges are greater in absolute value than the positive energies of pairs of like charges because the pairs of unlike charges are closer together, on average.

EXAMPLE 42.4 POTENTIAL ENERGY OF AN IONIC CRYSTAL

SOLUTION

Imagine a one-dimensional ionic crystal consisting of a very large number of alternating positive and negative ions with charges e and $-e$, with equal spacing a along a line. Show that the total interaction potential energy is negative, which means that such a "crystal" is stable.

SOLUTION

IDENTIFY and SET UP: We treat each ion as a point charge and use our results from Section 23.1 for the electric potential energy of a collection of point charges. Equations (23.10) and (23.11) tell us to consider the electric potential energy U of each pair of charges. The total potential energy of the system is the sum of the values of U for every possible pair; we take the number of pairs to be infinite.

EXECUTE: Let's pick an ion somewhere in the middle of the line and add the potential energies of its interactions with all the ions to one side of it. From Eq. (23.11), that sum is

$$\sum U = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{a} + \frac{e^2}{4\pi\epsilon_0}\frac{1}{2a} - \frac{e^2}{4\pi\epsilon_0}\frac{1}{3a} + \cdots$$

$$= -\frac{e^2}{4\pi\epsilon_0 a}\left(1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots\right)$$

You may notice that the series in parentheses resembles the Taylor series for the function $\ln(1 + x)$:

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

When $x = 1$ this becomes the series in parentheses above, so

$$\sum U = -\frac{e^2}{4\pi\epsilon_0 a}\ln 2$$

This is certainly a negative quantity. The atoms on the other side of the ion we're considering make an equal contribution to the potential energy. And if we include the potential energies of all pairs of atoms, the sum is certainly negative.

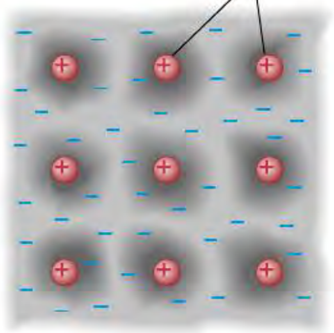
EVALUATE: We conclude that this one-dimensional ionic "crystal" is stable: It has lower energy than the zero electric potential energy that is obtained when all the ions are infinitely far apart from each other.

Types of Crystals

Carbon, silicon, germanium, and tin in the diamond structure are simple examples of *covalent crystals.* These elements are in Group IV of the periodic table, meaning that each atom has four electrons in its outermost shell. Each atom forms a covalent bond with each of four adjacent atoms at the corners of a tetrahedron (Fig. 42.14). These bonds are strongly directional because of the asymmetric electron distributions dictated by the exclusion principle (see Fig. 42.3), and the result is the tetrahedral diamond structure.

A third crystal type, less directly related to the chemical bond than are ionic or covalent crystals, is the **metallic crystal.** In this structure, one or more of the outermost electrons in each atom become detached from the parent atom (leaving a positive ion). The detached electrons are free to move through the crystal and are not localized near the ion from which they originated. So we can picture a metallic crystal as an array of positive ions immersed in a sea of freed electrons whose attraction for the positive ions holds the crystal together (**Fig. 42.15**).

42.15 In a metallic solid, one or more electrons are detached from each atom and are free to wander around the crystal, forming an "electron gas." The wave functions for these electrons extend over many atoms. The positive ions vibrate around fixed locations in the crystal.



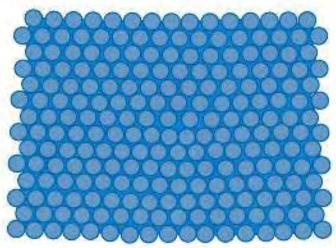
This sea of electrons, which gives metals their high electrical and thermal conductivities, has many of the properties of a gas. Indeed, we speak of the *electron-gas model* of metallic solids. The simplest version of this model is the *free-electron model*, which ignores interactions with the ions completely (except at the surface). We'll return to this model in Section 42.5.

In a metallic crystal the freed electrons are shared among *many* atoms. This gives a bonding that is neither localized nor strongly directional. The crystal structure is determined primarily by considerations of *close packing*—that is, the maximum number of atoms that can fit into a given volume. The two most common metallic crystal lattices are the face-centered cubic and hexagonal close-packed (see Figs. 42.11b, 42.11d, and 42.11e). In structures composed of these lattices with a basis of one atom, each atom has 12 nearest neighbors.

Hydrogen bonds and van der Waals forces also play a role in the structure of some solids. In polyethylene and similar polymers, covalent bonding of atoms forms long-chain molecules, and hydrogen bonding forms cross-links between adjacent chains. In solid water, both hydrogen bonds and van der Waals forces are significant in determining the crystal structures of ice.

Our discussion has centered on perfect crystals. Real crystals show a variety of departures from this idealized structure. Materials are often *polycrystalline*, composed of many small single crystals bonded together at *grain boundaries*. There may be *point defects* within a crystal: *Interstitial* atoms may occur in places where they do not belong, and there may be *vacancies*, positions that should be occupied by an atom but are not. A point defect of interest in semiconductors, which we will discuss in Section 42.6, is the *substitutional impurity*, a foreign atom replacing a regular atom (for example, arsenic in a silicon crystal).

42.16 An edge dislocation in two dimensions. In three dimensions an edge dislocation would look like an extra plane of atoms slipped partway into the crystal.



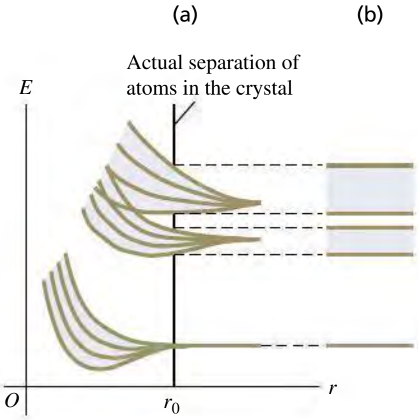
You can see the irregularity most easily by viewing the figure from various directions at a grazing angle with the page.

There are several basic types of extended defects called *dislocations.* One type is the *edge dislocation*, shown schematically in **Fig. 42.16**, in which one plane of atoms slips relative to another. The mechanical properties of metallic crystals are influenced strongly by the presence of dislocations. The ductility and malleability of some metals depend on the presence of dislocations that can move through the crystal during plastic deformations. The biggest extended defect of all, present in *all* real crystals, is the surface of the material with its dangling bonds and abrupt change in potential energy.

TEST YOUR UNDERSTANDING OF SECTION 42.3 If a is the distance in an NaCl crystal from an Na^+ ion to one of its nearest-neighbor Cl^- ions, what is the distance from an Na^+ ion to one of its *next-to-nearest*-neighbor Cl^- ions? (i) $a\sqrt{2}$; (ii) $a\sqrt{3}$; (iii) $2a$; (iv) none of these. ❙

42.17 The concept of energy bands was first developed by the Swiss-American physicist Felix Bloch (1905–1983) in his doctoral thesis. Our modern understanding of electrical conductivity stems from that landmark work. Bloch's work in nuclear physics brought him (along with Edward Purcell) the 1952 Nobel Prize in physics.

42.18 Origin of energy bands in a solid. (a) As the distance r between atoms decreases, the energy levels spread into bands. The vertical line at r_0 shows the actual atomic spacing in the crystal. (b) Symbolic representation of energy bands.



42.4 ENERGY BANDS

The **energy-band** concept, introduced in 1928 (**Fig. 42.17**), is a great help in understanding several properties of solids. To introduce the idea, suppose we have a large number N of identical atoms, far enough apart that their interactions are negligible. Every atom has the same energy-level diagram. We can draw an energy-level diagram for the *entire system*. It looks just like the diagram for a single atom, but the exclusion principle, applied to the entire system, permits each state to be occupied by N electrons (one per atom) instead of just one.

Now we begin to push the atoms uniformly closer together. Because of the electrical interactions and the exclusion principle, the wave functions begin to distort, especially those of the outer, or *valence*, electrons. The corresponding energies also shift, some upward and some downward, by varying amounts, as the valence electron wave functions become less localized and extend over more and more atoms. (The inner electrons in an atom are affected much less by nearby atoms than are the valence electrons, and their energy levels remain relatively sharp.) Thus the valence states that formerly gave the *system* a state with a sharp energy level that could accommodate N electrons now give a *band* containing N closely spaced levels (**Fig. 42.18**). Ordinarily, N is somewhere near the order of Avogadro's number (10^{24}), so we can accurately treat the levels as forming a *continuous* distribution of energies within a band. Between adjacent energy bands are gaps where there are *no* allowed energy levels.

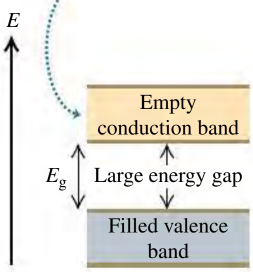
Insulators, Semiconductors, and Conductors

The nature of the energy bands determines whether the material is an electrical insulator, a semiconductor, or a conductor. In particular, what matters are the extent to which the states in each band are occupied and the spacing, called the *band gap* or *energy gap*, between adjacent bands.

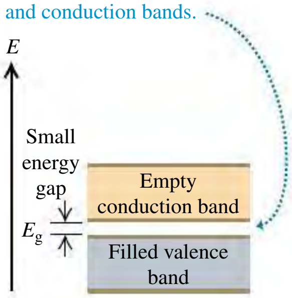
In an *insulator* at absolute zero temperature, the highest band that is completely filled, called the **valence band,** is also the highest band that has *any* electrons in it. The next higher band, called the **conduction band,** is completely empty; there are no electrons in its states (**Fig. 42.19a**). Imagine what happens if an electric field is applied to a material of this kind. To move in response to the field, an electron would have to go into a different quantum state with a slightly different energy. It can't do that, however, because all the neighboring states are already occupied. The only way such an electron can move is to jump across the energy gap into the conduction band, where there are plenty of nearby unoccupied states. At any temperature above absolute zero there is some probability this jump can happen, because an electron can gain energy from thermal motion. In an insulator, however, the energy gap between the valence and conduction bands can be 5 eV or more, and that much thermal energy is not ordinarily available. Hence little or no current flows in response to an applied electric field, and the electrical

42.19 Three types of energy-band structure.

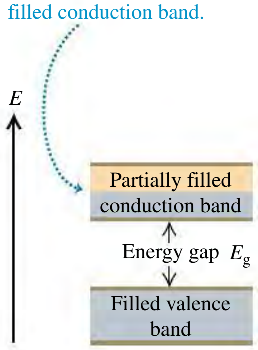
(a) In an insulator at absolute zero, there are no electrons in the conduction band.



(b) A semiconductor has the same band structure as an insulator but a smaller gap between the valence and conduction bands.



(c) A conductor has a partially filled conduction band.



conductivity (Section 25.2) is low. The thermal conductivity (Section 17.7), which also depends on mobile electrons, is likewise low.

We saw in Section 24.4 that an insulator becomes a conductor if it is subjected to a large enough electric field; this is called *dielectric breakdown*. If the electric field is of order 10^{10} V/m, there is a potential difference of a few volts over a distance comparable to atomic sizes. In this case the field can do enough work on a valence electron to boost it across the energy gap and into the conduction band. (In practice dielectric breakdown occurs for fields much less than 10^{10} V/m, because imperfections in the structure of an insulator provide some more accessible energy states *within* the energy gap.)

As in an insulator, a *semiconductor* at absolute zero has an empty conduction band above the full valence band. The difference is that in a semiconductor the energy gap between these bands is relatively small and electrons can more readily jump into the conduction band (Fig. 42.19b). As the temperature of a semiconductor increases, the population in the conduction band increases very rapidly, as does the electrical conductivity. For example, in a semiconductor near room temperature with an energy gap of 1 eV, the number of conduction electrons doubles when the temperature rises by just 10°C. We will use the concept of energy bands to explore semiconductors in more depth in Section 42.6.

In a *conductor* such as a metal, there are electrons in the conduction band even at absolute zero (Fig. 42.19c). The metal sodium is an example. An analysis of the atomic energy-level diagram for sodium (see Fig. 39.19a) shows that for an isolated sodium atom, the six lowest excited states (all 3*p* states) are about 2.1 eV above the two 3*s* ground states. In solid sodium, however, the atoms are so close together that the 3*s* and 3*p bands* spread out and overlap into a single band. Each sodium atom contributes one electron to the band, leaving an Na^+ ion behind. Each atom also contributes eight *states* to that band (two 3*s*, six 3*p*), so the band is only one-eighth occupied. We call this structure a *conduction* band because it is only partially occupied. Electrons near the top of the filled portion of the band have many adjacent unoccupied states available, and they can easily gain or lose small amounts of energy in response to an applied electric field. Therefore these electrons are mobile, giving solid sodium its high electrical and thermal conductivity. A similar description applies to other conducting materials.

MP

PhET: Band Structure

PhET: Conductivity

EXAMPLE 42.5 PHOTOCONDUCTIVITY IN GERMANIUM

SOLUTION

At room temperature, pure germanium has an almost completely filled valence band separated by a 0.67-eV gap from an almost completely empty conduction band. It is a poor electrical conductor, but its conductivity increases greatly when it is irradiated with electromagnetic waves of a certain maximum wavelength. What is that wavelength?

SOLUTION

IDENTIFY and SET UP: The conductivity of a semiconductor increases greatly when electrons are excited from the valence band into the conduction band. In germanium, the excitation occurs when an electron absorbs a photon with an energy of at least $E_{\min} = 0.67$ eV. From $E = hc/\lambda$, the *maximum* wavelength $\lambda_{\max}$ (our target variable) corresponds to this *minimum* photon energy.

EXECUTE: The wavelength of a photon with energy $E_{\min} = 0.67$ eV is

$$\lambda_{\max} = \frac{hc}{E_{\min}} = \frac{(4.136 \times 10^{-15}\ \text{eV} \cdot \text{s})(3.00 \times 10^{8}\ \text{m/s})}{0.67\ \text{eV}}$$
$$= 1.9 \times 10^{-6}\ \text{m} = 1.9\ \mu\text{m} = 1900\ \text{nm}$$

EVALUATE: This wavelength is in the infrared part of the spectrum, so visible-light photons (which have shorter wavelength and higher energy) will also induce conductivity in germanium. As we'll see in Section 42.7, semiconductor crystals are widely used as photocells and for many other applications.

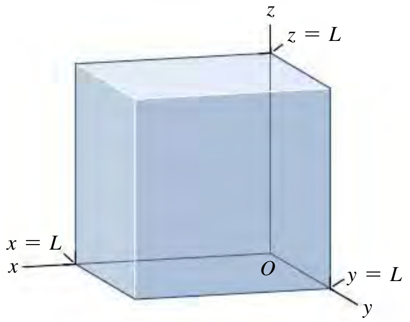
TEST YOUR UNDERSTANDING OF SECTION 42.4 One type of thermometer works by measuring the temperature-dependent electrical resistivity of a sample. Which of the following types of material displays the greatest change in resistivity for a given temperature change? (i) Insulator; (ii) semiconductor; (iii) conductor. ❙

42.5 FREE-ELECTRON MODEL OF METALS

Studying the energy states of electrons in metals can give us a lot of insight into their electrical and magnetic properties, the electron contributions to heat capacities, and other behavior. As we discussed in Section 42.3, one of the distinguishing features of a metal is that one or more valence electrons are detached from their home atom and can move freely within the metal, with wave functions that extend over many atoms.

The **free-electron model** assumes that these electrons don't interact at all with the ions or with each other, but that there are infinite potential-energy barriers at the surfaces. The idea is that a typical electron moves so rapidly within the metal that it "sees" the effect of the ions and other electrons as a uniform potential-energy function, whose value we can choose to be zero.

We can represent the surfaces of the metal by the same cubical box that we analyzed in Section 41.2 (the three-dimensional version of the particle in a box studied in Section 40.2). If the box has sides of length L (**Fig. 42.20**), the energies of the stationary states (quantum states of definite energy) are



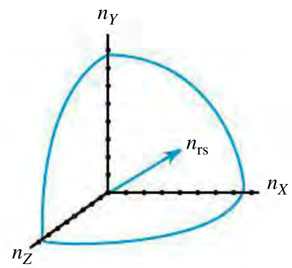
42.20 A cubical box with side length L. We studied this three-dimensional version of the infinite square well in Section 41.2. The energy levels for a particle in this box are given by Eq. (42.10).

$$E_{n_X,n_Y,n_Z} = \frac{(n_X{}^2 + n_Y{}^2 + n_Z{}^2)\pi^2\hbar^2}{2mL^2} \quad \begin{pmatrix} n_X = 1, 2, 3, \ldots; \\ n_Y = 1, 2, 3, \ldots; \\ n_Z = 1, 2, 3, \ldots \end{pmatrix} \tag{42.10}$$

Each state is labeled by the three positive-integer quantum numbers (n_X, n_Y, n_Z).

Density of States

Later we'll need to know the *number* dn of quantum states that have energies in a given range dE. The number of states per unit energy range dn/dE is called the **density of states,** denoted by $g(E)$. We'll begin by working out an expression for $g(E)$. Think of a three-dimensional space with coordinates (n_X, n_Y, n_Z) (**Fig. 42.21**). The radius n_{rs} of a sphere centered at the origin in that space is $n_{rs} = (n_X{}^2 + n_Y{}^2 + n_Z{}^2)^{1/2}$. Each point with integer coordinates in that space represents one spatial quantum state. Thus each point corresponds to one unit of volume in the space, and the total number of points with integer coordinates inside a sphere equals the volume of the sphere, $\frac{4}{3}\pi n_{rs}{}^3$. Because all our n's are positive, we must take only one *octant* of the sphere, with $\frac{1}{8}$ the total volume, or $(\frac{1}{8})(\frac{4}{3}\pi n_{rs}{}^3) = \frac{1}{6}\pi n_{rs}{}^3$. The particles are electrons, so each point corresponds to *two* states with opposite spin components ($m_s = \pm\frac{1}{2}$), and the total number n of electron states corresponding to points inside the octant is twice $\frac{1}{6}\pi n_{rs}{}^3$, or



42.21 The allowed values of n_X, n_Y, and n_Z are positive integers for the electron states in the free-electron gas model. Including spin, there are two states for each unit volume in n space.

$$n = \frac{\pi n_{rs}{}^3}{3} \tag{42.11}$$

The energy E of states at the surface of the sphere can be expressed in terms of n_{rs}. Equation (42.10) becomes

$$E = \frac{n_{rs}{}^2\pi^2\hbar^2}{2mL^2} \tag{42.12}$$

We can combine Eqs. (42.11) and (42.12) to get a relationship between E and n that doesn't contain n_{rs}. We'll leave the details for you to work out; the total number of states with energies of E or less is

$$n = \frac{(2m)^{3/2}VE^{3/2}}{3\pi^2\hbar^3} \tag{42.13}$$

where $V = L^3$ is the volume of the box.

To get the number of states dn in an energy interval dE, we treat n and E as continuous variables and take differentials of both sides of Eq. (42.13):

$$dn = \frac{(2m)^{3/2} V E^{1/2}}{2\pi^2 \hbar^3} dE \tag{42.14}$$

The density of states $g(E)$ is equal to dn/dE, so from Eq. (42.14) we get

Density of states, free-electron model:

Number of states per unit energy range near E; Electron mass; Volume

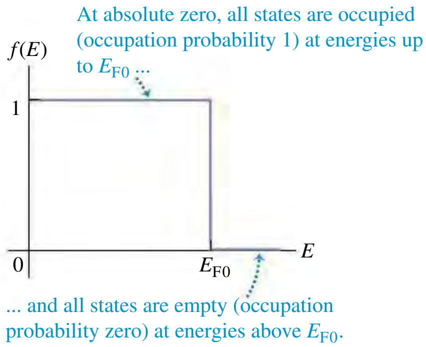
$$g(E) = \frac{(2m)^{3/2} V}{2\pi^2 \hbar^3} E^{1/2} \tag{42.15}$$

Planck's constant divided by 2π; Electron energy

Fermi–Dirac Distribution

Let's now see how the electrons are distributed among the various quantum states at any given temperature. The Maxwell–Boltzmann distribution states that the average number of particles in a state of energy E is proportional to $e^{-E/kT}$ (see Sections 18.5 and 39.4). However, there are two reasons why it wouldn't be right to use the Maxwell–Boltzmann distribution. The first reason is the exclusion principle. At absolute zero the Maxwell–Boltzmann function predicts that *all* the electrons would go into the two ground states of the system, with $n_X = n_Y = n_Z = 1$ and $m_s = \pm\frac{1}{2}$. But the exclusion principle allows only one electron in each state. At absolute zero the electrons can fill up the lowest *available* states, but they cannot *all* go into the lowest states. Thus at absolute zero the distribution function is as shown in **Fig. 42.22**. All states with energies E less than some value E_{F0} are occupied, so the occupation probability $f(E) = 1$; all states with energies greater than this value are unoccupied, so $f(E) = 0$.

42.22 The probability distribution for occupation of free-electron energy states at absolute zero.



The second reason we can't use the Maxwell–Boltzmann distribution is more subtle. That distribution assumes that the particles are *distinguishable*. But, as we discussed in Section 41.8, electrons are *indistinguishable*; it's impossible to "tag" electrons to know which is which. If one electron is in state A and the other is in state B, there's no way to tell whether electron 1 is in state A and electron 2 is in state B, or electron 1 is in state B and electron 2 is in state A.

The statistical distribution function that emerges from the exclusion principle and the indistinguishability requirement is called (after its inventors) the **Fermi–Dirac distribution.** This distribution gives the probability $f(E)$ that at temperature T, a particular state of energy E is occupied by an electron:

Fermi–Dirac distribution:

Probability that a given state is occupied by an electron; Absolute temperature

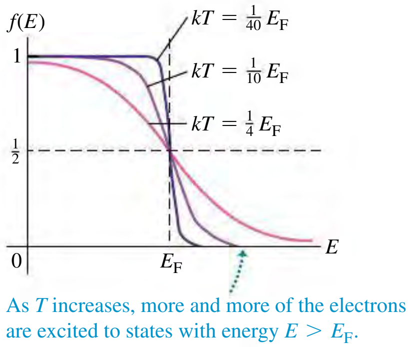
$$f(E) = \frac{1}{e^{(E - E_F)/kT} + 1} \tag{42.16}$$

Energy of state; Fermi energy; Boltzmann constant

The energy E_F is called the **Fermi energy** or the *Fermi level*; we'll discuss its significance below. We use E_{F0} for its value at absolute zero $(T = 0)$ and E_F for other temperatures. We can accurately let $E_F = E_{F0}$ for metals because the Fermi energy does not change much with temperature for solid conductors. However, it is not safe to assume that $E_F = E_{F0}$ for semiconductors, in which the Fermi energy usually does change with temperature.

Figure 42.23 shows graphs of Eq. (42.16) for three temperatures. The $kT = \frac{1}{40}E_F$ curve shows that for $kT \ll E_F$, $f(E)$ approaches the shape of the absolute-zero distribution function shown in Fig. 42.22.

42.23 Graphs of the Fermi–Dirac distribution function for various values of kT, assuming that the Fermi energy E_F is independent of the temperature T.



Note that when $E = E_F$, the exponent $(E - E_F)/kT$ in Eq. (42.16) is zero, so $f(E) = f(E_F) = \frac{1}{2}$ for any temperature T. That is, the probability is $\frac{1}{2}$ that a state at the Fermi energy contains an electron. This means that at $E = E_F$, half the states are filled (and half are empty). For states with $E < E_F$, the exponent is negative and the occupation probability $f(E)$ is greater than $\frac{1}{2}$; for states with $E > E_F$, $f(E)$ is less than $\frac{1}{2}$ and approaches zero for E much larger than kT.

EXAMPLE 42.6 PROBABILITIES IN THE FREE-ELECTRON MODEL

SOLUTION

For free electrons in a solid, at what energy is the probability that a particular state is occupied equal to (a) 0.01 and (b) 0.99?

SOLUTION

IDENTIFY and SET UP: This problem asks us to explore the Fermi–Dirac distribution. Equation (42.16) gives the occupation probability $f(E)$ for a given energy E. If we solve this equation for E, we get an expression for the energy that corresponds to a given occupation probability—which is just what we need.

EXECUTE: Using Eq. (42.16), you can show that

$$E = E_F + kT \ln\left(\frac{1}{f(E)} - 1\right)$$

(a) When $f(E) = 0.01$,

$$E = E_F + kT \ln\left(\frac{1}{0.01} - 1\right) = E_F + 4.6kT$$

The probability that a state $4.6kT$ above the Fermi level is occupied is only 0.01, or 1%.

(b) When $f(E) = 0.99$,

$$E = E_F + kT \ln\left(\frac{1}{0.99} - 1\right) = E_F - 4.6kT$$

The probability that a state $4.6kT$ below the Fermi level is occupied is 0.99, or 99%.

EVALUATE: At very low temperatures, $4.6kT$ is much less than E_F. Then the occupation probability of levels even slightly below E_F is nearly 1 (100%), and that for levels even slightly above E_F is nearly zero (see Fig. 42.23). In general, if the probability is P that a state with an energy ΔE *above* E_F is occupied, then the probability is $1 - P$ that a state ΔE *below* E_F is occupied. We leave the proof to you.

Electron Concentration and Fermi Energy

Equation (42.16) gives the probability that any specific state with energy E is occupied at a temperature T. To get the actual number of electrons in any energy range dE, we have to multiply this probability by the number dn of states in that range $g(E)\,dE$. Thus the number dN of electrons with energies in the range dE is

$$dN = g(E)f(E)\,dE = \frac{(2m)^{3/2}VE^{1/2}}{2\pi^2\hbar^3}\frac{1}{e^{(E-E_F)/kT} + 1}dE \tag{42.17}$$

The Fermi energy E_F is determined by the total number N of electrons; at any temperature the electron states are filled up to a point at which all electrons are accommodated. At absolute zero there is a simple relationship between E_{F0} and N. *All* states below E_{F0} are filled; in Eq. (42.13) we set n equal to the total number of electrons N and E to the Fermi energy at absolute zero E_{F0}:

$$N = \frac{(2m)^{3/2}VE_{F0}{}^{3/2}}{3\pi^2\hbar^3} \tag{42.18}$$

Solving Eq. (42.18) for E_{F0}, we get

$$E_{F0} = \frac{3^{2/3}\pi^{4/3}\hbar^2}{2m}\left(\frac{N}{V}\right)^{2/3} \tag{42.19}$$

The quantity N/V is the number of free electrons per unit volume. It is called the *electron concentration* and is usually denoted by n.

If we replace N/V with n, Eq. (42.19) becomes

$$E_{F0} = \frac{3^{2/3}\pi^{4/3}\hbar^2 n^{2/3}}{2m} \tag{42.20}$$

CAUTION Electron concentration and number of electrons Don't confuse the electron concentration n with any quantum number n. Furthermore, the number of states is *not* in general the same as the total number of electrons N.

SOLUTION

EXAMPLE 42.7 THE FERMI ENERGY IN COPPER

At low temperatures, copper has a free-electron concentration $n = 8.45 \times 10^{28}\ \text{m}^{-3}$. Using the free-electron model, find the Fermi energy for solid copper, and find the speed of an electron with a kinetic energy equal to the Fermi energy.

SOLUTION

IDENTIFY and SET UP: This problem uses the relationship between Fermi energy and free-electron concentration. Because copper is a solid conductor, its Fermi energy changes very little with temperature and we can use the expression for the Fermi energy at absolute zero, Eq. (42.20). We'll use the nonrelativistic formula $E_F = \frac{1}{2}mv_F^2$ to find the *Fermi speed* v_F that corresponds to kinetic energy E_F.

EXECUTE: Using the given value of n, we solve for E_F and v_F:

$$E_F = \frac{3^{2/3}\pi^{4/3}(1.055 \times 10^{-34}\ \text{J}\cdot\text{s})^2(8.45 \times 10^{28}\ \text{m}^{-3})^{2/3}}{2(9.11 \times 10^{-31}\ \text{kg})}$$
$$= 1.126 \times 10^{-18}\ \text{J} = 7.03\ \text{eV}$$

$$v_F = \sqrt{\frac{2E_F}{m}} = \sqrt{\frac{2(1.126 \times 10^{-18}\ \text{J})}{9.11 \times 10^{-31}\ \text{kg}}} = 1.57 \times 10^6\ \text{m/s}$$

EVALUATE: Our values of E_F and v_F are within the ranges of typical values for metals, 1.6–14 eV and $0.8\text{–}2.2 \times 10^6$ m/s, respectively. Note that the calculated Fermi speed is far less than the speed of light $c = 3.00 \times 10^8$ m/s, which justifies our use of the nonrelativistic formula $\frac{1}{2}mv_F^2 = E_F$.

Our calculated Fermi energy is much larger than kT at ordinary temperatures. (At room temperature $T = 20°\text{C} = 293$ K, the quantity kT equals $(1.381 \times 10^{-23}\ \text{J/K})(293\ \text{K}) = 4.04 \times 10^{-21}\ \text{J} = 0.0254$ eV.) So it is a good approximation to take almost all the states below E_F as completely full and almost all those above E_F as completely empty (see Fig. 42.22).

We can also use Eq. (42.15) to find $g(E)$ if E and V are known. You can show that if $E = 7.03$ eV and $V = 1\ \text{cm}^3$, $g(E)$ is about 2×10^{22} states/eV. This huge number shows why we were justified in treating n and E as continuous variables in our density-of-states derivation.

Average Free-Electron Energy

We can calculate the *average* free-electron energy in a metal at absolute zero by using the same ideas that we used to find E_{F0}. From Eq. (42.17) the number dN of electrons with energies in the range dE is $g(E)f(E)\,dE$. The energy of these electrons is $E\,dN = Eg(E)f(E)\,dE$. At absolute zero we substitute $f(E) = 1$ from $E = 0$ to $E = E_{F0}$ and $f(E) = 0$ for all other energies. Therefore the total energy E_{tot} of all the N electrons is

$$E_{tot} = \int_0^{E_{F0}} Eg(E)(1)\,dE + \int_{E_{F0}}^{\infty} Eg(E)(0)\,dE = \int_0^{E_{F0}} Eg(E)\,dE$$

The simplest way to evaluate this expression is to compare Eqs. (42.15) and (42.19). You'll see that

$$g(E) = \frac{3NE^{1/2}}{2E_{F0}^{3/2}}$$

Substituting this expression into the integral and using $E_{av} = E_{tot}/N$, we get

$$E_{av} = \frac{3}{2E_{F0}^{3/2}}\int_0^{E_{F0}} E^{3/2}\,dE = \tfrac{3}{5}E_{F0} \qquad (42.21)$$

At absolute zero the average free-electron energy equals $\frac{3}{5}$ of the Fermi energy.

SOLUTION

EXAMPLE 42.8 FREE-ELECTRON GAS VERSUS IDEAL GAS

(a) Find the average energy of the free electrons in copper at absolute zero (see Example 42.7). (b) What would be the average kinetic energy of electrons if they behaved like an ideal gas at room temperature, 20°C (see Section 18.3)? What would be the speed of an electron with this kinetic energy? Compare these ideal-gas values with the (correct) free-electron values.

SOLUTION

IDENTIFY and SET UP: Free electrons in a metal behave like a kind of gas. In part (a) we use Eq. (42.21) to determine the average kinetic energy of free electrons in terms of the Fermi energy

Continued

at absolute zero, which we know for copper from Example 42.7. In part (b) we treat electrons as an ideal gas at room temperature: Eq. (18.16) then gives the average kinetic energy per electron as $E_{av} = \frac{3}{2}kT$, and $E_{av} = \frac{1}{2}mv^2$ gives the corresponding electron speed v.

EXECUTE: (a) From Example 42.7, the Fermi energy in copper at absolute zero is 1.126×10^{-18} J = 7.03 eV. According to Eq. (42.21), the average energy is $\frac{3}{5}$ of this, or 6.76×10^{-19} J = 4.22 eV.

(b) In Example 42.7 we found that $kT = 4.04 \times 10^{-21}$ J = 0.0254 eV at room temperature $T = 20°C = 293$ K. If electrons behaved like an ideal gas at this temperature, the average kinetic energy per electron would be $\frac{3}{2}$ of this, or 6.07×10^{-21} J = 0.0379 eV. The speed of an electron with this kinetic energy is

$$v = \sqrt{\frac{2E_{av}}{m}} = \sqrt{\frac{2(6.07 \times 10^{-21}\ \text{J})}{9.11 \times 10^{-31}\ \text{kg}}}$$
$$= 1.15 \times 10^5\ \text{m/s}$$

EVALUATE: The ideal-gas model predicts an average energy that is about 1% of the value given by the free-electron model, and a speed that is about 7% of the free-electron Fermi speed $v_F = 1.57 \times 10^6$ m/s that we found in Example 42.7. Thus temperature plays a *very* small role in determining the properties of electrons in metals; their average energies are determined almost entirely by the exclusion principle.

A similar analysis allows us to determine the contributions of electrons to the heat capacities of a solid metal. If there is one conduction electron per atom, the principle of equipartition of energy (see Section 18.4) would predict that the kinetic energies of these electrons contribute $3R/2$ to the molar heat capacity at constant volume C_V. But when kT is much smaller than E_F, which is usually the situation in metals, only those few electrons near the Fermi level can find empty states and change energy appreciably when the temperature changes. The number of such electrons is proportional to kT/E_F, so we expect that the electron molar heat capacity at constant volume is proportional to $(kT/E_F)(3R/2) = (3kT/2E_F)R$. A more detailed analysis shows that the actual electron contribution to C_V for a solid metal is $(\pi^2 kT/2E_F)R$, not far from our prediction. You can verify that if $T = 293$ K and $E_F = 7.03$ eV, the electron contribution to C_V is $0.018R$, which is only 1.2% of the (incorrect) $3R/2$ prediction of the equipartition principle. Because the electron contribution is so small, the overall heat capacity of most solid metals is due primarily to vibration of the atoms in the crystal structure (see Fig. 18.18 in Section 18.4).

TEST YOUR UNDERSTANDING OF SECTION 42.5 An ideal gas obeys the relationship $pV = nRT$ (see Section 18.1): For a given volume V and a number of moles n, as the temperature T decreases, the pressure p decreases proportionately and tends to zero as T approaches absolute zero. Is this also true of the free-electron gas in a solid metal?

42.6 SEMICONDUCTORS

PhET: Semiconductors
PhET: Conductivity

A **semiconductor** has an electrical resistivity that is intermediate between those of good conductors and of good insulators. The tremendous importance of semiconductors in present-day electronics stems in part from the fact that their electrical properties are very sensitive to very small concentrations of impurities. We'll discuss the basic concepts, using the semiconductor elements silicon (Si) and germanium (Ge) as examples.

Silicon and germanium are in Group IV of the periodic table. Both have four electrons in the outermost atomic subshells ($3s^2 3p^2$ for silicon, $4s^2 4p^2$ for germanium), and both crystallize in the covalently bonded diamond structure discussed in Section 42.3 (see Fig. 42.14). Because all four of the outer electrons are involved in the bonding, at absolute zero the band structure (see Section 42.4) has a completely empty conduction band (see Fig. 42.19b). As we discussed in Section 42.4, at very low temperatures electrons cannot jump from the filled valence band into the conduction band. This property makes these materials insulators at very low temperatures; their electrons have no nearby states available into which they can move in response to an applied electric field.

However, in semiconductors the energy gap E_g between the valence and conduction bands is small in comparison to the gap of 5 eV or more for many insulators; room-temperature values are 1.12 eV for silicon and only 0.67 eV for germanium. Thus even at room temperature a substantial number of electrons can gain enough energy to jump the gap to the conduction band, where they are dissociated from their parent atoms and are free to move about the crystal. The number of these electrons increases rapidly with temperature.

EXAMPLE 42.9 JUMPING A BAND GAP

SOLUTION

Consider a material with the band structure described above, with its Fermi energy in the middle of the gap (**Fig. 42.24**). Find the probability that a state at the bottom of the conduction band is occupied at $T = 300$ K, and compare that with the probability at $T = 310$ K, for band gaps of (a) 0.200 eV; (b) 1.00 eV; (c) 5.00 eV.

SOLUTION

IDENTIFY and SET UP: The Fermi–Dirac distribution function gives the probability that a state of energy E is occupied at temperature T. Figure 42.24 shows that the state of interest at the bottom of the conduction band has an energy $E = E_F + E_g/2$ that is greater than the Fermi energy E_F, with $E - E_F = E_g/2$. Figure 42.23 shows that the higher the temperature, the larger the fraction of electrons with energies greater than the Fermi energy.

EXECUTE: (a) When $E_g = 0.200$ eV,

$$\frac{E - E_F}{kT} = \frac{E_g}{2kT} = \frac{0.100\text{ eV}}{(8.617 \times 10^{-5}\text{ eV/K})(300\text{ K})} = 3.87$$

$$f(E) = \frac{1}{e^{3.87} + 1} = 0.0205$$

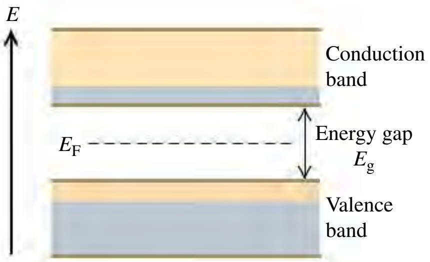
For $T = 310$ K, the exponent is 3.74 and $f(E) = 0.0231$, a 13% increase in probability for a temperature rise of 10 K.

(b) For $E_g = 1.00$ eV, both exponents are five times as large as in part (a), namely 19.3 and 18.7; the values of $f(E)$ are 4.0×10^{-9} and 7.4×10^{-9}. In this case the (low) probability nearly doubles with a temperature rise of 10-K.

(c) For $E_g = 5.0$ eV, the exponents are 96.7 and 93.6; the values of $f(E)$ are 1.0×10^{-42} and 2.3×10^{-41}. The (extremely low) probability increases by a factor of 23 for a 10-K temperature rise.

EVALUATE: This example illustrates two important points. First, the probability of finding an electron in a state at the bottom of the conduction band is extremely sensitive to the width of the band gap. At room temperature, the probability is about 2% for a 0.200-eV gap, a few in a thousand million for a 1.00-eV gap, and essentially zero for a 5.00-eV gap. (Pure diamond, with a 5.47-eV band gap, has essentially no electrons in the conduction band and is an excellent insulator.) Second, for any given band gap the probability depends strongly on temperature, and even more strongly for large gaps than for small ones.

42.24 Band structure of a semiconductor. At absolute zero a completely filled valence band is separated by a narrow energy gap E_g of 1 eV or so from a completely empty conduction band. At ordinary temperatures, a number of electrons are excited to the conduction band.



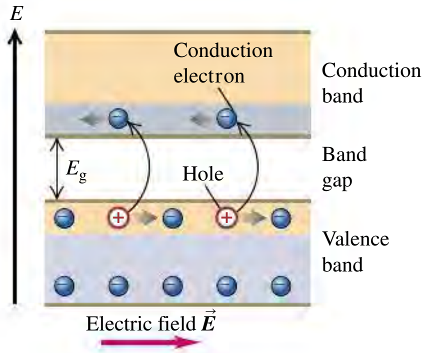
In principle, we could continue the calculation in Example 42.9 to find the actual density $n = N/V$ of electrons in the conduction band at any temperature. To do this, we would have to evaluate the integral $\int g(E)f(E)\,dE$ from the bottom of the conduction band to its top. [To do this we would need to know the density of states function $g(E)$. This function for a semiconductor is more complicated than the free-electron model expression given by Eq. (42.15).] Once we know n, we can *begin* to determine the resistivity of the material (and its temperature dependence) by using the analysis of Section 25.2, which you may want to review. But next we'll see that the electrons in the conduction band don't tell the whole story about conduction in semiconductors.

Holes

When an electron is removed from a covalent bond, it leaves a vacancy behind. An electron from a neighboring atom can move into this vacancy, leaving the neighbor with the vacancy. In this way the vacancy, called a **hole,** can travel through the material and serve as an additional current carrier. It's like describing the motion of a bubble in a liquid. In a pure, or *intrinsic*, semiconductor, valence-band holes and conduction-band electrons are always present in equal numbers. When an electric field is applied, they move in opposite directions (**Fig. 42.25**). Thus a hole in the valence band behaves like a positively charged particle, even though the moving charges in that band are electrons. The conductivity that we just described for a pure semiconductor is called *intrinsic conductivity*. Another kind of conductivity, to be discussed in the next subsection, is due to impurities.

An analogy helps to picture conduction in an intrinsic semiconductor. The valence band at absolute zero is like a floor of a parking garage that's filled bumper to bumper with cars (which represent electrons). No cars can move

42.25 Motion of electrons in the conduction band and of holes in the valence band of a semiconductor under the action of an applied electric field $\vec{E}$.



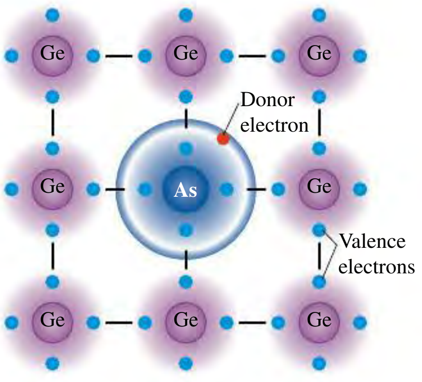
because there is nowhere for them to go. But if one car is moved to the vacant floor above, it can move freely, just as electrons can move freely in the conduction band. Also, the empty space that it leaves permits cars to move on the nearly filled floor, thereby moving the empty space just as holes move in the normally filled valence band.

Impurities

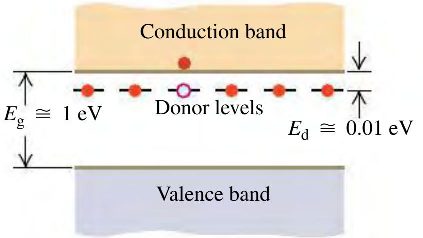
Suppose we mix into melted germanium $(Z = 32)$ a small amount of arsenic $(Z = 33)$, the next element after germanium in the periodic table. This deliberate addition of impurity elements is called *doping*. Arsenic is in Group V; it has *five* valence electrons. When one of these electrons is removed, the remaining electron structure is essentially identical to that of germanium. The only difference is that it is smaller; the arsenic nucleus has a charge of $+33e$ rather than $+32e$, and it pulls the electrons in a little more. An arsenic atom can comfortably take the place of a germanium atom as a substitutional impurity. Four of its five valence electrons form the necessary nearest-neighbor covalent bonds.

42.26 An n-type semiconductor.

(a) A donor (n-type) impurity atom has a fifth valence electron that does not participate in the covalent bonding and is very loosely bound.



(b) Energy-band diagram for an n-type semiconductor at a low temperature. One donor electron has been excited from the donor levels into the conduction band.



The fifth valence electron is very loosely bound (**Fig. 42.26a**); it doesn't participate in the covalent bonds, and it is screened from the nuclear charge of $+33e$ by the 32 electrons, leaving a net effective charge of about $+e$. We might guess that the binding energy would be of the same order of magnitude as the energy of the $n = 4$ level in hydrogen—that is, $(\frac{1}{4})^2(13.6\text{ eV}) = 0.85\text{ eV}$. In fact, it is much smaller than this, only about 0.01 eV, because the electron probability distribution actually extends over many atomic diameters and the polarization of intervening atoms provides additional screening.

The energy level of this fifth electron corresponds in the band picture to an isolated energy level lying in the gap, about 0.01 eV below the bottom of the conduction band (Fig. 42.26b). This level is called a *donor level*, and the impurity atom that is responsible for it is simply called a *donor*. All Group V elements, including N, P, As, Sb, and Bi, can serve as donors. At room temperature, kT is about 0.025 eV. This is substantially greater than 0.01 eV, so at ordinary temperatures, most electrons can gain enough energy to jump from donor levels into the conduction band, where they are free to wander through the material. The remaining ionized donor stays at its site in the structure and does not participate in conduction.

Example 42.9 shows that at ordinary temperatures and with a band gap of 1.0 eV, only a very small fraction (of the order of 10^{-9}) of the states at the bottom of the conduction band in a pure semiconductor contain electrons to participate in intrinsic conductivity. Thus we expect the conductivity of such a semiconductor to be about 10^{-9} as great as that of good metallic conductors, and measurements bear out this prediction. However, a concentration of donors as small as one part in 10^8 can increase the conductivity so drastically that conduction due to impurities becomes by far the dominant mechanism. In this case the conductivity is due almost entirely to *negative* charge (electron) motion. We call the material an ***n*-type semiconductor,** with n-type impurities.

Adding atoms of an element in Group III (B, Al, Ga, In, Tl), with only *three* valence electrons, has an analogous effect. An example is gallium $(Z = 31)$; as a substitutional impurity in germanium, the gallium atom would like to form four covalent bonds, but it has only three outer electrons. It can, however, steal an electron from a neighboring germanium atom to complete the required four covalent bonds (**Fig. 42.27a**). The resulting atom has the same electron configuration as Ge but is somewhat larger because gallium's nuclear charge is smaller, $+31e$ instead of $+32e$.

This theft leaves the neighboring atom with a *hole*, or missing electron. The hole acts as a positive charge that can move through the crystal just as with intrinsic conductivity. The stolen electron is bound to the gallium atom in a level called an *acceptor level* about 0.01 eV above the top of the valence band

(Fig. 42.27b). The gallium atom, called an *acceptor*, thus accepts an electron to complete its desire for four covalent bonds. This extra electron gives the previously neutral gallium atom a net charge of $-e$. The resulting gallium ion is *not* free to move. In a semiconductor that is doped with acceptors, we consider the conductivity to be almost entirely due to *positive* charge (hole) motion. We call the material a ***p*-type semiconductor,** with p-type impurities. Some semiconductors are doped with *both* n- and p-type impurities. Such materials are called *compensated* semiconductors.

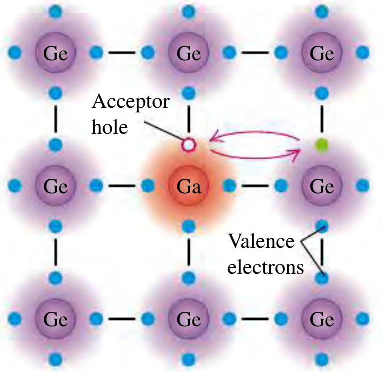
CAUTION **The meaning of "*p*-type" and "*n*-type"** Saying that a material is a p-type semiconductor does *not* mean that the material has a positive charge; ordinarily, it would be neutral. Rather, it means that its *majority carriers* of current are positive holes (and therefore its *minority carriers* are negative electrons). The same idea holds for an n-type semiconductor; ordinarily, it will *not* have a negative charge, but its majority carriers are negative electrons.

We can verify the assertion that the current in n- and p-type semiconductors really *is* carried by electrons and holes, respectively, by using the Hall effect (see Section 27.9). The sign of the Hall emf is opposite in the two cases. Hall-effect devices constructed from semiconductor materials are used in probes to measure magnetic fields and the currents that cause those fields.

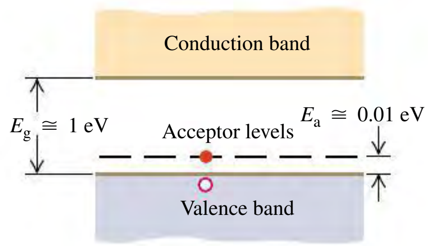
TEST YOUR UNDERSTANDING OF SECTION 42.6 Would there be any advantage to adding n-type or p-type impurities to copper?

42.27 A p-type semiconductor.

(a) An acceptor (p-type) impurity atom has only three valence electrons, so it can borrow an electron from a neighboring atom. The resulting hole is free to move about the crystal.



(b) Energy-band diagram for a p-type semiconductor at a low temperature. One acceptor level has accepted an electron from the valence band, leaving a hole behind.



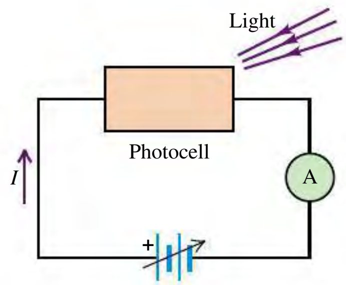
42.7 SEMICONDUCTOR DEVICES

Semiconductor devices play an indispensable role in contemporary electronics. In the early days of radio and television, transmitting and receiving equipment relied on vacuum tubes, but these have been replaced by solid-state devices, including transistors, diodes, integrated circuits, and other semiconductor devices. All modern consumer electronic devices use semiconductor devices of various kinds.

One simple semiconductor device is the *photocell* (**Fig. 42.28**). When a thin slab of semiconductor is irradiated with an electromagnetic wave whose photons have at least as much energy as the band gap between the valence and conduction bands, an electron in the valence band can absorb a photon and jump to the conduction band, where it and the hole it left behind contribute to the conductivity (see Example 42.5 in Section 42.4). The conductivity therefore increases with wave intensity, thus increasing the current I in the photocell circuit of Fig. 42.28. Hence the ammeter reading indicates the intensity of the light.

Detectors for charged particles operate on the same principle. An external circuit applies a voltage across a semiconductor. An energetic charged particle passing through the semiconductor collides inelastically with valence electrons, exciting them from the valence to the conduction band and creating pairs of holes and conduction electrons. The conductivity increases momentarily, causing a pulse of current in the external circuit. Solid-state detectors are widely used in nuclear and high-energy physics research.

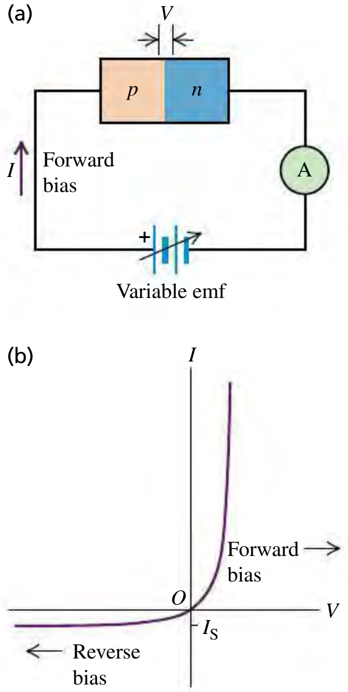
42.28 A semiconductor photocell in a circuit. The more intense the light falling on the photocell, the greater the conductivity of the photocell and the greater the current measured by the ammeter (A).



The *p-n* Junction

In many semiconductor devices the essential principle is the fact that the conductivity of the material is controlled by impurity concentrations, which can be varied within wide limits from one region of a device to another. An example is the ***p-n* junction** at the boundary between one region of a semiconductor with p-type impurities and another region containing n-type impurities. One way of fabricating a p-n junction is to deposit some n-type material on the *very* clean surface of some p-type material. (We can't just stick p- and n-type pieces together and expect the junction to work properly because of the impossibility of matching their surfaces at the atomic level.)

42.29 (a) A semiconductor *p-n* junction in a circuit. (b) Graph showing the asymmetric current–voltage relationship. The curve is described by Eq. (42.22).



When a *p-n* junction is connected to an external circuit, as in **Fig. 42.29a**, and the potential difference $V_p - V_n = V$ across the junction is varied, the current I varies as shown in Fig. 42.29b. In striking contrast to the symmetric behavior of resistors that obey Ohm's law and give a straight line on an I–V graph, a *p-n* junction conducts much more readily in the direction from p to n than the reverse. Such a (mostly) one-way device is called a **diode rectifier.** Later we'll discuss a simple model of *p-n* junction behavior that predicts a current–voltage relationship in the form

$$I = I_S\left(e^{eV/kT} - 1\right) \qquad (42.22)$$

Current through a *p-n* junction: I; Saturation current: I_S; Voltage: V; Absolute temperature: T; Boltzmann constant: k; Magnitude of electron charge: e (in the exponent); $e = 2.71828\ldots$ (base)

CAUTION **Two different uses of *e*** In $e^{eV/kT}$ the base of the exponent also uses the symbol e, standing for the base of the natural logarithms, 2.71828 . . . . This e is quite different from $e = 1.602 \times 10^{-19}$ C in the exponent.

Equation (42.22) is valid for both positive and negative values of V; note that V and I always have the same sign. As V becomes very negative, I approaches the value $-I_S$. The magnitude I_S (always positive) is called the *saturation current.*

Currents Through a *p-n* Junction

We can understand the behavior of a *p-n* junction diode qualitatively on the basis of the mechanisms for conductivity in the two regions. Suppose, as in Fig. 42.29a, you connect the positive terminal of the battery to the p region and the negative terminal to the n region. Then the p region is at higher potential than the n region, corresponding to positive V in Eq. (42.22), and the resulting electric field is in the direction p to n. This is called the *forward* direction, and the positive potential difference is called *forward bias*. Holes, plentiful in the p region, flow easily across the junction into the n region, and free electrons, plentiful in the n region, easily flow into the p region; these movements of charge constitute a *forward* current. Connecting the battery with the opposite polarity gives *reverse bias*, and the field tends to push electrons from p to n and holes from n to p. But there are very few free electrons in the p region and very few holes in the n region. As a result, the current in the *reverse* direction is much smaller than that with the same potential difference in the forward direction.

Suppose you have a box with a barrier separating the left and right sides: You fill the left side with oxygen gas and the right side with nitrogen gas. What happens if the barrier leaks? Oxygen diffuses to the right, and nitrogen diffuses to the left. A similar diffusion occurs across a *p-n* junction. First consider the equilibrium situation with no applied voltage (**Fig. 42.30**). The many holes in the p region act like a hole gas that diffuses across the junction into the n region. Once there, the holes recombine with some of the many free electrons. Similarly, electrons diffuse from the n region to the p region and fall into some of the many holes there. The hole and electron diffusion currents lead to a net positive charge in the n region and a net negative charge in the p region, causing an electric field in the direction from n to p at the junction. The potential energy associated with this field raises the electron energy levels in the p region relative to the same levels in the n region.

There are four currents across the junction, as shown. The diffusion processes lead to *recombination currents* of holes and electrons, labeled i_{pr} and i_{nr} in Fig. 42.30. At the same time, electron–hole pairs are generated in the junction region by thermal excitation. The electric field described above sweeps these electrons and holes out of the junction; electrons are swept opposite the field to the n side, and holes are swept in the same direction as the field to the p side. The corresponding currents, called *generation currents*, are labeled i_{pg} and i_{ng}. At equilibrium the magnitudes of the generation and recombination currents are equal:

$$|i_{pg}| = |i_{pr}| \quad \text{and} \quad |i_{ng}| = |i_{nr}| \qquad (42.23)$$

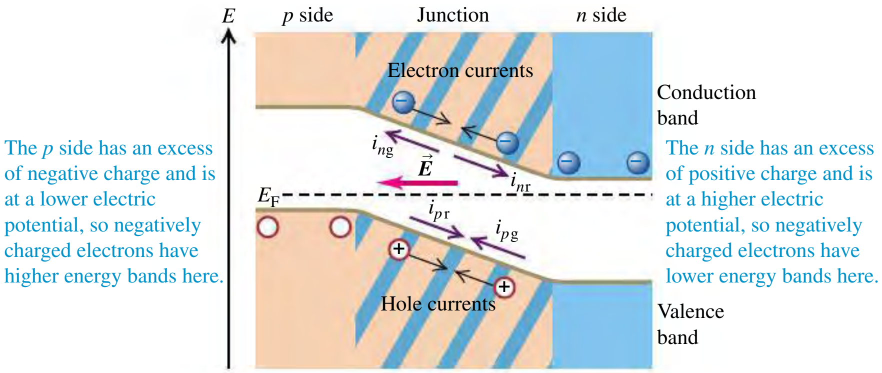
DATA *SPEAKS*

Semiconductors

When students were given a problem involving semiconductors, more than 24% gave an incorrect response. Common errors:

- Confusion about *p*-type and *n*-type semiconductors. The mobile charges (holes) in a *p*-type semiconductor are positive, but the semiconductor as a whole does not have a net positive charge. Similarly, the mobile charges (electrons) in an *n*-type semiconductor are negative, but the semiconductor does not have a net negative charge.
- Confusion about band gaps. An electron can absorb a photon and be promoted from the valence band to the conduction band, but only if it is given enough energy to bridge the gap between these bands. Hence a photon with an energy less than that of the band gap cannot be absorbed.

42.30 A p-n junction in equilibrium, with no externally applied field or potential difference. The generation currents (subscript g) and recombination currents (subscript r) exactly balance. The Fermi energy E_F is the same on both sides of the junction. The excess positive and negative charges on the n and p sides produce an electric field $\vec{E}$ in the direction shown.



In thermal equilibrium the Fermi energy is the same at each point across the junction.

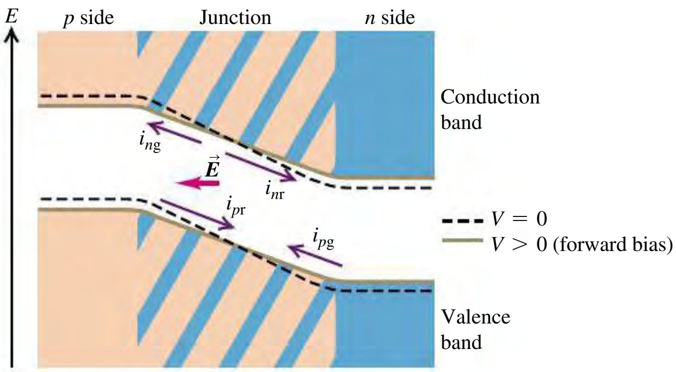
Now we apply a forward bias—that is, a positive potential difference V across the junction. A forward bias *decreases* the electric field in the junction region. It also decreases the difference between the energy levels on the p and n sides (**Fig. 42.31**) by an amount $\Delta E = -eV$. It becomes easier for the electrons in the n region to climb the potential-energy hill and diffuse into the p region and for the holes in the p region to diffuse into the n region. This effect increases both recombination currents by the Maxwell–Boltzmann factor $e^{-\Delta E/kT} = e^{eV/kT}$. (We don't have to use the Fermi–Dirac distribution because most of the available states for the diffusing electrons and holes are empty, so the exclusion principle has little effect.) The generation currents don't change appreciably, so the net hole current is

$$\begin{aligned} i_{p\text{tot}} &= i_{p\text{r}} - |i_{p\text{g}}| \\ &= |i_{p\text{g}}|e^{eV/kT} - |i_{p\text{g}}| \\ &= |i_{p\text{g}}|(e^{eV/kT} - 1) \end{aligned} \tag{42.24}$$

The net electron current $i_{n\text{tot}}$ is given by a similar expression, so the total current $I = i_{p\text{tot}} + i_{n\text{tot}}$ is

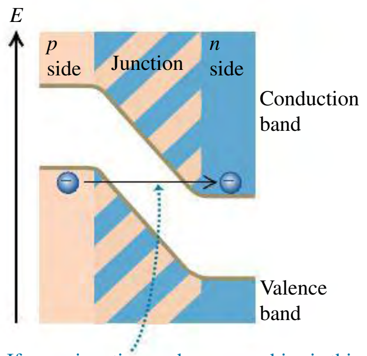
$$I = I_S(e^{eV/kT} - 1) \tag{42.25}$$

in agreement with Eq. (42.22). We can repeat this entire discussion for reverse bias (negative V and I) with the same result. Therefore Eq. (42.22) is valid for both positive and negative values.



42.31 A p-n junction under forward-bias conditions. The potential difference between p and n regions is reduced, as is the electric field within the junction. The recombination currents increase but the generation currents are nearly constant, causing a net current from left to right. (Compare Fig. 42.30.)

42.32 Under reverse-bias conditions the potential-energy difference between the p and n sides of a junction is greater than at equilibrium. If this difference is great enough, the bottom of the conduction band on the n side may actually be below the top of the valence band on the p side.



If a p-n junction under reverse bias is thin enough, electrons can tunnel from the valence band to the conduction band (a process called Zener breakdown).

BIO Application Swallow This Semiconductor Device This tiny capsule—designed to be swallowed by a patient—contains a miniature camera with a CCD light detector, plus six LEDs to illuminate the subject. The capsule radios high-resolution images to an external recording unit as it passes painlessly through the patient's stomach and intestines. This technique makes it possible to examine the small intestine, which is not readily accessible with conventional endoscopy.

Several effects make the behavior of practical p-n junction diodes more complex than this simple analysis predicts. One effect, *avalanche breakdown*, occurs under large reverse bias. The electric field in the junction is so great that the carriers can gain enough energy between collisions to create electron–hole pairs during inelastic collisions. The electrons and holes then gain energy and collide to form more pairs, and so on. (A similar effect occurs in dielectric breakdown in insulators, discussed in Section 42.4.)

A second type of breakdown begins when the reverse bias becomes large enough that the top of the valence band in the p region is just higher in energy than the bottom of the conduction band in the n region (**Fig. 42.32**). If the junction region is thin enough, the probability becomes large that electrons can *tunnel* from the valence band of the p region to the conduction band of the n region. This process is called *Zener breakdown*. It occurs in Zener diodes, which are used for voltage regulation and protection against voltage surges.

Semiconductor Devices and Light

A *light-emitting diode (LED)* is a p-n junction diode that emits light. When the junction is forward biased, many holes are pushed from their p region to the junction region, and many electrons are pushed from their n region to the junction region. In the junction region the electrons fall into holes (recombine). In recombining, the electron can emit a photon with energy approximately equal to the band gap. This energy (and therefore the photon wavelength and the color of the light) can be varied by using materials with different band gaps. Light-emitting diodes are very energy-efficient light sources and have many applications, including automobile lamps, traffic signals, and flat-screen displays.

The reverse process is called the *photovoltaic effect*. Here the material absorbs photons, and electron–hole pairs are created. Pairs that are created in the p-n junction, or close enough to migrate to it without recombining, are separated by the electric field we described above that sweeps the electrons to the n side and the holes to the p side. We can connect this device to an external circuit, where it becomes a source of emf and power. Such a device is often called a *solar cell*, although sunlight isn't required. *Any* light with photon energies greater than the band gap will do. You might have a calculator powered by such cells. Production of low-cost photovoltaic cells for large-scale solar energy conversion is a very active field of research. The same basic physics is used in charge-coupled device (CCD) image detectors, digital cameras, and video cameras.

Transistors

A *bipolar junction transistor* includes two p-n junctions in a "sandwich" configuration, which may be either p-n-p or n-p-n. **Figure 42.33** shows such a p-n-p transistor. The three regions are called the emitter, base, and collector, as shown. When there is no current in the left loop of the circuit, there is only a very small current through the resistor R because the voltage across the base–collector junction is in the reverse direction. But when a forward bias is applied between emitter and base, as shown, most of the holes traveling from emitter to base travel *through* the base (which is typically both narrow and lightly doped) to the second junction, where they come under the influence of the collector-to-base potential difference and flow on through the collector to give an increased current to the resistor.

In this way the current in the collector circuit is *controlled* by the current in the emitter circuit. Furthermore, V_c may be considerably larger than V_e, so the *power* dissipated in R may be much larger than the power supplied to the emitter circuit by the battery V_e. Thus the device functions as a *power amplifier*. If the potential drop across R is greater than V_e, it may also be a voltage amplifier.

In this configuration the *base* is the common element between the "input" and "output" sides of the circuit. Another widely used arrangement is the *common-emitter* circuit, shown in **Fig. 42.34**. In this circuit the current in the collector side of the circuit is much larger than that in the base side, and the result is current amplification.

The *field-effect transistor* (**Fig. 42.35**) is an important type. In one variation a slab of p-type silicon is made with two n-type regions on the top, called the *source* and the *drain*; a metallic conductor is fastened to each. A third electrode called the *gate* is separated from the slab, source, and drain by an insulating layer of SiO_2. When there is no charge on the gate and a potential difference of either polarity is applied between the source and the drain, there is very little current because one of the p-n junctions is reverse biased.

Now we place a positive charge on the gate. With dimensions of the order of 10^{-6} m, it takes little charge to provide a substantial electric field. Thus there is very little current into or out of the gate. There aren't many free electrons in the p-type material, but there are some, and the effect of the field is to attract them toward the positive gate. The resulting greatly enhanced concentration of electrons near the gate (and between the two junctions) permits current to flow between the source and the drain. The current is very sensitive to the gate charge and potential, and the device functions as an amplifier. The device just described is called an *enhancement-type MOSFET* (metal-oxide-semiconductor field-effect transistor).

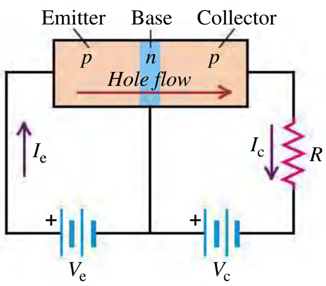
Integrated Circuits

A further refinement in semiconductor technology is the *integrated circuit*. By successively depositing layers of material and etching patterns to define current paths, we can combine the functions of several MOSFETs, capacitors, and resistors on a single square of semiconductor material that may be only a few millimeters on a side. An elaboration of this idea leads to *large-scale integrated circuits*. The resulting integrated circuit chips are the heart of all pocket calculators and present-day computers, large and small (**Fig. 42.36**).

The first semiconductor devices were invented in 1947. Since then, they have completely revolutionized the electronics industry through miniaturization, reliability, speed, energy usage, and cost. They have found applications in communications, computer systems, control systems, and many other areas. In transforming these areas, they have changed, and continue to change, human civilization itself.

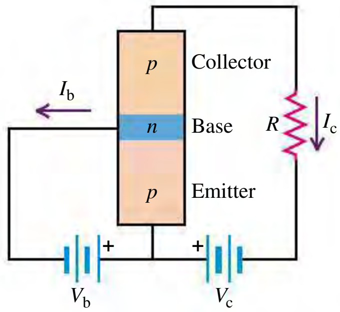
TEST YOUR UNDERSTANDING OF SECTION 42.7 Suppose a negative charge is placed on the gate of the MOSFET shown in Fig. 42.35. Will a substantial current flow between the source and the drain? ❙

42.33 Schematic diagram of a p-n-p transistor and circuit.



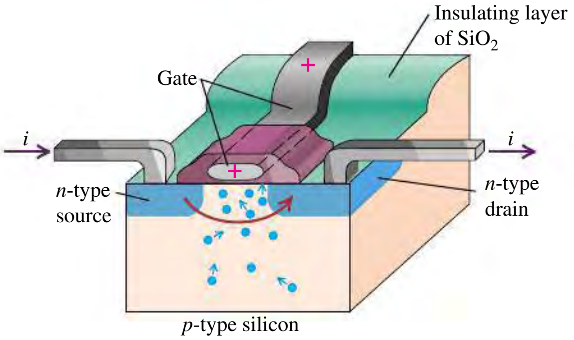
- When $V_e = 0$, the current is very small.
- When a potential V_e is applied between emitter and base, holes travel from the emitter to the base.
- When V_c is sufficiently large, most of the holes continue into the collector.

42.34 A common-emitter circuit.



- When $V_b = 0$, I_c is very small, and most of the voltage V_c appears across the base–collector junction.
- As V_b increases, the base–collector potential decreases, and more holes can diffuse into the collector; thus, I_c increases. Ordinarily, I_c is much larger than I_b.

42.35 A field-effect transistor. The current from source to drain is controlled by the potential difference between the source and the drain and by the charge on the gate; no current flows through the gate.



42.36 An integrated circuit chip smaller than your fingertip can contain millions of transistors.

42.8 SUPERCONDUCTIVITY

Superconductivity is the complete disappearance of all electrical resistance at low temperatures. We described this property at the end of Section 25.2 and the magnetic properties of type-I and type-II superconductors in Section 29.8. In this section we'll relate superconductivity to the structure and energy-band model of a solid.

Although superconductivity was discovered in 1911, it was not well understood on a theoretical basis until 1957. That year, the American physicists John Bardeen, Leon Cooper, and Robert Schrieffer published the theory of superconductivity, now called the BCS theory, that earned them the Nobel Prize in physics in 1972. (It was Bardeen's second; he shared his first for his work on the development of the transistor.) The key to the BCS theory is an interaction between *pairs* of conduction electrons, called *Cooper pairs*, caused by an interaction with the positive ions of the crystal. Here's a rough picture of what happens. A free electron exerts attractive forces on nearby positive ions, pulling them slightly closer together. The resulting slight concentration of positive charge then exerts an attractive force on another free electron with momentum opposite to the first. At ordinary temperatures this electron-pair interaction is very small in comparison to energies of thermal motion, but at very low temperatures it is significant.

Bound together this way, the pairs of electrons cannot *individually* gain or lose very small amounts of energy, as they would ordinarily be able to do in a partly filled conduction band. Their pairing gives an energy gap in the allowed electron quantum levels, and at low temperatures there is not enough collision energy to jump this gap. Therefore the electrons can move freely through the crystal without any energy exchange through collisions—that is, with zero resistance.

Since 1987 physicists have discovered a number of compounds that remain superconducting at temperatures above 77 K (the boiling point of liquid nitrogen). The original pairing mechanism of the BCS theory cannot explain the properties of these *high-temperature superconductors*. Instead, it appears that electrons in these materials form pairs due to magnetic interactions between their spins.

CHAPTER 42 SUMMARY

SOLUTIONS TO ALL EXAMPLES

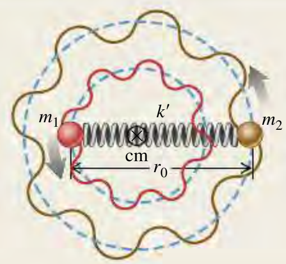
Molecular bonds and molecular spectra: The principal types of molecular bonds are ionic, covalent, van der Waals, and hydrogen bonds. In a diatomic molecule the rotational energy levels are given by Eq. (42.3), where I is the moment of inertia of the molecule, m_r is its reduced mass, and r_0 is the distance between the two atoms. The vibrational energy levels are given by Eq. (42.7), where k' is the effective force constant of the interatomic force. (See Examples 42.1–42.3.)

$$E_l = l(l+1)\frac{\hbar^2}{2I} \quad (l = 0, 1, 2, \ldots) \tag{42.3}$$

$$I = m_r r_0^2 \tag{42.6}$$

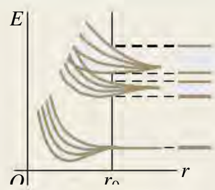
$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{42.4}$$

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\omega = \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m_r}} \quad (n = 0, 1, 2, \ldots) \tag{42.7}$$



Solids and energy bands: Interatomic bonds in solids are of the same types as in molecules plus one additional type, the metallic bond. Associating the basis with each lattice point gives the crystal structure. (See Example 42.4.)

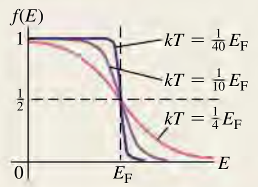
When atoms are bound together in condensed matter, their outer energy levels spread out into bands. At absolute zero, insulators and conductors have a completely filled valence band separated by an energy gap from an empty conduction band. Conductors, including metals, have partially filled conduction bands. (See Example 42.5.)



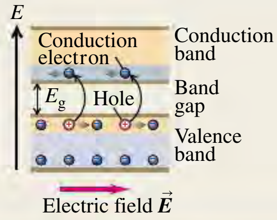
Free-electron model of metals: In the free-electron model of the behavior of conductors, the electrons are treated as completely free particles within the conductor. In this model the density of states is given by Eq. (42.15). The probability that an energy state of energy E is occupied is given by the Fermi–Dirac distribution, Eq. (42.16), which is a consequence of the exclusion principle. In Eq. (42.16), E_F is the Fermi energy. (See Examples 42.6–42.8.)

$$g(E) = \frac{(2m)^{3/2}V}{2\pi^2\hbar^3}E^{1/2} \tag{42.15}$$

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1} \tag{42.16}$$

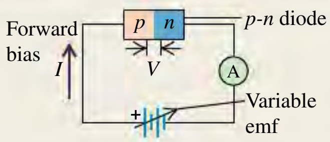


Semiconductors: A semiconductor has an energy gap of about 1 eV between its valence and conduction bands. Its electrical properties may be drastically changed by the addition of small concentrations of donor impurities, giving an n-type semiconductor, or acceptor impurities, giving a p-type semiconductor. (See Example 42.9.)



Semiconductor devices: Many semiconductor devices, including diodes, transistors, and integrated circuits, use one or more p-n junctions. The current–voltage relationship for an ideal p-n junction diode is given by Eq. (42.22).

$$I = I_S(e^{eV/kT} - 1) \tag{42.22}$$



BRIDGING PROBLEM MOLECULAR VIBRATION AND SEMICONDUCTOR BAND GAP

SOLUTION

At 80 K, the band gap in the semiconductor indium antimonide (InSb) is 0.230 eV. A photon emitted by a hydrogen fluoride (HF) molecule undergoing a vibration-rotation transition from $(n = 1, l = 0)$ to $(n = 0, l = 1)$ is absorbed by an electron at the top of the valence band of InSb. (a) How far above the top of the band gap (in eV) is the final state of the electron? (b) What is the probability that the final state was already occupied? The vibration frequency for HF is 1.24×10^{14} Hz, the mass of a hydrogen atom is 1.67×10^{-27} kg, the mass of a fluorine atom is 3.15×10^{-26} kg, and the equilibrium distance between the two nuclei is 0.092 nm. Assume that the Fermi energy for InSb is in the middle of the gap.

SOLUTION GUIDE

IDENTIFY and SET UP

1. This problem involves what you learned about molecular transitions in Section 42.2, about the Fermi–Dirac distribution in Section 42.5, and about semiconductors in Section 42.6.
2. Equation (42.9) gives the combined vibrational-rotational energy in the initial and final molecular states. The difference between the initial and final molecular energies equals the energy E of the emitted photon, which is in turn equal to the energy gained by the InSb valence electron when it absorbs that photon. The probability that the final state is occupied is given by the Fermi–Dirac distribution, Eq. (42.16).

EXECUTE

3. Before you can use Eq. (42.9), you'll first need to use the data given to calculate the moment of inertia I and the quantity $\hbar\omega$ for the HF molecule. (*Hint:* Be careful not to confuse frequency f and angular frequency ω.)
4. Use your results from step 3 to calculate the initial and final energies of the HF molecule. (*Hint:* Does the vibrational energy increase or decrease? What about the rotational energy?)
5. Use your result from step 4 to find the energy imparted to the InSb electron. Determine the final energy of this electron relative to the bottom of the conduction band.
6. Use your result from step 5 to determine the probability that the InSb final state is already occupied.

EVALUATE

7. Is the molecular transition of the HF molecule allowed? Which is larger: the vibrational energy change or the rotational energy change?
8. Is it likely that the excited InSb electron will be blocked from entering a state in the conduction band?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q42.1 Van der Waals bonds occur in many molecules, but hydrogen bonds occur only with materials that contain hydrogen. Why is this type of bond unique to hydrogen?

Q42.2 The bonding of gallium arsenide (GaAs) is said to be 31% ionic and 69% covalent. Explain.

Q42.3 The $H_2{}^+$ molecule consists of two hydrogen nuclei and a single electron. What kind of molecular bond do you think holds this molecule together? Explain.

Q42.4 The moment of inertia for an axis through the center of mass of a diatomic molecule calculated from the wavelength emitted in an $l = 19 \rightarrow l = 18$ transition is different from the moment of inertia calculated from the wavelength of the photon emitted in an $l = 1 \rightarrow l = 0$ transition. Explain this difference. Which transition corresponds to the larger moment of inertia?

Q42.5 Analysis of the photon absorption spectrum of a diatomic molecule shows that the vibrational energy levels for small values of n are very nearly equally spaced but the levels for large n are not equally spaced. Discuss the reason for this observation. Do you expect the adjacent levels to move closer together or farther apart as n increases? Explain.

Q42.6 Discuss the differences between the rotational and vibrational energy levels of the deuterium ("heavy hydrogen") molecule D_2 and those of the ordinary hydrogen molecule H_2. A deuterium atom has twice the mass of an ordinary hydrogen atom.

Q42.7 Various organic molecules have been discovered in interstellar space. Why were these discoveries made with radio telescopes rather than optical telescopes?

Q42.8 The air you are breathing contains primarily nitrogen (N_2) and oxygen (O_2). Many of these molecules are in excited rotational energy levels ($l = 1, 2, 3, \ldots$), but almost all of them are in the vibrational ground level ($n = 0$). Explain this difference between the rotational and vibrational behaviors of the molecules.

Q42.9 In what ways do atoms in a diatomic molecule behave as though they were held together by a spring? In what ways is this a poor description of the interaction between the atoms?

Q42.10 Individual atoms have discrete energy levels, but certain solids (which are made up of only individual atoms) show energy bands and gaps. What causes the solids to behave so differently from the atoms of which they are composed?

Q42.11 What factors determine whether a material is a conductor of electricity or an insulator? Explain.

Q42.12 Ionic crystals are often transparent, whereas metallic crystals are always opaque. Why?

Q42.13 Speeds of molecules in a gas vary with temperature, whereas speeds of electrons in the conduction band of a metal are nearly independent of temperature. Why are these behaviors so different?

Q42.14 Use the band model to explain how it is possible for some materials to undergo a semiconductor-to-metal transition as the temperature or pressure varies.

Q42.15 An isolated zinc atom has a ground-state electron configuration of filled 1s, 2s, 2p, 3s, 3p, and 4s subshells. How can zinc be a conductor if its valence subshell is full?

Q42.16 The assumptions of the *free-electron model* of metals may seem contrary to reason, since electrons exert powerful electric forces on each other. Give some reasons why these assumptions actually make physical sense.

Q42.17 Why are materials that are good thermal conductors also good electrical conductors? What kinds of problems does this pose for the design of appliances such as clothes irons and electric heaters? Are there materials that do not follow this general rule?

Q42.18 What is the essential characteristic for an element to serve as a donor impurity in a semiconductor such as Si or Ge? For it to serve as an acceptor impurity? Explain.

Q42.19 There are several methods for removing electrons from the surface of a semiconductor. Can holes be removed from the surface? Explain.

Q42.20 A student asserts that silicon and germanium become good insulators at very low temperatures and good conductors at very high temperatures. Do you agree? Explain your reasoning.

Q42.21 The electrical conductivities of most metals decrease gradually with increasing temperature, but the intrinsic conductivity of semiconductors always *increases* rapidly with increasing temperature. What causes the difference?

Q42.22 How could you make compensated silicon that has twice as many acceptors as donors?

Q42.23 The saturation current I_S for a p-n junction, Eq. (42.22), depends strongly on temperature. Explain why.

Q42.24 Why does tunneling limit the miniaturization of MOSFETs?

EXERCISES

Section 42.1 Types of Molecular Bonds

42.1 • If the energy of the H_2 covalent bond is -4.48 eV, what wavelength of light is needed to break that molecule apart? In what part of the electromagnetic spectrum does this light lie?

42.2 • **An Ionic Bond.** (a) Calculate the electric potential energy for a K^+ ion and a Br^- ion separated by a distance of 0.29 nm, the equilibrium separation in the KBr molecule. Treat the ions as point charges. (b) The ionization energy of the potassium atom is 4.3 eV. Atomic bromine has an electron affinity of 3.5 eV. Use these data and the results of part (a) to estimate the binding energy of the KBr molecule. Do you expect the actual binding energy to be higher or lower than your estimate? Explain your reasoning.

42.3 • For the H_2 molecule the equilibrium spacing of the two protons is 0.074 nm. The mass of a hydrogen atom is 1.67×10^{-27} kg. Calculate the wavelength of the photon emitted in the rotational transition $l = 2$ to $l = 1$.

42.4 • During each of these processes, a photon of light is given up. In each process, what wavelength of light is given up, and in what part of the electromagnetic spectrum is that wavelength? (a) A molecule decreases its vibrational energy by 0.198 eV; (b) an atom decreases its energy by 7.80 eV; (c) a molecule decreases its rotational energy by 4.80×10^{-3} eV.

Section 42.2 Molecular Spectra

42.5 • A hypothetical NH molecule makes a rotational-level transition from $l = 3$ to $l = 1$ and gives off a photon of wavelength 1.780 nm in doing so. What is the separation between the two atoms in this molecule if we model them as point masses? The mass of hydrogen is 1.67×10^{-27} kg, and the mass of nitrogen is 2.33×10^{-26} kg.

42.6 •• The H_2 molecule has a moment of inertia of 4.6×10^{-48} kg · m^2. What is the wavelength λ of the photon absorbed when H_2 makes a transition from the $l = 3$ to the $l = 4$ rotational level?

42.7 • The water molecule has an $l = 1$ rotational level 1.01×10^{-5} eV above the $l = 0$ ground level. Calculate the wavelength and frequency of the photon absorbed by water when it undergoes a rotational-level transition from $l = 0$ to $l = 1$. The magnetron oscillator in a microwave oven generates microwaves with a frequency of 2450 MHz. Does this make sense, in view of the frequency you calculated in this problem? Explain.

42.8 • Two atoms of cesium (Cs) can form a Cs_2 molecule. The equilibrium distance between the nuclei in a Cs_2 molecule is 0.447 nm. Calculate the moment of inertia about an axis through the center of mass of the two nuclei and perpendicular to the line joining them. The mass of a cesium atom is 2.21×10^{-25} kg.

42.9 •• **CP** The rotational energy levels of CO are calculated in Example 42.2. If the energy of the rotating molecule is described by the classical expression $K = \frac{1}{2}I\omega^2$, for the $l = 1$ level what are (a) the angular speed of the rotating molecule; (b) the linear speed of each atom; (c) the rotational period (the time for one rotation)?

42.10 •• The average kinetic energy of an ideal-gas atom or molecule is $\frac{3}{2}kT$, where T is the Kelvin temperature (Chapter 18). The rotational inertia of the H_2 molecule is 4.6×10^{-48} kg · m^2. What is the value of T for which $\frac{3}{2}kT$ equals the energy separation between the $l = 0$ and $l = 1$ energy levels of H_2? What does this tell you about the number of H_2 molecules in the $l = 1$ level at room temperature?

42.11 • A lithium atom has mass 1.17×10^{-26} kg, and a hydrogen atom has mass 1.67×10^{-27} kg. The equilibrium separation between the two nuclei in the LiH molecule is 0.159 nm. (a) What is the difference in energy between the $l = 3$ and $l = 4$ rotational levels? (b) What is the wavelength of the photon emitted in a transition from the $l = 4$ to the $l = 3$ level?

42.12 • If a sodium chloride (NaCl) molecule could undergo an $n \rightarrow n - 1$ vibrational transition with no change in rotational quantum number, a photon with wavelength 20.0 μm would be emitted. The mass of a sodium atom is 3.82×10^{-26} kg, and the mass of a chlorine atom is 5.81×10^{-26} kg. Calculate the force constant k' for the interatomic force in NaCl.

42.13 •• When a hypothetical diatomic molecule having atoms 0.8860 nm apart undergoes a rotational transition from the $l = 2$ state to the next lower state, it gives up a photon having energy 8.841×10^{-4} eV. When the molecule undergoes a vibrational transition from one energy state to the next lower energy state, it gives up 0.2560 eV. Find the force constant of this molecule.

42.14 • The vibrational and rotational energies of the CO molecule are given by Eq. (42.9). Calculate the wavelength of the photon absorbed by CO in each of these vibration-rotation transitions: (a) $n = 0,\ l = 2 \rightarrow n = 1,\ l = 3$; (b) $n = 0,\ l = 3 \rightarrow n = 1,\ l = 2$; (c) $n = 0,\ l = 4 \rightarrow n = 1,\ l = 3$.

Section 42.3 Structure of Solids

42.15 • **Density of NaCl.** The spacing of adjacent atoms in a crystal of sodium chloride is 0.282 nm. The mass of a sodium atom is 3.82×10^{-26} kg, and the mass of a chlorine atom is 5.89×10^{-26} kg. Calculate the density of sodium chloride.

42.16 • Potassium bromide (KBr) has a density of 2.75×10^3 kg/m^3 and the same crystal structure as NaCl. The mass of a potassium atom is 6.49×10^{-26} kg, and the mass of a bromine atom is 1.33×10^{-25} kg. (a) Calculate the average spacing between adjacent atoms in a KBr crystal. (b) How does the value calculated in part (a) compare with the spacing in NaCl (see Exercise 42.15)? Is the relationship between the two values qualitatively what you would expect? Explain.

Section 42.4 Energy Bands

42.17 • The maximum wavelength of light that a certain silicon photocell can detect is 1.11 μm. (a) What is the energy gap (in electron volts) between the valence and conduction bands for this photocell? (b) Explain why pure silicon is opaque.

42.18 • The gap between valence and conduction bands in diamond is 5.47 eV. (a) What is the maximum wavelength of a photon that can excite an electron from the top of the valence band into the conduction band? In what region of the electromagnetic spectrum does this photon lie? (b) Explain why pure diamond is transparent and colorless. (c) Most gem diamonds have a yellow color. Explain how impurities in the diamond can cause this color.

42.19 • The gap between valence and conduction bands in silicon is 1.12 eV. A nickel nucleus in an excited state emits a gamma-ray photon with wavelength 9.31×10^{-4} nm. How many electrons can be excited from the top of the valence band to the bottom of the conduction band by the absorption of this gamma ray?

Section 42.5 Free-Electron Model of Metals

42.20 • Calculate v_{rms} for free electrons with average kinetic energy $\frac{3}{2}kT$ at a temperature of 300 K. How does your result compare to the speed of an electron with a kinetic energy equal to the Fermi energy of copper, calculated in Example 42.7? Why is there such a difference between these speeds?

42.21 • Calculate the density of states $g(E)$ for the free-electron model of a metal if $E = 7.0$ eV and $V = 1.0$ cm^3. Express your answer in units of states per electron volt.

42.22 • The Fermi energy of sodium is 3.23 eV. (a) Find the average energy E_{av} of the electrons at absolute zero. (b) What is the speed of an electron that has energy E_{av}? (c) At what Kelvin temperature T is kT equal to E_F? (This is called the *Fermi temperature* for the metal. It is approximately the temperature at which molecules in a classical ideal gas would have the same kinetic energy as the fastest-moving electron in the metal.)

42.23 • **CP** Silver has a Fermi energy of 5.48 eV. Calculate the electron contribution to the molar heat capacity at constant volume of silver, C_V, at 300 K. Express your result (a) as a multiple of R and (b) as a fraction of the actual value for silver, $C_V = 25.3$ J/mol · K. (c) Is the value of C_V due principally to the electrons? If not, to what is it due? (*Hint:* See Section 18.4.)

42.24 •• At the Fermi temperature T_F, $E_F = kT_F$ (see Exercise 42.22). When $T = T_F$, what is the probability that a state with energy $E = 2E_F$ is occupied?

42.25 •• For a solid metal having a Fermi energy of 8.500 eV, what is the probability, at room temperature, that a state having an energy of 8.520 eV is occupied by an electron?

Section 42.6 Semiconductors

42.26 • Pure germanium has a band gap of 0.67 eV. The Fermi energy is in the middle of the gap. (a) For temperatures of 250 K, 300 K, and 350 K, calculate the probability $f(E)$ that a state at the bottom of the conduction band is occupied. (b) For each temperature in part (a), calculate the probability that a state at the top of the valence band is empty.

42.27 • Germanium has a band gap of 0.67 eV. Doping with arsenic adds donor levels in the gap 0.01 eV below the bottom of the conduction band. At a temperature of 300 K, the probability is 4.4×10^{-4} that an electron state is occupied at the bottom of the conduction band. Where is the Fermi level relative to the conduction band in this case?

Section 42.7 Semiconductor Devices

42.28 •• (a) Suppose a piece of very pure germanium is to be used as a light detector by observing, through the absorption of photons, the increase in conductivity resulting from generation of electron–hole pairs. If each pair requires 0.67 eV of energy, what is the maximum wavelength that can be detected? In what portion of the spectrum does it lie? (b) What are the answers to part (a) if the material is silicon, with an energy requirement of 1.12 eV per pair, corresponding to the gap between valence and conduction bands in that element?

42.29 • **CP** At a temperature of 290 K, a certain *p-n* junction has a saturation current $I_S = 0.500$ mA. (a) Find the current at this temperature when the voltage is (i) 1.00 mV, (ii) -1.00 mV, (iii) 100 mV, and (iv) -100 mV. (b) Is there a region of applied voltage where the diode obeys Ohm's law?

42.30 • For a certain *p-n* junction diode, the saturation current at room temperature $(20°C)$ is 0.950 mA. What is the resistance of this diode when the voltage across it is (a) 85.0 mV and (b) -50.0 mV?

42.31 •• (a) A forward-bias voltage of 15.0 mV produces a positive current of 9.25 mA through a *p-n* junction at 300 K. What does the positive current become if the forward-bias voltage is reduced to 10.0 mV? (b) For reverse-bias voltages of -15.0 mV and -10.0 mV, what is the reverse-bias negative current?

42.32 •• A *p-n* junction has a saturation current of 6.40 mA. (a) At a temperature of 300 K, what voltage is needed to produce a positive current of 40.0 mA? (b) For a voltage equal to the negative of the value calculated in part (a), what is the negative current?

PROBLEMS

42.33 •• A hypothetical diatomic molecule of oxygen (mass $= 2.656 \times 10^{-26}$ kg) and hydrogen (mass $= 1.67 \times 10^{-27}$ kg) emits a photon of wavelength 2.39 μm when it makes a transition from one vibrational state to the next lower state. If we model this molecule as two point masses at opposite ends of a massless spring, (a) what is the force constant of this spring, and (b) how many vibrations per second is the molecule making?

42.34 • When a diatomic molecule undergoes a transition from the $l = 2$ to the $l = 1$ rotational state, a photon with wavelength 54.3 μm is emitted. What is the moment of inertia of the molecule for an axis through its center of mass and perpendicular to the line connecting the nuclei?

42.35 •• **CP** (a) The equilibrium separation of the two nuclei in an NaCl molecule is 0.24 nm. If the molecule is modeled as charges $+e$ and $-e$ separated by 0.24 nm, what is the electric dipole moment of the molecule (see Section 21.7)? (b) The measured electric dipole moment of an NaCl molecule is 3.0×10^{-29} C·m. If this dipole moment arises from point charges $+q$ and $-q$ separated by 0.24 nm, what is q? (c) A definition of the *fractional ionic character* of the bond is q/e. If the sodium atom has charge $+e$ and the chlorine atom has charge $-e$, the fractional ionic character would be equal to 1. What is the actual fractional ionic character for the bond in NaCl? (d) The equilibrium distance between nuclei in the hydrogen iodide (HI) molecule is 0.16 nm, and the measured electric dipole moment of the molecule is 1.5×10^{-30} C·m. What is the fractional ionic character for the bond in HI? How does your answer compare to that for NaCl calculated in part (c)? Discuss reasons for the difference in these results.

42.36 • The binding energy of a potassium chloride molecule (KCl) is 4.43 eV. The ionization energy of a potassium atom is 4.3 eV, and the electron affinity of chlorine is 3.6 eV. Use these data to estimate the equilibrium separation between the two atoms in the KCl molecule. Explain why your result is only an estimate and not a precise value.

42.37 • (a) For the sodium chloride molecule (NaCl) discussed at the beginning of Section 42.1, what is the maximum separation of the ions for stability if they may be regarded as point charges? That is, what is the largest separation for which the energy of an Na^+ ion and a Cl^- ion, calculated in this model, is lower than the energy of the two separate atoms Na and Cl? (b) Calculate this distance for the potassium bromide molecule, described in Exercise 42.2.

42.38 • When a NaF molecule makes a transition from the $l = 3$ to the $l = 2$ rotational level with no change in vibrational quantum number or electronic state, a photon with wavelength 3.83 mm is emitted. A sodium atom has mass 3.82×10^{-26} kg, and a fluorine atom has mass 3.15×10^{-26} kg. Calculate the equilibrium separation between the nuclei in a NaF molecule. How does your answer compare with the value for NaCl given in Section 42.1? Is this result reasonable? Explain.

42.39 •• **CP** Consider a gas of diatomic molecules (moment of inertia I) at an absolute temperature T. If E_g is a ground-state energy and E_{ex} is the energy of an excited state, then the Maxwell–Boltzmann distribution (see Section 39.4) predicts that the ratio of the numbers of molecules in the two states is $n_{ex}/n_g = e^{-(E_{ex}-E_g)/kT}$. (a) Explain why the ratio of the number of molecules in the lth rotational energy *level* to the number of molecules in the ground-state $(l = 0)$ rotational level is

$$\frac{n_l}{n_0} = (2l + 1)e^{-[l(l+1)\hbar^2]/2IkT}$$

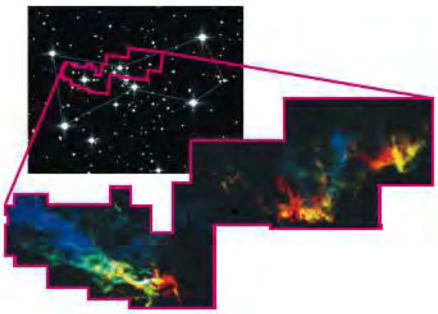
(*Hint:* For each value of l, how many states are there with different values of m_l?) (b) Determine the ratio n_l/n_0 for a gas of CO molecules at 300 K for (i) $l = 1$; (ii) $l = 2$; (iii) $l = 10$; (iv) $l = 20$; (v) $l = 50$. The moment of inertia of the CO molecule is given in Example 42.2 (Section 42.2). (c) Your results in part (b) show that as l is increased, the ratio n_l/n_0 first increases and then decreases. Explain why.

42.40 ••• **CALC** Part (a) of Problem 42.39 gives an equation for the number of diatomic molecules in the lth rotational level to the number in the ground-state rotational level. (a) Derive an expression for the value of l for which this ratio is the largest. (b) For the CO molecule at $T = 300$ K, for what value of l is this ratio a maximum? (The moment of inertia of the CO molecule is given in Example 42.2.)

42.41 • **Spectral Lines from Isotopes.** The equilibrium separation for NaCl is 0.2361 nm. The mass of a sodium atom is 3.8176×10^{-26} kg. Chlorine has two stable isotopes, ^{35}Cl and ^{37}Cl, that have different masses but identical chemical properties. The atomic mass of ^{35}Cl is 5.8068×10^{-26} kg, and the atomic mass of ^{37}Cl is 6.1384×10^{-26} kg. (a) Calculate the wavelength of the photon emitted in the $l = 2 \rightarrow l = 1$ and $l = 1 \rightarrow l = 0$ transitions for $Na^{35}Cl$. (b) Repeat part (a) for $Na^{37}Cl$. What are the differences in the wavelengths for the two isotopes?

42.42 •• Our galaxy contains numerous *molecular clouds*, regions many light-years in extent in which the density is high enough and the temperature low enough for atoms to form into molecules. Most of the molecules are H_2, but a small fraction of the molecules are carbon monoxide (CO). Such a molecular cloud in the constellation Orion is shown in **Fig. P42.42**. The upper image was made with an ordinary visible-light telescope; the lower image shows the molecular cloud in Orion as imaged with a radio telescope tuned to a wavelength emitted by CO in a rotational transition. The different colors in the radio image indicate regions of the cloud that are moving either toward us (blue) or away from us (red) relative to the motion of the cloud as a whole, as determined by the Doppler shift of the radiation. (Since a molecular cloud has about 10,000 hydrogen molecules for each CO molecule, it might seem more reasonable to tune a radio telescope to emissions from H_2 than to emissions from CO. Unfortunately, it turns out that the H_2 molecules in molecular clouds do not radiate in either the radio or visible portions of the electromagnetic spectrum.) (a) Using the data in Example 42.2 (Section 42.2), calculate the energy and wavelength of the photon emitted by a CO molecule in an $l = 1 \rightarrow l = 0$ rotational transition. (b) As a rule, molecules in a gas at temperature T will be found in a certain excited rotational energy level, provided the energy of that level is no higher than kT (see Problem 42.39). Use this rule to explain why astronomers can detect radiation from CO in molecular clouds even though the typical temperature of a molecular cloud is a very low 20 K.

Figure **P42.42**



42.43 • The force constant for the internuclear force in a hydrogen molecule (H_2) is $k' = 576$ N/m. A hydrogen atom has mass 1.67×10^{-27} kg. Calculate the zero-point vibrational energy for H_2 (that is, the vibrational energy the molecule has in the $n = 0$ ground vibrational level). How does this energy compare in magnitude with the H_2 bond energy of -4.48 eV?

42.44 • When an OH molecule undergoes a transition from the $n = 0$ to the $n = 1$ vibrational level, its internal vibrational energy increases by 0.463 eV. Calculate the frequency of vibration and the force constant for the interatomic force. (The mass of an oxygen atom is 2.66×10^{-26} kg, and the mass of a hydrogen atom is 1.67×10^{-27} kg.)

42.45 • The hydrogen iodide (HI) molecule has equilibrium separation 0.160 nm and vibrational frequency 6.93×10^{13} Hz. The mass of a hydrogen atom is 1.67×10^{-27} kg, and the mass of an iodine atom is 2.11×10^{-25} kg. (a) Calculate the moment of inertia of HI about a perpendicular axis through its center of mass. (b) Calculate the wavelength of the photon emitted in each of the following vibration-rotation transitions: (i) $n = 1$, $l = 1 \rightarrow n = 0$, $l = 0$; (ii) $n = 1$, $l = 2 \rightarrow n = 0$, $l = 1$; (iii) $n = 2$, $l = 2 \rightarrow n = 1$, $l = 3$.

42.46 • Suppose the hydrogen atom in HF (see the Bridging Problem for this chapter) is replaced by an atom of deuterium, an isotope of hydrogen with a mass of 3.34×10^{-27} kg. The force constant is determined by the electron configuration, so it is the same as for the normal HF molecule. (a) What is the vibrational frequency of this molecule? (b) What wavelength of light corresponds to the energy difference between the $n = 1$ and $n = 0$ levels? In what region of the spectrum does this wavelength lie?

42.47 •• Compute the Fermi energy of potassium by making the simple approximation that each atom contributes one free electron. The density of potassium is 851 kg/m^3, and the mass of a single potassium atom is 6.49×10^{-26} kg.

42.48 •• **CALC** The one-dimensional calculation of Example 42.4 (Section 42.3) can be extended to three dimensions. For the three-dimensional fcc NaCl lattice, the result for the potential energy of a pair of Na^+ and Cl^- ions due to the electrostatic interaction with all of the ions in the crystal is $U = -\alpha e^2/4\pi\epsilon_0 r$, where $\alpha = 1.75$ is the *Madelung constant*. Another contribution to the potential energy is a repulsive interaction at small ionic separation r due to overlap of the electron clouds. This contribution can be represented by A/r^8, where A is a positive constant, so the expression for the total potential energy is

$$U_{\text{tot}} = -\frac{\alpha e^2}{4\pi\epsilon_0 r} + \frac{A}{r^8}$$

(a) Let r_0 be the value of the ionic separation r for which U_{tot} is a minimum. Use this definition to find an equation that relates r_0 and A, and use this to write U_{tot} in terms of r_0. For NaCl, $r_0 = 0.281$ nm. Obtain a numerical value (in electron volts) of U_{tot} for NaCl. (b) The quantity $-U_{\text{tot}}$ is the energy required to remove a Na^+ ion and a Cl^- ion from the crystal. Forming a pair of neutral atoms from this pair of ions involves the release of 5.14 eV (the ionization energy of Na) and the expenditure of 3.61 eV (the electron affinity of Cl). Use the result of part (a) to calculate the energy required to remove a pair of neutral Na and Cl atoms from the crystal. The experimental value for this quantity is 6.39 eV; how well does your calculation agree?

42.49 ••• Metallic lithium has a bcc crystal structure. Each unit cell is a cube of side length $a = 0.35$ nm. (a) For a bcc lattice, what is the number of atoms per unit volume? Give your answer in terms of a. (*Hint:* How many atoms are there per unit cell?) (b) Use the result of part (a) to calculate the zero-temperature Fermi energy E_{F0} for metallic lithium. Assume there is one free electron per atom.

42.50 •• **DATA** To determine the equilibrium separation of the atoms in the HCl molecule, you measure the rotational spectrum of HCl. You find that the spectrum contains these wavelengths (among others): 60.4 μm, 69.0 μm, 80.4 μm, 96.4 μm, and 120.4 μm. (a) Use your measured wavelengths to find the moment of inertia of the HCl molecule about an axis through the center of mass and perpendicular to the line joining the two nuclei. (b) The value of l changes by ± 1 in rotational transitions. What value of l for the upper level of the transition gives rise to each of these wavelengths? (c) Use your result of part (a) to calculate the equilibrium separation of the atoms in the HCl molecule. The mass of a chlorine atom is 5.81×10^{-26} kg, and the mass of a hydrogen atom is 1.67×10^{-27} kg. (d) What is the longest-wavelength line in the rotational spectrum of HCl?

42.51 •• DATA The table gives the occupation probabilities $f(E)$ as a function of the energy E for a solid conductor at a fixed temperature T.

$f(E)$	0.064	0.173	0.390	0.661	0.856	0.950
E (eV)	3.0	2.5	2.0	1.5	1.0	0.5

To determine the Fermi energy of the solid material, you are asked to analyze this information in terms of the Fermi–Dirac distribution. (a) Graph the values in the table as E versus $\ln\{[1/f(E)] - 1\}$. Find the slope and y-intercept of the best-fit straight line for the data points when they are plotted this way. (b) Use your results of part (a) to calculate the temperature T and the Fermi energy of the material.

42.52 •• DATA A p-n junction is part of the control mechanism for a wind turbine that is used to generate electricity. The turbine has been malfunctioning, so you are running diagnostics. You can remotely change the bias voltage V applied to the junction and measure the current through the junction. With a forward bias voltage of $+5.00$ mV, the current is $I_f = 0.407$ mA. With a reverse bias voltage of -5.00 mV, the current is $I_r = -0.338$ mA. Assume that Eq. (42.22) accurately represents the current–voltage relationship for the junction, and use these two results to calculate the temperature T and saturation current I_S for the junction. [*Hint:* In your analysis, let $x = e^{eV/kT}$. Apply Eq. (42.22) to each measurement and obtain a quadratic equation for x.]

CHALLENGE PROBLEMS

42.53 •• CALC Consider a system of N free electrons within a volume V. Even at absolute zero, such a system exerts a pressure p on its surroundings due to the motion of the electrons. To calculate this pressure, imagine that the volume increases by a small amount dV. The electrons will do an amount of work $p\,dV$ on their surroundings, which means that the total energy E_{tot} of the electrons will change by an amount $dE_{tot} = -p\,dV$. Hence $p = -dE_{tot}/dV$. (a) Show that the pressure of the electrons at absolute zero is

$$p = \frac{3^{2/3}\pi^{4/3}\hbar^2}{5m}\left(\frac{N}{V}\right)^{5/3}$$

(b) Evaluate this pressure for copper, which has a free-electron concentration of $8.45 \times 10^{28}\ \text{m}^{-3}$. Express your result in pascals and in atmospheres. (c) The pressure you found in part (b) is extremely high. Why, then, don't the electrons in a piece of copper simply explode out of the metal?

42.54 •• CALC When the pressure p on a material increases by an amount Δp, the volume of the material will change from V to $V + \Delta V$, where ΔV is negative. The *bulk modulus B* of the material is defined to be the ratio of the pressure change Δp to the absolute value $|\Delta V/V|$ of the fractional volume change. The greater the bulk modulus, the greater the pressure increase required for a given fractional volume change, and the more incompressible the material (see Section 11.4). Since $\Delta V < 0$, the bulk modulus can be written as $B = -\Delta p/(\Delta V/V_0)$. In the limit that the pressure and volume changes are very small, this becomes

$$B = -V\frac{dp}{dV}$$

(a) Use the result of Problem 42.53 to show that the bulk modulus for a system of N free electrons in a volume V at low temperatures is $B = \frac{5}{3}p$. (*Hint:* The quantity p in the expression $B = -V(dp/dV)$ is the *external* pressure on the system. Can you explain why this is equal to the *internal* pressure of the system itself, as found in Problem 42.53?) (b) Evaluate the bulk modulus for the electrons in copper, which has a free-electron concentration of $8.45 \times 10^{28}\ \text{m}^{-3}$. Express your result in pascals. (c) The actual bulk modulus of copper is 1.4×10^{11} Pa. Based on your result in part (b), what fraction of this is due to the free electrons in copper? (This result shows that the free electrons in a metal play a major role in making the metal resistant to compression.) What do you think is responsible for the remaining fraction of the bulk modulus?

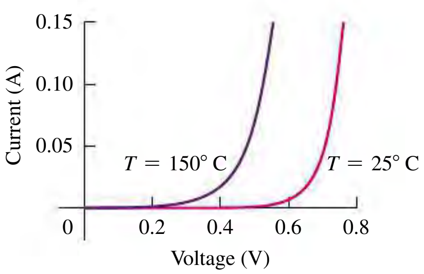
42.55 •• In the discussion of free electrons in Section 42.5, we assumed that we could ignore the effects of relativity. This is not a safe assumption if the Fermi energy is greater than about $\frac{1}{100}mc^2$ (that is, more than about 1% of the rest energy of an electron). (a) Assume that the Fermi energy at absolute zero, as given by Eq. (42.19), is equal to $\frac{1}{100}mc^2$. Show that the electron concentration is

$$\frac{N}{V} = \frac{2^{3/2}m^3c^3}{3000\pi^2\hbar^3}$$

and determine the numerical value of N/V. (b) Is it a good approximation to ignore relativistic effects for electrons in a metal such as copper, for which the electron concentration is $8.45 \times 10^{28}\ \text{m}^{-3}$? Explain. (c) A *white dwarf star* is what is left behind by a star like the sun after it has ceased to produce energy by nuclear reactions. (Our own sun will become a white dwarf star in another 6×10^9 years or so.) A typical white dwarf has mass 2×10^{30} kg (comparable to the sun) and radius 6000 km (comparable to that of the earth). The gravitational attraction of different parts of the white dwarf for each other tends to compress the star; what prevents it from compressing is the pressure of free electrons within the star (see Problem 42.53). Use both of the following assumptions to estimate the electron concentration within a typical white dwarf star: (i) the white dwarf star is made of carbon, which has a mass per atom of 1.99×10^{-26} kg; and (ii) all six of the electrons from each carbon atom are able to move freely throughout the star. (d) Is it a good approximation to ignore relativistic effects in the structure of a white dwarf star? Explain.

PASSAGE PROBLEMS

DIODE TEMPERATURE SENSOR. The current–voltage characteristics of a forward-biased p-n junction diode depend strongly on temperature, as shown in the figure. As a result, diodes can be used as temperature sensors. In actual operation, the voltage is adjusted to keep the current through the diode constant at a specified value, such as 100 mA, and the temperature is determined from a measurement of the voltage at that current.



42.56 The sensitivity of a diode thermometer depends on how much the voltage changes for a given temperature change, with the current remaining constant. What is the sensitivity for this diode thermometer, operated at 100 mA, for a temperature change from 25°C to 150°C? (a) $+0.2$ mV/°C; (b) $+2.0$ mV/°C; (c) -0.2 mV/°C; (d) -2.0 mV/°C.

42.57 Which statement best explains the temperature dependence of the current–voltage characteristics that the graph shows? At higher temperatures: (a) The band gap is larger, so the electron–hole pairs have more energy, which causes the current at a given voltage to be larger. (b) More electrons can move to the conduction band, which causes the current at a given voltage to be larger. (c) All of the electrons in the valence band move to the conduction band, and the diode behaves like a metal and follows Ohm's law. (d) The acceptor and donor impurity atoms are free to move through the material, which causes the current at a given voltage to be larger.

42.58 If the voltage rather than the current is kept constant, what happens as the temperature increases from 25°C to 150°C? (a) At first the current increases, then it decreases. (b) The current increases. (c) The current decreases, eventually approaching zero. (d) The current does not change unless the voltage also changes.

Answers

Chapter Opening Question ?

(i) Venus must radiate energy into space at the same rate that it receives energy in the form of sunlight. However, carbon dioxide (CO_2) molecules in the atmosphere absorb infrared radiation emitted by the surface of Venus and re-emit it toward the ground. This involves a transition between vibrational states of the CO_2 molecule (see Section 42.2). To compensate for this and to maintain the balance between emitted and received energy, the surface temperature of Venus and hence the rate of blackbody radiation from the surface increase.

Test Your Understanding Questions

42.1 (i) The exclusion principle states that only one electron can be in a given state. Real electrons have spin, so two electrons (one spin up, one spin down) can be in a given *spatial* state and hence two can participate in a given covalent bond between two atoms. If electrons obeyed the exclusion principle but did not have spin, that state of an electron would be completely described by its spatial distribution and only *one* electron could participate in a covalent bond. (We will learn in Chapter 44 that this situation is wholly imaginary: There are subatomic particles without spin, but they do *not* obey the exclusion principle.)

42.2 (ii) Figure 42.5 shows that the difference in energy between adjacent rotational levels increases with increasing l. Hence, as l increases, the energy E of the emitted photon increases and the wavelength $\lambda = hc/E$ decreases.

42.3 (ii) In Fig. 42.13 let a be the distance between adjacent Na^+ and Cl^- ions. This figure shows that the Cl^- ion that is the next nearest neighbor to a Na^+ ion is on the opposite corner of a cube of side a. The distance between these two ions is $\sqrt{a^2 + a^2 + a^2} = \sqrt{3a^2} = a\sqrt{3}$.

42.4 (ii) A small temperature change causes a substantial increase in the population of electrons in a semiconductor's conduction band and a comparably substantial increase in conductivity. The conductivity of conductors and insulators varies more gradually with temperature.

42.5 no The kinetic-molecular model of an ideal gas (Section 18.3) shows that the gas pressure is proportional to the average translational kinetic energy E_{av} of the particles that make up the gas. In a classical ideal gas, E_{av} is directly proportional to the average temperature T, so the pressure decreases as T decreases. In a free-electron gas, the average kinetic energy per electron is *not* related simply to T; as Example 42.8 shows, for the free-electron gas in a metal, E_{av} is almost completely a consequence of the exclusion principle at room temperature and colder. Hence the pressure of a free-electron gas in a solid metal does *not* change appreciably between room temperature and absolute zero.

42.6 no Pure copper is already an excellent conductor since it has a partially filled conduction band (Fig. 42.19c). Furthermore, copper forms a metallic crystal (Fig. 42.15) as opposed to the covalent crystals of silicon or germanium, so the scheme of using an impurity to donate or accept an electron does not work for copper. In fact, adding impurities to copper *decreases* the conductivity because an impurity tends to scatter electrons, impeding the flow of current.

42.7 no A negative charge on the gate will repel, not attract, electrons in the p-type silicon. Hence the electron concentration in the region between the two p-n junctions will be made even smaller. With so few charge carriers present in this region, very little current will flow between the source and the drain.

Bridging Problem

(a) 0.278 eV

(b) 1.74×10^{-25}

? This sculpture of a wooly mammoth, just 3.7 cm (1.5 in.) in length, was carved from a mammoth's ivory tusk by an artist who lived in southwestern Germany 35,000 years ago. It is possible to date biological specimens such as this because (i) ancient materials were more radioactive than modern ones; (ii) ancient materials were less radioactive than modern ones; (iii) biological specimens continue to take in radioactive substances after they die; (iv) biological specimens no longer take in radioactive substances after they die; (v) more than one of these.

43 NUCLEAR PHYSICS

LEARNING GOALS

Looking forward at ...

43.1 Some key properties of atomic nuclei, including radii, densities, spins, and magnetic moments.

43.2 How the binding energy of a nucleus depends on the numbers of protons and neutrons that it contains.

43.3 The most important ways in which unstable nuclei undergo radioactive decay.

43.4 How the decay rate of a radioactive substance depends on time.

43.5 Some of the biological hazards and medical uses of radiation.

43.6 How to analyze some important types of nuclear reactions.

43.7 What happens in a nuclear fission chain reaction, and how it can be controlled.

43.8 The nuclear reactions that allow the sun to shine.

Looking back at ...

5.5 The strong interaction.

18.3 Kinetic energy of gas molecules.

21.1 Proton and neutron.

26.4 Discharging capacitor.

37.8 Rest mass and rest energy.

39.2 Discovery of the nucleus.

40.3, 40.4 Square-well potential; tunneling.

41.4–41.6 Magnetic moments; spin-$\frac{1}{2}$ particles; central-field approximation.

Every atom contains at its center an extremely dense, positively charged *nucleus,* which is much smaller than the overall size of the atom but contains most of its total mass. In this chapter we'll look at several important general properties of nuclei and of the nuclear force that holds protons and neutrons together within a nucleus. The stability or instability of a particular nucleus is determined by the competition between the attractive nuclear force among the protons and neutrons and the repulsive electrical interactions among the protons. Unstable nuclei *decay,* transforming themselves spontaneously into other nuclei by a variety of processes. Nuclear reactions can also be induced by impact on a nucleus of a particle or another nucleus. Two classes of reactions of special interest are *fission* and *fusion*. Fission is the process that takes place within a nuclear reactor used for generating power. We could not survive without the energy released by one nearby fusion reactor, our sun.

43.1 PROPERTIES OF NUCLEI

As we described in Section 39.2, Rutherford found that the nucleus is tens of thousands of times smaller in radius than the atom itself. Since Rutherford's initial experiments, many additional scattering experiments have been performed with high-energy protons, electrons, neutrons, and alpha particles. These experiments show that we can model a nucleus as a sphere with a radius R that depends on the total number of *nucleons* (neutrons and protons) in the nucleus. This number is called the **nucleon number** A. The radii of most nuclei are represented quite well by the equation

Experimentally determined constant $= 1.2 \times 10^{-15}$ m $= 1.2$ fm

Radius of an atomic nucleus $$R = R_0 A^{1/3} \quad (43.1)$$ Nucleon number = total number of protons and neutrons

The nucleon number A in Eq. (43.1) is also called the **mass number** because it is the nearest whole number to the mass of the nucleus measured in unified atomic mass units (u). (The proton mass and the neutron mass are both approximately 1 u.) The best current conversion factor is

$$1\text{ u} = 1.660538921(73) \times 10^{-27}\text{ kg}$$

In Section 43.2 we'll discuss the masses of nuclei in more detail. Note that when we speak of the masses of nuclei and particles, we mean their *rest* masses.

Nuclear Density

The volume V of a sphere is equal to $4\pi R^3/3$, so Eq. (43.1) shows that the *volume* of a nucleus is proportional to A. Dividing A (the approximate mass in u) by the volume gives us the approximate density and cancels out A. Thus *all nuclei have approximately the same density.* This fact is of crucial importance in understanding nuclear structure.

EXAMPLE 43.1 CALCULATING NUCLEAR PROPERTIES

SOLUTION

The most common kind of iron nucleus has mass number $A = 56$. Find the radius, approximate mass, and approximate density of the nucleus.

SOLUTION

IDENTIFY and SET UP: Equation (43.1) tells us how the nuclear radius R depends on the mass number A. The mass of the nucleus in atomic mass units is approximately equal to the value of A, and the density ρ is mass divided by volume.

EXECUTE: The radius and approximate mass are

$$\begin{aligned} R &= R_0 A^{1/3} = (1.2 \times 10^{-15}\text{ m})(56)^{1/3} \\ &= 4.6 \times 10^{-15}\text{ m} = 4.6\text{ fm} \\ m &\approx (56\text{ u})(1.66 \times 10^{-27}\text{ kg/u}) = 9.3 \times 10^{-26}\text{ kg} \end{aligned}$$

The volume V of the nucleus (which we treat as a sphere of radius R) and its density ρ are

$$\begin{aligned} V &= \tfrac{4}{3}\pi R^3 = \tfrac{4}{3}\pi R_0^3 A = \tfrac{4}{3}\pi(4.6 \times 10^{-15}\text{ m})^3 \\ &= 4.1 \times 10^{-43}\text{ m}^3 \\ \rho &= \frac{m}{V} \approx \frac{9.3 \times 10^{-26}\text{ kg}}{4.1 \times 10^{-43}\text{ m}^3} = 2.3 \times 10^{17}\text{ kg/m}^3 \end{aligned}$$

EVALUATE: As we mentioned above, *all* nuclei have approximately this same density. The density of solid iron is about 7000 kg/m^3; the iron nucleus is more than 10^{13} times as dense as iron in bulk. Such densities are also found in *neutron stars,* which are similar to gigantic nuclei made almost entirely of neutrons. A 1-cm cube of material with this density would have a mass of 2.3×10^{11} kg, or 230 million metric tons!

Nuclides and Isotopes

The building blocks of the nucleus are the proton and the neutron. In a neutral atom, there is one electron for every proton in the nucleus. We introduced these particles in Section 21.1; we'll recount the discovery of the neutron and proton in Chapter 44. The masses of these particles are

$$\begin{aligned} &\text{Proton:} && m_\text{p} = 1.007276\text{ u} = 1.672622 \times 10^{-27}\text{ kg} \\ &\text{Neutron:} && m_\text{n} = 1.008665\text{ u} = 1.674927 \times 10^{-27}\text{ kg} \\ &\text{Electron:} && m_\text{e} = 0.000548580\text{ u} = 9.10938 \times 10^{-31}\text{ kg} \end{aligned}$$

The number of protons in a nucleus is the **atomic number** Z. The number of neutrons is the **neutron number** N. The nucleon number or mass number A is the sum of the number of protons Z and the number of neutrons N:

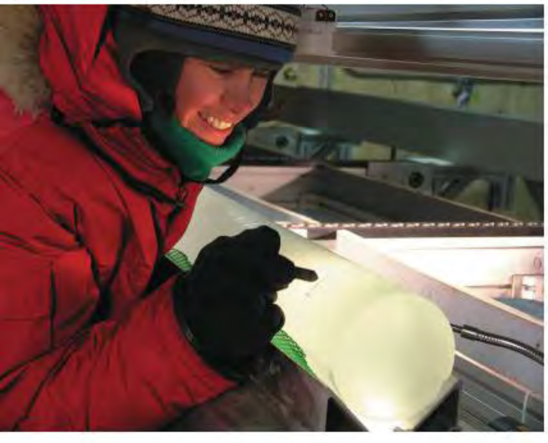
$$A = Z + N \tag{43.2}$$

TABLE 43.1 Compositions of Some Common Nuclides

Z = atomic number (number of protons)
N = neutron number
$A = Z + N$ = mass number (total number of nucleons)

Nucleus	Z	N	$A = Z + N$
${}^{1}_{1}H$	1	0	1
${}^{2}_{1}H$	1	1	2
${}^{4}_{2}He$	2	2	4
${}^{6}_{3}Li$	3	3	6
${}^{7}_{3}Li$	3	4	7
${}^{9}_{4}Be$	4	5	9
${}^{10}_{5}B$	5	5	10
${}^{11}_{5}B$	5	6	11
${}^{12}_{6}C$	6	6	12
${}^{13}_{6}C$	6	7	13
${}^{14}_{7}N$	7	7	14
${}^{16}_{8}O$	8	8	16
${}^{23}_{11}Na$	11	12	23
${}^{65}_{29}Cu$	29	36	65
${}^{200}_{80}Hg$	80	120	200
${}^{235}_{92}U$	92	143	235
${}^{238}_{92}U$	92	146	238

Application Using Isotopes to Measure Ancient Climate This sample of ice from Antarctica was deposited tens of thousands of years ago. The deeper the sample, the further in the past the ice was deposited. Most of the water molecules (H_2O) in the ice contain the oxygen isotope ${}^{16}_{8}O$, but a small percentage contain the heavier isotope ${}^{18}_{8}O$. Water molecules that contain the lighter isotope evaporate more readily, but condense less readily, than water molecules that include the heavier isotope, and these processes vary with temperature. Measuring the ratio of ${}^{18}_{8}O$ to ${}^{16}_{8}O$ in an ancient ice sample thus allows scientists to determine the average ocean temperature at the time the sample was deposited. Scientists also measure the amount of atmospheric carbon dioxide (CO_2) that was trapped in the ice when it was deposited. These observations have helped confirm the idea that high atmospheric CO_2 concentrations go hand in hand with high temperatures, a key principle for understanding 21st-century climate change (see Section 17.7).



A single nuclear species having specific values of both Z and N is called a **nuclide. Table 43.1** lists values of A, Z, and N for some nuclides. The electron structure of an atom, which is responsible for its chemical properties, is determined by the charge Ze of the nucleus. The table shows some nuclides that have the same number of protons Z but a different number of neutrons N. These nuclides are called **isotopes** of that element. A familiar example is chlorine (Cl, $Z = 17$). About 76% of chlorine nuclei have $N = 18$; the other 24% have $N = 20$. Different isotopes of an element usually have slightly different physical properties such as melting and boiling temperatures and diffusion rates. The two common isotopes of uranium with $A = 235$ and 238 are usually separated industrially by taking advantage of the different diffusion rates of gaseous uranium hexafluoride (UF_6) containing the two isotopes.

Table 43.1 also shows the usual notation for individual nuclides: the symbol of the element, with a pre-subscript equal to Z and a pre-superscript equal to the mass number A. The general format for an element El is ${}^{A}_{Z}El$. The isotopes of chlorine mentioned above, with $A = 35$ and 37, are written ${}^{35}_{17}Cl$ and ${}^{37}_{17}Cl$ and pronounced "chlorine-35" and "chlorine-37," respectively. This name of the element determines the atomic number Z, so the pre-subscript Z is sometimes omitted, as in ${}^{35}Cl$.

Table 43.2 gives the masses of some common atoms, including their electrons. Note that this table gives masses of *neutral* atoms (with Z electrons) rather than masses of *bare* nuclei, because it is much more difficult to measure masses of bare nuclei with high precision. The mass of a neutral carbon-12 atom is exactly 12 u; that's how the unified atomic mass unit is defined. The masses of other atoms are *approximately* equal to A atomic mass units, as we stated earlier. In fact, the atomic masses are *less* than the sum of the masses of their parts (the Z protons, the Z electrons, and the N neutrons). We'll explain this very important mass difference in the next section.

TABLE 43.2 Neutral Atomic Masses for Some Light Nuclides

Element and Isotope	Atomic Number, Z	Neutron Number, N	Atomic Mass (u)	Mass Number, A
Hydrogen (^1_1H)	1	0	1.007825	1
Deuterium (^2_1H)	1	1	2.014102	2
Tritium (^3_1H)	1	2	3.016049	3
Helium (^3_2He)	2	1	3.016029	3
Helium (^4_2He)	2	2	4.002603	4
Lithium (^6_3Li)	3	3	6.015123	6
Lithium (^7_3Li)	3	4	7.016005	7
Beryllium (^9_4Be)	4	5	9.012182	9
Boron ($^{10}_5B$)	5	5	10.012937	10
Boron ($^{11}_5B$)	5	6	11.009305	11
Carbon ($^{12}_6C$)	6	6	12.000000	12
Carbon ($^{13}_6C$)	6	7	13.003355	13
Nitrogen ($^{14}_7N$)	7	7	14.003074	14
Nitrogen ($^{15}_7N$)	7	8	15.000109	15
Oxygen ($^{16}_8O$)	8	8	15.994915	16
Oxygen ($^{17}_8O$)	8	9	16.999132	17
Oxygen ($^{18}_8O$)	8	10	17.999161	18

Source: G. Audi, A. H. Wapstra, and C. Thibault, *Nuclear Physics* **A729**, 337 (2003).

Nuclear Spins and Magnetic Moments

Like electrons, nucleons (protons and neutrons) are spin-$\frac{1}{2}$ particles with spin angular momenta given by the same equations as in Section 41.5. The magnitude of the spin angular momentum $\vec{S}$ of a nucleon is

$$S = \sqrt{\tfrac{1}{2}\left(\tfrac{1}{2} + 1\right)}\hbar = \sqrt{\tfrac{3}{4}}\hbar \tag{43.3}$$

and the z-component is

$$S_z = \pm\tfrac{1}{2}\hbar \tag{43.4}$$

In addition to its spin angular momentum, a nucleon may have *orbital* angular momentum $\vec{L}$ associated with its motion within the nucleus. The values of $\vec{L}$ and of its z-component L_z for a nucleon are quantized in the same way as for an electron in an atom.

The *total* angular momentum $\vec{J}$ of the nucleus is the vector sum of the individual spin and orbital angular momenta of all the nucleons. It has magnitude

$$J = \sqrt{j(j+1)}\,\hbar \tag{43.5}$$

and z-component

$$J_z = m_j\hbar \qquad (m_j = -j, -j+1, \ldots, j-1, j) \tag{43.6}$$

The total nuclear angular momentum quantum number j is usually called the *nuclear spin,* even though in general it refers to a combination of the orbital and spin angular momenta of the nucleons that make up the nucleus. When the total number of nucleons A is *even,* j is an integer; when it is *odd*, j is a half-integer. All nuclides for which both Z and N are even have $J = 0$. As we will see, this happens because nucleons tend to form pairs with opposite spin components.

Associated with nuclear angular momentum is a *magnetic moment*. When we discussed *electron* magnetic moments in Section 41.4, we introduced the Bohr magneton $\mu_B = e\hbar/2m_e$ as a natural unit of magnetic moment. We found that the magnitude of the z-component of the electron spin magnetic moment is almost exactly equal to μ_B; that is, $|\mu_{sz}|_{\text{electron}} \approx \mu_B$. In discussing *nuclear* magnetic moments, we can define an analogous quantity, the **nuclear magneton** μ_n:

$$\mu_n = \frac{e\hbar}{2m_p} = 5.05078 \times 10^{-27}\ \text{J/T} = 3.15245 \times 10^{-8}\ \text{eV/T} \qquad (43.7)$$

(nuclear magneton)

The proton mass m_p is 1836 times larger than the electron mass m_e, so the nuclear magneton μ_n is 1836 times smaller than the Bohr magneton μ_B.

We might expect the magnitude of the z-component of the spin magnetic moment of the proton to be approximately μ_n. Instead, it turns out to be

$$|\mu_{sz}|_{\text{proton}} = 2.7928\mu_n \qquad (43.8)$$

Even more surprising, the neutron, which has zero charge, has a spin magnetic moment; its z-component has magnitude

$$|\mu_{sz}|_{\text{neutron}} = 1.9130\mu_n \qquad (43.9)$$

The proton has a positive charge; as expected, its spin magnetic moment $\vec{\boldsymbol{\mu}}$ is parallel to its spin angular momentum $\vec{\boldsymbol{S}}$. However, $\vec{\boldsymbol{\mu}}$ and $\vec{\boldsymbol{S}}$ are opposite for a neutron, as would be expected for a *negative* charge distribution. These *anomalous* magnetic moments arise because the proton and neutron aren't really fundamental particles but are made of simpler particles called *quarks*. We'll discuss quarks in some detail in Chapter 44.

The magnetic moment of an entire nucleus is typically a few nuclear magnetons. When a nucleus is placed in an external magnetic field $\vec{\boldsymbol{B}}$, there is an interaction energy $U = -\vec{\boldsymbol{\mu}} \cdot \vec{\boldsymbol{B}} = -\mu_z B$ just as with atomic magnetic moments. The components of the magnetic moment in the direction of the field μ_z are quantized, so a series of energy levels results from this interaction.

EXAMPLE 43.2 PROTON SPIN FLIPS

SOLUTION

Protons are placed in a 2.30-T magnetic field that points in the positive z-direction. (a) What is the energy difference between states with the z-component of proton spin angular momentum parallel and antiparallel to the field? (b) A proton can make a transition from one of these states to the other by emitting or absorbing a photon with the appropriate energy. Find the frequency and wavelength of such a photon.

SOLUTION

IDENTIFY and SET UP: The proton is a spin-$\frac{1}{2}$ particle with a magnetic moment $\vec{\boldsymbol{\mu}}$ in the same direction as its spin $\vec{\boldsymbol{S}}$, so its energy depends on the orientation of its spin relative to an applied magnetic field $\vec{\boldsymbol{B}}$. If the z-component of $\vec{\boldsymbol{S}}$ is aligned with $\vec{\boldsymbol{B}}$, then μ_z is equal to the positive value given in Eq. (43.8). If the z-component of $\vec{\boldsymbol{S}}$ is opposite $\vec{\boldsymbol{B}}$, then μ_z is the negative of this value. The interaction energy in either case is $U = -\mu_z B$; the difference between these energies is our target variable in part (a). We find the photon frequency and wavelength by using $E = hf = hc/\lambda$.

EXECUTE: (a) When the z-components of $\vec{\boldsymbol{S}}$ and $\vec{\boldsymbol{\mu}}$ are parallel to $\vec{\boldsymbol{B}}$, the interaction energy is

$$U = -|\mu_z|B = -(2.7928)(3.152 \times 10^{-8}\ \text{eV/T})(2.30\ \text{T})$$
$$= -2.025 \times 10^{-7}\ \text{eV}$$

When the z-components of $\vec{\boldsymbol{S}}$ and $\vec{\boldsymbol{\mu}}$ are antiparallel to the field, the energy is $+2.025 \times 10^{-7}$ eV. Hence the energy *difference* between the states is

$$\Delta E = 2(2.025 \times 10^{-7}\ \text{eV}) = 4.05 \times 10^{-7}\ \text{eV}$$

(b) The corresponding photon frequency and wavelength are

$$f = \frac{\Delta E}{h} = \frac{4.05 \times 10^{-7}\ \text{eV}}{4.136 \times 10^{-15}\ \text{eV} \cdot \text{s}} = 9.79 \times 10^{7}\ \text{Hz} = 97.9\ \text{MHz}$$

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^{8}\ \text{m/s}}{9.79 \times 10^{7}\ \text{s}^{-1}} = 3.06\ \text{m}$$

EVALUATE: This frequency is in the middle of the FM radio band. When a hydrogen specimen is placed in a 2.30-T magnetic field and irradiated with radio waves of this frequency, proton *spin flips* can be detected by the absorption of energy from the radiation.

Nuclear Magnetic Resonance and MRI

Spin-flip experiments of the sort referred to in Example 43.2 are called *nuclear magnetic resonance* (NMR). They have been carried out with many different nuclides. Frequencies and magnetic fields can be measured very precisely, so this technique permits precise measurements of nuclear magnetic moments. An elaboration of this basic idea leads to *magnetic resonance imaging* (MRI), a noninvasive medical imaging technique that discriminates among various types of body tissues on the basis of the differing environments of protons in the tissues (**Fig. 43.1**).

The magnetic moment of a nucleus is also the *source* of a magnetic field. In an atom the interaction of an electron's magnetic moment with the field of the nucleus's magnetic moment causes additional splittings in atomic energy levels and spectra. We called this effect *hyperfine structure* in Section 41.5. Measurements of the hyperfine structure may be used to directly determine the nuclear spin.

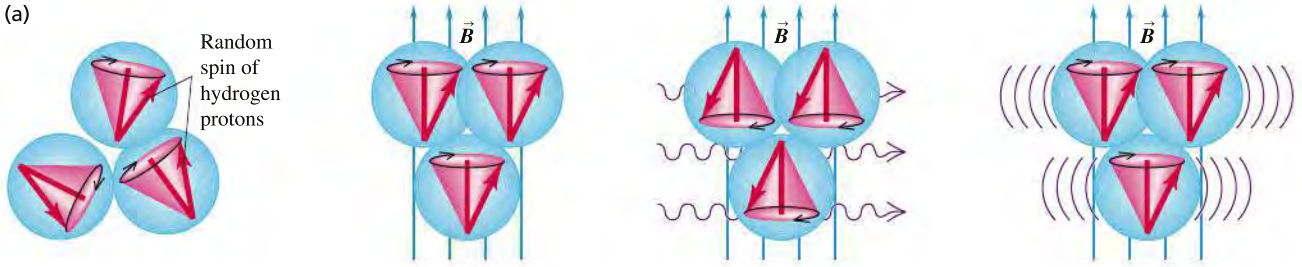
MP

PhET: Simplified MRI

TEST YOUR UNDERSTANDING OF SECTION 43.1 (a) By what factor must the mass number of a nucleus increase to double its volume? (i) $\sqrt[3]{2}$; (ii) $\sqrt{2}$; (iii) 2; (iv) 4; (v) 8. (b) By what factor must the mass number increase to double the radius of the nucleus? (i) $\sqrt[3]{2}$; (ii) $\sqrt{2}$; (iii) 2; (iv) 4; (v) 8. ❙

43.1 Magnetic resonance imaging (MRI).

(a)



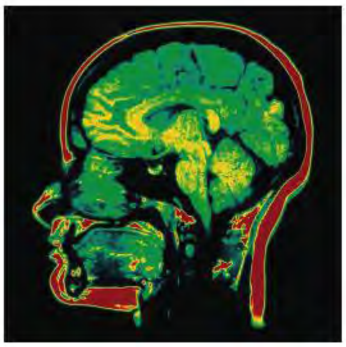
Protons, the nuclei of hydrogen atoms in the tissue under study, normally have random spin orientations.

In the presence of a strong magnetic field, the spins become aligned with a component parallel to $\vec{B}$.

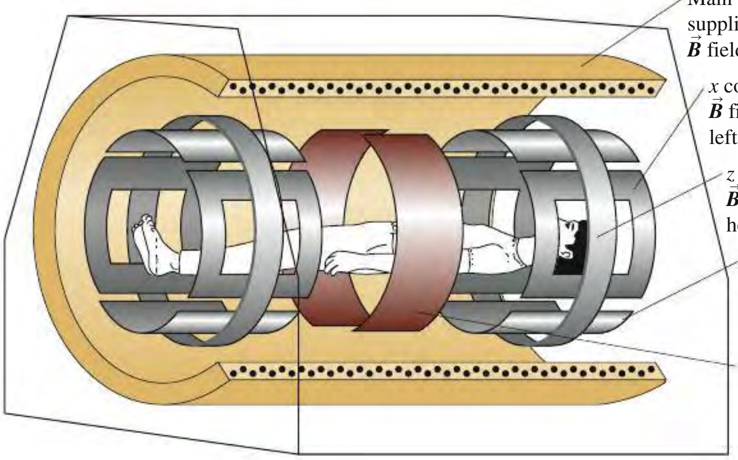
A brief radio signal causes the spins to flip orientation.

As the protons realign with the $\vec{B}$ field, they emit radio waves that are picked up by sensitive detectors.

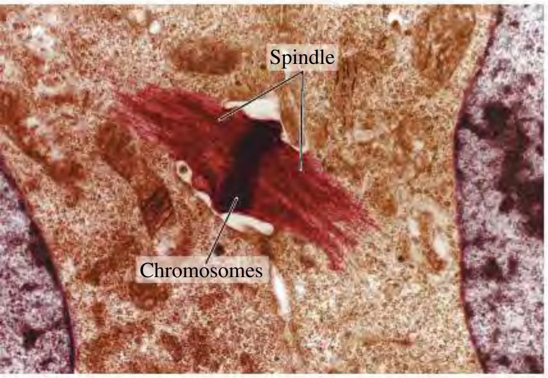
(b) Since $\vec{B}$ has a different value at different locations in the tissue, the radio waves from different locations have different frequencies. This makes it possible to construct an image.



(c) An electromagnet used for MRI



BIO Application Deuterium and Heavy Water Toxicity A crucial step in plant and animal cell division is the formation of a spindle, which separates the two sets of daughter chromosomes. If a plant is given only heavy water—in which one or both of the hydrogen atoms in an H_2O molecule are replaced with a deuterium atom—cell division stops and the plant stops growing. The reason is that deuterium is more massive than ordinary hydrogen, so the O–H bond in heavy water has a slightly different binding energy and heavy water has slightly different properties as a solvent. The biochemical reactions that occur during cell division are very sensitive to these solvent properties, so a spindle never forms and the cell cannot reproduce.



43.2 NUCLEAR BINDING AND NUCLEAR STRUCTURE

Because energy must be added to a nucleus to separate it into its individual protons and neutrons, the total rest energy E_0 of the separated nucleons is greater than the rest energy of the nucleus. The energy that must be added to separate the nucleons is called the **binding energy** E_B; it is the magnitude of the energy by which the nucleons are bound together. Thus the rest energy of the nucleus is $E_0 - E_B$. The binding energy is defined as

Binding energy of a nucleus with Z protons, N neutrons

$$E_B = (\overset{\text{Atomic number}}{Z}\underset{\text{Mass of hydrogen atom}}{M_H} + \overset{\text{Neutron number}}{N}\underset{\text{Neutron mass}}{m_n} - \underset{\text{Mass of neutral atom containing nucleus}}{{}^{A}_{Z}M})\overset{\text{(Speed of light in vacuum)}^2 = 931.5\text{ MeV/u}}{c^2} \tag{43.10}$$

Note that Eq. (43.10) does not include Zm_p, the mass of Z protons. Rather, it contains ZM_H, the mass of Z protons and Z electrons combined as Z neutral ${}^1_1H$ atoms, to balance the Z electrons included in ${}^A_ZM$, the mass of the neutral atom.

The simplest nucleus is that of hydrogen, a single proton. Next comes the nucleus of ${}^2_1H$, the isotope of hydrogen with mass number 2, usually called *deuterium*. Its nucleus consists of a proton and a neutron bound together to form a particle called the *deuteron*. By using values from Table 43.2 in Eq. (43.10), we find that the binding energy of the deuteron is

$$\begin{aligned} E_B &= (1.007825\text{ u} + 1.008665\text{ u} - 2.014102\text{ u})(931.5\text{ MeV/u}) \\ &= 2.224\text{ MeV} \end{aligned}$$

This much energy would be required to pull the deuteron apart into a proton and a neutron. An important measure of how tightly a nucleus is bound is the *binding energy per nucleon,* E_B/A. At $(2.224\text{ MeV})/(2\text{ nucleons}) = 1.112$ MeV per nucleon, ${}^2_1H$ has the lowest binding energy per nucleon of all nuclides.

Using the equivalence of rest mass and energy (Section 37.8), we see that the mass of a nucleus is always *less* than the total mass of its nucleons by an amount $\Delta M = E_B/c^2$, called the *mass defect*. For example, the mass defect of ${}^2_1H$ is $\Delta M = E_B/c^2 = (2.224\text{ MeV})/(931.5\text{ MeV/u}) = 0.002388$ u.

PROBLEM-SOLVING STRATEGY 43.1 NUCLEAR PROPERTIES

IDENTIFY *the relevant concepts:* The key properties of a nucleus are its mass, radius, binding energy, mass defect, binding energy per nucleon, and angular momentum.

SET UP *the problem:* Once you have identified the target variables, assemble the equations needed to solve the problem. A relatively small number of equations from this section and Section 43.1 are all you need.

EXECUTE *the solution:* Solve for the target variables. Binding-energy calculations that use Eq. (43.10) often involve subtracting two nearly equal quantities. To get enough precision in the difference, you may need to carry as many as nine significant figures, if that many are available.

EVALUATE *your answer:* It's useful to be familiar with the following benchmark magnitudes. Protons and neutrons are about 1840 times as massive as electrons. Nuclear radii are of the order of 10^{-15} m. The electric potential energy of two protons in a nucleus is roughly 10^{-13} J or 1 MeV, so nuclear interaction energies are typically a few MeV rather than a few eV as with atoms. The binding energy per nucleon is about 1% of the nucleon rest energy. (The ionization energy of the hydrogen atom is only 0.003% of the electron's rest energy.) Angular momenta are determined only by the value of $\hbar$, so they are of the same order of magnitude in both nuclei and atoms. Nuclear magnetic moments, however, are about a factor of 1000 *smaller* than those of electrons in atoms because nuclei are so much more massive than electrons.

EXAMPLE 43.3 THE MOST STRONGLY BOUND NUCLIDE

SOLUTION

Find the mass defect, the total binding energy, and the binding energy per nucleon of $^{62}_{28}\mathrm{Ni}$, which has the highest binding energy per nucleon of all nuclides (**Fig. 43.2**). The neutral atomic mass of $^{62}_{28}\mathrm{Ni}$ is 61.928345 u.

SOLUTION

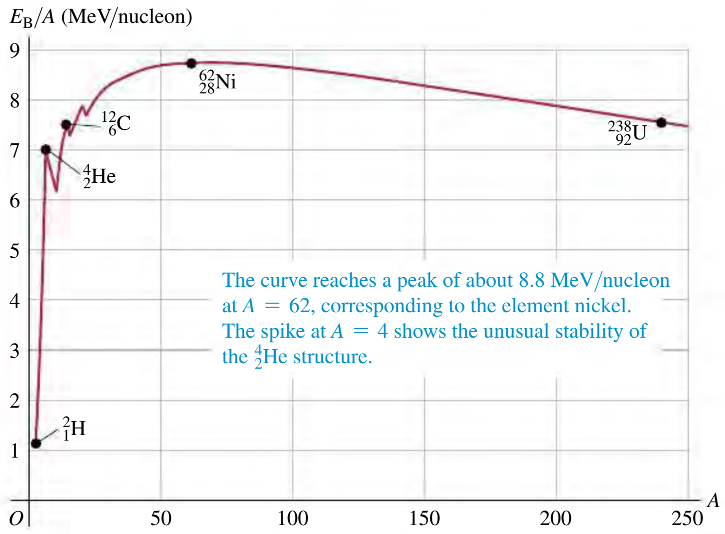
IDENTIFY and SET UP: The mass defect ΔM is the difference between the mass of the nucleus and the combined mass of its constituent nucleons. The binding energy E_B is this quantity multiplied by c^2, and the binding energy per nucleon is E_B divided by the mass number A. We use Eq. (43.10), $\Delta M = ZM_H + Nm_n - {}^A_ZM$, to determine both the mass defect and the binding energy.

EXECUTE: With $Z = 28$, $M_H = 1.007825$ u, $N = A - Z = 62 - 28 = 34$, $m_n = 1.008665$ u, and ${}^A_ZM = 61.928345$ u, Eq. (43.10) gives $\Delta M = 0.585365$ u. The binding energy is then

$$E_B = (0.585365\ \mathrm{u})(931.5\ \mathrm{MeV/u}) = 545.3\ \mathrm{MeV}$$

The binding energy *per nucleon* is $E_B/A = (545.3\ \mathrm{MeV})/62$, or 8.795 MeV per nucleon.

EVALUATE: Our result means that it would take a minimum of 545.3 MeV to pull a $^{62}_{28}\mathrm{Ni}$ completely apart into 28 protons and 34 neutrons. The mass defect of $^{62}_{28}\mathrm{Ni}$ is about 1% of the atomic (or the nuclear) mass. The binding energy is therefore about 1% of the rest energy of the nucleus, and the binding energy per nucleon is about 1% of the rest energy of a nucleon. Note that the mass defect is more than half the mass of a nucleon, which suggests how tightly bound nuclei are.



43.2 Approximate binding energy per nucleon as a function of mass number A (the total number of nucleons) for stable nuclides.

Nearly all stable nuclides, from the lightest to the most massive, have binding energies in the range of 7–9 MeV per nucleon. Figure 43.2 is a graph of binding energy per nucleon as a function of the mass number A. Note the spike at $A = 4$, showing the unusually large binding energy per nucleon of the $^{4}_{2}\mathrm{He}$ nucleus (alpha particle) relative to its neighbors. To explain this curve, we must consider the interactions among the nucleons.

The Nuclear Force

The force that binds protons and neutrons together in the nucleus, despite the electrical repulsion of the protons, is an example of the *strong interaction* that we mentioned in Section 5.5. In the context of nuclear structure, this interaction is called the *nuclear force*. Here are some of its characteristics. First, it does not depend on charge; neutrons as well as protons are bound, and the binding is the same for both. Second, it has short range, of the order of nuclear dimensions—that is, 10^{-15} m. (Otherwise, the nucleus would grow by pulling in additional protons and neutrons.) But within its range, the nuclear force is much stronger than electric forces; otherwise, the nucleus could never be stable. It would be nice if we could write a simple equation like Newton's law of gravitation or Coulomb's law for this force, but physicists have yet to fully determine its dependence on the separation r. Third, the nearly constant density of nuclear matter and the nearly constant binding energy per nucleon of larger nuclides show that a particular nucleon cannot interact simultaneously with *all* the other nucleons in a nucleus, but only with those few in its immediate vicinity. This is different from electric forces; *every* proton in the nucleus repels every other one. This limited number of interactions is called *saturation;* it is analogous to covalent bonding in molecules and solids. Finally, the nuclear force favors binding of *pairs* of protons or neutrons with opposite spins and of *pairs of pairs*—that is, a pair of protons and a pair of neutrons, each pair having opposite spins. Hence the alpha particle (two protons and two neutrons) is an exceptionally stable nucleus for its mass number. We'll see other evidence for pairing effects in nuclei in the next subsection. (In Section 42.8 we described an analogous pairing that binds opposite-spin electrons in Cooper pairs in the BCS theory of superconductivity.)

The analysis of nuclear structure is more complex than the analysis of many-electron atoms. Two different kinds of interactions are involved (electrical and nuclear). Even so, we can gain some insight into nuclear structure by the use of simple models. We'll discuss briefly two rather different but successful models, the *liquid-drop model* and the *shell model*.

The Liquid-Drop Model

The **liquid-drop model,** first proposed in 1928 by the Russian physicist George Gamow and later expanded on by Niels Bohr, is suggested by the observation that all nuclei have nearly the same density. The individual nucleons are analogous to molecules of a liquid, held together by short-range interactions and surface-tension effects. We can use this simple picture to derive a formula for the estimated total binding energy of a nucleus. We'll include five contributions:

1. We've remarked that nuclear forces show *saturation;* an individual nucleon interacts only with a few of its nearest neighbors. This effect gives a binding-energy term that is proportional to the number of nucleons. We write this term as C_1A, where C_1 is an experimentally determined constant.
2. The nucleons on the surface of the nucleus are less tightly bound than those in the interior because they have no neighbors outside the surface. This decrease in the binding energy gives a *negative* energy term proportional to the surface area $4\pi R^2$. Because R is proportional to $A^{1/3}$, this term is proportional to $A^{2/3}$; we write it as $-C_2A^{2/3}$, where C_2 is another constant.
3. Every one of the Z protons repels every one of the $(Z-1)$ other protons. The total repulsive electric potential energy is proportional to $Z(Z-1)$ and inversely proportional to the radius R and thus to $A^{1/3}$. This energy term is negative because the nucleons are less tightly bound than they would be without the electrical repulsion. We write this correction as $-C_3Z(Z-1)/A^{1/3}$.
4. Observations show that nuclei are most tightly bound if N is close to Z for small A and N is greater than Z (but not too much greater) for larger A. We need a negative energy term corresponding to the difference $|N-Z|$. The best agreement with observed binding energies is obtained if this term is proportional to $(N-Z)^2/A$. If we use $N = A - Z$ to express this energy in terms of A and Z, this correction is $-C_4(A-2Z)^2/A$.
5. Finally, the nuclear force favors *pairing* of protons and of neutrons. This energy term is positive (more binding) if both Z and N are even, negative (less binding) if both Z and N are odd, and zero otherwise. The best fit to the data occurs with the form $\pm C_5A^{-4/3}$ for this term.

The total estimated binding energy E_B is the sum of these five terms:

$$E_B = C_1A - C_2A^{2/3} - C_3\frac{Z(Z-1)}{A^{1/3}} - C_4\frac{(A-2Z)^2}{A} \pm C_5A^{-4/3} \tag{43.11}$$

(nuclear binding energy)

The constants C_1, C_2, C_3, C_4, and C_5, chosen to make this formula best fit the observed binding energies of nuclides, are

$$\begin{aligned} C_1 &= 15.75 \text{ MeV} \\ C_2 &= 17.80 \text{ MeV} \\ C_3 &= 0.7100 \text{ MeV} \\ C_4 &= 23.69 \text{ MeV} \\ C_5 &= 39 \text{ MeV} \end{aligned}$$

The constant C_1 is the binding energy per nucleon due to the saturated nuclear force. This energy is almost 16 MeV per nucleon, about double the *total* binding energy per nucleon in most nuclides.

If we use Eq. (43.11) to estimate the binding energy E_B, we can solve Eq. (43.10) to use it to estimate the mass of any neutral atom:

$$ {}^{A}_{Z}M = ZM_H + Nm_n - \frac{E_B}{c^2} \qquad \text{(semiempirical mass formula)} \qquad (43.12)$$

Equation (43.12) is called the *semiempirical mass formula*. The name is apt; the equation is *empirical* in the sense that the C's have to be determined empirically (experimentally), yet it does have a sound theoretical basis.

SOLUTION

EXAMPLE 43.4 ESTIMATING BINDING ENERGY AND MASS

For the nuclide ${}^{62}_{28}\text{Ni}$ of Example 43.3, (a) calculate the five terms in the binding energy and the total estimated binding energy, and (b) find the neutral atomic mass using the semiempirical mass formula.

SOLUTION

IDENTIFY and SET UP: We use the liquid-drop model of the nucleus and its five contributions to the binding energy, as given by Eq. (43.11), to calculate the total binding energy E_B. We then use Eq. (43.12) to find the neutral atomic mass ${}^{62}_{28}M$.

EXECUTE: (a) With $Z = 28$, $A = 62$, and $N = 34$, the five terms in Eq. (43.11) are

1. $C_1A = (15.75 \text{ MeV})(62) = 976.5 \text{ MeV}$
2. $-C_2A^{2/3} = -(17.80 \text{ MeV})(62)^{2/3} = -278.8 \text{ MeV}$
3. $-C_3\dfrac{Z(Z-1)}{A^{1/3}} = -(0.7100 \text{ MeV})\dfrac{(28)(27)}{(62)^{1/3}} = -135.6 \text{ MeV}$
4. $-C_4\dfrac{(A-2Z)^2}{A} = -(23.69 \text{ MeV})\dfrac{(62-56)^2}{62} = -13.8 \text{ MeV}$
5. $+C_5A^{-4/3} = (39 \text{ MeV})(62)^{-4/3} = 0.2 \text{ MeV}$

The pairing correction (term 5) is by far the smallest of all the terms; it is positive because both Z and N are even. The sum of all five terms is the total estimated binding energy, $E_B = 548.5$ MeV.

(b) We use $E_B = 548.5$ MeV in Eq. (43.12):

$$\begin{aligned} {}^{62}_{28}M &= 28(1.007825 \text{ u}) + 34(1.008665 \text{ u}) - \frac{548.5 \text{ MeV}}{931.5 \text{ MeV/u}} \\ &= 61.925 \text{ u} \end{aligned}$$

EVALUATE: The binding energy of ${}^{62}_{28}\text{Ni}$ calculated in part (a) is only about 0.6% larger than the true value of 545.3 MeV found in Example 43.3, and the mass calculated in part (b) is only about 0.005% smaller than the measured value of 61.928345 u. The semiempirical mass formula can be quite accurate!

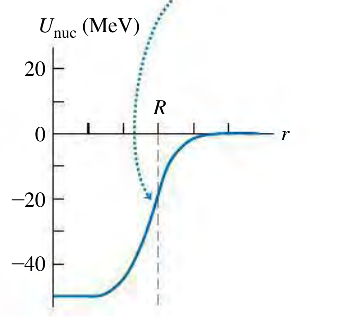
The liquid-drop model and the mass formula derived from it are quite successful in correlating nuclear masses, and we will see later that they help in understanding decay processes of unstable nuclides. Other aspects of nuclei, such as angular momentum and excited states, are better approached with different models.

The Shell Model

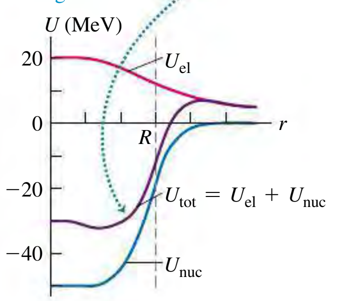
The **shell model** of nuclear structure is analogous to the central-field approximation in atomic physics (see Section 41.6). We picture each nucleon as moving in a potential that represents the averaged-out effect of all the other nucleons. Although this is a very simplified model, in several respects it works out quite well.

43.3 Approximate potential-energy functions for a nucleon in a nucleus. The approximate nuclear radius is R.

(a) The potential energy U_{nuc} due to the nuclear force is the same for protons and neutrons. For neutrons, it is the *total* potential energy.



(b) For protons, the total potential energy U_{tot} is the sum of the nuclear (U_{nuc}) and electric (U_{el}) potential energies.



The potential-energy function for the nuclear force is the same for protons as for neutrons. **Figure 43.3a** shows a reasonable assumption for the shape of this function: a spherical version of the square-well potential we discussed in Section 40.3. The corners are somewhat rounded because the nucleus doesn't have a sharply defined surface. For protons there is an additional potential energy associated with electrical repulsion. We consider each proton to interact with a sphere of uniform charge density, with radius R and total charge $(Z - 1)e$. Figure 43.3b shows the nuclear, electric, and total potential energies for a proton as functions of the distance r from the center of the nucleus.

In principle, we could solve the Schrödinger equation for a proton or neutron moving in such a potential. For any spherically symmetric potential energy, the angular-momentum states are the same as for the electrons in the central-field approximation in atomic physics. In particular, we can use the concept of *filled shells and subshells* and their relationship to stability. As we saw in Section 41.6, the exclusion principle forbids more than one electron from occupying any given quantum-mechanical state in a multi-electron atom. This explains why the values $Z = 2, 10, 18, 36, 54$, and 86 (the atomic numbers of the noble gases) correspond to atoms with particularly stable electron arrangements.

A comparable effect occurs in nuclear structure. Like electrons, protons and neutrons are spin-$\frac{1}{2}$ particles. So no more than one proton can be in any given quantum-mechanical state in a nucleus, and likewise for neutrons. Just as for electrons in atoms, there are certain numbers of protons *or* of neutrons, called *magic numbers,* that correspond to particularly stable nuclei—that is, nuclei with particularly high binding energies. The magic numbers are 2, 8, 20, 28, 50, 82, and 126. These numbers are different from those for electrons in atoms because the potential-energy function is different and the nuclear spin-orbit interaction is much stronger and of opposite sign than in atoms. So nuclear subshells fill up in a different order from those for electrons in an atom. Nuclides in which Z is a magic number tend to have an above-average number of stable isotopes. (Nuclides with $Z = 126$ have not been observed in nature.) There are several *doubly magic* nuclides for which both Z and N are magic, including

$$^{4}_{2}\mathrm{He} \qquad ^{16}_{8}\mathrm{O} \qquad ^{40}_{20}\mathrm{Ca} \qquad ^{48}_{20}\mathrm{Ca} \qquad ^{208}_{82}\mathrm{Pb}$$

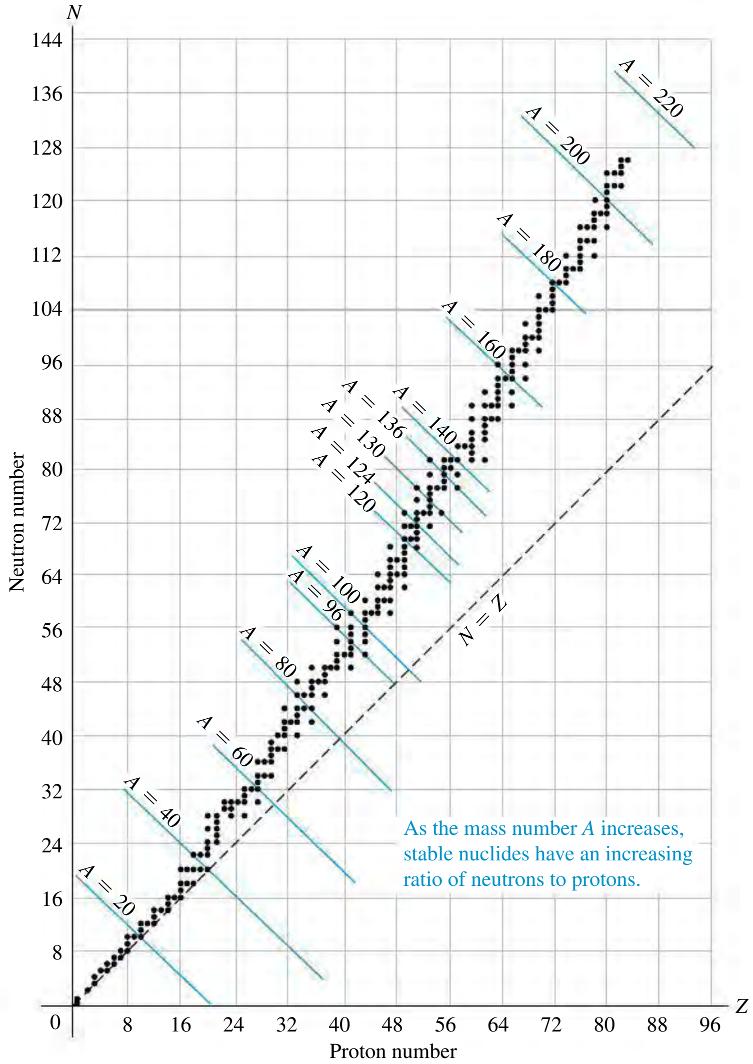
All these nuclides have substantially higher binding energy per nucleon than do nuclides with neighboring values of N or Z. They also all have zero nuclear spin. The magic numbers correspond to filled-shell or -subshell configurations of nucleon energy levels with a relatively large jump in energy to the next allowed level.

TEST YOUR UNDERSTANDING OF SECTION 43.2 Rank the following nuclei in order from largest to smallest value of the binding energy per nucleon. (i) $^{4}_{2}\mathrm{He}$; (ii) $^{52}_{24}\mathrm{Cr}$; (iii) $^{152}_{62}\mathrm{Sm}$; (iv) $^{200}_{80}\mathrm{Hg}$; (v) $^{252}_{92}\mathrm{Cf}$. ❙

43.3 NUCLEAR STABILITY AND RADIOACTIVITY

Among about 2500 known nuclides, fewer than 300 are stable. The others are unstable structures that decay to form other nuclides by emitting particles and electromagnetic radiation, a process called **radioactivity.** The time scale of these decay processes ranges from a small fraction of a microsecond to billions of years. The *stable* nuclides are shown by dots on the graph in **Fig. 43.4**, where the neutron number N and proton number (or atomic number) Z for each nuclide are plotted. Such a chart is called a *Segrè chart*, after its inventor, the Italian-American physicist Emilio Segrè (1905–1989).

Each blue line in Fig. 43.4 represents a specific value of the mass number $A = Z + N$. Most lines of constant A pass through only one or two stable nuclides; that is, there is usually a very narrow range of stability for a given mass



43.4 Segrè chart showing neutron number and proton number for stable nuclides.

number. The lines at $A = 20$, 40, 60, and 80 are examples. In four exceptional cases ($A = 94$, 124, 130, and 136), these lines pass through *three* stable nuclides.

Only four stable nuclides have both odd Z and odd N:

$$^{2}_{1}\mathrm{H} \qquad ^{6}_{3}\mathrm{Li} \qquad ^{10}_{5}\mathrm{B} \qquad ^{14}_{7}\mathrm{N}$$

These are called *odd-odd nuclides*. The absence of other odd-odd nuclides shows the importance of pairing in adding to nuclear stability. Also, there is *no* stable nuclide with $A = 5$ or $A = 8$. The doubly magic $^{4}_{2}\mathrm{He}$ nucleus, with a pair of protons and a pair of neutrons, has no interest in accepting a fifth particle into its structure. Collections of eight nucleons decay to smaller nuclides, with a $^{8}_{4}\mathrm{Be}$ nucleus immediately splitting into two $^{4}_{2}\mathrm{He}$ nuclei.

The stable nuclides define a rather narrow region on the Segrè chart. For low mass numbers, the numbers of protons and neutrons are approximately equal, $N \approx Z$. The ratio N/Z increases gradually with A, up to about 1.6 at large mass numbers, because of the increasing influence of the electrical repulsion of the protons. Points to the right of the stability region represent nuclides that have too many protons relative to neutrons. In these cases, repulsion wins, and the nucleus comes apart. To the left are nuclides with too many neutrons relative to protons. In these cases the energy associated with the neutrons is out of balance with that associated with the protons, and the nuclides decay in a process that converts neutrons to protons. The graph also shows that no nuclide with $A > 209$ or

$Z > 83$ is stable. A nucleus is unstable if it is too big. Note that there is no stable nuclide with $Z = 43$ (technetium) or 61 (promethium).

Nearly 90% of the 2500 known nuclides are *radioactive*; they are not stable but decay into other nuclides. Many of these radioactive nuclides occur in nature. For example, you are very slightly radioactive because of unstable nuclides such as carbon-14 ($^{14}_{6}C$) and potassium-40 ($^{40}_{19}K$) that are present throughout your body. The study of radioactivity began in 1896, one year after Wilhelm Röntgen discovered x rays (Section 36.6). Henri Becquerel discovered a radiation from uranium salts that seemed similar to x rays. Investigation in the following two decades by Marie and Pierre Curie, Ernest Rutherford, and many others revealed that the emissions consist of positively and negatively charged particles and neutral rays. These particles were given the names *alpha, beta,* and *gamma* because of their differing penetration characteristics.

MP
PhET: Alpha Decay

Alpha Decay

When unstable nuclides decay into different nuclides, they usually emit alpha (α) or beta (β) particles. An **alpha particle** is a ^4He nucleus, two protons and two neutrons bound together, with total spin zero. Alpha decay occurs principally with nuclei that are too large to be stable. When a nucleus emits an alpha particle, its N and Z values each decrease by 2 and A decreases by 4, moving it closer to stable territory on the Segrè chart.

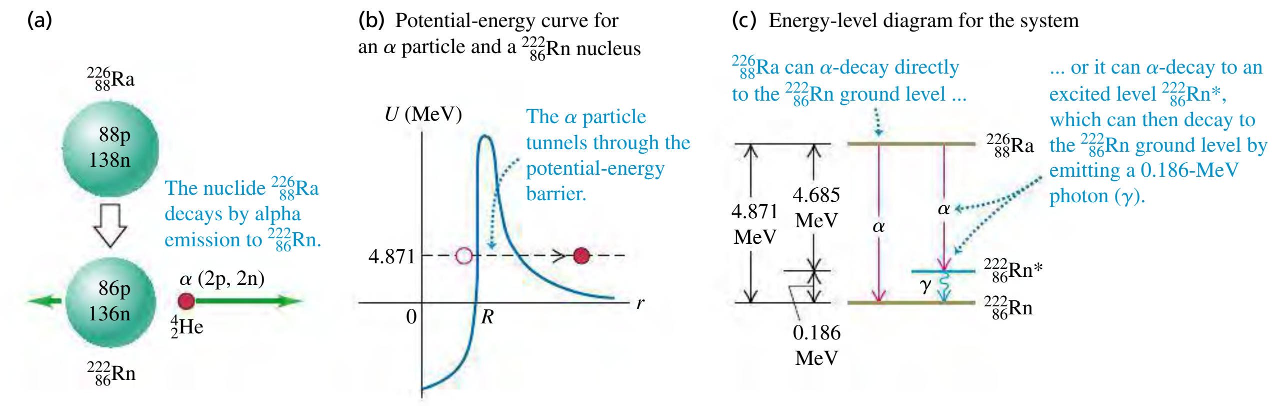
Figure 43.5a shows the alpha decay of radium-226 ($^{226}_{88}Ra$). Spontaneous alpha decay of this kind can occur only if energy is released in the process; this released energy goes into the kinetic energy of the emitted α particle and of the nucleus that remains, called the *daughter nucleus.* (For the decay shown in Fig. 43.5a, the daughter nucleus is radon-222, $^{222}_{86}Rn$.) The original nucleus (in this case, $^{226}_{88}Ra$) is called the *parent nucleus.* You can use mass-energy conservation to show that

alpha decay is possible whenever the mass of the original neutral atom is greater than the sum of the masses of the final neutral atom and a neutral ^4_2He atom.

In alpha decay, the α particle tunnels through a potential-energy barrier, as Fig. 43.5b shows. You may want to review the discussion of tunneling in Section 40.4.

Alpha particles are always emitted with definite kinetic energies, determined by conservation of momentum and energy in the alpha-decay process. As Fig. 43.5c shows, an α particle emitted in the decay of $^{226}_{88}Ra$ can have either of *two* possible energies, depending on the energy level of the $^{222}_{86}Rn$ daughter nucleus just after the decay. (Later in this section we'll discuss the photon-emission process shown in Fig. 43.5c.)

43.5 Alpha decay of the unstable radium nuclide $^{226}_{88}Ra$. The alpha particles used in the Rutherford scattering experiment (Section 39.2) were emitted by this nuclide.



Alpha particles are emitted at high speeds, typically a few percent of the speed of light (see the following example). Nonetheless, because of their charge and mass, alpha particles can travel only several centimeters in air, or a few tenths or hundredths of a millimeter through solids, before they are brought to rest by collisions.

SOLUTION

EXAMPLE 43.5 ALPHA DECAY OF RADIUM

Show that the α-emission process ${}^{226}_{88}\mathrm{Ra} \rightarrow {}^{222}_{86}\mathrm{Rn} + {}^{4}_{2}\mathrm{He}$ (Fig. 43.5a) is energetically possible, and calculate the kinetic energy of the emitted α particle. The neutral atomic masses are 226.025410 u for ${}^{226}_{88}\mathrm{Ra}$, 222.017578 u for ${}^{222}_{86}\mathrm{Rn}$, and 4.002603 u for ${}^{4}_{2}\mathrm{He}$.

SOLUTION

IDENTIFY and SET UP: Alpha emission is possible if the mass of the ${}^{226}_{88}\mathrm{Ra}$ atom is greater than the sum of the atomic masses of ${}^{222}_{86}\mathrm{Rn}$ and ${}^{4}_{2}\mathrm{He}$. The mass difference between the initial radium atom and the final radon and helium atoms corresponds (through $E = mc^2$) to the energy E released in the decay. Because momentum is conserved as well as energy, *both* the alpha particle and the ${}^{222}_{86}\mathrm{Rn}$ atom are in motion after the decay; we will have to account for this in determining the kinetic energy of the alpha particle.

EXECUTE: The difference in mass between the original nucleus and the decay products is

$$226.025410\ \mathrm{u} - (222.017578\ \mathrm{u} + 4.002603\ \mathrm{u}) = +0.005229\ \mathrm{u}$$

Since this is positive, α decay is energetically possible. The energy equivalent of this mass difference is

$$E = (0.005229\ \mathrm{u})(931.5\ \mathrm{MeV/u}) = 4.871\ \mathrm{MeV}$$

In this process the ${}^{222}_{86}\mathrm{Rn}$ nucleus is produced in its ground level (Fig. 43.5c). Thus we expect the decay products to emerge with total kinetic energy 4.871 MeV. Momentum is also conserved; if the parent ${}^{226}_{88}\mathrm{Ra}$ nucleus is at rest, the daughter ${}^{222}_{86}\mathrm{Rn}$ nucleus and the α particle have momenta of equal magnitude p but opposite direction. Kinetic energy is $K = \frac{1}{2}mv^2 = p^2/2m$: Since p is the same for the two particles, the kinetic energy divides inversely as their masses. Hence the α particle gets $222/(222 + 4)$ of the total, or 4.78 MeV.

EVALUATE: Experiment shows that ${}^{226}_{88}\mathrm{Ra}$ does emit α particles with a kinetic energy of 4.78 MeV. Check your results by verifying that the alpha particle and the ${}^{222}_{86}\mathrm{Rn}$ nucleus produced in the decay have the same magnitude of momentum $p = mv$. You can calculate the speed v of each of the decay products from its respective kinetic energy [note that the ${}^{222}_{86}\mathrm{Rn}$ nucleus gets $4/(222 + 4)$ of the 4.871 MeV released]. You'll find that the alpha particle moves at $0.0506c = 1.52 \times 10^7\ \mathrm{m/s}$; if momentum is conserved, you should find that the ${}^{222}_{86}\mathrm{Rn}$ nucleus moves $\frac{4}{222}$ as fast. Does it?

Beta Decay

There are three different simple types of *beta decay: beta-minus, beta-plus,* and *electron capture*. A **beta-minus particle** (β^-) is an electron. There are no electrons in the nucleus waiting to be emitted; instead, emission of a β^- involves *transformation* of a neutron into a proton, an electron, and a third particle called an *antineutrino*. In fact, if you freed a neutron from a nucleus, it would decay into a proton, an electron, and an antineutrino in an average time of about 15 minutes.

Beta particles can be identified and their speeds can be measured with techniques that are similar to the Thomson e/m experiment we described in Section 27.5. The speeds of beta particles range up to 0.9995 of the speed of light, so their motion is highly relativistic. They are emitted with a continuous spectrum of energies. This would not be possible if the only two particles were the β^- and the recoiling nucleus, since energy and momentum conservation would then require a definite speed for the β^-. Thus there must be a *third* particle involved. From conservation of charge, it must be neutral, and from conservation of angular momentum, it must be a spin-$\frac{1}{2}$ particle.

This third particle is an antineutrino, the *antiparticle* of a **neutrino.** The symbol for a neutrino is ν_e (the Greek letter nu). Both the neutrino and the antineutrino have zero charge and very small mass and therefore produce very little observable effect when passing through matter. Both evaded detection until 1953, when Frederick Reines and Clyde Cowan succeeded in observing the antineutrino directly. We now know that there are at least three varieties of neutrinos, each with its corresponding antineutrino; one is associated with beta decay and the other two are associated with the decay of two unstable particles, the muon and the tau particle. We'll discuss these particles in more detail in Chapter 44. The antineutrino that is emitted in β^- decay is denoted as $\bar{\nu}_e$. The basic process of β^- decay is

$$n \longrightarrow p + \beta^- + \bar{\nu}_e \qquad (43.13)$$

Beta-minus decay usually occurs with nuclides for which the neutron-to-proton ratio N/Z is too large for stability. In β^- decay, N decreases by 1, Z increases by 1, and A doesn't change. You can use mass-energy conservation to show that

beta-minus decay can occur whenever the mass of the original neutral atom is larger than that of the final atom.

EXAMPLE 43.6 WHY COBALT-60 IS A BETA-MINUS EMITTER

SOLUTION

The nuclide ${}^{60}_{27}\text{Co}$, an odd-odd unstable nucleus, is used in medical and industrial applications of radiation. Show that it is unstable relative to β^- decay. The atomic masses you need are 59.933817 u for ${}^{60}_{27}\text{Co}$ and 59.930786 u for ${}^{60}_{28}\text{Ni}$.

SOLUTION

IDENTIFY and SET UP: Beta-minus decay is possible if the mass of the original neutral atom is greater than that of the final atom. We must first identify the nuclide that will result if ${}^{60}_{27}\text{Co}$ undergoes β^- decay and then compare its neutral atomic mass to that of ${}^{60}_{27}\text{Co}$.

EXECUTE: In the presumed β^- decay of ${}^{60}_{27}\text{Co}$, Z increases by 1 from 27 to 28 and A remains at 60, so the final nuclide is ${}^{60}_{28}\text{Ni}$. The neutral atomic mass of ${}^{60}_{27}\text{Co}$ is greater than that of ${}^{60}_{28}\text{Ni}$ by 0.003031 u, so β^- decay *can* occur.

EVALUATE: With three decay products in β^- decay—the ${}^{60}_{28}\text{Ni}$ nucleus, the electron, and the antineutrino—the energy can be shared in many different ways that are consistent with conservation of energy and momentum. It's impossible to predict precisely how the energy will be shared for the decay of a particular ${}^{60}_{27}\text{Co}$ nucleus. By contrast, in alpha decay there are just two decay products, and their energies and momenta are determined uniquely (see Example 43.5).

We have noted that β^- decay occurs with nuclides that have too large a neutron-to-proton ratio N/Z. Nuclides for which N/Z is too *small* for stability can emit a *positron,* the electron's antiparticle, which is identical to the electron but with positive charge. (We'll discuss the positron in more detail in Chapter 44.) The basic process, called *beta-plus decay* (β^+), is

$$p \longrightarrow n + \beta^+ + \nu_e \qquad (43.14)$$

where β^+ is a positron and ν_e is the electron neutrino.

Beta-plus decay can occur whenever the mass of the original neutral atom is at least two electron masses larger than that of the final atom.

You can show this by using mass-energy conservation.

The third type of beta decay is *electron capture*. There are a few nuclides for which β^+ emission is not energetically possible but in which an orbital electron (usually in the innermost K shell) can combine with a proton in the nucleus to form a neutron and a neutrino. The neutron remains in the nucleus and the neutrino is emitted. The basic process is

$$p + \beta^- \longrightarrow n + \nu_e \qquad (43.15)$$

You can use mass-energy conservation to show that

electron capture can occur whenever the mass of the original neutral atom is larger than that of the final atom.

In all types of beta decay, A remains constant. However, in beta-plus decay and electron capture, N increases by 1 and Z decreases by 1 as the neutron-to-proton ratio increases toward a more stable value. The reaction of Eq. (43.15) also helps explain the formation of a neutron star, mentioned in Example 43.1.

CAUTION **Beta decay inside and outside nuclei** The beta-decay reactions given by Eqs. (43.13), (43.14), and (43.15) occur *within* a nucleus. Although the decay of a neutron outside the nucleus proceeds through the reaction of Eq. (43.13), the reaction of Eq. (43.14) is forbidden by mass-energy conservation for a proton outside the nucleus. The reaction of Eq. (43.15) can occur outside the nucleus only with the addition of some extra energy, as in a collision.

EXAMPLE 43.7 WHY COBALT-57 IS NOT A BETA-PLUS EMITTER

SOLUTION

The nuclide $^{57}_{27}\mathrm{Co}$ is an odd-even unstable nucleus. Show that it cannot undergo β^+ decay, but that it *can* decay by electron capture. The atomic masses you need are 56.936291 u for $^{57}_{27}\mathrm{Co}$ and 56.935394 u for $^{57}_{26}\mathrm{Fe}$.

SOLUTION

IDENTIFY and SET UP: Beta-plus decay is possible if the mass of the original neutral atom is greater than that of the final atom plus two electron masses (0.001097 u). Electron capture is possible if the mass of the original atom is greater than that of the final atom. We must first identify the nuclide that will result if $^{57}_{27}\mathrm{Co}$ undergoes β^+ decay or electron capture and then find the corresponding mass difference.

EXECUTE: The original nuclide is $^{57}_{27}\mathrm{Co}$. In both the presumed β^+ decay and electron capture, Z decreases by 1 from 27 to 26, and A remains at 57, so the final nuclide is $^{57}_{26}\mathrm{Fe}$. Its mass is less than that of $^{57}_{27}\mathrm{Co}$ by 0.000897 u, a value smaller than 0.001097 u (two electron masses), so β^+ decay *cannot* occur. However, the mass of the original atom is greater than the mass of the final atom, so electron capture *can* occur.

EVALUATE: In electron capture there are just two decay products, the final nucleus and the emitted neutrino. As in alpha decay (Example 43.5) but unlike in β^- decay (Example 43.6), the decay products of electron capture have unique energies and momenta. In Section 43.4 we'll see how to relate the probability that electron capture will occur to the *half-life* of this nuclide.

Gamma Decay

The energy of internal motion of a nucleus is quantized. A typical nucleus has a set of allowed energy levels, including a *ground state* (state of lowest energy) and several *excited states*. Because of the great strength of nuclear interactions, excitation energies of nuclei are typically of the order of 1 MeV, compared with a few eV for atomic energy levels. In ordinary physical and chemical transformations the nucleus always remains in its ground state. When a nucleus is placed in an excited state, either by bombardment with high-energy particles or by a radioactive transformation, it can decay to the ground state by emission of one or more photons called **gamma rays** or *gamma-ray photons,* with typical energies of 10 keV to 5 MeV. This process is called *gamma* (γ) *decay*. For example, alpha particles emitted from $^{226}\mathrm{Ra}$ have two possible kinetic energies, either 4.784 MeV or 4.602 MeV. Including the recoil energy of the resulting $^{222}\mathrm{Rn}$ nucleus, these correspond to a total released energy of 4.871 MeV or 4.685 MeV, respectively (see Fig. 43.5c). When an alpha particle with the smaller energy is emitted, the $^{222}\mathrm{Rn}$ nucleus is left in an excited state. It then decays to its ground state by emitting a gamma-ray photon with energy

$$(4.871 - 4.685)\ \mathrm{MeV} = 0.186\ \mathrm{MeV}$$

CAUTION **γ decay vs. α and β decay** In both α and β decay, the Z value of a nucleus changes and the nucleus of one element becomes the nucleus of a different element. In γ decay, the element does *not* change; the nucleus merely goes from an excited state to a less excited state.

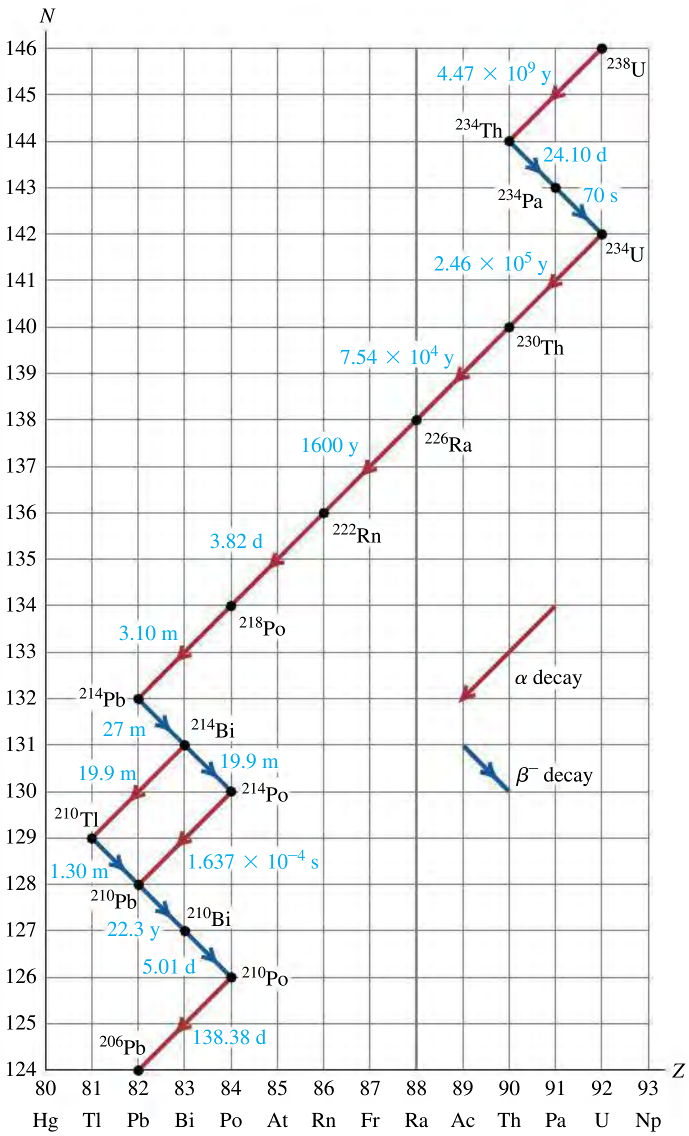
43.6 Earthquakes are caused in part by the radioactive decay of ^{238}U in the earth's interior. These decays release energy that helps produce convection currents in the earth's interior. Such currents drive the motions of the earth's crust, including the sudden sharp motions that we call earthquakes (like the one that caused this damage).

Radioactive Decay Series

When a radioactive nucleus decays, the resulting (daughter) nucleus may also be unstable. In this case a *series* of successive decays occurs until a stable configuration is reached. Several such series are found in nature. The most abundant radioactive nuclide found on earth is the uranium isotope ^{238}U, which undergoes a series of 14 decays, including eight α emissions and six β^- emissions, terminating at a stable isotope of lead, ^{206}Pb (**Fig. 43.6**).

Radioactive decay series can be represented on a Segrè chart, as in **Fig. 43.7**. The neutron number N is plotted vertically, and the atomic number Z is plotted horizontally. In alpha emission, both N and Z decrease by 2. In β^- emission, N decreases by 1 and Z increases by 1. The decays can also be represented in equation form; the first two decays in the series are written as

$$^{238}\mathrm{U} \rightarrow {}^{234}\mathrm{Th} + \alpha \qquad \text{and} \qquad {}^{234}\mathrm{Th} \rightarrow {}^{234}\mathrm{Pa} + \beta^- + \overline{\nu}_e$$



43.7 Segrè chart showing the uranium ^{238}U decay series, terminating with the stable nuclide ^{206}Pb. The times are half-lives (discussed in the next section), given in years (y), days (d), hours (h), minutes (m), or seconds (s).

DATA *SPEAKS*

Nuclear Decays

When students were given a problem involving radioactive decay, more than 24% gave an incorrect response. Common errors:

- Confusing alpha, beta, and gamma decays. In alpha decay, atomic number Z decreases by 2 and mass number A decreases by 4. In beta-minus decay, Z increases by 1 and A is unchanged; in beta-plus decay or electron capture, Z decreases by 1 and A is unchanged. In gamma decay, both Z and A remain unchanged; the final nucleus is simply a less excited state of the initial nucleus.
- Confusion about the fate of radioactive atoms. In alpha, beta, and gamma decays, atoms do not disappear; they transform into other atoms.

or more briefly as

$$^{238}\mathrm{U} \xrightarrow{\alpha} {}^{234}\mathrm{Th} \qquad \text{and} \qquad {}^{234}\mathrm{Th} \xrightarrow{\beta^-} {}^{234}\mathrm{Pa}$$

In the second process, the beta decay leaves the daughter nucleus ^{234}Pa in an excited state, from which it decays to the ground state by emitting a gamma-ray photon. An excited state is denoted by an asterisk, so we can represent the γ emission as

$$^{234}\mathrm{Pa}^* \rightarrow {}^{234}\mathrm{Pa} + \gamma \qquad \text{or} \qquad {}^{234}\mathrm{Pa}^* \xrightarrow{\gamma} {}^{234}\mathrm{Pa}$$

An interesting feature of the ^{238}U decay series is the branching that occurs at ^{214}Bi. This nuclide decays to ^{210}Pb by emission of an α and a β^-, which can occur in either order. We also note that the series includes unstable isotopes of several elements that also have stable isotopes, including thallium (Tl), lead (Pb), and bismuth (Bi). The unstable isotopes of these elements that occur in the ^{238}U series all have too many neutrons to be stable.

Many other decay series are known. Two of these occur in nature, one starting with the uncommon isotope ^{235}U and ending with ^{207}Pb, the other starting with thorium (^{232}Th) and ending with ^{208}Pb.

TEST YOUR UNDERSTANDING OF SECTION 43.3 A nucleus with atomic number Z and neutron number N undergoes two decay processes. The result is a nucleus with atomic number $Z - 3$ and neutron number $N - 1$. Which decay processes may have taken place? (i) Two β^- decays; (ii) two β^+ decays; (iii) two α decays; (iv) an α decay and a β^- decay; (v) an α decay and a β^+ decay.

43.4 ACTIVITIES AND HALF-LIVES

Suppose you have a certain number of nuclei of a particular radioactive nuclide. If no more are produced, that number decreases in a simple manner as the nuclei decay. This decrease is a statistical process; there is no way to predict when any individual nucleus will decay. No change in physical or chemical environment, such as chemical reactions or heating or cooling, greatly affects most decay rates. The rate varies over an extremely wide range for different nuclides.

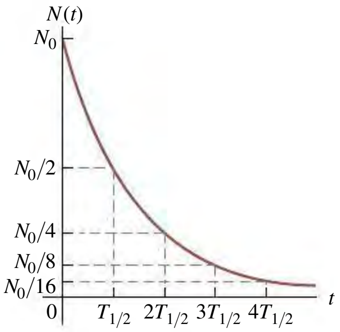
Radioactive Decay Rates

Let $N(t)$ be the (very large) number of radioactive nuclei in a sample at time t, and let $dN(t)$ be the (negative) change in that number during a short time interval dt. (We'll use $N(t)$ to minimize confusion with the neutron number N.) The number of decays during the interval dt is $-dN(t)$. The rate of change of $N(t)$ is the negative quantity $dN(t)/dt$; thus $-dN(t)/dt$ is called the *decay rate* or the **activity** of the specimen. The larger the number of nuclei in the specimen, the more nuclei decay during any time interval. That is, the activity is directly proportional to $N(t)$; it equals a constant λ multiplied by $N(t)$:

$$-\frac{dN(t)}{dt} = \lambda N(t) \qquad (43.16)$$

The constant λ is called the **decay constant,** and it has different values for different nuclides. A large value of λ corresponds to rapid decay; a small value corresponds to slower decay. Solving Eq. (43.16) for λ shows us that λ is the ratio of the number of decays per time to the number of remaining radioactive nuclei; λ can then be interpreted as the *probability per unit time* that any individual nucleus will decay.

43.8 The number of nuclei in a sample of a radioactive element as a function of time. The sample's activity has an exponential decay curve with the same shape.



The situation is reminiscent of a discharging capacitor, which we studied in Section 26.4. Equation (43.16) has the same form as the negative of Eq. (26.15), with q and $1/RC$ replaced by $N(t)$ and λ. Then we can make the same substitutions in Eq. (26.16), with the initial number of nuclei $N(0) = N_0$, to find

Number of remaining nuclei at time *t* in sample of radioactive element

Number of nuclei at $t = 0$

$$N(t) = N_0 e^{-\lambda t} \tag{43.17}$$

Time

Decay constant

Figure 43.8 is a graph of this function.

The **half-life** $T_{1/2}$ is the time required for the number of radioactive nuclei to decrease to one-half the original number N_0. Then half of the remaining radioactive nuclei decay during a second interval $T_{1/2}$, and so on. The numbers remaining after successive half-lives are $N_0/2, N_0/4, N_0/8, \ldots$.

To get the relationship between the half-life $T_{1/2}$ and the decay constant λ, we set $N(t)/N_0 = \frac{1}{2}$ and $t = T_{1/2}$ in Eq. (43.17), obtaining

$$\tfrac{1}{2} = e^{-\lambda T_{1/2}}$$

We take logarithms of both sides and solve for $T_{1/2}$:

$$T_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda} \tag{43.18}$$

The mean lifetime T_{mean}, generally called the *lifetime,* of an unstable nucleus or particle is proportional to the half-life $T_{1/2}$:

Half-life of nucleus or particle

Lifetime of unstable nucleus or particle

$$T_{\text{mean}} = \frac{1}{\lambda} = \frac{T_{1/2}}{\ln 2} = \frac{T_{1/2}}{0.693} \tag{43.19}$$

Decay constant of nucleus or particle

In particle physics the life of an unstable particle is usually described by the lifetime, not the half-life.

Because the activity $-dN(t)/dt$ at any time equals $\lambda N(t)$, Eq. (43.17) tells us that the activity also depends on time as $e^{-\lambda t}$. Thus the graph of activity versus time has the same shape as Fig. 43.8. Also, after successive half-lives, the activity is one-half, one-fourth, one-eighth, and so on of the original activity.

CAUTION **A half-life may not be enough** It is sometimes implied that any radioactive sample will be safe after a half-life has passed. That's wrong. If your radioactive waste initially has ten times too much activity for safety, it is not safe after one half-life, when it still has five times too much. Even after three half-lives it still has 25% more activity than is safe. The number of radioactive nuclei and the activity approach zero only as t approaches infinity.

A common unit of activity is the **curie,** abbreviated Ci, which is defined to be 3.70×10^{10} decays per second. This is approximately equal to the activity of one gram of radium-226. The SI unit of activity is the *becquerel,* abbreviated Bq. One becquerel is one decay per second, so

$$1\ \text{Ci} = 3.70 \times 10^{10}\ \text{Bq} = 3.70 \times 10^{10}\ \text{decays/s}$$

EXAMPLE 43.8 ACTIVITY OF ^{57}Co

The isotope ^{57}Co decays by electron capture to ^{57}Fe with a half-life of 272 d. The ^{57}Fe nucleus is produced in an excited state, and it almost instantaneously emits gamma rays that we can detect. (a) Find the mean lifetime and decay constant for ^{57}Co. (b) If the activity of a ^{57}Co radiation source is now 2.00 μCi, how many ^{57}Co nuclei does the source contain? (c) What will be the activity after one year?

SOLUTION

IDENTIFY and SET UP: This problem uses the relationships among decay constant λ, lifetime T_{mean}, and activity $-dN(t)/dt$. In part (a) we use Eq. (43.19) to find λ and T_{mean} from $T_{1/2}$. In part (b), we use Eq. (43.16) to calculate the number of nuclei $N(t)$ from the activity. Finally, in part (c) we use Eqs. (43.16) and (43.17) to find the activity after one year.

EXECUTE: (a) It's convenient to convert the half-life from days to seconds:

$$T_{1/2} = (272\ \text{d})(86{,}400\ \text{s/d})$$
$$= 2.35 \times 10^7\ \text{s}$$

From Eq. (43.19), we find that the mean lifetime and the decay constant are

$$T_{\text{mean}} = \frac{T_{1/2}}{\ln 2} = \frac{2.35 \times 10^7\ \text{s}}{0.693}$$
$$= 3.39 \times 10^7\ \text{s} = 392\ \text{days}$$
$$\lambda = \frac{1}{T_{\text{mean}}} = 2.95 \times 10^{-8}\ \text{s}^{-1}$$

(b) The activity $-dN(t)/dt$ is given as 2.00 μCi, so

$$-\frac{dN(t)}{dt} = 2.00\ \mu\text{Ci} = (2.00 \times 10^{-6})(3.70 \times 10^{10}\ \text{s}^{-1})$$
$$= 7.40 \times 10^4\ \text{decays/s}$$

From Eq. (43.16) this is equal to $\lambda N(t)$, so we find

$$N(t) = -\frac{dN(t)/dt}{\lambda} = \frac{7.40 \times 10^4\ \text{s}^{-1}}{2.95 \times 10^{-8}\ \text{s}^{-1}}$$
$$= 2.51 \times 10^{12}\ \text{nuclei}$$

If you feel we're being too cavalier about the "units" decays and nuclei, you can use decays/(nucleus $\cdot$ s) as the unit for λ.

(c) From Eq. (43.17) the number $N(t)$ of nuclei remaining after one year $(3.156 \times 10^7\ \text{s})$ is

$$N(t) = N_0 e^{-\lambda t} = N_0 e^{-(2.95\times10^{-8}\,\text{s}^{-1})(3.156\times10^{7}\,\text{s})}$$
$$= 0.394 N_0$$

The number of nuclei has decreased to 0.394 of the original number. Equation (43.16) says that the activity is proportional to the number of nuclei, so the activity has decreased by this same factor to $(0.394)(2.00\ \mu\text{Ci}) = 0.788\ \mu\text{Ci}$.

EVALUATE: The number of nuclei found in part (b) is equivalent to 4.17×10^{-12} mol, with a mass of 2.38×10^{-10} g. This is a far smaller mass than even the most sensitive balance can measure.

After one 272-day half-life, the number of ^{57}Co nuclei has decreased to $N_0/2$; after $2(272\ \text{d}) = 544$ d, it has decreased to $N_0/2^2 = N_0/4$. This result agrees with our answer to part (c), which says that after 365 d the number of nuclei is between $N_0/2$ and $N_0/4$.

Radioactive Dating

An important application of radioactivity is the dating of archaeological and geological specimens by measuring the concentration of radioactive isotopes. The most familiar example is *carbon dating*. The unstable isotope ^{14}C, produced during nuclear reactions in the atmosphere that result from cosmic-ray bombardment, gives a small proportion of ^{14}C in the CO_2 in the atmosphere. Plants that obtain their carbon from this source contain the same proportion of ^{14}C as the atmosphere. When a plant dies, it stops taking in carbon, and its ^{14}C β^- decays to ^{14}N with a half-life of 5730 years. By measuring the proportion of ^{14}C in the remains, we can determine how long ago the organism died.

One difficulty with radiocarbon dating is that the ^{14}C concentration in the atmosphere changes over long time intervals. Corrections can be made on the basis of other data such as measurements of tree rings that show annual growth cycles. Similar radioactive techniques are used with other isotopes for dating geological specimens. Some rocks, for example, contain the unstable potassium isotope ^{40}K, a beta emitter that decays to the stable nuclide ^{40}Ar with a half-life of 2.4×10^8 y. The age of the rock can be determined by comparing the concentrations of ^{40}K and ^{40}Ar.

EXAMPLE 43.9 RADIOCARBON DATING

SOLUTION

Before 1900 the activity per unit mass of atmospheric carbon due to the presence of ^{14}C averaged about 0.255 Bq per gram of carbon. (a) What fraction of carbon atoms were ^{14}C? (b) In analyzing an archaeological specimen containing 500 mg of carbon, you observe 174 decays in one hour. What is the age of the specimen, assuming that its activity per unit mass of carbon when it died was that average value of the air?

SOLUTION

IDENTIFY and SET UP: The key idea is that the present-day activity of a biological sample containing ^{14}C is related to both the elapsed time since it stopped taking in atmospheric carbon and its activity at that time. We use Eqs. (43.16) and (43.17) to solve for the age t of the specimen. In part (a) we determine the number of ^{14}C atoms $N(t)$ from the activity $-dN(t)/dt$ by using Eq. (43.16). We find the total number of carbon atoms in 500 mg by using the molar mass of carbon (12.011 g/mol, given in Appendix D), and we use the result to calculate the fraction of carbon atoms that are ^{14}C. The activity decays at the same rate as the number of ^{14}C nuclei; we use this and Eq. (43.17) to solve for the age t of the specimen.

EXECUTE: (a) To use Eq. (43.16), we must first find the decay constant λ from Eq. (43.18):

$$T_{1/2} = 5730\text{ y} = (5730\text{ y})(3.156 \times 10^7\text{ s/y}) = 1.808 \times 10^{11}\text{ s}$$

$$\lambda = \frac{\ln 2}{T_{1/2}} = \frac{0.693}{1.808 \times 10^{11}\text{ s}} = 3.83 \times 10^{-12}\text{ s}^{-1}$$

Then, from Eq. (43.16),

$$N(t) = \frac{-dN/dt}{\lambda} = \frac{0.255\text{ s}^{-1}}{3.83 \times 10^{-12}\text{ s}^{-1}} = 6.65 \times 10^{10}\text{ atoms}$$

The *total* number of C atoms in 1 gram (1/12.011 mol) is $(1/12.011)(6.022 \times 10^{23}) = 5.01 \times 10^{22}$. The ratio of ^{14}C atoms to all C atoms is

$$\frac{6.65 \times 10^{10}}{5.01 \times 10^{22}} = 1.33 \times 10^{-12}$$

Only four carbon atoms in every 3×10^{12} are ^{14}C.

(b) Assuming that the activity per gram of carbon in the specimen when it died ($t = 0$) was $0.255\text{ Bq/g} = (0.255\text{ s}^{-1} \cdot \text{g}^{-1}) \times (3600\text{ s/h}) = 918\text{ h}^{-1} \cdot \text{g}^{-1}$, the activity of 500 mg of carbon then was $(0.500\text{ g})(918\text{ h}^{-1} \cdot \text{g}^{-1}) = 459\text{ h}^{-1}$. The observed activity now, at time t, is 174 h^{-1}. Since the activity is proportional to the number of radioactive nuclei, the activity ratio $174/459 = 0.379$ equals the number ratio $N(t)/N_0$.

Now we solve Eq. (43.17) for t and insert values for $N(t)/N_0$ and λ:

$$t = \frac{\ln(N(t)/N_0)}{-\lambda} = \frac{\ln 0.379}{-3.83 \times 10^{-12}\text{ s}^{-1}} = 2.53 \times 10^{11}\text{ s} = 8020\text{ y}$$

EVALUATE: After 8020 y the ^{14}C activity has decreased from 459 to 174 decays per hour. The specimen died and stopped taking CO_2 out of the air about 8000 years ago.

Radiation in the Home

A serious health hazard in some areas is the accumulation in houses of ^{222}Rn, an inert, colorless, odorless radioactive gas. Looking at the ^{238}U decay chain in Fig. 43.7, we see that the half-life of ^{222}Rn is 3.82 days. If so, why not just move out of the house for a while and let it decay away? The answer is that ^{222}Rn is continuously being *produced* by the decay of ^{226}Ra, which is found in minute quantities in the rocks and soil on which some houses are built. It's a dynamic equilibrium situation, in which the rate of production equals the rate of decay. The reason ^{222}Rn is a bigger hazard than the other elements in the ^{238}U decay series is that it's a gas. During its short half-life of 3.82 days it can migrate from the soil into your house. If a ^{222}Rn nucleus decays in your lungs, it emits a damaging α particle and its daughter nucleus ^{218}Po, which is *not* chemically inert and is likely to stay in your lungs until it decays, emits another damaging α particle and so on down the ^{238}U decay series.

How much of a hazard is radon? Although reports indicate values as high as 3500 pCi/L, the average activity per volume in the air inside American homes due to ^{222}Rn is about 1.5 pCi/L (over a thousand decays each second in an average-sized room). *If* your environment has this level of activity, it has been estimated that a lifetime exposure would reduce your life expectancy by about 40 days. For comparison, smoking one pack of cigarettes per day reduces life expectancy by 6 years, and it is estimated that the average emission from all the nuclear power plants in the world reduces life expectancy by anywhere from 0.01 day to 5 days. These figures include catastrophes such as the nuclear reactor disasters at Chernobyl, Ukraine (1986), and Fukushima, Japan (2011), for which the *local* effect on life expectancy is much greater.

TEST YOUR UNDERSTANDING OF SECTION 43.4 Which sample contains a greater number of nuclei: a 5.00-μCi sample of ^{240}Pu (half-life 6560 y) or a 4.45-μCi sample of ^{243}Am (half-life 7370 y)? (i) The ^{240}Pu sample; (ii) the ^{243}Am sample; (iii) both have the same number of nuclei.

43.5 BIOLOGICAL EFFECTS OF RADIATION

The above discussion of radon introduced the interaction of radiation with living organisms, a topic of vital interest and importance. Under *radiation* we include radioactivity (alpha, beta, gamma, and neutrons) and electromagnetic radiation such as x rays. As these particles pass through matter, they lose energy, breaking molecular bonds and creating ions—hence the term *ionizing radiation*. Charged particles interact directly with the electrons in the material. X rays and γ rays interact by the photoelectric effect, in which an electron absorbs a photon and breaks loose from its site, or by Compton scattering (see Section 38.3). Neutrons cause ionization indirectly through collisions with nuclei or absorption by nuclei with subsequent radioactive decay of the resulting nuclei.

These interactions are extremely complex. It is well known that excessive exposure to radiation, including sunlight, x rays, and all the nuclear radiations, can destroy tissues. In mild cases it results in a burn, as with common sunburn. Greater exposure can cause very severe illness or death by a variety of mechanisms, including massive destruction of tissue cells, alterations of genetic material, and destruction of the components in bone marrow that produce red blood cells.

Calculating Radiation Doses

Radiation dosimetry is the quantitative description of the effect of radiation on living tissue. The *absorbed dose* of radiation is defined as the energy delivered to the tissue per unit mass. The SI unit of absorbed dose, the joule per kilogram, is called the *gray* (Gy); 1 Gy = 1 J/kg. Another unit is the *rad,* defined as

$$1 \text{ rad} = 0.01 \text{ J/kg} = 0.01 \text{ Gy}$$

Absorbed dose by itself is not an adequate measure of biological effect because equal energies of different kinds of radiation cause different extents of biological effect. This variation is described by a numerical factor called the **relative biological effectiveness (RBE)**, also called the *quality factor* (QF), of each specific radiation. X rays with 200 keV of energy are defined to have an RBE of unity, and the effects of other radiations can be compared experimentally. **Table 43.3** shows approximate values of RBE for several radiations. All these values depend somewhat on the kind of tissue in which the radiation is absorbed and on the energy of the radiation.

TABLE 43.3 Relative Biological Effectiveness (RBE) for Several Types of Radiation

Radiation	RBE (Sv/Gy or rem/rad)
X rays and γ rays	1
Electrons	1.0–1.5
Slow neutrons	3–5
Protons	10
α particles	20
Heavy ions	20

The biological effect is described by the product of the absorbed dose and the RBE of the radiation; this quantity is called the *biologically equivalent dose,* or simply the equivalent dose. The SI unit of equivalent dose for humans is the sievert (Sv):

$$\text{Equivalent dose (Sv)} = \text{RBE} \times \text{Absorbed dose (Gy)} \qquad (43.20)$$

A more common unit, corresponding to the rad, is the rem (an abbreviation of *röntgen equivalent for man*):

$$\text{Equivalent dose (rem)} = \text{RBE} \times \text{Absorbed dose (rad)} \qquad (43.21)$$

Thus the unit of the RBE is 1 Sv/Gy or 1 rem/rad, and 1 rem = 0.01 Sv.

EXAMPLE 43.10 DOSE FROM A MEDICAL X RAY

During a diagnostic x-ray examination a 1.2-kg portion of a broken leg receives an equivalent dose of 0.40 mSv. (a) What is the equivalent dose in mrem? (b) What is the absorbed dose in mrad and in mGy? (c) If the x-ray energy is 50 keV, how many x-ray photons are absorbed?

SOLUTION

IDENTIFY and SET UP: We are asked to relate the equivalent dose (the biological effect of the radiation, measured in sieverts or rems) to the absorbed dose (the energy absorbed per mass, measured in grays or rads). In part (a) we use the conversion factor 1 rem = 0.01 Sv for equivalent dose. Table 43.3 gives the RBE for x rays; we use this value in part (b) to determine the absorbed dose from Eqs. (43.20) and (43.21). Finally, in part (c) we use the mass and the definition of absorbed dose to find the total energy absorbed and the total number of photons absorbed.

EXECUTE: (a) The equivalent dose in mrem is

$$\frac{0.40\ \text{mSv}}{0.01\ \text{Sv/rem}} = 40\ \text{mrem}$$

(b) For x rays, RBE = 1 rem/rad or 1 Sv/Gy, so the absorbed dose is

$$\frac{40\ \text{mrem}}{1\ \text{rem/rad}} = 40\ \text{mrad}$$

$$\frac{0.40\ \text{mSv}}{1\ \text{Sv/Gy}} = 0.40\ \text{mGy} = 4.0 \times 10^{-4}\ \text{J/kg}$$

(c) The total energy absorbed is

$$(4.0 \times 10^{-4}\ \text{J/kg})(1.2\ \text{kg}) = 4.8 \times 10^{-4}\ \text{J} = 3.0 \times 10^{15}\ \text{eV}$$

The number of x-ray photons is

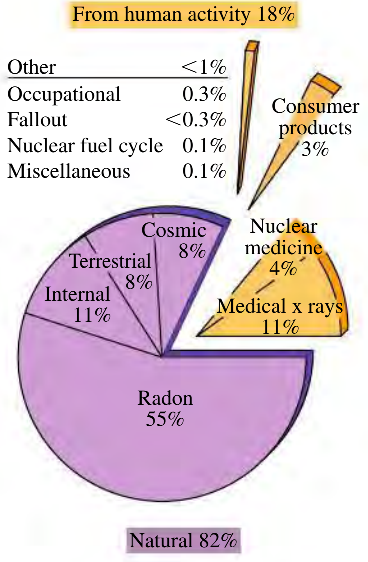
$$\frac{3.0 \times 10^{15}\ \text{eV}}{5.0 \times 10^{4}\ \text{eV/photon}} = 6.0 \times 10^{10}\ \text{photons}$$

EVALUATE: The absorbed dose is relatively large because x rays have a low RBE. If the ionizing radiation had been a beam of α particles, for which RBE = 20, the absorbed dose needed for an equivalent dose of 0.40 mSv would be only 0.020 mGy, corresponding to a smaller total absorbed energy of 2.4×10^{-5} J.

Radiation Hazards

Here are a few numbers for perspective. To convert from Sv to rem, simply multiply by 100. An ordinary chest x-ray exam delivers about 0.20–0.40 mSv to about 5 kg of tissue. Radiation exposure from cosmic rays and natural radioactivity in soil, building materials, and so on is of the order of 2–3 mSv per year at sea level and twice that at an elevation of 1500 m (5000 ft). A whole-body dose of up to about 0.20 Sv causes no immediately detectable effect. A short-term whole-body dose of 5 Sv or more usually causes death within a few days or weeks. A localized dose of 100 Sv causes complete destruction of the exposed tissues.

43.9 Contribution of various sources to the total average radiation exposure in the U.S. population, expressed as percentages of the total.



The long-term hazards of radiation exposure in causing various cancers and genetic defects have been widely publicized, and the question of whether there is any "safe" level of radiation exposure has been hotly debated. U.S. government regulations are based on a maximum *yearly* exposure, from all except natural resources, of 2 to 5 mSv. Workers with occupational exposure to radiation are permitted an average of 20 mSv per year. Recent studies suggest that these limits are too high and that even extremely small exposures carry hazards, but it is very difficult to gather reliable statistics on the effects of low doses. It has become clear that any use of x rays for medical diagnosis should be preceded by a very careful estimation of the relationship of risk to possible benefit.

Another sharply debated question is that of radiation hazards from nuclear power plants. The radiation level from these plants is *not* negligible. However, to make a meaningful evaluation of hazards, we must compare these levels with the alternatives, such as coal-powered plants. The health hazards of coal smoke are serious and well documented, and the natural radioactivity in the smoke from a coal-fired power plant is believed to be roughly 100 times as great as that from a properly operating nuclear plant with equal capacity. But the comparison is not this simple; the possibility of a nuclear accident and the very serious problem of safe disposal of radioactive waste from nuclear plants must also be considered. **Figure 43.9** shows one estimate of the various sources of radiation exposure for the U.S. population. Ionizing radiation is a two-edged sword; it poses very serious health hazards, yet it also provides many benefits to humanity, including the diagnosis and treatments of disease and a wide variety of analytical techniques.

Beneficial Uses of Radiation

Radiation is widely used in medicine for intentional selective destruction of tissue such as tumors. The hazards are considerable, but if the disease would be fatal without treatment, any hazard may be preferable. Artificially produced isotopes are often used as radiation sources. Such isotopes have several advantages over naturally radioactive isotopes. They may have shorter half-lives and correspondingly greater activity. Isotopes can be chosen that emit the type and energy of radiation desired. Some artificial isotopes have been replaced by photon, proton, and electron beams from linear accelerators.

Nuclear medicine is an expanding field of application. Radioactive isotopes have virtually the same electron configurations and resulting chemical behavior as stable isotopes of the same element. But the location and concentration of radioactive isotopes can easily be detected by measurements of the radiation they emit. A familiar example is the use of radioactive iodine for thyroid studies. Nearly all the iodine ingested is either eliminated or stored in the thyroid, and the body's chemical reactions do not discriminate between the unstable isotope ^{131}I and the stable isotope ^{127}I. A minute quantity of ^{131}I is fed or injected into the patient, and the speed with which it becomes concentrated in the thyroid provides a measure of thyroid function. The half-life is 8.02 days, so there are no long-lasting radiation hazards. By use of more sophisticated scanning detectors, one can also obtain a "picture" of the thyroid, which shows enlargement and other abnormalities. This procedure, a type of *autoradiography,* is comparable to photographing the glowing filament of an incandescent light bulb by using the light emitted by the filament itself. If this process discovers cancerous thyroid nodules, they can be destroyed by much larger quantities of ^{131}I.

Another useful nuclide for nuclear medicine is technetium-99 (^{99}Tc), which is formed in an excited state by the β^- decay of molybdenum (^{99}Mo). The technetium then decays to its ground state by emitting a γ-ray photon with energy 143 keV. The half-life is 6.01 hours, unusually long for γ emission. (The ground state of ^{99}Tc is also unstable, with a half-life of 2.11×10^5 y; it decays by β^- emission to the stable ruthenium nuclide ^{99}Ru.) The chemistry of technetium is such that it can readily be attached to organic molecules that are taken up by various organs of the body. A small quantity of such technetium-bearing molecules is injected into a patient, and a scanning detector or *gamma camera* is used to produce an image, or *scintigram,* that reveals which parts of the body take up these γ-emitting molecules. This technique, in which ^{99}Tc acts as a radioactive *tracer,* plays an important role in locating cancers, embolisms, and other pathologies (**Fig. 43.10**).

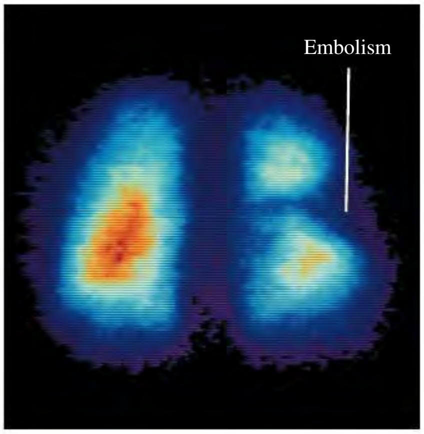
Tracer techniques have many other applications. Tritium (^3H), a radioactive hydrogen isotope, is used to tag molecules in complex organic reactions; radioactive tags on pesticide molecules, for example, can be used to trace their passage through food chains. In the world of machinery, radioactive iron can be used to study piston-ring wear. Laundry detergent manufacturers have even used radioactive dirt to test the effectiveness of their products.

Many direct effects of radiation are also useful, such as strengthening polymers by cross-linking, sterilizing surgical tools, dispersing unwanted static electricity in the air, and intentionally ionizing the air in smoke detectors. Gamma rays are also used to sterilize and preserve some food products.

TEST YOUR UNDERSTANDING OF SECTION 43.5 Alpha particles have 20 times the relative biological effectiveness of 200-keV x rays. Which would be better to use to radiate tissue deep inside the body? (i) A beam of alpha particles; (ii) a beam of 200-keV x rays; (iii) both are equally effective.

BIO Application A Radioactive Building The United States Capitol building in Washington, DC, is made of granite that contains a small amount of naturally radioactive uranium. As a result, the radiation exposure to someone working in the Capitol is 0.85 mSv per year. The health effects of this are negligible; it is estimated that a person who spent 20 years inside the Capitol would have an extra 0.1% chance of getting cancer, above the 10% chance due to all other causes during the same 20-year period.

43.10 This colored scintigram shows where a chemical containing radioactive ^{99}Tc was taken up by a patient's lungs. The orange color in the lung on the left indicates strong γ-ray emission by the ^{99}Tc, which shows that the chemical was able to pass into this lung through the bloodstream. The lung on the right shows weaker emission, indicating the presence of an embolism (a blood clot or other obstruction in an artery) that is restricting the flow of blood to this lung.



43.6 NUCLEAR REACTIONS

In the preceding sections we studied the decay of unstable nuclei, especially spontaneous emission of an α or β particle, sometimes followed by γ emission. Nothing needs to be done to initiate this decay, and nothing can be done to control it. This section examines some *nuclear reactions,* rearrangements of nuclear components that result from a bombardment by a particle rather than a spontaneous natural process. Rutherford suggested in 1919 that a massive particle with sufficient kinetic energy might be able to penetrate a nucleus. The result would be either a new nucleus with greater atomic number and mass number or a decay of the original nucleus. Rutherford verified this when he bombarded nitrogen $(^{14}\mathrm{N})$ with α particles and obtained an oxygen $(^{17}\mathrm{O})$ nucleus and a proton:

$$^4_2\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{17}_{8}\mathrm{O} + {}^1_1\mathrm{H} \qquad (43.22)$$

Rutherford used alpha particles from naturally radioactive sources. In Chapter 44 we'll describe some of the particle accelerators that are now used to initiate nuclear reactions.

Nuclear reactions are subject to several *conservation laws*. The classical conservation principles for charge, momentum, angular momentum, and energy (including rest energies) are obeyed in all nuclear reactions. An additional conservation law, not anticipated by classical physics, is conservation of the total number of nucleons. The numbers of protons and neutrons need not be conserved separately; in β decay, neutrons and protons change into one another. We'll study the basis of the conservation of nucleon number in Chapter 44.

When two nuclei interact, charge conservation requires that the sum of the initial atomic numbers must equal the sum of the final atomic numbers. Because of conservation of nucleon number, the sum of the initial mass numbers must also equal the sum of the final mass numbers. In general, these are *not* elastic collisions, and the total initial mass does *not* equal the total final mass.

Reaction Energy

The difference between the masses before and after the reaction corresponds to the **reaction energy,** according to the mass–energy relationship $E = mc^2$. If initial particles A and B interact to produce final particles C and D, the reaction energy Q is defined as

$$Q = (M_A + M_B - M_C - M_D)c^2 \quad \text{(reaction energy)} \qquad (43.23)$$

To balance the electrons, we use the neutral atomic masses in Eq. (43.23). That is, we use the mass of $^1_1\mathrm{H}$ for a proton, $^2_1\mathrm{H}$ for a deuteron, $^4_2\mathrm{He}$ for an α particle, and so on. When Q is positive, the total mass decreases and the total kinetic energy increases. Such a reaction is called an *exoergic reaction*. When Q is negative, the mass increases and the kinetic energy decreases, and the reaction is called an *endoergic reaction*. The terms *exothermal* and *endothermal,* borrowed from chemistry, are also used. In an endoergic reaction the reaction cannot occur at all unless the initial kinetic energy in the center-of-mass reference frame is at least as great as $|Q|$. That is, there is a **threshold energy,** the minimum kinetic energy to make an endoergic reaction go.

EXAMPLE 43.11 EXOERGIC AND ENDOERGIC REACTIONS

(a) When a lithium-7 nucleus is bombarded by a proton, two alpha particles $(^4\mathrm{He})$ are produced. Find the reaction energy. (b) Calculate the reaction energy for the reaction $^4_2\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{17}_{8}\mathrm{O} + {}^1_1\mathrm{H}$.

SOLUTION

IDENTIFY and SET UP: The reaction energy Q for any nuclear reaction equals c^2 times the difference between the total initial mass and the total final mass, as in Eq. (43.23). Table 43.2 gives the required masses.

EXECUTE: (a) The reaction is $^1_1\mathrm{H} + {}^7_3\mathrm{Li} \rightarrow {}^4_2\mathrm{He} + {}^4_2\mathrm{He}$. The initial and final masses and their respective sums are

A: $^1_1\mathrm{H}$	1.007825 u	C: $^4_2\mathrm{He}$	4.002603 u
B: $^7_3\mathrm{Li}$	7.016005 u	D: $^4_2\mathrm{He}$	4.002603 u
	8.023830 u		8.005206 u

The mass decreases by 0.018624 u. From Eq. (43.23), the reaction energy is

$$Q = (0.018624\ \text{u})(931.5\ \text{MeV/u}) = +17.35\ \text{MeV}$$

(b) The initial and final masses are

A: ${}^{4}_{2}\text{He}$	4.002603 u	C: ${}^{17}_{8}\text{O}$	16.999132 u
B: ${}^{14}_{7}\text{N}$	14.003074 u	D: ${}^{1}_{1}\text{H}$	1.007825 u
	18.005677 u		18.006957 u

The mass increases by 0.001280 u, and the corresponding reaction energy is

$$Q = (-0.001280\ \text{u})(931.5\ \text{MeV/u}) = -1.192\ \text{MeV}$$

EVALUATE: The reaction in part (a) is *exoergic:* The final total kinetic energy of the two separating alpha particles is 17.35 MeV greater than the initial total kinetic energy of the proton and the lithium nucleus. The reaction in part (b) is *endoergic:* In the center-of-mass system—that is, in a head-on collision with zero total momentum—the minimum total initial kinetic energy required for this reaction to occur is 1.192 MeV.

Ordinarily, the endoergic reaction of part (b) of Example 43.11 would be produced by bombarding stationary ^{14}N nuclei with alpha particles from an accelerator. In this case an alpha's kinetic energy must be *greater than* 1.192 MeV. If all the alpha's kinetic energy went solely to increasing the rest energy, the final kinetic energy would be zero, and momentum would not be conserved. When a particle with mass m and kinetic energy K collides with a stationary particle with mass M, the total kinetic energy K_{cm} in the center-of-mass coordinate system (the energy available to cause reactions) is

$$K_{\text{cm}} = \frac{M}{M + m}K \qquad (43.24)$$

This expression assumes that the kinetic energies of the particles and nuclei are much less than their rest energies. We leave the derivation of Eq. (43.24) to you. In part (b) of Example 43.11, $M = 14.003074$ u and $m = 4.002603$ u, so $M/(M + m) = (14.003074\ \text{u})/(18.005677\ \text{u}) = 0.7777$ and $K_{\text{cm}} = 0.7777K$. Since K_{cm} must be at least 1.192 MeV, the α particle's kinetic energy K must be at least $(1.192\ \text{MeV})/0.7777 = 1.533\ \text{MeV}$.

For a charged particle such as a proton or an α particle to penetrate the nucleus of another atom and cause a reaction, it must usually have enough initial kinetic energy to overcome the potential-energy barrier caused by the repulsive electrostatic forces. In the reaction of part (a) of Example 43.11, if we treat the proton and the ^{7}Li nucleus as spherically symmetric charges with radii given by Eq. (43.1), their centers will be 3.5×10^{-15} m apart when they touch. The repulsive potential energy of the proton (charge $+e$) and the ^{7}Li nucleus (charge $+3e$) at this separation r is

$$U = \frac{1}{4\pi\epsilon_0}\frac{(e)(3e)}{r} = (9.0 \times 10^{9}\ \text{N}\cdot\text{m}^2/\text{C}^2)\frac{(3)(1.6 \times 10^{-19}\ \text{C})^2}{3.5 \times 10^{-15}\ \text{m}}$$
$$= 2.0 \times 10^{-13}\ \text{J} = 1.2\ \text{MeV}$$

Even though the reaction is exoergic, the proton must have a minimum kinetic energy of about 1.2 MeV for the reaction to occur, unless the proton *tunnels* through the barrier (see Section 40.4).

Neutron Absorption

Absorption of *neutrons* by nuclei forms an important class of nuclear reactions. Heavy nuclei bombarded by neutrons can undergo a series of neutron absorptions alternating with beta decays, in which the mass number A increases by as much as 25. Some of the *transuranic elements,* elements having Z larger than 92, are produced in this way. These elements have not been found in nature. Many transuranic elements, having Z possibly as high as 118, have been identified.

The analytical technique of *neutron activation analysis* uses similar reactions. When bombarded by neutrons, many stable nuclides absorb a neutron to become unstable and then undergo β^- decay. The energies of the β^- and associated γ emissions depend on the unstable nuclide and provide a means of identifying it and the original stable nuclide. Quantities of elements that are far too small for conventional chemical analysis can be detected in this way.

TEST YOUR UNDERSTANDING OF SECTION 43.6 The reaction described in part (a) of Example 43.11 is exoergic. Can it happen naturally when a sample of solid lithium is placed in a flask of hydrogen gas?

43.7 NUCLEAR FISSION

MP

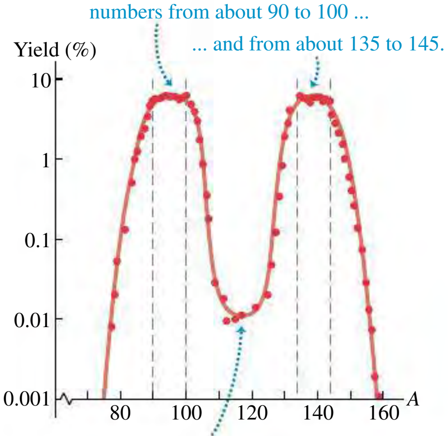
PhET: Nuclear Fission

Nuclear fission is a decay process in which an unstable nucleus splits into two fragments of comparable mass. Fission was discovered in 1938 through the experiments of Otto Hahn and Fritz Strassman in Germany. Pursuing earlier work by Enrico Fermi, they bombarded uranium $(Z = 92)$ with neutrons. The resulting radiation did not coincide with that of any known radioactive nuclide. Urged on by their colleague Lise Meitner, they used meticulous chemical analysis to reach the astonishing but inescapable conclusion that they had found a radioactive isotope of barium $(Z = 56)$. Later, radioactive krypton $(Z = 36)$ was also found. Meitner and Otto Frisch correctly interpreted these results as showing that uranium nuclei were splitting into two massive fragments called *fission fragments*. Two or three free neutrons usually appear along with the fission fragments and, very occasionally, a light nuclide such as ^{3}H.

Both the common isotope (99.3%) ^{238}U and the uncommon isotope (0.7%) ^{235}U (as well as several other nuclides) can be easily split by neutron bombardment: ^{235}U by slow neutrons (kinetic energy less than 1 eV) but ^{238}U only by fast neutrons with a minimum of about 1 MeV of kinetic energy. Fission resulting from neutron absorption is called *induced fission*. Some nuclides can also undergo *spontaneous fission* without initial neutron absorption, but this is quite rare. When ^{235}U absorbs a neutron, the resulting nuclide ^{236}U* is in a highly excited state and splits into two fragments almost instantaneously. Strictly speaking, it is ^{236}U*, not ^{235}U, that undergoes fission, but it's usual to speak of the fission of ^{235}U.

Over 100 different nuclides, representing more than 20 different elements, have been found among the fission products. **Figure 43.11** shows the distribution of mass numbers for fission fragments from the fission of ^{235}U.

43.11 Mass distribution of fission fragments from the fission of ^{236}U* (an excited state of ^{236}U), which is produced when ^{235}U absorbs a neutron. The vertical scale is logarithmic.



Fission Reactions

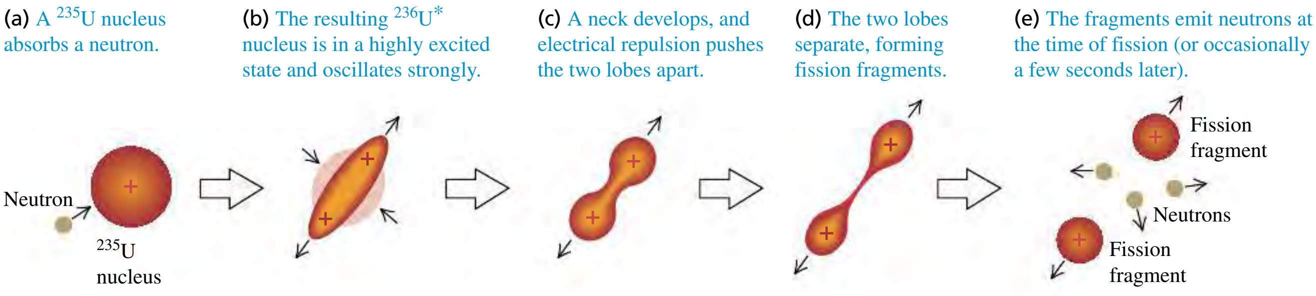
You should check the following two typical fission reactions for conservation of nucleon number and charge:

$$^{235}_{92}\text{U} + {}^1_0\text{n} \rightarrow {}^{236}_{92}\text{U}^* \rightarrow {}^{144}_{56}\text{Ba} + {}^{89}_{36}\text{Kr} + 3\,{}^1_0\text{n}$$
$$^{235}_{92}\text{U} + {}^1_0\text{n} \rightarrow {}^{236}_{92}\text{U}^* \rightarrow {}^{140}_{54}\text{Xe} + {}^{94}_{38}\text{Sr} + 2\,{}^1_0\text{n}$$

The total kinetic energy of the fission fragments is enormous, about 200 MeV (compared to typical α and β energies of a few MeV). The reason for this is that nuclides at the high end of the mass spectrum (near $A = 240$) are less tightly bound than those nearer the middle ($A = 90$ to 145). Referring to Fig. 43.2, we see that the average binding energy per nucleon is about 7.6 MeV at $A = 240$ but about 8.5 MeV at $A = 120$. Therefore a rough estimate of the expected *increase* in binding energy during fission is about 8.5 MeV $-$ 7.6 MeV $=$ 0.9 MeV per nucleon, or a total of $(235)(0.9\text{ MeV}) \approx 200$ MeV.

CAUTION **Binding energy and rest energy** It may seem to be a violation of conservation of energy to have an increase in both the binding energy and the kinetic energy during a fission reaction. But relative to the total rest energy E_0 of the separated nucleons, the rest energy of the nucleus is E_0 *minus* E_B. Thus an *increase* in binding energy corresponds to a *decrease* in rest energy as rest energy is converted to the kinetic energy of the fission fragments.

43.12 A liquid-drop model of fission.



Fission fragments always have too many neutrons to be stable. We noted in Section 43.3 that the neutron-to-proton ratio (N/Z) for stable nuclides is about 1 for light nuclides but almost 1.6 for the heaviest nuclides because of the increasing influence of the electrical repulsion of the protons. The N/Z value for stable nuclides is about 1.3 at $A = 100$ and 1.4 at $A = 150$. The fragments have about the same N/Z as ^{235}U, about 1.55. They usually respond to this surplus of neutrons by undergoing a series of β^- decays (each of which increases Z by 1 and decreases N by 1) until a stable value of N/Z is reached. A typical example is

$$^{140}_{54}\text{Xe} \xrightarrow{\beta^-} {}^{140}_{55}\text{Cs} \xrightarrow{\beta^-} {}^{140}_{56}\text{Ba} \xrightarrow{\beta^-} {}^{140}_{57}\text{La} \xrightarrow{\beta^-} {}^{140}_{58}\text{Ce}$$

The nuclide ^{140}Ce is stable. This series of β^- decays produces, on average, about 15 MeV of additional kinetic energy. The neutron excess of fission fragments also explains why two or three free neutrons are released during the fission.

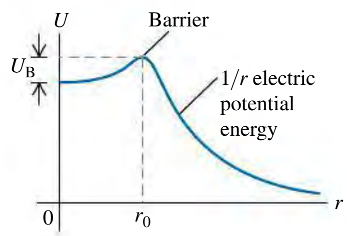
Fission appears to set an upper limit on the production of transuranic nuclei, mentioned in Section 43.6, that are relatively stable. There are theoretical reasons to expect that nuclei near $Z = 114$, $N = 184$ or 196, might be stable with respect to spontaneous fission. In the shell model (see Section 43.2), these numbers correspond to filled shells and subshells in the nuclear energy-level structure. Such *superheavy nuclei* would still be unstable with respect to alpha emission. In 2009 it was confirmed that there are at least four isotopes with $Z = 114$, the longest-lived of which has a half-life due to alpha decay of about 2.6 s.

Liquid-Drop Model

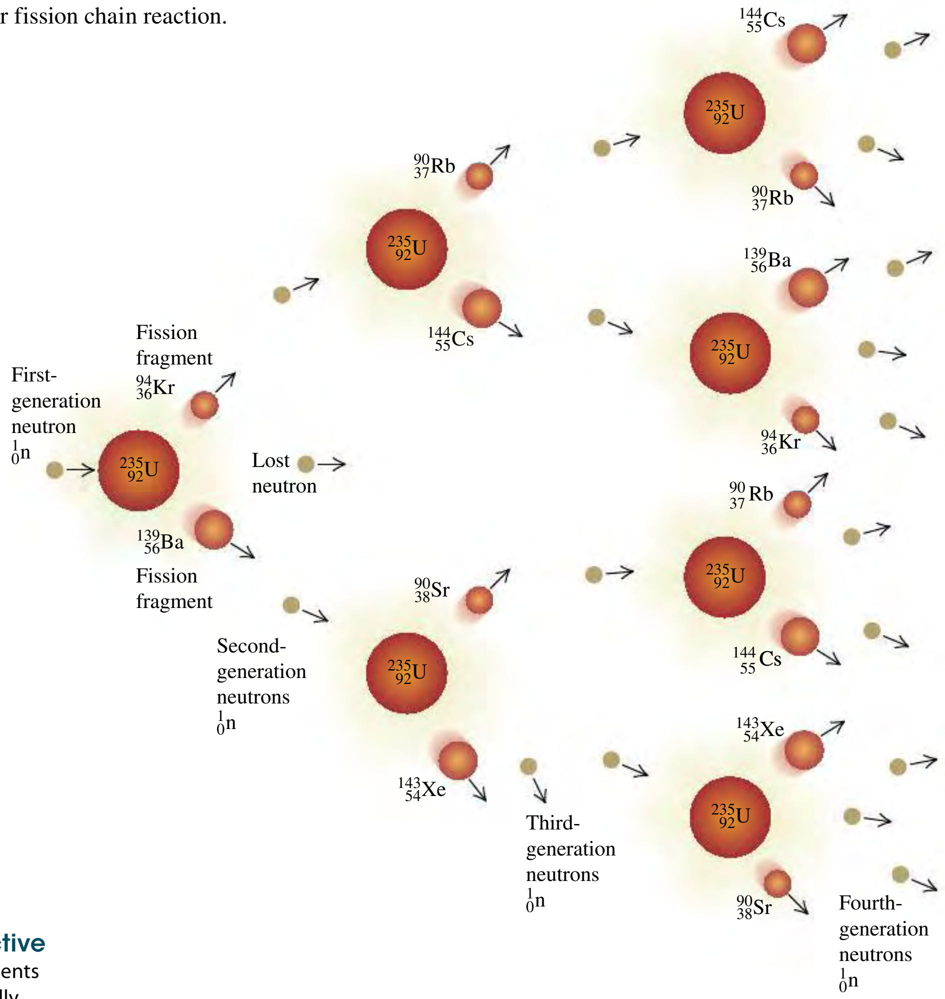
We can understand fission qualitatively on the basis of the liquid-drop model of the nucleus (see Section 43.2). The process is shown in **Fig. 43.12** in terms of an electrically charged liquid drop. These sketches shouldn't be taken too literally, but they may help to develop your intuition about fission. A ^{235}U nucleus absorbs a neutron (Fig. 43.12a), becoming a ^{236}U* nucleus with excess energy (Fig. 43.12b). This excess energy causes violent oscillations, during which a neck between two lobes develops (Fig. 43.12c). The electrical repulsion of these two lobes stretches the neck farther (Fig. 43.12d), and finally two smaller fragments are formed (Fig. 43.12e) that move rapidly apart.

This qualitative picture has been developed into a more quantitative theory to explain why some nuclei undergo fission and others don't. **Figure 43.13** shows a hypothetical potential-energy function for two possible fission fragments. If neutron absorption results in an excitation energy greater than the energy barrier height U_B, fission occurs immediately. Even when there isn't quite enough energy to surmount the barrier, fission can take place by quantum-mechanical *tunneling,* discussed in Section 40.4. In principle, many stable heavy nuclei can fission by tunneling. But the probability depends very critically on the height and width of the barrier. For most nuclei this process is so unlikely that it is never observed.

43.13 Hypothetical potential-energy function for two fission fragments in a fissionable nucleus. At distances r beyond the range of the nuclear force, the potential energy varies approximately as $1/r$. Fission occurs if there is an excitation energy greater than U_B or an appreciable probability for tunneling through the potential-energy barrier.



43.14 Schematic diagram of a nuclear fission chain reaction.



BIO Application Making Radioactive Isotopes for Medicine The fragments that result from nuclear fission are typically unstable, neutron-rich isotopes. A number of these are useful for medical diagnosis and cancer radiotherapy (see Section 43.5). This photograph shows a nuclear fission reactor used for producing such isotopes. The uranium fuel is kept in a large tank of water for cooling. Some of the neutron-rich fission fragments undergo beta decay and emit electrons that move faster than the speed of light in water (about 0.75*c*). Like an airplane that produces an intense sonic boom when it flies faster than sound (see Section 16.9), these ultrafast electrons produce a "light boom" called *Čerenkov radiation* that has a characteristic blue color.

Chain Reactions

Fission of a uranium nucleus, triggered by neutron bombardment, releases other neutrons that can trigger more fissions, suggesting the possibility of a **chain reaction** (**Fig. 43.14**). The chain reaction may be made to proceed slowly and in a controlled manner in a nuclear reactor or explosively in a bomb. The energy release in a nuclear chain reaction is enormous, far greater than that in any chemical reaction. (In a sense, *fire* is a chemical chain reaction.) For example, when uranium is "burned" to uranium dioxide in the chemical reaction

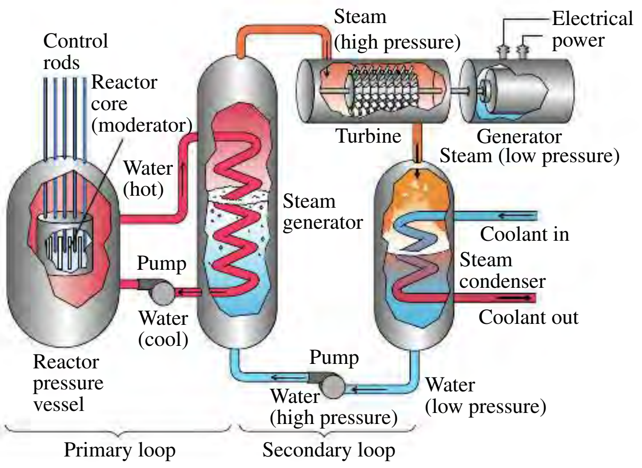
$$U + O_2 \rightarrow UO_2$$

the heat of combustion is about 4500 J/g. Expressed as energy per atom, this is about 11 eV per atom. By contrast, fission liberates about 200 MeV per atom, nearly 20 million times as much energy.

Nuclear Reactors

A *nuclear reactor* is a system in which a controlled nuclear chain reaction is used to liberate energy. In a nuclear power plant, this energy is used to generate steam, which operates a turbine and turns an electrical generator.

On average, each fission of a ^{235}U nucleus produces about 2.5 free neutrons, so 40% of the neutrons are needed to sustain a chain reaction. A ^{235}U nucleus is much more likely to absorb a low-energy neutron (less than 1 eV) than one of the higher-energy neutrons (1 MeV or so) that are liberated during fission. In a nuclear reactor the higher-energy neutrons are slowed down by collisions with nuclei in the surrounding material, called the *moderator,* so they are much more likely to cause further fissions. In nuclear power plants, the moderator is often



43.15 Schematic diagram of a nuclear power plant.

water, occasionally graphite. The *rate* of the reaction is controlled by inserting or withdrawing *control rods* made of elements (such as boron or cadmium) whose nuclei *absorb* neutrons without undergoing any additional reaction. The isotope ^{238}U can also absorb neutrons, leading to $^{239}U^*$, but not with high enough probability for it to sustain a chain reaction by itself. Thus uranium that is used in reactors is often "enriched" by increasing the proportion of ^{235}U above the natural value of 0.7%, typically to 3% or so, by isotope-separation processing.

The most familiar application of nuclear reactors is for the generation of electrical power. As was noted above, the fission energy appears as kinetic energy of the fission fragments, and its immediate result is to increase the internal energy of the fuel elements and the surrounding moderator. This increase in internal energy is transferred as heat to generate steam to drive turbines, which spin the electrical generators. **Figure 43.15** is a schematic diagram of a nuclear power plant. The energetic fission fragments heat the water surrounding the reactor core. The steam generator is a heat exchanger that takes heat from this highly radioactive water and generates nonradioactive steam to run the turbines.

A typical nuclear plant has an electric-generating capacity of 1000 MW (or 10^9 W). The turbines are heat engines and are subject to the efficiency limitations imposed by the second law of thermodynamics, discussed in Chapter 20. In modern nuclear plants the overall efficiency is about one-third, so 3000 MW of thermal power from the fission reaction is needed to generate 1000 MW of electrical power.

EXAMPLE 43.12 URANIUM CONSUMPTION IN A NUCLEAR REACTOR

SOLUTION

What mass of ^{235}U must undergo fission each day to provide 3000 MW of thermal power?

SOLUTION

IDENTIFY and SET UP: Fission of ^{235}U liberates about 200 MeV per atom. We use this and the mass of the ^{235}U atom to determine the required amount of uranium.

EXECUTE: Each second, we need 3000 MJ or 3000×10^6 J. Each fission provides 200 MeV, or

$$(200\ \text{MeV/fission})(1.6 \times 10^{-13}\ \text{J/MeV}) = 3.2 \times 10^{-11}\ \text{J/fission}$$

The number of fissions needed each second is

$$\frac{3000 \times 10^6\ \text{J}}{3.2 \times 10^{-11}\ \text{J/fission}} = 9.4 \times 10^{19}\ \text{fissions}$$

Each ^{235}U atom has a mass of $(235\ \text{u})(1.66 \times 10^{-27}\ \text{kg/u}) = 3.9 \times 10^{-25}$ kg, so the mass of ^{235}U that undergoes fission each second is

$$(9.4 \times 10^{19})(3.9 \times 10^{-25}\ \text{kg}) = 3.7 \times 10^{-5}\ \text{kg} = 37\ \mu\text{g}$$

In one day (86,400 s), the total consumption of ^{235}U is

$$(3.7 \times 10^{-5}\ \text{kg/s})(86{,}400\ \text{s}) = 3.2\ \text{kg}$$

EVALUATE: For comparison, a 1000-MW coal-fired power plant burns 10,600 tons (about 10 million kg) of coal per day!

We mentioned above that about 15 MeV of the energy released after fission of a ^{235}U nucleus comes from the β^- decays of the fission fragments. This fact poses a serious problem with respect to control and safety of reactors. Even after the chain reaction has been completely stopped by insertion of control rods into the core, heat continues to be evolved by the β^- decays, which cannot be stopped. For a 3000-MW reactor this heat power is initially very large, about 200 MW. In the event of total loss of cooling water, this power is more than enough to cause a catastrophic meltdown of the reactor core and possible penetration of the containment vessel. The difficulty in achieving a "cold shutdown" following an accident at the Three Mile Island nuclear power plant in Pennsylvania in March 1979 was a result of the continued evolution of heat due to β^- decays.

The catastrophe of April 26, 1986, at Chernobyl reactor No. 4 in Ukraine resulted from a combination of an inherently unstable design and several human errors committed during a test of the emergency core cooling system. Too many control rods were withdrawn to compensate for a decrease in power caused by a buildup of neutron absorbers such as ^{135}Xe. The power level rose from 1% of normal to 100 times normal in 4 seconds; a steam explosion ruptured pipes in the core cooling system and blew the heavy concrete cover off the reactor. The graphite moderator caught fire and burned for several days, and there was a meltdown of the core. The total activity of the radioactive material released into the atmosphere has been estimated as about 10^8 Ci.

TEST YOUR UNDERSTANDING OF SECTION 43.7 The fission of ^{235}U can be triggered by the absorption of a slow neutron by a nucleus. Can a slow *proton* be used to trigger ^{235}U fission?

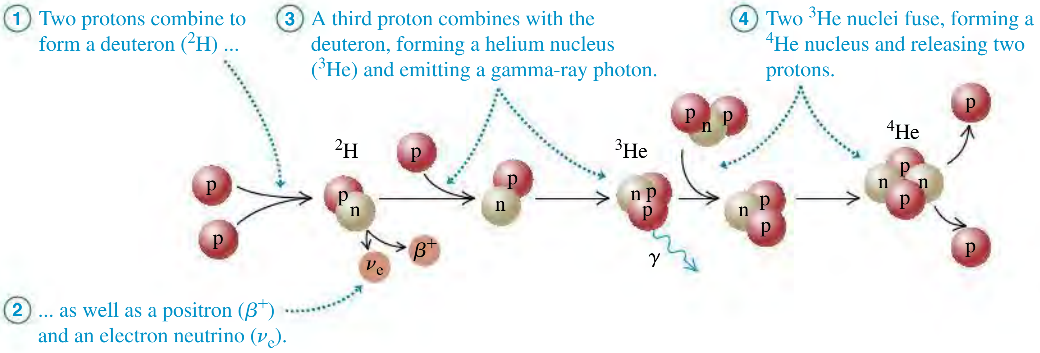
43.8 NUCLEAR FUSION

In a **nuclear fusion** reaction, two or more small light nuclei come together, or *fuse,* to form a larger nucleus. Fusion reactions release energy for the same reason as fission reactions: The binding energy per nucleon after the reaction is greater than before. Referring to Fig. 43.2, we see that the binding energy per nucleon increases with A up to about $A = 60$, so fusion of nearly any two light nuclei to make a nucleus with A less than 60 is likely to be an exoergic reaction. In comparison to fission, we are moving toward the peak of this curve from the opposite side. Another way to express the energy relationships is that the total mass of the products is less than that of the initial particles.

Here are three examples of energy-liberating fusion reactions, written in terms of the neutral atoms:

$$
\begin{aligned}
{}^1_1\mathrm{H} + {}^1_1\mathrm{H} &\rightarrow {}^2_1\mathrm{H} + \beta^+ + \nu_e \\
{}^2_1\mathrm{H} + {}^1_1\mathrm{H} &\rightarrow {}^3_2\mathrm{He} + \gamma \\
{}^3_2\mathrm{He} + {}^3_2\mathrm{He} &\rightarrow {}^4_2\mathrm{He} + {}^1_1\mathrm{H} + {}^1_1\mathrm{H}
\end{aligned}
$$

In the first reaction, two protons combine to form a deuteron (^2H), with the emission of a positron (β^+) and an electron neutrino. In the second, a proton and a deuteron combine to form the nucleus of the light isotope of helium, ^3He, with the emission of a gamma ray. Now double the first two reactions to provide the two ^3He nuclei that fuse in the third reaction to form an alpha particle (^4He) and two protons. Together the reactions make up the process called the *proton-proton chain* (**Fig. 43.16**).



43.16 The proton-proton chain.

The net effect of the chain is the conversion of four protons into one α particle, two positrons, two electron neutrinos, and two γ's. We can calculate the energy release from this part of the process: The mass of an α particle plus two positrons is the mass of neutral ^{4}He, the neutrinos have zero (or negligible) mass, and the gammas have zero mass.

Mass of four protons	4.029106 u
Mass of ^{4}He	4.002603 u
Mass difference and energy release	0.026503 u and 24.69 MeV

The two positrons that are produced during the first step of the proton-proton chain collide with two electrons; mutual annihilation of the four particles takes place, and their rest energy is converted into $4(0.511\text{ MeV}) = 2.044\text{ MeV}$ of gamma radiation. Thus the total energy released is $(24.69 + 2.044)\text{ MeV} = 26.73\text{ MeV}$. The proton-proton chain takes place in the interior of the sun and other stars (**Fig. 43.17**). Each gram of the sun's mass contains about 4.5×10^{23} protons. If all of these protons were fused into helium, the energy released would be about 130,000 kWh. If the sun were to continue to radiate at its present rate, it would take about 75×10^9 years to exhaust its supply of protons. As we will soon see, fusion reactions can occur only at extremely high temperatures; in the sun, these temperatures are found only deep within the interior. Hence the sun cannot fuse *all* of its protons and can sustain fusion for a total of only about 10×10^9 years in total. The present age of the solar system (including the sun) is 4.54×10^9 years, so the sun is about halfway through its available store of protons.

43.17 The energy released as starlight comes from fusion reactions deep within a star's interior. When a star is first formed and for most of its life, it converts the hydrogen in its core into helium. As a star ages, the core temperature can become high enough for additional fusion reactions that convert helium into carbon, oxygen, and other elements.

EXAMPLE 43.13 A FUSION REACTION

SOLUTION

Two deuterons fuse to form a *triton* (a nucleus of tritium, or ^{3}H) and a proton. How much energy is liberated?

SOLUTION

IDENTIFY and SET UP: This is a nuclear reaction of the type discussed in Section 43.6. We use Eq. (43.23) to find the energy released.

EXECUTE: Adding one electron to each nucleus makes each a neutral atom; we find their masses in Table 43.2. Substituting into Eq. (43.23), we find

$$Q = [2(2.014102\text{ u}) - 3.016049\text{ u} - 1.007825\text{ u}] \times (931.5\text{ MeV/u}) = 4.03\text{ MeV}$$

EVALUATE: Thus 4.03 MeV is released in the reaction; the triton and proton together have 4.03 MeV more kinetic energy than the two deuterons had together.

Achieving Fusion

For two nuclei to undergo fusion, they must come together to within the range of the nuclear force, typically of the order of 2×10^{-15} m. To do this, they must overcome the electrical repulsion of their positive charges. For two protons at this distance, the corresponding potential energy is about 1.2×10^{-13} J or 0.7 MeV; this represents the total initial *kinetic* energy that the fusion nuclei must have—for example, 0.6×10^{-13} J each in a head-on collision.

Atoms have this much energy only at extremely high temperatures. The discussion of Section 18.3 showed that the average translational kinetic energy of a gas molecule at temperature T is $\frac{3}{2}kT$, where k is Boltzmann's constant. The temperature at which this is equal to $E = 0.6 \times 10^{-13}$ J is determined by the relationship

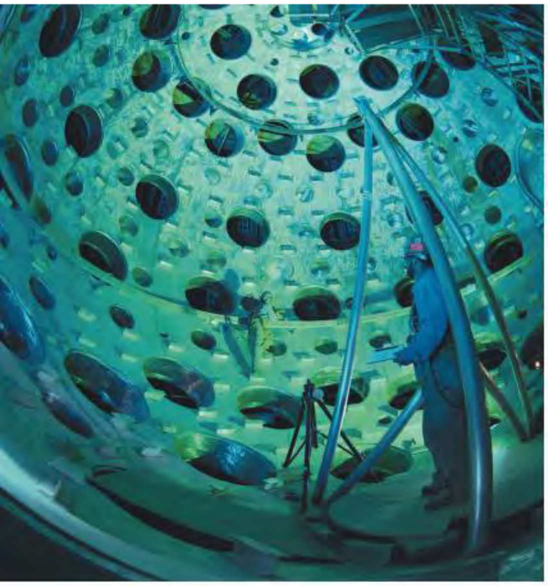
$$E = \tfrac{3}{2}kT$$

$$T = \frac{2E}{3k} = \frac{2(0.6 \times 10^{-13}\ \text{J})}{3(1.38 \times 10^{-23}\ \text{J/K})} = 3 \times 10^{9}\ \text{K}$$

Fusion reactions are possible at lower temperatures because the Maxwell–Boltzmann distribution function (see Section 18.5) gives a small fraction of protons with kinetic energies much higher than the average value. The proton-proton reaction occurs at "only" 1.5×10^{7} K at the center of the sun, making it an extremely low-probability process; but that's why the sun is expected to last so long. At these temperatures the fusion reactions are called *thermonuclear* reactions.

Intensive efforts are under way to achieve controlled fusion reactions, which potentially represent an enormous new resource of energy (see Fig. 24.11). At the temperatures mentioned, light atoms are fully ionized, and the resulting state of matter is called a *plasma*. In one kind of experiment using *magnetic confinement,* a plasma is heated to extremely high temperature by an electrical discharge, while being contained by appropriately shaped magnetic fields. In another, through *inertial confinement,* pellets of the material to be fused are heated by a high-intensity laser beam (see **Fig. 43.18**). Some of the reactions being studied are

43.18 This target chamber at the National Ignition Facility in California has apertures for 192 powerful laser beams. The lasers deliver 5×10^{14} W of power for a few nanoseconds to a millimeter-sized pellet of deuterium and tritium at the center of the chamber, thus triggering thermonuclear fusion.



$$ {}^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{1}\text{H} + {}^{1}_{1}\text{H} + 4.0\ \text{MeV} \qquad (1)$$

$$ {}^{3}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{0}\text{n} + 17.6\ \text{MeV} \qquad (2)$$

$$ {}^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{2}\text{He} + {}^{1}_{0}\text{n} + 3.3\ \text{MeV} \qquad (3)$$

$$ {}^{3}_{2}\text{He} + {}^{2}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{1}\text{H} + 18.3\ \text{MeV} \qquad (4)$$

We considered reaction (1) in Example 43.13; two deuterons fuse to form a triton and a proton. In reaction (2) a triton combines with another deuteron to form an alpha particle and a neutron. The result of these two reactions together is the conversion of three deuterons into an alpha particle, a proton, and a neutron, with 21.6 MeV of energy liberated. Reactions (3) and (4) together achieve the same conversion. In a plasma that contains deuterons, the two pairs of reactions occur with roughly equal probability. As yet, no one has succeeded in producing these reactions under controlled conditions in such a way as to yield a net surplus of usable energy.

Methods of achieving fusion that don't require high temperatures are also being studied; these are called *cold fusion*. One successful scheme uses an unusual hydrogen molecule ion. The usual H_2^+ ion consists of two protons bound by one shared electron; the nuclear spacing is about 0.1 nm. If the protons are replaced by a deuteron (^2H) and a triton (^3H) and the electron by a *muon,* which is 208 times as massive as the electron, the spacing is reduced by a factor of 208. The probability then becomes appreciable for the two nuclei to tunnel through the narrow repulsive potential-energy barrier and fuse in reaction (2) above. The prospect of making this process, called *muon-catalyzed fusion,* into a practical energy source is still distant.

TEST YOUR UNDERSTANDING OF SECTION 43.8 Are *all* fusion reactions exoergic? ❙

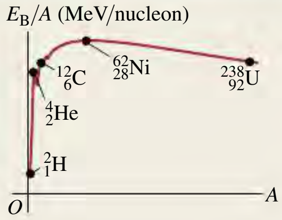
SOLUTIONS TO ALL EXAMPLES

Nuclear properties: A nucleus is composed of A nucleons (Z protons and N neutrons). All nuclei have about the same density. The radius of a nucleus with mass number A is given approximately by Eq. (43.1). A single nuclear species of a given Z and N is called a nuclide. Isotopes are nuclides of the same element (same Z) that have different numbers of neutrons. Nuclear masses are measured in atomic mass units. Nucleons have angular momentum and a magnetic moment. (See Examples 43.1 and 43.2.)

$$R = R_0 A^{1/3} \quad (R_0 = 1.2 \times 10^{-15}\ \text{m}) \tag{43.1}$$

Nuclear binding and structure: The mass of a nucleus is always less than the mass of the protons and neutrons within it. The mass difference multiplied by c^2 gives the binding energy E_B. The binding energy for a given nuclide is determined by the nuclear force, which is short range and favors pairs of particles, and by the electrical repulsion between protons. A nucleus is unstable if A or Z is too large or if the ratio N/Z is wrong. Two widely used models of the nucleus are the liquid-drop model and the shell model; the latter is analogous to the central-field approximation for atomic structure. (See Examples 43.3 and 43.4.)

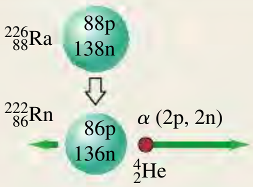
$$E_B = (ZM_H + Nm_n - {}^{A}_{Z}M)c^2 \tag{43.10}$$



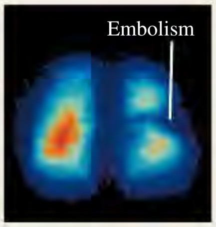
Radioactive decay: Unstable nuclides usually emit an alpha particle (a ${}^{4}_{2}He$ nucleus) or a beta particle (an electron) in the process of changing to another nuclide, sometimes followed by a gamma-ray photon. The rate of decay of an unstable nucleus is described by the decay constant λ, the half-life $T_{1/2}$, or the lifetime T_{mean}. If the number of nuclei at time $t = 0$ is N_0 and no more are produced, the number at time t is given by Eq. (43.17). (See Examples 43.5–43.9.)

$$N(t) = N_0 e^{-\lambda t} \tag{43.17}$$

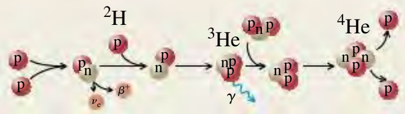
$$T_{mean} = \frac{1}{\lambda} = \frac{T_{1/2}}{\ln 2} = \frac{T_{1/2}}{0.693} \tag{43.19}$$



Biological effects of radiation: The biological effect of any radiation depends on the product of the energy absorbed per unit mass and the relative biological effectiveness (RBE), which is different for different radiations. (See Example 43.10.)



Nuclear reactions: In a nuclear reaction, two nuclei or particles collide to produce two new nuclei or particles. Reactions can be exoergic or endoergic. Several conservation laws, including charge, energy, momentum, angular momentum, and nucleon number, are obeyed. Energy is released by the fission of a heavy nucleus into two lighter, always unstable, nuclei. Energy is also released by the fusion of two light nuclei into a heavier nucleus. (See Examples 43.11–43.13.)



BRIDGING PROBLEM SATURATION OF ^{128}I PRODUCTION

In an experiment, the iodine isotope ^{128}I is created by irradiating a sample of ^{127}I with a beam of neutrons, yielding 1.50×10^6 ^{128}I nuclei per second. Initially no ^{128}I nuclei are present. A ^{128}I nucleus decays by β^- emission with a half-life of 25.0 min. (a) To what nuclide does ^{128}I decay? (b) Could that nuclide decay back to ^{128}I by β^+ emission? Why or why not? (c) After the sample has been irradiated for a long time, what is the maximum number of ^{128}I atoms that can be present in the sample? What is the maximum activity that can be produced? (This steady-state situation is called *saturation*.) (d) Find an expression for the number of ^{128}I atoms present in the sample as a function of time.

SOLUTION GUIDE

IDENTIFY and SET UP

1. What happens to the values of Z, N, and A in β^- decay? What must be true for β^- decay to be possible? For β^+ decay to be possible?
2. You'll need to write an equation for the rate of change dN/dt of the number N of ^{128}I atoms in the sample, taking account of both the creation of ^{128}I by the neutron irradiation and the decay of any ^{128}I present. In the steady state, how do the rates of these two processes compare?
3. List the unknown quantities for each part of the problem and identify your target variables.

EXECUTE

4. Find the values of Z and N of the nuclide produced by the decay of ^{128}I. What element is this?
5. Decide whether this nuclide can decay back to ^{128}I.
6. Inspect your equation for dN/dt. What is the value of dN/dt in the steady state? Use this to solve for the steady-state values of N and the activity.
7. Solve your dN/dt equation for the function $N(t)$. (*Hint:* See Section 26.4.)

EVALUATE

8. Your result from step 6 tells you the value of N after a long time (that is, for large values of t). Is this consistent with your result from step 7? What would constitute a "long time" under these conditions?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **DATA**: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q43.1 **BIO** Neutrons have a magnetic dipole moment and can undergo spin flips by absorbing electromagnetic radiation. Why, then, are protons rather than neutrons used in MRI of body tissues? (See Fig. 43.1.)

Q43.2 In Eq. (43.11), as the total number of nucleons becomes larger, the importance of the second term in the equation decreases relative to that of the first term. Does this make physical sense? Explain.

Q43.3 Why aren't the masses of all nuclei integer multiples of the mass of a single nucleon?

Q43.4 The only two stable nuclides with more protons than neutrons are $^1_1$H and $^3_2$He. Why is $Z > N$ so uncommon?

Q43.5 What are the six known elements for which Z is a magic number? Discuss what properties these elements have as a consequence of their special values of Z.

Q43.6 The binding energy per nucleon for most nuclides doesn't vary much (see Fig. 43.2). Is there similar consistency in the *atomic* energy of atoms, on an "energy per electron" basis? If so, why? If not, why not?

Q43.7 Heavy, unstable nuclei usually decay by emitting an α or a β particle. Why don't they usually emit a single proton or neutron?

Q43.8 As stars age, they use up their supply of hydrogen and eventually begin producing energy by a reaction that involves the fusion of three helium nuclei to form a carbon nucleus. Would you expect the interiors of these old stars to be hotter or cooler than the interiors of younger stars? Explain.

Q43.9 Since lead is a stable element, why doesn't the ^{238}U decay series shown in Fig. 43.7 stop at lead, ^{214}Pb?

Q43.10 In the ^{238}U decay series shown in Fig. 43.7, some nuclides in the series are found much more abundantly in nature than others, even though every ^{238}U nucleus goes through every step in the series before finally becoming ^{206}Pb. Why don't the intermediate nuclides all have the same abundance?

Q43.11 Compared to α particles with the same energy, β particles can much more easily penetrate through matter. Why is this?

Q43.12 If $^A_Z\mathrm{El_i}$ represents the initial nuclide, what is the decay process or processes if the final nuclide is (a) $_{Z+1}^{A}\mathrm{El_f}$; (b) $^{A-4}_{Z-2}\mathrm{El_f}$; (c) $_{Z-1}^{A}\mathrm{El_f}$?

Q43.13 In a nuclear decay equation, why can we represent an electron as $^{0}_{-1}\beta^-$? What are the equivalent representations for a positron, a neutrino, and an antineutrino?

Q43.14 Why is the alpha, beta, or gamma decay of an unstable nucleus unaffected by the *chemical* situation of the atom, such as the nature of the molecule or solid in which it is bound? The chemical situation of the atom can, however, have an effect on the half-life in electron capture. Why is this?

Q43.15 In the process of *internal conversion*, a nucleus decays from an excited state to a ground state by giving the excitation energy directly to an atomic electron rather than emitting a gamma-ray photon. Why can this process also produce x-ray photons?

Q43.16 In Example 43.9 (Section 43.4), the activity of atmospheric carbon *before* 1900 was given. Discuss why this activity may have changed since 1900.

Q43.17 BIO One problem in radiocarbon dating of biological samples, especially very old ones, is that they can easily be contaminated with modern biological material during the measurement process. What effect would such contamination have on the estimated age? Why is such contamination a more serious problem for samples of older material than for samples of younger material?

Q43.18 The most common radium isotope found on earth, ^{226}Ra, has a half-life of about 1600 years. If the earth was formed well over 10^9 years ago, why is there any radium left now?

Q43.19 Fission reactions occur only for nuclei with large nucleon numbers, while exoergic fusion reactions occur only for nuclei with small nucleon numbers. Why is this?

Q43.20 When a large nucleus splits during nuclear fission, the daughter nuclei of the fission fly apart with enormous kinetic energy. Why does this happen?

EXERCISES

Section 43.1 Properties of Nuclei

43.1 • How many protons and how many neutrons are there in a nucleus of the most common isotope of (a) silicon, $^{28}_{14}$Si; (b) rubidium, $^{85}_{37}$Rb; (c) thallium, $^{205}_{81}$Tl?

43.2 •• Neutrons are placed in a magnetic field with magnitude 2.30 T. (a) What is the energy difference between the states with the nuclear spin angular momentum components parallel and antiparallel to the field? Which state is lower in energy: the one with its spin component parallel to the field or the one with its spin component antiparallel to the field? How do your results compare with the energy states for a proton in the same field (see Example 43.2)? (b) The neutrons can make transitions from one of these states to the other by emitting or absorbing a photon with energy equal to the energy difference of the two states. Find the frequency and wavelength of such a photon.

43.3 • Hydrogen atoms are placed in an external magnetic field. The protons can make transitions between states in which the nuclear spin component is parallel and antiparallel to the field by absorbing or emitting a photon. What magnetic-field magnitude is required for this transition to be induced by photons with frequency 22.7 MHz?

Section 43.2 Nuclear Binding and Nuclear Structure

43.4 •• The nuclei $^{11}_{5}$B and $^{11}_{6}$C are called *mirror nuclei,* because the number of protons in $^{11}_{5}$B equals the number of neutrons in $^{11}_{6}$C and the number of neutrons in $^{11}_{5}$B equals the number of protons in $^{11}_{6}$C. The atomic mass of $^{11}_{6}$C is 11.011434 u, and the atomic mass of $^{11}_{5}$B is given in Table 43.2. (a) Calculate the binding energies of $^{11}_{5}$B and $^{11}_{6}$C. (b) Which of the two nuclei has the larger binding energy? Why is this so?

43.5 • The most common isotope of boron is $^{11}_{5}$B. (a) Determine the total binding energy of $^{11}_{5}$B from Table 43.2 in Section 43.1. (b) Calculate this binding energy from Eq. (43.11). (Why is the fifth term zero?) Compare to the result you obtained in part (a). What is the percent difference? Compare the accuracy of Eq. (43.11) for $^{11}_{5}$B to its accuracy for $^{62}_{28}$Ni (see Example 43.4).

43.6 • The most common isotope of uranium, $^{238}_{92}$U, has atomic mass 238.050788 u. Calculate (a) the mass defect; (b) the binding energy (in MeV); (c) the binding energy per nucleon.

43.7 • Calculate (a) the total binding energy and (b) the binding energy per nucleon of ^{12}C. (c) What percent of the rest mass of this nucleus is its total binding energy?

43.8 •• An alpha particle is strongly bound. The $^{12}_{6}$C nucleus might be modeled as a composite of three alpha particles. Compare the binding energy of $^{12}_{6}$C with three times the binding energy of an alpha particle. Which of these quantities is larger, and why might this be so?

43.9 • CP A photon with a wavelength of 3.50×10^{-13} m strikes a deuteron, splitting it into a proton and a neutron. (a) Calculate the kinetic energy released in this interaction. (b) Assuming the two particles share the energy equally, and taking their masses to be 1.00 u, calculate their speeds after the photodisintegration.

43.10 • Calculate the mass defect, the binding energy (in MeV), and the binding energy per nucleon of (a) the nitrogen nucleus, $^{14}_{7}$N, and (b) the helium nucleus, $^{4}_{2}$He. (c) How does the binding energy per nucleon compare for these two nuclei?

43.11 • Use Eq. (43.11) to calculate the binding energy per nucleon for the nuclei $^{86}_{36}$Kr and $^{180}_{73}$Ta. Do your results confirm what is shown in Fig. 43.2—that for A greater than 62 the binding energy per nucleon decreases as A increases?

Section 43.3 Nuclear Stability and Radioactivity

43.12 • (a) Is the decay $n \rightarrow p + \beta^- + \bar{\nu}_e$ energetically possible? If not, explain why not. If so, calculate the total energy released. (b) Is the decay $p \rightarrow n + \beta^+ + \nu_e$ energetically possible? If not, explain why not. If so, calculate the total energy released.

43.13 • What nuclide is produced in the following radioactive decays? (a) α decay of $^{239}_{94}$Pu; (b) β^- decay of $^{24}_{11}$Na; (c) β^+ decay of $^{15}_{8}$O.

43.14 •• CP ^{238}U decays spontaneously by α emission to ^{234}Th. Calculate (a) the total energy released by this process and (b) the recoil velocity of the ^{234}Th nucleus. The atomic masses are 238.050788 u for ^{238}U and 234.043601 u for ^{234}Th.

43.15 •• The atomic mass of ^{14}C is 14.003242 u. Show that the β^- decay of ^{14}C is energetically possible, and calculate the energy released in the decay.

43.16 • What particle (α particle, electron, or positron) is emitted in the following radioactive decays? (a) $^{27}_{14}\text{Si} \rightarrow {}^{27}_{13}\text{Al}$; (b) $^{238}_{92}\text{U} \rightarrow {}^{234}_{90}\text{Th}$; (c) $^{74}_{33}\text{As} \rightarrow {}^{74}_{34}\text{Se}$.

43.17 •• (a) Calculate the energy released by the electron-capture decay of $^{57}_{27}$Co (see Example 43.7). (b) A negligible amount of this energy goes to the resulting $^{57}_{26}$Fe atom as kinetic energy. About 90% of the time, the $^{57}_{26}$Fe nucleus emits two successive gamma-ray photons after the electron-capture process, of energies 0.122 MeV and 0.014 MeV, respectively, in decaying to its ground state. What is the energy of the neutrino emitted in this case?

43.18 • Tritium ($^{3}_{1}$H) is an unstable isotope of hydrogen; its mass, including one electron, is 3.016049 u. (a) Show that tritium must be unstable with respect to beta decay because the decay products ($^{3}_{2}$He plus an emitted electron) have less total mass than the tritium. (b) Determine the total kinetic energy (in MeV) of the decay products, taking care to account for the electron masses correctly.

Section 43.4 Activities and Half-Lives

43.19 • If a 6.13-g sample of an isotope having a mass number of 124 decays at a rate of 0.350 Ci, what is its half-life?

43.20 • BIO Radioactive isotopes used in cancer therapy have a "shelf-life," like pharmaceuticals used in chemotherapy. Just after it has been manufactured in a nuclear reactor, the activity of a sample of ^{60}Co is 5000 Ci. When its activity falls below 3500 Ci, it is considered too weak a source to use in treatment. You work in the radiology department of a large hospital. One of these ^{60}Co sources in your inventory was manufactured on October 6, 2011. It is now April 6, 2014. Is the source still usable? The half-life of ^{60}Co is 5.271 years.

43.21 •• The common isotope of uranium, ^{238}U, has a half-life of 4.47×10^9 years, decaying to ^{234}Th by alpha emission. (a) What is the decay constant? (b) What mass of uranium is required for an activity of 1.00 curie? (c) How many alpha particles are emitted per second by 10.0 g of uranium?

43.22 •• BIO **Radiation Treatment of Prostate Cancer.** In many cases, prostate cancer is treated by implanting 60 to 100 small seeds of radioactive material into the tumor. The energy released from the decays kills the tumor. One isotope that is used (there are others) is palladium (^{103}Pd)), with a half-life of 17 days. If a typical grain contains 0.250 g of ^{103}Pd, (a) what is its initial activity rate in Bq, and (b) what is the rate 68 days later?

43.23 •• A 12.0-g sample of carbon from living matter decays at the rate of 184 decays/minute due to the radioactive ^{14}C in it. What will be the decay rate of this sample in (a) 1000 years and (b) 50,000 years?

43.24 •• BIO **Radioactive Tracers.** Radioactive isotopes are often introduced into the body through the bloodstream. Their spread through the body can then be monitored by detecting the appearance of radiation in different organs. One such tracer is ^{131}I, a β^- emitter with a half-life of 8.0 d. Suppose a scientist introduces a sample with an activity of 325 Bq and watches it spread to the organs. (a) Assuming that all of the sample went to the thyroid gland, what will be the decay rate in that gland 24 d (about $3\frac{1}{2}$ weeks) later? (b) If the decay rate in the thyroid 24 d later is measured to be 17.0 Bq, what percentage of the tracer went to that gland? (c) What isotope remains after the I-131 decays?

43.25 •• The unstable isotope ^{40}K is used for dating rock samples. Its half-life is 1.28×10^9 y. (a) How many decays occur per second in a sample containing 1.63×10^{-6} g of ^{40}K? (b) What is the activity of the sample in curies?

43.26 • As a health physicist, you are being consulted about a spill in a radiochemistry lab. The isotope spilled was 400 μCi of ^{131}Ba, which has a half-life of 12 days. (a) What mass of ^{131}Ba was spilled? (b) Your recommendation is to clear the lab until the radiation level has fallen 1.00 μCi. How long will the lab have to be closed?

43.27 • Measurements on a certain isotope tell you that the decay rate decreases from 8318 decays/min to 3091 decays/min in 4.00 days. What is the half-life of this isotope?

43.28 • A radioactive isotope has a half-life of 43.0 min. At $t = 0$ its activity is 0.376 Ci. What is its activity at $t = 2.00$ h?

43.29 • The radioactive nuclide ^{199}Pt has a half-life of 30.8 minutes. A sample is prepared that has an initial activity of 7.56×10^{11} Bq. (a) How many ^{199}Pt nuclei are initially present in the sample? (b) How many are present after 30.8 minutes? What is the activity at this time? (c) Repeat part (b) for a time 92.4 minutes after the sample is first prepared.

43.30 •• **Radiocarbon Dating.** At an archeological site, a sample from timbers containing 500 g of carbon provides 2690 decays/min. What is the age of the sample?

Section 43.5 Biological Effects of Radiation

43.31 •• BIO (a) If a chest x ray delivers 0.25 mSv to 5.0 kg of tissue, how many *total* joules of energy does this tissue receive? (b) Natural radiation and cosmic rays deliver about 0.10 mSv per year at sea level. Assuming an RBE of 1, how many rem and rads is this dose, and how many joules of energy does a 75-kg person receive in a year? (c) How many chest x rays like the one in part (a) would it take to deliver the same *total* amount of energy to a 75-kg person as she receives from natural radiation in a year at sea level, as described in part (b)?

43.32 • BIO **Radiation Overdose.** If a person's entire body is exposed to 5.0 J/kg of x rays, death usually follows within a few days. (a) Express this lethal radiation dose in Gy, rad, Sv, and rem. (b) How much total energy does a 70.0-kg person absorb from such a dose? (c) If the 5.0 J/kg came from a beam of protons instead of x rays, what would be the answers to parts (a) and (b)?

43.33 •• BIO A nuclear chemist receives an accidental radiation dose of 5.0 Gy from slow neutrons $(\text{RBE} = 4.0)$. What does she receive in rad, rem, and J/kg?

43.34 •• BIO A person exposed to fast neutrons receives a radiation dose of 300 rem on part of his hand, affecting 25 g of tissue. The RBE of these neutrons is 10. (a) How many rad did he receive? (b) How many joules of energy did he receive? (c) Suppose the person received the same rad dosage, but from beta rays with an RBE of 1.0 instead of neutrons. How many rem would he have received?

43.35 • BIO **Food Irradiation.** Food is often irradiated with either x rays or electron beams to help prevent spoilage. A low dose of 5–75 kilorads (krad) helps to reduce and kill inactive parasites, a medium dose of 100–400 krad kills microorganisms and pathogens such as salmonella, and a high dose of 2300–5700 krad sterilizes food so that it can be stored without refrigeration. (a) A dose of 175 krad kills spoilage microorganisms in fish. If x rays are used, what would be the dose in Gy, Sv, and rem, and how much energy would a 220-g portion of fish absorb? (See Table 43.3.) (b) Repeat part (a) if electrons of RBE 1.50 are used instead of x rays.

43.36 • BIO **To Scan or Not to Scan?** It has become popular for some people to have yearly whole-body scans (CT scans, formerly called CAT scans) using x rays, just to see if they detect anything suspicious. A number of medical people have recently questioned the advisability of such scans, due in part to the radiation they impart. Typically, one such scan gives a dose of 12 mSv, applied to the *whole body*. By contrast, a chest x ray typically administers 0.20 mSv to only 5.0 kg of tissue. How many chest x rays would deliver the same *total* amount of energy to the body of a 75-kg person as one whole-body scan?

43.37 •• BIO A 67-kg person accidentally ingests 0.35 Ci of tritium. (a) Assume that the tritium spreads uniformly throughout the body and that each decay leads on the average to the absorption of 5.0 keV of energy from the electrons emitted in the decay. The half-life of tritium is 12.3 y, and the RBE of the electrons is 1.0. Calculate the absorbed dose in rad and the equivalent dose in rem during one week. (b) The β^- decay of tritium releases more than 5.0 keV of energy. Why is the average energy absorbed less than the total energy released in the decay?

43.38 • BIO In an industrial accident a 65-kg person receives a lethal whole-body equivalent dose of 5.4 Sv from x rays. (a) What is the equivalent dose in rem? (b) What is the absorbed dose in rad? (c) What is the total energy absorbed by the person's body? How does this amount of energy compare to the amount of energy required to raise the temperature of 65 kg of water 0.010 C°?

43.39 •• **CP BIO** In a diagnostic x-ray procedure, 5.00×10^{10} photons are absorbed by tissue with a mass of 0.600 kg. The x-ray wavelength is 0.0200 nm. (a) What is the total energy absorbed by the tissue? (b) What is the equivalent dose in rem?

Section 43.6 Nuclear Reactions
Section 43.7 Nuclear Fission
Section 43.8 Nuclear Fusion

43.40 •• Calculate the reaction energy Q for the reaction $p + {}^3_1H \rightarrow {}^2_1H + {}^2_1H$. Is this reaction exoergic or endoergic?

43.41 • Consider the nuclear reaction

$${}^2_1H + {}^9_4Be \rightarrow X + {}^4_2He$$

where X is a nuclide. (a) What are the values of Z and A for the nuclide X? (b) How much energy is liberated? (c) Estimate the threshold energy for this reaction.

43.42 • **Energy from Nuclear Fusion.** Calculate the energy released in the fusion reaction

$${}^3_2He + {}^2_1H \rightarrow {}^4_2He + {}^1_1H$$

43.43 •• At the beginning of Section 43.7 the equation of a fission process is given in which ${}^{235}U$ is struck by a neutron and undergoes fission to produce ${}^{144}Ba$, ${}^{89}Kr$, and three neutrons. The measured masses of these isotopes are 235.043930 u (${}^{235}U$), 143.922953 u (${}^{144}Ba$), 88.917631 u (${}^{89}Kr$), and 1.0086649 u (neutron). (a) Calculate the energy (in MeV) released by each fission reaction. (b) Calculate the energy released per gram of ${}^{235}U$, in MeV/g.

43.44 • The United States uses about 1.4×10^{19} J of electrical energy per year. If all this energy came from the fission of ${}^{235}U$, which releases 200 MeV per fission event, (a) how many kilograms of ${}^{235}U$ would be used per year, and (b) how many kilograms of uranium would have to be mined per year to provide that much ${}^{235}U$? (Recall that only 0.70% of naturally occurring uranium is ${}^{235}U$.)

43.45 • Consider the nuclear reaction

$${}^4_2He + {}^7_3Li \rightarrow X + {}^1_0n$$

where X is a nuclide. (a) What are Z and A for the nuclide X? (b) Is energy absorbed or liberated? How much?

43.46 •• Consider the nuclear reaction

$${}^{28}_{14}Si + \gamma \rightarrow {}^{24}_{12}Mg + X$$

where X is a nuclide. (a) What are Z and A for the nuclide X? (b) Ignoring the effects of recoil, what minimum energy must the photon have for this reaction to occur? The mass of a ${}^{28}_{14}Si$ atom is 27.976927 u, and the mass of a ${}^{24}_{12}Mg$ atom is 23.985042 u.

PROBLEMS

43.47 • **Comparison of Energy Released per Gram of Fuel.** (a) When gasoline is burned, it releases 1.3×10^8 J of energy per gallon (3.788 L). Given that the density of gasoline is 737 kg/m^3, express the quantity of energy released in J/g of fuel. (b) During fission, when a neutron is absorbed by a ${}^{235}U$ nucleus, about 200 MeV of energy is released for each nucleus that undergoes fission. Express this quantity in J/g of fuel. (c) In the proton-proton chain that takes place in stars like our sun, the overall fusion reaction can be summarized as six protons fusing to form one ${}^4He$ nucleus with two leftover protons and the liberation of 26.7 MeV of energy. The fuel is the six protons. Express the energy produced here in units of J/g of fuel. Notice the huge difference between the two forms of nuclear energy, on the one hand, and the chemical energy from gasoline, on the other. (d) Our sun produces energy at a measured rate of 3.86×10^{26} W. If its mass of 1.99×10^{30} kg were all gasoline, how long could it last before consuming all its fuel? (*Historical note:* Before the discovery of nuclear fusion and the vast amounts of energy it releases, scientists were confused. They knew that the earth was at least many millions of years old, but could not explain how the sun could survive that long if its energy came from chemical burning.)

43.48 •• (a) Calculate the minimum energy required to remove one proton from the nucleus ${}^{12}_6C$. This is called the proton-removal energy. (*Hint:* Find the difference between the mass of a ${}^{12}_6C$ nucleus and the mass of a proton plus the mass of the nucleus formed when a proton is removed from ${}^{12}_6C$.) (b) How does the proton-removal energy for ${}^{12}_6C$ compare to the binding energy per nucleon for ${}^{12}_6C$, calculated using Eq. (43.10)?

43.49 •• (a) Calculate the minimum energy required to remove one neutron from the nucleus ${}^{17}_8O$. This is called the neutron-removal energy. (See Problem 43.48.) (b) How does the neutron-removal energy for ${}^{17}_8O$ compare to the binding energy per nucleon for ${}^{17}_8O$, calculated using Eq. (43.10)?

43.50 •• The isotope ${}^{110}_{47}Ag$ is created by irradiation of a sample of ${}^{109}_{47}Ag$ nuclei with neutrons. The ${}^{109}_{47}Ag$ nucleus is stable and the number of ${}^{109}_{47}Ag$ nuclei in the sample is large, so we take the number of ${}^{109}_{47}Ag$ nuclei to be constant. Therefore, with a constant flux of neutrons, the ${}^{110}_{47}Ag$ nuclei are produced at a constant rate of 8.40×10^3 nuclei per second. The ${}^{110}_{47}Ag$ isotope decays by β^- emission to the stable nucleus ${}^{110}_{48}Cd$, with a half-life of 24.6 s. Initially, at $t = 0$, only ${}^{109}_{47}Ag$ nuclei are present. (a) After steady state has been reached and the number of ${}^{110}_{47}Ag$ nuclei in the sample is constant, how many ${}^{110}_{47}Ag$ nuclei are there? (b) How many ${}^{110}_{47}Ag$ are there in the sample at $t = 24.6$ s? (*Hint:* Refer to the Bridging Problem for this chapter.)

43.51 • **BIO Radioactive Fallout.** One of the problems of in-air testing of nuclear weapons (or, even worse, the *use* of such weapons!) is the danger of radioactive fallout. One of the most problematic nuclides in such fallout is strontium-90 (${}^{90}Sr$), which breaks down by β^- decay with a half-life of 28 years. It is chemically similar to calcium and therefore can be incorporated into bones and teeth, where, due to its rather long half-life, it remains for years as an internal source of radiation. (a) What is the daughter nucleus of the ${}^{90}Sr$ decay? (b) What percentage of the original level of ${}^{90}Sr$ is left after 56 years? (c) How long would you have to wait for the original level to be reduced to 6.25% of its original value?

43.52 •• **CP** Thorium ${}^{230}_{90}Th$ decays to radium ${}^{226}_{88}Ra$ by α emission. The masses of the neutral atoms are 230.033134 u for ${}^{230}_{90}Th$ and 226.025410 u for ${}^{226}_{88}Ra$. If the parent thorium nucleus is at rest, what is the kinetic energy of the emitted α particle? (Be sure to account for the recoil of the daughter nucleus.)

43.53 •• The atomic mass of ${}^{25}_{12}Mg$ is 24.985837 u, and the atomic mass of ${}^{25}_{13}Al$ is 24.990428 u. (a) Which of these nuclei will decay into the other? (b) What type of decay will occur? Explain how you determined this. (c) How much energy (in MeV) is released in the decay?

43.54 •• The polonium isotope ${}^{210}_{84}Po$ has atomic mass 209.982874 u. Other atomic masses are ${}^{206}_{82}Pb$, 205.974465 u; ${}^{209}_{83}Bi$, 208.980399 u; ${}^{210}_{83}Bi$, 209.984120 u; ${}^{209}_{84}Po$, 208.982430 u; and ${}^{210}_{85}At$, 209.987148 u. (a) Show that the alpha decay of ${}^{210}_{84}Po$ is energetically possible, and find the energy of the emitted

α particle. (b) Is $^{210}_{84}Po$ energetically stable with respect to emission of a proton? Why or why not? (c) Is $^{210}_{84}Po$ energetically stable with respect to emission of a neutron? Why or why not? (d) Is $^{210}_{84}Po$ energetically stable with respect to β^- decay? Why or why not? (e) Is $^{210}_{84}Po$ energetically stable with respect to β^+ decay? Why or why not?

43.55 •• BIO **Irradiating Ourselves!** The radiocarbon in our bodies is one of the naturally occurring sources of radiation. Let's see how large a dose we receive. ^{14}C decays via β^- emission, and 18% of our body's mass is carbon. (a) Write out the decay scheme of carbon-14 and show the end product. (A neutrino is also produced.) (b) Neglecting the effects of the neutrino, how much kinetic energy (in MeV) is released per decay? The atomic mass of ^{14}C is 14.003242 u. (c) How many grams of carbon are there in a 75-kg person? How many decays per second does this carbon produce? (*Hint:* Use data from Example 43.9.) (d) Assuming that all the energy released in these decays is absorbed by the body, how many MeV/s and J/s does the ^{14}C release in this person's body? (e) Consult Table 43.3 and use the largest appropriate RBE for the particles involved. What radiation dose does the person give himself in a year, in Gy, rad, Sv, and rem?

43.56 •• BIO **Pion Radiation Therapy.** A neutral pion (π^0) has a mass of 264 times the electron mass and decays with a lifetime of 8.4×10^{-17} s to two photons. Such pions are used in the radiation treatment of some cancers. (a) Find the energy and wavelength of these photons. In which part of the electromagnetic spectrum do they lie? What is the RBE for these photons? (b) If you want to deliver a dose of 200 rem (which is typical) in a single treatment to 25 g of tumor tissue, how many π^0 mesons are needed?

43.57 •• Calculate the mass defect for the β^+ decay of $^{11}_{6}C$. Is this decay energetically possible? Why or why not? The atomic mass of $^{11}_{6}C$ is 11.011434 u.

43.58 •• BIO A person ingests an amount of a radioactive source that has a very long lifetime and activity 0.52 μCi. The radioactive material lodges in her lungs, where all of the emitted 4.0-MeV α particles are absorbed within a 0.50-kg mass of tissue. Calculate the absorbed dose and the equivalent dose for one year.

43.59 • **We Are Stardust.** In 1952 spectral lines of the element technetium-99 (^{99}Tc) were discovered in a red giant star. Red giants are very old stars, often around 10 billion years old, and near the end of their lives. Technetium has *no* stable isotopes, and the half-life of ^{99}Tc is 200,000 years. (a) For how many half-lives has the ^{99}Tc been in the red giant star if its age is 10 billion years? (b) What fraction of the original ^{99}Tc would be left at the end of that time? This discovery was extremely important because it provided convincing evidence for the theory (now essentially known to be true) that most of the atoms heavier than hydrogen and helium were made inside of stars by thermonuclear fusion and other nuclear processes. If the ^{99}Tc had been part of the star since it was born, the amount remaining after 10 billion years would have been so minute that it would not have been detectable. This knowledge is what led the late astronomer Carl Sagan to proclaim that "we are stardust."

43.60 • BIO A 70.0-kg person experiences a whole-body exposure to α radiation with energy 4.77 MeV. A total of 7.75×10^{12} α particles are absorbed. (a) What is the absorbed dose in rad? (b) What is the equivalent dose in rem? (c) If the source is 0.0320 g of ^{226}Ra (half-life 1600 y) somewhere in the body, what is the activity of this source? (d) If all of the alpha particles produced are absorbed, what time is required for this dose to be delivered?

43.61 •• Measurements indicate that 27.83% of all rubidium atoms currently on the earth are the radioactive ^{87}Rb isotope. The rest are the stable ^{85}Rb isotope. The half-life of ^{87}Rb is 4.75×10^{10} y. Assuming that no rubidium atoms have been formed since, what percentage of rubidium atoms were ^{87}Rb when our solar system was formed 4.6×10^9 y ago?

43.62 • The nucleus $^{15}_{8}O$ has a half-life of 122.2 s; $^{19}_{8}O$ has a half-life of 26.9 s. If at some time a sample contains equal amounts of $^{15}_{8}O$ and $^{19}_{8}O$, what is the ratio of $^{15}_{8}O$ to $^{19}_{8}O$ (a) after 3.0 min and (b) after 12.0 min?

43.63 •• BIO A ^{60}Co source with activity 2.6×10^{-4} Ci is embedded in a tumor that has mass 0.200 kg. The source emits γ photons with average energy 1.25 MeV. Half the photons are absorbed in the tumor, and half escape. (a) What energy is delivered to the tumor per second? (b) What absorbed dose (in rad) is delivered per second? (c) What equivalent dose (in rem) is delivered per second if the RBE for these γ rays is 0.70? (d) What exposure time is required for an equivalent dose of 200 rem?

43.64 •• **An Oceanographic Tracer.** Nuclear weapons tests in the 1950s and 1960s released significant amounts of radioactive tritium ($^{3}_{1}H$, half-life 12.3 years) into the atmosphere. The tritium atoms were quickly bound into water molecules and rained out of the air, most of them ending up in the ocean. For any of this tritium-tagged water that sinks below the surface, the amount of time during which it has been isolated from the surface can be calculated by measuring the ratio of the decay product, $^{3}_{2}He$, to the remaining tritium in the water. For example, if the ratio of $^{3}_{2}He$ to $^{3}_{1}H$ in a sample of water is 1:1, the water has been below the surface for one half-life, or approximately 12 years. This method has provided oceanographers with a convenient way to trace the movements of subsurface currents in parts of the ocean. Suppose that in a particular sample of water, the ratio of $^{3}_{2}He$ to $^{3}_{1}H$ is 4.3 to 1.0. How many years ago did this water sink below the surface?

43.65 • A bone fragment found in a cave believed to have been inhabited by early humans contains 0.29 times as much ^{14}C as an equal amount of carbon in the atmosphere when the organism containing the bone died. (See Example 43.9 in Section 43.4.) Find the approximate age of the fragment.

43.66 •• BIO In the 1986 disaster at the Chernobyl reactor in eastern Europe, about $\frac{1}{8}$ of the ^{137}Cs present in the reactor was released. The isotope ^{137}Cs has a half-life of 30.07 y for β decay, with the emission of a total of 1.17 MeV of energy per decay. Of this, 0.51 MeV goes to the emitted electron; the remaining 0.66 MeV goes to a γ ray. The radioactive ^{137}Cs is absorbed by plants, which are eaten by livestock and humans. How many ^{137}Cs atoms would need to be present in each kilogram of body tissue if an equivalent dose for one week is 3.5 Sv? Assume that all of the energy from the decay is deposited in 1.0 kg of tissue and that the RBE of the electrons is 1.5.

43.67 •• Consider the fusion reaction $^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{2}He + ^{1}_{0}n$. (a) Estimate the barrier energy by calculating the repulsive electrostatic potential energy of the two $^{2}_{1}H$ nuclei when they touch. (b) Compute the energy liberated in this reaction in MeV and in joules. (c) Compute the energy liberated *per mole* of deuterium, remembering that the gas is diatomic, and compare with the heat of combustion of hydrogen, about 2.9×10^5 J/mol.

43.68 •• DATA As a scientist in a nuclear physics research lab, you are conducting a photodisintegration experiment to verify the binding energy of a deuteron. A photon with wavelength λ in air is absorbed by a deuteron, which breaks apart into a neutron and a proton. The two fragments share the released kinetic energy equally, and the deuteron can be assumed to be initially at rest. You measure the speed of the proton after the disintegration as a function of the wavelength λ of the photon. Your experimental results are given in the table.

λ (10^{-13} m)	3.50	3.75	4.00	4.25	4.50	4.75	5.00
v (10^6 m/s)	11.3	10.2	9.1	8.1	7.2	6.1	4.9

(a) Graph the data as v^2 versus $1/\lambda$. Explain why the data points, when graphed this way, should follow close to a straight line. Find the slope and y-intercept of the straight line that gives the best fit to the data. (b) Assume that h and c have their accepted values. Use your results from part (a) to calculate the mass of the proton and the binding energy (in MeV) of the deuteron.

43.69 •• **DATA** Your company develops radioactive isotopes for medical applications. In your work there, you measure the activity of a radioactive sample. Your results are given in the table.

Time (h)	Decays/s
0	20,000
0.5	14,800
1.0	11,000
1.5	8130
2.0	6020
2.5	4460
3.0	3300
4.0	1810
5.0	1000
6.0	550
7.0	300

(a) Find the half-life of the sample. (b) How many radioactive nuclei were present in the sample at $t = 0$? (c) How many were present after 7.0 h?

43.70 ••• **DATA** In your job as a health physicist, you measure the activity of a mixed sample of radioactive elements. Your results are given in the table.

Time (h)	Decays/s
0	7500
0.5	4120
1.0	2570
1.5	1790
2.0	1350
2.5	1070
3.0	872
4.0	596
5.0	404
6.0	288
7.0	201
8.0	140
9.0	98
10.0	68
12.0	33

(a) What minimum number of different nuclides are present in the mixture? (b) What are their half-lives? (c) How many nuclei of each type are initially present in the sample? (d) How many of each type are present at $t = 5.0$ h?

CHALLENGE PROBLEMS

43.71 ••• **Industrial Radioactivity.** Radioisotopes are used in a variety of manufacturing and testing techniques. Wear measurements can be made using the following method. An automobile engine is produced using piston rings with a total mass of 100 g, which includes 9.4 μCi of ^{59}Fe whose half-life is 45 days. The engine is test-run for 1000 hours, after which the oil is drained and its activity is measured. If the activity of the engine oil is 84 decays/s, how much mass was worn from the piston rings per hour of operation?

43.72 ••• Many radioactive decays occur within a sequence of decays—for example, $^{234}_{92}\text{U} \rightarrow {}^{230}_{88}\text{Th} \rightarrow {}^{226}_{84}\text{Ra}$. The half-life for the $^{234}_{92}\text{U} \rightarrow {}^{230}_{88}\text{Th}$ decay is 2.46×10^5 y, and the half-life for the $^{230}_{88}\text{Th} \rightarrow {}^{226}_{84}\text{Ra}$ decay is 7.54×10^4 y. Let 1 refer to $^{234}_{92}\text{U}$, 2 to $^{230}_{88}\text{Th}$, and 3 to $^{226}_{84}\text{Ra}$; let λ_1 be the decay constant for the $^{234}_{92}\text{U} \rightarrow {}^{230}_{88}\text{Th}$ decay and λ_2 be the decay constant for the $^{230}_{88}\text{Th} \rightarrow {}^{226}_{84}\text{Ra}$ decay. The amount of $^{230}_{88}\text{Th}$ present at any time depends on the rate at which it is produced by the decay of $^{234}_{92}\text{U}$ and the rate by which it is depleted by its decay to $^{226}_{84}\text{Ra}$. Therefore, $dN_2(t)/dt = \lambda_1 N_1(t) - \lambda_2 N_2(t)$. If we start with a sample that contains N_{10} nuclei of $^{234}_{92}\text{U}$ and nothing else, then $N(t) = N_{01}e^{-\lambda_1 t}$. Thus $dN_2(t)/dt = \lambda_1 N_{10}e^{-\lambda_1 t} - \lambda_2 N_2(t)$. This differential equation for $N_2(t)$ can be solved as follows. Assume a trial solution of the form $N_2(t) = N_{10}[h_1 e^{-\lambda_1 t} + h_2 e^{-\lambda_2 t}]$, where h_1 and h_2 are constants. (a) Since $N_2(0) = 0$, what must be the relationship between h_1 and h_2? (b) Use the trial solution to calculate $dN_2(t)/dt$, and substitute that into the differential equation for $N_2(t)$. Collect the coefficients of $e^{-\lambda_1 t}$ and $e^{-\lambda_2 t}$. Since the equation must hold at all t, each of these coefficients must be zero. Use this requirement to solve for h_1 and thereby complete the determination of $N_2(t)$. (c) At time $t = 0$, you have a pure sample containing 30.0 g of $^{234}_{92}\text{U}$ and nothing else. What mass of $^{230}_{88}\text{Th}$ is present at time $t = 2.46 \times 10^5$ y, the half-life for the $^{234}_{92}\text{U}$ decay?

PASSAGE PROBLEMS

BIO RADIOACTIVE IODINE IN MEDICINE. Iodine in the body is preferentially taken up by the thyroid gland. Therefore, radioactive iodine in small doses is used to image the thyroid and in large doses is used to kill thyroid cells to treat some types of cancer or thyroid disease. The iodine isotopes used have relatively short half-lives, so they must be produced in a nuclear reactor or accelerator. One isotope frequently used for imaging is ^{123}I; it has a half-life of 13.2 h and emits a 0.16-MeV gamma-ray photon. One method of producing ^{123}I is in the nuclear reaction $^{123}\text{Te} + \text{p} \rightarrow {}^{123}\text{I} + \text{n}$. The atomic masses relevant to this reaction are ^{123}Te, 122.904270 u; ^{123}I, 122.905589 u; n, 1.008665 u; and ^{1}H, 1.007825 u.

The iodine isotope commonly used for treatment of disease is ^{131}I, which is produced by irradiating ^{130}Te in a nuclear reactor to form ^{131}Te. The ^{131}Te then decays to ^{131}I. ^{131}I undergoes β^- decay with a half-life of 8.04 d, emitting electrons with energies up to 0.61 MeV and gamma-ray photons of energy 0.36 MeV. A typical thyroid cancer treatment might involve administration of 3.7 GBq of ^{131}I.

43.73 Which reaction produces ^{131}Te in the nuclear reactor? (a) $^{130}\text{Te} + \text{n} \rightarrow {}^{131}\text{Te}$; (b) $^{130}\text{I} + \text{n} \rightarrow {}^{131}\text{Te}$; (c) $^{132}\text{Te} + \text{n} \rightarrow {}^{131}\text{Te}$; (d) $^{132}\text{I} + \text{n} \rightarrow {}^{131}\text{Te}$.

43.74 Which type of radioactive decay produces ^{131}I from ^{131}Te? (a) Alpha decay; (b) β^- decay; (c) β^+ decay; (d) gamma decay.

43.75 How many ^{131}I atoms are administered in a typical thyroid cancer treatment? (a) 4.2×10^{10}; (b) 1.0×10^{12}; (c) 2.5×10^{14}; (d) 3.7×10^{15}.

43.76 In the reaction that produces ^{123}I, is there a minimum kinetic energy the protons need to make the reaction go? (a) No, because the proton has a smaller mass than the neutron. (b) No, because the total initial mass is smaller than the total final mass. (c) Yes, because the proton has a smaller mass than the neutron. (d) Yes, because the total initial mass is smaller than the total final mass.

43.77 Why might ^{123}I be preferred for imaging over ^{131}I? (a) The atomic mass of ^{123}I is smaller, so the ^{123}I particles travel farther through tissue. (b) Because ^{123}I emits only gamma-ray photons, the radiation dose to the body is lower with that isotope. (c) The beta particles emitted by ^{131}I can leave the body, whereas the gamma-ray photons emitted by ^{123}I cannot. (d) ^{123}I is radioactive, whereas ^{131}I is not.

Answers

Chapter Opening Question ?

(iv) When an organism dies, it stops taking in carbon from atmospheric CO_2. Some of this carbon is radioactive ^{14}C, which decays with a half-life of 5730 years. By measuring the proportion of ^{14}C that remains in the specimen, scientists can determine how long ago the organism died. (See Section 43.4.)

Test Your Understanding Questions

43.1 (a) (iii), (b) (v) The radius R is proportional to the cube root of the mass number A, while the volume is proportional to R^3 and hence to $(A^{1/3})^3 = A$. Therefore, doubling the volume requires increasing the mass number by a factor of 2; doubling the radius implies increasing both the volume and the mass number by a factor of $2^3 = 8$.

43.2 (ii), (iii), (iv), (v), (i) You can find the answers by inspecting Fig. 43.2. The binding energy per nucleon is lowest for very light nuclei such as $^4_2$He, is greatest around $A = 60$, and then decreases with increasing A.

43.3 (v) Two protons and two neutrons are lost in an α decay, so Z and N each decrease by 2. A β^+ decay changes a proton to a neutron, so Z decreases by 1 and N increases by 1. The net result is that Z decreases by 3 and N decreases by 1.

43.4 (iii) The activity $-dN(t)/dt$ of a sample is the product of the number of nuclei in the sample $N(t)$ and the decay constant $\lambda = (\ln 2)/T_{1/2}$. Hence $N(t) = [-dN(t)/dt]T_{1/2}/(\ln 2)$. Taking the ratio of this expression for ^{240}Pu to this same expression for ^{243}Am, the factors of $\ln 2$ cancel and we get

$$\frac{N_{\text{Pu}}}{N_{\text{Am}}} = \frac{(-dN_{\text{Pu}}/dt)T_{1/2\text{-Pu}}}{(-dN_{\text{Am}}/dt)T_{1/2\text{-Am}}} = \frac{(5.00\ \mu\text{Ci})(6560\ \text{y})}{(4.45\ \mu\text{Ci})(7370\ \text{y})} = 1.00$$

The two samples contain *equal* numbers of nuclei. The ^{243}Am sample has a longer half-life and hence a slower decay rate, so it has a lower activity than the ^{240}Pu sample.

43.5 (ii) We saw in Section 43.3 that alpha particles can travel only a very short distance before they are stopped. By contrast, x-ray photons are very penetrating, so they can easily pass into the body.

43.6 no The reaction $^1_1\text{H} + ^7_3\text{Li} \rightarrow ^4_2\text{He} + ^4_2\text{He}$ is a *nuclear* reaction that can take place only if a proton (a hydrogen nucleus) comes into contact with a lithium nucleus. If the hydrogen is in atomic form, the interaction between its electron cloud and the electron cloud of a lithium atom keeps the two nuclei from getting close to each other. Even if isolated protons are used, they must be fired at the lithium atoms with enough kinetic energy to overcome the electrical repulsion between the protons and the lithium nuclei. The statement that the reaction is exoergic means that more energy is released by the reaction than had to be put in to make the reaction occur.

43.7 no Because the neutron has no electric charge, it experiences no electrical repulsion from a ^{235}U nucleus. Hence a slow-moving neutron can approach and enter a ^{235}U nucleus, thereby providing the excitation needed to trigger fission. By contrast, a slow-moving *proton* (charge $+e$) feels a strong electrical repulsion from a ^{235}U nucleus (charge $+92e$). It never gets close to the nucleus, so it cannot trigger fission.

43.8 no Fusion reactions between sufficiently light nuclei are exoergic because the binding energy per nucleon E_B/A increases. If the nuclei are too massive, however, E_B/A decreases and fusion is *endoergic* (i.e., it takes in energy rather than releasing it). As an example, imagine fusing together two nuclei of $A = 100$ to make a single nucleus with $A = 200$. From Fig. 43.2, E_B/A is more than 8.5 MeV for the $A = 100$ nuclei but is less than 8 MeV for the $A = 200$ nucleus. Such a fusion reaction is possible, but requires a substantial input of energy.

Bridging Problem

(a) ^{128}Xe
(b) no; β^+ emission would be endoergic
(c) 3.25×10^9 atoms, 1.50×10^6 Bq
(d) $N(t) = (3.25 \times 10^9\ \text{atoms})(1 - e^{-(4.62\times10^{-4}\ \text{s}^{-1})t})$

? This image shows a portion of the Eagle Nebula, a region some 6500 light-years away where new stars are forming. The luminous stars, glowing gas, and opaque dust clouds are all made of "normal" matter—that is, atoms and their constituents. What percentage of the mass and energy in the universe is composed of "normal" matter? (i) 75% to 100%; (ii) 50% to 75%; (iii) 25% to 50%; (iv) 5% to 25%; (v) less than 5%.

44 PARTICLE PHYSICS AND COSMOLOGY

LEARNING GOALS

Looking forward at ...

44.1 The key varieties of fundamental subatomic particles and how they were discovered.

44.2 How physicists use accelerators and detectors to probe the properties of subatomic particles.

44.3 The four ways in which subatomic particles interact with each other.

44.4 How the structure of protons, neutrons, and other particles can be explained in terms of quarks.

44.5 The standard model of particles and interactions.

44.6 The evidence that the universe is expanding and that the expansion is speeding up.

44.7 The history of the first 380,000 years after the Big Bang.

Looking back at ...

13.3 Escape speed.

27.4 Motion of charged particles in a magnetic field.

32.1 Radiation from accelerated charges.

38.1, 38.3, 38.4 Photons; electron–positron annihilation; uncertainty principle.

39.1, 39.2 Electron waves; discovery of the nucleus.

41.5, 41.6 Electron spin; exclusion principle.

42.6 Valence bands and holes.

43.1, 43.3 Neutrons and protons; β^+ decay.

What are the most fundamental constituents of matter? How did the universe begin? And what is the fate of our universe? In this chapter we will explore what physicists and astronomers have learned in their quest to answer these questions.

The chapter title, "Particle Physics and Cosmology," may seem strange. Fundamental particles are the *smallest* things in the universe, and cosmology deals with the *biggest* thing there is—the universe itself. Nonetheless, we'll see in this chapter that physics on the most microscopic scale plays an essential role in determining the nature of the universe on the largest scale.

The development of high-energy accelerators and associated detectors has been crucial in our emerging understanding of particles. We can classify particles and their interactions in several ways in terms of conservation laws and symmetries, some of which are absolute and others of which are obeyed only in certain kinds of interactions. We'll conclude by discussing our present understanding of the nature and evolution of the universe as a whole.

44.1 FUNDAMENTAL PARTICLES—A HISTORY

The idea that the world is made of fundamental particles has a long history. In about 400 B.C. the Greek philosophers Democritus and Leucippus suggested that matter is made of indivisible particles that they called *atoms,* a word derived from *a-* (not) and *tomos* (cut or divided). This idea lay dormant until about 1804, when the English scientist John Dalton (1766–1844), often called the father of modern chemistry, discovered that many chemical phenomena could be explained if atoms of each element are the basic, indivisible building blocks of matter.

The Electron and the Proton

Toward the end of the 19th century it became clear that atoms are *not* indivisible. The characteristic spectra of elements suggested that atoms have internal structure, and J. J. Thomson's discovery of the negatively charged *electron* in 1897 showed that atoms could be taken apart into charged particles. Rutherford's experiments in 1910–11 (see Section 39.2) revealed that an atom's positive charge resides in a small, dense nucleus. In 1919 Rutherford made an additional discovery: When alpha particles are fired into nitrogen, one product is hydrogen gas. He reasoned that the hydrogen nucleus is a constituent of the nuclei of heavier atoms such as nitrogen, and that a collision with a fast-moving alpha particle can dislodge one of those hydrogen nuclei. Thus the hydrogen nucleus is an elementary particle that Rutherford named the *proton*. The following decade saw the blossoming of quantum mechanics, including the Schrödinger equation. Physicists were on their way to understanding the principles that underlie atomic structure.

The Photon

Einstein explained the photoelectric effect in 1905 by assuming that the energy of electromagnetic waves is quantized; that is, it comes in little bundles called *photons* with energy $E = hf$. Atoms and nuclei can emit (create) and absorb (destroy) photons (see Section 38.1). Considered as particles, photons have zero charge and zero rest mass. (Note that any discussions of a particle's mass in this chapter will refer to its rest mass.) In particle physics, a photon is denoted by the symbol γ (the Greek letter gamma).

The Neutron

In 1930 the German physicists Walther Bothe and Herbert Becker observed that when beryllium, boron, or lithium was bombarded by alpha particles, the target material emitted a radiation that had much greater penetrating power than the original alpha particles. Experiments by the English physicist James Chadwick in 1932 showed that the emitted particles were electrically neutral, with mass approximately equal to that of the proton. Chadwick christened these particles *neutrons* (symbol n or ${}^{1}_{0}\mathrm{n}$). A typical reaction of the type studied by Bothe and Becker, with a beryllium target, is

$$ {}^{4}_{2}\mathrm{He} + {}^{9}_{4}\mathrm{Be} \rightarrow {}^{12}_{6}\mathrm{C} + {}^{1}_{0}\mathrm{n} \tag{44.1} $$

Elementary particles are usually detected by their electromagnetic effects—for instance, by the ionization that they cause when they pass through matter. (This is the principle of the cloud chamber, described below.) Because neutrons have no charge, they are difficult to detect directly; they interact hardly at all with electrons and produce little ionization when they pass through matter. However, neutrons can be slowed down by scattering from nuclei, and they can penetrate a nucleus. Hence slow neutrons can be detected by means of a nuclear reaction in which a neutron is absorbed and an alpha particle is emitted. An example is

$$ {}^{1}_{0}\mathrm{n} + {}^{10}_{5}\mathrm{B} \rightarrow {}^{7}_{3}\mathrm{Li} + {}^{4}_{2}\mathrm{He} \tag{44.2} $$

The ejected alpha particle is easy to detect because it is charged. Later experiments showed that neutrons and protons, like electrons, are spin-$\frac{1}{2}$ particles (see Section 43.1).

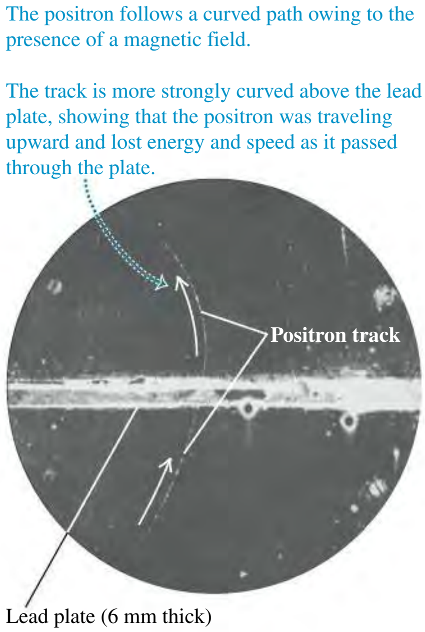
The discovery of the neutron cleared up a mystery about the composition of the nucleus. Before 1930 the mass of a nucleus was thought to be due only to protons, but no one understood why the charge-to-mass ratio was not the same for all nuclides. It soon became clear that all nuclides (except ${}^{1}_{1}\mathrm{H}$) contain both protons and neutrons. Hence the proton, the neutron, and the electron are the building blocks of atoms. However, that is not the end of the particle story; these are not the only particles, and particles can do more than build atoms.

The Positron

The positive electron, or positron, was discovered by the American physicist Carl D. Anderson in 1932, during an investigation of particles bombarding the earth from space. **Figure 44.1** shows a historic photograph made with a *cloud chamber,* an instrument used to visualize the tracks of charged particles. The chamber contained a supercooled vapor; a charged particle passing through the vapor causes ionization, and the ions trigger the condensation of liquid droplets from the vapor. The droplets make a visible track showing the charged particle's path.

The cloud chamber in Fig. 44.1 is in a magnetic field directed into the plane of the photograph. The particle has passed through a thin lead plate (which extends from left to right in the figure) that lies within the chamber. The track is more tightly curved above the plate than below it, showing that the speed was less above the plate than below it. Therefore the particle had to be moving upward; it could not have gained energy passing through the lead. The thickness and curvature of the track suggested that its mass and the magnitude of its charge equaled those of the electron. But the directions of the magnetic field and the velocity in the magnetic force equation $\vec{F} = q\vec{v} \times \vec{B}$ showed that the particle had *positive* charge. Anderson christened this particle the *positron.*

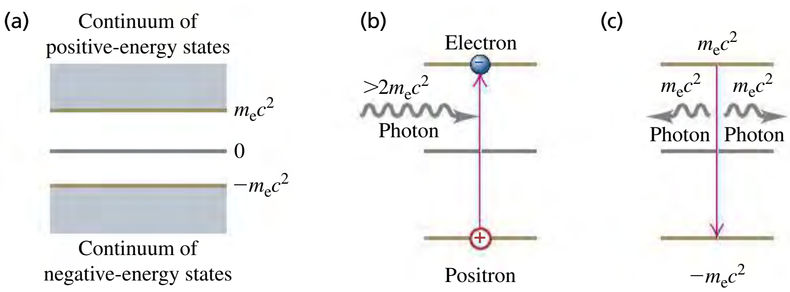
44.1 Photograph of the cloud-chamber track made by the first positron ever identified. The photograph was made by Carl D. Anderson in 1932.



To theorists, the appearance of the positron was a welcome development. In 1928 the English physicist Paul Dirac had developed a relativistic generalization of the Schrödinger equation for the electron. In Section 41.5 we discussed how Dirac's ideas helped explain the spin magnetic moment of the electron.

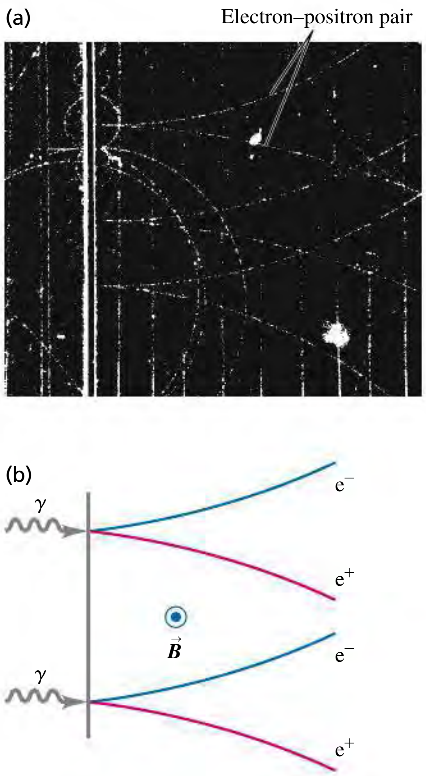
One of the puzzling features of the Dirac equation was that for a *free* electron it predicted not only a continuum of energy states greater than its rest energy $m_e c^2$, as expected, but also a continuum of *negative* energy states *less than* $-m_e c^2$ (**Fig. 44.2a**). That posed a problem. What was to prevent an electron from emitting a photon with energy $2m_e c^2$ or greater and hopping from a positive state to a negative state? It wasn't clear what these negative-energy states meant, and there was no obvious way to get rid of them. Dirac's ingenious interpretation was that all the negative-energy states were filled with electrons, and that these electrons were for some reason unobservable. The exclusion principle (see Section 41.6) would forbid a transition to a state that was already occupied.

A vacancy in a negative-energy state would act like a positive charge, just as a hole in the valence band of a semiconductor (see Section 42.6) acts like a positive charge. Initially, Dirac tried to argue that such vacancies were protons. But after Anderson's discovery it became clear that the vacancies were observed physically as *positrons.* Furthermore, the Dirac energy-state picture provides a mechanism for the *creation* of positrons. When an electron in a negative-energy state absorbs a photon with energy greater than $2m_e c^2$, it goes to a positive state (Fig. 44.2b), in which it becomes observable. The vacancy that it leaves behind is observed as a positron; the result is the creation of an electron–positron pair. Similarly, when an electron in a positive-energy state falls into a vacancy, both the electron and the vacancy (that is, the positron) disappear, and photons are emitted (Fig. 44.2c). Thus the Dirac theory leads naturally to the conclusion that, like photons, *electrons can be created and destroyed.* While photons can be created and destroyed singly, electrons can be produced or destroyed only in electron–positron pairs or in association with other particles. (Creating or destroying an electron alone would mean creating or destroying an amount of charge $-e$, which would violate the conservation of electric charge.)

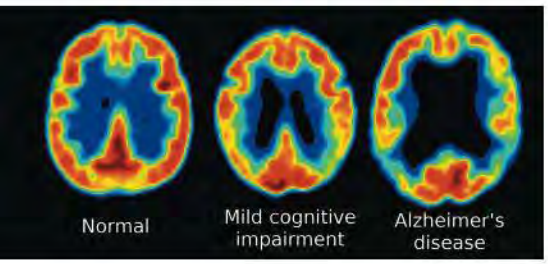


44.2 (a) Energy states for a free electron predicted by the Dirac equation. (b) Raising an electron from an $E < 0$ state to an $E > 0$ state corresponds to electron–positron pair production. (c) An electron dropping from an $E > 0$ state to a vacant $E < 0$ state corresponds to electron–positron pair annihilation.

44.3 (a) Photograph of bubble-chamber tracks of electron–positron pairs that are produced when 300-MeV photons strike a lead sheet. A magnetic field directed out of the photograph made the electrons and positrons curve in opposite directions. (b) Diagram showing the pair-production process for two of the photons.



BIO Application Pair Annihilation in Medical Diagnosis A technique called positron emission tomography (PET) can be used to identify the early stages of Alzheimer's disease. A patient is administered a glucose-like compound called FDG in which one oxygen atom is replaced by radioactive ^{18}F. FDG accumulates in active areas of the brain, where glucose metabolism is high. The ^{18}F undergoes β^+ decay (positron emission) with a half-life of 110 minutes, and the emitted positron immediately annihilates with an atomic electron to produce two gamma-ray photons. A scanner detects both photons, then calculates where the annihilation took place—the site of FDG accumulation. These PET images—which show areas of strongest emission, and hence greatest glucose metabolism, in red—reveal changes in the brains of patients.



In 1949 the American physicist Richard Feynman showed that a positron could be described mathematically as an electron traveling backward in time. His reformulation of the Dirac theory eliminated difficult calculations involving the infinite sea of negative-energy states and put electrons and positrons on the same footing. But the creation and destruction of electron–positron pairs remain. The Dirac theory provides the beginning of a theoretical framework for creation and destruction of all fundamental particles.

Experiment and theory tell us that the masses of the positron and electron are identical and that their charges are equal in magnitude but opposite in sign. The positron's spin angular momentum $\vec{S}$ and magnetic moment $\vec{\mu}$ are parallel; they are opposite for the electron. However, $\vec{S}$ and $\vec{\mu}$ have the same magnitude for both particles because they have the same spin. We use the term **antiparticle** for a particle that is related to another particle as the positron is to the electron. Each kind of particle has a corresponding antiparticle. For a few kinds of particles (necessarily all neutral) the particle and antiparticle are identical, and we can say that they are their own antiparticles. The photon is an example; there is no way to distinguish a photon from an antiphoton. We'll use the standard symbols e^- for the electron and e^+ for the positron. The generic term "electron" often includes both electrons and positrons. Other antiparticles may be denoted by a bar over the particle's symbol; for example, an antiproton is $\overline{p}$. We'll see other examples of antiparticles later.

Positrons do not occur in ordinary matter. Electron–positron pairs are produced during high-energy collisions of charged particles or γ rays with matter. This process is called e^+e^- *pair production* (**Fig. 44.3**). The minimum available energy required for electron–positron pair production equals the rest energy $2m_ec^2$ of the two particles:

$$E_{\text{min}} = 2m_ec^2 = 2(9.109 \times 10^{-31}\ \text{kg})(2.998 \times 10^8\ \text{m/s})^2$$
$$= 1.637 \times 10^{-13}\ \text{J} = 1.022\ \text{MeV}$$

The inverse process, e^+e^- *pair annihilation*, occurs when a positron and an electron collide (see Example 38.6 in Section 38.3). Both particles disappear, and two (or occasionally three) photons can appear, with total energy of at least $2m_ec^2 = 1.022$ MeV. Decay into a *single* photon is impossible: Such a process could not conserve both energy and momentum.

Positrons also occur in the decay of some unstable nuclei, in which they are called beta-plus particles (β^+). We discussed β^+ decay in Section 43.3.

We'll frequently represent particle masses in terms of the equivalent rest energy by using $m = E/c^2$. Then typical mass units are MeV/c^2; for example, $m = 0.511\ \text{MeV}/c^2$ for an electron or positron.

Particles As Force Mediators

In classical physics we describe the interaction of charged particles in terms of electric and magnetic forces. In quantum mechanics we can describe this interaction in terms of emission and absorption of photons. Two electrons repel each other as one emits a photon and the other absorbs it, just as two skaters can push each other apart by tossing a heavy ball back and forth between them (**Fig. 44.4a**). For an electron and a proton, in which the charges are opposite and the force is attractive, we imagine the skaters trying to grab the ball away from each other (Fig. 44.4b). The electromagnetic interaction between two charged particles is *mediated* or transmitted by photons.

If charged-particle interactions are mediated by photons, where does the energy to create the photons come from? Recall from our discussion of the uncertainty principle (see Sections 38.4 and 39.6) that a state that exists for a short time Δt has an uncertainty ΔE in its energy such that

$$\Delta E\,\Delta t \geq \frac{\hbar}{2} \qquad (44.3)$$

This uncertainty permits the creation of a photon with energy ΔE, provided that it lives no longer than the time Δt given by Eq. (44.3). A photon that can exist for a short time because of this energy uncertainty is called a *virtual photon*. It's as though there were an energy bank; you can borrow energy, provided that you pay it back within the time limit. According to Eq. (44.3), the more you borrow, the sooner you have to pay it back.

Mesons

Is there a particle that mediates the *nuclear* force? By the mid-1930s the nuclear force between two nucleons (neutrons or protons) appeared to be described by a potential energy $U(r)$ with the general form

$$U(r) = -f^2\left(\frac{e^{-r/r_0}}{r}\right) \qquad \text{(nuclear potential energy)} \qquad (44.4)$$

The constant f characterizes the strength of the interaction, and r_0 describes its range. **Figure 44.5** compares the absolute value of this function with the function f^2/r, which would be analogous to the *electric* interaction of two protons:

$$U(r) = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r} \qquad \text{(electric potential energy)} \qquad (44.5)$$

In 1935 the Japanese physicist Hideki Yukawa suggested that a hypothetical particle that he called a **meson** might mediate the nuclear force. He showed that the range of the force was related to the mass of the particle. Yukawa argued that the particle must live for a time Δt long enough to travel a distance comparable to the range r_0 of the nuclear force. This range was known from the sizes of nuclei and other information to be about 1.5×10^{-15} m $=$ 1.5 fm. If we assume that an average particle's speed is comparable to c and travels about half the range, its lifetime Δt must be about

$$\Delta t = \frac{r_0}{2c} = \frac{1.5 \times 10^{-15}\ \text{m}}{2(3.0 \times 10^8\ \text{m/s})} = 2.5 \times 10^{-24}\ \text{s}$$

From Eq. (44.3), the minimum necessary uncertainty ΔE in energy is

$$\Delta E = \frac{\hbar}{2\Delta t} = \frac{1.05 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(2.5 \times 10^{-24}\ \text{s})} = 2.1 \times 10^{-11}\ \text{J} = 130\ \text{MeV}$$

The mass equivalent Δm of this energy is about 250 times the electron mass:

$$\Delta m = \frac{\Delta E}{c^2} = \frac{2.1 \times 10^{-11}\ \text{J}}{(3.00 \times 10^8\ \text{m/s})^2} = 2.3 \times 10^{-28}\ \text{kg} = 130\ \text{MeV}/c^2$$

Yukawa postulated that an as yet undiscovered particle with this mass serves as the messenger for the nuclear force.

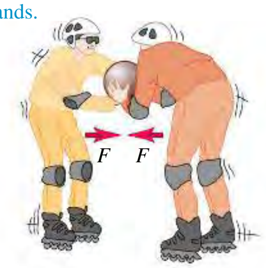
A year later, Carl Anderson and his colleague Seth Neddermeyer discovered in cosmic radiation two new particles, now called **muons.** The μ^- has charge equal to that of the electron, and its antiparticle the μ^+ has a positive charge with equal magnitude. The two particles have equal mass, about 207 times the electron mass. But it soon became clear that muons were *not* Yukawa's particles because they interacted with nuclei only very weakly.

In 1947 a family of three particles, called π *mesons* or **pions,** were discovered. Their charges are $+e$, $-e$, and zero, and their masses are about 270 times the electron mass. The pions interact strongly with nuclei, and they *are* the particles predicted by Yukawa. Other, heavier mesons, the ω and ρ, evidently also act as shorter-range messengers of the nuclear force. The complexity of this explanation suggests that the nuclear force has simpler underpinnings; these involve the quarks

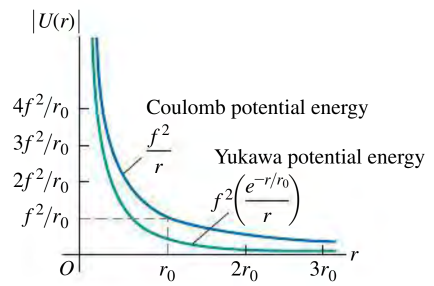
44.4 An analogy for how particles act as force mediators.

(a) Two skaters exert repulsive forces on each other by tossing a ball back and forth.

(b) Two skaters exert attractive forces on each other when one tries to grab the ball out of the other's hands.



44.5 Graph of the magnitude of the Yukawa potential-energy function for nuclear forces, $|U(r)| = f^2e^{-r/r_0}/r$. The function $U(r) = f^2/r$, proportional to the potential energy for Coulomb's law, is also shown. The two functions are similar at small r, but the Yukawa potential energy drops off much more quickly at large r.



and gluons that we'll discuss in Section 44.4. Before discussing mesons further, we'll describe some particle accelerators and detectors to see how mesons and other particles are created in a controlled fashion and observed.

TEST YOUR UNDERSTANDING OF SECTION 44.1 Each of the following particles can be exchanged between two protons, two neutrons, or a neutron and a proton as part of the nuclear force. Rank the particles in order of the range of the interaction that they mediate, from largest to smallest range. (i) The π^+ (pi-plus) meson, mass 140 MeV/c^2; (ii) the ρ^+ (rho-plus) meson, mass 776 MeV/c^2; (iii) the η^0 (eta-zero) meson, mass 548 MeV/c^2; (iv) the ω^0 (omega-zero) meson, mass 783 MeV/c^2. ▌

44.2 PARTICLE ACCELERATORS AND DETECTORS

Early nuclear physicists used alpha and beta particles from naturally occurring radioactive elements for their experiments, but they were restricted in energy to the few MeV that are available in such random decays. Present-day particle accelerators can produce precisely controlled beams of particles, from electrons and positrons up to heavy ions, with a wide range of energies. These beams have three main uses. First, high-energy particles can collide to produce new particles, just as a collision of an electron and a positron can produce photons. Second, a high-energy particle has a short de Broglie wavelength and so can probe the small-scale interior structure of other particles, just as electron microscopes (see Section 39.1) can give better resolution than optical microscopes. Third, they can be used to produce nuclear reactions of scientific or medical use.

BIO Application Linear Accelerators in Medicine Electron linear accelerators that provide a kinetic energy of 4–20 MeV are important tools in the treatment of many cancers. The electrons themselves are used to irradiate superficial tumors. Alternatively, the electrons can be directed at a metal target; then bremsstrahlung (see Section 38.2) produces x rays that are used to irradiate tumors that lie deeper inside the patient.

Linear Accelerators

Particle accelerators use electric and magnetic fields to accelerate and guide beams of charged particles. A *linear accelerator* (linac) accelerates particles in a straight line. J. J. Thomson's cathode-ray tubes were early examples of linacs. Modern linacs use a series of electrodes with gaps to give the particles a series of boosts. Most present-day high-energy linear accelerators use a traveling electromagnetic wave; the charged particles "ride" the wave in more or less the way that a surfer rides an incoming ocean wave. In the highest-energy linac in the world today, at the SLAC National Accelerator Laboratory, electrons and positrons can be accelerated to 50 GeV in a tube 3 km long. At this energy their de Broglie wavelengths are 0.025 fm, much smaller than the size of a proton or a neutron.

The Cyclotron

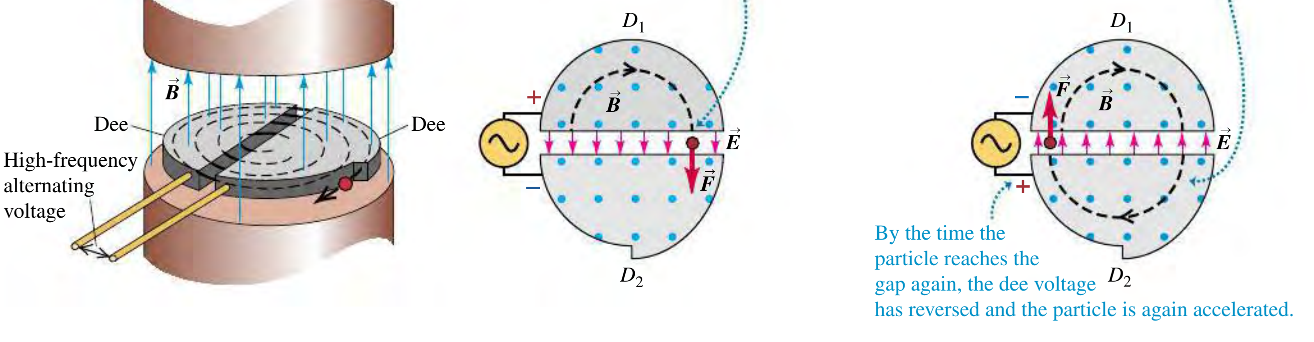
Many accelerators use magnets to deflect the charged particles into circular paths. The first was the *cyclotron,* invented in 1931 by E. O. Lawrence and M. Stanley Livingston at the University of California (**Fig. 44.6a**). Particles with mass m and

44.6 Layout and operation of a cyclotron.

(a) Schematic diagram of a cyclotron

(b) As the positive particle reaches the gap, it is accelerated by the electric-field force ...

(c) ... and the next semicircular orbit has a larger radius.



charge q move inside a vacuum chamber in a uniform magnetic field $\vec{B}$ that is perpendicular to the plane of their paths. In Section 27.4 we showed that in such a field, a particle with speed v moves in a circular path with radius r given by

$$r = \frac{mv}{|q|B} \tag{44.6}$$

and with angular speed (angular frequency) ω given by

$$\omega = \frac{v}{r} = \frac{|q|B}{m} \tag{44.7}$$

An alternating potential difference is applied between the two hollow electrodes D_1 and D_2 (called *dees*), creating an electric field in the gap between them. The polarity of the potential difference and electric field is changed precisely twice each revolution (Figs. 44.6b and 44.6c), so that the particles get a push each time they cross the gap. The pushes increase their speed and kinetic energy, boosting them into paths of larger radius. The maximum speed v_{max} and kinetic energy K_{max} are determined by the radius R of the largest possible path. Solving Eq. (44.6) for v, we find $v = |q|Br/m$ and $v_{max} = |q|BR/m$. Assuming nonrelativistic speeds, we have

$$K_{max} = \tfrac{1}{2}mv_{max}{}^2 = \frac{q^2B^2R^2}{2m} \tag{44.8}$$

EXAMPLE 44.1 FREQUENCY AND ENERGY IN A PROTON CYCLOTRON

One cyclotron built during the 1930s has a path of maximum radius 0.500 m and a magnetic field of magnitude 1.50 T. If it is used to accelerate protons, find (a) the frequency of the alternating voltage applied to the dees and (b) the maximum particle energy.

SOLUTION

IDENTIFY and SET UP: The frequency f of the applied voltage must equal the frequency of the proton orbital motion. Equation (44.7) gives the *angular* frequency ω of the proton orbital motion; we find f from $f = \omega/2\pi$. The proton reaches its maximum energy K_{max}, given by Eq. (44.8), when the radius of its orbit equals the radius of the dees.

EXECUTE: (a) For protons, $q = 1.60 \times 10^{-19}$ C and $m = 1.67 \times 10^{-27}$ kg. From Eq. (44.7),

$$f = \frac{\omega}{2\pi} = \frac{|q|B}{2\pi m} = \frac{(1.60 \times 10^{-19}\ \text{C})(1.50\ \text{T})}{2\pi(1.67 \times 10^{-27}\ \text{kg})}$$
$$= 2.3 \times 10^7\ \text{Hz} = 23\ \text{MHz}$$

(b) From Eq. (44.8) the maximum kinetic energy is

$$K_{max} = \frac{(1.60 \times 10^{-19}\ \text{C})^2(1.50\ \text{T})^2(0.50\ \text{m})^2}{2(1.67 \times 10^{-27}\ \text{kg})}$$
$$= 4.3 \times 10^{-12}\ \text{J} = 2.7 \times 10^7\ \text{eV} = 27\ \text{MeV}$$

This proton kinetic energy is much larger than that available from natural radioactive sources.

EVALUATE: From Eq. (44.6) or Eq. (44.7), the proton speed is $v = 7.2 \times 10^7$ m/s, which is about 25% of the speed of light. At such speeds, relativistic effects are beginning to become important. Since we ignored these effects in our calculation, the results for f and K_{max} are in error by a few percent; this is why we kept only two significant figures.

The maximum energy that can be attained with a cyclotron is limited by relativistic effects. The relativistic version of Eq. (44.7) is

$$\omega = \frac{|q|B}{m}\sqrt{1 - v^2/c^2}$$

As the particles speed up, their angular frequency ω *decreases,* and their motion gets out of phase with the alternating dee voltage. In the *synchrocyclotron* the particles are accelerated in bursts. For each burst, the frequency of the alternating voltage is decreased as the particles speed up, maintaining the correct phase relationship with the particles' motion.

Another limitation of the cyclotron is the difficulty of building very large electromagnets. The largest synchrocyclotron ever built has a vacuum chamber that is about 8 m in diameter and accelerates protons to energies of about 600 MeV.

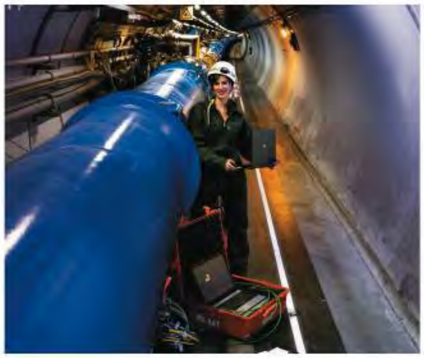
The Synchrotron

To attain higher energies, a type of machine called the *synchrotron* is more practical. Particles move in a vacuum chamber in the form of a thin doughnut called the *accelerating ring.* The particle beam is bent to follow the ring by a series of electromagnets placed around the ring. As the particles speed up, the magnetic field is increased so that the particles retrace the same trajectory over and over. The Large Hadron Collider (LHC) near Geneva, Switzerland, is the highest-energy accelerator in the world (**Fig. 44.7**). It is designed to accelerate protons to a maximum energy of 7 TeV, or 7×10^{12} eV. (As we'll see in Section 44.3, *hadrons* are a class of elementary particles that includes protons and neutrons.)

44.7 (a) The Large Hadron Collider at the European Organization for Nuclear Research (CERN). The underground accelerating ring (shown by the red circle) is 100 m underground and 8.5 km in diameter, so large that it spans the border between Switzerland and France. (Note the Alps in the background.) When accelerated to 7 TeV, protons travel around the ring more than 11,000 times per second. (b) An engineer working on one of the 9593 superconducting electromagnets around the LHC ring.

(a)

(b)



As we pointed out in Section 32.1, accelerated charges radiate electromagnetic energy. In an accelerator in which the particles move in curved paths, this radiation is often called *synchrotron radiation*. High-energy accelerators are typically constructed underground to provide protection from this radiation. From the accelerator standpoint, synchrotron radiation is undesirable, since the energy given to an accelerated particle is radiated right back out. It can be minimized by making the accelerator radius r large so that the centripetal acceleration v^2/r is small. On the positive side, synchrotron radiation is used as a source of well-controlled high-frequency electromagnetic waves.

Available Energy

When a beam of high-energy particles collides with a stationary target, not all the kinetic energy of the incident particles is *available* to form new particle states. Because momentum must be conserved, the particles emerging from the collision must have some net motion and thus some kinetic energy. The discussion following Example 43.11 (Section 43.6) presented a nonrelativistic example of this principle. The maximum available energy is the kinetic energy in the frame of reference in which the total momentum is zero. We call this the *center-of-momentum system;* it is the relativistic generalization of the center-of-mass system that we discussed in Section 8.5. In this system the total kinetic energy after the collision can be zero, so that the maximum amount of the initial kinetic energy becomes available to cause the reaction being studied.

Consider the *laboratory system,* in which a target particle with mass M is initially at rest and is bombarded by a particle with mass m and total energy (including rest energy) E_m. The total available energy E_a in the center-of-momentum system (including rest energies of all the particles) can be shown to be

$$E_a^2 = 2Mc^2E_m + (Mc^2)^2 + (mc^2)^2 \qquad \text{(available energy)} \tag{44.9}$$

When the masses of the target and projectile particles are equal, we can simplify:

$$E_a^2 = 2mc^2(E_m + mc^2) \qquad \text{(available energy, equal masses)} \tag{44.10}$$

If in addition E_m is much greater than mc^2, we can ignore the second term in the parentheses in Eq (44.10). Then E_a is

$$E_a = \sqrt{2mc^2E_m} \qquad \text{(available energy, equal masses, } E_m \gg mc^2\text{)} \tag{44.11}$$

The square root in Eq. (44.11) is a disappointing result for an accelerator designer: Doubling the energy E_m of the bombarding particle increases the available energy E_a by only a factor of $\sqrt{2} = 1.414$. Examples 44.2 and 44.3 explore the limitations of having a stationary target particle.

EXAMPLE 44.2 THRESHOLD ENERGY FOR PION PRODUCTION

SOLUTION

A proton (rest energy 938 MeV) with kinetic energy K collides with a proton at rest. Both protons survive the collision, and a neutral pion (π^0, rest energy 135 MeV) is produced. What is the threshold energy (minimum value of K) for this process?

SOLUTION

IDENTIFY and SET UP: The final state includes the two original protons (mass m) and the pion (mass m_π). The threshold energy corresponds to the minimum-energy case in which all three particles are at rest in the center-of-momentum system. The total available energy E_a in that system must be at least the total rest energy, $2mc^2 + m_\pi c^2$. We use this to solve Eq. (44.10) for the total energy E_m of the bombarding proton; the kinetic energy K (our target variable) is then E_m minus the proton rest energy mc^2.

EXECUTE: We substitute $E_a = 2mc^2 + m_\pi c^2$ into Eq. (44.10), simplify, and solve for E_m:

$$4m^2c^4 + 4mm_\pi c^4 + m_\pi^2 c^4 = 2mc^2 E_m + 2(mc^2)^2$$

$$E_m = mc^2 + m_\pi c^2\left(2 + \frac{m_\pi}{2m}\right) = mc^2 + K$$

$$K = m_\pi c^2\left(2 + \frac{m_\pi}{2m}\right)$$

We see that the bombarding proton's kinetic energy K must be somewhat greater than twice the pion rest energy $m_\pi c^2$. With $mc^2 = 938$ MeV and $m_\pi c^2 = 135$ MeV, we have $m_\pi/2m = 0.072$ and

$$K = (135\text{ MeV})(2 + 0.072) = 280\text{ MeV}$$

EVALUATE: Compare this result with the result of Example 37.11 (Section 37.8), where we found that a pion can be produced in a head-on collision of two protons, each with only 67.5 MeV of kinetic energy. We discuss the energy advantage of such collisions in the next subsection.

EXAMPLE 44.3 INCREASING THE AVAILABLE ENERGY

SOLUTION

The Fermilab accelerator in Illinois was designed to bombard stationary targets with 800-GeV protons. (a) What is the available energy E_a in a proton-proton collision? (b) What is E_a if the beam energy is increased to 980 GeV?

SOLUTION

IDENTIFY and SET UP: Our target variable is the available energy E_a in a stationary-target collision between identical particles. In both parts (a) and (b) the beam energy E_m is much larger than the proton rest energy $mc^2 = 938\text{ MeV} = 0.938$ GeV, so we can safely use the approximation of Eq. (44.11).

EXECUTE: (a) For $E_m = 800$ GeV, Eq. (44.11) gives

$$E_a = \sqrt{2(0.938\text{ GeV})(800\text{ GeV})} = 38.7\text{ GeV}$$

(b) For $E_m = 980$ GeV,

$$E_a = \sqrt{2(0.938\text{ GeV})(980\text{ GeV})} = 42.9\text{ GeV}$$

EVALUATE: With a stationary-proton target, increasing the proton beam energy by 180 GeV increases the available energy by only 4.2 GeV! This shows a major limitation of experiments in which one of the colliding particles is initially at rest. Below we describe how physicists can overcome this limitation.

Colliding Beams

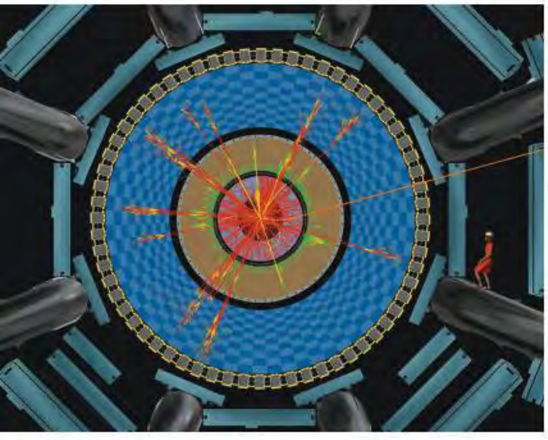
The limitation illustrated by Example 44.3 is circumvented in *colliding-beam* experiments. In these experiments there is no stationary target; instead, beams of particles moving in opposite directions are tightly focused onto one another so that head-on collisions can occur. Usually the two colliding particles have momenta of equal magnitude and opposite direction, so the total momentum is zero. Hence the laboratory system is also the center-of-momentum system, and the available energy is maximized.

The highest-energy colliding beams available are those at the Large Hadron Collider (see Fig. 44.7). In operation, 2808 bunches of 7-TeV protons circulate around the ring, half in one direction and half in the opposite direction. Each bunch contains about 10^{11} protons. Magnets steer the oppositely moving bunches to collide at interaction points. The available energy E_a in the resulting head-on collisions is the *total* energy of the two colliding particles: $E_a = 2 \times 7\text{ TeV} = 14$ TeV. (Strictly, E_a is 14 TeV minus the rest energy of the two colliding protons. But this rest energy is only $2mc^2 = 2(938\text{ MeV}) = 1.876 \times 10^{-3}$ TeV, which is so small compared to 14 TeV that it can be ignored.) The very large available energy at the Large Hadron Collider makes it possible to produce particles that have never been seen before (see Section 44.5).

Detectors

A wide variety of devices have been designed to measure the properties of subatomic particles. Many detectors use the ionization caused by charged particles as they move through a gas, liquid, or solid. The ions along the particle's path give rise to droplets of liquid in the supersaturated vapor of a cloud chamber (Fig. 44.1) or cause small volumes of vapor in the superheated liquid of a bubble chamber (Fig. 44.3a). In a semiconducting solid the ionization can take the form of electron–hole pairs. We discussed their detection in Section 42.7. *Wire chambers* contain arrays of closely spaced wires that detect the ions. The charge collected and time information from each wire are processed by using computers to reconstruct the particle trajectories. The detectors at the Large Hadron Collider use an array of devices to follow the tracks of particles produced by collisions between protons (**Fig. 44.8**). The giant solenoid in the photo that opens Chapter 28 is at the heart of one these detector arrays. The intense magnetic field of the solenoid helps identify newly produced particles, which curve in different directions and along paths of different radii depending on their charge and energy.

44.8 This computer-generated image shows the result of a simulated collision between two protons (not shown) in one of the interaction regions at the Large Hadron Collider. The view is along the beampipe. The different color tracks show different types of particles emerging from the collision. A variety of different detectors surround the collision region. (Note the woman in a red dress, drawn for scale.)



Cosmic-Ray Experiments

Large numbers of particles called *cosmic rays* continually bombard the earth from sources both within and beyond our galaxy. These particles consist mostly of neutrinos, protons, and heavier nuclei, with energies ranging from less than 1 MeV to more than 10^{20} eV. The earth's atmosphere and magnetic field protect us from much of this radiation. This means that cosmic-ray experimentation often must be carried out above all or most of the atmosphere by means of rockets or high-altitude balloons.

In contrast, neutrino detectors are buried below the earth's surface in tunnels or mines or submerged deep in the ocean. This is done to screen out all other types of particles so that only neutrinos, which interact only very weakly with matter, reach the detector. Because neutrino interactions with matter are so weak, neutrino detectors must consist of huge amounts of matter: The Super-Kamiokande detector looks for flashes of light produced when a neutrino interacts in a tank containing 5×10^7 kg of water (see Section 44.5).

Cosmic rays were important in early particle physics, and their study currently brings us important information about the rest of the universe. Although cosmic rays provide a source of high-energy particles that does not depend on expensive accelerators, most particle physicists use accelerators because the high-energy cosmic-ray particles they want are too few and too random.

DATA *SPEAKS*

Particle Collisions

When students were given a problem involving collisions between elementary particles, more than 46% gave an incorrect response. Common errors:

- Confusion about the energy released in particle–antiparticle annihilation. If a particle of mass m collides with and annihilates its antiparticle, the released energy is greater than or equal to the *combined* rest energy $2mc^2$ of the particle and antiparticle.
- Confusion about available energy. To produce a new particle of mass m in a collision, the *available* energy must be at least mc^2. If the target particle is at rest, the available energy can be far less than the kinetic energy of the bombarding particle.

TEST YOUR UNDERSTANDING OF SECTION 44.2 In a colliding-beam experiment, a 90-GeV electron collides head-on with a 90-GeV positron. The electron and the positron annihilate each other, forming a single virtual photon that then transforms into other particles. Does the virtual photon obey the same relationship $E = pc$ as real photons do?

44.3 PARTICLES AND INTERACTIONS

We have mentioned the array of subatomic particles that were known as of 1947: photons, electrons, positrons, protons, neutrons, muons, and pions. Since then, literally hundreds of additional particles have been discovered in accelerator experiments. The vast majority of known particles are *unstable* and decay spontaneously into other particles. Particles of all kinds, whether stable or unstable, can be created or destroyed in interactions between particles. Each such interaction involves the exchange of virtual particles, which exist on borrowed energy allowed by the uncertainty principle.

Although the world of subatomic particles and their interactions is complex, some key results bring order and simplicity to the seeming chaos. One key simplification is that there are only four fundamental types of interactions, each mediated or transmitted by the exchange of certain characteristic virtual particles. Furthermore, not all particles respond to all four kinds of interaction. In this section we will examine the fundamental interactions more closely and see how physicists classify particles in terms of the ways in which they interact.

Four Forces and Their Mediating Particles

In Section 5.5 we first described the four fundamental types of forces or interactions (**Fig. 44.9**). They are, in order of decreasing strength:

1. The strong interaction
2. The electromagnetic interaction
3. The weak interaction
4. The gravitational interaction

44.9 The ties that bind us together originate in the fundamental interactions of nature. The nuclei within our bodies are held together by the strong interaction. The electromagnetic interaction binds nuclei and electrons together to form atoms, binds atoms together to form molecules, and binds molecules together to form us.

The *electromagnetic* and *gravitational* interactions are familiar from classical physics. Both are characterized by a $1/r^2$ dependence on distance. In this scheme, the mediating particles for both interactions have mass zero and are stable as ordinary particles. The mediating particle for the electromagnetic interaction is the familiar photon, which has spin 1. (That means its spin quantum number is $s = 1$, so the magnitude of its spin angular momentum is $S = \sqrt{s(s+1)}\,\hbar = \sqrt{2}\,\hbar$.) The mediating particle for the gravitational force is the spin-2 *graviton* $(s = 2, S = \sqrt{s(s+1)}\,\hbar = \sqrt{6}\,\hbar)$. The graviton has not yet been observed experimentally because the gravitational force is very much weaker than the electromagnetic force. For example, the gravitational attraction of two protons is smaller than their electrical repulsion by a factor of about 10^{36}. The gravitational force is of primary importance in the structure of stars and the large-scale behavior of the universe, but it is not believed to play a significant role in particle interactions at the energies that are currently attainable.

The other two forces are less familiar. One, usually called the *strong interaction,* is responsible for the nuclear force and also for the production of pions and several other particles in high-energy collisions. At the most fundamental level, the mediating particle for the strong interaction is called a *gluon*. However, the force between nucleons is more easily described in terms of mesons as the mediating particles. We'll discuss the spin-1, massless gluon in Section 44.4.

Equation (44.4) is a possible potential-energy function for the nuclear force. The strength of the interaction is described by the constant f^2, which has units of energy times distance. A better basis for comparison with other forces is the dimensionless ratio $f^2/\hbar c$, called the *coupling constant* for the interaction. (We invite you to verify that this ratio is a pure number and so must have the same value in all systems of units.) The observed behavior of nuclear forces suggests that $f^2/\hbar c \approx 1$. The dimensionless coupling constant for *electromagnetic* interactions is the fine-structure constant, which we introduced in Section 41.5:

$$\frac{1}{4\pi\epsilon_0}\frac{e^2}{\hbar c} = 7.2974 \times 10^{-3} = \frac{1}{137.04} \tag{44.12}$$

Thus the strong interaction is roughly 100 times as strong as the electromagnetic interaction; however, it drops off with distance more quickly than $1/r^2$.

The fourth interaction is called the *weak* interaction. It is responsible for beta decay, such as the conversion of a neutron into a proton, an electron, and an antineutrino. It is also responsible for the decay of many unstable particles (pions into muons, muons into electrons, and so on). Its mediating particles are the short-lived particles W^+, W^-, and Z^0. The existence of these particles

TABLE 44.1 Four Fundamental Interactions

Interaction	Relative Strength	Range	Mediating Particle Name	Mass	Charge	Spin
Strong	1	Short ($\sim$1 fm)	Gluon	0	0	1
Electromagnetic	$\frac{1}{137.04}$	Long ($1/r^2$)	Photon	0	0	1
Weak	10^{-9}	Short ($\sim$0.001 fm)	$W^{\pm}$, Z^0	80.4, 91.2 GeV/c^2	$\pm e$, 0	1
Gravitational	10^{-38}	Long ($1/r^2$)	Graviton	0	0	2

was confirmed in 1983 in experiments at CERN, for which Carlo Rubbia and Simon van der Meer were awarded the Nobel Prize in 1984. The $W^{\pm}$ and Z^0 have spin 1 like the photon and the gluon, but they are *not* massless. In fact, they have enormous masses, $80.4\ \text{GeV}/c^2$ for the W's and $91.2\ \text{GeV}/c^2$ for the Z^0. With such massive mediating particles the weak interaction has a much shorter range than the strong interaction. It also lives up to its name by being weaker than the strong interaction by a factor of about 10^9.

Table 44.1 compares the main features of these four fundamental interactions.

More Particles

In Section 44.1 we mentioned the discoveries of muons in 1937 and of pions in 1947. The electric charges of the muons and the charged pions have the same magnitude e as the electron charge. The positive muon μ^+ is the antiparticle of the negative muon μ^-. Each has spin $\frac{1}{2}$, like the electron, and a mass of about $207m_e = 106\ \text{MeV}/c^2$. Muons are unstable; each decays with a lifetime of 2.2×10^{-6} s into an electron of the same sign, a neutrino, and an antineutrino.

There are three kinds of pions, all with spin 0; they have *no* spin angular momentum. The π^+ and π^- have masses of $273m_e = 140\ \text{MeV}/c^2$. They are unstable; each $\pi^{\pm}$ decays with a lifetime of 2.6×10^{-8} s into a muon of the same sign along with a neutrino for the π^+ and an antineutrino for the π^-. The π^0 is somewhat less massive, $264m_e = 135\ \text{MeV}/c^2$, and it decays with a lifetime of 8.4×10^{-17} s into two photons. The π^+ and π^- are antiparticles of one another, while the π^0 is its own antiparticle. (That is, there is no distinction between particle and antiparticle for the π^0.)

The existence of the *antiproton* $\overline{\text{p}}$ had been suspected ever since the discovery of the positron. The $\overline{\text{p}}$ was found in 1955, when proton–antiproton $(\text{p}\overline{\text{p}})$ pairs were created by use of a beam of 6-GeV protons from the Bevatron at the University of California, Berkeley. The *antineutron* $\overline{\text{n}}$ was found soon afterward. After 1960, as higher-energy accelerators and more sophisticated detectors were developed, a veritable blizzard of new unstable particles were identified. To describe and classify them, we need a small blizzard of new terms.

Initially, particles were classified by mass into three categories: (1) leptons ("light ones" such as electrons); (2) mesons ("intermediate ones" such as pions); and (3) baryons ("heavy ones" such as nucleons and more massive particles). But this scheme has been superseded by a more useful one in which particles are classified in terms of their *interactions*. For instance, *hadrons* (which include mesons and baryons) have strong interactions, and *leptons* do not.

In the following discussion we will also distinguish between **fermions,** which have half-integer spins, and **bosons,** which have zero or integer spins. Fermions obey the exclusion principle, on which the Fermi-Dirac distribution function (see Section 42.5) is based. Bosons do not obey the exclusion principle (there is no limit on how many bosons can occupy the same quantum state) and have a different distribution function, the Bose-Einstein distribution.

TABLE 44.2 The Six Leptons

Particle Name	Symbol	Anti-particle	Mass (MeV/c^2)	L_e	L_μ	L_τ	Lifetime (s)	Principal Decay Modes
Electron	e^-	e^+	0.511	+1	0	0	Stable	
Electron neutrino	ν_e	$\overline{\nu}_e$	$<2 \times 10^{-6}$	+1	0	0	Stable	
Muon	μ^-	μ^+	105.7	0	+1	0	2.20×10^{-6}	$e^- \overline{\nu}_e \nu_\mu$
Muon neutrino	ν_μ	$\overline{\nu}_\mu$	<0.19	0	+1	0	Stable	
Tau	τ^-	τ^+	1777	0	0	+1	2.9×10^{-13}	$\mu^- \overline{\nu}_\mu \nu_\tau$ or $e^- \overline{\nu}_e \nu_\tau$
Tau neutrino	ν_τ	$\overline{\nu}_\tau$	<18.2	0	0	+1	Stable	

Note: In addition to the limits on the individual neutrino masses, there is a much more stringent limit on the *sum* of the masses of the three types of neutrinos. Evidence suggests that this sum is less than about $3 \times 10^{-7}\ \mathrm{MeV}/c^2 = 0.3\ \mathrm{eV}/c^2$.

Leptons

The **leptons,** which do not have strong interactions, include six particles: the electron (e^-) and its neutrino (ν_e), the muon (μ^-) and its neutrino (ν_μ), and the tau particle (τ^-) and its neutrino (ν_τ). Each of these has a distinct antiparticle. All leptons have spin $\frac{1}{2}$ and thus are fermions. **Table 44.2** shows the family of leptons. The taus have mass $3478m_e = 1777\ \mathrm{MeV}/c^2$. Taus and muons are unstable; a τ^- decays into a μ^- plus a tau neutrino and a muon antineutrino, or an electron plus a tau neutrino and an electron antineutrino. A μ^- decays into an electron plus a muon neutrino and an electron antineutrino. They have relatively long lifetimes because their decays are mediated by the weak interaction. Despite their zero charge, a neutrino is distinct from an antineutrino; the spin angular momentum of a neutrino has a component that is opposite its linear momentum, while for an antineutrino that component is parallel to its linear momentum. Because neutrinos are so elusive, physicists have only been able to place upper limits on the rest masses of the ν_e, the ν_μ, and the ν_τ. It was thought that the rest masses of the neutrinos were zero; compelling recent evidence indicates that they have small but nonzero masses. We'll return to this point and its implications later.

Leptons obey a *conservation principle.* Corresponding to the three pairs of leptons are three lepton numbers L_e, L_μ, and L_τ. The electron e^- and the electron neutrino ν_e are assigned $L_e = 1$, and their antiparticles e^+ and $\overline{\nu}_e$ are given $L_e = -1$. Corresponding assignments of L_μ and L_τ are made for the μ and τ particles and their neutrinos. **In all interactions, each lepton number is separately conserved.** For example, in the decay of the μ^-, the lepton numbers are

$$\begin{array}{ccccccc} \mu^- & \rightarrow & e^- & + & \overline{\nu}_e & + & \nu_\mu \\ L_\mu = 1 & & L_e = 1 & & L_e = -1 & & L_\mu = 1 \end{array}$$

These conservation principles have no counterpart in classical physics.

EXAMPLE 44.4 LEPTON NUMBER CONSERVATION

SOLUTION

Check conservation of lepton numbers for these decay schemes:

(a) $\mu^+ \rightarrow e^+ + \nu_e + \overline{\nu}_\mu$

(b) $\pi^- \rightarrow \mu^- + \overline{\nu}_\mu$

(c) $\pi^0 \rightarrow \mu^- + e^+ + \nu_e$

SOLUTION

IDENTIFY and SET UP: Lepton number conservation requires that L_e, L_μ, and L_τ (given in Table 44.2) separately have the same sums after the decay as before.

EXECUTE: We tabulate L_e and L_μ for each decay scheme. An antiparticle has the opposite lepton number from its corresponding particle listed in Table 44.2. No τ particles or τ neutrinos appear in any of the schemes, so $L_\tau = 0$ both before and after each decay and L_τ is conserved.

(a) $\mu^+ \rightarrow e^+ + \nu_e + \overline{\nu}_\mu$

L_e: $0 = -1 + 1 + 0$

L_μ: $-1 = 0 + 0 + (-1)$

Continued

(b) $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$

L_e: $0 = 0 + 0$

L_μ: $0 = 1 + (-1)$

(c) $\pi^0 \rightarrow \mu^- + e^+ + \nu_e$

L_e: $0 = 0 + (-1) + 1$

L_μ: $0 \neq 1 + 0 + 0$

EVALUATE: Decays (a) and (b) are consistent with lepton number conservation and are observed. Decay (c) violates the conservation of L_μ and has *never* been observed. Physicists used these and other experimental results to deduce the principle that all three lepton numbers must separately be conserved.

Hadrons

Hadrons, the strongly interacting particles, are a more complex family than leptons. Each hadron has an antiparticle, often denoted with an overbar, as with the antiproton $\overline{\text{p}}$. There are two subclasses of hadrons: *mesons* and *baryons*. **Table 44.3** shows some of the many hadrons that are currently known. (We'll explain *strangeness* and *quark content* later in this section and in the next one.)

Mesons include the pions that have already been mentioned, K mesons or *kaons,* η mesons, and others that we will discuss later. Mesons have spin 0 or 1 and therefore are all bosons. There are no stable mesons; all mesons decay to less massive particles, obeying all the conservation laws for such decays.

Baryons include the nucleons and several particles called *hyperons,* including the Λ, Σ, Ξ, and Ω. These resemble nucleons but are more massive. Baryons have half-integer spin, and therefore all are fermions. The only stable baryon is the proton; a free neutron decays to a proton, and hyperons decay to other hyperons or to nucleons by various processes. Baryons obey the *conservation of baryon number,* analogous to conservation of lepton numbers, again with no counterpart in classical physics. We assign a baryon number $B = 1$ to each baryon (p, n, Λ, Σ, and so on) and $B = -1$ to each antibaryon ($\overline{\text{p}}$, $\overline{\text{n}}$, $\overline{\Lambda}$, $\overline{\Sigma}$, and so on).

In all interactions, the total baryon number is conserved.

This principle is the reason the mass number A was conserved in all of the nuclear reactions that we studied in Chapter 43.

TABLE 44.3 Some Hadrons and Their Properties

Particle	Mass (MeV/c^2)	Charge Ratio, Q/e	Spin	Baryon Number, B	Strangeness, S	Mean Lifetime (s)	Typical Decay Modes	Quark Content
Mesons								
π^0	135.0	0	0	0	0	8.4×10^{-17}	$\gamma\gamma$	$u\bar{u}$, $d\bar{d}$
π^+	139.6	+1	0	0	0	2.60×10^{-8}	$\mu^+\nu_\mu$	$u\bar{d}$
π^-	139.6	−1	0	0	0	2.60×10^{-8}	$\mu^-\bar{\nu}_\mu$	$\bar{u}d$
K^+	493.7	+1	0	0	+1	1.24×10^{-8}	$\mu^+\nu_\mu$	$u\bar{s}$
K^-	493.7	−1	0	0	−1	1.24×10^{-8}	$\mu^-\bar{\nu}_\mu$	$\bar{u}s$
η^0	547.3	0	0	0	0	$\approx 10^{-18}$	$\gamma\gamma$	$u\bar{u}$, $d\bar{d}$, $s\bar{s}$
Baryons								
p	938.3	+1	$\frac{1}{2}$	1	0	Stable	—	*uud*
n	939.6	0	$\frac{1}{2}$	1	0	886	$pe^-\bar{\nu}_e$	*udd*
Λ^0	1116	0	$\frac{1}{2}$	1	−1	2.63×10^{-10}	$p\pi^-$ or $n\pi^0$	*uds*
Σ^+	1189	+1	$\frac{1}{2}$	1	−1	8.02×10^{-11}	$p\pi^0$ or $n\pi^+$	*uus*
Σ^0	1193	0	$\frac{1}{2}$	1	−1	7.4×10^{-20}	$\Lambda^0\gamma$	*uds*
Σ^-	1197	−1	$\frac{1}{2}$	1	−1	1.48×10^{-10}	$n\pi^-$	*dds*
Ξ^0	1315	0	$\frac{1}{2}$	1	−2	2.90×10^{-10}	$\Lambda^0\pi^0$	*uss*
Ξ^-	1321	−1	$\frac{1}{2}$	1	−2	1.64×10^{-10}	$\Lambda^0\pi^-$	*dss*
Δ^{++}	1232	+2	$\frac{3}{2}$	1	0	$\approx 10^{-23}$	$p\pi^+$	*uuu*
Ω^-	1672	−1	$\frac{3}{2}$	1	−3	8.2×10^{-11}	Λ^0K^-	*sss*
Λ_c^+	2285	+1	$\frac{1}{2}$	1	0	2.0×10^{-13}	$pK^-\pi^+$	*udc*

EXAMPLE 44.5 BARYON NUMBER CONSERVATION

SOLUTION

Check conservation of baryon number for these reactions:

(a) $n + p \rightarrow n + p + p + \overline{p}$

(b) $n + p \rightarrow n + p + \overline{n}$

SOLUTION

IDENTIFY and SET UP: This example is similar to Example 44.4. We compare the total baryon number before and after each reaction, using data from Table 44.3.

EXECUTE: We tabulate the baryon numbers, noting that a baryon has $B = 1$ and an antibaryon has $B = -1$:

(a) $n + p \rightarrow n + p + p + \overline{p}$: $1 + 1 = 1 + 1 + 1 + (-1)$

(b) $n + p \rightarrow n + p + \overline{n}$: $1 + 1 \neq 1 + 1 + (-1)$

EVALUATE: Reaction (a) is consistent with baryon number conservation. It can occur if enough energy is available in the $n + p$ collision. Reaction (b) violates baryon number conservation and has never been observed.

EXAMPLE 44.6 ANTIPROTON CREATION

SOLUTION

What is the minimum proton energy required to produce an antiproton in a collision with a stationary proton?

SOLUTION

IDENTIFY and SET UP: The reaction must conserve baryon number, charge, and energy. Since the target and bombarding protons are of equal mass and the target is at rest, we determine the minimum energy E_m of the bombarding proton from Eq. (44.10).

EXECUTE: Conservation of charge and conservation of baryon number forbid the creation of an antiproton by itself; it must be created as part of a proton–antiproton pair. The complete reaction is

$$p + p \rightarrow p + p + p + \overline{p}$$

For this reaction to occur, the minimum available energy E_a in Eq. (44.10) is the final rest energy $4mc^2$ of three protons and an antiproton. Equation (44.10) then gives

$$(4mc^2)^2 = 2mc^2(E_m + mc^2)$$
$$E_m = 7mc^2$$

EVALUATE: The energy E_m of the bombarding proton includes its rest energy mc^2, so its minimum *kinetic* energy must be $6mc^2 = 6(938\text{ MeV}) = 5.63\text{ GeV}$.

The search for the antiproton was a principal reason for the construction of the Bevatron at the University of California, Berkeley, with beam energy of 6 GeV. The search succeeded in 1955, and Emilio Segrè and Owen Chamberlain were later awarded the Nobel Prize for this discovery.

Strangeness

The K mesons and the Λ and Σ hyperons were discovered during the late 1950s. Because of their unusual behavior they were called *strange particles*. They were produced in high-energy collisions such as $\pi^- + p$, and a K meson and a hyperon were always produced *together*. The relatively high rate of production of these particles suggested that it was a *strong*-interaction process, but their relatively long lifetimes suggested that their decay was a *weak*-interaction process. The K^0 appeared to have *two* lifetimes, one about 9×10^{-11} s and another nearly 600 times longer. Were the K mesons strongly interacting hadrons or not?

The search for the answer to this question led physicists to introduce a new quantity called **strangeness.** The hyperons Λ^0 and $\Sigma^{\pm,0}$ were assigned a strangeness quantum number $S = -1$, and the associated K^0 and K^+ mesons were assigned $S = +1$. The corresponding antiparticles had opposite strangeness, $S = +1$ for $\overline{\Lambda}^0$ and $\overline{\Sigma}^{\pm,0}$ and $S = -1$ for $\overline{K}^0$ and K^-. Then strangeness was *conserved* in production processes such as

$$p + \pi^- \rightarrow \Sigma^- + K^+$$
$$p + \pi^- \rightarrow \Lambda^0 + K^0$$

The process

$$p + \pi^- \rightarrow p + K^-$$

does not conserve strangeness and it does not occur.

When strange particles decay individually, strangeness is usually *not* conserved. Typical processes include

$$\Sigma^+ \rightarrow n + \pi^+$$
$$\Lambda^0 \rightarrow p + \pi^-$$
$$K^- \rightarrow \pi^+ + \pi^- + \pi^-$$

CAUTION **Strangeness vs. spin** Take care not to confuse the symbol S for strangeness with the identical symbol for the magnitude of the spin angular momentum.

In each of these decays, the initial strangeness is 1 or -1, and the final value is zero. All observations of these particles are consistent with the conclusion that *strangeness is conserved in strong interactions but it can change by zero or one unit in weak interactions.* There is no counterpart to the strangeness quantum number in classical physics.

Conservation Laws

The decay of strange particles provides our first example of a *conditional conservation law,* one that is obeyed in some interactions and not in others. By contrast, several conservation laws are obeyed in *all* interactions. These include the familiar conservation laws; energy, momentum, angular momentum, and electric charge. These are called *absolute conservation laws.* Baryon number and the three lepton numbers are also conserved in all interactions. Strangeness is conserved in strong and electromagnetic interactions but *not* in all weak interactions.

Two other quantities, which are conserved in some but not all interactions, are useful in classifying particles and their interactions. One is *isospin,* a quantity that is used to describe the charge independence of the strong interactions. The other is *parity,* which describes the comparative behavior of two systems that are mirror images of each other. Isospin is conserved in strong interactions, which are charge independent, but not in electromagnetic or weak interactions. (The electromagnetic interaction is certainly *not* charge independent.) Parity is conserved in strong and electromagnetic interactions but not in weak ones. The Chinese-American physicists T. D. Lee and C. N. Yang received the Nobel Prize in 1957 for laying the theoretical foundations for nonconservation of parity in weak interactions.

This discussion shows that conservation laws provide another basis for classifying particles and their interactions. Each conservation law is also associated with a *symmetry* property of the system. A familiar example is angular momentum. If a system is in an environment that has spherical symmetry, no torque can act on it because the direction of the torque would violate the symmetry. In such a system, total angular momentum is *conserved.* When a conservation law is violated, the interaction may be described as a *symmetry-breaking interaction.*

TEST YOUR UNDERSTANDING OF SECTION 44.3 From conservation of energy, a particle of mass m and rest energy mc^2 can decay only if the decay products have a total mass less than m. (The remaining energy goes into the kinetic energy of the decay products.) Can a proton decay into less massive mesons?

44.4 QUARKS AND GLUONS

The leptons form a fairly neat package: three particles and three neutrinos, each with its antiparticle, and a conservation law relating their numbers. Physicists believe that leptons are genuinely fundamental particles. The hadron family, by comparison, is a mess. Table 44.3 (in Section 44.3) contains only a sample of well over 100 hadrons that have been discovered since 1960, and it has become clear that these particles *do not* represent the most fundamental level of the structure of matter.

Our present understanding of the structure of hadrons is based on a proposal made initially in 1964 by the American physicist Murray Gell-Mann and his collaborators. In this proposal, hadrons are not fundamental particles but are composite structures whose constituents are spin-$\frac{1}{2}$ fermions called **quarks.** (The name is from the line "Three quarks for Muster Mark!" from *Finnegans Wake,* by James Joyce.) Each baryon is composed of three quarks (qqq), each antibaryon of three antiquarks ($\overline{q}\overline{q}\overline{q}$), and each meson of a quark–antiquark pair ($q\overline{q}$). Table 44.3 gives the quark content of many hadrons. No other compositions seem to be necessary. This scheme requires that quarks have electric charges with magnitudes $\frac{1}{3}$ and $\frac{2}{3}$ of the electron charge e, which had been thought to be the smallest unit of charge. Each quark also has a fractional value $\frac{1}{3}$ for its baryon number B, and each antiquark has a baryon-number value $-\frac{1}{3}$. In a meson, a quark and antiquark combine with net baryon number 0 and can have their spin angular momentum components parallel to form a spin-1 meson or antiparallel to form a spin-0 meson. Similarly, the three quarks in a baryon combine with net baryon number 1 and can form a spin-$\frac{1}{2}$ baryon or a spin-$\frac{3}{2}$ baryon.

The Three Original Quarks

The first (1964) quark theory included three types (called *flavors*) of quarks, labeled u (up), d (down), and s (strange). Their principal properties are listed in **Table 44.4**. The corresponding antiquarks $\overline{u}$, $\overline{d}$, and $\overline{s}$ have opposite values of charge Q, B, and S. Protons, neutrons, π and K mesons, and several hyperons can be constructed from these three quarks. For example, the proton quark content is uud. Checking Table 44.4, we see that the values of Q/e add to 1 and that the values of the baryon number B also add to 1, as we should expect. The neutron is udd, with total $Q = 0$ and $B = 1$. The π^+ meson is $u\overline{d}$, with $Q/e = 1$ and $B = 0$, and the K^+ meson is $u\overline{s}$. Checking the values of S for the quark content, we see that the proton, neutron, and π^+ have strangeness 0 and that the K^+ has strangeness 1, in agreement with Table 44.3. The antiproton is $\overline{p} = \overline{u}\overline{u}\overline{d}$, the negative pion is $\pi^- = \overline{u}d$, and so on. The quark content can also be used to explain hadron excited states and magnetic moments. **Figure 44.10** shows the quark content of two baryons and two mesons.

44.10 Quark content of four hadrons. The various color combinations that are needed for color neutrality are not shown.

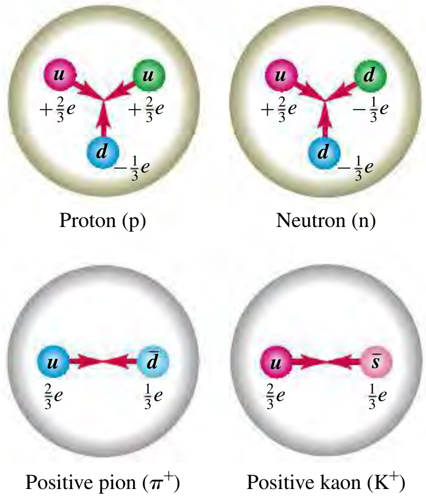


TABLE 44.4 Properties of the Three Original Quarks

Symbol	Q/e	Spin	Baryon Number, B	Strangeness, S	Charm, C	Bottomness, B'	Topness, T
u	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0	0	0
d	$-\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0	0	0
s	$-\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	-1	0	0	0

EXAMPLE 44.7 DETERMINING THE QUARK CONTENT OF BARYONS

SOLUTION

Given that they contain only u, d, s, $\overline{u}$, $\overline{d}$, and/or $\overline{s}$, find the quark content of (a) Σ^+ and (b) $\overline{\Lambda}{}^0$. The Σ^+ and Λ^0 (the antiparticle of the $\overline{\Lambda}{}^0$) are both baryons with strangeness $S = -1$.

SOLUTION

IDENTIFY and SET UP: We use the idea that the total charge of each baryon is the sum of the individual quark charges, and similarly for the baryon number and strangeness. We use the quark properties given in Table 44.4.

EXECUTE: Baryons contain three quarks. If $S = -1$, exactly *one* of the three must be an s quark, which has $S = -1$ and $Q/e = -\frac{1}{3}$.

(a) The Σ^+ has $Q/e = +1$, so the other two quarks must both be u quarks (each of which has $Q/e = +\frac{2}{3}$). Hence the quark content of Σ^+ is uus.

(b) First we find the quark content of the Λ^0. To yield zero total charge, the other two quarks must be u ($Q/e = +\frac{2}{3}$) and d ($Q/e = -\frac{1}{3}$), so the quark content of the Λ^0 is uds. The quark content of the $\overline{\Lambda}{}^0$ is therefore $\overline{u}\,\overline{d}\,\overline{s}$.

EVALUATE: Although the Λ^0 and $\overline{\Lambda}{}^0$ are both electrically neutral and have the same mass, they are different particles: Λ^0 has $B = 1$ and $S = -1$, while $\overline{\Lambda}{}^0$ has $B = -1$ and $S = 1$.

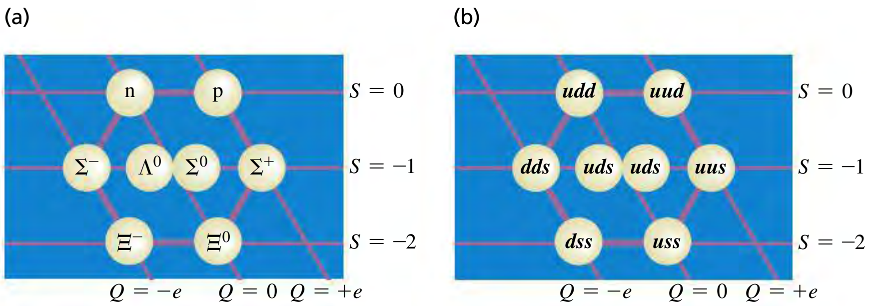
Motivating the Quark Model

What caused physicists to suspect that hadrons were made up of something smaller? The magnetic moment of the neutron (see Section 43.1) was one of the first reasons. In Section 27.7 we learned that a magnetic moment results from a circulating current (a motion of electric charge). But the neutron has *no* charge, or, to be more accurate, no *total* charge. It could be made up of smaller particles whose charges add to zero. The quantum motion of these particles within the neutron would then give its surprising nonzero magnetic moment. To verify this hypothesis by "seeing" inside a neutron, we need a probe with a wavelength that is much less than the neutron's size of about a femtometer. This probe should not be affected by the strong interaction, so that it won't interact with the neutron as a whole but will penetrate into it and interact electromagnetically with these supposed smaller charged particles. A probe with these properties is an electron with energy above 10 GeV. In experiments carried out at SLAC, such electrons were scattered from neutrons and protons to help show that nucleons are indeed made up of fractionally charged, spin-$\frac{1}{2}$ pointlike particles.

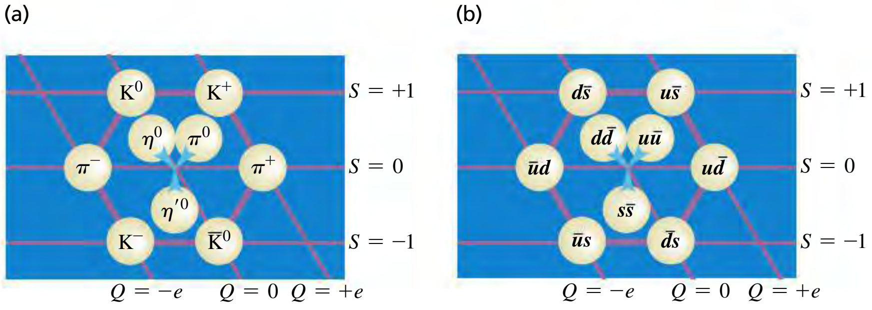
The Eightfold Way

Symmetry considerations play a very prominent role in particle theory. Here are two examples. Consider the eight spin-$\frac{1}{2}$ baryons we've mentioned: the familiar p and n; the strange Λ^0, Σ^+, Σ^0, and Σ^-; and the doubly strange Ξ^0 and Ξ^-. For each we plot the value of strangeness S versus the value of charge Q in **Fig. 44.11**. The result is a hexagonal pattern. A similar plot for the nine spin-0 mesons (six shown in Table 44.3 plus three others not included in that table) is shown in **Fig. 44.12**; the particles fall in exactly the same hexagonal pattern! In each plot, all the particles have masses that are within about $\pm 200\ \text{MeV}/c^2$ of the median mass value of that plot, with variations due to differences in quark masses and internal potential energies.

The symmetries that lead to these and similar patterns are collectively called the **eightfold way.** They were discovered in 1961 by Murray Gell-Mann and independently by Yu'val Ne'eman. (The name is a slightly irreverent reference to the Noble Eightfold Path, a set of principles for right living in Buddhism.) A similar pattern for the spin-$\frac{3}{2}$ baryons contains *ten* particles, arranged in a triangular



44.11 (a) Plot of S and Q values for spin-$\frac{1}{2}$ baryons, showing the symmetry pattern of the eightfold way. (b) Quark content of each spin-$\frac{1}{2}$ baryon. The quark contents of the Σ^0 and Λ^0 are the same; the Σ^0 is an excited state of the Λ^0 and can decay into it by photon emission.



44.12 (a) Plot of S and Q values for nine spin-0 mesons, showing the symmetry pattern of the eightfold way. Each particle is on the opposite side of the hexagon from its antiparticle; each of the three particles in the center is its own antiparticle. (b) Quark content of each spin-0 meson. The particles in the center are different mixtures of the three quark–antiquark pairs shown.

pattern like pins in a bowling alley. When this pattern was first discovered, one of the particles was missing. But Gell-Mann gave it a name anyway (Ω^-), predicted the properties it should have, and told experimenters what they should look for. Three years later, the particle was found during an experiment at Brookhaven National Laboratory, a spectacular success for Gell-Mann's theory. The whole series of events is reminiscent of the way in which Mendeleev used gaps in the periodic table of the elements to predict properties of undiscovered elements and to guide chemists in their search for these elements.

What binds quarks to one another? The attractive interactions among quarks are mediated by massless spin-1 bosons called **gluons** in much the same way that photons mediate the electromagnetic interaction or that pions mediated the nucleon–nucleon force in the old Yukawa theory.

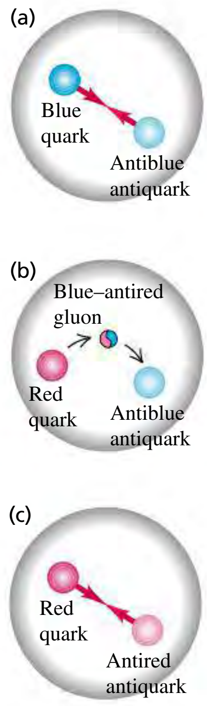
Color

Quarks, having spin $\frac{1}{2}$, are fermions and so are subject to the exclusion principle. This would seem to forbid a baryon having two or three quarks with the same flavor and same spin component. To avoid this difficulty, it is assumed that each quark comes in three varieties, which are whimsically called *colors*. Red, green, and blue are the usual choices. The exclusion principle applies separately to each color. A baryon always contains one red, one green, and one blue quark, so the baryon itself has no net color. Each gluon has a color–anticolor combination (for example, blue–antired) that allows it to transmit color when exchanged, and color is conserved during emission and absorption of a gluon by a quark. The gluon-exchange process changes the colors of the quarks in such a way that there is always one quark of each color in every baryon. The color of an individual quark changes continually as gluons are exchanged.

Similar processes occur in mesons such as pions. The quark–antiquark pairs of mesons have canceling color and anticolor (for example, blue and antiblue), so mesons also have no net color. Suppose a pion initially consists of a blue quark and an antiblue antiquark. The blue quark can become a red quark by emitting a blue–antired virtual gluon. The gluon is then absorbed by the antiblue antiquark, converting it to an antired antiquark (**Fig. 44.13**). Color is conserved in each emission and absorption, but a blue–antiblue pair has become a red–antired pair. Such changes occur continually, so we have to think of a pion as a superposition of three quantum states: blue–antiblue, green–antigreen, and red–antired. On a larger scale, the strong interaction between nucleons was described in Section 44.3 as due to the exchange of virtual mesons. In terms of quarks and gluons, these mediating virtual mesons are quark–antiquark systems bound together by the exchange of gluons.

The theory of strong interactions is known as *quantum chromodynamics* (QCD). No one has been able to isolate an individual quark, and indeed QCD predicts that quarks are bound in such a way that it is impossible to obtain a free quark. An impressive body of experimental evidence supports the correctness of the quark model and the idea that quantum chromodynamics is the key to understanding the strong interactions.

44.13 (a) A pion containing a blue quark and an antiblue antiquark. (b) The blue quark emits a blue–antired gluon, changing to a red quark. (c) The gluon is absorbed by the antiblue antiquark, which becomes an antired antiquark. The pion now consists of a red–antired quark–antiquark pair. The actual quantum state of the pion is an equal superposition of red–antired, green–antigreen, and blue–antiblue pairs.



Three More Quarks

Before the tau particles were discovered, there were four known leptons. This fact, together with some puzzling decay rates, led to the speculation that there might be a fourth quark flavor. This quark is labeled ***c*** (the *charmed* quark); it has $Q/e = \frac{2}{3}$, $B = \frac{1}{3}$, $S = 0$, and a new quantum number **charm** $C = +1$. This was confirmed in 1974 by the observation at both SLAC and the Brookhaven National Laboratory of a meson, now named ψ, with mass 3097 MeV/c^2. This meson was found to have several decay modes, decaying into e^+e^-, $\mu^+\mu^-$, or hadrons. The mean lifetime was found to be about 10^{-20} s. These results are consistent with ψ being a spin-1 $c\bar{c}$ system. Almost immediately after this,

TABLE 44.5 Properties of the Six Quarks

Symbol	Q/e	Spin	Baryon Number, B	Strangeness, S	Charm, C	Bottomness, B'	Topness, T
u	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0	0	0
d	$-\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0	0	0
s	$-\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	-1	0	0	0
c	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	$+1$	0	0
b	$-\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0	-1	0
t	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0	0	$+1$

similar mesons of greater mass were observed and identified as excited states of the $c\bar{c}$ system. A few years later, individual mesons with a nonzero net charm quantum number, D^0 $(c\bar{u})$ and D^+ $(c\bar{d})$, and a charmed baryon, Λ_c^+ (udc), were also observed.

In 1977 a meson with mass 9460 MeV$/c^2$, called upsilon (Υ), was discovered at Brookhaven. Because it had properties similar to ψ, it was conjectured that the meson was really the bound system of a new quark, b (the *bottom* quark), and its antiquark, $\bar{b}$. The bottom quark has the value -1 of a new quantum number B' called *bottomness*. Excited states of the Υ were soon observed, as were the B^+ $(\bar{b}u)$ and B^0 $(\bar{b}d)$ mesons.

CAUTION **Bottomness vs. baryon number** Don't confuse the bottomness quantum number B' with baryon number B. For example, the proton (which has zero bottomness and is a baryon) has $B' = 0$ and $B = +1$; the B^+ meson (which includes an antibottom quark but is not a baryon) has $B' = +1$ and $B = 0$.

With the five flavors of quarks ($u, d, s, c,$ and b) and the six flavors of leptons (e, μ, τ, ν_e, ν_μ, and ν_τ) it was an appealing conjecture that nature is symmetric in its building blocks and that therefore there should be a *sixth* quark. This quark, labeled t (top), would have $Q/e = \frac{2}{3}$, $B = \frac{1}{3}$, and a new quantum number, $T = 1$. In 1995, groups using two different detectors at Fermilab's Tevatron announced the discovery of the top quark. The groups collided 0.9-TeV protons with 0.9-TeV antiprotons, but even with 1.8 TeV of available energy, a top–antitop $(t\bar{t})$ pair was detected in fewer than two of every 10^{11} collisions! **Table 44.5** lists some properties of the six quarks. Each has a corresponding antiquark with opposite values of Q, B, S, C, B', and T.

TEST YOUR UNDERSTANDING OF SECTION 44.4 Is it possible to have a baryon with charge $Q = +e$ and strangeness $S = -2$?

44.5 THE STANDARD MODEL AND BEYOND

The particles and interactions that we've discussed in this chapter provide a reasonably comprehensive picture of the fundamental building blocks of nature. There is enough confidence in the basic correctness of this picture that it is called the **standard model.**

The standard model includes three families of particles: (1) the six leptons, which have no strong interactions; (2) the six quarks, from which all hadrons are made; and (3) the particles that mediate the various interactions. These mediators are gluons for the strong interaction among quarks, photons for the electromagnetic interaction, the $W^{\pm}$ and Z^0 particles for the weak interaction, and the graviton for the gravitational interaction.

Electroweak Unification

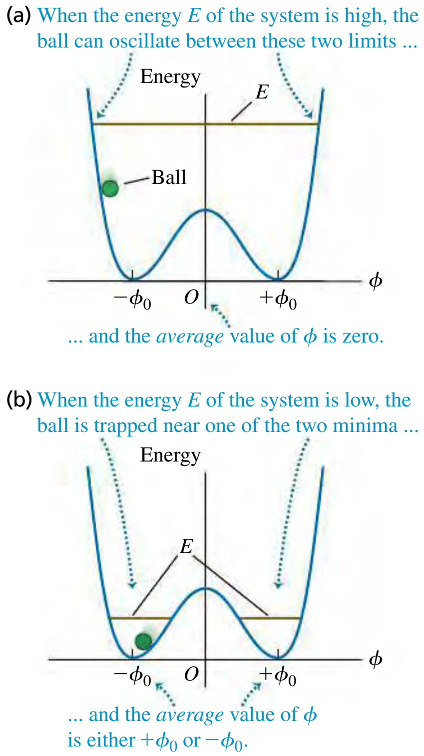
Theoretical physicists have long dreamed of combining all the interactions of nature into a single unified theory. As a first step, Einstein spent much of his later life trying to develop a field theory that would unify gravitation and electromagnetism. He was only partly successful.

Between 1961 and 1967, Sheldon Glashow, Abdus Salam, and Steven Weinberg developed a theory that unifies the weak and electromagnetic forces. One outcome of their **electroweak theory** is a prediction of the weak-force mediator particles, the $W^{\pm}$ and Z^0 bosons, including their masses. The basic idea is that the mass difference between photons (zero mass) and the weak bosons ($\approx 100\ \text{GeV}/c^2$) makes the electromagnetic and weak interactions behave quite differently at low energies. At sufficiently high energies (well above 100 GeV), however, the distinction disappears, and the two merge into a single interaction. This prediction was verified in 1983 in experiments with proton-antiproton collisions at CERN. The weak bosons were found, again with the help provided by the theoretical description, and their observed masses agreed with the predictions of the electroweak theory, a wonderful convergence of theory and experiment. The electroweak theory and quantum chromodynamics form the backbone of the standard model. Glashow, Salam, and Weinberg received the Nobel Prize in 1979.

In the electroweak theory photons are massless but the weak bosons are very massive. To account for the broken symmetry among these interaction mediators, a field called the *Higgs field* was proposed by theoretical physicists in the 1960s. We use the symbol $\phi(\vec{r}, t)$ to denote the value of this field at position $\vec{r}$ and time t. (Unlike the electric and magnetic fields, which are vectors, the Higgs field is a scalar quantity.) According to the theory, the mass of the weak bosons is proportional to the absolute value of ϕ_{av}, where ϕ_{av} is the average value of $\phi(\vec{r}, t)$ over space. **Figure 44.14** shows a simplified model of how ϕ_{av} depends on energy. At very high energies, the value of the Higgs field $\phi(\vec{r}, t)$ oscillates between positive and negative values, so its average value is $\phi_{\text{av}} = 0$ (Fig. 44.14a). But at low energies, $\phi(\vec{r}, t)$ oscillates around either a positive average value $\phi_{\text{av}} = +\phi_0$ or a negative average value $\phi_{\text{av}} = -\phi_0$ (Fig. 44.14b). The oscillation is no longer symmetric around $\phi = 0$, so the symmetry has been broken. Hence at low energies, the weak bosons acquire a nonzero mass proportional to $|\phi_{\text{av}}|$, which is equal to ϕ_0 for either of the cases shown in Fig. 44.14b.

This theory also predicts that there should be a particle called the *Higgs boson* associated with the Higgs field itself. (In an analogous way, the photon is the particle associated with the electromagnetic field.) The Higgs boson was predicted to be unstable, have zero charge and spin 0, and have a large mass. An important mission of the Large Hadron Collider at CERN was to produce the Higgs boson from the available energy in proton-proton collisions and thereby verify the existence of the Higgs field. In 2012 the first Higgs bosons were detected in such collisions, with the predicted properties. This suggests that the concept of the Higgs field—that all of space is filled with a field that gives mass to the weak bosons—is indeed correct. The Nobel Prize was awarded in 2013 to François Englert and Peter Higgs, two of the theorists who first proposed the idea of the Higgs field in 1964. Current experiments show that the mass of the Higgs boson is about 125 GeV/c^2, even greater than the masses of the $W^{\pm}$ and Z^0 weak bosons.

44.14 The Higgs-field value ϕ can oscillate, much like the value of the coordinate of a ball rolling within a trough with two minima. The average value of ϕ determines the masses of the $W^{\pm}$ and Z^0 weak bosons. **(a)** At high energies, the average value of ϕ is zero and the $W^{\pm}$ and Z^0 are massless (like the photon). **(b)** At low energies, the symmetry is broken. The average value of ϕ is nonzero, and the $W^{\pm}$ and Z^0 acquire nonzero masses.

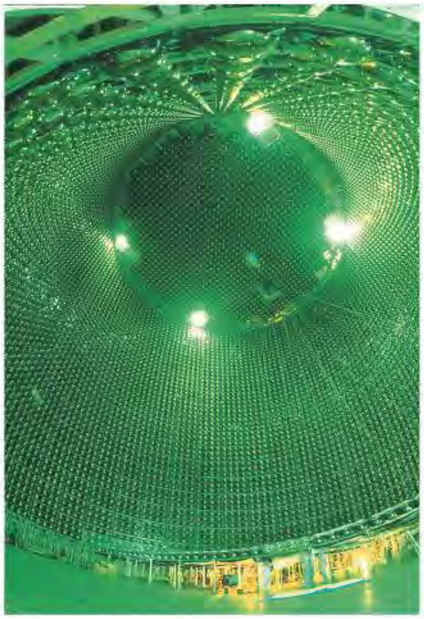


Grand Unified Theories

Perhaps at sufficiently high energies the strong interaction and the electroweak interaction have a convergence similar to that between the electromagnetic and weak interactions. If so, they can be unified to give a comprehensive theory of strong, weak, and electromagnetic interactions. Such schemes, called **grand unified theories** (GUTs), are still speculative.

Some grand unified theories predict the decay of the proton (in violation of conservation of baryon number), with an estimated lifetime of more than 10^{28} years. (For comparison the age of the universe is known to be 1.38×10^{10} years.) With a lifetime of 10^{28} years, six metric tons of protons would be expected to have only one decay per day, so huge amounts of material must be examined. Some of the neutrino detectors that we mentioned in Section 44.2 originally looked for, and failed to find, evidence of proton decay. Current estimates set the proton lifetime well over 10^{33} years. Some GUTs also predict the existence of magnetic monopoles, which we mentioned in Chapter 27. At present there is no confirmed experimental evidence that magnetic monopoles exist.

44.15 This photo shows the interior of the Super-Kamiokande neutrino detector in Japan. When in operation, the detector is filled with 5×10^7 kg of water. A neutrino passing through the detector can produce a faint flash of light, which is detected by the 13,000 photomultiplier tubes lining the detector walls. Data from this detector were the first to indicate that neutrinos have mass.



In the standard model, the neutrinos have zero mass. Nonzero values are controversial because experiments to determine neutrino masses are difficult both to perform and to analyze. In most GUTs the neutrinos *must* have nonzero masses. If neutrinos do have mass, transitions called *neutrino oscillations* can occur, in which one type of neutrino (ν_e, ν_μ, or ν_τ) changes into another type. In 1998, scientists using the Super-Kamiokande neutrino detector in Japan (**Fig. 44.15**) reported the discovery of oscillations between muon neutrinos and tau neutrinos. Subsequent measurements at the Sudbury Neutrino Observatory in Canada have confirmed the existence of neutrino oscillations. This discovery is evidence for exciting physics beyond that predicted by the standard model.

The discovery of neutrino oscillations cleared up a long-standing mystery. Since the 1960s, physicists have been using sensitive detectors to look for electron neutrinos produced by nuclear fusion reactions in the sun's core (see Section 43.8). However, the observed flux of solar electron neutrinos is only one-third of the predicted value. The explanation was provided in 2002 by the Sudbury Neutrino Observatory, which can detect neutrinos of all three flavors. The results showed that the combined flux of solar neutrinos of *all* flavors is equal to the theoretical prediction for the flux of *electron* neutrinos. The explanation is that the sun is producing electron neutrinos at the predicted rate, but that two-thirds of these electron neutrinos are transformed into muon or tau neutrinos during their flight from the sun's core to a detector on earth.

Supersymmetric Theories and TOEs

The ultimate dream of theorists is to unify all four fundamental interactions, adding gravitation to the strong and electroweak interactions that are included in GUTs. Such a unified theory is whimsically called a Theory of Everything (TOE). It turns out that an essential ingredient of such theories is a space-time continuum with more than four dimensions. The additional dimensions are "rolled up" into extremely tiny structures that we ordinarily do not notice. Depending on the scale of these structures, it may be possible for the next generation of particle accelerators to reveal the presence of extra dimensions.

Another ingredient of many theories is *supersymmetry,* which gives every boson and fermion a "superpartner" of the other spin type. For example, the proposed supersymmetric partner of the spin-$\frac{1}{2}$ electron is a spin-0 particle called the *selectron,* and that of the spin-1 photon is a spin-$\frac{1}{2}$ *photino.* As yet, no superpartner particles have been discovered, perhaps because they are too massive to be produced by the present generation of accelerators. Within a few years, new data from the Large Hadron Collider will help us decide whether these intriguing theories have merit.

TEST YOUR UNDERSTANDING OF SECTION 44.5 One aspect of the standard model is that a $\boldsymbol{d}$ quark can transform into a $\boldsymbol{u}$ quark, an electron, and an antineutrino by means of the weak interaction. If this happens to a $\boldsymbol{d}$ quark inside a neutron, what kind of particle remains afterward in addition to the electron and antineutrino? (i) A proton; (ii) a Σ^-; (iii) a Σ^+; (iv) a Λ^0 or a Σ^0; (v) any of these. ❙

44.6 THE EXPANDING UNIVERSE

In the last two sections of this chapter we'll explore briefly the connections between the early history of the universe and the interactions of fundamental particles. It is remarkable that there are such close ties between physics on the smallest scale that we've explored experimentally (the range of the weak interaction, of the order of 10^{-18} m) and physics on the largest scale (the universe itself, of the order of at least 10^{26} m).

Gravitational interactions play an essential role in the large-scale behavior of the universe. We saw in Chapter 13 how the law of gravitation explains the motions of planets in the solar system. Astronomical evidence shows that gravitational forces also dominate in larger systems such as galaxies and clusters of galaxies (**Fig. 44.16**).

Until early in the 20th century it was usually assumed that the universe was *static;* stars might move relative to each other, but there was not thought to be any overall expansion or contraction. But measurements that were begun in 1912 by Vesto Slipher at Lowell Observatory in Arizona, and continued in the 1920s by Edwin Hubble with the help of Milton Humason at Mount Wilson in California, indicated that the universe is *not* static. The motions of galaxies relative to the earth can be measured by observing the shifts in the wavelengths of their spectra. For distant galaxies these shifts are always toward longer wavelength, so they appear to be receding from us and from each other. Astronomers first assumed that these were Doppler shifts and used a relationship between the wavelength λ_0 of light measured now on earth from a source receding at speed v and the wavelength λ_S measured in the rest frame of the source when it was emitted. We can derive this relationship by inverting Eq. (37.25) for the Doppler effect, making subscript changes, and using $\lambda = c/f$; the result is

$$\lambda_0 = \lambda_S \sqrt{\frac{c+v}{c-v}} \tag{44.13}$$

Wavelengths from receding sources are always shifted toward longer wavelengths; this increase in λ is called the **redshift.** We can solve Eq. (44.13) for v:

$$v = \frac{(\lambda_0/\lambda_S)^2 - 1}{(\lambda_0/\lambda_S)^2 + 1}c \tag{44.14}$$

CAUTION **Redshift, not Doppler shift** Equations (44.13) and (44.14) are from the *special* theory of relativity and refer to the Doppler effect. As we'll see, the redshift from *distant* galaxies is caused by an effect that is explained by the *general* theory of relativity and is *not* a Doppler shift. However, as the ratio v/c and the fractional wavelength change $(\lambda_0 - \lambda_S)/\lambda_S$ become small, the general theory's equations approach Eqs. (44.13) and (44.14), and those equations may be used.

44.16 **(a)** The galaxy M101 is a larger version of the Milky Way galaxy of which our solar system is a part. Like all galaxies, M101 is held together by the mutual gravitational attraction of its stars, gas, dust, and other matter, all of which orbit around the galaxy's center of mass. M101 is 25 million light-years away. **(b)** This image shows part of the Coma cluster, an immense grouping of over 1000 galaxies that lies 300 million light-years from us. The galaxies within the cluster are all in motion. Gravitational forces between the galaxies prevent them from escaping from the cluster.

(a)

(b)

EXAMPLE 44.8 RECESSION SPEED OF A GALAXY

SOLUTION

The spectral lines of various elements are detected in light from a galaxy in the constellation Ursa Major. An ultraviolet line from singly ionized calcium ($\lambda_S = 393$ nm) is observed at wavelength $\lambda_0 = 414$ nm, redshifted into the visible portion of the spectrum. At what speed is this galaxy receding from us?

SOLUTION

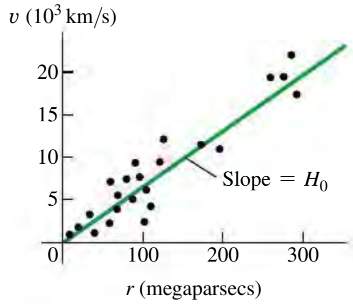
IDENTIFY and SET UP: This example uses the relationship between redshift and recession speed for a distant galaxy. We can use the wavelengths λ_S at which the light is emitted and λ_0 that we detect on earth in Eq. (44.14) to determine the galaxy's recession speed v if the fractional wavelength shift is not too great.

EXECUTE: The fractional wavelength redshift for this galaxy is $\lambda_0/\lambda_S = (414\text{ nm})/(393\text{ nm}) = 1.053$. This is only a 5.3% increase, so we can use Eq. (44.14) with reasonable accuracy:

$$v = \frac{(1.053)^2 - 1}{(1.053)^2 + 1}c = 0.0516c = 1.55 \times 10^7\text{ m/s}$$

EVALUATE: The galaxy is receding from the earth at 5.16% of the speed of light. Rather than going through this calculation, astronomers often just state the *redshift* $z = (\lambda_0 - \lambda_S)/\lambda_S = (\lambda_0/\lambda_S) - 1$. This galaxy has redshift $z = 0.053$.

44.17 Graph of recession speed versus distance for several galaxies. The best-fit straight line illustrates Hubble's law. The slope of the line is the Hubble constant, H_0.



The Hubble Law

Analysis of redshifts from many distant galaxies led Edwin Hubble to a remarkable conclusion: The speed of recession v of a galaxy is proportional to its distance r from us (**Fig. 44.17**). This relationship is now called the **Hubble law;** expressed as an equation,

$$v = H_0 r \tag{44.15}$$

where H_0 is an experimental quantity commonly called the *Hubble constant,* since at any given time it is constant over space. Determining H_0 has been a key goal of the Hubble Space Telescope, which can measure distances to galaxies with unprecedented accuracy. The current best value is $2.18 \times 10^{-18}\ \text{s}^{-1}$, with an uncertainty of 2%.

Astronomical distances are often measured in *parsecs* (pc); one parsec is the distance at which there is a one-arcsecond $(1/3600°)$ angular separation between two objects 1.50×10^{11} m apart (the average distance from the earth to the sun). A distance of 1 pc is equal to 3.26 *light-years* (ly), where 1 ly $= 9.46 \times 10^{12}$ km is the distance that light travels in one year. The Hubble constant is then commonly expressed in the mixed units (km/s)/Mpc (kilometers per second per megaparsec), where 1 Mpc $= 10^6$ pc:

$$H_0 = (2.18 \times 10^{-18}\ \text{s}^{-1})\left(\frac{9.46 \times 10^{12}\ \text{km}}{1\ \text{ly}}\right)\left(\frac{3.26\ \text{ly}}{1\ \text{pc}}\right)\left(\frac{10^6\ \text{pc}}{1\ \text{Mpc}}\right) = 67.3\,\frac{\text{km/s}}{\text{Mpc}}$$

EXAMPLE 44.9 DETERMINING DISTANCE WITH THE HUBBLE LAW

SOLUTION

Use the Hubble law to find the distance from earth to the galaxy in Ursa Major described in Example 44.8.

SOLUTION

IDENTIFY and SET UP: The Hubble law relates the redshift of a distant galaxy to its distance r from earth. We solve Eq. (44.15) for r and substitute the recession speed v from Example 44.8.

EXECUTE: Using $H_0 = 67.3\ (\text{km/s})/\text{Mpc} = 6.73 \times 10^4\ (\text{m/s})/\text{Mpc}$,

$$r = \frac{v}{H_0} = \frac{1.55 \times 10^7\ \text{m/s}}{6.73 \times 10^4\ (\text{m/s})/\text{Mpc}} = 230\ \text{Mpc}$$

$$= 2.3 \times 10^8\ \text{pc} = 7.5 \times 10^8\ \text{ly} = 7.1 \times 10^{24}\ \text{m}$$

EVALUATE: A distance of 230 million parsecs (750 million light-years) is truly stupendous, but many galaxies lie much farther away. To appreciate the immensity of this distance, consider that our farthest-ranging unmanned spacecraft have traveled only about 0.002 ly from our planet.

Another aspect of Hubble's observations was that, *in all directions,* distant galaxies appeared to be receding from us. There is no particular reason to think that our galaxy is at the very center of the universe; if we lived in some other galaxy, every distant galaxy would still seem to be moving away. That is, at any given time, *the universe looks more or less the same, no matter where in the universe we are.* This important idea is called the **cosmological principle.** There are local fluctuations in density, but on average, the universe looks the same from all locations. Thus the Hubble constant is constant in space although not necessarily constant in time, and the laws of physics are the same everywhere.

The Big Bang

The Hubble law suggests that at some time in the past, all the matter in the universe was far more concentrated than it is today. It was then blown apart in a rapid expansion called the **Big Bang,** giving all observable matter more or less the velocities that we observe today. When did this happen? According to the

Hubble law, matter at a distance r away from us is traveling with speed $v = H_0 r$. The time t needed to travel a distance r is

$$t = \frac{r}{v} = \frac{r}{H_0 r} = \frac{1}{H_0} = 4.59 \times 10^{17}\ \text{s} = 1.45 \times 10^{10}\ \text{y}$$

By this hypothesis the Big Bang occurred about 14 billion years ago. It assumes that all speeds are *constant* after the Big Bang; that is, it ignores any change in the expansion rate due to gravitational attraction or other effects. We'll return to this point later. For now, however, notice that the age of the earth determined from radioactive dating (see Section 43.4) is 4.54 billion (4.54×10^9) years. It's encouraging that our hypothesis tells us that the universe is older than the earth!

Expanding Space

The general theory of relativity takes a radically different view of the expansion just described. According to this theory, the increased wavelength is *not* caused by a Doppler shift as the universe expands into a previously empty void. Rather, the increase comes from *the expansion of space itself* and everything in intergalactic space, including the wavelengths of light traveling to us from distant sources. This is not an easy concept to grasp, and if you haven't encountered it before, it may sound like doubletalk.

An analogy may help you develop some intuition on this point. Imagine we are all bugs crawling around on a horizontal surface. We can't leave the surface, and we can see in any direction along the surface, but not up or down. We are then living in a two-dimensional world; some writers have called such a world *flatland*. If the surface is a plane, we can locate our position with two Cartesian coordinates (x, y). If the plane extends indefinitely in both the x- and y-directions, we described our space as having *infinite* extent, or as being *unbounded*. No matter how far we go, we never reach an edge or a boundary.

An alternative habitat for us bugs would be the surface of a sphere of radius R. The space would still seem infinite—we could crawl forever and never reach an edge or a boundary. Yet in this case the space is *finite* or *bounded*. To describe the location of a point in this space, we could still use two coordinates: latitude and longitude, or the spherical coordinates θ and ϕ shown in Fig. 41.5.

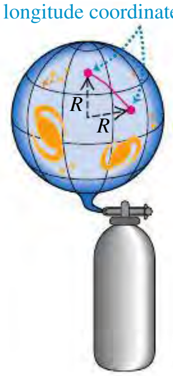
Now suppose the spherical surface is that of a balloon (**Fig. 44.18**). As we inflate the balloon more and more, increasing the radius R, the coordinates of a point don't change, yet the distance between any two points gets larger and larger. Furthermore, as R increases, the *rate of change* of distance between two points (their recession speed) is proportional to their distance apart. *The recession speed is proportional to the distance,* just as with the Hubble law. For example, the distance from Pittsburgh to Miami is twice as great as the distance from Pittsburgh to Boston. If the earth were to begin to swell, Miami would recede from Pittsburgh twice as fast as Boston would.

Although the quantity R isn't one of the two coordinates giving the position of a point on the balloon's surface, it nevertheless plays an essential role in any discussion of distance. It is the radius of curvature of our two-dimensional universe, and it is also a varying *scale factor* that changes as this universe expands.

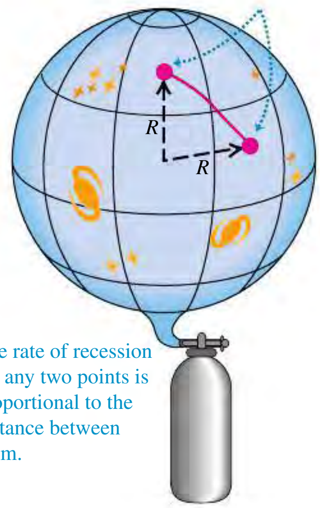
Generalizing this picture to three dimensions isn't so easy. We have to think of our three-dimensional space as being embedded in a space with four or more dimensions, just as we visualized the two-dimensional spherical flatland as being embedded in a three-dimensional Cartesian space. Our real three-space is *not Cartesian;* to describe its characteristics in any small region requires at least one additional parameter, the curvature of space, which is analogous to the radius of the sphere. In a sense, this scale factor, which we'll continue to call R, describes the *size* of the universe, just as the radius of the sphere described the size of our two-dimensional spherical universe. We'll return later to the question of whether the universe is bounded or unbounded.

44.18 An inflating balloon as an analogy for an expanding universe.

(a) Points (representing galaxies) on the surface of a balloon are described by their latitude and longitude coordinates.



(b) The radius R of the balloon has increased. The coordinates of the points are the same, but the distance between them has increased.



Any length that is measured in intergalactic space is proportional to R, so the wavelength of light traveling to us from a distant galaxy increases along with every other dimension as the universe expands. That is,

$$\frac{\lambda_0}{\lambda} = \frac{R_0}{R} \tag{44.16}$$

The zero subscripts refer to the values of the wavelength and scale factor *now,* just as H_0 is the current value of the Hubble constant. The quantities λ and R without subscripts are the values at *any* time—past, present, or future. For the galaxy described in Examples 44.8 and 44.9, we have $\lambda_0 = 414$ nm and $\lambda = \lambda_S = 393$ nm, so Eq. (44.16) gives $R_0/R = 1.053$. That is, the scale factor *now* (R_0) is 5.3% larger than it was 750 million years ago when the light was emitted from that galaxy in Ursa Major. This increase of wavelength with time as the scale factor increases in our expanding universe is called the *cosmological redshift.* The farther away an object is, the longer its light takes to get to us and the greater the change in R and λ. The current largest measured wavelength ratio for galaxies is about 8.6, meaning that the volume of space itself is about $(8.6)^3 \approx 640$ times larger than it was when the light was emitted. Do *not* attempt to substitute $\lambda_0/\lambda_S = 8.6$ into Eq. (44.14) to find the recession speed; that equation is accurate only for small cosmological redshifts and $v \ll c$. The actual value of v depends on the density of the universe, the value of H_0, and the expansion history of the universe.

Here's a surprise: If the distance from us in the Hubble law is large enough, then the speed of recession is greater than the speed of light! This does *not* violate the special theory of relativity because the recession speed is *not* caused by the motion of the astronomical object relative to some coordinates in its region of space. Rather, $v > c$ when two sets of coordinates move apart fast enough as space itself expands. In other words, there are objects whose coordinates have been moving away from our coordinates so fast that light from them hasn't had enough time in the entire history of the universe to reach us. What we see is just the *observable* universe; we have no direct evidence about what lies beyond its horizon.

CAUTION **The universe isn't expanding into emptiness** The balloon shown in Fig. 44.18 is expanding into the empty space around it. It's a common misconception to picture the universe in the same way as a large but finite collection of galaxies that's expanding into unoccupied space. The reality is quite different! All evidence shows that our universe is *infinite:* It has no edges, so there is nothing "outside" it and it isn't "expanding into" anything. The expansion of the universe simply means that the scale factor of the universe is increasing. A good two-dimensional analogy is to think of the universe as a flat, infinitely large rubber sheet that's stretching and expanding much like the surface of the balloon in Fig. 44.18. In a sense, the infinite universe is becoming more infinite!

Critical Density

In an expanding universe, gravitational attractions between galaxies should slow the initial expansion. But by how much? If these attractions are strong enough, the universe should expand more and more slowly, eventually stop, and then begin to contract, perhaps all the way down to what's been called a *Big Crunch.* On the other hand, if gravitational forces are much weaker, they slow the expansion only a little, and the universe should continue to expand forever.

The situation is analogous to the problem of escape speed of a projectile launched from the earth. We studied this problem in Example 13.5 (Section 13.3). The total energy $E = K + U$ when a projectile of mass m and speed v is at a distance r from the center of the earth (mass m_E) is

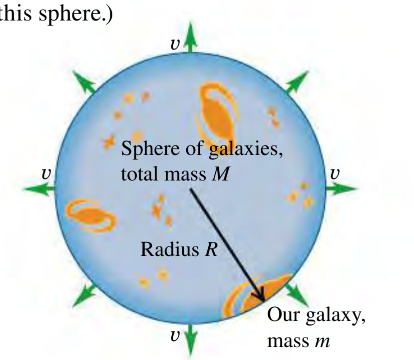
$$E = \tfrac{1}{2}mv^2 - \frac{Gmm_E}{r}$$

If E is positive, the projectile has enough kinetic energy to move infinitely far from the earth $(r \rightarrow \infty)$ and have some kinetic energy left over. If E is negative, the kinetic energy $K = \frac{1}{2}mv^2$ becomes zero and the projectile stops when $r = -Gmm_E/E$. In that case, no greater value of r is possible, and the projectile can't escape the earth's gravity.

We can carry out a similar analysis for the universe. Whether the universe continues to expand indefinitely should depend on the average *density* of matter. If matter is relatively dense, there is a lot of gravitational attraction to slow and eventually stop the expansion and make the universe contract again. If not, the expansion should continue indefinitely. We can derive an expression for the *critical density* ρ_c needed to just barely stop the expansion.

Here's a calculation based on Newtonian mechanics; it isn't relativistically correct, but it illustrates the idea. Consider a large sphere with radius R, containing many galaxies (**Fig. 44.19**), with total mass M. Suppose our own galaxy has mass m and is located at the surface of this sphere. According to the cosmological principle, the average distribution of matter within the sphere is uniform. The total gravitational force on our galaxy is just the force due to the mass M inside the sphere. The force on our galaxy and potential energy U due to this spherically symmetric distribution are the same as though m and M were both points, so $U = -GmM/R$, just as in Section 13.3. The net force from all the uniform distribution of mass *outside* the sphere is zero, so we'll ignore it.

44.19 An imaginary sphere of galaxies. The net gravitational force exerted on our galaxy (at the surface of the sphere) by the other galaxies is the same as if all of their mass were concentrated at the center of the sphere. (Since the universe is infinite, there's also an infinity of galaxies outside this sphere.)



The total energy E (kinetic plus potential) for our galaxy is

$$E = \frac{1}{2}mv^2 - \frac{GmM}{R} \tag{44.17}$$

If E is *positive*, our galaxy has enough energy to escape from the gravitational attraction of the mass M inside the sphere; in this case the universe should keep expanding forever. If E is negative, our galaxy cannot escape and the universe should eventually pull back together. The crossover between these two cases occurs when $E = 0$, so

$$\frac{1}{2}mv^2 = \frac{GmM}{R} \tag{44.18}$$

The total mass M inside the sphere is the volume $4\pi R^3/3$ times the density ρ_c:

$$M = \frac{4}{3}\pi R^3 \rho_c$$

We'll assume that the speed v of our galaxy relative to the center of the sphere is given by the Hubble law: $v = H_0 R$. Substituting these expressions for m and v into Eq. (44.18), we get

$$\frac{1}{2}m(H_0 R)^2 = \frac{Gm}{R}\left(\frac{4}{3}\pi R^3 \rho_c\right) \qquad \text{or}$$

$$\rho_c = \frac{3H_0^2}{8\pi G} \qquad \text{(critical density of the universe)} \tag{44.19}$$

This is the *critical density*. If the average density is less than ρ_c, the universe should continue to expand indefinitely; if it is greater, the universe should eventually stop expanding and begin to contract.

Putting numbers into Eq. (44.19), we find

$$\rho_c = \frac{3(2.18 \times 10^{-18}\ \text{s}^{-1})^2}{8\pi(6.67 \times 10^{-11}\ \text{N} \cdot \text{m}^2/\text{kg}^2)} = 8.50 \times 10^{-27}\ \text{kg/m}^3$$

The mass of a hydrogen atom is 1.67×10^{-27} kg, so this density is equivalent to about five hydrogen atoms per cubic meter.

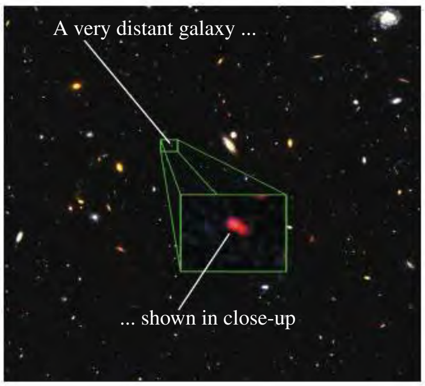
Dark Matter, Dark Energy, and the Accelerating Universe

Astronomers have made extensive studies of the average density of matter in the universe. One way to do so is to count the number of galaxies in a patch of sky. Based on the mass of an average star and the number of stars in an average galaxy, this effort gives an estimate of the average density of *luminous* matter in the universe—that is, matter that emits electromagnetic radiation. (You are made of luminous matter because you emit infrared radiation as a consequence of your temperature; see Sections 17.7 and 39.5.) It's also necessary to take into account other luminous matter within a galaxy, including the tenuous gas and dust between the stars.

Another technique is to study the motions of galaxies within clusters of galaxies (see Fig. 44.16b). The motions are so slow that we can't actually see galaxies changing positions within a cluster. However, observations show that different galaxies within a cluster have somewhat different redshifts, which indicates that the galaxies are moving relative to the center of mass of the cluster. The speeds of these motions are related to the gravitational force exerted on each galaxy by the other members of the cluster, which in turn depends on the total mass of the cluster. By measuring these speeds, astronomers can determine the average density of *all* kinds of matter within the cluster, whether or not the matter emits electromagnetic radiation.

Observations using these and other techniques show that the average density of *all* matter in the universe is 31.5% of the critical density, but the average density of *luminous* matter is only 4.9% of the critical density. In other words, most of the matter in the universe is not luminous: It does not emit electromagnetic radiation of *any* kind. At present, the nature of this **dark matter** remains an outstanding mystery. Some proposed candidates for dark matter are WIMPs (weakly interacting massive particles, which are hypothetical subatomic particles far more massive than those produced in accelerator experiments) and MACHOs (massive compact halo objects, which include objects such as black holes that might form "halos" around galaxies). Whatever the true nature of dark matter, it is by far the dominant form of matter in the universe. For every kilogram of the conventional matter that has been our subject for most of this book—including electrons, protons, atoms, molecules, blocks on inclined planes, planets, and stars—there are about *five and a half* kilograms of dark matter.

44.20 The bright spots in this image are not stars but entire galaxies. We see the most distant of these, magnified in the inset, as it was 13.1 billion years ago, when the universe was just 700 million years old. At that time the scale factor of the universe was only about 12% as large as it is now. (The red color of this galaxy is due to its very large redshift.) By comparison, we see the relatively nearby Coma cluster (see Fig. 44.16b) as it was 300 million years ago, when the scale factor was 98% of the present-day value.



Since the average density of matter in the universe is less than the critical density, it might seem fair to conclude that the universe will continue to expand indefinitely, and that gravitational attraction between matter in different parts of the universe should slow the expansion down (albeit not enough to stop it). One way to test this prediction is to examine the redshifts of extremely distant objects. The more distant a galaxy is, the more time it takes that light to reach us from that galaxy, so the further back in time we look when we observe that galaxy. If the expansion of the universe has been slowing down, the expansion must have been more rapid in the distant past. Thus we would expect very distant galaxies to have *greater* redshifts than predicted by the Hubble law, Eq. (44.15).

Only since the 1990s has it become possible to measure accurately both the distances and the redshifts of extremely distant galaxies. The results have been totally surprising: Very distant galaxies, seen as they were when the universe was a small fraction of its present age (**Fig. 44.20**), have *smaller* redshifts than predicted by the Hubble law! The implication is that the expansion of the universe was slower in the past than it is now, so the expansion has been *speeding up* rather than slowing down.

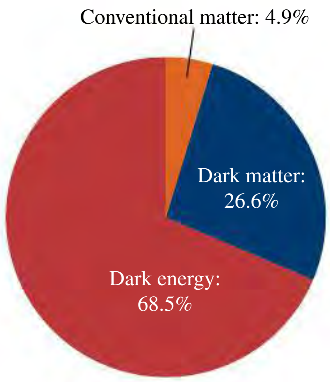
If gravitational attraction should make the expansion slow down, why is it speeding up instead? Our best explanation is that space is suffused with a kind of energy that has no gravitational effect and emits no electromagnetic radiation, but rather acts as a kind of "antigravity" that produces a universal *repulsion*. This invisible, immaterial energy is called **dark energy.** As the name suggests, the nature of dark energy is poorly understood but is the subject of very active research.

Observations show that the *energy* density of dark energy (measured in, say, joules per cubic meter) is 68.5% of the critical density times c^2; that is, it is equal to $0.685\rho_c c^2$. As described above, the average density of matter of all kinds is 31.5% of the critical density. From the Einstein relationship $E = mc^2$, the average *energy* density of matter in the universe is therefore $0.315\rho_c c^2$. Because the energy density of dark energy is nearly three times greater than that of matter, the expansion of the universe will continue to accelerate. This expansion will never stop, and the universe will never contract.

If we account for energy of *all* kinds, the average energy density of the universe is equal to $0.685\rho_c c^2 + 0.315\rho_c c^2 = 1.00\rho_c c^2$. Of this, 68.5% is the mysterious dark energy, 26.6% is the no less mysterious dark matter, and a mere 4.9% is well-understood conventional matter. How little we know about the contents of our universe (**Fig. 44.21**)! When we take account of the density of matter in the universe (which tends to slow the expansion of space) and the density of dark energy (which tends to speed up the expansion), the age of the universe turns out to be 13.8 billion (1.38×10^{10}) years.

What is the significance of the result that within observational error, the average energy density of the universe is equal to $\rho_c c^2$? It tells us that the universe is infinite and unbounded, but just barely so. If the average energy density were even slightly larger than $\rho_c c^2$, the universe would be finite like the surface of the balloon depicted in Fig. 44.18. As of this writing, the observational error in the average energy density is less than 1%, but we can't be totally sure that the universe *is* unbounded. Improving these measurements will be an important task for physicists and astronomers in the years ahead.

44.21 The composition of our universe. Conventional matter includes all of the familiar sorts of matter that you see around you, including your body, our planet, and the sun and stars.



TEST YOUR UNDERSTANDING OF SECTION 44.6 Is it accurate to say that your body is made of "ordinary" matter?

BIO Application A Fossil Both Ancient and Recent This fossil trilobite is an example of a group of marine arthropods that flourished in earth's oceans from 540 to 250 million years ago. (By comparison, the first dinosaurs did not appear until 230 million years ago.) From our perspective, this makes trilobites almost unfathomably ancient. But compared to the time that has elapsed since the Big Bang, 13.8 billion years, even trilobites are a very recent phenomenon: They first appeared when the universe was already 96% of its present age.

44.7 THE BEGINNING OF TIME

What an odd title for the very last section of a book! We will describe in general terms some of the current theories about the very early history of the universe and their relationship to fundamental particle interactions. We'll find that an astonishing amount happened in the very first second.

Temperatures

The early universe was extremely dense and extremely hot, and the average particle energies were extremely large, all many orders of magnitude beyond anything that exists in the present universe. We can compare particle energy E and absolute temperature T by using the equipartition principle (see Section 18.4):

$$E = \tfrac{3}{2}kT \qquad (44.20)$$

In this equation k is Boltzmann's constant, which we'll often express in eV/K:

$$k = 8.617 \times 10^{-5}\ \text{eV/K}$$

Thus we can replace Eq. (44.20) by $E \approx (10^{-4}\ \text{eV/K})T = (10^{-13}\ \text{GeV/K})T$ when we're discussing orders of magnitude.

EXAMPLE 44.10 TEMPERATURE AND ENERGY

(a) What is the average kinetic energy E (in eV) of particles at room temperature ($T = 290\ \text{K}$) and at the surface of the sun ($T = 5800\ \text{K}$)? (b) What approximate temperature corresponds to the ionization energy of the hydrogen atom and to the rest energies of the electron and the proton?

SOLUTION

IDENTIFY and SET UP: In this example we are to apply the equipartition principle. We use Eq. (44.20) to relate the target variables E and T.

EXECUTE: (a) At room temperature, from Eq. (44.20),

$$E = \tfrac{3}{2}kT = \tfrac{3}{2}(8.617 \times 10^{-5}\ \text{eV/K})(290\ \text{K}) = 0.0375\ \text{eV}$$

The temperature at the sun's surface is higher than room temperature by a factor of $(5800\ \text{K})/(290\ \text{K}) = 20$, so the average kinetic energy there is $20(0.0375\ \text{eV}) = 0.75\ \text{eV}$.

(b) The ionization energy of hydrogen is 13.6 eV. Using the approximation $E \approx (10^{-4}\ \text{eV/K})T$, we have

$$T \approx \frac{E}{10^{-4}\ \text{eV/K}} = \frac{13.6\ \text{eV}}{10^{-4}\ \text{eV/K}} \approx 10^{5}\ \text{K}$$

Repeating this calculation for the rest energies of the electron ($E = 0.511\ \text{MeV}$) and proton ($E = 938\ \text{MeV}$) gives temperatures of 10^{10} K and 10^{13} K, respectively.

EVALUATE: Temperatures in excess of 10^5 K are found in the sun's interior, so most of the hydrogen there is ionized. Temperatures of 10^{10} K or 10^{13} K are not found anywhere in the solar system; as we will see, temperatures were this high in the very early universe.

Uncoupling of Interactions

We've characterized the expansion of the universe by a continual increase of the scale factor R, which we can think of very roughly as characterizing the *size* of the universe, and by a corresponding decrease in average density. As the total gravitational potential energy increased during expansion, there were corresponding *decreases* in temperature and average particle energy. As this happened, the basic interactions became progressively uncoupled.

To understand the uncouplings, recall that the unification of the electromagnetic and weak interactions occurs at energies that are large enough that the differences in mass among the various spin-1 bosons that mediate the interactions become insignificant by comparison. The electromagnetic interaction is mediated by the massless photon, and the weak interaction is mediated by the weak bosons $W^{\pm}$ and Z^0 with masses of the order of 100 GeV/c^2. At energies much *less* than 100 GeV, the two interactions seem quite different. But at energies much *greater* than 100 GeV, they become part of a single interaction, because the $W^{\pm}$ and Z^0 weak bosons become massless like the photon. (This occurs because the average value ϕ_{av} of the Higgs field is zero at high energy, as in Fig. 44.14a.)

The grand unified theories (GUTs) provide a similar behavior for the strong interaction. It becomes unified with the electroweak interaction at energies of the order of 10^{14} GeV, but at lower energies the two appear quite distinct. One of the reasons GUTs are still very speculative is that there is no way to do controlled experiments in this energy range, which is larger by a factor of 10^{11} than energies available with any current accelerator.

Finally, at sufficiently high energies and short distances, it is assumed that gravitation becomes unified with the other three interactions. The distance at which this happens is thought to be of the order of 10^{-35} m. This distance, called the *Planck length* l_P, is determined by the speed of light c and the fundamental constants of quantum mechanics and gravitation, h and G, respectively:

$$l_P = \sqrt{\frac{\hbar G}{c^3}} = 1.616 \times 10^{-35}\ \text{m} \tag{44.21}$$

You should verify that this combination of constants has units of length. The *Planck time* $t_P = l_P/c$ is the time required for light to travel a distance l_P:

$$t_P = \frac{l_P}{c} = \sqrt{\frac{\hbar G}{c^5}} = 0.539 \times 10^{-43}\ \text{s} \tag{44.22}$$

If we mentally go backward in time, we have to stop when we reach $t = 10^{-43}$ s because we have no adequate theory that unifies all four interactions. So as yet we have no way of knowing what happened or how the universe behaved at times earlier than the Planck time or when its size was less than the Planck length.

The Standard Model of the History of the Universe

The description that follows is called the *standard model* of the history of the universe. The title indicates that there are substantial areas of theory that rest on solid experimental foundations and are quite generally accepted. The figure on pages 1512–1513 is a graphical description of this history, with the characteristic temperature, particle energy, and scale factor at various times. Referring to this figure frequently will help you to understand the following discussion.

In this standard model, the temperature of the universe at time $t = 10^{-43}$ s (the Planck time) was about 10^{32} K, and the average energy per particle was approximately

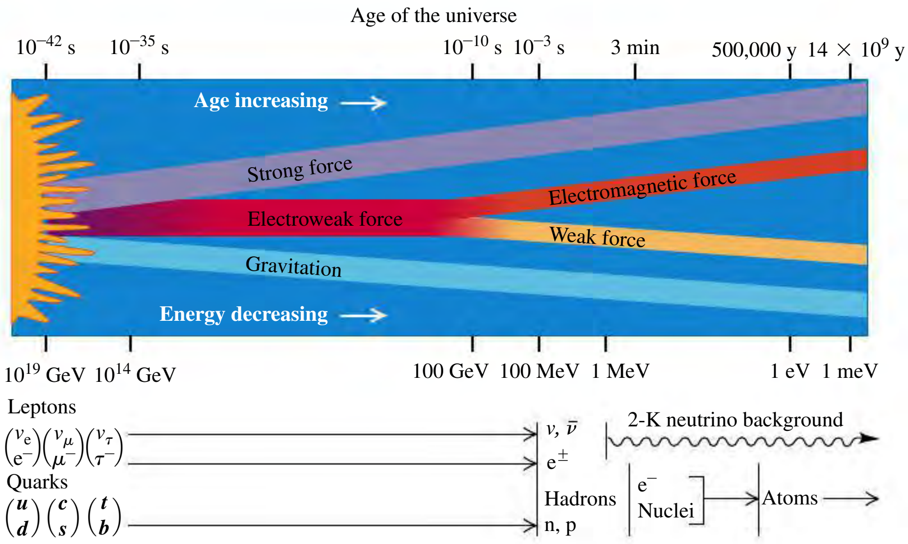
$$E \approx (10^{-13}\ \text{GeV/K})(10^{32}\ \text{K}) = 10^{19}\ \text{GeV}$$

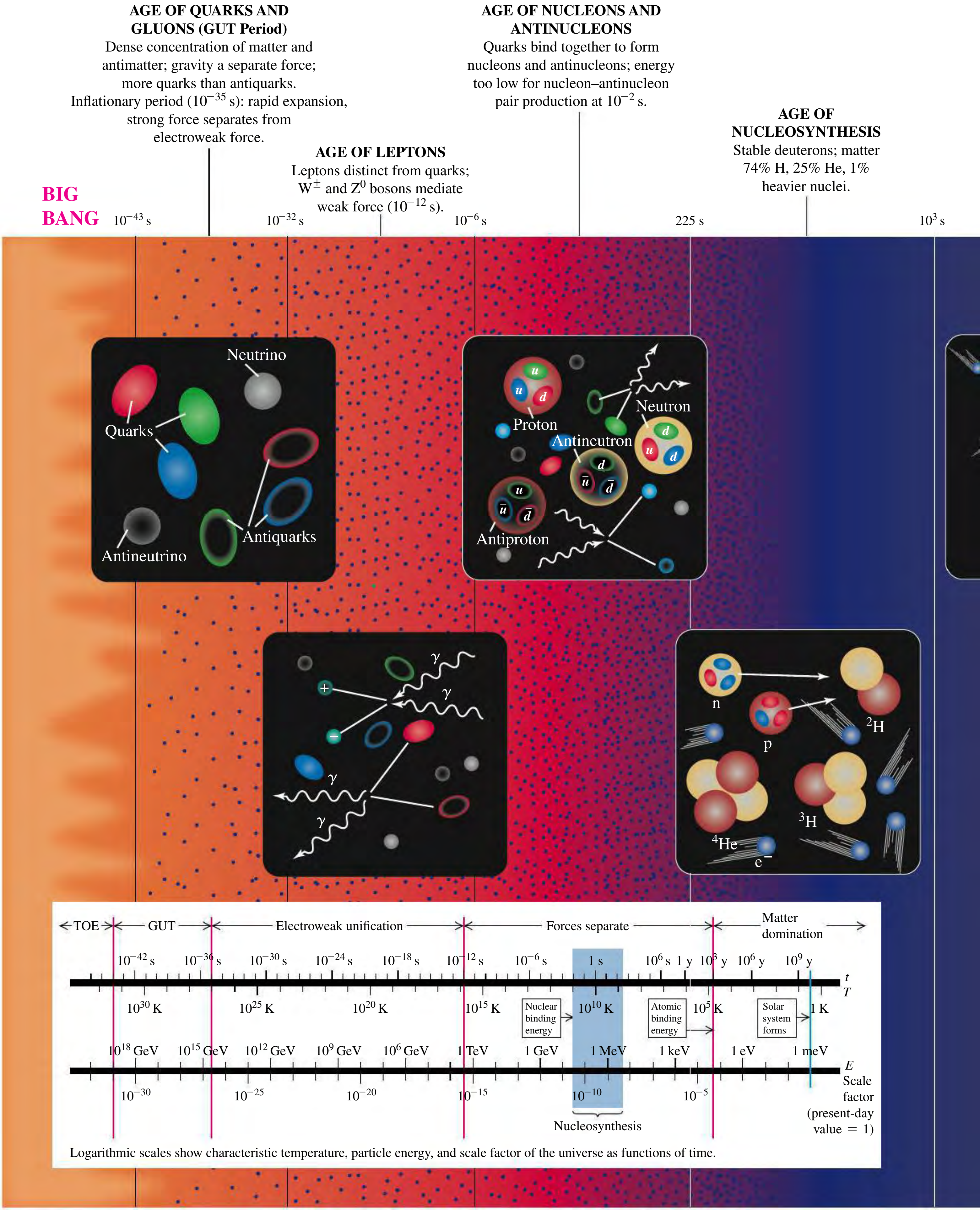
In a totally unified theory this is about the energy below which gravity begins to behave as a separate interaction. This time therefore marked the transition from any proposed TOE to the GUT period.

During the GUT period, roughly $t = 10^{-43}$ to 10^{-35} s, the strong and electroweak forces were still unified, and the universe consisted of a soup of quarks and leptons transforming into each other so freely that there was no distinction between the two families of particles. Other, much more massive particles may also have been freely created and destroyed. One important characteristic of GUTs is that at sufficiently high energies, baryon number is not conserved. (We mentioned earlier the proposed decay of the proton, which has not yet been observed.) Thus by the end of the GUT period the numbers of quarks and antiquarks may have been unequal. This point has important implications; we'll return to it at the end of the section.

By $t = 10^{-35}$ s the temperature had decreased to about 10^{27} K and the average energy to about 10^{14} GeV. At this energy the strong force separated from the electroweak force (**Fig. 44.22**), and baryon number and lepton numbers began to be separately conserved. This separation of the strong force was analogous to a phase change such as boiling a liquid, with an associated heat of vaporization. Think of it as being similar to boiling a heavy nucleus, pulling the particles apart

44.22 Schematic diagram showing the times and energies at which the various interactions are thought to have uncoupled. The energy scale is backward because the average energy decreased as the age of the universe increased.




AGE OF QUARKS AND GLUONS (GUT Period)
Dense concentration of matter and antimatter; gravity a separate force; more quarks than antiquarks. Inflationary period (10^{-35} s): rapid expansion, strong force separates from electroweak force.
AGE OF LEPTONS
Leptons distinct from quarks; $W^{\pm}$ and Z^0 bosons mediate weak force (10^{-12} s).
AGE OF NUCLEONS AND ANTINUCLEONS
Quarks bind together to form nucleons and antinucleons; energy too low for nucleon–antinucleon pair production at 10^{-2} s.
AGE OF NUCLEOSYNTHESIS
Stable deuterons; matter 74% H, 25% He, 1% heavier nuclei.
BIG BANG
10^{-43} s
10^{-32} s
10^{-6} s
225 s
10^{3} s
Neutrino
Quarks
Antiquarks
Antineutrino
Proton
Neutron
Antineutron
Antiproton
u
d
$\bar{u}$
$\bar{d}$
+
−
γ
n
p
^{2}H
^{3}H
^{4}He
e^-
TOE
GUT
Electroweak unification
Forces separate
Matter domination
10^{-42} s
10^{-36} s
10^{-30} s
10^{-24} s
10^{-18} s
10^{-12} s
10^{-6} s
1 s
10^{6} s
1 y
10^{3} y
10^{6} y
10^{9} y
t
T
10^{30} K
10^{25} K
10^{20} K
10^{15} K
Nuclear binding energy
10^{10} K
Atomic binding energy
10^{5} K
Solar system forms
1 K
10^{18} GeV
10^{15} GeV
10^{12} GeV
10^{9} GeV
10^{6} GeV
1 TeV
1 GeV
1 MeV
1 keV
1 eV
1 meV
E
10^{-30}
10^{-25}
10^{-20}
10^{-15}
10^{-10}
10^{-5}
Scale factor (present-day value = 1)
Nucleosynthesis
Logarithmic scales show characteristic temperature, particle energy, and scale factor of the universe as functions of time.

A Brief History of the Universe

AGE OF IONS
Expanding, cooling gas of ionized H and He.

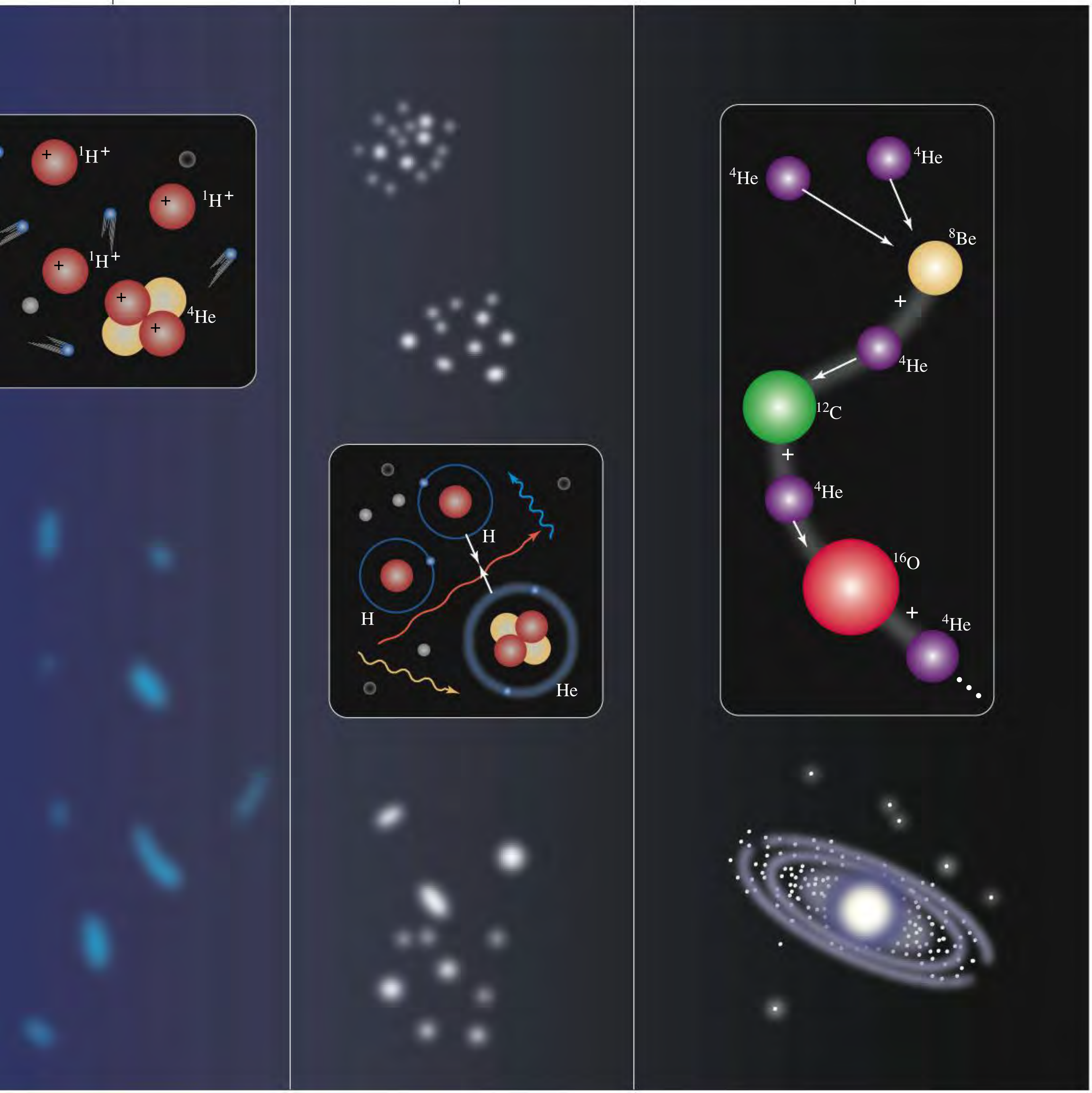
10^{13} s

AGE OF ATOMS
Neutral atoms form; universe becomes transparent to most light.

10^{15} s

AGE OF STARS AND GALAXIES
Thermonuclear fusion begins in stars, forming heavier nuclei.

NOW



beyond the short range of the nuclear force. As a result, the universe underwent a dramatic expansion (far more rapid than the present-day expansion rate) called *cosmic inflation*. In one model, the scale factor R increased by a factor of 10^{50} in 10^{-32} s.

At $t = 10^{-32}$ s the universe was a mixture of quarks, leptons, and the mediating bosons (gluons, photons, and the weak bosons $W^{\pm}$ and Z^0). It continued to expand and cool from the inflationary period to $t = 10^{-6}$ s, when the temperature was about 10^{13} K and typical energies were about 1 GeV (comparable to the rest energy of a nucleon; see Example 44.11). The quarks then began to bind together, forming nucleons and antinucleons. There were still enough photons of sufficient energy to produce nucleon–antinucleon pairs to balance the process of nucleon–antinucleon annihilation. However, by about $t = 10^{-2}$ s, most photon energies fell well below the threshold energy for such pair production. There was a slight excess of nucleons over antinucleons; as a result, virtually all of the antinucleons and most of the nucleons annihilated one another. A similar equilibrium occurred later between the production of electron–positron pairs from photons and the annihilation of such pairs. At about $t = 14$ s the average energy dropped to around 1 MeV, below the threshold for e^+e^- pair production. After pair production ceased, virtually all of the remaining positrons were annihilated, leaving the universe with many more protons and electrons than the antiparticles of each.

Up until about $t = 1$ s, neutrons and neutrinos could be produced in the endoergic reaction

$$e^- + p \rightarrow n + \nu_e$$

After this time, most electrons no longer had enough energy for this reaction. The average neutrino energy also decreased, and as the universe expanded, equilibrium reactions that involved *absorption* of neutrinos (which occurred with decreasing probability) became inoperative. At this time, in effect, the flux of neutrinos and antineutrinos throughout the universe uncoupled from the rest of the universe. Because of the extraordinarily low probability for neutrino absorption, most of this flux is still present today, although cooled greatly by expansion. The standard model of the universe predicts a present neutrino temperature of about 2 K, but no experiment has yet been able to test this prediction.

Nucleosynthesis

At about $t = 1$ s, the ratio of protons to neutrons was determined by the Boltzmann distribution factor $e^{-\Delta E/kT}$, where ΔE is the difference between the neutron and proton rest energies: $\Delta E = 1.294$ MeV. At a temperature of about 10^{10} K, this distribution factor gives about 4.5 times as many protons as neutrons. However, as we have discussed, free neutrons (with a half-life of 887 s) decay spontaneously to protons. This decay caused the proton-to-neutron ratio to increase until about $t = 225$ s. At this time, the temperature was about 10^9 K, and the average energy was well below 2 MeV.

This energy distribution was critical because the binding energy of the *deuteron* (a neutron and a proton bound together) is 2.22 MeV (see Section 43.2). A neutron bound in a deuteron does not decay spontaneously. As the average energy decreased, a proton and a neutron could combine to form a deuteron, and there were fewer and fewer photons with 2.22 MeV or more of energy to dissociate the deuterons again. Therefore the combining of protons and neutrons into deuterons halted the decay of free neutrons.

The formation of deuterons starting at about $t = 225$ s marked the beginning of the period of formation of nuclei, or *nucleosynthesis*. At this time, there were about seven protons for each neutron. The deuteron (^2H) can absorb a neutron and form a triton (^3H), or it can absorb a proton and form ^3He. Then ^3H can absorb a proton and ^3He can absorb a neutron, each yielding ^4He (the alpha particle). A few ^{7}Li nuclei may also have formed by fusion of ^{3}H and ^{4}He nuclei. According to the theory, essentially all the ^{1}H and ^{4}He in the present universe formed at this time. But then the building of nuclei almost ground to a halt. The reason is that *no* nuclide with mass number $A = 5$ has a half-life greater than 10^{-21} s. Alpha particles simply do not permanently absorb neutrons or protons. The nuclide ^{8}Be that is formed by fusion of two ^{4}He nuclei is unstable, with an extremely short half-life, about 7×10^{-17} s. At this time, the average energy was still much too large for electrons to be bound to nuclei; there were not yet any atoms.

CONCEPTUAL EXAMPLE 44.11 THE RELATIVE ABUNDANCE OF HYDROGEN AND HELIUM IN THE UNIVERSE

SOLUTION

Nearly all of the protons and neutrons in the seven-to-one ratio at $t = 225$ s either formed ^{4}He or remained as ^{1}H. After this time, what was the resulting relative abundance of ^{1}H and ^{4}He, by mass?

SOLUTION

The ^{4}He nucleus contains two protons and two neutrons. For every two neutrons present at $t = 225$ s there were 14 protons. The two neutrons and two of the 14 protons make up one ^{4}He nucleus, leaving 12 protons (^{1}H nuclei). So there were eventually 12 ^{1}H nuclei for every ^{4}He nucleus. The masses of ^{1}H and ^{4}He are about 1 u and 4 u, respectively, so there were 12 u of ^{1}H for every 4 u of ^{4}He. Therefore the relative abundance, by mass, was 75% ^{1}H and 25% ^{4}He. This result agrees very well with estimates of the present H–He ratio in the universe, an important confirmation of this part of the theory.

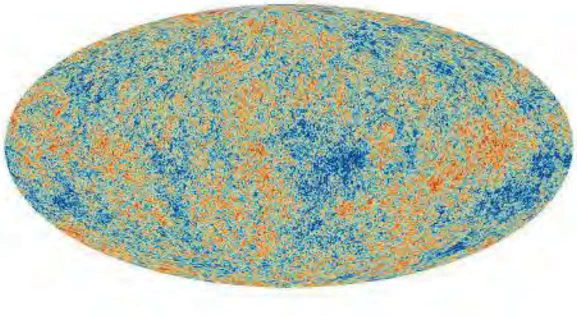
Further nucleosynthesis did not occur until very much later, well after $t = 10^{13}$ s (about 380,000 y). At that time, the temperature was about 3000 K, and the average energy was a few tenths of an electron volt. Because the ionization energies of hydrogen and helium atoms are 13.6 eV and 24.5 eV, respectively, almost all the hydrogen and helium was electrically neutral (not ionized). With the electrical repulsions of the nuclei canceled out, gravitational attraction could slowly pull the neutral atoms together to form clouds of gas and eventually stars. Thermonuclear reactions in stars then produced all of the more massive nuclei. In Section 43.8 we discussed one cycle of thermonuclear reactions in which ^{1}H becomes ^{4}He.

For stars whose mass is 40% of the sun's mass or greater, as the hydrogen is consumed the star's core begins to contract as the inward gravitational pressure exceeds the outward gas and radiation pressure. The gravitational potential energy decreases as the core contracts, so the kinetic energy of nuclei in the core increases. Eventually the core temperature becomes high enough to begin another process, *helium fusion*. First two ^{4}He nuclei fuse to form ^{8}Be, which is highly unstable. But because a star's core is so dense and collisions among nuclei are so frequent, there is a nonzero probability that a third ^{4}He nucleus will fuse with the ^{8}Be nucleus before it can decay. The result is the stable nuclide ^{12}C. This is called the *triple-alpha process,* since three ^{4}He nuclei (that is, alpha particles) fuse to form one carbon nucleus. Then successive fusions with ^{4}He give ^{16}O, ^{20}Ne, and ^{24}Mg. All these reactions are exoergic. They release energy to heat up the star, and ^{12}C and ^{16}O can fuse to form elements with higher and higher atomic number.

For nuclides that can be created in this manner, the binding energy per nucleon peaks at mass number $A = 56$ with the nuclide ^{56}Fe, so exoergic fusion reactions stop with Fe. But successive neutron captures followed by beta decays can continue the synthesis of more massive nuclei. If the star is massive enough, it may eventually explode as a *supernova,* sending out into space the heavy elements that were produced by the earlier processes (**Fig. 44.23**; see also Fig. 37.7). In space, the debris and other interstellar matter can gravitationally bunch together to form a new generation of stars and planets. Our sun is one such "second-generation" star. The sun's planets and everything on them (including you) contain matter that was long ago blasted into space by an exploding supernova.

44.23 The Veil Nebula in the constellation Cygnus is a remnant of a supernova explosion that occurred more than 20,000 years ago. The gas ejected from the supernova is still moving very rapidly. Collisions between this fast-moving gas and the tenuous material of interstellar space excite the gas and cause it to glow. The portion of the nebula shown here is about 40 ly (12 pc) in length.

44.24 This false-color map shows microwave radiation from the entire sky mapped onto an oval. When this radiation was emitted 380,000 years after the Big Bang, the regions shown in blue were slightly cooler and denser than average. Within these cool, dense regions formed galaxies, including the Milky Way galaxy of which our solar system, our earth, and our selves are part.



Background Radiation

In 1965 Arno Penzias and Robert Wilson, working at Bell Telephone Laboratories in New Jersey on satellite communications, turned a microwave antenna skyward and found a background signal that had no apparent preferred direction. (This signal produces about 1% of the "hash" you see on an analog TV that's tuned to an unused channel.) Further research has shown that the radiation that is received has a frequency spectrum that fits Planck's blackbody radiation law, Eq. (39.24) (Section 39.5). The wavelength of peak intensity is 1.063 mm (in the microwave region of the spectrum), with a corresponding absolute temperature $T = 2.725$ K. Penzias and Wilson contacted physicists at Princeton University who had begun the design of an antenna to search for radiation that was a remnant from the early evolution of the universe. We mentioned above that neutral atoms began to form at about $t = 380{,}000$ years when the temperature was 3000 K. With far fewer charged particles present than previously, the universe became transparent at this time to electromagnetic radiation of long wavelength. The 3000-K blackbody radiation therefore survived, cooling to its present 2.725-K temperature as the universe expanded. The *cosmic background radiation* is among the most clear-cut experimental confirmations of the Big Bang theory. **Figure 44.24** shows a modern map of the cosmic background radiation.

EXAMPLE 44.12 EXPANSION OF THE UNIVERSE

SOLUTION

By approximately what factor has the universe expanded since $t = 380{,}000$ y?

SOLUTION

IDENTIFY and SET UP: We use the idea that as the universe has expanded, all intergalactic wavelengths have expanded with it. The Wien displacement law, Eq. (39.21), relates the peak wavelength λ_m in blackbody radiation to the temperature T. Given the temperatures of the cosmic background radiation today (2.725 K) and at $t = 380{,}000$ y (3000 K) we can determine the factor by which wavelengths have changed and hence determine the factor by which the universe has expanded.

EXECUTE: We rewrite Eq. (39.21) as

$$\lambda_m = \frac{2.90 \times 10^{-3}\ \text{m}\cdot\text{K}}{T}$$

Hence the peak wavelength λ_m is inversely proportional to T. As the universe expands, all intergalactic wavelengths (including λ_m) increase in proportion to the scale factor R. The temperature has decreased by the factor $(3000\ \text{K})/(2.725\ \text{K}) \approx 1100$, so λ_m and the scale factor must both have *increased* by this factor. Thus, between $t = 380{,}000$ y and the present, the universe has expanded by a factor of about 1100.

EVALUATE: Our results show that since $t = 380{,}000$ y, any particular intergalactic *volume* has increased by a factor of about $(1100)^3 = 1.3 \times 10^9$. They also show that when the cosmic background radiation was emitted, its peak wavelength was $\frac{1}{1100}$ of the present-day value of 1.063 mm, or 967 nm. This is in the infrared region of the spectrum.

Matter and Antimatter

One of the most remarkable features of our universe is the asymmetry between matter and antimatter. You might think that the universe should have equal numbers of protons and antiprotons and of electrons and positrons, but this doesn't appear to be the case. Theories of the early universe must explain this imbalance.

We've mentioned that most GUTs include violation of conservation of baryon number at energies at which the strong and electroweak interactions have converged. If particle–antiparticle symmetry is also violated, we have a mechanism for making more quarks than antiquarks, more leptons than antileptons, and eventually more matter than antimatter. One serious problem is that any asymmetry that is created in this way during the GUT era might be wiped out by the electroweak interaction after the end of the GUT era. If so, there must be some mechanism that creates particle–antiparticle asymmetry at a much *later* time. The problem of the matter–antimatter asymmetry is still very much an open one.

There are still many unanswered questions at the intersection of particle physics and cosmology. Is the energy density of the universe precisely equal to $\rho_c c^2$, or are there small but important differences? What is dark energy? Has the density of dark energy remained constant over the history of the universe, or has the density changed? What is dark matter? What happened during the first 10^{-43} s after the Big Bang? Can we see evidence that the strong and electroweak interactions undergo a grand unification at high energies? The search for the answers to these and many other questions about our physical world continues to be one of the most exciting adventures of the human mind.

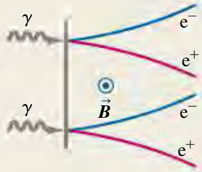
TEST YOUR UNDERSTANDING OF SECTION 44.7 Given a sufficiently powerful telescope, could we detect photons emitted earlier than $t = 380{,}000$ y?

CHAPTER 44 SUMMARY

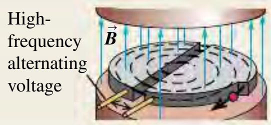
SOLUTIONS TO ALL EXAMPLES

Fundamental particles: Each particle has an antiparticle; some particles are their own antiparticles. Particles can be created and destroyed, some of them (including electrons and positrons) only in pairs or in conjunction with other particles and antiparticles.

Particles serve as mediators for the fundamental interactions. The photon is the mediator of the electromagnetic interaction. Yukawa proposed the existence of mesons to mediate the nuclear interaction. Mediating particles that can exist only because of the uncertainty principle for energy are called virtual particles.



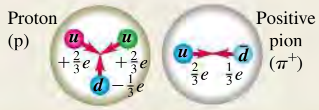
Particle accelerators and detectors: Cyclotrons, synchrotrons, and linear accelerators are used to accelerate charged particles to high energies for experiments with particle interactions. Only part of the beam energy is available to cause reactions with targets at rest. This problem is avoided in colliding-beam experiments. (See Examples 44.1–44.3.)



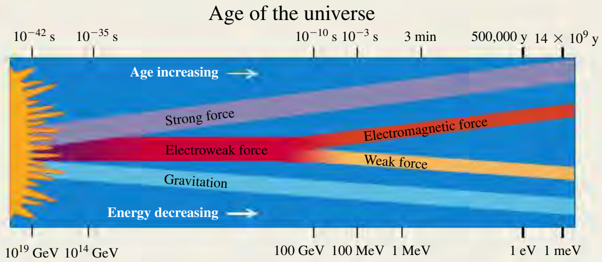
Particles and interactions: Four fundamental interactions are found in nature: the strong, electromagnetic, weak, and gravitational interactions. Particles can be described in terms of their interactions and of quantities that are conserved in all or some of the interactions.

Fermions have half-integer spins; bosons have integer spins. Leptons, which are fermions, have no strong interactions. Strongly interacting particles are called hadrons. They include mesons, which are always bosons, and baryons, which are always fermions. There are conservation laws for three different lepton numbers and for baryon number. Additional quantum numbers, including strangeness and charge, are conserved in some interactions. (See Examples 44.4–44.6.)

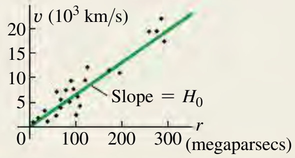
Quarks: Hadrons are composed of quarks. There are thought to be six types of quarks. The interaction between quarks is mediated by gluons. Quarks and gluons have an additional attribute called color. (See Example 44.7.)



Symmetry and the unification of interactions: Symmetry considerations play a central role in all fundamental-particle theories. The electromagnetic and weak interactions become unified at high energies into the electroweak interaction. In grand unified theories the strong interaction is also unified with these interactions, but at much higher energies.



The expanding universe and its composition: The Hubble law shows that galaxies are receding from each other and that the universe is expanding. Observations show that the rate of expansion is accelerating due to the presence of dark energy, which makes up 68.5% of the energy in the universe. Only 4.9% of the energy in the universe is in the form of conventional matter; the remaining 26.6% is dark matter, whose nature is poorly understood. (See Examples 44.8 and 44.9.)



The history of the universe: In the standard model of the universe, a Big Bang gave rise to the first fundamental particles. They eventually formed into the lightest atoms as the universe expanded and cooled. The cosmic background radiation is a relic of the time when these atoms formed. The heavier elements were manufactured much later by fusion reactions inside stars. (See Examples 44.10–44.12.)

BRIDGING PROBLEM HYPERONS, PIONS, AND THE EXPANDING UNIVERSE

SOLUTION

(a) A Λ^0 hyperon decays into a neutron and a neutral pion (π^0). Find the kinetic energies of the decay products and the fraction of the kinetic energy that is carried off by each particle. (b) A π^0 is at rest in the galaxy shown in Fig. 44.20. If a physicist on earth detects one of the two photons emitted in the decay of this π^0, what is the energy of this detected photon?

SOLUTION GUIDE

IDENTIFY and SET UP

1. Which quantities are conserved in the Λ^0 decay? In the π^0 decay?
2. The universe expanded during the time that the photon traveled from the cluster to earth. How does this affect the wavelength and energy of the photon that the physicist detects?
3. List the unknown quantities for each part of the problem and identify the target variables.
4. Select the equations that will allow you to solve for the target variables.

EXECUTE

5. Write the conservation equations for the decay of the Λ^0. [*Hint:* It's useful to write the energy E of a particle in terms of its momentum p and mass m with $E = (p^2c^2 + m^2c^4)^{1/2}$.]
6. Solve the conservation equations for the energy of one of the decay products. (*Hint:* Rearrange the energy conservation equation so that one of the $(p^2c^2 + m^2c^4)^{1/2}$ terms is on one side of the equation. Then square both sides.) Then use $K = E - mc^2$.
7. Find the fraction of the total kinetic energy that goes into the neutron and into the pion.
8. Write the conservation equations for the decay of the π^0 at rest and find the energy of each emitted photon. By what factor does the wavelength of this photon change as it travels from the galaxy to earth? By what factor does the photon *energy* change? (*Hint:* See Fig. 44.20.)

EVALUATE

9. Which of the Λ^0 decay products should have the greater kinetic energy? Should the detected π^0 decay photon have more or less energy than when it was emitted?

Problems

For assigned homework and other learning materials, go to MasteringPhysics®.

•, ••, •••: Difficulty levels. CP: Cumulative problems incorporating material from earlier chapters. CALC: Problems requiring calculus. DATA: Problems involving real data, scientific evidence, experimental design, and/or statistical reasoning. BIO: Biosciences problems.

DISCUSSION QUESTIONS

Q44.1 Is it possible that some parts of the universe contain antimatter whose atoms have nuclei made of antiprotons and antineutrons, surrounded by positrons? How could we detect this condition without actually going there? Can we detect these antiatoms by identifying the light they emit as composed of antiphotons? Explain. What problems might arise if we actually *did* go there?

Q44.2 Given the Heisenberg uncertainty principle, is it possible to create particle–antiparticle pairs that exist for extremely short periods of time before annihilating? Does this mean that empty space is really empty?

Q44.3 When they were first discovered during the 1930s and 1940s, there was confusion as to the identities of pions and muons. What are the similarities and most significant differences?

Q44.4 The gravitational force between two electrons is weaker than the electric force by the order of 10^{-40}. Yet the gravitational interactions of matter were observed and analyzed long before electrical interactions were understood. Why?

Q44.5 When a π^0 decays to two photons, what happens to the quarks of which it was made?

Q44.6 Why can't an electron decay to two photons? To two neutrinos?

Q44.7 According to the standard model of the fundamental particles, what are the similarities between baryons and leptons? What are the most important differences?

Q44.8 According to the standard model of the fundamental particles, what are the similarities between quarks and leptons? What are the most important differences?

Q44.9 The quark content of the neutron is ***udd***. (a) What is the quark content of the antineutron? Explain your reasoning. (b) Is the neutron its own antiparticle? Why or why not? (c) The quark content of the ψ is $c\bar{c}$. Is the ψ its own antiparticle? Explain your reasoning.

Q44.10 Does the universe have a center? Explain.

Q44.11 Does it make sense to ask, "If the universe is expanding, what is it expanding into?"

Q44.12 Assume that the universe has an edge. Placing yourself at that edge in a thought experiment, explain why this assumption violates the cosmological principle.

Q44.13 Explain why the cosmological principle requires that H_0 must have the same value everywhere in space, but does not require that it be constant in time.

EXERCISES

Section 44.1 Fundamental Particles—A History

44.1 • A neutral pion at rest decays into two photons. Find the energy, frequency, and wavelength of each photon. In which part of the electromagnetic spectrum does each photon lie? (Use the pion mass given in terms of the electron mass in Section 44.1.)

44.2 •• CP Two equal-energy photons collide head-on and annihilate each other, producing a $\mu^+\mu^-$ pair. The muon mass is given in terms of the electron mass in Section 44.1. (a) Calculate the maximum wavelength of the photons for this to occur. If the photons have this wavelength, describe the motion of the μ^+ and μ^- immediately after they are produced. (b) If the wavelength of each photon is half the value calculated in part (a), what is the speed of each muon after they have moved apart? Use correct relativistic expressions for momentum and energy.

44.3 •• A positive pion at rest decays into a positive muon and a neutrino. (a) Approximately how much energy is released in the decay? (Assume the neutrino has zero rest mass. Use the muon and pion masses given in terms of the electron mass in Section 44.1.) (b) Why can't a positive muon decay into a positive pion?

44.4 • A proton and an antiproton annihilate, producing two photons. Find the energy, frequency, and wavelength of each photon (a) if the p and $\overline{\text{p}}$ are initially at rest and (b) if the p and $\overline{\text{p}}$ collide head-on, each with an initial kinetic energy of 620 MeV.

44.5 •• CP For the nuclear reaction given in Eq. (44.2) assume that the initial kinetic energy and momentum of the reacting particles are negligible. Calculate the speed of the α particle immediately after it leaves the reaction region.

44.6 •• Estimate the range of the force mediated by an ω^0 meson that has mass 783 MeV/c^2.

44.7 • The starship *Enterprise*, of television and movie fame, is powered by combining matter and antimatter. If the entire 400-kg antimatter fuel supply of the *Enterprise* combines with matter, how much energy is released? How does this compare to the U.S. yearly energy use, which is roughly 1.0×10^{20} J?

Section 44.2 Particle Accelerators and Detectors

44.8 • An electron with a total energy of 30.0 GeV collides with a stationary positron. (a) What is the available energy? (b) If the electron and positron are accelerated in a collider, what total energy corresponds to the same available energy as in part (a)?

44.9 • Deuterons in a cyclotron travel in a circle with radius 32.0 cm just before emerging from the dees. The frequency of the applied alternating voltage is 9.00 MHz. Find (a) the magnetic field and (b) the kinetic energy and speed of the deuterons upon emergence.

44.10 • The magnetic field in a cyclotron that accelerates protons is 1.70 T. (a) How many times per second should the potential across the dees reverse? (This is twice the frequency of the circulating protons.) (b) The maximum radius of the cyclotron is 0.250 m. What is the maximum speed of the proton? (c) Through what potential difference must the proton be accelerated from rest to give it the speed that you calculated in part (b)?

44.11 • (a) A high-energy beam of alpha particles collides with a stationary helium gas target. What must the total energy of a beam particle be if the available energy in the collision is 16.0 GeV? (b) If the alpha particles instead interact in a colliding-beam experiment, what must the energy of each beam be to produce the same available energy?

44.12 •• Let ω_{nr} be the nonrelativistic cyclotron angular frequency given by Eq. (44.7), and let ω_{r} be the corresponding relativistic value, $\omega_{\text{r}} = (|q|B/m)\sqrt{1 - v^2/c^2}$. (a) What is the speed v of a proton for which $\omega_{\text{r}} = 0.90\omega_{\text{nr}}$ so that the two expressions differ by 10%? (b) What is the kinetic energy (in MeV) of a proton with the speed calculated in part (a)? Use the nonrelativistic expression for kinetic energy.

44.13 •• (a) What is the speed of a proton that has total energy 1000 GeV? (b) What is the angular frequency ω of a proton with the speed calculated in part (a) in a magnetic field of 4.00 T? Use both the nonrelativistic Eq. (44.7) and the correct relativistic expression, and compare the results.

44.14 •• Calculate the minimum beam energy in a proton-proton collider to initiate the $p + p \rightarrow p + p + \eta^0$ reaction. The rest energy of the η^0 is 547.3 MeV (see Table 44.3).

44.15 • In Example 44.3 it was shown that a proton beam with an 800-GeV beam energy gives an available energy of 38.7 GeV for collisions with a stationary proton target. (a) You are asked to design an upgrade of the accelerator that will double the available energy in stationary-target collisions. What beam energy is required? (b) In a colliding-beam experiment, what total energy of each beam is needed to give an available energy of $2(38.7\text{ GeV}) = 77.4\text{ GeV}$?

44.16 •• You work for a start-up company that is planning to use antiproton annihilation to produce radioactive isotopes for medical applications. One way to produce antiprotons is by the reaction $p + p \rightarrow p + p + p + \overline{p}$ in proton-proton collisions. (a) You first consider a colliding-beam experiment in which the two proton beams have equal kinetic energies. To produce an antiproton via this reaction, what is the required minimum kinetic energy of the protons in each beam? (b) You then consider the collision of a proton beam with a stationary proton target. For this experiment, what is the required minimum kinetic energy of the protons in the beam?

Section 44.3 Particles and Interactions

44.17 • A K^+ meson at rest decays into two π mesons. (a) What are the allowed combinations of π^0, π^+, and π^- as decay products? (b) Find the total kinetic energy of the π mesons.

44.18 • How much energy is released when a μ^- muon at rest decays into an electron and two neutrinos? Neglect the small masses of the neutrinos.

44.19 • What is the mass (in kg) of the Z^0? What is the ratio of the mass of the Z^0 to the mass of the proton?

44.20 • Table 44.3 shows that a Σ^0 decays into a Λ^0 and a photon. (a) Calculate the energy of the photon emitted in this decay, if the Λ^0 is at rest. (b) What is the magnitude of the momentum of the photon? Is it reasonable to ignore the final momentum and kinetic energy of the Λ^0? Explain.

44.21 • If a Σ^+ at rest decays into a proton and a π^0, what is the total kinetic energy of the decay products?

44.22 • The discovery of the Ω^- particle helped confirm Gell-Mann's eightfold way. If an Ω^- decays into a Λ^0 and a K^-, what is the total kinetic energy of the decay products?

44.23 • In which of the following decays are the three lepton numbers conserved? In each case, explain your reasoning. (a) $\mu^- \rightarrow e^- + \nu_e + \overline{\nu}_\mu$; (b) $\tau^- \rightarrow e^- + \overline{\nu}_e + \nu_\tau$; (c) $\pi^+ \rightarrow e^+ + \gamma$; (d) $n \rightarrow p + e^- + \overline{\nu}_e$.

44.24 • Which of the following reactions obey the conservation of baryon number? (a) $p + p \rightarrow p + e^+$; (b) $p + n \rightarrow 2e^+ + e^-$; (c) $p \rightarrow n + e^- + \overline{\nu}_e$; (d) $p + \overline{p} \rightarrow 2\gamma$.

44.25 • In which of the following reactions or decays is strangeness conserved? In each case, explain your reasoning. (a) $K^+ \rightarrow \mu^+ + \nu_\mu$; (b) $n + K^+ \rightarrow p + \pi^0$; (c) $K^+ + K^- \rightarrow \pi^0 + \pi^0$; (d) $p + K^- \rightarrow \Lambda^0 + \pi^0$.

Section 44.4 Quarks and Gluons

44.26 •• Determine the electric charge, baryon number, strangeness quantum number, and charm quantum number for the following quark combinations: (a) uus, (b) $c\overline{s}$, (c) $\overline{d}\overline{d}\overline{u}$, and (d) $\overline{c}b$.

44.27 • Determine the electric charge, baryon number, strangeness quantum number, and charm quantum number for the following quark combinations: (a) uds; (b) $c\overline{u}$; (c) ddd; and (d) $d\overline{c}$. Explain your reasoning.

44.28 • What is the total kinetic energy of the decay products when an upsilon particle at rest decays to $\tau^+ + \tau^-$?

44.29 • Given that each particle contains only combinations of $u, d, s, \overline{u}, \overline{d}$, and $\overline{s}$, use the method of Example 44.7 to deduce the quark content of (a) a particle with charge $+e$, baryon number 0, and strangeness $+1$; (b) a particle with charge $+e$, baryon number -1, and strangeness $+1$; (c) a particle with charge 0, baryon number $+1$, and strangeness -2.

Section 44.5 The Standard Model and Beyond

44.30 • Section 44.5 states that current experiments show that the mass of the Higgs boson is about $125\text{ GeV}/c^2$. What is the ratio of the mass of the Higgs boson to the mass of a proton?

Section 44.6 The Expanding Universe

44.31 • The spectrum of the sodium atom is detected in the light from a distant galaxy. (a) If the 590.0-nm line is redshifted to 658.5 nm, at what speed is the galaxy receding from the earth? (b) Use the Hubble law to calculate the distance of the galaxy from the earth.

44.32 •• In an experiment done in a laboratory on the earth, the wavelength of light emitted by a hydrogen atom in the $n = 4$ to $n = 2$ transition is 486.1 nm. In the light emitted by the quasar 3C273 (see Problem 36.60), this spectral line is redshifted to 563.9 nm. Assume the redshift is described by Eq. (44.14) and use the Hubble law to calculate the distance in light-years of this quasar from the earth.

44.33 • A galaxy in the constellation Pisces is 5210 Mly from the earth. (a) Use the Hubble law to calculate the speed at which this galaxy is receding from earth. (b) What redshifted ratio λ_0/λ_S is expected for light from this galaxy?

44.34 • **Redshift.** The definition of the redshift z is given in Example 44.8. (a) Show that Eq. (44.13) can be written as $1 + z = ([1 + \beta]/[1 - \beta])^{1/2}$, where $\beta = v/c$. (b) The observed redshift for a certain galaxy is $z = 0.700$. Find the speed of the galaxy relative to the earth; assume the redshift is described by Eq. (44.14). (c) Use the Hubble law to find the distance of this galaxy from the earth.

Section 44.7 The Beginning of Time

44.35 • Calculate the reaction energy Q (in MeV) for the reaction $e^- + p \rightarrow n + \nu_e$. Is this reaction endoergic or exoergic?

44.36 • Calculate the energy (in MeV) released in the triple-alpha process $3\,{}^4\text{He} \rightarrow {}^{12}\text{C}$.

44.37 • CP The 2.728-K blackbody radiation has its peak wavelength at 1.062 mm. What was the peak wavelength at $t = 700{,}000$ y when the temperature was 3000 K?

44.38 • Calculate the reaction energy Q (in MeV) for the nucleosynthesis reaction

$$ {}^{12}_{6}\text{C} + {}^{4}_{2}\text{He} \rightarrow {}^{16}_{8}\text{O} $$

Is this reaction endoergic or exoergic?

PROBLEMS

44.39 •• CP BIO Radiation Therapy with π^- Mesons. Beams of π^- mesons are used in radiation therapy for certain cancers. The energy comes from the complete decay of the π^- to *stable* particles. (a) Write out the complete decay of a π^- meson to stable particles. What are these particles? (b) How much energy is released from the complete decay of a single π^- meson to stable particles? (You can ignore the very small masses of the neutrinos.) (c) How many π^- mesons need to decay to give a dose of 50.0 Gy to 10.0 g of tissue? (d) What would be the equivalent dose in part (c) in Sv and in rem? Consult Table 43.3 and use the largest appropriate RBE for the particles involved in this decay.

44.40 •• A proton and an antiproton collide head-on with equal kinetic energies. Two γ rays with wavelengths of 0.720 fm are produced. Calculate the kinetic energy of the incident proton.

44.41 •• Calculate the threshold kinetic energy for the reaction $p + p \rightarrow p + p + K^+ + K^-$ if a proton beam is incident on a stationary proton target.

44.42 •• Calculate the threshold kinetic energy for the reaction $\pi^- + p \rightarrow \Sigma^0 + K^0$ if a π^- beam is incident on a stationary proton target. The K^0 has a mass of 497.7 MeV/c^2.

44.43 • Each of the following reactions is missing a single particle. Calculate the baryon number, charge, strangeness, and the three lepton numbers (where appropriate) of the missing particle, and from this identify the particle. (a) $p + p \rightarrow p + \Lambda^0 + ?$; (b) $K^- + n \rightarrow \Lambda^0 + ?$; (c) $p + \overline{p} \rightarrow n + ?$; (d) $\overline{\nu}_\mu + p \rightarrow n + ?$

44.44 •• An η^0 meson at rest decays into three π mesons. (a) What are the allowed combinations of π^0, π^+, and π^- as decay products? (b) Find the total kinetic energy of the π mesons.

44.45 • The ϕ meson has mass 1019.4 MeV/c^2 and a measured energy width of 4.4 MeV/c^2. Using the uncertainty principle, estimate the lifetime of the ϕ meson.

44.46 • Estimate the energy width (energy uncertainty) of the ψ if its mean lifetime is 7.6×10^{-21} s. What fraction is this of its rest energy?

44.47 •• CP BIO One proposed proton decay is $p^+ \rightarrow e^+ + \pi^0$, which violates both baryon and lepton number conservation, so the proton lifetime is expected to be very long. Suppose the proton half-life were 1.0×10^{18} y. (a) Calculate the energy deposited per kilogram of body tissue (in rad) due to the decay of the protons in your body in one year. Model your body as consisting entirely of water. Only the two protons in the hydrogen atoms in each H_2O molecule would decay in the manner shown; do you see why? Assume that the π^0 decays to two γ rays, that the positron annihilates with an electron, and that all the energy produced in the primary decay and these secondary decays remains in your body. (b) Calculate the equivalent dose (in rem) assuming an RBE of 1.0 for all the radiation products, and compare with the 0.1 rem due to the natural background and the 5.0-rem guideline for industrial workers. Based on your calculation, can the proton lifetime be as short as 1.0×10^{18} y?

44.48 •• A ϕ meson (see Problem 44.45) at rest decays via $\phi \rightarrow K^+ + K^-$. It has strangeness 0. (a) Find the kinetic energy of the K^+ meson. (Assume that the two decay products share kinetic energy equally, since their masses are equal.) (b) Suggest a reason the decay $\phi \rightarrow K^+ + K^- + \pi^0$ has not been observed. (c) Suggest reasons the decays $\phi \rightarrow K^+ + \pi^-$ and $\phi \rightarrow K^+ + \mu^-$ have not been observed.

44.49 •• Cosmic Jerk. The densities of ordinary matter and dark matter have decreased as the universe has expanded, since the same amount of mass occupies an ever-increasing volume. Yet observations suggest that the density of dark energy has remained constant over the entire history of the universe. (a) Explain why the expansion of the universe actually slowed down in its early history but is speeding up today. "Jerk" is the term for a change in acceleration, so the change in cosmic expansion from slowing down to speeding up is called *cosmic jerk*. (b) Calculations show that the change in acceleration took place when the combined density of matter of all kinds was equal to twice the density of dark energy. Compared to today's value of the scale factor, what was the scale factor at that time? (c) We see the galaxies in Figs. 44.16b and 44.20 as they were 300 million years ago and 13.1 billion years ago. Was the expansion of the universe slowing down or speeding up at these times? (*Hint:* See the caption for Fig. 44.20.)

44.50 ••• CP A Ξ^- particle at rest decays to a Λ^0 and a π^-. (a) Find the total kinetic energy of the decay products. (b) What fraction of the energy is carried off by each particle? (Use relativistic expressions for momentum and energy.)

44.51 ••• CP A Σ^- particle moving in the $+x$-direction with kinetic energy 180 MeV decays into a π^- and a neutron. The π^- moves in the $+y$-direction. What is the kinetic energy of the neutron, and what is the direction of its velocity? Use relativistic expressions for energy and momentum.

44.52 ••• CP The K^0 meson has rest energy 497.7 MeV. A K^0 meson moving in the $+x$-direction with kinetic energy 225 MeV decays into a π^+ and a π^-, which move off at equal angles above and below the $+x$-axis. Calculate the kinetic energy of the π^+ and the angle it makes with the $+x$-axis. Use relativistic expressions for energy and momentum.

44.53 •• DATA While tuning up a medical cyclotron for use in isotope production, you obtain the data given in the table.

B (T)	0.10	0.20	0.30	0.40
K_{max} (MeV)	0.068	0.270	0.608	1.080

B is the uniform magnetic field in the cyclotron, and K_{max} is the maximum kinetic energy of the particle being accelerated, which is a proton. The radius R of the proton path at maximum kinetic energy has the same value for each magnetic-field value. (a) Compare the kinetic energy values in the table to the rest energy mc^2 of a proton. Is it necessary to use relativistic expressions in your analysis? Explain. (b) Graph your data as K_{max} versus B^2. Use the slope of the best-fit straight line to your data to find R. (c) What is the maximum kinetic energy for a 0.25-T magnetic field? (d) What is the angular frequency ω of the proton when $B = 0.40$ T?

44.54 •• DATA The decay products from the decay of short-lived unstable particles can provide evidence that these particles have been produced in a collision experiment. As an initial step in designing an experiment to detect short-lived hadrons, you make a literature study of their decays. Table 44.3 gives experimental data for the mass and typical decay modes of the particles Σ^-, Ξ^0, Δ^{++}, and Ω^-. (a) Which of these four particles has the largest mass? The smallest? (b) By the decay modes shown in the table, for which of these particles do the decay products have the greatest total kinetic energy? The least?

44.55 •• DATA You have entered a graduate program in particle physics and are learning about the use of symmetry. You begin by repeating the analysis that led to the prediction of the Ω^- particle. Nine of the spin-$\frac{3}{2}$ baryons are four Δ particles, each with mass 1232 MeV/c^2, strangeness 0, and charges $+2e$, $+e$, 0, and $-e$; three Σ^* particles, each with mass 1385 MeV/c^2, strangeness -1, and charges $+e$, 0, and $-e$; and two Ξ^* particles, each with mass 1530 MeV/c^2, strangeness -2, and charges 0 and $-e$.

(a) Place these particles on a plot of S versus Q. Deduce the Q and S values of the tenth spin-$\frac{3}{2}$ baryon, the Ω^- particle, and place it on your diagram. Also label the particles with their masses. The mass of the Ω^- is 1672 MeV/c^2; is this value consistent with your diagram? (b) Deduce the three-quark combinations (of $\boldsymbol{u}$, $\boldsymbol{d}$, and $\boldsymbol{s}$) that make up each of these ten particles. Redraw the plot of S versus Q from part (a) with each particle labeled by its quark content. What regularities do you see?

CHALLENGE PROBLEM

44.56 ••• **CP** Consider a collision in which a stationary particle with mass M is bombarded by a particle with mass m, speed v_0, and total energy (including rest energy) E_m. (a) Use the Lorentz transformation to write the velocities v_m and v_M of particles m and M in terms of the speed v_{cm} of the center of momentum. (b) Use the fact that the total momentum in the center-of-momentum frame is zero to obtain an expression for v_{cm} in terms of m, M, and v_0. (c) Combine the results of parts (a) and (b) to obtain Eq. (44.9) for the total energy in the center-of-momentum frame.

PASSAGE PROBLEMS

BIO LOOKING UNDER THE HOOD OF PET. In the imaging method called positron emission tomography (PET), a patient is injected with molecules containing atoms that have nuclei with an excess of protons. As they decay into neutrons, these protons emit positrons. An emitted positron travels a short distance and slows to near-zero velocity; when it encounters an electron, they may annihilate each other and emit two photons in opposite directions. The patient is enclosed in a circular array of detectors, with the tissue to be imaged centered in the array. If two photons of the proper energy strike two detectors simultaneously (within 10 ns), we can conclude that the photons were produced by positron–electron annihilation somewhere along a line connecting the detectors. By observing many such simultaneous events, we can create a map of the distribution of positron-emitting atoms in the tissue. However, photons can be absorbed or scattered as they pass through tissue. The number of photons remaining after they travel a distance x through tissue is given by $N = N_0 e^{-\mu x}$, where N_0 is the initial number of photons and μ is the attenuation coefficient, which is approximately 0.1 cm^{-1} for photons of this energy. The index of refraction of biological tissue for x rays is 1.

44.57 What is the energy of each photon produced by positron–electron annihilation? (a) $\frac{1}{2}m_e v^2$, where v is the speed of the emitted positron; (b) $m_e v^2$; (c) $\frac{1}{2}m_e c^2$; (d) $m_e c^2$.

44.58 Suppose that positron–electron annihilations occur on the line 3 cm from the center of the line connecting two detectors. Will the resultant photons be counted as having arrived at these detectors simultaneously? (a) No, because the time difference between their arrivals is 100 ms; (b) no, because the time difference is 200 ms; (c) yes, because the time difference is 0.1 ns; (d) yes, because the time difference is 0.2 ns.

44.59 If the annihilation photons come from a part of the body that is separated from the detector by 20 cm of tissue, what percentage of the photons that originally travelled toward the detector remains after they have passed through the tissue? (a) 1.4%; (b) 8.6%; (c) 14%; (d) 86%.

Answers

Chapter Opening Question ?

(v) Only 4.9% of the mass and energy of the universe is in the form of "normal" matter. Of the rest, 26.6% is poorly understood dark matter and 68.5% is even more mysterious dark energy.

Test Your Understanding Questions

44.1 **(i), (iii), (ii), (iv)** The more massive the virtual particle, the shorter its lifetime and the shorter the distance that it can travel during its lifetime.

44.2 **no** In a head-on collision between an electron and a positron of equal energy, the net momentum is zero. Since both momentum and energy are conserved in the collision, the virtual photon also has momentum $p = 0$ but has energy $E = 90\text{ GeV} + 90\text{ GeV} = 180\text{ GeV}$. Hence the relationship $E = pc$ is definitely *not* true for this virtual photon.

44.3 **no** Mesons all have baryon number $B = 0$, while a proton has $B = 1$. The decay of a proton into one or more mesons would require that baryon number *not* be conserved. No violation of this conservation principle has ever been observed, so the proposed decay is impossible.

44.4 **no** Only the $\boldsymbol{s}$ quark, with $S = -1$, has nonzero strangeness. For a baryon to have $S = -2$, it must have two $\boldsymbol{s}$ quarks and one quark of a different flavor. Since each $\boldsymbol{s}$ quark has charge $-\frac{1}{3}e$, the nonstrange quark must have charge $+\frac{5}{3}e$ to make the net charge equal to $+e$. But *no* quark has charge $+\frac{5}{3}e$, so the proposed baryon is impossible.

44.5 **(i)** If a $\boldsymbol{d}$ quark in a neutron (quark content $\boldsymbol{udd}$) undergoes the process $\boldsymbol{d} \rightarrow \boldsymbol{u} + e^- + \bar{\nu}_e$, the remaining baryon has quark content $\boldsymbol{uud}$ and hence is a proton (see Fig. 44.11). An electron is the same as a β^- particle, so the net result is beta-minus decay: $n \rightarrow p + \beta^- + \bar{\nu}_e$.

44.6 **yes . . . and no** The material of which your body is made is ordinary to us on earth. But from a cosmic perspective your material is quite *extraordinary:* Only 4.9% of the mass and energy in the universe is in the form of atoms.

44.7 **no** Prior to $t = 380{,}000$ y the temperature was so high that atoms could not form, so free electrons and protons were plentiful. These charged particles are very effective at scattering photons, so light could not propagate very far and the universe was opaque. The oldest photons that we can detect date from the time $t = 380{,}000$ y when atoms formed and the universe became transparent.

Bridging Problem

(a) Neutron: 5.78 MeV (0.140 of total);
pion: 35.62 MeV (0.860 of total)

(b) 8.1 MeV

APPENDIX A

THE INTERNATIONAL SYSTEM OF UNITS

The Système International d'Unités, abbreviated SI, is the system developed by the General Conference on Weights and Measures and adopted by nearly all the industrial nations of the world. The following material is adapted from the National Institute of Standards and Technology (**http://physics.nist.gov/cuu**).

Quantity	Name of unit	Symbol	
	SI base units		
length	meter	m	
mass	kilogram	kg	
time	second	s	
electric current	ampere	A	
thermodynamic temperature	kelvin	K	
amount of substance	mole	mol	
luminous intensity	candela	cd	
	SI derived units		**Equivalent units**
area	square meter	m^2	
volume	cubic meter	m^3	
frequency	hertz	Hz	s^{-1}
mass density (density)	kilogram per cubic meter	kg/m^3	
speed, velocity	meter per second	m/s	
angular velocity	radian per second	rad/s	
acceleration	meter per second squared	m/s^2	
angular acceleration	radian per second squared	rad/s^2	
force	newton	N	$kg \cdot m/s^2$
pressure (mechanical stress)	pascal	Pa	N/m^2
kinematic viscosity	square meter per second	m^2/s	
dynamic viscosity	newton-second per square meter	$N \cdot s/m^2$	
work, energy, quantity of heat	joule	J	$N \cdot m$
power	watt	W	J/s
quantity of electricity	coulomb	C	$A \cdot s$
potential difference, electromotive force	volt	V	J/C, W/A
electric field strength	volt per meter	V/m	N/C
electrical resistance	ohm	Ω	V/A
capacitance	farad	F	$A \cdot s/V$
magnetic flux	weber	Wb	$V \cdot s$
inductance	henry	H	$V \cdot s/A$
magnetic flux density	tesla	T	Wb/m^2
magnetic field strength	ampere per meter	A/m	
magnetomotive force	ampere	A	
luminous flux	lumen	lm	$cd \cdot sr$
luminance	candela per square meter	cd/m^2	
illuminance	lux	lx	lm/m^2
wave number	1 per meter	m^{-1}	
entropy	joule per kelvin	J/K	
specific heat capacity	joule per kilogram-kelvin	$J/kg \cdot K$	
thermal conductivity	watt per meter-kelvin	$W/m \cdot K$	

Quantity	Name of unit	Symbol	Equivalent units
radiant intensity	watt per steradian	W/sr	
activity (of a radioactive source)	becquerel	Bq	s^{-1}
radiation dose	gray	Gy	J/kg
radiation dose equivalent	sievert	Sv	J/kg
	SI supplementary units		
plane angle	radian	rad	
solid angle	steradian	sr	

Definitions of SI Units

meter (m) The *meter* is the length equal to the distance traveled by light, in vacuum, in a time of 1/299,792,458 second.

kilogram (kg) The *kilogram* is the unit of mass; it is equal to the mass of the international prototype of the kilogram. (The international prototype of the kilogram is a particular cylinder of platinum-iridium alloy that is preserved in a vault at Sévres, France, by the International Bureau of Weights and Measures.)

second (s) The *second* is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom.

ampere (A) The *ampere* is that constant current that, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per meter of length.

kelvin (K) The *kelvin,* unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

ohm (Ω) The *ohm* is the electric resistance between two points of a conductor when a constant difference of potential of 1 volt, applied between these two points, produces in this conductor a current of 1 ampere, this conductor not being the source of any electromotive force.

coulomb (C) The *coulomb* is the quantity of electricity transported in 1 second by a current of 1 ampere.

candela (cd) The *candela* is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian.

mole (mol) The *mole* is the amount of substance of a system that contains as many elementary entities as there are carbon atoms in 0.012 kg of carbon 12. The elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.

newton (N) The *newton* is that force that gives to a mass of 1 kilogram an acceleration of 1 meter per second per second.

joule (J) The *joule* is the work done when the point of application of a constant force of 1 newton is displaced a distance of 1 meter in the direction of the force.

watt (W) The *watt* is the power that gives rise to the production of energy at the rate of 1 joule per second.

volt (V) The *volt* is the difference of electric potential between two points of a conducting wire carrying a constant current of 1 ampere, when the power dissipated between these points is equal to 1 watt.

weber (Wb) The *weber* is the magnetic flux that, linking a circuit of one turn, produces in it an electromotive force of 1 volt as it is reduced to zero at a uniform rate in 1 second.

lumen (lm) The *lumen* is the luminous flux emitted in a solid angle of 1 steradian by a uniform point source having an intensity of 1 candela.

farad (F) The *farad* is the capacitance of a capacitor between the plates of which there appears a difference of potential of 1 volt when it is charged by a quantity of electricity equal to 1 coulomb.

henry (H) The *henry* is the inductance of a closed circuit in which an electromotive force of 1 volt is produced when the electric current in the circuit varies uniformly at a rate of 1 ampere per second.

radian (rad) The *radian* is the plane angle between two radii of a circle that cut off on the circumference an arc equal in length to the radius.

steradian (sr) The *steradian* is the solid angle that, having its vertex in the center of a sphere, cuts off an area of the surface of the sphere equal to that of a square with sides of length equal to the radius of the sphere.

SI Prefixes To form the names of multiples and submultiples of SI units, apply the prefixes listed in Appendix F.

APPENDIX B

USEFUL MATHEMATICAL RELATIONS

Algebra

$a^{-x} = \frac{1}{a^x} \qquad a^{(x+y)} = a^x a^y \qquad a^{(x-y)} = \frac{a^x}{a^y}$

Logarithms: If $\log a = x$, then $a = 10^x$. $\quad \log a + \log b = \log(ab) \quad \log a - \log b = \log(a/b) \quad \log(a^n) = n \log a$

If $\ln a = x$, then $a = e^x$. $\quad \ln a + \ln b = \ln(ab) \quad \ln a - \ln b = \ln(a/b) \quad \ln(a^n) = n \ln a$

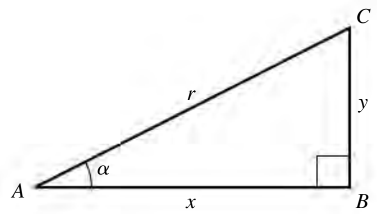
Quadratic formula: If $ax^2 + bx + c = 0$, $\qquad x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

Binomial Theorem

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)a^{n-2}b^2}{2!} + \frac{n(n-1)(n-2)a^{n-3}b^3}{3!} + \cdots$$

Trigonometry

In the right triangle ABC, $x^2 + y^2 = r^2$.



Definitions of the trigonometric functions:

$\sin\alpha = y/r \qquad \cos\alpha = x/r \qquad \tan\alpha = y/x$

Identities:

$\sin^2\alpha + \cos^2\alpha = 1 \qquad \tan\alpha = \dfrac{\sin\alpha}{\cos\alpha}$

$\sin 2\alpha = 2\sin\alpha\cos\alpha \qquad \cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 2\cos^2\alpha - 1 = 1 - 2\sin^2\alpha$

$\sin\frac{1}{2}\alpha = \sqrt{\dfrac{1-\cos\alpha}{2}} \qquad \cos\frac{1}{2}\alpha = \sqrt{\dfrac{1+\cos\alpha}{2}}$

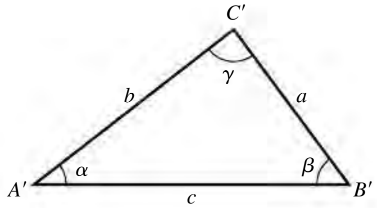
$\sin(-\alpha) = -\sin\alpha \qquad \sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$

$\cos(-\alpha) = \cos\alpha \qquad \cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$

$\sin(\alpha \pm \pi/2) = \pm\cos\alpha \qquad \sin\alpha + \sin\beta = 2\sin\frac{1}{2}(\alpha+\beta)\cos\frac{1}{2}(\alpha-\beta)$

$\cos(\alpha \pm \pi/2) = \mp\sin\alpha \qquad \cos\alpha + \cos\beta = 2\cos\frac{1}{2}(\alpha+\beta)\cos\frac{1}{2}(\alpha-\beta)$

For *any* triangle $A'B'C'$ (not necessarily a right triangle) with sides a, b, and c and angles α, β, and γ:



Law of sines: $\dfrac{\sin\alpha}{a} = \dfrac{\sin\beta}{b} = \dfrac{\sin\gamma}{c}$

Law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\gamma$

Geometry

Circumference of circle of radius r:	$C = 2\pi r$	Surface area of sphere of radius r:	$A = 4\pi r^2$
Area of circle of radius r:	$A = \pi r^2$	Volume of cylinder of radius r and height h:	$V = \pi r^2 h$
Volume of sphere of radius r:	$V = 4\pi r^3/3$		

Calculus

Derivatives:

$$\frac{d}{dx}x^n = nx^{n-1} \qquad \frac{d}{dx}\ln ax = \frac{1}{x} \qquad \frac{d}{dx}e^{ax} = ae^{ax}$$

$$\frac{d}{dx}\sin ax = a\cos ax \qquad \frac{d}{dx}\cos ax = -a\sin ax$$

Integrals:

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} \quad (n \neq -1) \qquad \int \frac{dx}{x} = \ln x \qquad \int e^{ax}\,dx = \frac{1}{a}e^{ax}$$

$$\int \sin ax\,dx = -\frac{1}{a}\cos ax \qquad \int \cos ax\,dx = \frac{1}{a}\sin ax \qquad \int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsin\frac{x}{a}$$

$$\int \frac{dx}{\sqrt{x^2 + a^2}} = \ln\left(x + \sqrt{x^2 + a^2}\right) \qquad \int \frac{dx}{x^2 + a^2} = \frac{1}{a}\arctan\frac{x}{a} \qquad \int \frac{dx}{(x^2 + a^2)^{3/2}} = \frac{1}{a^2}\frac{x}{\sqrt{x^2 + a^2}}$$

$$\int \frac{x\,dx}{(x^2 + a^2)^{3/2}} = -\frac{1}{\sqrt{x^2 + a^2}}$$

Power series (convergent for range of x shown):

$$(1 + x)^n = 1 + nx + \frac{n(n-1)x^2}{2!} + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots \; (|x| < 1)$$

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \cdots \; (|x| < \pi/2)$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \; (\text{all } x)$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \; (\text{all } x)$$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \; (|x| < 1)$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \; (\text{all } x)$$

APPENDIX C

THE GREEK ALPHABET

Name	Capital	Lowercase	Name	Capital	Lowercase	Name	Capital	Lowercase
Alpha	A	α	Iota	I	ι	Rho	P	ρ
Beta	B	β	Kappa	K	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	T	τ
Delta	Δ	δ	Mu	M	μ	Upsilon	Υ	υ
Epsilon	E	ϵ	Nu	N	ν	Phi	Φ	ϕ
Zeta	Z	ζ	Xi	Ξ	ξ	Chi	X	χ
Eta	H	η	Omicron	O	o	Psi	Ψ	ψ
Theta	Θ	θ	Pi	Π	π	Omega	Ω	ω

APPENDIX D

PERIODIC TABLE OF THE ELEMENTS

Group / Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 **H** 1.008																	2 **He** 4.003
2	3 **Li** 6.941	4 **Be** 9.012											5 **B** 10.811	6 **C** 12.011	7 **N** 14.007	8 **O** 15.999	9 **F** 18.998	10 **Ne** 20.180
3	11 **Na** 22.990	12 **Mg** 24.305											13 **Al** 26.982	14 **Si** 28.086	15 **P** 30.974	16 **S** 32.065	17 **Cl** 35.453	18 **Ar** 39.948
4	19 **K** 39.098	20 **Ca** 40.078	21 **Sc** 44.956	22 **Ti** 47.867	23 **V** 50.942	24 **Cr** 51.996	25 **Mn** 54.938	26 **Fe** 55.845	27 **Co** 58.933	28 **Ni** 58.693	29 **Cu** 63.546	30 **Zn** 65.409	31 **Ga** 69.723	32 **Ge** 72.64	33 **As** 74.922	34 **Se** 78.96	35 **Br** 79.904	36 **Kr** 83.798
5	37 **Rb** 85.468	38 **Sr** 87.62	39 **Y** 88.906	40 **Zr** 91.224	41 **Nb** 92.906	42 **Mo** 95.94	43 **Tc** (98)	44 **Ru** 101.07	45 **Rh** 102.906	46 **Pd** 106.42	47 **Ag** 107.868	48 **Cd** 112.411	49 **In** 114.818	50 **Sn** 118.710	51 **Sb** 121.760	52 **Te** 127.60	53 **I** 126.904	54 **Xe** 131.293
6	55 **Cs** 132.905	56 **Ba** 137.327	71 **Lu** 174.967	72 **Hf** 178.49	73 **Ta** 180.948	74 **W** 183.84	75 **Re** 186.207	76 **Os** 190.23	77 **Ir** 192.217	78 **Pt** 195.078	79 **Au** 196.967	80 **Hg** 200.59	81 **Tl** 204.383	82 **Pb** 207.2	83 **Bi** 208.980	84 **Po** (209)	85 **At** (210)	86 **Rn** (222)
7	87 **Fr** (223)	88 **Ra** (226)	103 **Lr** (262)	104 **Rf** (261)	105 **Db** (262)	106 **Sg** (266)	107 **Bh** (270)	108 **Hs** (269)	109 **Mt** (278)	110 **Ds** (281)	111 **Rg** (281)	112 **Cn** (285)	113 **Uut** (284)	114 **Fl** (289)	115 **Uup** (288)	116 **Lv** (292)	117 **Uus** (294)	118 **Uuo** (294)

Lanthanoids	57 **La** 138.905	58 **Ce** 140.116	59 **Pr** 140.908	60 **Nd** 144.24	61 **Pm** (145)	62 **Sm** 150.36	63 **Eu** 151.964	64 **Gd** 157.25	65 **Tb** 158.925	66 **Dy** 162.500	67 **Ho** 164.930	68 **Er** 167.259	69 **Tm** 168.934	70 **Yb** 173.04
Actinoids	89 **Ac** (227)	90 **Th** (232)	91 **Pa** (231)	92 **U** (238)	93 **Np** (237)	94 **Pu** (244)	95 **Am** (243)	96 **Cm** (247)	97 **Bk** (247)	98 **Cf** (251)	99 **Es** (252)	100 **Fm** (257)	101 **Md** (258)	102 **No** (259)

For each element, the average atomic mass of the mixture of isotopes occurring in nature is shown. For elements having no stable isotope, the approximate atomic mass of the longest-lived isotope is shown in parentheses. All atomic masses are expressed in atomic mass units $(1\ \text{u} = 1.660538921(73) \times 10^{-27}\ \text{kg})$, equivalent to grams per mole (g/mol).

APPENDIX E

UNIT CONVERSION FACTORS

Length

$1\ \text{m} = 100\ \text{cm} = 1000\ \text{mm} = 10^6\ \mu\text{m} = 10^9\ \text{nm}$
$1\ \text{km} = 1000\ \text{m} = 0.6214\ \text{mi}$
$1\ \text{m} = 3.281\ \text{ft} = 39.37\ \text{in.}$
$1\ \text{cm} = 0.3937\ \text{in.}$
$1\ \text{in.} = 2.540\ \text{cm}$
$1\ \text{ft} = 30.48\ \text{cm}$
$1\ \text{yd} = 91.44\ \text{cm}$
$1\ \text{mi} = 5280\ \text{ft} = 1.609\ \text{km}$
$1\ \text{Å} = 10^{-10}\ \text{m} = 10^{-8}\ \text{cm} = 10^{-1}\ \text{nm}$
$1\ \text{nautical mile} = 6080\ \text{ft}$
$1\ \text{light-year} = 9.461 \times 10^{15}\ \text{m}$

Area

$1\ \text{cm}^2 = 0.155\ \text{in.}^2$
$1\ \text{m}^2 = 10^4\ \text{cm}^2 = 10.76\ \text{ft}^2$
$1\ \text{in.}^2 = 6.452\ \text{cm}^2$
$1\ \text{ft}^2 = 144\ \text{in.}^2 = 0.0929\ \text{m}^2$

Volume

$1\ \text{liter} = 1000\ \text{cm}^3 = 10^{-3}\ \text{m}^3 = 0.03531\ \text{ft}^3 = 61.02\ \text{in.}^3$
$1\ \text{ft}^3 = 0.02832\ \text{m}^3 = 28.32\ \text{liters} = 7.477\ \text{gallons}$
$1\ \text{gallon} = 3.788\ \text{liters}$

Time

$1\ \text{min} = 60\ \text{s}$
$1\ \text{h} = 3600\ \text{s}$
$1\ \text{d} = 86{,}400\ \text{s}$
$1\ \text{y} = 365.24\ \text{d} = 3.156 \times 10^7\ \text{s}$

Angle

$1\ \text{rad} = 57.30° = 180°/\pi$
$1° = 0.01745\ \text{rad} = \pi/180\ \text{rad}$
$1\ \text{revolution} = 360° = 2\pi\ \text{rad}$
$1\ \text{rev/min (rpm)} = 0.1047\ \text{rad/s}$

Speed

$1\ \text{m/s} = 3.281\ \text{ft/s}$
$1\ \text{ft/s} = 0.3048\ \text{m/s}$
$1\ \text{mi/min} = 60\ \text{mi/h} = 88\ \text{ft/s}$
$1\ \text{km/h} = 0.2778\ \text{m/s} = 0.6214\ \text{mi/h}$
$1\ \text{mi/h} = 1.466\ \text{ft/s} = 0.4470\ \text{m/s} = 1.609\ \text{km/h}$
$1\ \text{furlong/fortnight} = 1.662 \times 10^{-4}\ \text{m/s}$

Acceleration

$1\ \text{m/s}^2 = 100\ \text{cm/s}^2 = 3.281\ \text{ft/s}^2$
$1\ \text{cm/s}^2 = 0.01\ \text{m/s}^2 = 0.03281\ \text{ft/s}^2$
$1\ \text{ft/s}^2 = 0.3048\ \text{m/s}^2 = 30.48\ \text{cm/s}^2$
$1\ \text{mi/h} \cdot \text{s} = 1.467\ \text{ft/s}^2$

Mass

$1\ \text{kg} = 10^3\ \text{g} = 0.0685\ \text{slug}$
$1\ \text{g} = 6.85 \times 10^{-5}\ \text{slug}$
$1\ \text{slug} = 14.59\ \text{kg}$
$1\ \text{u} = 1.661 \times 10^{-27}\ \text{kg}$
1 kg has a weight of 2.205 lb when $g = 9.80\ \text{m/s}^2$

Force

$1\ \text{N} = 10^5\ \text{dyn} = 0.2248\ \text{lb}$
$1\ \text{lb} = 4.448\ \text{N} = 4.448 \times 10^5\ \text{dyn}$

Pressure

$1\ \text{Pa} = 1\ \text{N/m}^2 = 1.450 \times 10^{-4}\ \text{lb/in.}^2 = 0.0209\ \text{lb/ft}^2$
$1\ \text{bar} = 10^5\ \text{Pa}$
$1\ \text{lb/in.}^2 = 6895\ \text{Pa}$
$1\ \text{lb/ft}^2 = 47.88\ \text{Pa}$
$1\ \text{atm} = 1.013 \times 10^5\ \text{Pa} = 1.013\ \text{bar}$
$\quad = 14.7\ \text{lb/in.}^2 = 2117\ \text{lb/ft}^2$
$1\ \text{mm Hg} = 1\ \text{torr} = 133.3\ \text{Pa}$

Energy

$1\ \text{J} = 10^7\ \text{ergs} = 0.239\ \text{cal}$
$1\ \text{cal} = 4.186\ \text{J}$ (based on 15° calorie)
$1\ \text{ft} \cdot \text{lb} = 1.356\ \text{J}$
$1\ \text{Btu} = 1055\ \text{J} = 252\ \text{cal} = 778\ \text{ft} \cdot \text{lb}$
$1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}$
$1\ \text{kWh} = 3.600 \times 10^6\ \text{J}$

Mass–Energy Equivalence

$1\ \text{kg} \leftrightarrow 8.988 \times 10^{16}\ \text{J}$
$1\ \text{u} \leftrightarrow 931.5\ \text{MeV}$
$1\ \text{eV} \leftrightarrow 1.074 \times 10^{-9}\ \text{u}$

Power

$1\ \text{W} = 1\ \text{J/s}$
$1\ \text{hp} = 746\ \text{W} = 550\ \text{ft} \cdot \text{lb/s}$
$1\ \text{Btu/h} = 0.293\ \text{W}$

APPENDIX F

NUMERICAL CONSTANTS

Fundamental Physical Constants*

Name	Symbol	Value
Speed of light in vacuum	c	2.99792458×10^{8} m/s
Magnitude of charge of electron	e	$1.602176565(35) \times 10^{-19}$ C
Gravitational constant	G	$6.67384(80) \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$
Planck's constant	h	$6.62606957(29) \times 10^{-34}\ \mathrm{J \cdot s}$
Boltzmann constant	k	$1.3806488(13) \times 10^{-23}$ J/K
Avogadro's number	N_A	$6.02214129(27) \times 10^{23}$ molecules/mol
Gas constant	R	$8.3144621(75)\ \mathrm{J/mol \cdot K}$
Mass of electron	m_e	$9.10938291(40) \times 10^{-31}$ kg
Mass of proton	m_p	$1.672621777(74) \times 10^{-27}$ kg
Mass of neutron	m_n	$1.674927351(74) \times 10^{-27}$ kg
Magnetic constant	μ_0	$4\pi \times 10^{-7}\ \mathrm{Wb/A \cdot m}$
Electric constant	$\epsilon_0 = 1/\mu_0 c^2$	$8.854187817\ldots \times 10^{-12}\ \mathrm{C^2/N \cdot m^2}$
	$1/4\pi\epsilon_0$	$8.987551787\ldots \times 10^{9}\ \mathrm{N \cdot m^2/C^2}$

Other Useful Constants*

Mechanical equivalent of heat		4.186 J/cal (15° calorie)
Standard atmospheric pressure	1 atm	1.01325×10^{5} Pa
Absolute zero	0 K	−273.15°C
Electron volt	1 eV	$1.602176565(35) \times 10^{-19}$ J
Atomic mass unit	1 u	$1.660538921(73) \times 10^{-27}$ kg
Electron rest energy	$m_e c^2$	0.510998928(11) MeV
Volume of ideal gas (0°C and 1 atm)		22.413968(20) liter/mol
Acceleration due to gravity (standard)	g	$9.80665\ \mathrm{m/s^2}$

*Source: National Institute of Standards and Technology (**http://physics.nist.gov/cuu**). Numbers in parentheses show the uncertainty in the final digits of the main number; for example, the number 1.6454(21) means 1.6454 ± 0.0021. Values shown without uncertainties are exact.

Astronomical Data†

Body	Mass (kg)	Radius (m)	Orbit radius (m)	Orbital period
Sun	1.99×10^{30}	6.96×10^{8}	—	—
Moon	7.35×10^{22}	1.74×10^{6}	3.84×10^{8}	27.3 d
Mercury	3.30×10^{23}	2.44×10^{6}	5.79×10^{10}	88.0 d
Venus	4.87×10^{24}	6.05×10^{6}	1.08×10^{11}	224.7 d
Earth	5.97×10^{24}	6.37×10^{6}	1.50×10^{11}	365.3 d
Mars	6.42×10^{23}	3.39×10^{6}	2.28×10^{11}	687.0 d
Jupiter	1.90×10^{27}	6.99×10^{7}	7.78×10^{11}	11.86 y
Saturn	5.68×10^{26}	5.82×10^{7}	1.43×10^{12}	29.45 y
Uranus	8.68×10^{25}	2.54×10^{7}	2.87×10^{12}	84.02 y
Neptune	1.02×10^{26}	2.46×10^{7}	4.50×10^{12}	164.8 y
Pluto‡	1.31×10^{22}	1.15×10^{6}	5.91×10^{12}	247.9 y

†Source: NASA (**http://solarsystem.nasa.gov/planets/**). For each body, "radius" is its average radius and "orbit radius" is its average distance from the sun or (for the moon) from the earth.

‡In August 2006, the International Astronomical Union reclassified Pluto and similar small objects that orbit the sun as "dwarf planets."

Prefixes for Powers of 10

Power of ten	Prefix	Abbreviation	Pronunciation
10^{-24}	yocto-	y	*yoc*-toe
10^{-21}	zepto-	z	*zep*-toe
10^{-18}	atto-	a	*at*-toe
10^{-15}	femto-	f	*fem*-toe
10^{-12}	pico-	p	*pee*-koe
10^{-9}	nano-	n	*nan*-oe
10^{-6}	micro-	μ	*my*-crow
10^{-3}	milli-	m	*mil*-i
10^{-2}	centi-	c	*cen*-ti
10^{3}	kilo-	k	*kil*-oe
10^{6}	mega-	M	*meg*-a
10^{9}	giga-	G	*jig*-a or *gig*-a
10^{12}	tera-	T	*ter*-a
10^{15}	peta-	P	*pet*-a
10^{18}	exa-	E	*ex*-a
10^{21}	zetta-	Z	*zet*-a
10^{24}	yotta-	Y	*yot*-a

Examples:

1 femtometer $= 1$ fm $= 10^{-15}$ m

1 picosecond $= 1$ ps $= 10^{-12}$ s

1 nanocoulomb $= 1$ nC $= 10^{-9}$ C

1 microkelvin $= 1$ μK $= 10^{-6}$ K

1 millivolt $= 1$ mV $= 10^{-3}$ V

1 kilopascal $= 1$ kPa $= 10^{3}$ Pa

1 megawatt $= 1$ MW $= 10^{6}$ W

1 gigahertz $= 1$ GHz $= 10^{9}$ Hz

ANSWERS TO ODD-NUMBERED PROBLEMS

Chapter 1

1.1 a) 1.61 km b) 3.28×10^3 ft
1.3 1.02 ns
1.5 5.36 L
1.7 31.7 y
1.9 a) 23.4 km/L b) 1.4 tanks
1.11 9.0 cm
1.13 4.2×10^{-12} cm^3, 1.3×10^{-5} mm^2
1.15 0.45%
1.17 a) no b) no c) no d) no e) no
1.19 $\approx 4 \times 10^8$
1.21 $\approx$\$70 million
1.23 2×10^5
1.25 7.8 km, 38° north of east
1.27 $A_x = 0$, $A_y = -8.00$ m, $B_x = 7.50$ m, $B_y = 13.0$ cm, $C_x = -10.9$ cm, $C_y = -5.07$ m, $D_x = -7.99$ m, $D_y = 6.02$ m
1.29 a) –6.00 m b) 11.3 m
1.31 a) 9.01 m, 33.7° b) 9.01 m, 33.7° c) 22.3 m, 250° d) 22.3 m, 70.3°
1.33 2.81 km, 38.5° north of west
1.35 a) 2.48 cm, 18.4° b) 4.09 cm, 83.7° c) 4.09 cm, 264°
1.37 $\vec{A} = -(8.00\text{ m})\hat{\jmath}$, $\vec{B} = (7.50\text{ m})\hat{\imath} + (+13.0\text{ m})\hat{\jmath}$, $\vec{C} = (-10.9\text{ m})\hat{\imath} + (-5.07\text{ m})\hat{\jmath}$, $\vec{D} = (-7.99\text{ m})\hat{\imath} + (6.02\text{ m})\hat{\jmath}$
1.39 a) $\vec{A} = (1.23\text{ m})\hat{\imath} + (3.38\text{ m})\hat{\jmath}$, $\vec{B} = (-2.08\text{ m})\hat{\imath} + (-1.20\text{ m})\hat{\jmath}$ b) $\vec{C} = (12.0\text{ m})\hat{\imath} + (14.9\text{ m})\hat{\jmath}$ c) 19.2 m, 51.2°
1.41 a) $A = 5.38$, $B = 4.36$ b) $-5.00\hat{\imath} + 2.00\hat{\jmath} + 7.00\hat{k}$ c) 8.83, yes
1.43 a) -104 m^2 b) -148 m^2 c) 40.6 m^2
1.45 a) 165° b) 28° c) 90°
1.47 a) $(-63.9\text{ m}^2)\hat{k}$ b) $(63.9\text{ m}^2)\hat{k}$
1.49 a) 5.51 g/cm^3 b) 1.1×10^6 g/cm^3 c) 4.7×10^{14} g/cm^3
1.51 a) 1.64×10^4 km b) $2.57r_E$
1.53 a) 2200 g b) 2.1 m
1.55 a) (2.8 ± 0.3) cm^3 b) 170 ± 20
1.57 $\approx 6 \times 10^{27}$
1.59 179 N, 358 N, 45.8° east of north, or 393 N, 786 N, 45.8° south of east
1.61 144 m, 41° south of west
1.63 7.55 N
1.65 60.9 km, 33.0° south of west
1.67 28.8 m, 11.4° north of east
1.69 71.9 m, 64.1° north of west
1.71 160 N, 13° below horizontal
1.73 a) 818 m, 15.8° west of south
1.75 18.6° east of south, 29.6 m
1.77 28.2 m
1.79 124°
1.81 156 m^2
1.83 28.0 m
1.85 $C_x = -8.0$, $C_y = -6.1$
1.87 D, F, B, C, A, E
1.89 b) (i) 0.9857 AU (ii) 1.3820 AU (iii) 1.695 AU c) 54.6°
1.91 a) 76.2 ly b) 129°
1.93 choice (a)

Chapter 2

2.1 25.0 m
2.3 55 min
2.5 a) 0.312 m/s b) 1.56 m/s
2.7 a) 12.0 m/s b) (i) 0 (ii) 15.0 m/s (iii) 12.0 m/s c) 13.3 m/s
2.9 a) 2.33 m/s, 2.33 m/s b) 2.33 m/s, 0.33 m/s
2.11 6.7 m/s, 6.7 m/s, 0, -40.0 m/s, -40.0 m/s, -40.0 m/s, 0
2.13 a) no b) (i) 12.8 m/s^2 (ii) 3.50 m/s^2 (iii) 0.718 m/s^2
2.15 a) 2.00 cm/s, 50.0 cm, -0.125 cm/s^2 b) 16.0 s c) 32.0 s d) 6.20 s, 1.23 cm/s; 25.8 s, -1.23 cm/s; 36.4 s, -2.55 cm/s
2.17 a) 0.500 m/s^2 b) 0, 1.00 m/s^2
2.19 a) 8.33 m/s b) 1.11 m/s^2
2.21 a) 675 m/s^2 b) 0.0667 s
2.23 1.70 m
2.25 0.38 m
2.27 a) 3.1×10^6 m/s^2 $= 3.2 \times 10^5\, g$ b) 1.6 ms c) no
2.29 a) (i) 5.59 m/s^2 (ii) 7.74 m/s^2 b) (i) 179 m (ii) 1.28×10^4 m
2.31 a) 0, 6.3 m/s^2, -11.2 m/s^2 b) 100 m, 230 m, 320 m
2.33 2.69 m/s
2.35 a) 2.94 m/s b) 0.600 s
2.37 1.67 s
2.39 a) 33.5 m b) 15.8 m/s
2.41 a) $t = \sqrt{2d/g}$ b) 0.190 s
2.43 a) 646 m b) 16.4 s, 112 m/s
2.45 a) 249 m/s^2 b) 25.4 c) 101 m d) no (if a is constant)
2.47 0.0868 m/s^2
2.49 37.6 m/s
2.51 a) 467 m b) 110 m/s
2.53 a) $x = (0.25\text{ m/s}^3)t^3 - (0.010\text{ m/s}^4)t^4$, $v_x = (0.75\text{ m/s}^3)t^2 - (0.040\text{ m/s}^4)t^3$ b) 39.1 m/s
2.55 a) 10.0 m b) (i) 8.33 m/s (ii) 9.09 m/s (iii) 9.52 m/s
2.57 250 km
2.59 a) 197 m/s b) 169 m/s
2.61 a) 92.0 m b) 92.0 m
2.63 67 m
2.65 a) 7.56 s b) 37.2 m c) 25.7 m/s (car), 15.9 m/s (truck)
2.67 a) 15.9 s b) 393 m c) 29.5 m/s
2.69 a) -4.00 m/s b) 12.0 m/s
2.71 a) $2.64H$ b) $2.64T$
2.73 a) 6.69 m/s b) 4.49 m c) 1.42 s
2.75 a) 3.3 s b) $9H$
2.77 6.75 s
2.79 a) 380 m b) 184 m
2.81 a) 0.625 m/s^3 b) 107 m
2.83 a) car A b) 2.27 s, 5.73 s c) 1.00 s, 4.33 s d) 2.67 s
2.85 a) 0.0510 s^2/m b) lower than c) no
2.87 4.8
2.89 a) 8.3 m/s b) (i) 0.411 m (ii) 1.15 km c) 9.8 m/s d) 4.9 m/s
2.91 choice (b)

Chapter 3

3.1 a) 1.4 m/s, -1.3 m/s b) 1.9 m/s, 317°
3.3 a) 7.1 cm/s, 45° b) 5.0 cm/s, 90°; 7.1 cm/s, 45°; 11 cm/s, 27°
3.5 b) -8.67 m/s^2, -2.33m/s^2 c) 8.98 m/s^2, 195°
3.7 b) $\vec{v} = \alpha\hat{\imath} - 2\beta t\hat{\jmath}$, $\vec{a} = -2\beta\hat{\jmath}$ c) 5.4 m/s, 297°; 2.4 m/s^2, 270° d) speeding up and turning right
3.9 a) 1.13 m b) 0.528 m c) $v_x = 1.10$ m/s, $v_y = -4.70$ m/s, 4.83 m/s, 76.8° below the horizontal
3.11 2.57 m
3.13 a) 24.1 m/s b) 31.0 m/s
3.15 1.28 m/s^2
3.17 a) 0.683 s, 2.99 s b) 24.0 m/s, 11.3 m/s; 24.0 m/s, -11.3 m/s c) 30.0 m/s, 36.9° below the horizontal
3.19 a) 1.5 m b) -0.89 m/s
3.21 a) 13.6 m b) 34.6 m/s c) 103 m
3.23 a) 0.034 m/s^2 $= 0.0034g$ b) 1.4 h
3.25 120 m/s^2, 270 mph
3.27 a) 2.57 m/s^2 upward b) 2.57 m/s^2 downward c) 14.7 s
3.29 a) 32.9 m/s b) 27.7 m/s^2 c) 35.5 rpm
3.31 a) 14 s b) 70 s
3.33 0.36 m/s, 52.5° south of west
3.35 a) 4.7 m/s, 25° south of east b) 120 s c) 240 m
3.37 a) 24° west of south b) 5.5 h
3.39 a) $A = 0$, $B = 2.00$ m/s^2, $C = 50.0$ m, $D = 0.500$ m/s^3 b) $\vec{a} = (4.00\text{ m/s}^2)\hat{\imath}$, $\vec{v} = 0$ c) $v_x = 40.0$ m/s, $v_y = 150$ m/s, 155 m/s d) $\vec{r} = (200\text{ m})\hat{\imath} + (550\text{ m})\hat{\jmath}$
3.41 $2b/3c$
3.43 a) 128 m b) 315 m
3.45 31 m/s
3.47 274 m
3.49 795 m
3.51 33.7 m
3.53 a) 42.8 m/s b) 42.0 m
3.55 a) 16.6 m/s b) 10.9 m/s, 40.5° below the horizontal
3.57 a) 1.50 m/s b) 4.66 m
3.59 a) 6.91 m c) no
3.61 a) 4.25 m/s b) 10.6 m
3.63 a) 17.8 m/s b) in the river, 28.4 m horizontally from his launch point
3.65 a) 49.5 m/s b) 50 m
3.67 a) 81.6 m b) 245 m c) in the cart
3.69 a) 13.3 m/s b) 3.8 m
3.71 a) 44.7 km/h, 26.6° west of south b) 10.5° north of west
3.73 7.39 m/s, 12.4° north of east
3.75 3.01 m/s, 33.7° north of east
3.77 a) graph R^2 versus h b) 16.4 m/s c) 23.8 m
3.79 70.5°
3.81 5.15 s
3.83 choice (b)
3.85 choice (c)

Chapter 4

4.1 494 N, 31.8°
4.3 3.15 N
4.5 a) –8.10 N, 3.00 N b) 8.64 N
4.7 46.7 N, opposite to the motion of the skater
4.9 21.8 kg
4.11 a) 3.12 m, 3.12 m/s b) 21.9 m, 6.24 m/s
4.13 a) 45.0 N, between 2.0 s and 4.0 s b) between 2.0 s and 4.0 s c) 0 s, 6.0 s
4.15 a) $A = 100$ N, $B = 12.5$ N/s^2 b) (i) 21.6 N, 2.70 m/s^2 (ii) 134 N, 16.8 m/s^2 c) 26.6 m/s^2
4.17 2940 N
4.19 a) 4.49 kg b) 4.49 kg, 8.13 N
4.21 825 N, blocks
4.23 50 N
4.25 b) yes
4.27 a) yes b) no
4.29 b) 142 N
4.31 2.58 s
4.33 a) 17 N, 90° clockwise from the $+x$-axis b) 840 N
4.35 a) 4.85 m/s b) 16.2 m/s^2 upward c) 1470 N upward (on him), 2360 N downward (on ground)
4.37 a) 153 N
4.39 a) 2.50 m/s^2 b) 10.0 N c) to the right, $F > T$ d) 25.0 N
4.41 a) 4.4 m b) 300 m/s c) (i) 2.7×10^4 N (ii) 9.0×10^3 N
4.43 b) 0.049 N c) $410mg$
4.45 a) 0.603 m/s^2, upward b) 1.26 m/s^2, downward
4.47 a) 7.79 m/s b) 50.6 m/s^2 upward c) $F_{\text{ground}} - mg$ upward, 4530 N upward, $6.16mg$
4.49 a) 4.34 kg b) 5.30 kg
4.51 7.78 m
4.53 a) largest: Ferrari; smallest: Alpha Romeo and Honda Civic b) largest: Ferrari; smallest: Volvo c) 7.5 kN, smaller d) zero
4.55 b) 26 kg, 8.3 m/s^2
4.57 choice (d)
4.59 choice (a)

Chapter 5

5.1 a) 25.0 N b) 50.0 N
5.3 a) 990 N, 735 N b) 926 N
5.5 48°
5.7 a) $T_A = 0.732w$, $T_B = 0.897w$, $T_C = w$ b) $T_A = 2.73w$, $T_B = 3.35w$, $T_C = w$
5.9 a) 574 N b) 607 N
5.11 a) 1.10×10^8 N b) $5w$ c) 8.4 s
5.13 a) 4610 m/s^2 = $470g$ b) 9.70×10^5 N = $471w$ c) 0.0187 s
5.15 b) 2.96 m/s^2 c) 191 N; greater than; less than
5.17 b) 3.75 m/s^2 c) 2.48 kg d) $T <$ weight of the hanging block
5.19 a) 0.832 m/s^2 b) 17.3 s
5.21 a) 3.4 m/s c) $2.2w$
5.23 a) 14.0 m b) 18.0 m/s
5.25 50°
5.27 a) 33 N b) 3.1 m
5.29 a) μ_s: 0.710; μ_k: 0.472 b) 258 N c) (i) 51.8 N (ii) 4.97 m/s^2
5.31 a) 18.3 m/s^2 b) 2.29 m/s^2
5.33 a) 57.1 N b) 146 N up the ramp
5.35 a) 52.5 m b) 16.0 m/s
5.37 a) $\mu_k(m_A + m_B)g$ b) $\mu_k m_A g$
5.39 a) 0.218 m/s b) 11.7 N
5.41 a) $\dfrac{\mu_k mg}{\cos\theta - \mu_k \sin\theta}$ b) $1/\tan\theta$
5.43 b) 8.2 m/s
5.45 a) 61.8 N b) 30.4 N
5.47 3.66 s
5.49 a) 21.0°, no b) 11,800 N (car), 23,600 N (truck)
5.51 6200 N (horizontal cable), 1410 N (upper cable)
5.53 a) 1.5 rev/min b) 0.92 rev/min
5.55 a) 38.3 m/s = 138 km/h b) 3580 N
5.57 2.42 m/s
5.59 a) 1.73 m/s^2 c) 0.0115 N upward d) 0.0098 N
5.61 a) rope making 60° angle b) 6400 N
5.63 $T_B = 4960$ N, $T_C = 1200$ N
5.65 a) 470 N b) 163 N
5.67 762 N
5.69 a) (i) −3.80 m/s (ii) 24.6 m/s b) 4.36 m c) 2.45 s
5.71 a) 11.4 N b) 2.57 kg
5.73 12.3 m/s
5.75 1.78 m/s
5.77 a) $m_1(\sin\alpha + \mu_k \cos\alpha)$ b) $m_1(\sin\alpha - \mu_k \cos\alpha)$ c) $m_1(\sin\alpha - \mu_s \cos\alpha) \le m_2 \le m_1(\sin\alpha + \mu_s \cos\alpha)$
5.79 a) 1.44 N b) 1.80 N
5.81 920 N
5.83 a) 88.0 N northward b) 78 N southward
5.85 a) 294 N (18.0-cm wire), 152 N, 152 N b) 40.0 N
5.87 3.0 N
5.89 a) 12.9 kg b) $T_{AB} = 47.2$ N, $T_{BC} = 101$ N
5.91 $a_1 = \dfrac{2m_2 g}{4m_1 + m_2}$, $a_2 = \dfrac{m_2 g}{4m_1 + m_2}$
5.93 1.46 m above the floor
5.95 g/μ_s
5.97 b) 0.452
5.99 0.34
5.101 b) 8.8 N c) 31.0 N d) 1.54 m/s^2
5.103 $v = (2mg/k)\left[\frac{1}{2} + e^{-(k/m)t}\right]$
5.105 b) 0.28 c) no
5.107 a) 81.1° b) no c) The bead rides at the bottom of the hoop.
5.109 a) 0.371 b) 0.290 c) yes, same slope, less-negative intercept
5.111 a) 5/8 in. b) 23.9 kN c) 3.57 kN, smaller d) larger; accurate
5.113 $F = (M + m)g\tan\alpha$
5.115 $\cos^2\beta$
5.117 choice (b)

Chapter 6

6.1 a) 3.60 J b) −0.900 J c) 0 d) 0 e) 2.70 J
6.3 a) 74 N b) 333 J c) −330 J d) 0, 0 e) 0
6.5 a) −1750 J b) no
6.7 a) (i) 9.00 J (ii) −9.00 J b) (i) 0 (ii) 9.00 J (iii) −9.00 J (iv) 0 c) zero for each block
6.9 a) (i) 0 (ii) 0 b) (i) 0 (ii) −25.1 J
6.11 a) 374 J b) −333 J c) 0 d) 41 J e) 352 J
6.13 −572 J
6.15 a) 120 J b) –108 J c) 24.3 J
6.17 a) 36,000 J b) 4
6.19 a) 1.0×10^{16} J b) 2.4 times
6.21 a) 43.2 m/s b) 101 m/s
6.23 $\sqrt{2gh(1 + \mu_k/\tan\alpha)}$
6.25 48.0 N
6.27 a) 4.48 m/s b) 3.61 m/s
6.29 a) 4.96 m/s b) 1.43 m/s^2; 4.96 m/s, same
6.31 a) $v_0^2/2\mu_k g$ b) (i) $\frac{1}{2}$ (ii) 4 (iii) 2
6.33 a) 40.0 N/m b) 0.456 N
6.35 b) 13.1 cm (bottom), 14.1 cm (middle), 15.2 cm (top)
6.37 a) 2.83 m/s b) 3.46 m/s
6.39 8.5 cm
6.41 a) 1.76 b) 0.666 m/s
6.43 a) 4.0 J b) 0 c) −1.0 J d) 3.0 J e) −1.0 J
6.45 a) 2.83 m/s b) 2.40 m/s
6.47 a) 0.0565 m b) no, 0.57 J
6.49 8.17 m/s
6.51 360,000 J; 100 m/s
6.53 $(3.9 \times 10^{13})P$
6.55 745 W ≈ 1 hp
6.57 a) 84.6/min b) 22.7/min
6.59 29.6 kW
6.61 0.20 W
6.63 a) 608 J b) −395 J c) 0 d) −189 J e) 24 J f) 1.5 m/s
6.65 a) 5.62 J (20.0-N block), 3.38 J (12.0-N block) b) 2.58 J (20.0-N block), 1.54 J (12.0-N block)
6.67 a) 1.8 m/s = 4.0 mi/h b) 180 m/s^2 ≈ $18g$, 900 N
6.69 a) 5.11 m b) 0.304 c) 10.3 m
6.71 a) 0.074 N b) 4.7 N c) 0.22 J
6.73 6.3×10^4 N/m
6.75 1.1 m
6.77 a) 2.39 m/s b) 9.42 m/s, away from the wall
6.79 a) 0.600 m b) 1.50 m/s
6.81 0.786
6.83 1.3 m
6.85 a) 1.10×10^5 J b) 1.30×10^5 J c) 3.99 kW
6.87 3.6 h
6.89 a) 1.26×10^5 J b) 1.46 W
6.91 b) $v^2 = -\dfrac{k}{m}d^2 + 2d\left[\dfrac{k}{m}(0.400\text{ m}) - \mu_k g\right]$ c) 1.29 m/s, 0.204 m d) 12.0 N/m, 0.800
6.93 a) $Mv^2/6$ b) 6.1 m/s c) 3.9 m/s d) 0.40 J, 0.60 J
6.95 choice (a)
6.97 choice (d)

Chapter 7

7.1 a) 6.6×10^5 J b) -7.7×10^5 J
7.3 a) 610 N b) (i) 0 (ii) 550 J
7.5 a) 24.0 m/s b) 24.0 m/s c) part (b)
7.7 a) 2.0 m/s b) $9.8/10^{-7}$ J, 2.0 J/kg c) 200 m, 63 m/s d) 5.9 J/kg e) in its tensed legs
7.9 a) (i) 0 (ii) 0.98 J b) 2.8 m/s c) Only gravity is constant. d) 5.1 N
7.11 −5400 J
7.13 a) 660 J b) –118 J c) 353 J d) 190 J e) 3.16 m/s^2, 6.16 m/s, 190 J
7.15 a) 52.0 J b) 3.25 J
7.17 a) (i) $4U_0$ (ii) $U_0/4$ b) (i) $x_0\sqrt{2}$ (ii) $x_0/\sqrt{2}$
7.19 a) 5.48 cm b) 3.92 cm
7.21 a) 6.32 cm b) 12 cm
7.23 a) 3.03 m/s, as it leaves the spring b) 95.9 m/s^2, when the spring has its maximum compression
7.25 a) 4.46×10^5 N/m b) 0.128 m
7.27 a) −5.4 J b) −5.4 J c) −10.8 J d) nonconservative

7.29 a) 8.16 m/s b) 766 J
7.31 1.29 N, $+x$-direction
7.33 130 m/s^2, 132° counterclockwise from the $+x$-axis
7.35 a) $F(r) = (12a/r^{13}) - (6b/r^7)$
b) $(2a/b)^{1/6}$, yes c) $b^2/4a$
d) $a = 6.67 \times 10^{-138}$ J · m^{12}, $b = 6.41 \times 10^{-78}$ J · m^6
7.37 a) zero (gravel), 637 N (box) b) 2.99 m/s
7.39 0.41
7.41 a) 16.0 m/s b) 11,500 N
7.43 a) 20.0 m along the rough bottom b) −78.4 J
7.45 a) 22.2 m/s b) 16.4 m c) no
7.47 0.602 m
7.49 15.5 m/s
7.51 4.4 m/s
7.53 a) 7.00 m/s b) 8.82 N
7.55 48.2°
7.57 a) 0.392 b) −0.83 J
7.59 a) $U(x) = \frac{1}{2}\alpha x^2 + \frac{1}{3}\beta x^3$ b) 7.85 m/s
7.61 a) $\alpha/(x + x_0)$ b) 3.27 m/s
7.63 7.01 m/s
7.65 a) 0.747 m/s b) 0.931 m/s
7.67 a) 0.480 m/s b) 0.566 m/s
7.69 a) 3.87 m/s b) 0.10 m
7.71 0.456 N
7.73 119 J
7.75 a) −50.6 J b) –67.5 J c) nonconservative
7.77 a) 57.0 m b) 16.5 m
c) negative work done by air resistance
7.79 a) yes b) 0.14 J d) −1.0 m, 0, 1.0 m
e) positive: -1.5 m $< x < -1.0$ m and $0 < x < 1.0$ m; negative: -1.0 m $< x < 0$ and 1.0 m $< x <$ 1.5 m f) −0.55 m, 0.12 J
7.81 choice (c)
7.83 choice (b)

Chapter 8

8.1 a) 1.20×10^5 kg · m/s
b) (i) 60.0 m/s (ii) 26.8 m/s
8.3 a) −30 kg · m/s, −55 kg · m/s
b) 0, 52 kg · m/s c) 0, −3.0 kg · m/s
8.5 a) 22.5 kg · m/s, to the left b) 838 J
8.7 562 N, not significant
8.9 a) 10.8 m/s, to the right
b) 0.750 m/s, to the left
8.11 a) 500 N/s^2 b) 5810 N · s c) 2.70 m/s
8.13 a) 2.50 N · s, in the direction of the force
b) (i) 6.25 m/s, to the right (ii) 3.75 m/s, to the right
8.15 0.593 kg · m/s
8.17 0.87 kg · m/s, in the same direction as the bullet is traveling
8.19 a) 6.79 m/s b) 55.2 J
8.21 a) 0.790 m/s b) −0.0023 J
8.23 1.97 m/s
8.25 a) 0.0559 m/s b) 0.0313 m/s
8.27 a) 7.20 m/s, 38.0° from Rebecca's original direction b) −680 J
8.29 a) 4.3 m/s c) 4.3 m/s
8.31 a) A: 29.3 m/s; B: 20.7 m/s b) 19.6%
8.33 a) 0.846 m/s b) 2.10 J
8.35 a) -1.4×10^{-6} km/h, no
b) -6.7×10^{-8} km/h, no
8.37 5.9 m/s, 58° north of east
8.39 5.46 m/s, 36.0° south of east
8.41 19.5 m/s (car), 21.9 m/s (truck)
8.43 a) 2.93 cm b) 866 J c) 1.73 J
8.45 13.6 N
8.47 a) 3.00 J; 0.500 m/s for both
b) A: −1.00 m/s; B: 1.00 m/s
8.49 a) $v_1/3$ b) $K_1/9$ c) 10
8.51 (0.0444 m, 0.0556 m)
8.53 2520 km
8.55 0.700 m to the right and 0.700 m upward
8.57 0.73 m/s
8.59 $F_x = -(1.50\text{ N/s})t$, $F_y = 0.25$ N, $F_z = 0$
8.61 a) 0.053 kg b) 5.19 N
8.63 a) 7.2×10^{-66} b) 0.223
8.65 a) −1.14 N · s, 0.330 N · s
b) 0.04 m/s, 1.8 m/s
8.67 a) 5.21 J, −0.0833 m/s
b) −2.17 m/s (A), 0.333 m/s (B)
8.69 a) 1.75 m/s, 0.260 m/s b) −0.092 J
8.71 0.946 m
8.73 1.8 m
8.75 a) $a_A = 162$ m/s^2, $a_B = 54.0$ m/s^2
b) $v_A = 5.23$ m/s, $v_B = 1.74$ m/s
8.77 12 m/s (SUV), 21 m/s (sedan)
8.79 a) 2.60 m/s b) 325 m/s
8.81 a) 5.3 m/s b) 5.7 m
8.83 53.7°
8.85 a) 0.0781 b) 248 J c) 0.441 J
8.87 a) 9.35 m/s b) 3.29 m/s
8.89 1.61×10^{-22} kg · m/s, to the left
8.91 1.33 m
8.93 0.400 m/s
8.95 250 J
8.97 a) 71.6 m/s (0.28-kg piece), 14.3 m/s (1.40-kg piece) b) 347 m
8.99 a) yes b) no, decreases by 4800 J
8.101 a) maximum: C, minimum: B
b) 69 N/m c) 0.12 m
8.103 a) $g/3$ b) 14.7 m c) 29.4 g
8.105 0, $4a/3\pi$
8.107 choice (b)
8.109 choice (b)

Chapter 9

9.1 a) 0.600 rad, 34.4° b) 6.27 cm c) 1.05 m
9.3 a) rad/s, rad/s^3 b) (i) 0 (ii) 15.0 rad/s^2
c) 9.50 rad
9.5 a) $\omega_z = \gamma + 3\beta t^2$ b) 0.400 rad/s
c) 1.30 rad/s, 0.700 rad/s
9.7 a) $\pi/4$ rad, 2.00 rad/s, −0.139 rad/s^3
b) 0 c) 19.5 rad, 9.36 rad/s
9.9 a) 2.00 rad/s b) 4.38 rad
9.11 a) 24.0 s b) 68.8 rev
9.13 3.00 rad/s
9.15 a) 300 rpm b) 75.0 s, 312 rev
9.17 9.00 rev
9.19 a) 1.99×10^{-7} rad/s
b) 7.27×10^{-5} rad/s c) 2.98×10^4 m/s
d) 463 m/s e) 0.0337 m/s^2, 0
9.21 a) 15.1 m/s^2 b) 15.1 m/s^2
9.23 a) 0.180 m/s^2, 0, 0.180 m/s^2
b) 0.180 m/s^2, 0.377 m/s^2, 0.418 m/s^2
c) 0.180 m/s^2, 0.754 m/s^2, 0.775 m/s^2
9.25 0.107 m, no
9.27 a) 0.831 m/s b) 109 m/s^2
9.29 a) (i) 0.469 kg · m^2 (ii) 0.117 kg · m^2 (iii) 0
b) (i) 0.0433 kg · m^2 (ii) 0.0722 kg · m^2
c) (i) 0.0288 kg · m^2 (ii) 0.0144 kg · m^2
9.31 a) 1.93 kg · m^2 b) 6.53 kg · m^2
c) 1.15 kg · m^2
9.33 0.193 kg · m^2
9.35 8.52 kg · m^2
9.37 6.49 m/s
9.39 0.600 kg · m^2
9.41 7.35×10^4 J
9.43 a) 0.673 m b) 45.5%
9.45 46.5 kg
9.47 a) f^5 b) 6.37×10^8 J
9.49 an axis that is parallel to a diameter and is 0.516R from the center
9.51 $M(a^2 + b^2)/3$
9.53 $\frac{1}{2}MR^2$
9.55 a) $\gamma L^2/2$ b) $ML^2/2$ c) $ML^2/6$
9.57 7.68 m
9.59 a) 0.600 m/s^3 b) $\alpha = (2.40\text{ rad/s}^3)t$
c) 3.54 s d) 17.7 rad
9.61 13.8 rad/s^2
9.63 a) 1.70 m/s b) 94.2 rad/s
9.65 2.99 cm
9.67 a) 7.36 m b) 327 m/s^2
9.69 4.65 kg · m^2
9.71 a) −0.882 J b) 5.42 rad/s c) 5.42 m/s
d) 5.42 m/s compared to 4.43 m/s
9.73 1.46 m/s
9.75 $\sqrt{\dfrac{2gd(m_B - \mu_k m_A)}{m_A + m_B + I/R^2}}$
9.77 a) 2.25×10^{-3} kg · m^2 b) 3.40 m/s
c) 4.95 m/s
9.79 13.9 m
9.81 a) 1.05 rad/s b) 5.0 J c) 78.5 J d) 6.4%
9.85 a) 55.3 kg b) 0.804 kg · m^2
9.87 a) 4.0 rev, no b) 15 rad/s c) 9.5 rad/s
9.89 a) yes b) 3.15 m/s c) 0.348 kg · m^2
d) 36.4 N
9.91 a) $s(\theta) = r_0\theta + \dfrac{\beta}{2}\theta^2$
b) $\theta(t) = \dfrac{1}{\beta}\left(\sqrt{r_0^2 + 2\beta vt} - r_0\right)$
c) $\omega_z(t) = \dfrac{v}{\sqrt{r_0^2 + 2\beta vt}}$,
$\alpha_z(t) = -\dfrac{\beta v^2}{(r_0^2 + 2\beta vt)^{3/2}}$, no
d) 25.0 mm, 0.247 μm/rad, 2.13×10^4 rev
9.93 choice (d)
9.95 choice (d)

Chapter 10

10.1 a) 40.0 N · m, out of the page
b) 34.6 N · m, out of the page
c) 20.0 N · m, out of the page
d) 17.3 N · m, into the page e) 0 f) 0
10.3 2.50 N · m, out of the page
10.5 b) $-\hat{\boldsymbol{k}}$ c) $(-1.05\text{ N}\cdot\text{m})\hat{\boldsymbol{k}}$
10.7 a) 2.56 N · m
b) 4.25 N · m, perpendicular to handle
10.9 8.38 N · m
10.11 a) 14.8 rad/s^2 b) 1.52 s
10.13 a) 7.5 N (at book on table), 18.2 N (at hanging book) b) 0.16 kg · m^2
10.15 0.255 kg · m^2
10.17 a) 1.56 m/s b) 5.35 J
c) (i) 3.12 m/s to the right (ii) 0
(iii) 2.21 m/s at 45° below the horizontal
d) (i) 1.56 m/s to the right (ii) 1.56 m/s to the left (iii) 1.56 m/s downward
10.19 a) $\frac{1}{3}$ b) $\frac{2}{7}$ c) $\frac{2}{5}$ d) $\frac{5}{13}$
10.21 a) 0.613 b) no c) no slipping
10.23 14.0 m
10.25 a) 3.76 m b) 8.58 m/s
10.27 a) 67.9 rad/s b) 8.35 J
10.29 a) 0.309 rad/s b) 100 J c) 6.67 W
10.31 a) 0.704 N · m b) 157 rad c) 111 J
d) 111 J

10.33 a) 358 N · m b) 1790 N c) 83.8 m/s
10.35 a) 115 kg · m^2/s into the page
b) 125 kg · m^2/s^2 out of the page
10.37 4.71×10^{-6} kg · m^2/s
10.39 a) A: rad/s^2; B: rad/s^4
b) (i) 59.0 kg · m^2/s (ii) 56.1 N · m
10.41 4600 rad/s
10.43 1.14 rev/s
10.45 a) 1.38 rad/s b) 1080 J, 495 J
10.47 a) 0.120 rad/s b) 3.20×10^{-4} J
c) work done by the bug
10.49 a) 5.88 rad/s
10.51 a) 1.62 N b) 1800 rev/min
10.53 2.4×10^{-12} N · m
10.55 0.483
10.57 a) 16.3 rad/s^2 b) no, decrease
c) 5.70 rad/s
10.59 0.921 m/s^2, 7.68 rad/s^2, 35.5 N (at A),
21.4 N (at B)
10.61 a) 293 N b) 16.2 rad/s^2
10.63 a) 2.88 m/s^2 b) 6.13 m/s^2
10.65 270 N
10.67 $a = \dfrac{2g}{2 + (R/b)^2}, \alpha = \dfrac{2g}{2b + R^2/b},$
$T = \dfrac{2mg}{2(b/R)^2 + 1}$
10.69 a) $3H_0/5$
10.71 29.0 m/s
10.73 a) 26.0 m/s b) no change
10.75 $g/3$
10.77 1.87 m
10.79 a) $\frac{6}{19}v/L$ b) $\frac{3}{19}$
10.81 3200 J
10.83 5.41 m
10.85 a) 2.00 rad/s b) 6.58 rad/s
10.87 0.776 rad/s
10.89 a) A: solid sphere, B: solid cylinder,
C: hollow sphere, D: hollow cylinder
b) same c) D d) 0.350
10.91 a) $mv_1^2r_1^2/r^3$ b) $\dfrac{mv_1^2}{2}r_1^2\left(\dfrac{1}{r_2^2} - \dfrac{1}{r_1^2}\right)$
c) same
10.93 a) 39.2 N upward, 39.2 N upward
b) 60.0 N upward, 18.4 N upward c) 165 N
upward, 86.2 N downward d) 0.0940 rev/s
10.95 choice (c)
10.97 choice (a)

Chapter 11

11.1 29.8 cm
11.3 1.35 m
11.5 6.6 kN
11.7 a) 1000 N, 0.800 m from end where 600-N
force is applied b) 800 N, 0.75 m from end
where 600-N force is applied
11.9 a) 550 N b) 0.614 m from A
11.11 a) 1920 N b) 1140 N
11.13 a) $T = 2.60w$; $3.28w$, 37.6°
b) $T = 4.10w$; $5.39w$, 48.8°
11.15 a) 3410 N b) 3410 N, 7600 N
11.17 b) 533 N c) 600 N, 267 N; downward
11.19 220 N (left), 255 N (right), 42°
11.21 a) 0.800 m b) clockwise
c) 0.800 m, clockwise
11.23 b) 208 N
11.25 1.9 mm
11.27 2.0×10^{11} Pa
11.29 a) 3.1×10^{-3} (upper), 2.0×10^{-3} (lower)
b) 1.6 mm (upper), 1.0 mm (lower)
11.31 a) 150 atm b) 1.5 km, no
11.33 4.8×10^9 Pa, 2.1×10^{-10} Pa^{-1}
11.35 b) 6.6×10^5 N c) 1.8 mm
11.37 7.36×10^6 Pa
11.39 3.41×10^7 Pa
11.41 10.2 m/s^2
11.43 20.0 kg
11.45 a) 525 N b) 222 N, 328 N c) 1.48
11.47 a) 140 N b) 6 cm to the right
11.49 a) 409 N b) 161 N
11.51 49.9 cm
11.53 a) 370 N b) when he starts to raise his leg
c) no
11.55 a) 3 cm b) lean backward
11.57 5500 N
11.59 b) 2000 N = $2.72mg$ c) 4.4 mm
11.61 a) 4.90 m b) 60 N
11.63 a) 175 N at each hand, 200 N at each foot
b) 91 N at each hand and at each foot
11.65 a) 1150 N b) 1940 N c) 918 N d) 0.473
11.67 590 N (person above), 1370 N (person below);
person above
11.69 a) $\dfrac{T_{max}hD}{L\sqrt{h^2 + D^2}}$
b) $\dfrac{T_{max}h}{L\sqrt{h^2 + D^2}}\left(1 - \dfrac{D^2}{h^2 + D^2}\right)$, positive
11.71 a) 71.5 kg
b) 380 N, 25.2° above the horizontal
11.73 a) 375 N b) 325 N c) 512 N
11.75 a) 0.424 N (A), 1.47 N (B), 0.424 N (C)
b) 0.848 N
11.77 a) 27° to tip, 31° to slip, tips first
b) 27° to tip, 22° to slip, slips first
11.79 a) 80 N (A), 870 N (B) b) 1.92 m
11.81 a) 1.0 cm b) 0.86 cm
11.83 a) 0.70 m from A b) 0.60 m from A
11.85 a) 4.2×10^4 N b) 65 m
11.87 b) $x = 1.50\text{ m} + \dfrac{(1.30\text{ m})m_1 - (0.38\text{ m})M}{m_2}$
c) 1.59 kg d) 1.50 m
11.89 a) 391 N (4.00-m ladder), 449 N (3.00-m ladder)
b) 322 N c) 334 N d) 937 N
11.91 a) 0.66 mm b) 0.022 J c) 8.35×10^{-3} J
d) -3.04×10^{-2} J e) 3.04×10^{-2} J
11.93 choice (a)
11.95 choice (d)

Chapter 12

12.1 no (41.8 N)
12.3 7020 kg/m^3; yes
12.5 1.6
12.7 61.6 N
12.9 a) 1.86×10^6 Pa b) 184 m
12.11 0.581 m
12.13 a) 1.90×10^4 Pa
b) causes additional force on their walls
12.15 2.8 m
12.17 6.0×10^4 Pa
12.19 2.27×10^5 N
12.21 a) 636 Pa b) (i) 1170 Pa (ii) 1170 Pa
12.23 10.9
12.25 a) 2.19×10^7 N b) 2.17×10^7 N
c) 5.79×10^8 N
12.27 0.122 m
12.29 6.43×10^{-4} m^3, 2.78×10^3 kg/m^3
12.31 10.5 N
12.33 a) 116 Pa b) 921 Pa
c) 0.822 kg, 822 kg/m^3
12.35 1640 kg/m^3
12.37 9.6 m/s
12.39 a) 17.0 m/s b) 0.317 m
12.41 28.4 m/s
12.43 1.47×10^5 Pa
12.45 2.03×10^4 Pa
12.47 2.25×10^5 Pa
12.49 $1.19D$
12.51 a) $(p_0 - p)\pi D^2/4$ b) 776 N
12.53 a) 5.9×10^5 N b) 1.8×10^5 N
12.55 2.61×10^4 N · m
12.57 0.964 cm, rise
12.59 a) 1470 Pa b) 13.9 cm
12.61 a) 0.0500 m^3 b) 10.0 kg
12.63 9.8×10^6 kg, yes
12.65 a) 0.30 b) 0.70
12.67 a) 8.27×10^3 m^3 b) 83.8 kN
12.69 a) 16.5 cm b) 1.75 m
12.71 a) 5.07 m/s, 1.28 b) 32.4 min, 2.08
12.73 a) 53.9 N b) 31.0 m/s^2
12.75 a) $1 - \dfrac{\rho_B}{\rho_L}$ b) $\left(\dfrac{\rho_L - \rho_B}{\rho_L - \rho_w}\right)L$ c) 4.60 cm
12.77 a) $2\sqrt{h(H - h)}$ b) h
12.79 5.47 m
12.81 a) 0.200 m^3/s b) 6.97×10^4 Pa
12.83 $3h_1$
12.85 b) no
12.87 a) 2.5×10^{-4} m^2/Pa (slope), 16 m^2
(intercept) b) 8.2 m, 800 kg/m^3
12.89 choice (b)
12.91 choice (a)

Chapter 13

13.1 a) 2.18
13.3 a) 1.2×10^{-11} m/s^2 b) 15 days
c) no, increase
13.5 2.1×10^{-9} m/s^2, downward
13.7 a) 2.4×10^{-3} N
b) $F_{moon}/F_{earth} = 3.5 \times 10^{-6}$
13.9 a) 0.634 m from $3m$
b) (i) unstable (ii) stable
13.11 1.38×10^7 m
13.13 a) 0.37 m/s^2 b) 1700 kg/m^3
13.15 610 N, 735 N (on earth), astronaut and satellite
have same acceleration; no
13.17 a) 5030 m/s b) 60,200 m/s
13.19 9.03 m/s^2
13.21 a) 7410 m/s b) 1.71 h
13.23 7330 m/s
13.25 a) 4.1 m/s = 9.1 mph, yes b) 2.6 h
13.27 a) 82,700 m/s b) 14.5 days
13.29 a) 7.84×10^9 s = 248 y
b) 4.44×10^{12} m, 7.38×10^{12} m
13.31 2.3×10^{30} kg = $1.2M_S$
13.33 a) (i) 5.31×10^{-9} N (ii) 2.67×10^{-9} N
13.35 a) $-\dfrac{GmM}{\sqrt{x^2 + a^2}}$ b) $-GmM/x$
c) $\dfrac{GmMx}{(x^2 + a^2)^{3/2}}$, toward the ring d) GmM/x^2
e) $U = -GmM/a, F_x = 0$
13.37 a) 33.7 N b) 32.8 N
13.39 a) 4.3×10^{37} kg = $(2.1 \times 10^7)M_S$ b) no
c) 6.32×10^{10} m, yes
13.41 9.16×10^{13} N
13.43 a) 9.67×10^{-12} N, at 45° above $+x$-axis
b) 3.02×10^{-5} m/s
13.45 a) 2.00×10^{-10} N, 161° above $+x$-axis
b) $x = 0, y = 1.32$ m
13.47 b) (i) 1.49×10^{-5} m/s (50.0-kg sphere),
7.46×10^{-6} m/s (100.0-kg sphere)
(ii) 2.24×10^{-5} m/s
c) 26.6 m
13.49 a) 3.59×10^7 m

13.51 177 m/s
13.53 a) 7.36 h b) 2.47 h
13.55 1.83×10^{27} kg
13.57 22.8 m
13.59 6060 km/h
13.61 $\sqrt{v = \dfrac{2Gm_E h}{R_E(R_E + h)}}$
13.63 a) $GM^2/4R^2$
b) $v = \sqrt{GM/4R}$, $T = 4\pi\sqrt{R^3/GM}$
c) $GM^2/4R$
13.65 6.8×10^4 m/s
13.67 a) 7900 s b) 1.53
c) 8430 m/s (perigee), 5510 m/s (apogee)
d) 2420 m/s; 3250 m/s; perigee
13.69 5.38×10^9 J
13.71 9.34 m/s^2
13.73 $GmMx/(a^2 + x^2)^{3/2}$
13.75 a) $U(r) = \dfrac{Gm_E m}{2R_E^3} r^2$ b) 7.91×10^3 m/s
13.77 a) It is considerable and shows no apparent pattern.
b) Earth (5500 kg/m^3), Mercury (5400 kg/m^3), Venus (5300 kg/m^3), Mars (3900 kg/m^3), Neptune (1600 kg/m^3), Uranus (1200 kg/m^3), Jupiter (1200 kg/m^3), Saturn (530 kg/m^3)
c) no effect d) 93 m/s^2
13.79 a) opposite; opposite b) 259 days
c) 44.1°
13.81 $\dfrac{2GMm}{a^2}\left(1 - \dfrac{x}{\sqrt{a^2 + x^2}}\right)$
13.83 choice (c)

Chapter 14

14.1 a) 2.15 ms, 2930 rad/s
b) 2.00×10^4 Hz, 1.26×10^5 rad/s
c) 1.3×10^{-15} s $\le T \le 2.3 \times 10^{-15}$ s, 4.3×10^{14} Hz $\le f \le 7.5 \times 10^{14}$ Hz
d) 2.0×10^{-7} s, 3.1×10^7 rad/s
14.3 5530 rad/s, 1.14 ms
14.5 0.0625 s
14.7 a) 0.80 s b) 1.25 Hz c) 7.85 rad/s
d) 3.0 cm e) 148 N/m
14.9 a) 0.167 s b) 37.7 rad/s c) 0.0844 kg
14.11 a) 0.150 s b) 0.0750 s
14.13 a) 0.98 m b) $\pi/2$ rad
c) $x = (-0.98\text{ m})\sin[(12.2\text{ rad/s})t]$
14.15 a) -2.71 m/s^2 b) $x = (1.46\text{ cm}) \times \cos[(15.7\text{ rad/s})t + 0.715\text{ rad}]$, $v_x = (-22.9\text{ cm/s}) \times \sin[(15.7\text{ rad/s})t + 0.715\text{ rad}]$, $a_x = (-359\text{ cm/s}^2) \times \cos[(15.7\text{ rad/s})t + 0.715\text{ rad}]$
14.17 120 kg
14.19 a) 0.253 kg b) 1.21 cm c) 3.03 N
14.21 a) 1.51 s b) 26.0 N/m
c) 30.8 cm/s d) 1.92 N
e) -0.0125 m, 30.4 cm/s, 0.216 m/s^2
f) 0.324 N
14.23 a) $x = (0.0030\text{ m})\cos[(2760\text{ rad/s})t]$
b) 8.3 m/s, 2.3×10^4 m/s^2
c) $da_x/dt = (6.3 \times 10^7\text{ m/s}^3) \times \sin[(2760\text{ rad/s})t]$, 6.3×10^7 m/s^3
14.25 92.2 m/s^2
14.27 a) 0.0336 J b) 0.0150 m c) 0.669 m/s
14.29 a) 1.20 m/s b) 1.11 m/s c) 36 m/s^2
d) 13.5 m/s^2 e) 0.36 J
14.31 $3M$; $\frac{3}{4}$
14.33 0.240 m
14.35 a) 0.376 m b) 59.3 m/s^2 c) 119 N
14.37 a) 4.06 cm b) 1.21 m/s c) 29.8 rad/s
14.39 a) 0, 0, 3.92 J, 3.92 J b) 3.92 J, 0, 0, 3.92 J
c) 0.98 J, 0.98 J, 1.96 J, 3.92 J
14.41 a) 2.7×10^{-8} kg·m^2
b) 4.3×10^{-6} N·m/rad
14.43 0.0294 kg·m^2
14.45 a) 0.25 s b) 0.25 s
14.47 0.407 swing per second
14.49 10.7 m/s^2
14.51 a) 2.84 s b) 2.89 s c) 2.89 s; −2%
14.53 A: $2\pi\sqrt{\dfrac{L}{g}}$, B: $\dfrac{2\sqrt{2}}{3}\left(2\pi\sqrt{\dfrac{L}{g}}\right)$; pendulum A
14.55 0.129 kg·m^2
14.57 A: $2\pi\sqrt{\dfrac{L}{g}}$, B: $\sqrt{\dfrac{11}{10}}\left(2\pi\sqrt{\dfrac{L}{g}}\right)$, pendulum B
14.59 a) 0.30 J
14.61 a) 0.393 Hz b) 1.73 kg/s
14.63 a) $A_1/3$ b) $2A_1$
14.65 0.353 m
14.67 a) 1.34 m/s b) 1.90 m/s^2
14.69 a) 24.4 cm b) 0.221 s c) 1.19 m/s
14.71 2.00 m
14.73 $0.921\left(\dfrac{1}{2\pi}\sqrt{\dfrac{g}{L}}\right)$
14.75 a) 0.784 s b) -1.12×10^{-4} s per s; shorter
c) 0.419 s
14.77 a) 0.150 m/s b) 0.112 m/s^2 downward
c) 0.700 s d) 4.38 m
14.79 a) 2.6 m/s b) 0.21 m c) 0.49 s
14.81 1.17 s
14.83 0.421 s
14.85 0.705 Hz, 14.5°
14.87 $2\pi\sqrt{\dfrac{M}{3k}}$
14.89 a) 1.60 s b) 0.625 Hz c) 3.93 rad/s
d) 5.1 cm; 0.4 s, 1.2 s, 1.8 s
e) 79 cm/s^2; 0.4 s, 1.2 s, 1.8 s f) 4.9 kg
14.91 b) The angular amplitude increases as L decreases. c) about 53°
14.93 a) $Mv^2/6$ c) $\omega = \sqrt{3k/M}$, $M' = M/3$
14.95 choice (a)

Chapter 15

15.1 a) 0.439 m, 1.28 ms b) 0.219 m
15.3 220 m/s = 800 km/h
15.5 a) 1.7 cm to 17 m
b) 4.3×10^{14} Hz to 7.5×10^{14} Hz
c) 1.5 cm d) 6.4 cm
15.7 a) 25.0 Hz, 0.0400 s, 19.6 rad/m
b) $y(x, t) = (0.0700\text{ m}) \times \cos[(19.6\text{ m}^{-1})x + (157\text{ rad/s})t]$
c) 4.95 cm d) 0.0050 s
15.9 a) yes b) yes c) no
d) $v_y = \omega A\cos(kx + \omega t)$, $a_y = -\omega^2 A\sin(kx + \omega t)$
15.11 a) 4 mm b) 0.040 s c) 0.14 m, 3.6 m/s
d) 0.24 m, 6.0 m/s e) no
15.13 b) $+x$-direction
15.15 a) 17.5 m/s b) 0.146 m
c) both would increase by a factor of $\sqrt{2}$
15.17 0.337 kg
15.19 a) 9.53 N b) 20.8 m/s
15.21 a) 10.0 m/s b) 0.250 m
c) $y(x, t) = (3.00\text{ cm}) \times \cos[(8.00\pi\text{ rad/m})x - (80.0\pi\text{ rad/s})t]$
d) 1890 m/s^2 e) yes
15.23 4.10 mm
15.25 a) 95 km b) 0.25 $\mu W/m^2$ c) 110 kW
15.27 a) 0.050 W/m^2 b) 22 kJ
15.29 9.48×10^{27} W
15.37 a) $(1.33\text{ m})n$, $n = 0, 1, 2, \ldots$
b) $(1.33\ m)\left(n + \frac{1}{2}\right)$, $n = 0, 1, 2, \ldots$
15.39 a) 96.0 m/s b) 461 N
c) 1.13 m/s, 4.26 m/s^2
15.41 b) 2.80 cm c) 277 cm
d) 185 cm, 7.96 Hz, 0.126 s, 1470 cm/s
e) 280 cm/s
f) $y(x, t) = (5.60\text{ cm}) \times \sin[(0.0906\text{ rad/cm})x]\sin[(133\text{ rad/s})t]$
15.43 4.0 m, 2.0 m, 1.33 m
15.45 a) 45.0 cm b) no
15.47 a) 311 m/s b) 246 Hz c) 245 Hz, 1.40 m
15.49 a) 20.0 Hz, 126 rad/s, 3.49 rad/m
b) $y(x, t) = (2.50 \times 10^{-3}\text{ m}) \times \cos[(3.49\text{ rad/m})x - (126\text{ rad/s})t]$
c) $y(0, t) = (2.50 \times 10^{-3}\text{ m})\cos[(126\text{ rad/s})t]$
d) $y(1.35\text{ m}, t) = (2.50 \times 10^{-3}\text{ m}) \times \cos[(126\text{ rad/s})t - 3\pi/2\text{ rad}]$
e) 0.315 m/s f) -2.50×10^{-3} m, 0
15.51 a) $\dfrac{7L}{2}\sqrt{\dfrac{\mu_1}{F}}$ b) no
15.53 a) 62.1 m
15.55 13.7 Hz, 25.0 m
15.57 1.83 m
15.59 361 Hz (copper), 488 Hz (aluminum)
15.61 a) 18.8 cm b) 0.0169 kg
15.63 a) 7.07 cm b) 0.400 kW
15.65 $(0.800\text{ Hz})n$, $n = 1, 2, 3, \ldots$
15.67 a) 2.22 g b) 2.24×10^4 m/s^2
15.69 233 N
15.71 1780 kg/m^3
15.73 a) 148 N b) 26%
15.75 c) 47.5 Hz d) 138 g
15.77 a) 392 N b) 392 N + (7.70 N/m)x
c) 3.89 s
15.79 choice (b)

Chapter 16

16.1 a) 0.344 m b) 1.2×10^{-5} m
c) 6.9 m, 50 Hz
16.3 a) 7.78 Pa b) 77.8 Pa c) 778 Pa
16.5 a) 90 m b) 102 kHz c) 1.4 cm
d) 4.4 mm to 8.8 mm e) 6.2 MHz
16.7 90.8 m
16.9 81.4°C
16.11 0.16 s
16.13 a) 5.5×10^{-15} J b) 0.074 mm/s
16.15 15.0 cm
16.17 a) 4.14 Pa b) 0.0208 W/m^2 c) 103 dB
16.19 a) 4.4×10^{-12} W/m^2 b) 6.4 dB
c) 5.8×10^{-11} m
16.21 14.0 dB
16.23 a) 2.0×10^{-7} W/m^2 b) 6.0 m
c) 290 m d) yes, no
16.25 a) fundamental: 0.60 m; 0, 1.20 m; first overtone: 0.30 m, 0.90 m; 0, 0.60 m, 1.20 m; second overtone: 0.20 m, 0.60 m, 1.00 m; 0, 0.40 m, 0.80 m, 1.20 m
b) fundamental: 0; 1.20 m; first overtone: 0, 0.80 m; 0.40 m, 1.20 m; second overtone: 0, 0.48 m, 0.96 m; 0.24 m, 0.72 m, 1.20 m
16.27 506 Hz, 1517 Hz, 2529 Hz
16.29 a) 35.2 Hz b) 17.6 Hz
16.31 a) 614 Hz b) 1230 Hz
16.33 a) 137 Hz, 0.50 m b) 137 Hz, 2.51 m
16.35 a) 172 Hz b) 86 Hz
16.37 0.125 m
16.39 a) $(820\text{ Hz})n$, $n = 1, 2, 3, \ldots$
b) $(410\text{ Hz})(2n + 1)$, $n = 0, 1, 2, \ldots$
16.41 a) 433 Hz b) loosen
16.43 1.3 Hz

16.45 780 m/s
16.47 a) 375 Hz b) 371 Hz c) 4 Hz
16.49 a) 0.25 m/s b) 0.91 m
16.51 19.8 m/s
16.53 a) 1910 Hz b) 0.188 m
16.55 a) 7.02 m/s, toward b) 1404 Hz
16.57 a) 36.0° b) 2.94 s
16.59 a) 1.00 b) 8.00
c) 4.73×10^{-8} m = 47.3 nm
16.61 flute harmonic $3n$ resonates with string harmonic $4n$, $n = 1, 3, 5, \ldots$
16.63 a) stopped b) 7th and 9th c) 0.439 m
16.65 a) 0.026 m, 0.53 m, 1.27 m, 2.71 m, 9.01 m
b) 0.26 m, 0.86 m, 1.84 m, 4.34 m c) 86 Hz
16.67 a) 0.0823 m b) 120 Hz
16.69 b) 2.0 m/s
16.71 a) 38 Hz b) no
16.73 a) 375 m/s b) 1.39 c) 0.8 cm
16.75 d) 9.69 cm/s, 667 m/s^2
16.77 choice (b)
16.79 choice (a)
16.81 choice (b)

Chapter 17

17.1 a) −81.0°F b) 134.1°F c) 88.0°F
17.3 a) 27.2 C° b) −55.6 C°
17.5 a) −18.0 F° b) −10.0 C°
17.7 0.964 atm
17.9 a) −282°C b) no, 47,600 Pa
17.11 0.39 m
17.13 1.9014 cm; 1.8964 cm
17.15 49.4°C
17.17 1.7×10^{-5} (C°)$^{-1}$
17.19 a) 1.431 cm^2 b) 1.436 cm^2
17.21 a) 6.0 mm b) -1.0×10^8 Pa
17.23 555 kJ
17.25 23 min
17.27 240 J/kg · K
17.29 0.526 C°
17.31 45.2 C°
17.33 0.0613 C°
17.35 a) 215 J/kg · K b) water c) too small
17.37 0.114 kg
17.39 27.5°C
17.41 150°C
17.43 7.6 min
17.45 54.5 kJ, 13.0 kcal, 51.7 Btu
17.47 357 m/s
17.49 3.45 L
17.51 5.05×10^{15} kg
17.53 0.0674 kg
17.55 190 g
17.57 a) 222 K/m b) 10.7 W c) 73.3°C
17.59 a) −0.86°C b) 24 W/m^2
17.61 4.0×10^{-3} W/m · C°
17.63 105.5°C
17.65 a) 21 kW b) 6.4 kW
17.67 15 W
17.69 2.1 cm^2
17.71 35.0°C
17.73 a) 35.1°M b) 39.6 C°
17.75 69.4°C
17.77 23.0 cm (first rod), 7.0 cm (second rod)
17.79 b) 1.9×10^8 Pa
17.81 a) 87°C b) −80°C
17.83 460 s
17.85 a) 83.6 J b) 1.86 J/mol · K
c) 5.60 J/mol · K
17.87 a) 4.20×10^7 J b) 10.7 C° c) 30.0 C°
17.89 a) 0.60 kg b) 0.80 bottle/h
17.91 3.4×10^5 J/kg
17.93 a) no b) 0.0°C, 0.156 kg
17.95 a) 86.1°C
b) no ice, 0.130 kg liquid water, no steam
17.97 a) 100°C
b) 0.0214 kg steam, 0.219 kg liquid water
17.99 a) 93.9 W b) 1.35
17.101 2.9
17.103 a) 59.8°C b) 42.7°C c) 8.40 W
17.105 c) 170 h d) 1.5×10^{10} s ≈ 500 y, no
17.107 5.82 g
17.109 a) 1.04 kW b) 87.1 W c) 1.13 kW
d) 28 g e) 1.1 bottles
17.111 a) 3.00×10^4 J/kg b) 1.00×10^3 J/kg · K (liquid), 1.33×10^3 J/kg · K (solid)
17.113 A: 216 W/m · K, B: 130 W/m · K
17.115 a) $H = \dfrac{(T_2 - T_1)2\pi kL}{\ln(b/a)}$
b) $T = T_2 - \dfrac{(T_2 - T_1)\ln(r/a)}{\ln(b/a)}$
d) 73°C e) 49 W
17.117 choice (a)
17.119 choice (a)

Chapter 18

18.1 a) 0.122 mol b) 14,700 Pa, 0.145 atm
18.3 0.100 atm
18.5 a) 0.0136 kg/m^3, 67.6 kg/m^3, 5.39 kg/m^3
b) $0.011\rho_E$, $56\rho_E$, $4.5\rho_E$
18.7 503°C
18.9 19.7 kPa
18.11 0.159 L
18.13 $0.0508V$
18.15 a) 70.2°C b) yes
18.17 850 m
18.19 a) 6.95×10^{-16} kg b) 2.32×10^{-13} kg/m^3
18.21 55.6 mol, 3.35×10^{25} molecules
18.23 a) 2.20×10^6 molecules
b) 2.44×10^{19} molecules
18.25 6.4×10^{-6} m
18.27 a) 5.83×10^7 J b) 242 m/s
18.29 (d) must be true; the others could be true.
18.31 a) 1.93×10^6 m/s, no b) 7.3×10^{10} K
18.33 a) 6.21×10^{-21} J b) 2.34×10^5 m^2/s^2
c) 484 m/s d) 2.57×10^{-23} kg · m/s
e) 1.24×10^{-19} N f) 1.24×10^{-17} Pa
g) 8.17×10^{21} molecules
h) 2.45×10^{22} molecules
18.35 3800°C
18.37 a) 1870 J b) 1120 J
18.39 a) 741 J/kg · K, $c_w = 5.65c_{N_2}$
b) 5.65 kg; 4850 L
18.41 a) 337 m/s b) 380 m/s c) 412 m/s
18.43 a) 610 Pa b) 22.12 MPa
18.45 18.0 cm^3, $V_{20°C} = 0.32V_{cp}$
18.47 a) 11.8 kPa b) 0.566 L
18.49 272°C
18.51 0.195 kg
18.53 a) −179°C b) 1.2×10^{26} molecules/m^3
c) $\rho_T = 4.8\rho_e$
18.55 1.92 atm
18.57 a) 30.7 cylinders b) 8420 N c) 7800 N
18.59 a) 26.2 m/s b) 16.1 m/s, 5.44 m/s c) 1.74 m
18.61 $\approx 5 \times 10^{27}$ atoms
18.63 a) A b) B c) 4250°C d) B
18.65 a) 6.00×10^3 Pa b) 32.8 m/s
18.67 a) 4.65×10^{-26} kg b) 6.11×10^{-21} J
c) 2.04×10^{24} molecules d) 12.5 kJ
18.69 b) r_2 c) $r_1 = \dfrac{R_0}{2^{1/6}}$, $r_2 = R_0$, $2^{-1/6}$ d) U_0
18.71 a) $2R$ = 16.6 J/mol · K b) less than
18.73 b) 1.40×10^5 K (N_2), 1.01×10^4 K (H_2)
c) 6370 K (N_2), 459 K (H_2)
18.75 $3kT/m$, same
18.77 b) $0.0421N$ c) $(2.94 \times 10^{-21})N$
d) $0.0297N$, $(2.08 \times 10^{-21})N$
e) $0.0595N$, $(4.15 \times 10^{-21})N$
18.79 a) $p_0 + \dfrac{mg}{\pi r^2}$ b) $-\left(\dfrac{y}{h}\right)(p_0\pi r^2 + mg)$
c) $\dfrac{1}{2\pi}\sqrt{\dfrac{g}{h}\left(1 + \dfrac{p_0\pi r^2}{mg}\right)}$, no
18.81 a) 42.6% b) 3 km c) 1 km
18.83 a) 4.5×10^{11} m
b) 703 m/s, 6.4×10^8 s (≈20 y)
c) 1.4×10^{-14} Pa
d) 650 m/s, $v_H > v_{esc}$, evaporate
f) 2×10^5 K, $>3T_{sun}$, no
18.85 choice (a)
18.87 choice (c)

Chapter 19

19.1 b) 1330 J
19.3 b) −6180 J
19.5 a) 1.04 atm
19.7 a) $(p_1 - p_2)(V_2 - V_1)$
b) negative of work done in reverse direction
19.9 a) 34.7 kJ b) 80.4 kJ c) no
19.11 a) 278 K, at a b) 0; 162 J c) 53 J
19.13 a) T_a = 535 K, T_b = 9350 K, T_c = 15,000 K
b) 21 kJ done by gas c) 36 kJ
19.15 a) 0 b) $T_b = 2T_a$ c) $U_b = U_a$ + 700 J
19.17 b) 208 J c) on the piston d) 712 J
e) 920 J f) 208 J
19.19 a) 948 K b) 900 K
19.21 $\frac{2}{5}$
19.23 a) 747 J b) 1.30
19.25 a) −605 J b) 0 c) yes, 605 J, liberate
19.27 a) 476 kPa b) −10.6 kJ c) 1.59, heated
19.29 b) 314 J c) −314 J
19.31 11.6°C
19.33 a) increase b) 4.8 kJ
19.35 a) 0.681 mol b) 0.0333 m^3
c) 2.23 kJ d) 0
19.37 a) 45.0 J b) liberate, 65.0 J c) 23.0 J, 22.0 J
19.39 a) the same b) 4.0 kJ, absorb c) 8.0 kJ
19.41 b) −2460 J
19.43 a) 0.80 L b) 305 K, 1220 K, 1220 K
c) ab: 76 J, into the gas
ca: −107 J, out of the gas
bc: 56 J, into the gas
d) ab: 76 J, increased
bc: 0, no change
ca: −76 J, decreased
19.45 a) 837°C b) 11.5 kJ c) 40.3 kJ d) 42.4 kJ
19.47 b) 6.00 L, 2.5×10^4 Pa, 75.0 K
c) 95 J d) heat it at constant volume
19.49 b) 11.9 C°
19.51 a) 0.168 m b) 196°C c) 70.1 kJ
19.53 a) Q = 450 J, ΔU = 0
b) Q = 0, ΔU = −450 J
c) Q = 1125 J, ΔU = 675 J
19.55 a) W = 738 J, Q = 2590 J, ΔU = 1850 J
b) W = 0, Q = −1850 J, ΔU = −1850 J
c) ΔU = 0
19.57 a) W = −187 J, Q = −654 J, ΔU = −467 J
b) W = 113 J, Q = 0, ΔU = −113 J
c) W = 0, Q = 580 J, ΔU = 580 J
19.59 a) a: adiabatic, b: isochoric, c: isobaric
b) 28.0°C c) a: −30.0 J, a: 0, a: 20.0 J
d) a
e) a: decrease, b: stay the same, c: increase

19.61 b) -300 J, out of the gas
19.63 choice (c)
19.65 choice (d)

Chapter 20

20.1 a) 6500 J b) 34%
20.3 a) 23% b) 12,400 J c) 0.350 g d) 222 kW = 298 hp
20.5 a) 12.3 atm b) 5470 J, *ca* c) 3723 J, *bc* d) 1747 J e) 31.9%
20.7 a) 58% b) 1.4%
20.9 a) 14.8 kJ b) 45.8 kJ
20.11 1.2 h
20.13 a) 215 J b) 378 K c) 39.0%
20.15 a) 38 kJ b) 590°C
20.17 a) 492 J b) 212 W c) 5.4
20.19 44.5 hp
20.21 a) 429 J/K b) -393 J/K c) 36 J/K
20.23 a) irreversible b) 1250 J/K
20.25 -6.31 J/K
20.27 a) 6.05×10^3 J/K b) about five times greater for vaporization
20.29 a) no b) 18.3 J/K c) 18.3 J/K
20.31 10.0 J/K
20.33 a) 121 J b) 3800 cycles
20.35 a) 90.2 J b) 320 J c) 45°C d) 0 e) 263 g
20.37 -5.8 J/K, decrease
20.39 b) absorbed: *bc*; rejected: *ab* and *ca* c) $T_a = T_b = 241$ K, $T_c = 481$ K d) 610 J, 610 J e) 8.7%
20.41 a) 21.0 kJ (enters), 16.6 kJ (leaves) b) 4.4 kJ, 21% c) $e = 0.31e_{\max}$
20.43 a) 7.0% b) 3.0 MW; 2.8 MW c) 6×10^5 kg/h $= 6 \times 10^5$ L/h
20.45 a) 1: 2.00 atm, 4.00 L; 2: 2.00 atm, 6.00 L; 3: 1.11 atm, 6.00 L; 4: 1.67 atm, 4.00 L b) $1 \rightarrow 2$: 1422 J, 405 J; $2 \rightarrow 3$: -1355 J, 0; $3 \rightarrow 4$: -274 J, -274 J; $4 \rightarrow 1$: 339 J, 0 c) 131 J d) 7.44%; $e = 0.168e_C$
20.47 $1 - T_C/T_H$, same
20.49 a) 122 J, -78 J b) 5.10×10^{-4} m^3 c) *b*: 2.32 MPa, 4.81×10^{-5} m^3, 771 K; *c*: 4.01 MPa, 4.81×10^{-5} m^3, 1332 K; *d*: 0.147 MPa, 5.10×10^{-4} m^3, 518 K d) 61.1%, 77.5%
20.51 6.23
20.55 a) A: 28.9%, B: 38.3%, C: 53.8%, D: 24.4% b) C c) B > D > A
20.57 a) 4.83% b) 4.83% c) 6.25% d) $e = \dfrac{0.80T_d - 200}{12T_d - 2700}$, 6.67%
20.59 choice (b)
20.61 choice (d)

Chapter 21

21.1 a) 2.00×10^{10} b) 8.59×10^{-13}
21.3 3.4×10^{18} m/s^2 (proton), 6.3×10^{21} m/s^2 (electron)
21.5 1.3 nC
21.7 3.7 km
21.9 a) 0.742 μC b) 0.371 μC, 1.48 μC
21.11 a) 2.21×10^4 m/s^2
21.13 +0.750 nC
21.15 1.8×10^{-4} N, in the $+x$-direction
21.17 $x = -0.144$ m
21.19 2.58 μN, in the $-y$-direction
21.21 a) 8.80×10^{-9} N, attractive b) 8.22×10^{-8} N; about 10 times larger than the bonding force
21.23 a) 4.40×10^{-16} N b) 2.63×10^{11} m/s^2 c) 2.63×10^5 m/s
21.25 a) 3.30×10^6 N/C, to the left b) 14.2 ns c) 1.80×10^3 N/C, to the right
21.27 a) -21.9 μC b) 1.02×10^{-7} N/C
21.29 a) 364 N/C b) no; 2.73 μm, downward
21.31 1.79×10^6 m/s
21.33 a) $-\hat{\jmath}$ b) $\dfrac{\sqrt{2}}{2}\hat{\imath} + \dfrac{\sqrt{2}}{2}\hat{\jmath}$ c) $-0.39\hat{\imath} + 0.92\hat{\jmath}$
21.35 a) 633 km/s b) 15.9 km/s
21.37 a) 0 b) for $|x| < a$: $E_x = -\dfrac{q}{\pi\epsilon_0}\dfrac{ax}{(x^2 - a^2)^2}$; for $x > a$: $E_x = \dfrac{q}{2\pi\epsilon_0}\dfrac{x^2 + a^2}{(x^2 - a^2)^2}$; for $x < -a$: $E_x = -\dfrac{q}{2\pi\epsilon_0}\dfrac{x^2 + a^2}{(x^2 - a^2)^2}$
21.39 a) (i) 574 N/C, $+x$-direction (ii) 268 N/C, $-x$-direction (iii) 404 N/C, $-x$-direction b) (i) 9.20×10^{-17} N, $-x$-direction (ii) 4.30×10^{-17} N, $+x$-direction (iii) 6.48×10^{-17} N, $+x$-direction
21.41 1.04×10^7 N/C, toward the -2.00-μC charge
21.43 a) 8740 N/C, to the right b) 6540 N/C, to the right c) 1.40×10^{-15} N, to the right
21.45 1.73×10^{-8} N, toward the point midway between the electrons
21.47 a) $E_x = E_y = E = 0$ b) $E_x = 2660$ N/C, $E_y = 0$, $E = 2660$ N/C, $+x$-direction c) $E_x = 129$ N/C, $E_y = -510$ N/C, $E = 526$ N/C, 284° counterclockwise from the $+x$-axis d) $E_x = 0$, $E_y = 1380$ N/C, $E = 1380$ N/C, $+y$-direction
21.49 a) 1.14×10^5 N/C, toward the center of the disk b) 8.92×10^4 N/C, toward the center of the disk c) 1.46×10^5 N/C, toward the charge
21.51 a) $(7.0\ \text{N/C})\hat{\imath}$ b) $(1.75 \times 10^{-5}\ \text{N})\hat{\imath}$
21.53 a) 1.4×10^{-11} C·m, from q_1 toward q_2 b) 860 N/C
21.55 a) $\vec{p}$ aligned in the same or the opposite direction as $\vec{E}$ b) stable: $\vec{p}$ aligned in the same direction as $\vec{E}$; unstable: $\vec{p}$ aligned in the opposite direction
21.57 a) 1680 N, from the $+5.00$-μC charge toward the -5.00-μC charge b) 22.3 N·m, clockwise
21.59 b) $\dfrac{Q^2}{8\pi\epsilon_0 L^2}(1 + 2\sqrt{2})$, away from the center of the square
21.61 a) 8.63×10^{-5} N, -5.52×10^{-5} N b) 1.02×10^{-4} N, 32.6° below the $+x$-axis
21.63 b) 2.80 μC c) 39.5°
21.65 3.41×10^4 N/C, to the left
21.67 between the charges, 0.24 m from the 0.500-nC charge
21.69 at $x = d/3$, $q = -4Q/9$
21.71 a) $\dfrac{6q^2}{4\pi\epsilon_0 L^2}$, away from the vacant corner b) $\dfrac{3q^2}{4\pi\epsilon_0 L^2}\left(\sqrt{2} + \dfrac{1}{2}\right)$, toward the center of the square
21.73 a) 6.0×10^{23} b) 4.1×10^{-31} N (gravitational), 510 kN (electric) c) yes (electric), no (gravitational)
21.75 2190 km/s
21.77 a) $\dfrac{mv_0^2 \sin^2\alpha}{2eE}$ b) $\dfrac{mv_0^2 \sin 2\alpha}{eE}$ d) $h_{\max}$: 0.418 m, d: 2.89 m
21.79 a) $E_x = \dfrac{Q}{4\pi\epsilon_0 a}\left(\dfrac{1}{x - a} - \dfrac{1}{x}\right)$, $E_y = 0$ b) $\dfrac{qQ}{4\pi\epsilon_0 a}\left(\dfrac{1}{r} - \dfrac{1}{r + a}\right)\hat{\imath}$
21.81 a) -7.99 nC b) -24.0 nC
21.83 a) 1.56 N/C, $+x$-direction c) smaller d) 4.7%
21.85 $E_x = E_y = \dfrac{Q}{2\pi^2\epsilon_0 a^2}$
21.87 a) 6.25×10^4 N/C, 225° counterclockwise from an axis pointing to the right at point P b) 1.00×10^{-14} N, opposite the electric field direction
21.89 a) 1.15×10^6 N/C, to the left b) 1.58×10^5 N/C, to the left c) 1.58×10^5 N/C, to the right
21.91 a) $\pi(R_2^2 - R_1^2)\sigma$ b) $\dfrac{\sigma}{2\epsilon_0}\left(\dfrac{1}{\sqrt{(R_1/x)^2 + 1}} - \dfrac{1}{\sqrt{(R_2/x)^2 + 1}}\right)\dfrac{|x|}{x}\hat{\imath}$ c) $\dfrac{\sigma}{2\epsilon_0}\left(\dfrac{1}{R_1} - \dfrac{1}{R_2}\right)x\hat{\imath}$; $x \ll R_1$ d) $\dfrac{1}{2\pi}\sqrt{\dfrac{q\sigma}{2\epsilon_0 m}\left(\dfrac{1}{R_1} - \dfrac{1}{R_2}\right)}$
21.93 a) $q_1 = 8.00$ μC, $q_2 = 3.00$ μC b) 7.49 N, in the $-x$-direction c) $x = 0.248$ m
21.95 b) $q_1 < 0$, $q_2 > 0$ c) 0.843 μC d) 56.2 N
21.97 a) $\dfrac{Q}{2\pi\epsilon_0 L}\left(\dfrac{1}{2x + a} - \dfrac{1}{2L + 2x + a}\right)$
21.99 choice (c)
21.101 choice (b)

Chapter 22

22.1 a) 1.8 N·m^2/C b) no c) (i) 0° (ii) 90°
22.3 a) 3.53×10^5 N·m^2/C b) 3.13 μC
22.5 $\pi r^2 E$
22.7 0.977 N·m^2/C, inward
22.9 a) 0 b) 1.22×10^8 N/C, radially inward c) 3.64×10^7 N/C, radially inward
22.11 a) 1.17×10^5 N·m^2/C b) no change
22.13 0.0810 N
22.15 1.35×10^{10}
22.17 a) 6.47×10^5 N/C, $+y$-direction b) 7.2×10^4 N/C, $-y$-direction
22.19 a) 5.73 μC/m^2 b) 6.47×10^5 N/C c) -5.65×10^4 N·m^2/C
22.21 a) 0.260 μC/m^3 b) 1960 N/C
22.23 a) 6.56×10^{-21} J b) 1.20×10^5 m/s
22.25 a) 6.00 nC b) -1.00 nC
22.27 σ/ϵ_0 (between), 0 (outside)
22.29 a) $2\pi R\sigma$ b) $\dfrac{\sigma R}{\epsilon_0 r}$ c) $\dfrac{\lambda}{2\pi\epsilon_0 r}$
22.31 1.16 km/s
22.33 10.2°
22.35 a) 750 N·m^2/C b) 0 c) 577 N/C, $+x$-direction d) within and outside
22.37 a) -0.598 nC b) within and outside
22.39 a) $\dfrac{\lambda}{2\pi\epsilon_0 r}$, radially outward b) $\dfrac{\lambda}{2\pi\epsilon_0 r}$, radially outward d) $-\lambda$ (inner), $+\lambda$ (outer)
22.41 a) $\dfrac{\rho r}{2\epsilon_0}$ b) $\dfrac{\lambda}{2\pi\epsilon_0 r}$ c) They are equal.

22.43 a) $r < R$: 0; $R < r < 2R$: $\frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$, radially outward; $r > 2R$: $\frac{1}{4\pi\epsilon_0}\frac{2Q}{r^2}$, radially outward

22.45 a) (i) 0 (ii) 0 (iii) $\frac{q}{2\pi\epsilon_0 r^2}$, radially outward (iv) 0 (v) $\frac{3q}{2\pi\epsilon_0 r^2}$, radially outward
b) (i) 0 (ii) $+2q$ (iii) $-2q$ (iv) $+6q$

22.47 a) $\frac{qQ}{4\pi\epsilon_0 r^2}$, toward the center of the shell b) 0

22.49 a) $\frac{\alpha}{2\epsilon_0}\left(1 - \frac{a^2}{r^2}\right)$ b) $q = +2\pi\alpha a^2$, $E = \frac{\alpha}{2\epsilon_0}$

22.51 $R/2$

22.53 c) $\frac{Qr}{4\pi\epsilon_0 R^3}\left(4 - \frac{3r}{R}\right)$ e) $2R/3$, $\frac{Q}{3\pi\epsilon_0 R^2}$

22.55 b) $|x| > d$ (outside the slab): $\frac{\rho_0 d}{3\epsilon_0}\frac{x}{|x|}\hat{\imath}$; $|x| < d$ (inside the slab): $\frac{\rho_0 x^3}{3\epsilon_0 d^2}\hat{\imath}$

22.57 b) $\frac{\rho\vec{b}}{3\epsilon_0}$

22.59 a) uniform line of charge: A; uniformly charged sphere: B b) $\lambda = 1.50 \times 10^{-7}$ C/m, $\rho = 2.81 \times 10^{-3}$ C/m^3

22.61 (i) 377 N/C (ii) 653 N/C (iii) 274 N/C (iv) 0

22.63 choice (a)

22.65 choice (b)

Chapter 23

23.1 −0.356 J

23.3 3.46×10^{-13} J = 2.16 MeV

23.5 a) 12.5 m/s b) 0.323 m

23.7 1.94×10^{-5} N

23.9 a) 13.6 km/s; very long after release b) 2.45×10^{17} m/s^2; just after release

23.11 $-q/2$

23.13 7.42 m/s, faster

23.15 a) 0 b) 0.750 mJ c) −2.06 mJ

23.17 a) 0 b) −175 kV c) −0.875 J

23.19 a) −737 V b) −704 V c) 8.2×10^{-8} J

23.21 b) $V = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{|x|} - \frac{2}{|x-a|}\right)$ c) $x = -a, a/3$ e) $V = \frac{q}{4\pi\epsilon_0 x}$

23.23 a) b b) 800 V/m c) −48.0 μJ

23.25 a) (i) 180 V (ii) −270 V (iii) −450 V b) 719 V, inner shell

23.27 a) oscillatory b) 1.67×10^7 m/s

23.29 150 m/s

23.31 a) 94.9 nC/m b) no, less c) 0

23.33 a) 78.2 kV b) 0

23.35 0.474 J

23.37 a) 8.00 kV/m b) 19.2 μN c) 0.864 μJ d) −0.864 μJ

23.39 −760 V

23.41 a) (i) $V = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r_a} - \frac{1}{r_b}\right)$ (ii) $V = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{r} - \frac{1}{r_b}\right)$ (iii) $V = 0$
d) 0 e) $E = \frac{q - Q}{4\pi\epsilon_0 r^2}$

23.43 a) $E_x = -Ay + 2Bx$, $E_y = -Ax - C$, $E_z = 0$ b) $x = -C/A$, $y = -2BC/A^2$, any value of z

23.45 a) 0.762 nC

23.47 a) −0.360 μJ b) $x = 0.074$ m

23.49 4.2×10^6 V

23.51 a) 4.79 MeV, 7.66×10^{-13} J b) 5.17×10^{-14} m

23.53 a) −21.5 μJ b) −2.83 kV c) 35.4 kV/m

23.55 a) 7.85×10^4 V/m$^{4/3}$ b) $E_x(x) = -(1.05 \times 10^5\ \text{V/m}^{4/3})x^{1/3}$ c) 3.13×10^{-15} N, toward the anode

23.57 a) $-\frac{1.46q^2}{\pi\epsilon_0 d}$

23.59 47.8 V

23.61 a) (i) $V = (\lambda/2\pi\epsilon_0)\ln(b/a)$ (ii) $V = (\lambda/2\pi\epsilon_0)\ln(b/r)$ (iii) $V = 0$ d) $(\lambda/2\pi\epsilon_0)\ln(b/a)$

23.63 a) 1.76×10^{-16} N, downward b) 1.93×10^{14} m/s^2, downward c) 0.822 cm d) 15.3° e) 3.29 cm

23.65 a) 97.1 kV/m b) 30.3 pC

23.67 $\frac{3}{5}\left(\frac{Q^2}{4\pi\epsilon_0 R}\right)$

23.69 360 kV

23.71 a) 50.0 g: 216 m/s^2, 12.7 m/s; 150.0 g: 7.20 m/s^2, 4.24 m/s

23.73 a) $\frac{Q}{4\pi\epsilon_0 a}\ln\left(\frac{x+a}{x}\right)$ b) $\frac{Q}{4\pi\epsilon_0 a}\ln\left(\frac{a + \sqrt{a^2 + y^2}}{y}\right)$ c) (a): $\frac{Q}{4\pi\epsilon_0 x}$, (b): $\frac{Q}{4\pi\epsilon_0 y}$

23.75 a) $\frac{1}{3}$ b) 3

23.77 a) 7580 km/s b) 7260 km/s c) 2.3×10^9 K; 6.4×10^9 K

23.79 a) $A = -6.0$ V/m^2, $B = -4.0$ V/m^3, $C = -2.0$ V/m^6, $D = 10$ V, $l = 2.0$, $m = 3.0$, $n = 6.0$
b) (0, 0, 0): 10.0 V, 0; (0.50 m, 0.50 m, 0.50 m): 8.0 V, 6.7 V/m; (1.00 m, 1.00 m, 1.00 m): −2.0 V, 21 V/m

23.81 c) 4.79×10^{-19} C (drop 1), 1.59×10^{-19} C (drop 2), 8.09×10^{-19} C (drop 3), 3.23×10^{-19} C (drop 4)
d) 3 (drop 1), 5 (drop 3), 2 (drop 4)
e) 1.60×10^{-19} C (drop 1), 1.59×10^{-19} C (drop 2), 1.62×10^{-19} C (drops 3 and 4); 1.61×10^{-19} C

23.83 1.01×10^{-12} m, 1.11×10^{-13} m, 2.54×10^{-14} m

23.85 choice (b)

Chapter 24

24.1 a) 10.0 kV b) 22.6 cm^2 c) 8.00 pF

24.3 a) 604 V b) 90.8 cm^2 c) 1840 kV/m d) 16.3 μC/m^2

24.5 a) 120 μC b) 60 μC c) 480 μC

24.7 a) 1.05 mm b) 84.0 V

24.9 a) 4.35 pF b) 2.30 V

24.11 a) 15.0 pF b) 3.09 cm c) 31.2 kN/C

24.13 a) 17.5 cm b) 25.5 nC

24.15 a) series b) 5000

24.17 a) $Q_1 = Q_2 = 22.4$ μC, $Q_3 = 44.8$ μC, $Q_4 = 67.2$ μC b) $V_1 = V_2 = 5.6$ V, $V_3 = 11.2$ V, $V_4 = 16.8$ V c) 11.2 V

24.19 a) $Q_1 = 156$ μC, $Q_2 = 260$ μC b) $V_1 = V_2 = 52.0$ V

24.21 a) 19.3 nF b) 482 nC c) 162 nC d) 25 V

24.23 0.0283 J/m^3

24.25 a) 90.0 pF b) 0.0152 m^3 c) 4.5 kV d) 1.80 μJ

24.27 a) $U_p = 4U_s$ b) $Q_p = 2Q_s$ c) $E_p = 2E_s$

24.29 a) 24.2 μC b) $Q_{35} = 7.7$ μC, $Q_{75} = 16.5$ μC c) 2.66 mJ d) $U_{35} = 0.85$ mJ, $U_{75} = 1.81$ mJ e) 220 V

24.31 a) 1.60 nC b) 8.05

24.33 a) 3.60 mJ (before), 13.5 mJ (after) b) 9.9 mJ, increase

24.35 a) 0.620 μC/m^2 b) 1.28

24.37 0.0135 m^2

24.39 a) 6.3 μC b) 6.3 μC c) none

24.41 a) 10.1 V b) 2.25

24.43 a) $\frac{Q}{\epsilon_0 AK}$ b) $\frac{Qd}{\epsilon_0 AK}$ c) $K\frac{\epsilon_0 A}{d} = KC_0$

24.45 a) 421 J b) 0.054 F

24.47 a) 0.531 pF b) 0.224 mm

24.49 a) 0.0160 C b) 533 V c) 4.26 J d) 2.14 J

24.51 a) 158 μJ b) 72.1 μJ

24.53 a) 2.5 μF b) $Q_1 = 550$ μC, $Q_2 = 370$ μC, $Q_3 = Q_4 = 180$ μC, $Q_5 = 550$ μC; $V_1 = 65$ V, $V_2 = 87$ V, $V_3 = V_4 = 43$ V, $V_5 = 65$ V

24.55 $C_2 = 6.00$ μF, $C_3 = 4.50$ μF

24.57 a) 76 μC b) 1.4 mJ c) 11 V d) 1.3 mJ

24.59 a) 2.3 μF b) $Q_1 = 970$ μC, $Q_2 = 640$ μC c) 47 V

24.61 a) 3.91 b) 22.8 V

24.63 1.67 μF

24.65 0.185 μJ

24.67 b) 2.38 nF

24.69 a) $C_1 = 6.00$ μF, $C_2 = 3.00$ μF b) same charge; C_2 stores more energy c) C_1 stores more charge and energy

24.71 a) first (connected) b) 144 cm^2 c) disconnected

24.73 choice (c)

24.75 choice (a)

Chapter 25

25.1 1.0 C

25.3 a) 3.12×10^{19} b) 1.51×10^6 A/m^2 c) 0.111 mm/s d) both (b) and (c) would increase

25.5 a) 110 min b) 440 min c) $v_d \propto 1/d^2$

25.7 a) 330 C b) 41 A

25.9 9.0 μA

25.11 a) 1.06×10^{-5} Ω·m b) 0.00105 (C°)$^{-1}$

25.13 a) 0.206 mV b) 0.176 mV

25.15 a) 1.21 V/m b) 0.0145 Ω c) 0.182 V

25.17 0.125 Ω

25.19 a) 4.67×10^{-8} Ω b) 6.72×10^{-4} Ω

25.21 a) 11 A b) 3.1 V c) 0.28 Ω

25.23 a) 99.54 Ω b) 0.0158 Ω

25.25 a) 27.4 V b) 12.3 MJ

25.27 a) 0 b) 5.0 V c) 5.0 V

25.29 3.08 V, 0.067 Ω, 1.80 Ω

25.31 a) 1.41 A, clockwise b) 13.7 V c) –1.0 V

25.33 a) 0.471 A, counterclockwise b) 15.2 V

25.35 a) 144 Ω b) 240 Ω c) 100 W: 0.833 A; 60 W: 0.500 A

25.37 a) 29.8 W b) 0.248 A

25.39 a) 3.1 W b) 7.2 W c) 4.1 W

25.41 a) 300 W b) 0.90 J

25.43 a) 2.6 MJ b) 0.063 L c) 1.6 h

25.45 12.3%

25.47 a) 24.0 W b) 4.0 W c) 20.0 W

25.49 a) 1.55×10^{-12} s

25.51 a) 3.65×10^{-8} Ω·m b) 172 A c) 2.58 mm/s

25.53 $0.060\ \Omega$
25.55 a) 2.5 mA b) $21.4\ \mu$V/m c) $85.5\ \mu$V/m d) 0.180 mV
25.57 a) 80 C° b) no
25.59 a) $\dfrac{\rho h}{\pi r_1 r_2}$
25.61 a) $0.36\ \Omega$ b) 8.94 V
25.63 a) $1.0\ \mathrm{k}\Omega$ b) 100 V c) 10 W
25.65 a) \$78.90 b) \$140.27
25.67 a) $I_A\left(1+\dfrac{R_A}{r+R}\right)$ b) $0.0429\ \Omega$
25.69 a) $171\ \mu\Omega$ b) $176\ \mu$V/m c) left: $54.7\ \mu\Omega$; right: $116\ \mu\Omega$
25.71 a) 204 V b) 199 J
25.73 6.67 V
25.75 b) no c) yes d) 9.40 W e) 4.12 W
25.77 a) $R=\dfrac{\rho_0 L}{A}\left(1-\dfrac{1}{e}\right), I=\dfrac{V_0 A}{\rho_0 L\left(1-\dfrac{1}{e}\right)}$
b) $E(x)=\dfrac{V_0 e^{-x/L}}{L\left(1-\dfrac{1}{e}\right)}$
c) $V(x)=\dfrac{V_0\left(e^{-x/L}-\dfrac{1}{e}\right)}{1-\dfrac{1}{e}}$
25.79 choice (c)
25.81 choice (d)

Chapter 26

26.1 $3R/4$
26.3 22.5 W
26.5 a) 3.50 A b) 4.50 A c) 3.15 A d) 3.25 A
26.7 0.769 A
26.9 a) $8.80\ \Omega$ b) 3.18 A c) 3.18 A d) $V_{1.60}=5.09$ V, $V_{2.40}=7.63$ V, $V_{4.80}=15.3$ V e) $P_{1.60}=16.2$ W, $P_{2.40}=24.3$ W, $P_{4.80}=48.5$ W f) greatest
26.11 a) $I_1=8.00$ A, $I_3=12.0$ A b) 84.0 V
26.13 $5.00\ \Omega$; $I_{3.00}=8.00$ A, $I_{4.00}=9.00$ A, $I_{6.00}=4.00$ A, $I_{12.0}=3.00$ A
26.15 a) $I_1=1.50$ A, $I_2=I_3=I_4=0.500$ A b) $P_1=10.1$ W, $P_2=P_3=P_4=1.12$ W; bulb R_1 c) $I_1=1.33$ A, $I_2=I_3=0.667$ A d) $P_1=8.00$ W, $P_2=P_3=2.00$ W e) brighter: R_2 and R_3; less bright: R_1
26.17 18.0 V, 3.00 A
26.19 1010 s
26.21 a) 0.100 A b) $P_{400}=4.0$ W, $P_{800}=8.0$ W c) 12.0 W d) $I_{400}=0.300$ A, $I_{800}=0.150$ A e) $P_{400}=36.0$ W, $P_{800}=18.0$ W f) 54.0 W g) series: 800-Ω bulb; parallel: 400-Ω bulb h) parallel
26.23 a) $20.0\ \Omega$ b) A_2: 4.00 A; A_3: 12.0 A; A_4: 14.0 A; A_5: 8.00 A
26.25 a) 2.00 A b) $5.00\ \Omega$ c) 42.0 V d) 3.50 A
26.27 a) 8.00 A b) $\mathcal{E}_1=36.0$ V, $\mathcal{E}_2=54.0$ V c) $9.00\ \Omega$
26.29 a) 1.60 A (top), 1.40 A (middle), 0.20 A (bottom) b) 10.4 V
26.31 a) 36.4 V b) 0.500 A
26.33 a) 2.14 V, a b) 0.050 A, 0; down
26.35 a) $0.641\ \Omega$ b) $975\ \Omega$
26.37 a) 17.9 V b) 22.7 V c) 21.4%
26.39 a) $0.849\ \mu$F b) 2.89 s
26.41 a) 0 b) 245 V c) 0 d) 32.7 mA e) (a): 245 V; (b): 0; (c): 1.13 mC; (d): 0
26.43 a) 4.21 ms b) 0.125 A
26.45 $192\ \mu$C
26.47 13.6 A
26.49 a) 0.937 A b) 0.606 A
26.51 a) $165\ \mu$C b) $463\ \Omega$ c) 12.6 ms
26.53 900 W
26.55 a) 2.2 A, 4.4 V, 9.7 W b) 16.3 W; more brightly
26.57 a) +0.22 V b) 0.464 A
26.59 $I_1=0.848$ A, $I_2=2.14$ A, $I_3=0.171$ A
26.61 $I_{2.00}=5.21$ A, $I_{4.00}=1.11$ A, $I_{5.00}=6.32$ A
26.63 a) 109 V; no b) 13.5 s
26.65 a) 186 V, upper terminal positive b) 3.00 A, upward c) $20.0\ \Omega$
26.67 a) −12.0 V b) 1.71 A c) $4.21\ \Omega$
26.69 a) P_1+P_2 b) $\dfrac{P_1P_2}{P_1+P_2}$
26.71 a) 1.35 W b) 8.31 ms c) 0.337 W
26.73 a) 114 V b) 263 V c) 266 V
26.75 a) 18.0 V b) a c) 6.00 V d) both decrease by $36.0\ \mu$C
26.77 a) $V_{224}=24.8$ V, $V_{589}=65.2$ V b) $3840\ \Omega$ c) 62.6 V d) no
26.79 $1.7\ \mathrm{M}\Omega$, $3.1\ \mu$F
26.81 a) −1.23 ms (slope), $79.5\ \mu$C (y-intercept) b) $247\ \Omega$, 15.9 V c) 1.22 ms d) 11.9 V
26.85 b) 4 c) $3.2\ \mathrm{M}\Omega$, 4.0×10^{-3} d) 3.4×10^{-4} e) 0.88
26.87 choice (d)

Chapter 27

27.1 a) $(-6.68\times10^{-4}\ \mathrm{N})\hat{\boldsymbol{k}}$ b) $(6.68\times10^{-4}\ \mathrm{N})\hat{\imath}+(7.27\times10^{-4}\ \mathrm{N})\hat{\jmath}$
27.3 a) positive b) 0.0505 N
27.5 9490 km/s
27.7 a) $B_x=-0.175$ T, $B_z=-0.256$ T b) B_y c) 0; 90°
27.9 a) 1.46 T, in the xz-plane at 40° from the $+x$-axis toward the $-z$-axis b) 7.47×10^{-16} N, in the xz-plane at 50° from the $-x$-axis toward the $-z$-axis
27.11 a) 3.05 mWb b) 1.83 mWb c) 0
27.13 −0.78 mWb
27.15 a) 0.160 mT, into the page b) $0.111\ \mu$s
27.17 7.93×10^{-10} N, toward the south
27.19 a) 2.84×10^{6} m/s, negative b) yes c) same
27.21 a) 835 km/s b) 26.2 ns c) 7.27 kV
27.23 0.838 mT
27.25 a) $(1.60\times10^{-14}\ \mathrm{N})\hat{\jmath}$ b) yes c) helix; no d) 1.40 cm
27.27 a) 7900 N/C, $\hat{\imath}$ b) 7900 N/C, $\hat{\imath}$
27.29 0.0445 T, out of the page
27.31 a) 4.92 km/s b) 9.96×10^{-26} kg
27.33 2.0 cm
27.35 0.724 N, 63.4° below the current direction in the upper wire segment
27.37 a) 817 V b) 113 m/s^2
27.39 a) a b) 3.21 kg
27.41 b) $F_{cd}=1.20$ N c) 0.420 N · m
27.43 a) A_2 b) 290 rad/s^2
27.45 a) $-NIAB\hat{\imath}$, 0 b) 0, $-NIAB$ c) $+NIAB\hat{\imath}$, 0 d) 0, $+NIAB$
27.47 a) 1.13 A b) 3.69 A c) 98.2 V d) 362 W
27.49 a) 4.7 mm/s b) $+4.5\times10^{-3}$ V/m, $+z$-direction c) $53\ \mu$V
27.51 a) $-\dfrac{F_2}{qv_1}\hat{\jmath}$ b) $F_2/\sqrt{2}$
27.53 a) 8.3×10^{6} m/s b) 0.14 T
27.55 3.45 T, perpendicular to the coin's initial velocity
27.57 a) $-3.89\ \mu$C b) $(7.60\times10^{14}\ \mathrm{m/s^2})\hat{\imath}+(5.70\times10^{14}\ \mathrm{m/s^2})\hat{\jmath}$ c) 2.90 cm d) 2.88×10^{7} Hz e) (0.0290 m, 0, 0.874 m)
27.59 1.6 mm
27.61 $\dfrac{Mg\tan\theta}{LB}$, right to left
27.63 a) 8.46 mT b) 27.2 cm c) 2.2 cm; yes
27.65 a) ILB, to the right b) $\dfrac{v^2m}{2ILB}$ c) 1960 km
27.67 1.97 N, 68.3° clockwise from the left-hand segment
27.69 0.024 T, $+y$-direction
27.71 a) $F_{PQ}=0$; $F_{RP}=12.0$ N, into the page; $F_{QR}=12.0$ N, out of the page b) 0 c) $\tau_{PQ}=\tau_{RP}=0$; $\tau_{QR}=3.60$ N · m d) 3.60 N · m; yes e) out
27.73 $-(0.444\ \mathrm{N})\hat{\jmath}$
27.75 b) left: $(B_0LI/2)\hat{\imath}$; top: $-IB_0L\hat{\jmath}$; right: $-(B_0LI/2)\hat{\imath}$; bottom: 0 c) $-IB_0L\hat{\jmath}$
27.77 a) $-IA\hat{\boldsymbol{k}}$ b) $B_x=\dfrac{3D}{IA}, B_y=\dfrac{4D}{IA}, B_z=-\dfrac{12D}{IA}$
27.79 b) 1.85×10^{-28} kg c) 1.20 kV d) 8.32×10^{5} m/s
27.81 a) 5.14 m b) $1.72\ \mu$s c) 6.08 mm d) 3.05 cm
27.83 choice (c)
27.85 choice (a)

Chapter 28

28.1 a) $-(19.2\ \mu\mathrm{T})\hat{\boldsymbol{k}}$ b) 0 c) $(19.2\ \mu\mathrm{T})\hat{\imath}$ d) $(6.79\ \mu\mathrm{T})\hat{\imath}$
28.3 a) 60.0 nT, out of the page at A and B b) $0.120\ \mu$T, out of the page c) 0
28.5 a) 0 b) $-(1.31\ \mu\mathrm{T})\hat{\boldsymbol{k}}$ c) $-(0.462\ \mu\mathrm{T})\hat{\boldsymbol{k}}$ d) $(1.31\ \mu\mathrm{T})\hat{\jmath}$
28.7 $(97.5\ \mathrm{nT})\hat{\boldsymbol{k}}$
28.9 a) $0.440\ \mu$T, out of the page b) 16.7 nT, out of the page c) 0
28.11 a) $(50.0\ \mathrm{pT})\hat{\jmath}$ b) $-(50.0\ \mathrm{pT})\hat{\imath}$ c) $-(17.7\ \mathrm{pT})(\hat{\imath}-\hat{\jmath})$ d) 0
28.13 $17.6\ \mu$T, into the page
28.15 a) 0.8 mT b) $40\ \mu$T (20 times larger)
28.17 $250\ \mu$A
28.19 a) 10.0 A b) at all points directly above the wire c) at all points directly east of the wire
28.21 a) $-(0.10\ \mu\mathrm{T})\hat{\imath}$ b) $2.19\ \mu$T, at 46.8° from the $+x$-axis to the $+z$-axis c) $(7.9\ \mu\mathrm{T})\hat{\imath}$
28.23 a) 0 b) $6.67\ \mu$T, toward the top of the page c) $7.54\ \mu$T, to the left
28.25 a) 0 b) 0 c) 0.40 mT, to the left
28.27 a) P: $41\ \mu$T, into the page; Q: $25\ \mu$T, out of the page b) P: $9.0\ \mu$T, out of the page; Q: $9.0\ \mu$T, into the page
28.29 a) $6.00\ \mu$N; repulsive b) $24.0\ \mu$N
28.31 $46\ \mu$N/m; repulsive; no
28.33 $0.38\ \mu$A
28.35 $\dfrac{\mu_0|I_1-I_2|}{4R}$; 0
28.37 a) $25.1\ \mu$T b) $503\ \mu$T; no
28.39 18.0 A, counterclockwise
28.41 a) 305 A b) -3.83×10^{-4} T · m
28.43 a) $\mu_0 I/2\pi r$ b) 0
28.45 a) 2.83 mT b) $35.0\ \mu$T; no
28.47 a) 1790 turns per meter b) 63.0 m
28.49 a) 3.72 MA b) 249 kA c) 237 A

28.51 1.11 mT
28.53 a) (i) 1.13 mT (ii) 4.68 MA/m (iii) 5.88 T
28.55 a) 1.00 μT, into the page b) $(74.9 \text{ nN})\hat{\jmath}$
28.57 a) in the plane of the wires, between them, 0.300 m from the 75.0-A wire
b) in the plane of the wires, 0.200 m from the 25.0-A wire and 0.600 m from the 75.0-A wire
28.59 a) 5.7×10^{12} m/s^2, away from the wire
b) 32.5 N/C, away from the wire c) no
28.61 a) 81 A b) 2.4×10^{-3} N/m
28.63 a) 2.00 A, out of the page
b) 2.13 μT, upward c) 2.06 μT
28.65 23.2 A
28.67 a) $\dfrac{\mu_0 N I a^2}{2} \times \left\{ \dfrac{1}{[(x+a/2)^2+a^2]^{3/2}} + \dfrac{1}{[(x-a/2)^2+a^2]^{3/2}} \right\}$
c) $\left(\dfrac{4}{5}\right)^{3/2} \dfrac{\mu_0 N I}{a}$ d) 20.2 mT e) 0, 0
28.69 a) $\dfrac{3I}{2\pi R^3}$ b) (i) $B = \dfrac{\mu_0 I r^2}{2\pi R^3}$ (ii) $B = \dfrac{\mu_0 I}{2\pi r}$
28.71 b) $B = \dfrac{\mu_0 I_0}{2\pi r}$ c) $\dfrac{I_0 r^2}{a^2}\left(2 - \dfrac{r^2}{a^2}\right)$
d) $B = \dfrac{\mu_0 I_0 r}{2\pi a^2}\left(2 - \dfrac{r^2}{a^2}\right)$
28.73 a) $B = \mu_0 I n/2$, $+x$-direction
b) $B = \mu_0 I n/2$, $-x$-direction
28.75 a) $I_0 = 2\pi b\delta(1 - e^{-a/\delta})$, 81.5 A b) $\dfrac{\mu_0 I_0}{2\pi r}$
c) $\left(\dfrac{e^{r/\delta}-1}{e^{a/\delta}-1}\right) I_0$ d) $\dfrac{\mu_0 I}{2\pi r}\left(\dfrac{e^{r/\delta}-1}{e^{a/\delta}-1}\right)$
e) $r = \delta$: 175 μT; $r = a$: 326 μT; $r = 2a$: 163 μT
28.77 a) no c) 65 A, 1.2 cm
28.79 b) $\dfrac{1}{2g}\left(\dfrac{\mu_0 Q_0{}^2}{4\pi\lambda R C d}\right)^2$
28.81 choice (b)
28.83 choice (c)

Chapter 29

29.1 a) 17.1 mV b) 28.5 mA
29.3 a) $Q = NBA/R$ c) no
29.5 a) 34 V b) counterclockwise
29.7 a) $\mu_0 i/2\pi r$, into the page b) $\dfrac{\mu_0 i}{2\pi r} L\,dr$
c) $\dfrac{\mu_0 i L}{2\pi}\ln(b/a)$ d) $\dfrac{\mu_0 L}{2\pi}\ln(b/a)\dfrac{di}{dt}$
e) 0.506 μV
29.9 a) 5.44 mV b) clockwise
29.11 a) bAv b) clockwise
c) bAv, counterclockwise
29.13 10.4 rad/s
29.15 a) counterclockwise b) clockwise
c) no induced current
29.17 a) C: counterclockwise; A: clockwise
b) toward the wire
29.19 a) a to b b) b to a c) b to a
29.21 a) clockwise b) no induced current
c) counterclockwise
29.23 13.2 mA, counterclockwise
29.25 a) 0.675 V b) b c) 2.25 V/m, b to a
d) b e) (i) 0 (ii) 0
29.27 46.2 m/s = 103 mph; no
29.29 a) 3.00 V b) b to a
c) 0.800 N, to the right d) 6.00 W for each
29.31 a) counterclockwise b) 42.4 mW
29.33 35.0 m/s, to the right
29.35 a) 0.225 A, clockwise b) 0
c) 0.225 A, counterclockwise
29.37 a) $\pi r_1{}^2 \dfrac{dB}{dt}$ b) $\dfrac{r_1}{2}\dfrac{dB}{dt}$ c) $\dfrac{R^2}{2r_2}\dfrac{dB}{dt}$ e) $\dfrac{\pi R^2}{4}\dfrac{dB}{dt}$
f) $\pi R^2 \dfrac{dB}{dt}$ g) $\pi R^2 \dfrac{dB}{dt}$
29.39 9.21 A/s
29.41 0.950 mV
29.43 a) 0.599 nC b) 6.00 mA c) 6.00 mA
29.45 a) inside: $B = 0$, $\vec{M} = -(0.103 \text{ MA/m})\hat{\imath}$;
outside: $\vec{B} = (0.130 \text{ T})\hat{\imath}$, $M = 0$
b) inside and outside: $\vec{B} = (0.260 \, T)\hat{\imath}$, $M = 0$
29.47 a) 3.7 A b) 1.33 mA c) counterclockwise
29.49 16.2 μV
29.51 a) $\dfrac{\mu_0 I a b v}{2\pi r(r+a)}$ b) clockwise
29.53 a) 17.9 mV b) a to b
29.55 $\mu_0 I W/4\pi$
29.57 a) $\dfrac{\mu_0 I v}{2\pi}\ln(1 + L/d)$ b) a c) 0
29.59 a) 0.165 V b) 0.165 V c) 0; 0.0412 V
29.61 a) B^2L^2v/R
29.63 a: $\dfrac{qr}{2}\dfrac{dB}{dt}$, to the left; b: $\dfrac{qr}{2}\dfrac{dB}{dt}$, toward the top of the page; c: 0
29.65 5.0 s
29.67 a) 0.3071 s^{-1} b) 3.69 T c) a d) 2.26 s
29.69 a) a to b b) $\dfrac{Rmg\tan\phi}{L^2B^2\cos\phi}$ c) $\dfrac{mg\tan\phi}{LB}$
d) $\dfrac{Rm^2g^2\tan^2\phi}{L^2B^2}$ e) $\dfrac{Rm^2g^2\tan^2\phi}{L^2B^2}$; same
29.71 choice (c)
29.73 choice (c)

Chapter 30

30.1 a) 0.270 V; yes b) 0.270 V
30.3 6.32 μH
30.5 a) 1.96 H b) 7.11 mWb
30.7 a) 1940 b) 800 A/s
30.9 a) 0.250 H b) 0.450 mWb
30.11 a) 4.68 mV b) a
30.13 a) 1000 b) 2.09 Ω
30.15 b) 0.111 μH
30.17 2850
30.19 a) 0.161 T b) 10.3 kJ/m^3 c) 0.129 J
d) 40.2 μH
30.21 91.7 J
30.23 a) 2.40 A/s b) 0.800 A/s c) 0.413 A
d) 0.750 A
30.25 a) 17.3 μs b) 30.7 μs
30.27 a) 0.250 A b) 0.137 A c) 32.9 V; c
d) 0.462 ms
30.29 15.3 V
30.31 a) 443 nC b) 358 nC
30.33 a) 25.0 mH b) 90.0 nC c) 0.540 μJ
d) 6.58 mA
30.35 a) 105 rad/s, 59.6 ms b) 0.720 mC
c) 4.32 mJ d) -0.542 mC
e) -0.050 A, counterclockwise
f) $U_C = 2.45$ mJ, $U_L = 1.87$ mJ
30.37 a) 7.50 μC b) 15.9 kHz c) 21.2 mJ
30.39 a) 298 rad/s b) 83.8 Ω
30.41 a) 8.76 kHz b) 1.35 ms c) 2420 Ω
30.43 a) 0.288 μH b) 14.2 μV
30.45 20 km/s; about 30 times smaller
30.47 a) $\dfrac{\mu_0 i}{2\pi r}$ b) $\dfrac{\mu_0 i^2 l}{4\pi r}\,dr$ c) $\dfrac{\mu_0 i^2 l}{4\pi}\ln(b/a)$
d) $\dfrac{\mu_0 l}{2\pi}\ln(b/a)$
30.49 a) 5.00 H b) 31.7 m; no
30.51 222 μF, 9.31 μH
30.53 a) 0.896 mJ b) 0.691 A; 0
30.55 a) 24.0 mV b) 1.55 mA c) 72.1 nJ
d) 5.20 μC, 18.0 nJ
30.57 a) 0, 20.0 V b) 0.267 A, 0 c) 0.147 A, 9.0 V
30.59 a) $A_1 = A_4 = 0.800$ A, $A_2 = A_3 = 0$;
$V_1 = 40.0$ V, $V_2 = V_3 = V_4 = V_5 = 0$
b) $A_1 = 0.480$ A, $A_2 = 0.160$ A,
$A_3 = 0.320$ A, $A_4 = 0$; $V_1 = 24.0$ V, $V_2 = 0$,
$V_3 = V_4 = V_5 = 16.0$ V c) 192 μC
30.61 a) 60.0 V b) a c) 60.0 V d) c
e) -96.0 V f) b g) 156 V h) d
30.63 a) 0; $v_{ac} = 0$, $v_{cb} = 36.0$ V
b) 0.180 A, $v_{ac} = 9.0$ V, $v_{cb} = 27.0$ V
c) $i_0 = (0.180 \text{ A})(1 - e^{-t/(0.020 \text{ s})})$,
$v_{ac} = (9.0 \text{ V})(1 - e^{-t/(0.020 \text{ s})})$,
$v_{cb} = (9.0 \text{ V})(3.00 + e^{-t/(0.020 \text{ s})})$
30.65 a) $A_1 = A_4 = 0.455$ A, $A_2 = A_3 = 0$
b) $A_1 = 0.585$ A, $A_2 = 0.320$ A,
$A_3 = 0.160$ A, $A_4 = 0.107$ A
30.67 a) $v_L = \dfrac{\epsilon}{R + R_L}(R_L + Re^{-(R+R_L)t/L})$
b) 50.0 V c) 30.0 V; 3.00 A d) 6.67 Ω
e) 40.0 mH
30.69 b) 5.0 Ω, 8.5 H c) 1.7 kJ; 2.0 kW
30.71 a) $i_1 = \dfrac{\epsilon}{R_1}(1 - e^{-R_1 t/L})$, $i_2 = \dfrac{\epsilon}{R_2}e^{-t/R_2C}$,
$q_2 = \epsilon C(1 - e^{-t/R_2C})$
b) $i_1 = 0$, $i_2 = 9.60$ mA
c) $i_1 = 1.92$ A, $i_2 = 0$; $t \gg L/R_1$ and $t \gg R_2C$
d) 1.6 ms e) 9.4 mA f) 0.22 s
30.73 choice (b)
30.75 choice (c)

Chapter 31

31.1 1.06 A
31.3 a) 31.8 V b) 0
31.5 a) 90°; lead b) 193 Hz
31.7 13.3 μF
31.9 a) 1510 Ω b) 0.239 H c) 497 Ω
d) 16.6 μF
31.11 a) $(12.5 \text{ V})\cos[(480 \text{ rad/s})t]$ b) 7.17 V
31.13 a) $i = (0.0253 \text{ A})\cos[(720 \text{ rad/s})t]$
b) 180 Ω
c) $v_L = -(4.56 \text{ V})\sin[(720 \text{ rad/s})t]$
31.15 a) 601 Ω b) 49.9 mA c) $-70.6°$; lag
d) $V_R = 9.98$ V, $V_L = 4.99$ V, $V_C = 33.3$ V
31.17 50.0 V
31.19 a) 40.0 W b) 0.167 A c) 720 Ω
31.21 b) 76.7 V
31.23 a) 45.8°, 0.697 b) 344 Ω c) 155 V
d) 48.6 W e) 48.6 W f) 0 g) 0
31.25 a) 0.302 b) 0.370 W
c) 0.370 W (resistor), 0, 0
31.27 a) 113 Hz; 15.0 mA b) 7.61 mA; lag
31.29 a) 150 V b) $V_R = 150$ V,
$V_L = V_C = 1290$ V c) 37.5 W
31.31 a) 1.00 b) 75.0 W c) 75.0 W
31.33 a) 945 rad/s b) 70.6 Ω
c) $V_L = V_C = 450$ V, $V_R = 120$ V
31.35 a) 10 b) 2.40 A c) 28.8 W d) 500 Ω
31.37 0.124 H
31.39 230 Ω
31.41 3.59×10^7 rad/s

31.43 a) inductor b) 0.133 H
31.45 a) 0.831 b) 161 W
31.47 $\dfrac{V_{\text{out}}}{V_s} = \sqrt{\dfrac{R^2 + \omega^2 L^2}{R^2 + \left(\omega L - \dfrac{1}{\omega C}\right)^2}}$
31.51 a) 102 Ω b) 0.882 A c) 270 V
31.53 a) $V_R = 48.6$ V, $V_L = 155$ V, $V_C = 243$ V, $-60.9°$
b) $V_R = 100$ V, $V_L = V_C = 400$ V, 0°
c) $V_R = 48.6$ V, $V_L = 243$ V, $V_C = 155$ V, $+60.9°$
31.55 b) 5770 rad/s c) 2.40 A d) 2.40 A
e) 0.139 A f) 0.139 A
31.57 a) $\omega = 28{,}800$ rad/s so $\phi = 60°$
b) $P_R = 0.375$ W, $P_L = P_C = 0$; 0.100 A
31.59 a) 0.750 A b) 160 Ω c) 341 Ω, 619 Ω
d) 341 Ω
31.61 a) $\dfrac{V}{R}$ b) $\dfrac{V}{R}\sqrt{\dfrac{L}{C}}$ c) $\dfrac{V}{R}\sqrt{\dfrac{L}{C}}$ d) $\frac{1}{2}L\dfrac{V^2}{R^2}$
e) $\frac{1}{2}L\dfrac{V^2}{R^2}$
31.63 a) 20.6 Ω b) 105 μF c) 699 W
31.65 20.0 Ω, 0.18 H
31.67 a) $\frac{1}{2}V_R I$ b) 0 c) 0
31.69 choice (b)
31.71 choice (d)

Chapter 32

32.1 a) 1.28 s b) 8.15×10^{13} km
32.3 13.3 nT, $+y$-direction
32.5 3.0×10^{18} Hz, 3.3×10^{-19} s, 6.3×10^{10} rad/s
32.7 a) 6.94×10^{14} Hz b) 375 V/m
c) $E(x, t) = (375 \text{ V/m}) \times \cos[(1.45 \times 10^7 \text{ rad/m})x - (4.36 \times 10^{15} \text{ rad/s})t]$,
$B(x, t) = (1.25\ \mu\text{T}) \times \cos[(1.45 \times 10^7 \text{ rad/m})x - (4.36 \times 10^{15} \text{ rad/s})t]$
32.9 a) (i) 60 kHz (ii) 6.0×10^{13} Hz (iii) 6.0×10^{16} Hz
b) (i) 4.62×10^{-14} m $= 4.62 \times 10^{-5}$ nm
(ii) 508 m $= 5.08 \times 10^{11}$ nm
32.11 a) $+y$-direction b) 0.149 mm
c) $\vec{B} = (1.03 \text{ mT}) \cos[(4.22 \times 10^4 \text{ rad/m})y - (1.265 \times 10^{13} \text{ rad/s})t]\hat{\imath}$
32.13 a) 361 m b) 0.0174 rad/m
c) 5.22×10^6 rad/s d) 0.0144 V/m
32.15 a) 0.381 μm b) 0.526 μm
c) 1.38 d) 1.90
32.17 a) 330 W/m^2 b) 500 V/m, 1.7 μT
32.19 2.5×10^{25} W
32.21 a) 0.24 mW b) 17.4 V/m
32.23 12.0 V/m, 40.0 nT
32.25 850 kW
32.27 a) 0.18 mW b) 274 V/m, 0.913 μT
c) 0.18 mJ/s d) 0.010 W/cm^2
32.29 a) 637 W/m^2 b) 693 V/m, 2.31 μT
c) 2.12 μJ/m^3
32.31 a) 30.5 cm b) 2.46 GHz c) 2.11 GHz
32.33 a) 0.375 mJ b) 4.08 mPa c) 604 nm, 3.70×10^{14} Hz d) 30.3 kV/m, 101 μT
32.35 a) 6.02×10^{-9} W/m^2 b) 2.13×10^{-3} N/C, 7.10×10^{-12} T c) 1.20×10^{-18} N; no
32.37 a) at $r = R$: 64 MW/m^2, 0.21 Pa; at $r = R/2$: 260 MW/m^2, 0.85 Pa b) no
32.39 3.89×10^{-13} rad/s^2
32.41 a) $\rho I/\pi a^2$, in the direction of the current
b) $\mu_0 I/2\pi a$, counterclockwise if the current is out of the page
c) $\dfrac{\rho I^2}{2\pi^2 a^3}$, radially inward d) $\dfrac{\rho I^2}{\pi a^2} = I^2 R$
32.43 a) 1.363 m b) 10.90 m
32.45 a) 9.75×10^{-15} W/m^2
b) 2.71 μV/m, 9.03×10^{-15} T, 67.3 ms
c) 3.25×10^{-23} Pa d) 0.190 m
32.47 a) $\dfrac{4\rho G\pi MR^3}{3r^2}$ b) $\dfrac{LR^2}{4cr^2}$ c) 0.19 μm; no
32.49 b) 3.00×10^8 m/s
32.51 b) 1.39×10^{-11} c) 2.54×10^{-8}
32.53 c) 66.0 μm
32.55 choice (d)

Chapter 33

33.1 39.4°
33.3 a) 2.04×10^8 m/s b) 442 nm
33.5 a) 1.55 b) 550 nm
33.7 a) 47.5° b) 66.0°
33.9 2.51×10^8 m/s
33.11 a) 2.34 b) 82°
33.13 71.8°
33.15 a) 51.3° b) 33.6°
33.17 a) 58.1° b) 22.8°
33.19 1.77
33.21 a) 48.9° b) 28.7°
33.23 0.6°
33.25 $0.375I_0$
33.27 a) A: $I_0/2$, B: $0.125I_0$, C: $0.0938I_0$ b) 0
33.29 a) 1.40 b) 35.5°
33.31 $\arccos\left(\dfrac{\cos\theta}{\sqrt{2}}\right)$
33.33 6.38 W/cm^2
33.35 a) $0.364I$ b) $2.70I$
33.37 a) 46.7° b) 13.4°
33.39 72.1°
33.41 1.28
33.43 3.52×10^4
33.45 1.84
33.47 a) 48.6° b) 48.6°
33.49 39.1°
33.51 b) 0.23°; about the same
33.53 b) 38.9° c) 5.0°
33.55 23.3°
33.57 a) A: 1.46, carbon tetrachloride; B: 1.33, water; C: 1.63, carbon disulfide; D: 1.50, benzene
b) A: 2.13, B: 1.77, C: 2.66, D: 2.25
c) all: 5.09×10^{14} Hz
33.59 a) 35° b) $I_0 = 10$ W/m^2, $I_p = 20$ W/m^2
33.61 a) $\Delta = 2\theta_a^A - 6\arcsin\left(\dfrac{\sin\theta_a^A}{n}\right) + 2\pi$
b) $\theta_2 = \arccos\sqrt{\dfrac{n^2 - 1}{8}}$
c) violet: $\theta_2 = 71.55°$, $\Delta = 233.2°$
red: $\theta_2 = 71.94°$, $\Delta = 230.1°$; violet
33.63 choice (d)

Chapter 34

34.1 39.2 cm to the right of the mirror, 4.85 cm
34.3 9.0 cm; tip of the lead
34.5 b) 33.0 cm to the left of the vertex, 1.20 cm, inverted, real
34.7 0.213 mm
34.9 18.0 cm from the vertex; 0.50 cm, erect, virtual
34.11 a) +4.00
b) 48.0 cm to the right of the mirror; virtual
34.13 a) concave b) $f = 2.50$ cm, $R = 5.00$ cm
34.15 a) 10.0 cm to the left of the shell vertex, 2.20 mm
b) 4.29 cm to the right of the shell vertex, 0.944 mm
34.17 2.67 cm
34.19 3.30 m
34.21 a) at the center of the bowl, 1.33 b) no
34.23 39.5 cm
34.25 8.35 cm to the left of the vertex, 0.326 mm; erect
34.27 a) 107 cm to the right of the lens, 17.8 mm; real; inverted b) the same
34.29 71.2 cm to the right of the lens; −2.97
34.31 3.69 cm; 2.82 cm to the left of the lens
34.33 1.67
34.35 a) 18.6 mm b) 19 mm from the cornea
c) 0.61 mm; real; inverted
34.37 a) 36.0 cm to the right of the lens b) 180 cm to the left of the lens c) 7.20 cm to the left of the lens d) 13.8 cm to the left of the lens
34.39 26.3 cm from the lens, 12.4 mm; erect; same side
34.41 a) 200 cm to the right of the first lens, 4.80 cm
b) 150 cm to the right of the second lens, 7.20 cm
34.43 a) 53.0 cm b) real c) 2.50 mm; inverted
34.45 10.2 m
34.47 8.69 cm; no
34.49 a) $f/11$ b) $1/480$ s $= 2.1$ ms
34.51 a) 80.0 cm b) 76.9 cm
34.53 49.4 cm, 2.02 diopters
34.55 –1.37 diopters
34.57 a) 6.06 cm b) 4.12 mm
34.59 a) 8.37 mm b) 21.4 c) −297
34.61 a) −6.33 b) 1.90 cm c) 0.127 rad
34.63 a) 0.661 m b) 59.1
34.65 7.20 m/s
34.67 a) 20.0 cm b) 39.0 cm
34.69 51 m/s
34.71 a) 1.49 cm
34.73 b) 2.4 cm; −0.133
34.75 2.00
34.77 a) converging, 52.5 cm from the lens
b) converging, 17.5 cm from the lens
34.79 converging, +50.2 cm
34.81 a) 58.7 cm, converging b) 4.48 mm; virtual
34.83 a) 6.48 mm b) no, behind the retina
c) 19.3 mm from the cornea; in front of the retina
34.85 10.6 cm
34.87 a) 0.24 m b) 0.24 m
34.89 b) first image: (i) 51.3 cm to the right of the lens (ii) real (iii) inverted
second image: (i) 51.3 cm to the right of the lens (ii) real (iii) erect
34.91 −26.7 cm
34.93 7.06 cm to the left of the spherical mirror vertex, 0.177 cm tall; 13.3 cm to the left of the spherical mirror vertex, 0.111 cm tall
34.95 134 cm to the left of the object
34.97 4.17 diopters
34.99 a) 30.9 cm b) 29.2 cm
34.101 d) 36.0 cm, 21.6 cm; $d = 1.2$ cm
34.103 a) −16.6 cm b) 20.0 cm to the right
34.105 a) $4f$
34.107 b) 1.74 cm
34.109 choice (d)
34.111 choice (b)

Chapter 35

35.1 a) 14 cm, 48 cm, 82 cm, 116 cm, 150 cm
b) 31 cm, 65 cm, 99 cm, 133 cm
35.3 0.75 m, 2.00 m, 3.25 m, 4.50 m, 5.75 m, 7.00 m, 8.25 m

35.5 a) 2.0 m b) constructively
c) 1.0 m, destructively
35.7 1.14 mm
35.9 0.83 mm
35.11 a) 39 b) $\pm 73.3°$
35.13 12.6 cm
35.15 1200 nm
35.17 a) $0.750I_0$ b) 80 nm
35.19 1670 rad
35.21 a) 4.52 rad b) $0.404I_0$
35.23 114 nm
35.25 0.0234°
35.27 a) 55.6 nm
b) (i) 2180 nm (ii) 11.0 wavelengths
35.29 a) 514 nm; green b) 603 nm; orange
35.31 0.11 μm
35.33 0.570 mm
35.35 1.57
35.37 a) 96.0 nm b) no, no
35.39 a) 1.58 mm (green), 1.72 mm (orange)
b) 3.45 mm (violet), 4.74 mm (green),
5.16 mm (orange) c) 9.57 μm
35.41 1.730
35.43 761 m, 219 m, 90.1 m, 20.0 m
35.45 $6.8 \times 10^{-5}\ (\mathrm{C°})^{-1}$
35.47 1.33 μm
35.49 600 nm, 467 nm; no
35.51 a) 1.54 b) $\pm 15.0°$
35.53 a) 50 MHz b) 237.0 m
35.55 14.0
35.57 choice (d)
35.59 choice (c)

Chapter 36

36.1 506 nm
36.3 a) 226 b) $\pm 83.0°$
36.5 9.07 m
36.7 a) 63.8 cm
b) $\pm 22.1°, \pm 34.3°, \pm 48.8°, \pm 70.1°$
36.9 $\pm 16.0°, \pm 33.4°, \pm 55.6°$
36.11 a) 10.9 mm b) 5.4 mm
36.13 a) 580 nm b) 0.128
36.15 a) 6.75 mm b) 2.43 $\mu\mathrm{W/m^2}$
36.17 a) 668 nm b) $(9.36 \times 10^{-5})I_0$
36.19 a) 3 b) 2
36.21 a) 0.0627°, 0.125° b) $0.249I_0$, $0.0256I_0$
36.23 a) 4830 lines/cm b) 4; $\pm 37.7°, \pm 66.5°$
36.25 a) 4790 slits/cm b) 19.1°, 40.8° c) no
36.27 a) yes b) 13.3 nm
36.29 a) 467 nm b) 27.8°
36.31 a) 17,500 b) yes
c) (i) 587.8170 nm (ii) 587.7834 nm
(iii) 587.7834 nm $< \lambda <$ 587.8170 nm
36.33 2752 slits/cm
36.35 0.232 nm
36.37 92 cm
36.39 1.88 m
36.41 220 m
36.43 a) 73 m (Hubble), 1100 km (Arecibo)
b) 1600 km
36.45 1.45 m
36.47 30.2 μm
36.49 a) 78 b) $\pm 80.8°$ c) 555 $\mu\mathrm{W/m^2}$
36.51 1.68
36.53 b) 4.49 rad, 7.73 rad c) 3.14 rad, 6.28 rad,
9.42 rad; no d) 4.78°, 6.84°, 9.59°
36.55 −0.033 mm; decrease
36.57 360 nm
36.59 second
36.61 1.40
36.63 a) 1.03 mm b) 0.148 mm
36.65 a) 12.1 μm b) 10.4 cm, 15.2 cm
36.69 choice (d)
36.71 choice (a)

Chapter 37

37.1 bolt A
37.3 $0.867c$; no
37.5 a) $0.998c$ b) 126 m
37.7 1.12 h, in the spacecraft
37.9 92.5 m
37.11 a) 0.66 km b) 49 μs; 15 km c) 0.45 km
37.13 a) 3570 m b) 90.0 μs c) 89.2 μs
37.15 a) $0.806c$ b) $0.974c$ c) $0.997c$
37.17 a) toward b) $0.385c$
37.19 $0.784c$
37.21 $0.611c$
37.23 a) $0.159c$ b) \$172 million
37.25 $0.220c$; toward you
37.27 $3.06p_0$
37.29 a) $0.866c$ b) $0.608c$
37.31 a) $0.866c$ b) $0.986c$
37.33 a) 0.450 nJ b) 1.94×10^{-18} kg·m/s
c) $0.968c$
37.35 a) 1110 kg b) 52.1 cm
37.37 a) 0.867 nJ b) 0.270 nJ c) 0.452
37.39 a) 5.34 pJ (nonrel), 5.65 pJ (rel), 1.06
b) 67.8 pJ (nonrel), 331 pJ (rel), 4.88
37.41 a) 2.06 MV b) 0.330 pJ = 2.06 MeV
37.43 a) $\Delta = 8.42 \times 10^{-6}$ b) 34.0 GeV
37.45 $0.700c$
37.47 42.5 y
37.49 a) $\Delta = 9 \times 10^{-9}$ b) $7000m$
37.51 5.01 ns, clock on plane
37.53 0.168 MeV
37.55 a) 1.08×10^{14} J b) 2.70×10^{19} W
c) 1.10×10^{10} kg
37.57 a) $0.999929c$ b) $-0.9965c$ c) (i) 42.4 MeV
(a), 5.60 MeV (b) (ii) 15.7 MeV (a) and (b)
37.59 $0.357c$; receding
37.61 154 km/h
37.63 2.04×10^{-13} N
37.65 a) 2.6×10^{-8} s b) 0.97
37.67 a) 2.0×10^{-18} kg b) $4.0 \times 10^4\ \mathrm{m/s^2}$
37.69 a) 2494 MeV b) 2.526 times
c) 987.4 MeV, twice as much
37.71 choice (c)
37.73 choice (b)

Chapter 38

38.1 5.77×10^{14} Hz, 1.28×10^{-27} kg·m/s,
3.84×10^{-19} J = 2.40 eV
38.3 a) 5.00×10^{14} Hz b) 1.13×10^{19} photons/s
c) no
38.5 a) 2.47×10^{-19} J = 1.54 eV
b) 804 nm; infrared
38.7 249 km/s
38.9 2.14 eV
38.11 a) 264 nm b) 4.70 eV, same
38.13 0.311 nm; same
38.15 1.13 keV
38.17 0.0714 nm; 180°
38.19 a) 4.39×10^{-4} nm b) 0.04294 nm
c) 300 eV, loss d) 300 eV
38.21 51.0°
38.23 1.19×10^{-27} kg·m/s, 1.96×10^{-29} kg·m/s
38.25 16.6 fs
38.27 a) 5.07 mJ b) 11.3 W
c) 1.49×10^{16} photons/s
38.29 a) 6.99×10^{-24} kg·m/s b) 705 eV
38.31 6.28×10^{-24} kg·m/s, 59.4°
38.33 a) 5×10^{-33} m b) $(4 \times 10^{-9})°$ c) 0.1 mm
38.35 a) 319 eV; 1.06×10^7 m/s b) 3.89 nm
38.37 a) V_0 versus $1/\lambda$; 1.23×10^{-6} V·m (slope),
−4.76 V (y-intercept) b) 6.58×10^{-34} J·s,
4.76 eV c) 260 nm d) 84.0 nm
38.39 a) 2.40 pm (slope), 5.21 pm (y-intercept)
b) 2.40 pm c) 5.21 pm
38.41 choice (c)
38.43 choice (a)
38.45 choice (b)

Chapter 39

39.1 a) 0.155 nm b) 8.46×10^{-14} m
39.3 a) 2.37×10^{-24} kg·m/s
b) 3.08×10^{-18} J = 19.3 eV
39.5 4.36 km/s
39.7 a) 62.0 nm (photon), 0.274 nm (electron)
b) 4.96 eV (photon), 2.41×10^{-5} eV (electron)
c) $\approx$250 nm, electron
39.9 3.90×10^{-34} m, no
39.11 a) 0.0607 V b) 248 eV c) 20.5 μm
39.13 0.432 eV
39.15 a) 2.07°, 4.14° b) 1.81 cm
39.17 a) 5.82×10^{-13} J = 3.63 MeV
b) 5.82×10^{-13} J = 3.63 MeV
c) 1.32×10^7 m/s
39.19 $3.16 \times 10^{-34}\ \mathrm{kg \cdot m^2/s}$
39.21 a) −218 eV; 16 times b) 218 eV; 16 times
c) 7.60 nm d) $\frac{1}{4}$ hydrogen radius
39.23 a) 2.18×10^6 m/s, 1.09×10^6 m/s,
7.27×10^5 m/s b) 1.53×10^{-16} s,
1.22×10^{-15} s, 4.13×10^{-15} s c) 8.2×10^6
39.25 a) 20 eV b) 3 eV, 5 eV, 8 eV, 10 eV, 15 eV,
18 eV c) photo will not be absorbed
d) 3 eV $< \phi <$ 5 eV
39.27 a) −17.50 eV, −4.38 eV, −1.95 eV, −1.10 eV,
−0.71 eV b) 378 nm
39.29 a) −5.08 eV b) −5.64 eV
39.31 5.32×10^{21} photons/s
39.33 4.00×10^{17} photons/s
39.35 a) 1.2×10^{-33} b) 3.5×10^{-17}
c) 5.9×10^{-9}
39.37 a) 2060 K b) 1410 nm
39.39 1.06 mm; microwave
39.41 a) $1.7T$ b) 0.58
39.43 a) $1.9 \times 10^{10}\ \mathrm{W/m^2}$ b) 20 nm; no
c) 6510 km = $0.0093R_{\mathrm{sun}}$ d) sun; 39
39.45 a) 1.6×10^4 m/s b) 2.3×10^{-4} m
39.47 not valid
39.49 6.34×10^{-14} eV
39.51 a) 1.69×10^{-28} kg b) −2.53 keV
c) 0.655 nm
39.53 a) 12.1 eV b) 3; 103 nm, 122 nm, 657 nm
39.55 a) 0.90 eV
39.57 a) 5×10^{49} photons/s b) 30,000
39.59 29,800 K
39.61 a) $I(f) = \dfrac{2\pi h f^5}{c^3(e^{hf/kT} - 1)}$
39.63 a) 12 eV b) 0.15 mV; 7.3 km/s
c) 0.082 μV; 4.0 m/s
39.65 a) no b) 2.52 V
39.67 a) $E = c\sqrt{2mK}$ b) photon
39.69 b) $\Delta = \dfrac{m^2c^2\lambda^2}{2h^2}$
c) $v = (1 - 8.50 \times 10^{-8})c$, $\Delta = 8.50 \times 10^{-8}$
39.71 a) $\dfrac{h}{mc\sqrt{15}}$
b) (i) 1.53 MeV, 6.26×10^{-13} m
(ii) 2810 MeV, 3.41×10^{-16} m

39.73 a) 1.1×10^{-20} kg·m/s b) 19 MeV
c) $|U_{\text{Coul}}| = 0.015K$; no
39.75 a) 1.1×10^{-35} m/s b) 2.3×10^{27} y; no
39.77 20.9°
39.79 a) 248 eV b) 0.0603 eV
39.81 a) $F = -\dfrac{A|x|}{x}$, where $x \neq 0$
b) $E = \dfrac{3}{2}\left(\dfrac{h^2A^2}{m}\right)^{1/3}$
39.83 a) 3 b) 11.44 nm c) 60.5 eV
39.85 a) Antares b) Polaris and α Centauri B
c) α Centauri B
39.89 choice (a)
39.91 choice (a)

Chapter 40

40.1 $\Psi(x, t) = Ae^{-i(4.27\times10^{10}\,\text{m}^{-1})x}e^{-i(1.05\times10^{17}\,\text{s}^{-1})t}$
40.3 a) $8\pi/k$ b) $4\omega/k$; same
40.5 a) $\lambda/4, 3\lambda/4, 5\lambda/4, \ldots$ b) $0, \lambda/2, 3\lambda/2, \ldots$
40.7 no
40.9 a) 1.6×10^{-67} J b) 1.3×10^{-33} m/s; 1.0×10^{33} s c) 4.9×10^{-67} J d) no
40.11 0.166 nm
40.13 0.61 nm
40.15 b) no; no c) $\sqrt{2b}$
40.17 a) $0, L/2, L$ b) $L/4, 3L/4$ c) yes
40.19 a) 6.0×10^{-10} m (twice the width of the box), 1.1×10^{-24} kg·m/s b) 3.0×10^{-10} m (same as the width of the box), 2.2×10^{-24} kg·m/s
c) 2.0×10^{-10} m (2/3 the width of the box), 3.3×10^{-24} kg·m/s
40.21 3.43×10^{-10} m
40.23 1.38 μm
40.25 22 fm
40.27 a) 0.0013 b) 10^{-143}
40.29 a) 4.4×10^{-8} b) 4.2×10^{-4}
40.31 $1/\sqrt{2}$
40.33 1.11×10^{-33} J $= 6.93 \times 10^{-15}$ eV, 2.22×10^{-33} J $= 1.39 \times 10^{-14}$ eV; no
40.35 a) 0.21 eV b) 5900 N/m
40.37 111 nm
40.39 $(2n + 1)\dfrac{\hbar}{2}$, increases with n
40.41 a) 5.89×10^{-3} eV b) 106 μm c) 0.0118 eV
40.43 a) $|\Psi(x, t)|^2 = \dfrac{2}{L}\left[1 - \cos\left(\dfrac{4\pi^2\hbar t}{mL^2}\right)\right]$
b) $\dfrac{4\pi^2\hbar}{mL^2}$
40.45 $B = \left(\dfrac{k_1 - k_2}{k_1 + k_2}\right)A$, $C = \left(\dfrac{2k_2}{k_1 + k_2}\right)A$
40.47 a) 19.2 μm b) 11.5 μm
40.49 a) $(2/L)dx$ b) 0 c) $(2/L)dx$
40.51 a) 0.818 b) 0.500 c) yes
40.55 a) $A = C$, $B\sin kL + A\cos kL = De^{-\kappa L}$, where $k = \dfrac{\sqrt{2mE}}{\hbar}$
b) $kB = \kappa C$, $kB\cos kL - kA\sin kL = -\kappa De^{-\kappa L}$
40.57 6.63×10^{-34} J $= 4.14 \times 10^{-15}$ eV, 1.33×10^{-33} J $= 8.30 \times 10^{-15}$ eV, no
40.59 b) 134 eV
40.61 a) 3, 4 b) 0.90 nm c) 890 nm
40.63 22 eV, 56 eV, 110 eV
40.65 a) $x = \pm\sqrt{2E/k'}$ c) underestimate
40.67 choice (c)
40.69 choice (a)

Chapter 41

41.1 a) 1 b) 3
41.3 3.51 nm
41.5 (2, 2, 1): $x = L/2$, $y = L/2$; (2, 1, 1): $x = L/2$; (1, 1, 1): none
41.7 a) 0 b) $\sqrt{12}\,\hbar$, 3.65×10^{-34} kg·m²/s
c) $3\hbar$, 3.16×10^{-34} kg·m²/s
d) $\frac{1}{2}\hbar$, 5.27×10^{-35} kg·m²/s e) $\frac{1}{6}$
41.9 4
41.11 4
41.13 $1.414\hbar$, $19.49\hbar$, $199.5\hbar$; as n increases, the maximum L gets closer to $n\hbar$.
41.15 a) 18 b) $m_l = -4$, 153.4°
c) $m_l = +4$, 26.6°
41.19 a) 0.468 T b) 3
41.21 a) 9 b) 3.47×10^{-5} eV c) 2.78×10^{-4} eV
41.23 a) 2.5×10^{30} rad/s
b) 2.5×10^{13} m/s; not valid since $v > c$
41.25 1.68×10^{-4} eV; $m_s = +\frac{1}{2}$
41.27 $n = 1, l = 0, m_l = 0, m_s = \pm\frac{1}{2}$: 2 states;
$n = 2, l = 0, m_l = 0, m_s = \pm\frac{1}{2}$: 2 states;
$n = 2, l = 1, m_l = 0, \pm1, m_s = \pm\frac{1}{2}$: 6 states
41.29 a) $1s^22s^2$ b) magnesium; $1s^22s^22p^63s^2$
c) calcium, $1s^22s^22p^63s^23p^64s^2$
41.31 4.18 eV
41.33 a) $1s^22s^22p$ b) −30.6 eV
c) $1s^22s^22p^63s^23p$ d) −13.6 eV
41.35 a) −13.6 eV b) −3.4 eV
41.37 a) 8.95×10^{17} Hz, 3.71 keV, 3.35×10^{-10} m
b) 1.68×10^{18} Hz, 6.96 keV, 1.79×10^{-10} m
c) 5.48×10^{18} Hz, 22.7 keV, 5.47×10^{-11} m
41.39 $3E_{1,1,1}$
41.41 a) $\frac{1}{64} = 0.0156$ b) 7.50×10^{-4}
c) 2.06×10^{-3}
41.43 a) 0.500 b) 0.409
41.45 a) $E = \hbar\left[(n_x + n_y + 1)\omega_1^2 + \left(n_z + \frac{1}{2}\right)\omega_2^2\right]$, with n_x, n_y, n_z nonnegative integers
b) $\hbar\left(\omega_1^2 + \frac{1}{2}\omega_2^2\right)$, $\hbar\left(\omega_1^2 + \frac{3}{2}\omega_2^2\right)$ c) 1
41.47 b) $n = 5$ shell
41.49 a) $2a$ b) 0.238
41.51 $4a$; same
41.53 b) $(\theta_L)_{\max} = \arccos\left(-\sqrt{1 - 1/n}\right)$
41.55 3.00 T
41.57 a) $0.99999978 = 1 - 2.2 \times 10^{-7}$
b) 0.9978 c) 0.978
41.59 a) 4, 20 b) $1s^42s^42p^3$
41.61 a) 122 nm b) 1.52 pm; increase
41.63 a) 0.188 nm, 0.250 nm
b) 0.0471 nm, 0.0624 nm
41.65 a) Li: 5.391 eV; Na: 5.139 eV; K: 4.341 eV; Rb: 4.177 eV; Cs: 3.894 eV; Fr: 3.9 eV
b) Li: 3; 2; Na: 11; 3; K: 19; 4; Rb: 37; 5; Cs: 55; 6; Fr: 87; 7
c) Li: 1.26; Na: 1.84; K: 2.26; Rb: 2.77; Cs: 3.21; Fr: 3.8 d) increase
41.67 a) 2.84×10^{10} Hz/T b) 9.41×10^{-24} J/T
c) 1.78×10^{11} Hz/T; 2.03
41.69 a) 3.02×10^{-11} m, 3.83×10^6 m/s
b) 83.5 eV c) −166.9 eV d) 83.4 eV
41.71 choice (b)
41.73 choice (d)

Chapter 42

42.1 277 nm; ultraviolet
42.3 40.8 μm
42.5 5.65×10^{-13} m
42.7 2440 MHz, 0.123 m; yes
42.9 a) 1.03×10^{12} rad/s
b) 66.3 m/s (C), 49.8 m/s (O) c) 6.10 ps
42.11 a) 7.49×10^{-3} eV b) 166 μm
42.13 30.27 N/m
42.15 2170 kg/m³
42.17 a) 1.12 eV
42.19 1.20×10^6
42.21 1.5×10^{22} states per electron volt
42.23 a) $0.0233R$ b) 0.00767 = 0.767%
c) no, motion of the ions
42.25 0.312 = 31.2%
42.27 0.20 eV below the bottom of the conduction band
42.29 a) (i) 0.0204 mA (ii) −0.0196 mA (iii) 26.8 mA (iv) −0.491 mA
b) yes, where -1.0 mV $< V < +1.0$ mV
42.31 a) 5.56 mA b) −5.18 mA, −3.77 mA
42.33 a) 977 N/m b) 1.25×10^{14} Hz
42.35 a) 3.8×10^{-29} C·m b) 1.3×10^{-19} C
c) 0.81 d) 0.058, much less
42.37 a) 0.96 nm b) 1.8 nm
42.39 b) (i) 2.95 (ii) 4.73 (iii) 7.57 (iv) 0.838 (v) 5.69×10^{-9}
42.41 a) 1.146 cm, 2.291 cm
b) 1.171 cm, 2.341 cm; 0.025 cm ($2 \to 1$), 0.050 cm ($1 \to 0$)
42.43 0.274 eV; much less
42.45 a) 4.24×10^{-47} kg·m²
b) (i) 4.30 μm (ii) 4.28 μm (iii) 4.40 μm
42.47 2.03 eV
42.49 a) $2/a^3$ b) 4.7 eV
42.51 a) 0.445 eV (slope), 1.80 eV (y-intercept)
b) 5170 K, 1.80 eV
42.53 b) 3.81×10^{10} Pa $= 3.76 \times 10^5$ atm
42.55 a) 1.67×10^{33} m⁻³ b) yes
c) 6.66×10^{35} m⁻³ d) no
42.57 choice (b)

Chapter 43

43.1 a) 14 p, 14 n b) 37 p, 48 n c) 81 p, 124 n
43.3 0.533 T
43.5 a) 76.21 MeV
b) 76.68 MeV; 0.6%; greater accuracy for $^{62}_{28}$Ni
43.7 a) 92.16 MeV b) 7.680 MeV/nucleon
c) 0.8245%
43.9 a) 1.32 MeV b) 1.13×10^7 m/s
43.11 $^{86}_{36}$K: 8.73 MeV/nucleon;
$^{180}_{73}$Ta: 8.08 MeV/nucleon; yes
43.13 a) $^{235}_{92}$U b) $^{24}_{12}$Mg c) $^{15}_{7}$N
43.15 156 keV
43.17 a) 0.836 MeV b) 0.700 MeV
43.19 5.01×10^4 y
43.21 a) 4.92×10^{-18} s⁻¹ b) 2990 kg
c) 1.24×10^5 decays/s
43.23 a) 163 decays/min b) 0.435 decay/min
43.25 a) 0.421 decay/s b) 11.4 pCi
43.27 2.80 days
43.29 a) 2.02×10^{15}
b) 1.01×10^{15}; 3.78×10^{11} decays/s
c) 2.53×10^{14}; 9.45×10^{10} decays/s
43.31 a) 1.2 mJ b) 10 mrem, 10 mrad, 7.5 mJ
c) 6.2
43.33 500 rad, 2000 rem, 5.0 J/kg
43.35 a) 1.75 kGy, 1.75 kSv, 175 krem, 385 J
b) 1.75 kGy, 2.625 kSv, 262.5 krem, 385 J
43.37 a) 9.32 rad, 9.32 rem
43.39 a) 0.497 mJ b) 0.0828 rem
43.41 a) $Z = 3$, $A = 7$ b) 7.152 MeV
c) 1.4 MeV
43.43 a) 173.3 MeV b) 4.42×10^{23} MeV/g
43.45 a) $Z = 5$, $A = 10$ b) absorbed; 2.79 MeV
43.47 a) 4.7×10^4 J/g b) 8.2×10^{10} J/g
c) 4.3×10^{11} J/g d) 7600 y
43.49 a) 4.14 MeV b) 7.75 MeV/nucleon, about half the binding energy per nucleon
43.51 a) $^{90}_{39}$Y b) 25% c) 112 y

43.53 a) $^{25}_{13}$Al will decay into $^{25}_{12}$Mg.
b) β^+ or electron capture c) 3.255 MeV (β^+), 4.277 MeV (electron capture)
43.55 a) $^{14}_{6}C \rightarrow e^- + ^{14}_{7}N + \overline{\nu}_e$ b) 0.156 MeV
c) 13.5 kg; 3400 decays/s
d) 530 MeV/s = 8.5×10^{-11} J/s
e) 36 μGy, 3.6 mrad, 36 μSv, 3.6 mrem
43.57 1.03×10^{-3} u; yes
43.59 a) 5.0×10^4 b) $10^{-15{,}000}$
43.61 29.2%
43.63 a) 0.96 μJ/s b) 0.48 mrad/s c) 0.34 mrem
d) 6.9 days
43.65 1.0×10^4 y
43.67 a) 0.48 MeV
b) 3.270 MeV = 5.239×10^{-13} J
c) 3.155×10^{-11} J/mol, more than a million times larger
43.69 a) 1.16 h b) 1.20×10^8 c) 1.81×10^6
43.71 4.59×10^{-5} g/h
43.73 choice (a)
43.75 choice (d)
43.77 choice (b)

Chapter 44

44.1 a) 69 MeV, 1.7×10^{22} Hz, 18 fm; gamma ray
44.3 a) 32 MeV
44.5 9.26×10^6 m/s
44.7 7.2×10^{19} J; 70%
44.9 a) 1.18 T b) 3.42 MeV, 1.81×10^7 m/s
44.11 a) 30.6 GeV b) 8.0 GeV
44.13 a) $0.999999559c$ b) 3.83×10^8 rad/s (nonrel), 3.59×10^5 rad/s (rel)
44.15 a) 3200 GeV b) 38.7 GeV
44.17 a) π^0, π^+ b) 219.1 MeV
44.19 1.63×10^{-25} kg; 97.2
44.21 116 MeV
44.23 (b) and (d)
44.25 (c) and (d)
44.27 a) 0, 1, −1, 0 b) 0, 0, 0, 1 c) $-e$, 1, 0, 0
d) $-e$, 0, 0, −1
44.29 a) $u\overline{s}$ b) $\overline{d}\,\overline{d}\,\overline{s}$ c) uss
44.31 a) 3.28×10^7 m/s b) 1590 Mly
44.33 a) 1.08×10^5 km/s b) 1.46
44.35 −0.783 MeV; endoergic
44.37 966 nm
44.39 a) $\pi^- \rightarrow \mu^- + \upsilon \rightarrow e^- + 3\upsilon$; an electron and neutrinos b) 139 MeV c) 2.24×10^{10}
d) 50 Sv, 5.0 krem
44.41 2.494 GeV
44.43 a) 0, $+e$, 1, $L_e = L_\mu = L_\tau = 0$, K^+
b) 0, $-e$, 0, $L_e = L_\mu = L_\tau = 0$, π^-
c) −1, 0, 0, $L_e = L_\mu = L_\tau = 0$, antineutron ($\overline{n}$)
d) 0, $+e$, 0, $L_\mu = -1$, $L_e = L_\tau = 0$, μ^+
44.45 7.5×10^{-23} s
44.47 a) 0.70 rad b) 0.70 rem, 7 times, 2%; no
44.49 b) $R/R_0 = 0.574$ c) speeding up at 300 My, slowing down at 13.1 Gy
44.51 230 MeV, 12.5° below the $+x$-axis
44.53 a) all are much less; no b) 37.5 cm
c) 0.42 MeV d) 3.8×10^7 rad/s
44.55 a) $Q = -1$, $S = -3$; yes
b) Δ: ddd, udd, uud, uuu, Σ^*: dds, uds, uus, Ξ^*: dss, uss, Ω^-: sss
44.57 choice (d)
44.59 choice (c)

CREDITS

Chapter 1 Opener: Minerva Studio/Shutterstock; 1.1a: Michele Perbellini/Shutterstock; 1.1b: Studio Bazile/Thales/ESA; 1.4: AFP/Getty Images/Newscom; 1.5ab: NASA; 1.5c: JPL/NASA; 1.5d: Photodisc/Getty Images; 1.5e: Chad Baker/Photodisc/Getty Images; 1.5f: Veeco Instruments, Inc; 1.6: Pearson; 1.7: ND/Roger Viollet/Getty Images; Appl. p. 10: Tyler Olsen/Shutterstock

Chapter 2 Opener: Vibe Images/Fotolia; 2.4: Michael Dalder/Reuters/Landov; 2.5: Wolfgang Rattay/Reuters; Appl. p. 46: NASA; 2.22: Richard Megna/Fundamental Photographs; 2.26: Andreas Stirnberg/Getty Images; 2.27: Guichaoua/Alamy; E2.54 (chart): Source: "The Flying Leap of the Flea" by M. Rothschild, Y. Schlein, K. Parker, C. Neville, and S. Sternberg in the November 1973 *Scientific American*

Chapter 3 Opener: Feng Li/Getty Images; Appl. p. 71: Luca Lozzi/Getty Images; 3.8: Dominique Douieb/PhotoAlto Agency/Getty Images; 3.16: Richard Megna/Fundamental Photographs; 3.19: Fundamental Photographs; Appl. p. 85: David Wall/Alamy; 3.31: Hart Matthews JHM/GAC/Reuters

Chapter 4 Opener: Monkey Business/Fotolia; p. 105 (law): Newton, Isaac. 1845. *Newton's Principia: The Mathematical Principles*. Andrew Mott (trans.) New York: Daniel Adee; Appl. p. 106: Prisca Koller/Fotolia; 4.11: Wayne Eastep/The Image Bank/Getty Images; 4.16: Albert Gea/Reuters; p. 111 (law): Newton, Isaac. 1845. *Newton's Principia: The Mathematical Principles*. Andrew Mott (trans.) New York: Daniel Adee; Appl. p. 111: Kadmy/Fotolia; 4.19: Cheryl A. Meyer/Shutterstock; p. 116 (law): Newton, Isaac. 1845. *Newton's Principia: The Mathematical Principles*. Andrew Mott (trans.) New York: Daniel Adee; 4.28: Maksym Gorpenyuk/Shutterstock; 4.29a: PCN Black Photography/Alamy; 4.29b: John W. McDonough/Sports Illustrated/Getty Images; 4.29c: Roy Pedersen/Fotolia; P4.53 (table): Data from www.autosnout.com

Chapter 5 Opener: Brian A Jackson/Shutterstock; 5.11: NASA; 5.16: Efired/Shutterstock; Appl. p. 144: Alex Kosev/Shutterstock; Appl. p. 148: Eye of Science/Science Source; 5.26b: 2happy/Shutterstock; Appl. p. 155: Suthep/Shutterstock; 5.38a: JPL/Space Science Institute/NASA; 5.38b: Jason Stitt/Shutterstock; 5.38c: FikMik/Shutterstock; 5.38d: Shots Studio/Shutterstock; P5.78 (chart): Source: "The Flying Leap of the Flea" by M. Rothschild, Y. Schlein, K. Parker, C. Neville, and S. Sternberg in the November 1973 *Scientific American;* P5.111 (table): Courtesy of The Engineering Toolbox. www.engineeringtoolbox.com/wire-rope-strength-d_1518.html

Chapter 6 Opener: Brocreative/Shutterstock; 6.1: mariiya/Fotolia; Appl. p. 173: Steve Gschmeissner/Science Source; 6.13: Mikadun/Shutterstock; Appl. p. 186: Steve Gschmeissner/Science Source; 6.26: Fox Photos/Hulton Archive/Getty Images; Appl. p. 190: Gayvoronskaya_Yana/Shutterstock; 6.27a: Keystone/Hulton Archive/Getty Images; 6.27b: Anthony Hall/Fotolia; 6.28: Fandu/Fotolia; P6.90 (chart): Data from Science Buddies. www.sciencebuddies.org/science-fair-projects/project_ideas/Physics_Springs_Tutorial.shtml

Chapter 7 Opener: Nagel Photography/Shutterstock; 7.1: Alistair Michael Thomas/Shutterstock; Appl. p. 205: Erni/Shutterstock; 7.3: Robert F. Bukaty/AP Images; 7.5: hinnamsaisuy/Shutterstock; 7.12: ejwhite/Shutterstock; Appl. p. 214: G. Ronald Austing/Science Source; 7.15: JNP/Shutterstock; Appl. p. 220: fotoedu/Shutterstock; 7.21: Kletr/Shutterstock; Appl. p. 223: LaiQuocAnh/Shutterstock; Appl. p. 224: Peter Menzel/Science Source; P7.78 (table): Courtesy of the EngineersHandbook.com

Chapter 8 Opener: Colorful High Speed Photographs/Moment Select/Getty Images; 8.2: Alex Emanuel Koch/Shutterstock; Appl. p. 239: Willie Linn/Shutterstock; 8.4: Vereshchagin Dmitry/Shutterstock; 8.5: Jim Cummins/The Image Bank/Getty Images; 8.7ab: Andrew Davidhazy; 8.17: FPG/Archive Photos/Getty Images; 8.22: David Leah/The Image Bank/Getty Images; 8.30: Richard Megna/Fundamental Photographs; Appl. p. 258: Elliotte Rusty Harold/Shutterstock; 8.34: NASA; P8.101 (table): Courtesy of Chuck Hawks. www.chuckhawks.com/handgun_power_chart.htm

Chapter 9 Opener: Blanscape/Shutterstock; 9.3b: connel/Shutterstock; Appl. p. 278: Hybrid Medical Animation/Science Source; Appl. p. 285 (top): Dan Rodney/Shutterstock; Appl. p. 285 (bottom): Ammit Jack/Shutterstock; 9.18: David J. Phillip/AP Images; 9.19: DenisNata/Shutterstock; 9.22: NASA; P9.86: NASA

Chapter 10 Opener: Lonely Planet Images; 10.7: 68/Ocean/Corbis; Appl. p. 311 (right): Bruce MacQueen/Shutterstock; Appl. p. 311 (left): David Lentink/Science Source; 10.14: Robert Young/Fotolia; 10.17: gorillaimages/Shutterstock; Appl. p. 315: Chris DeRidder/Shutterstock; 10.22: Bjorn Heller/Shutterstock; 10.28: Gerard Lacz/Photoshot; P10.88: Data from Chevrolet

Chapter 11 Opener: nito/Shutterstock; 11.3: Turleyt/Fotolia; 11.8a: Maridav/Shutterstock; 11.12a: Benjamin Marin Rubio/Shutterstock; 11.12b: Richard Carey/Fotolia; 11.12c: Andrew Bret Wallis/Photodisc/Getty Images; 11.13: Djomas/Shutterstock; Appl. p. 350: Dante Fenolio/Science Source

Chapter 12 Opener (left): orlandin/Shutterstock; Opener (right): Isabelle Kuehn/Shutterstock; Appl. p. 370: Jorg Hackemann/Shutterstock; 12.6: SPL/Science Source; p. 373 (law): Pascal, Blaise. 1653. *Treatise on the Equilibrium of Liquids*; 12.9b: Andrey Jitkov/Shutterstock; Appl. p. 375: Lisa F. Young/Shutterstock; p. 376 (principle): Source: Archimedes; 12.14: Pali A/Fotolia; 12.19: Pearson; 12.20: Popov Nikolay/Shutterstock; 12.22: Fribus Ekaterina/Shutterstock; Appl. p. 382: Kairos69/Shutterstock; 12.28: Shutterstock; Appl. p. 386: Edward Lara/Shutterstock; 12.30a: jupeart/Shutterstock; 12.30b: Btrseller/Shutterstock; 12.31f: Harold Edgerton at MIT, copyright 2014. Courtesy of Palm Press, Inc.

Chapter 13 Opener: JPL-Caltech/SSI/Cornell/NASA; p. 398 (law): Source: Newton published the law of gravitation in 1687. *Philosophiae naturalis principia mathematica* (MacLehose, 1726) [1871] Reprinted for Sir William Thomson and Hugh Blackburn; 13.3 (Jupiter): JPL/University of Arizona/NASA; 13.3 (inset): JPL/Cornell University/NASA; 13.6: ESA/Hubble/NASA; Appl. p. 403: NASA; 13.7: swisshippo/Fotolia; 13.13: National Aeronautics and Space Administration/NASA; 13.14: Based on Newton, Isaac. 1728. *A Treatise on the System of the World*; 13.16: John F. Kennedy/Space Center/NASA; 13.17: ESA and M. Showalter/NASA; p. 411 (laws): Kepler, Johannes. 1597. *Mysterium Cosmographicum*; Appl. p. 412: NASA; 13.21b: NASA; p. 419 (quote): Source: John Michell, (1784). "On the Means of Discovering the Distance, Magnitude, & c. of the Fixed Stars, in Consequence of the Diminution of the Velocity of Their Light, in Case Such a Diminution Should be Found to Take Place in any of Them, and Such Other Data Should be Procured from Observations, as Would be Farther Necessary for That Purpose". *Philosophical Transactions of the Royal Society* 74 (0): 35–57; 13.27ab: NASA; 13.28: NASA; 13.29: Andrea Ghez/UCLA; P13.76 (table): Source: ssd.jpl.nasa.gov; P13.77 (table): Source: nssdc.gsfc.nasa.gov/planetary/factsheet

Chapter 14 Opener: blurAZ/Shutterstock; Appl. p. 434: Steve Byland/Shutterstock; 14.7: American Diagnostic Corporation; 14.21a: Zurijeta/Shutterstock; 14.25: Christopher Griffin/Alamy; Appl. p. 455: Symbiot/Shutterstock; Appl. p. 456: Sonya Etchison/Shutterstock

Chapter 15 Opener: Walter D. Mooney/U.S. Geological Survey; Appl. p. 469: Marco PoloCollection/Balan Madhavan/Alamy; 15.2: Charles Platiau/Reuters; 15.5: EpicStockMedia/Shutterstock; 15.12: Kodda/Shutterstock; Appl. p. 482: Christian Delbert/Shutterstock; 15.23a–d: Richard Megna/Fundamental Photographs; 15.25: Shmeliova Natalia/Shutterstock; 15.27: National Optical Astronomy Observatories

Chapter 16 Opener: Eduard Kyslynskyy/Shutterstock; 16.5 (line art): Berg, Richard E.; Stork, David G., *The Physics Of Sound*, 1st Ed., ©1982. Reprinted and electronically reproduced by permission of Pearson Education, Inc., Upper Saddle River, New Jersey; 16.5a: olly/Shutterstock; 16.5b: Lebrecht Music and Arts Photo Library/Alamy; Appl. p. 510: Steve Thorne/Getty Images; 16.6: Geoff Dann/DK Images; 16.9: Kretztechnik/Science Source; 16.10: auremar/Shutterstock; 16.15: Digoarpi/Shutterstock; Appl. p. 523: Piotr Marcinski/Shutterstock; 16.20: Martin Bough/Fundamental

Photographs; 16.24: Roger A. Freedman; 16.26: R. Gino Santa Maria/Shutterstock; 16.29: Aaron Kohr/Shutterstock; 16.36c: NASA; 16.37: NASA

Chapter 17 Opener: huyangshu/Shutterstock; 17.4: PRNewsFoto/Exergen TemporalScanner/AP Images; Appl. p. 548 (top): David Thyberg/Shutterstock; Appl. p. 548 (bottom): Hofhauser/Shutterstock; 17.5a: Roger A. Freedman; 17.11: Eugene Sergeev/Shutterstock; 17.13: Eric Schrader/Pearson Education; 17.16: Hugh D. Young; 17.18: Roman Sigaev/Shutterstock; 17.19: Science Source; 17.21: Richard Megna/Fundamental Photographs; 17.22: NASA; 17.23: MaszaS/Shutterstock; Appl. p. 566 (top): Dmitry Deshevykh/Getty Images; Appl. p. 566 (bottom): Paul Nicklen/National Geographic/Getty Images; 17.28: Hugh D. Young; 17.29: Steve Gschmeissner/Science Source; 17.30a: Source: *Trends: A Compendium of Data on Global Change*, U.S. Department of Energy's Carbon Dioxide Information Analysis Center, http://cdiac.esd.ornl.gov/trends/co2/contents.htm; 17.30b: Source: http://data.giss.nasa.gov/gistemp/graphs_v3/, first graph; P17.91: Satoshi Kuribayashi/Nature Production/Nature Picture Library

Chapter 18 Opener: stockyimages/Shutterstock; 18.2: Topham/The Image Works; 18.10: Courtesy of Bruker Corporation; 18.13: Rich Iwasaki/Stone/Getty Images; 18.14: Arctic Images/Alamy; 18.16: David Grossman/The Image Works; Appl. p. 604: Ray Coleman/Science Source; 18.25: PhotoDisc/Getty Images; E18.24: National Optical Astronomy Observatory; P18.83: National Optical Astronomy Observatory

Chapter 19 Opener: Ken Paul/All Canada Photos/Getty Images; 19.1: Roger A. Freedman via John P. Surey; 19.2a: Stocktrek/Photodisc/Getty Images; 19.2b: Alin Dragulin/Alamy; Appl. p. 624: Diego Cervo/Shutterstock; 19.10: Liv friis-larsen/Shutterstock; 19.14: Tom Branch/Science Source; 19.15: Rob Byron/Shutterstock; Appl. p. 635: Dean Bertoncelj/Shutterstock

Chapter 20 Opener: George Burba/Shutterstock; 20.2: James Morgan/Shutterstock; Appl. p. 650: Tyler Olson/Shutterstock; 20.12: Bill Bachman/Science Source; 20.16: U.S. Air Force photo by Staff Sgt. Robert Zoellner; 20.17: cowardlion/Shutterstock; Appl. p. 669: DenisNata/Shutterstock; 20.20 abc: Eric Schrader/Pearson Education; Summ. p. 675: Eric Schrader/Pearson Education; PP20.59–62 (chart): Source: Public domain from Wikimedia Commons: http://commons.wikimedia.org/wiki/File:Thermocline.jpg

Chapter 21 Opener: Mikael Damkier/Shutterstock; 21.5: Clayton Stalter/Journal-Courier/The Image Works; 21.8a: Richard Megna/Fundamental Photographs; Appl. p. 690: Knape/Getty Images; Appl. p. 695: Matt9122/Shutterstock; 21.29a: Richard Megna/Fundamental Photographs; 21.30b: OtnaYdur/Shutterstock; Appl. p. 709: Mark Smith/Science Source; PP21.98–101: Data from "Detection and Learning of Floral Electric Fields by Bumblebees," Dominic Clarke et al., *Science*, April 5, 2013: Vol. 340, no. 6128, pp. 66–69: http://www.sciencemag.org

Chapter 22 Opener: U. S. Department of Energy; Appl. p. 726: Oxford Scientific/Getty Images; 22.10: Akg-images; Appl. p. 738: David Furness, Keele University/Science Source; 22.27b: Peter Terren/Tesladownunder; Appl. p. 741: Yevgeniy11/Shutterstock

Chapter 23 Opener: Ted Kurihara Photography/The Image Bank/Getty Images; 23.9: John F. Kennedy/NASA; 23.11: Vladyslav Danilin/Fotolia; Appl. p. 760: StockLite/Shutterstock; Appl. p. 762: BSIP/Newscom; 23.17: Keystone/Hulton Archive/Getty Images; 23.22: USGS; Appl. p. 773: 3d4Medical.com/Motif/Corbis

Chapter 24 Opener: Damir Sagolj/Reuters; 24.3: Nejron Photo/Shutterstock; 24.4: Andrew Lambert Photography/Science Source; 24.7: Eric Schrader/Pearson Education; Appl. p. 791: OtnaYdur/Shutterstock; 24.11: Sandia National Laboratories; Appl. p. 800: Ernest Prim/Getty Images; Appl. p. 802: Pearson Science

Chapter 25 Opener: JGI/Tom Grill/Getty Images; 25.4: Eric Schrader/Pearson Education; 25.5: Stockbyte/Stockbyte/Getty Images; Appl. p. 822: Dr. David Furness, Keele University/Science Source; 25.8: TFoxFoto/Shutterstock; 25.12: Laurence Gough/Shutterstock; Appl. p. 828: Andrew J. Martinez/Science Source; 25.15: Richard Megna/Fundamental Photographs; p. 847 (table): Source: Based on F. Vollrath and D. Edmonds, "Consequences of electrical conductivity in an orb spider's capture web," *Naturwissenschaften*: December 2013, 100:12, pp. 1163–69

Chapter 26 Opener: 06photo/Shutterstock; 26.2: Claro Cortes/Reuters; 26.5: John P. Surey; 26.13: John P. Surey; Appl. p. 859: Gary Ombler/Dorling Kindersley; 26.18: Richard Megna/Fundamental Photographs; Appl. p. 864: Science Source; 26.25a: Vladimir Arndt/Shutterstock; 26.25b: Evgeny Tomeev/Shutterstock

Chapter 27 Opener: Michael Flippo/Fotolia; 27.2b: Fotolia; Appl. p. 884: Robert Smith/age fotostock/Getty Images; Appl. p. 885: luminaimages/Shutterstock; 27.14ab: Richard Megna/Fundamental Photographs; 27.17b: Richard Megna/Fundamental Photographs; 27.20a: NASA; 27.20b: Roman Krochuk/Shutterstock; 27.21: Lawrence Berkeley National Laboratory; Appl. p. 901: Simon Fraser/Science Source; 27.40: Jeremy Burgess/Science Source

Chapter 28 Opener: CERN/European Organization for Nuclear Research; Appl. p. 925 (earth): NASA; Appl. p. 925 (moon): Godrick/Shutterstock; 28.8: Pearson Education; 28.11: Jeremy Walker/Science Source; Appl. p. 932: James Steidl/Fotolia; Appl. p. 939: Pearson Education; Appl. p. 944: Dr. Kenneth Scarberry

Chapter 29 Opener: Jabejon/E+/Getty Images; Appl. p. 956: Simon Fraser/University of Durham/Science Source; 29.9: Maximilian Stock Ltd/Science Source; 29.19a: Kim Kulish/Corbis News/Corbis; 29.19b: Susan & Allan Parker/Alamy; Appl. p. 973: NASA/JPL/Caltech; 29.27: Richard Megna/Fundamental Photographs

Chapter 30 Opener: EschCollection/Photonica/Getty Images; 30.2: Shutterstock; Appl. p. 994: Ron Chapple/The Image Bank/Getty Images; 30.7: Turleyt/Fotolia; Appl. p. 999: SOHO/NASA; 30.10: Science Source

Chapter 31 Opener: Roger A. Freedman; 31.5: Rubberball/Mike Kemp/Getty Images; 31.14: Lawrence Migdale/Science Source; Appl. p. 1030: Mauro Fermariello/Science Source; Appl. p. 1039: Katharina Wittfeld/Shutterstock; 31.22: Thomas Barrat/Shutterstock; 31.23: Roger A. Freedman

Chapter 32 Opener: JLImages/Alamy; 32.1 (quote): Source: Albert Einstein in James Clerk Maxwell, *A Commemoration Volume* (New York: The Macmillan Company, 1931), p. 71; 32.1: Bettmann/Corbis; 32.2a: Daniel Laflor/Vetta/Getty Images; 32.2b: sss78/Fotolia; Appl. p. 1054: Ted Kinsman/Science Source; 32.16: Dudarev Mikhail/Shutterstock; 32.18: Craig Holmes Premium/Alamy; Appl. p. 1065: Alexander Tsiaras/Science Source; 32.20: NASA; 32.23: grzymkiewicz/Getty Images; PP32.54–56: Data from ICNIRP Guidelines For Limiting Exposure To Time-Varying Electric, Magnetic and Electromagnetic Fields (up to 300 Ghz). *Health Physics* 74 (4): 494–522; 1998

Chapter 33 Opener: Rozaliya/Getty Images; 33.1: bradleym/E+/Getty Images; 33.2: National Institute of Health/Science Source; 33.9a: Roger A. Freedman; Appl. p. 1083: Image Quest Marine; 33.10a: NASA; 33.10b: Stephen Meese/Getty Images; 33.13b: Ken Kay/Fundamental Photographs; 33.16: VEM/Science Source; 33.19a: mtilghma/Fotolia; 33.21a: Dougal Waters/Photographer's Choice/Getty Images; 33.21b: Pixtal/AGE Fotostock; 33.25: Diane Hirsch/Fundamental Photographs; Appl. p. 1097: Kzenon/Shutterstock; Appl. p. 1098 (top): Linda Hirst; Appl. p. 1098 (bottom): Cefo design/Shutterstock; 33.30: Peter Aprahamian/Sharples Stress Engineers Ltd./Science Source; Appl. p. 1099: Laurie Knight/Getty Images; 33.32: Roger A. Freedman; 33.36 (left): K.Nomachi/Science Source

Chapter 34 Opener: Cardinal/Corbis; 34.8: Pearson Education; 34.12abc: NASA; Appl. p. 1118: OlegD/Shutterstock; 34.23: Vaclav Volrab/Shutterstock; 34.27: Richard Megna/Fundamental Photographs; 34.34: Hoby Finn/Photodisc/Getty Images; 34.41abcd: John Hedgecoe/The Image Works; Appl. p. 1138: kak2s/Shutterstock; 34.43c: THK Photo Products; 34.44b: Science Source; Appl. p. 1140: ethylalkohol/Shutterstock; Appl. p. 1143: Jan Zoetekouw/Shutterstock; 34.52c: Jan Hinsch/Science Source; 34.54d: National Optical Astronomy Observatories

Chapter 35 Opener: Ilya Andriyanov/Shutterstock; Appl. p. 1163: Peter Dazeley/The Image Bank/Getty Images; 35.4: Roger A. Freedman; 35.6: Pearson; 35.11b: Rob Friedman/E+/Getty Images; 35.16b: Bausch & Lomb Incorporated; 35.17: Bausch & Lomb Incorporated; Appl. p. 1176 (top left): Laurie Knight/Getty Images; Appl. p. 1176 (top right, bottom): Peter Vukusic; Appl. p. 1177: Jason Reed

Chapter 36 Opener: Tomatito/Shutterstock; 36.2ab: Pearson; 36.6: Pearson; 36.9b: Richard Megna/Fundamental Photographs; 36.11: Roger A. Freedman; 36.12d: Pearson; 36.17: Dr.OGA/Shutterstock; 36.18a: National Solar Observatory; 36.18b: Science Source; Appl. p. 1200: Tek Image/Science Source; 36.20b: Estate of Bertram Eugene Warren; 36.24: SPL/Science Source; 36.26: Pearson; 36.27abc: Pearson; Appl. p. 1206: National Radio Astronomy Observatory; Appl. p. 1207: Michal Ninger/Shutterstock; 36.3: Paul Silverman/Fundamental Photographs; Summ. p. 1210: Pearson; P36.90: NASA

Chapter 37 Opener: Brookhaven National Laboratory; 37.4: Adimas/Fotolia; 37.7: John F. Kennedy Space Center/NASA; Appl. p. 1228: Getty Images; 37.11: Science Source; Appl. p. 1234: bikeriderlondon/Shutterstock; 37.18: Science Source; 37.19: John F. Kennedy Space Center/NASA; Appl. p. 1242: Imaginechina/AP Images; 37.25: Matt Cooper/Shutterstock

Chapter 38 Opener: Photolibrary/Getty Images; 38.2a: U.S. Air Force; 38.2b: Bill Corwin; Appl. p. 1260: National Cancer Institute; 38.9: Jupiterimages/Getty Images; Appl. p. 1263: Gianluca Padovani/E+/Getty Images; 38.14a: Lawrence Berkeley National Laboratory; 38.16: Pearson; Appl. p. 1270: glenda/Shutterstock; P38.44 (table): Source: http://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html

Chapter 39 Opener: Eye of Science/Science Source; 39.1: American Institute of Physics/Emilio Segre Visual Archives; 39.4: Education Development Center; 39.6: Scimat/Science Source; 39.8: Ted Kinsman/Science Source; Appl. p. 1286: NASA; 39.9: Kitt Peak National Observatory; 39.10: American Institute of Physics/Emilio Segre Visual Archives; p. 1288 (quote): Source: Ernest Rutherford quoted in Ratcliffe, J. A. (1938). "Forty Years of Physics." In Needham, J., Pagel, W. *Background to Modern Science*. Cambridge University Press, p. 61; 39.15: Boyer/Roger Viollet/Getty Images; 39.19b: Andrew Lambert Photography/Science Source; Appl. p. 1293: Louise Murray/Science Source; 39.31: Ambio-Royal Swedish Academy of Sciences; Appl. p. 1305 (a): Piotr Krzeslak/Shutterstock; Appl. p. 1305 (b): Africa Studio/Shutterstock; Appl. p. 1305 (c): Kletr/Shutterstock; Appl. p. 1306 (top): NASA Images; 39.34b: From Huggins, *Physics I,* © 1968 by W.A. Benjamin, Inc. Reprinted by permission of Addison Wesley Longman; 39.36: NASA; PP39.88–91: Source: "Whole-Cell Imaging at Nanometer Resolutions Using Fast and Slow Focused Helium Ions," Xiao Chen et al., *Biophysical Journal*, Oct. 5, 2011; 101(7): 1788–1793

Chapter 40 Opener: Dmitri Talapin; 40.1: JGI/Blend Images/Age Fotostock; 40.2: Bettmann/Corbis; 40.4: Churchill Archives Centre; 40.17: Media Services/NASA; Appl. p. 1342: Dr. Mark J. Winter/Science Source; 40.21b: Hewlett-Packard Laboratories/Science Source

Chapter 41 Opener (left): Fundamental Photographs; Opener (right): Syda Productions/Fotolia; 41.7: NASA; 41.11ab: National Optical Astronomy Observatories; Appl. p. 1381: Creativemarc/Fotolia; 41.19a: National Radio Astronomy Observatory; 41.20: Courtesy AIP Emilio Segre Visual Archives; 41.22: Felipe Oliveira/Brazil/Fotolia; Appl. p. 1389: SPL/Science Source; Appl. p. 1393 (top): Andrea Danti/Getty Images; Appl. p. 1393 (bottom): Science Source; PP41.70–73: Source: G. Günter et al., "Observing the Dynamics of Dipole-Mediated Energy Transport by Interaction Enhanced Imaging." *Science*, Nov. 2013, 342(6161): 954–956. http://www.physi.uni-heidelberg.de/Forschung/QD/datafiles/publications/2013_Guenter.pdf

Chapter 42 Opener: JPL/NASA; Appl. p. 1410: Alfred Pasieka/Science Source; 42.9: Roger A. Freedman; Appl. p. 1414: Marshall Space Flight Center/NASA; 42.17: Mondadori/Contributor/Getty Images; Appl. p. 1430: Andy Crump/Science Source; 42.36: full image/Fotolia; P42.42: NASA

Chapter 43 Opener: Newscom; Appl. p. 1442: National Science Foundation; T43.2 Source: G. Audi, A. H. Wapstra, and C. Thibault, *Nuclear Physics* A729, 337 (2003); 43.1b: Science Source; Appl. p. 1446: Steve Gschmeissner & Carol Upton/Science Source; 43.6: Lloyd Cluff/Encyclopedia/Corbis; Appl. p. 1463: f11photo/Fotolia; 43.10: Science Source; Appl. p. 1468: Argonne National Laboratory; 43.17: National Optical Astronomy Observatory; 43.18: LLNL/Science Source; Summ. p. 1473: Science Source

Chapter 44 Opener: ESA and The Hubble Heritage Team (STScl/AURA)/NASA; 44.1: Ernest Orlando Lawrence Berkeley National Laboratory; 44.3a: Ernest Orlando Lawrence Berkeley National Laboratory; Appl. p. 1484: Susan Landau; Appl. p. 1486: Mark Kostich/E+/Getty Images; 44.7ab: CERN/European Organization for Nuclear Research; 44.8: CERN/European Organization for Nuclear Research; 44.9: Roger A. Freedman via Mark R. Lowenstine; p. 1497 (quote): From *Finnegans Wake* by James Joyce quoted in J. Joyce (1982) [1939]. *Finnegans Wake*. Penguin Books. p. 383. ISBN 0-14-006286-6; 44.15: ICRR Institute for Cosmic Ray Research; 44.16ab: NASA; 44.20: NASA; Appl. p. 1509: Natursports/Shutterstock; 44.23: National Optical Astronomy Observatories; 44.24: European Space; Summ. p. 1518: National Optical Astronomy Observatories

INDEX

For users of the three-volume edition: pages 1–682 are in Volume 1; pages 683–1253 are in Volume 2; pages 1218–1522 are in Volume 3. Pages 1254–1522 are not in the Standard Edition.

NOTE: Page numbers followed by f indicate figures; those followed by t indicate tables.

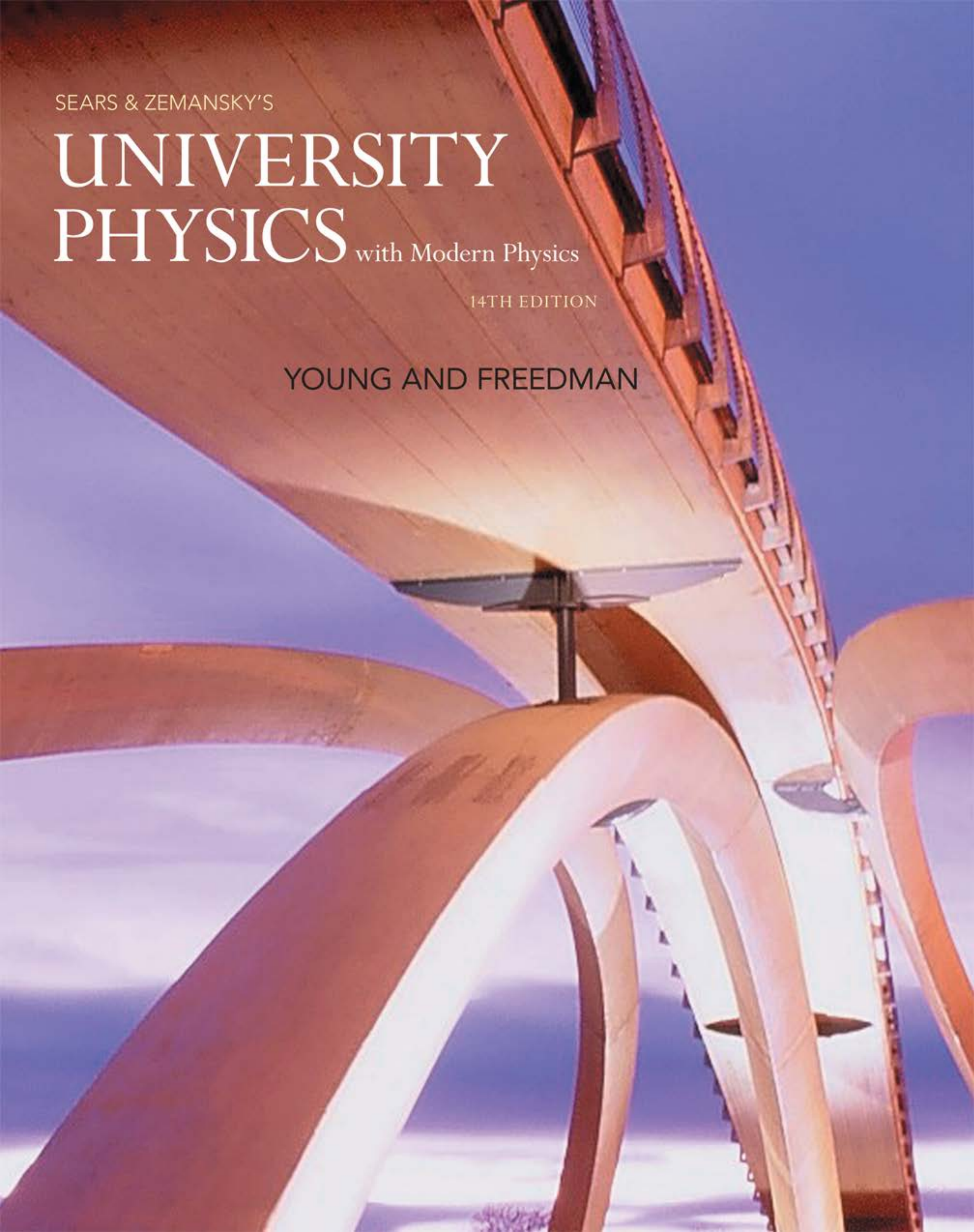
A

D

E

F

Q


SEARS & ZEMANSKY'S
UNIVERSITY
PHYSICS
with Modern Physics
14TH EDITION
YOUNG AND FREEDMAN