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 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 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 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 magnetic flux through each turn of the coil before it is rotated? After it is rotated? (b) What is the average emf induced in the coil?

29.3 • The magnetic flux through a coil is given by $\Phi_B = \alpha t - \beta t^3$, where α and β are constants. (a) What are the units of α and β? (b) If the induced emf is zero at $t = 0.500$ s, how is α related to β? (c) If the emf at $t = 0$ is -1.60 V, what is the emf at $t = 0.250$ s?

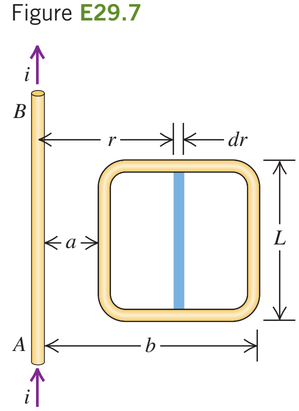
29.4 •• A small, closely wound coil has N turns, area A, and resistance R. The coil is initially in a uniform magnetic field that has magnitude B and a direction perpendicular to the plane of the loop. The coil is then rapidly pulled out of the field so that the flux through the coil is reduced to zero in time Δt. (a) What are the magnitude of the average emf $\mathcal{E}_{av}$ and average current I_{av} induced in the coil? (b) The total charge Q that flows through the coil is given by $Q = I_{av}\Delta t$. Derive an expression for Q in terms of N, A, B, and R. Note that Q does not depend on Δt. (c) What is Q if $N = 150$ turns, $A = 4.50 \text{ cm}^2$, $R = 30.0 \ \Omega$, and $B = 0.200$ T?

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-Ω 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?

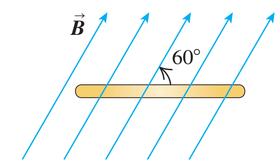
29.7 • **CALC** The current in the long, straight wire AB shown in **Fig. E29.7** is upward and is increasing steadily at a rate di/dt. (a) At an instant when the current is i, what are the magnitude and direction of the field $\vec{B}$ at a distance r to the right of the wire? (b) What is the flux $d\Phi_B$ 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 $B(t) = (1.4 \text{ T})e^{-(0.057 \text{ s}^{-1})t}$. (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



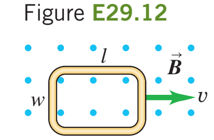
29.9 • **Shrinking Loop.** A circular loop of flexible iron wire has an initial circumference of 165.0 cm, but its circumference is decreasing at a constant rate of 12.0 cm/s due to a tangential pull on the wire. The loop is in a constant, uniform magnetic field oriented perpendicular to the plane of 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** A circular loop of wire with radius 2.00 cm and resistance $0.600 \ \Omega$ is in a region of a spatially uniform magnetic field $\vec{B}$ that is perpendicular to the plane of the loop. At $t = 0$ the magnetic field has magnitude $B_0 = 3.00$ T. The magnetic field then decreases according to the equation $B(t) = B_0 e^{-t/\tau}$, where $\tau = 0.500$ s. (a) What is the maximum magnitude of the current I induced in the loop? (b) What is the induced current I when $t = 1.50$ s?

29.12 • 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. E29.12**). (a) Find the emf induced in this coil. (b) If the speed and magnetic field are both tripled, what is the induced emf?

Figure E29.12

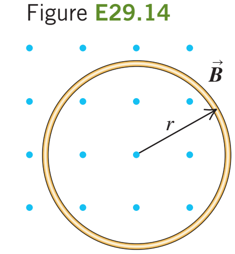


29.13 •• The armature of a small generator consists of a flat, square coil with 120 turns and sides with a length of 1.60 cm. The coil rotates in a magnetic field of 0.0750 T. What is the angular speed of the coil if the maximum emf produced is 24.0 mV?

Section 29.3 Lenz's Law

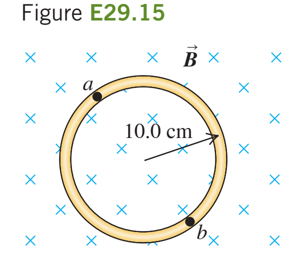
29.14 • A circular loop of wire with radius $r = 0.0480$ m and resistance $R = 0.160 \ \Omega$ is in a region of spatially uniform magnetic field, as shown in **Fig. E29.14**. The magnetic field is directed out of the plane of the figure. The magnetic field has an initial value of 8.00 T and is decreasing 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 E29.14



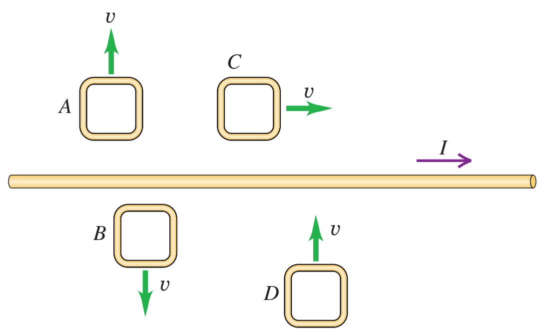
29.15 • A circular loop of wire is in a region of spatially uniform magnetic field, as shown in **Fig. E29.15**. The magnetic field is directed into the plane of the figure. Determine the direction (clockwise or counterclockwise) of the induced current in the loop when (a) B is increasing; (b) B is decreasing; (c) B is constant with value B_0. Explain your reasoning.

Figure E29.15



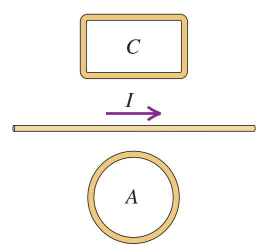
29.16 • The current I in a long, straight wire is constant and is directed toward the right as in **Fig. E29.16**. Conducting loops A, B, C, and D are moving, in the directions shown, near the wire. (a) For each loop, is the direction of the induced current clockwise or counterclockwise, or is the induced 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**



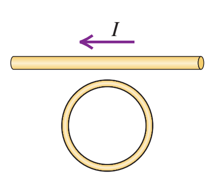
29.17 • Two closed loops A and C are close to a long wire carrying a current I (**Fig. E29.17**). (a) Find the direction (clockwise or counterclockwise) of the current induced in each loop if I is steadily decreasing. (b) While I is decreasing, what is the direction of the net force that the wire exerts on each loop? Explain how you obtain your answer.

Figure **E29.17**



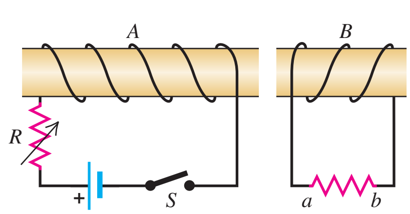
29.18 • The current in **Fig. E29.18** obeys the equation $I(t) = I_0 e^{-bt}$, where $b > 0$. Find the direction (clockwise or counterclockwise) of the current induced in the round coil for $t > 0$.

Figure **E29.18**



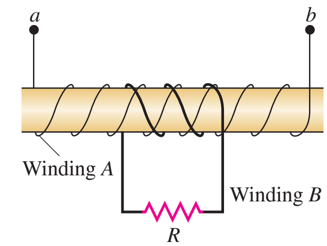
29.19 • Using Lenz's law, determine the direction of the current in resistor ab of **Fig. E29.19** when (a) switch S is opened after having been closed for several minutes; (b) coil B is brought closer to coil A with the switch closed; (c) the resistance of R is decreased while the switch remains closed.

Figure **E29.19**



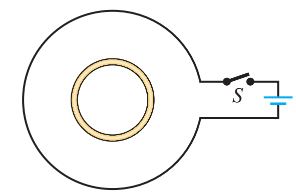
29.20 • A cardboard tube is wrapped with two windings of insulated wire wound in opposite directions, as shown in **Fig. E29.20**. Terminals a and b of winding A may be connected to a battery through a reversing switch. State whether the induced current in the resistor R is from left to right or from 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**



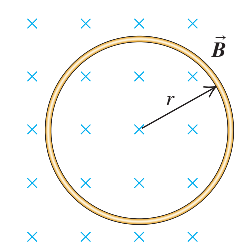
29.21 • A small, circular ring is inside a larger loop that is connected to a battery and a switch (**Fig. E29.21**). Use Lenz's law to find the direction of the current induced in the small ring (a) just after switch S is closed; (b) after S has been closed a long time; (c) just after S has been reopened after it was closed for a long time.

Figure **E29.21**



29.22 • **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. E29.22**. The magnetic field is directed into the plane of the figure. At $t = 0$, $B = 0$. The magnetic field then begins increasing, with $B(t) = (0.380\ \mathrm{T/s^3})t^3$. What is the current in the loop (magnitude and direction) at the instant when $B = 1.33$ T?

Figure **E29.22**



Section 29.4 Motional Electromotive Force

29.23 • A magnetic field of 0.080 T is in the y-direction. The velocity of wire segment S has a magnitude of 78 m/s and components of 18 m/s in the x-direction, 24 m/s in the y-direction, and 72 m/s in the z-direction. The segment has length 0.50 m and is parallel to the z-axis as it moves. (a) Find the motional emf induced between the ends of the segment. (b) What would the motional emf be if the wire segment was parallel to the y-axis?

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?

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 **E29.24**

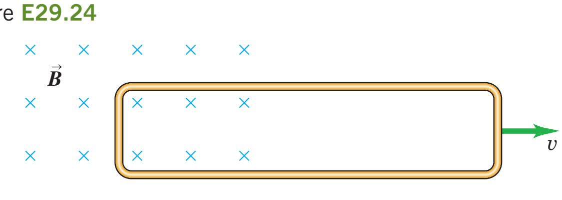
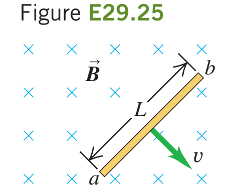
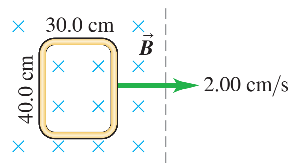


Figure **E29.25**



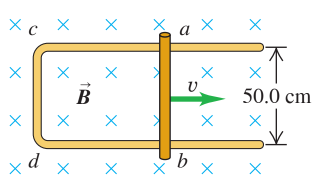
29.26 • A rectangle measuring 30.0 cm by 40.0 cm is located inside a region of a spatially uniform magnetic field of 1.25 T, with the field perpendicular to the plane of the coil (**Fig. E29.26**). The coil is pulled out at a steady rate of 2.00 cm/s traveling perpendicular to the field lines. The region 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**



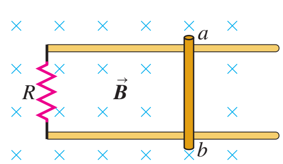
29.27 • The conducting rod *ab* shown in **Fig. E29.27** makes contact with metal rails *ca* and *db*. The apparatus is in a uniform magnetic field of 0.800 T, perpendicular to the plane of the figure. (a) Find the magnitude of the emf induced in the rod when it is moving toward the right with a speed 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 **E29.27**



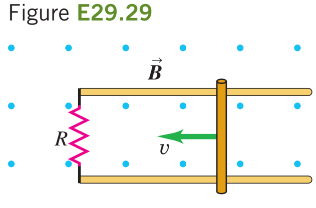
29.28 • A 0.650-m-long metal bar is pulled to the right at a steady 5.0 m/s perpendicular to a uniform, 0.750 T magnetic field. The bar rides on parallel metal rails connected through a 25.0 Ω resistor (**Fig. E29.28**), so the apparatus makes a complete circuit. Ignore the resistance of the bar and 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 **E29.28**



29.29 • 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. E29.29**, 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?

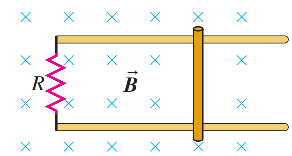
Figure **E29.29**



29.30 • Consider the circuit shown in Fig. E29.29, but with the bar moving to the right with speed v. As in Exercise 29.29, 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?

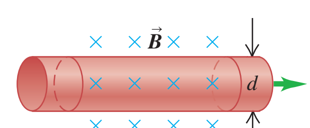
29.31 • A 0.250-m-long bar moves on parallel rails that are connected through a 6.00 Ω resistor, as shown in **Fig. E29.31**, so the apparatus makes a complete circuit. You can ignore the resistance of the bar and rails. The circuit is in a uniform magnetic field $B = 1.20$ T that is directed into the 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 **E29.31**



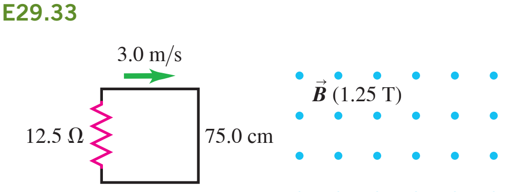
29.32 •• **BIO Measuring Blood Flow.** Blood contains positive and negative ions and thus is a conductor. A blood vessel, therefore, can be viewed as an electrical wire. We can even picture the flowing blood as a series of parallel conducting slabs whose thickness is the diameter d of the vessel moving with speed v. (See **Fig. E29.32**.) (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 **E29.32**



29.33 •• 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. E29.33**. 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 **E29.33**



Section 29.5 Induced Electric Fields

29.34 • 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?

29.35 •• 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.

29.36 •• 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?

29.37 • 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

29.38 • 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?

29.39 • **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?

29.40 • 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?

29.41 • CALC The electric flux is $(4.0\ \text{V}\cdot\text{m/s}^5)t^5$ through a certain area of a dielectric that has dielectric constant 2.5. (a) Find the displacement current through that area at $t = 1.5$ s. (b) At what time was the displacement current $\frac{1}{6}$ as much?

Section 29.8 Superconductivity

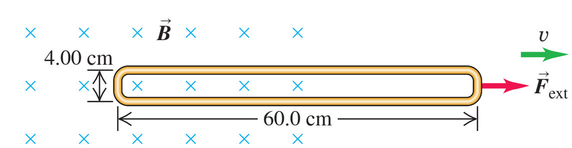
29.42 • At temperatures near absolute zero, B_c approaches 0.142 T for vanadium, a type-I superconductor. The normal phase of vanadium has a magnetic susceptibility close to zero. Consider a long, thin vanadium cylinder with its axis parallel to an external magnetic field $\vec{B}_0$ in the $+x$-direction. 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

29.43 •• CP A motor vehicle generates electrical power using an alternator, which employs electromagnetic induction to convert mechanical energy to electrical energy. The alternator acts as a dc generator (Example 29.4). The alternator maintains and replenishes charge on the car's battery and operates headlights, radiator fans, windshield wipers, power windows, computer systems, sensors, sound systems, and other components. (a) A typical car battery provides 70 amp-hours of charge. How many coulombs is that? (b) If headlights each draw 20 A of current, a radiator fan draws 10 A, and windshield wipers each draw 5 A, estimate the peak current needed for a car to operate on a rainy night. (c) A car's alternator supplies an average emf of 14 V as emf induced in a sequence of stator coils in the presence of a magnetic field created by rotor coil electromagnets turned by a pulley system. A stator coil may have 42 windings and a cross-sectional diameter of 5.0 cm, and it rotates at 400 Hz. Estimate the strength of the magnetic field generated by a rotor coil.

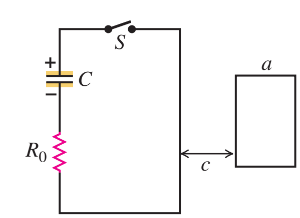
29.44 •• A very long, rectangular loop of wire can slide without friction on a horizontal surface. Initially the loop has part of its area in a region of uniform magnetic field that has magnitude $B = 2.90$ T and is perpendicular to the plane of the loop. The loop has dimensions 4.00 cm by 60.0 cm, mass 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. P29.44**). (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 P29.44



29.45 •• CP CALC In the circuit shown in **Fig. P29.45**, 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 P29.45

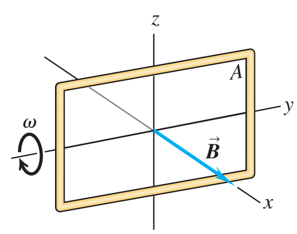


29.46 •• CP CALC In the circuit in Fig. P29.45, an emf of 90.0 V is added in series with the capacitor and the resistor, and the capacitor is initially uncharged. The emf is placed between the capacitor and switch S, with the positive terminal of the emf adjacent to the capacitor. Otherwise, the two circuits are the same as in Problem 29.45. 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?

29.47 •• **CALC** A very long, straight solenoid with a cross-sectional area of $2.00\ \text{cm}^2$ 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?

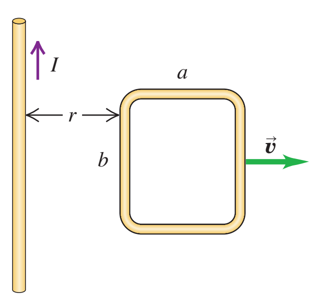
29.48 • Suppose the loop in **Fig. P29.48** is (a) rotated about the y-axis; (b) rotated about the x-axis; (c) rotated about an edge parallel to the z-axis. What is the maximum induced emf in each case if $A = 600\ \text{cm}^2$, $\omega = 35.0\ \text{rad/s}$, and $B = 0.320\ \text{T}$?

Figure **P29.48**



29.49 • In **Fig. P29.49** 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 $a \to 0$; (iii) the loop gets very far from the wire.

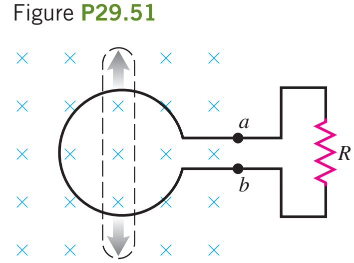
Figure **P29.49**



29.50 •• If you secure a refrigerator magnet about 2 mm from the metallic surface of a refrigerator door and then move the magnet sideways, you can feel a resistive force, indicating the presence of eddy currents in the surface. (a) Estimate the magnetic field strength B of the magnet to be 5 mT (Problem 28.53) and assume the magnet is rectangular with dimensions 4 cm wide by 2 cm high, so its area A is $8\ \text{cm}^2$. Now estimate the magnetic flux Φ_B into the refrigerator door behind the magnet. (b) If you move the magnet sideways at a speed of 2 cm/s, what is a corresponding estimate of the time rate at which the magnetic flux through an area A fixed on the refrigerator is changing as the magnet passes over? Use this estimate to estimate the emf induced under the rectangle on the door's surface.

29.51 • 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. P29.51**. 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.

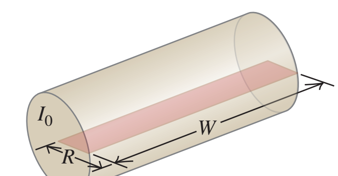
Figure **P29.51**



29.52 ••• **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?

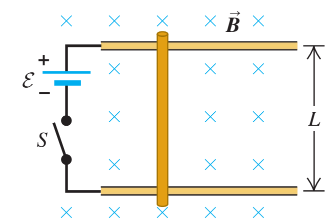
29.53 •• **CALC** A very long, cylindrical wire of radius R carries a current I_0 uniformly distributed across the cross section of the wire. Calculate the magnetic flux through a rectangle that has one side of length W running down the center of the wire and another side of length R, as shown in **Fig. P29.53** (see Exercise 29.7).

Figure **P29.53**



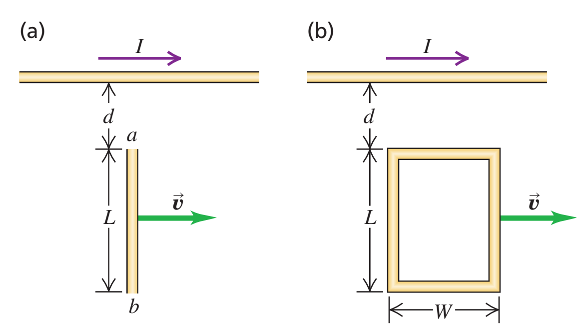
29.54 •• **CP** **CALC** **Terminal Speed.** A bar of length $L = 0.36\ \text{m}$ is free to slide without friction on horizontal rails as shown in **Fig. P29.54**. A uniform magnetic field $B = 2.4\ \text{T}$ is directed into the plane of the figure. At one end of the rails there is a battery with emf $\mathcal{E} = 12\ \text{V}$ 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 **P29.54**



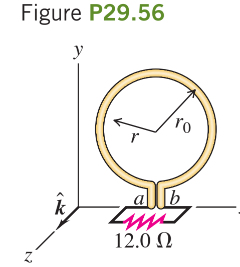
29.55 • **CALC** The long, straight wire shown in **Fig. P29.55a** carries constant current I. A metal bar with length L is moving at constant velocity $\vec{v}$, as shown in the figure. Point a is a distance d from the wire. (a) Calculate the emf induced in the bar. (b) Which point, a or b, is at higher potential? (c) If the bar is replaced by a rectangular wire loop of resistance R (Fig. P29.55b), what is the magnitude of the current induced in the loop?

Figure **P29.55**



29.56 • **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. P29.56**) 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 $t = 5.00 \times 10^{-3}\ \text{s}$. 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 $t = 1.21 \times 10^{-2}\ \text{s}$. What is the polarity of the emf? (e) Determine the time at which the current through R reverses its direction.

Figure **P29.56**

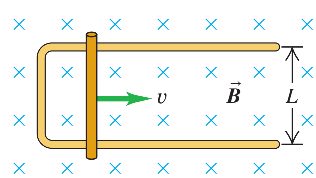


29.57 • CALC A slender rod, 0.240 m long, rotates with an angular speed of 8.80 rad/s about an axis through one end and perpendicular to the rod. The plane of rotation of the rod is perpendicular to a uniform magnetic field with a magnitude of 0.650 T. (a) What is the induced emf in the rod? (b) What 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?

29.58 •• 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.

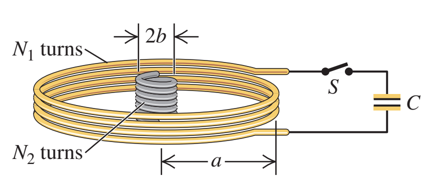
29.59 •• CP CALC A rectangular loop with width L and a slidewire with mass m are as shown in **Fig. P29.59**. A uniform magnetic field $\vec{B}$ is directed perpendicular to the plane of the loop into the plane of the figure. The slidewire is given an initial speed of v_0 and then released. There is no friction between the slidewire and the loop, and the resistance of the loop is negligible in comparison to the resistance R of the slidewire. (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 **P29.59**



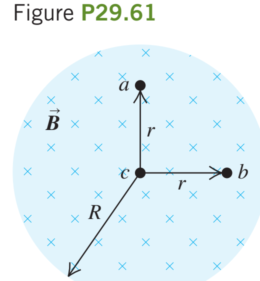
29.60 ••• CP A circular coil with $N_1 = 5000$ turns is made of a conducting material with resistance 0.0100 Ω/m and radius $a = 40.0$ cm. The coil is attached to a $C = 10.00\ \mu\text{F}$ capacitor as shown in **Fig. P29.60**. A second coil with radius $b = 4.00$ cm, made of the same wire, with $N_2 = 100$ turns, is concentric with the first coil and parallel to it. The capacitor has a charge of $+100\ \mu\text{C}$ on its upper plate, and the switch S is open. At time $t = 0$ the switch is closed. (a) What is the magnitude of the current in the larger coil immediately after the switch is closed? (b) What is the magnetic flux through each turn of the smaller coil immediately after the switch is closed? (Since $b << a$, we may treat the magnetic field in the smaller coil due to the larger coil as uniform.) (c) What is the direction of the current in the smaller coil immediately after the switch is closed? (d) What is the direction of the current in the smaller coil at $t = 1.26$ ms? (e) What is the magnitude of the current in the smaller coil at $t = 1.26$ ms?

Figure **P29.60**



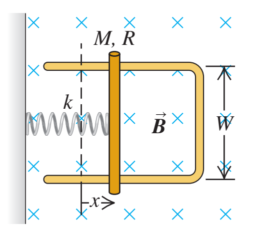
29.61 • The magnetic field $\vec{B}$, at all points within a circular region of radius R, is uniform in space and directed into the plane of the page as shown in **Fig. P29.61**. (The region could be a cross section inside the windings of a long, straight solenoid.) If the magnetic field is increasing 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 **P29.61**



29.62 ••• CP A bar with mass $M = 1.20$ kg and resistance $R = 0.500\ \Omega$ slides without friction on a horizontal U-shaped rail with width $W = 40.0$ cm and negligible resistance. The bar is attached to a spring with spring constant $k = 90.0$ N/m, as shown in **Fig. P29.62**. A constant magnetic field with magnitude 1.00 T points into the plane everywhere in the vicinity. At time $t = 0$ the bar is stretched beyond its equilibrium position by an amount $x = 10.0$ cm and released from rest. (a) This system behaves like a damped oscillator, described by Eq. (14.41). What is the damping coefficient b? (b) With what frequency does the bar oscillate around its equilibrium position? (c) What is the amplitude of the motion at time $t = 5.00$ s? (d) What is the magnitude of the current in the bar when it passes the equilibrium position for the first time? (e) What is the direction of that current?

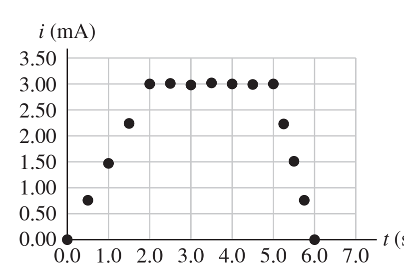
Figure **P29.62**



29.63 ••• 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?

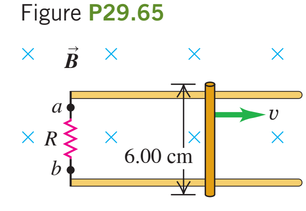
29.64 •• DATA You are evaluating the performance of a large electromagnet. The magnetic field of the electromagnet is zero at $t = 0$ and increases as the current through the windings of the electromagnet is increased. You determine the magnetic field as a function of time by measuring the time dependence 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. P29.64**. 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 **P29.64**



29.65 •• 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. P29.65**). 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 (next page):

Figure **P29.65**

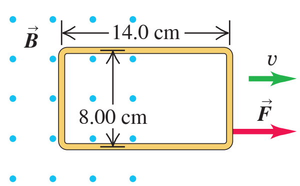


v (cm/s)	20.0	16.0	14.0	12.0	10.0	8.0
a (cm/s^2)	6.2	4.9	4.3	3.7	3.1	2.5

(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?

29.66 ••• 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. P29.66**. 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 **P29.66**



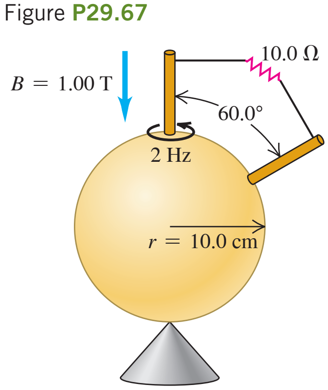
F (N)	0.10	0.21	0.31	0.41	0.52
v (cm/s)	2.0	4.0	6.0	8.0	10.0

(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. P29.66, 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

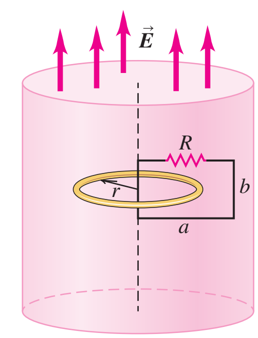
29.67 ••• CP A conducting spherical shell with radius 10.0 cm spins about a vertical axis twice every second in the presence of a constant magnetic field $\vec{B}$ with magnitude 1.00 T that points downward. Two conducting rods supported by a frame contact the sphere with conducting brushes and extend away from the sphere radially. One rod extends from the top of the sphere and the other forms a 60.0° angle with vertical, as shown in **Fig. P29.67**. The outer ends of the rods are connected to each other by a conducting wire that includes a 10.0 Ω resistor. (a) Construct a Cartesian coordinate system with the origin at the center of the sphere, the z-axis pointing upward, and the y-axis pointing rightward so that both rods lie in the yz-plane. What is the velocity $\vec{v}$ of a point on the sphere in the yz-plane at angle θ measured from the positive z-axis? (b) What is the vector product $\vec{v} \times \vec{B}$? (c) What is the line element $d\vec{l}$ along the shortest path on the sphere from the upper rod to the angled rod at angle θ? (d) What is the magnitude of the current in the wire? (e) What direction does the current flow in the vertical rod?

Figure **P29.67**



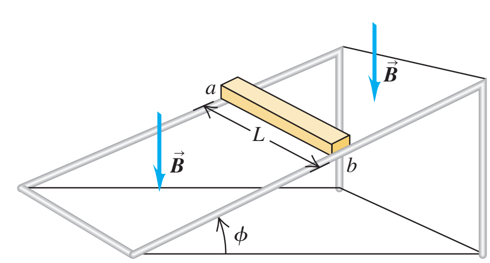
29.68 ••• CP A uniform electric field is directed axially in a cylindrical region that includes a rectangular loop of wire with total resistance R. This loop has radially oriented width a and axially oriented length b, and sits tight against the cylinder axis, as shown in **Fig. P29.68**. The electric field is zero at time $t = 0$ and then increases in time according to $\vec{E} = \eta t^2 \hat{k}$, where η is a constant with units of $\mathrm{V/(m \cdot s^2)}$. (a) What is the magnitude of the displacement current through a circular loop centered on the cylinder axis with radius $r \le a$, at time t? (b) Use Ampere's law to determine the magnitude of the magnetic field a distance $r \le a$ from the cylinder axis at time t. (c) What is the magnetic flux at time t through the rectangular wire loop? (d) What magnitude of current flows in the wire? (e) Does the current flow clockwise or counterclockwise from the perspective shown in the figure?

Figure **P29.68**



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**

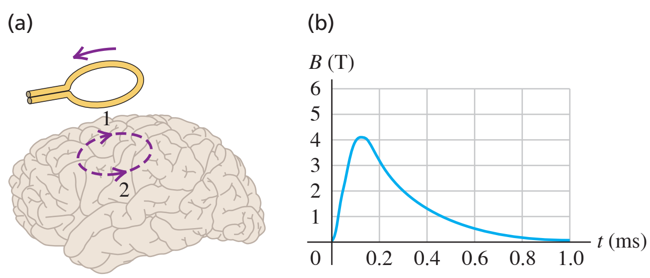


29.70 ••• CP CALC A square, conducting, wire loop of side L, total mass m, and total resistance R initially lies in the horizontal xy-plane, with corners at $(x, y, z) = (0, 0, 0)$, $(0, L, 0)$, $(L, 0, 0)$, and $(L, L, 0)$. There is a uniform, upward magnetic field $\vec{B} = B\hat{k}$ in the space 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.

MCAT-STYLE 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 MCAT-style 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.

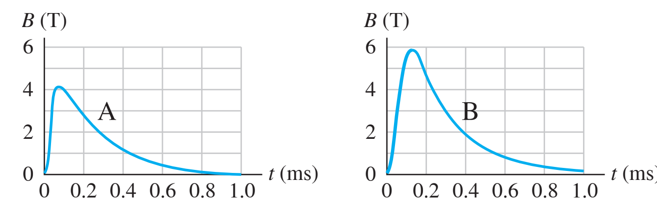


29.71 In part (a) of the figure, a current pulse increases to a peak and then decreases to zero in the direction shown in the stimulating coil. What will be the direction of the induced current (dashed line) in the brain tissue? (a) 1; (b) 2; (c) 1 while the current increases in the stimulating coil, 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.

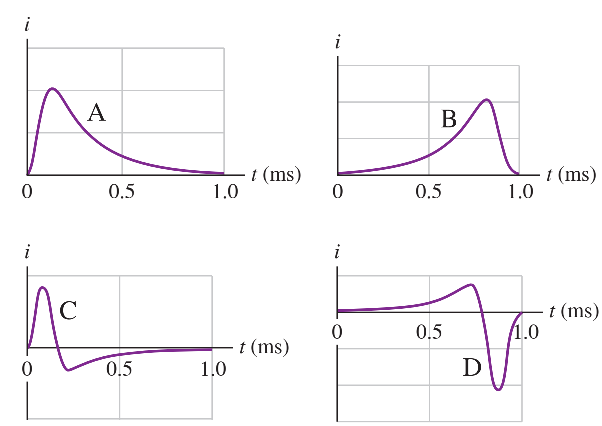
29.73 It may be desirable to increase the maximum induced current in the brain tissue. In **Fig. P29.73**, which time-dependent graph of the magnetic field B in the coil achieves that goal? Assume that everything else remains constant. (a) A; (b) B; (c) either A or B; (d) neither A nor B.

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.

Key Example VARIATION Problems

VP29.8.1 (a) $+2.53 \times 10^{-4}\ \mathrm{T \cdot m^2}$ at $t = 0$, $-1.99 \times 10^{-4}\ \mathrm{T \cdot m^2}$ at $t = 2.00$ s (b) 2.26×10^{-4} V
VP29.8.2 (a) 1.01×10^{-2} V (b) 6.95×10^{-4} A
VP29.8.3 (a) 0.163 V (b) counterclockwise
VP29.8.4 (a) $-\omega N B_0 A \cos \omega t$ (b) clockwise at $t = 0$, counterclockwise at $t = \pi/\omega$
VP29.9.1 (a) 0.187 V (b) 2.34 V/m, $+x$-direction
VP29.9.2 0.900 A, clockwise
VP29.9.3 (a) 2.40×10^{-2} V (b) end b
VP29.9.4 4.12 m/s
VP29.11.1 (a) 1.81×10^3 turns/meter (b) 7.70×10^{-5} V/m
VP29.11.2 2.70×10^{-5} V/m
VP29.11.3 (a) -7.77 A/s (b) $+11.7$ A/s
VP29.11.4 5.52×10^{-6} A

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 OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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'll 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. 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

Mutually induced emfs:

$$\underbrace{\mathcal{E}_2}_{\text{Induced emf in coil 2}} = -\underbrace{M}_{\text{Mutual inductance of coils 1 and 2}}\underbrace{\frac{di_1}{dt}}_{\text{Rate of change of current in coil 1}} \quad \text{and} \quad \underbrace{\mathcal{E}_1}_{\text{Induced emf in coil 1}} = -M\underbrace{\frac{di_2}{dt}}_{\text{Rate of change of current in coil 2}} \tag{30.4}$$

Figure **30.1** A current i_1 in coil 1 gives rise to a magnetic flux through coil 2. (For clarity, only one-half of each coil is shown.)

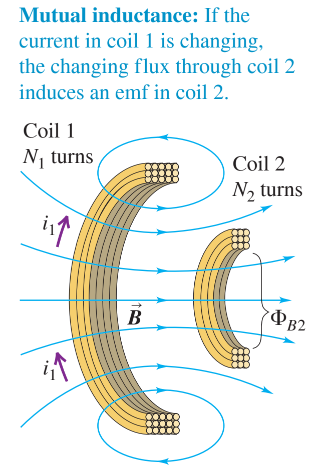
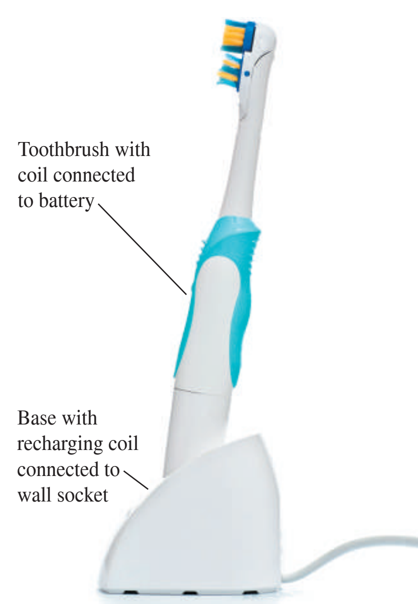


Figure **30.2** This electric toothbrush makes use of mutual inductance. The base contains a coil that is supplied with alternating current from a wall socket. Even though there is no direct electrical contact between the base and the toothbrush, this varying current induces an emf in a coil within the toothbrush itself, recharging the toothbrush battery.



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.)

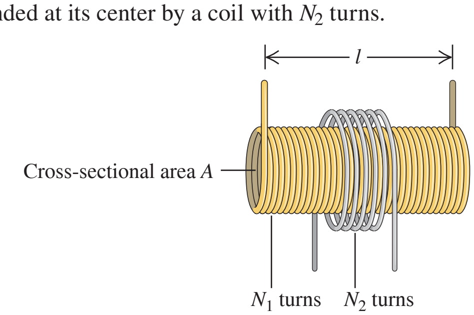
EXAMPLE 30.1 Calculating mutual inductance

WITH VARIATION PROBLEMS

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.

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.

Figure 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 $B_1 A$. 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

$$M = \frac{N_2 \Phi_{B2}}{i_1} = \frac{N_2 B_1 A}{i_1} = \frac{N_2}{i_1}\frac{\mu_0 N_1 i_1}{l}A = \frac{\mu_0 A N_1 N_2}{l}$$

EVALUATE The mutual inductance M of any two coils is proportional to the product $N_1 N_2$ 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

$$M = \frac{(4\pi \times 10^{-7}\ \text{Wb/A}\cdot\text{m})(1.0 \times 10^{-3}\ \text{m}^2)(1000)(10)}{0.50\ \text{m}}$$

$$= 25 \times 10^{-6}\ \text{Wb/A} = 25 \times 10^{-6}\ \text{H} = 25\ \mu\text{H}$$

KEYCONCEPT A current in one coil causes a magnetic flux through an adjacent coil. How much flux is caused in the other coil by a given current in the first coil is determined by the mutual inductance M of the two coils. The value of M doesn't depend on the currents in the coils, only on their geometry.

EXAMPLE 30.2 Emf due to mutual inductance

WITH VARIATION PROBLEMS

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 $t = 3.0\ \mu\text{s}$, 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?

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 $M = 25\ \mu\text{H}$ 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:

$$\Phi_{B1} = \frac{M i_2}{N_1} = \frac{(25 \times 10^{-6}\ \text{H})(6.0\ \text{A})}{1000} = 1.5 \times 10^{-7}\ \text{Wb}$$

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.

KEYCONCEPT When the current in a coil changes, it causes a changing magnetic flux and hence an induced emf in an adjacent coil. The mutual inductance of the two coils determines how much emf is induced in one coil for a given rate of change of the current in the other coil.

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.

ANSWER

| **(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.

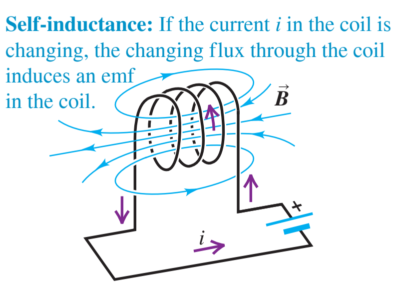
30.2 SELF-INDUCTANCE AND INDUCTORS

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

Figure 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

Current in coil

$$L = \frac{N\Phi_B}{i} \tag{30.6}$$

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; after we rearrange Eq. (30.6) and differentiate 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

Rate of change of current in circuit

$$\mathcal{E} = -L\frac{di}{dt} \tag{30.7}$$

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.

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.

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

$$V_{ab} = V_a - V_b = L\frac{di}{dt} \tag{30.8}$$

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.

Figure **30.5** A circuit containing a source of emf and an ideal inductor. The source is variable, so the current i and its rate of change di/dt can be varied.

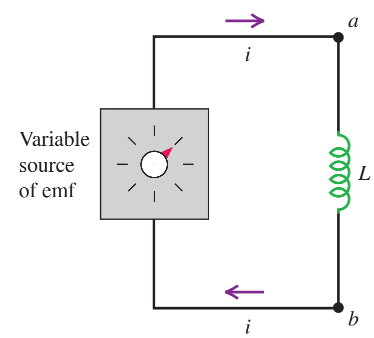
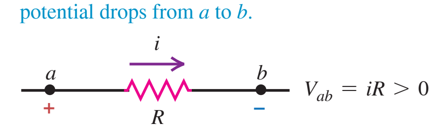
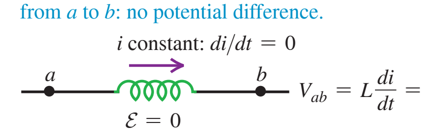


Figure **30.6** (a) The potential difference across a resistor depends on the current, whereas (b), (c), (d) the potential difference across an inductor depends on the rate of change of the current.

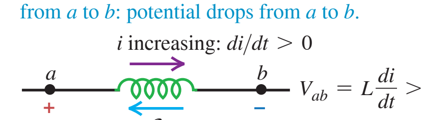
(a) Resistor with current i flowing from a to b: potential drops from a to b.



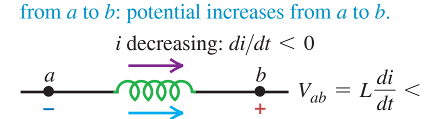
(b) Inductor with *constant* current i flowing from a to b: no potential difference.



(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

Figure 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'll 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'll 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.

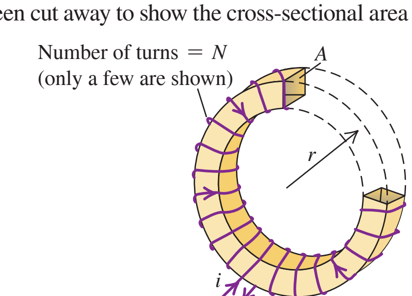
Because automobiles contain steel, a ferromagnetic material, driving an automobile over a coil causes an appreciable increase in the coil's inductance. This effect is used in traffic light sensors, which use a large, current-carrying coil embedded under the road surface near an intersection. The circuitry 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 — WITH VARIATION PROBLEMS

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).

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 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.

Figure 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.



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

$$L = \frac{(4\pi \times 10^{-7}\ \text{Wb/A}\cdot\text{m})(200)^2(5.0 \times 10^{-4}\ \text{m}^2)}{2\pi(0.10\ \text{m})}$$
$$= 40 \times 10^{-6}\ \text{H} = 40\ \mu\text{H}$$

KEYCONCEPT The self-inductance L of a coil (or inductor) tells you how much magnetic flux is produced through that coil by the current that flows in it. Like mutual inductance, the value of L depends only on the coil's geometry, not on how much current it carries.

EXAMPLE 30.4 Calculating self-induced emf

WITH VARIATION PROBLEMS

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.

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.

KEYCONCEPT If the current in a coil changes, the resulting change in magnetic flux through the coil produces an induced emf. The self-inductance of the coil determines how much emf is induced in the coil for a given rate of change of the current in the coil.

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.

ANSWER

(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 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

Inductance

Energy stored in an inductor

$$U = L\int_0^I i\,di = \tfrac{1}{2}LI^2 \qquad (30.9)$$

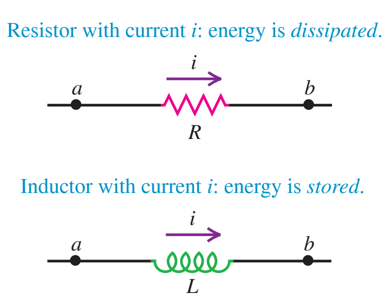
Final current

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.

Figure **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.



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.

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'll 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 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 $V = 2\pi rA$ enclosed by the windings. The energy *per unit volume,* or *magnetic energy density,* is $u = U/V$:

$$u = \frac{U}{2\pi rA} = \tfrac{1}{2}\mu_0\frac{N^2I^2}{(2\pi r)^2}$$

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

$$B = \frac{\mu_0 NI}{2\pi r}$$

and so

$$\frac{N^2I^2}{(2\pi r)^2} = \frac{B^2}{\mu_0^2}$$

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

$$u = \frac{\overbrace{B^2}^{\text{Magnetic-field magnitude}}}{\underbrace{2\mu_0}_{\text{Magnetic constant}}} \qquad (30.10)$$

This is the magnetic analog of the energy per unit volume in an *electric* field in vacuum, $u = \frac{1}{2}\epsilon_0 E^2$, 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 $u = B^2/2\mu_0 = (1.5\ \text{T})^2/(2 \times 4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A}) = 9.0 \times 10^5\ \text{J/m}^3$.

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

$$u = \frac{\overbrace{B^2}^{\text{Magnetic-field magnitude}}}{\underbrace{2\mu}_{\text{Permeability of material}}} \qquad (30.11)$$

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'll 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.

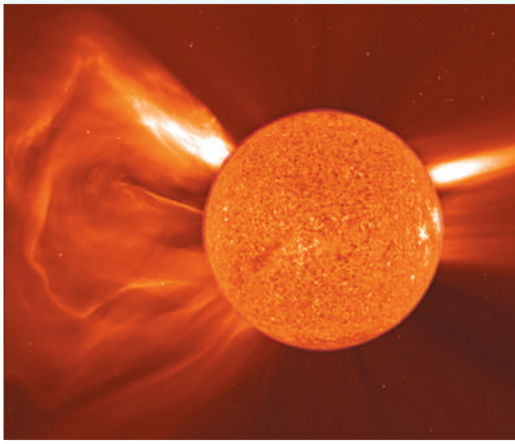


Figure 30.10 The energy required to fire an automobile spark plug is derived from magnetic-field energy stored in the ignition coil.

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?

IDENTIFY and SET UP We are given the required amount of stored energy U and the current $I = 200\ \text{A}$. 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 can be much smaller, since the resistance of a superconductor is zero and much thinner wires can be used. However, such wires must be kept at low temperature to remain superconducting, and maintaining this temperature itself requires energy. Despite these limitations, superconducting inductors are very efficient for energy storage and are presently in small-scale use by several electric utilities.

KEYCONCEPT Energy is stored in any magnetic field, including the field of a current-carrying inductor. The magnetic energy of an inductor depends on its self-inductance and current.

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?

ANSWER

(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 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 $\mathcal{E}$.)

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 30.6c.) We apply Kirchhoff's loop rule, starting at the negative terminal and proceeding counterclockwise around the loop:

$$\mathcal{E} - iR - L\frac{di}{dt} = 0 \qquad (30.12)$$

Solving this for di/dt, we find that the rate of increase of current is

$$\frac{di}{dt} = \frac{\mathcal{E} - iR}{L} = \frac{\mathcal{E}}{L} - \frac{R}{L}i \qquad (30.13)$$

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

$$\frac{di}{i - (\mathcal{E}/R)} = -\frac{R}{L}dt$$

Figure **30.11** An *R-L* circuit.

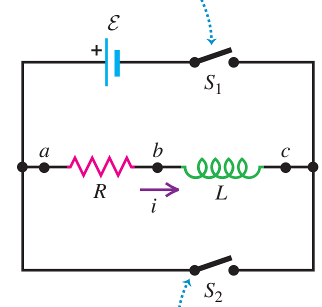
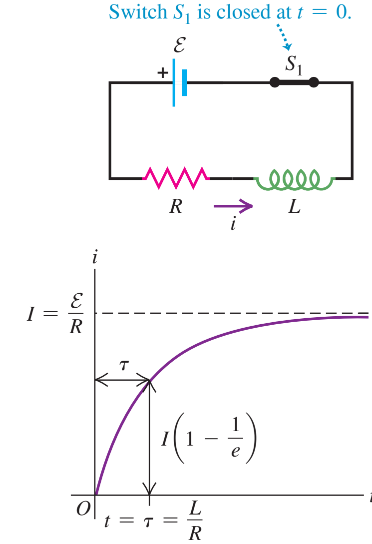


Figure **30.12** Graph of i versus t for growth of current in an *R-L* circuit with an emf in series. The final current is $I = \mathcal{E}/R$; after one time constant τ, the current is $1 - 1/e$ of this value.



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:

$$i = \frac{\mathcal{E}}{R}\left(1 - e^{-(R/L)t}\right) \quad \text{(current in an } R\text{-}L \text{ circuit with emf)} \tag{30.14}$$

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:

CAUTION Current and its rate of change in an inductor Note that the current i cannot change abruptly in a circuit that contains an inductor, so i must be a continuous function of time t. However, the rate of change of the current di/dt *can* change abruptly (for example, when the switch in Fig. 30.12a is closed and the current suddenly starts to increase, as shown at $t = 0$ in Fig. 30.12b).

$$\underbrace{\tau}_{\text{Time constant for an } R\text{-}L \text{ circuit}} = \frac{\overbrace{L}^{\text{Inductance}}}{\underbrace{R}_{\text{Resistance}}} \tag{30.16}$$

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, $\tau = 1.0 \times 10^{-4}$ 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

WITH VARIATION PROBLEMS

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?

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.

KEYCONCEPT The presence of an inductor in a circuit that contains a source of emf and a resistor slows the growth of the current when the circuit is completed. The equilibrium value of the current after a very long time is the same as if the inductor were not present.

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%.

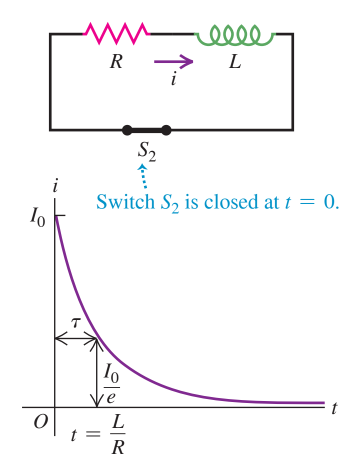
The energy that is needed to maintain the current during this decay is provided by the energy stored in the magnetic field of the inductor. The detailed energy analysis is simpler this time. In place of Eq. (30.17) we have

$$0 = i^2R + Li\frac{di}{dt} \tag{30.19}$$

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.

Figure 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.



EXAMPLE 30.7 Energy in an *R-L* circuit

WITH VARIATION PROBLEMS

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?

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 $L = 69$ mH and $I_0 = 36$ mA, we have

$$U_0 = \tfrac{1}{2}LI_0^2 = \tfrac{1}{2}(0.069\ \text{H})(0.036\ \text{A})^2$$
$$= 4.5 \times 10^{-5}\ \text{J}$$

Of this, 99.0% or 4.4×10^{-5} J is dissipated in $2.3(390\ \mu\text{s}) = 9.0 \times 10^{-4}\ \text{s} = 0.90$ 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.

KEYCONCEPT The presence of an inductor in a circuit that contains a resistor slows the decay of the current when the source of emf is removed. As the current decays, the magnetic energy stored in the inductor is dissipated in the resistor.

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) $v_{ab} < 0, v_{bc} < 0$.

ANSWER

(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 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 **30.14** 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.)

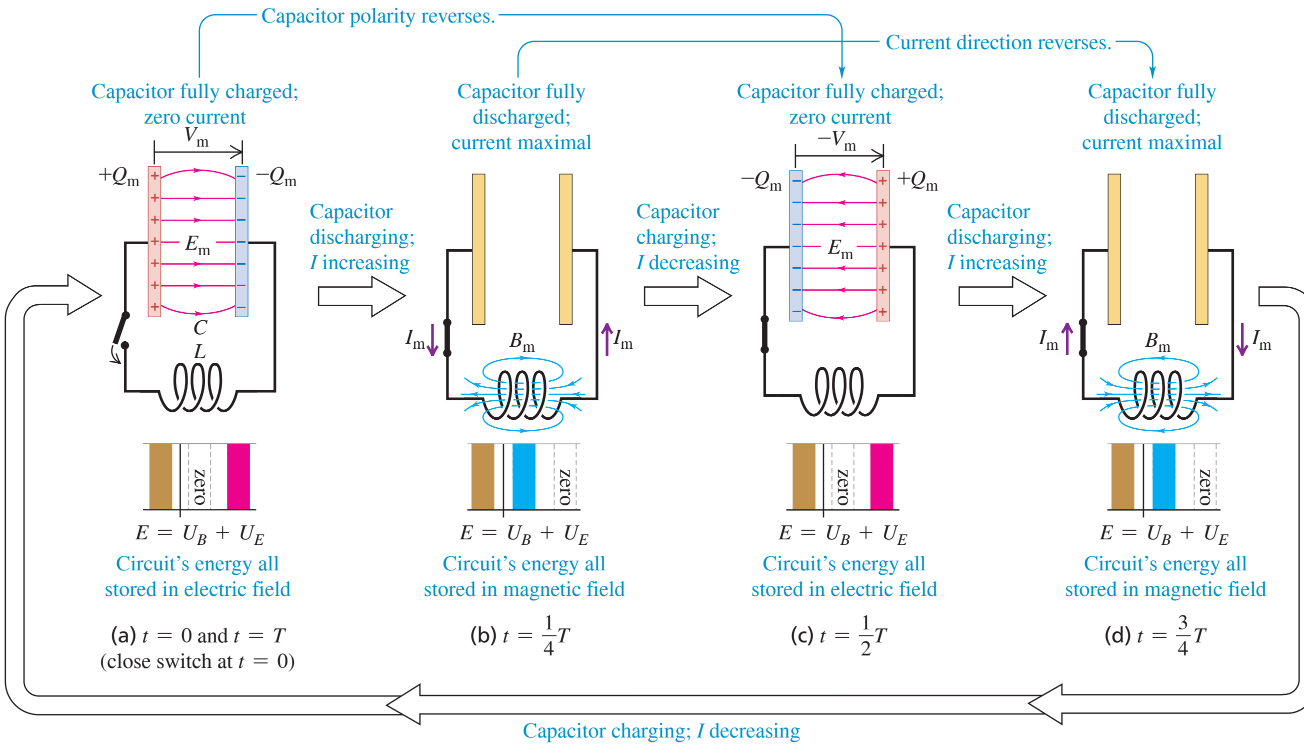


Figure 30.14b shows this situation; the capacitor has completely discharged. The potential difference between its terminals (and those of the inductor) has decreased to zero, and the current has reached its maximum value I_m.

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 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.

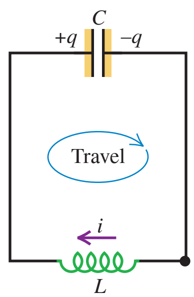
The process now repeats in the reverse direction; a little later, the capacitor has again discharged, and there is a current in the inductor in the opposite direction (Fig. 30.14d). Still later, the capacitor charge returns to its original value (Fig. 30.14a), and the whole process repeats. If there 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'll 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.

Figure 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).



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.

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

$$\frac{d^2q}{dt^2} + \frac{1}{LC}q = 0 \qquad (L\text{-}C \text{ circuit}) \qquad (30.20)$$

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

$$q = Q\cos(\omega t + \phi) \qquad (30.21)$$

and the angular frequency ω of oscillation is given by

$$\underbrace{\omega}_{\text{Angular frequency of oscillation in an } L\text{-}C \text{ circuit}} = \sqrt{\frac{1}{\underbrace{L}_{\text{Inductance}}\underbrace{C}_{\text{Capacitance}}}} \qquad (30.22)$$

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'll find that the instantaneous current $i = dq/dt$ is

$$i = -\omega Q\sin(\omega t + \phi) \qquad (30.23)$$

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 an object with mass m is attached to a spring with force constant k. Suppose we displace the object 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 object, 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'll 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

WITH VARIATION PROBLEMS

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.

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)].

Continued

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

$$f = \frac{\omega}{2\pi} = \frac{2.0 \times 10^3\ \text{rad/s}}{2\pi\ \text{rad/cycle}} = 320\ \text{Hz}$$

$$T = \frac{1}{f} = \frac{1}{320\ \text{Hz}} = 3.1 \times 10^{-3}\ \text{s} = 3.1\ \text{ms}$$

(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 $Q = C\mathcal{E} = (25 \times 10^{-6}\ \text{F})(300\ \text{V}) = 7.5 \times 10^{-3}$ 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.

KEYCONCEPT An L-C circuit that contains an inductor and a capacitor (and no resistance) is analogous to an ideal mass-spring system. The capacitor charge and the current in the circuit both oscillate sinusoidally, just like the displacement and the velocity for a mass on a spring.

EXAMPLE 30.9 Energy in an oscillating circuit

WITH VARIATION PROBLEMS

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.

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.

KEYCONCEPT In an L-C circuit, the inductor is analogous to the mass in an ideal mass-spring system and the capacitor is analogous to the spring. Magnetic energy in the inductor behaves like the kinetic energy of the mass; electrical energy in the capacitor behaves like the potential energy of the spring.

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?

ANSWER

(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).

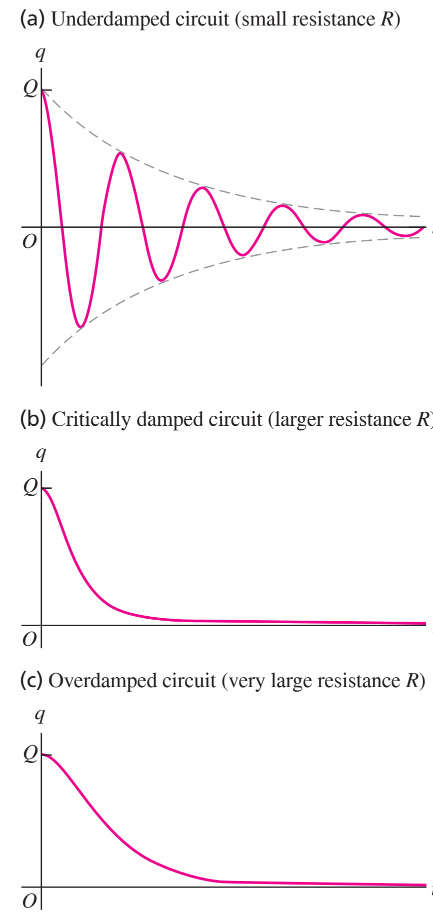
30.6 THE *L-R-C* SERIES CIRCUIT

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.

Figure **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.

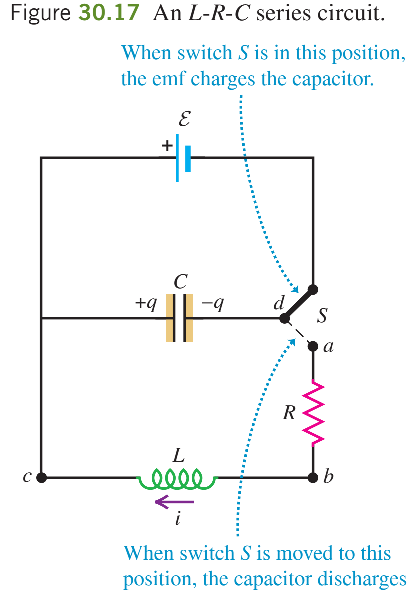
There are general methods for obtaining solutions of Eq. (30.27). The form of the solution is different for the underdamped (small R) and overdamped (large R) cases. When R^2 is less than $4L/C$, the solution has the form

$$q = Ae^{-(R/2L)t}\cos\left(\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\,t + \phi\right) \qquad (30.28)$$

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)

Figure **30.17** An *L-R-C* series circuit.



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

$$\omega' = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \tag{30.29}$$

Resistance (R); Inductance ($4L^2$); Inductance (L in LC); Capacitance (C)

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 30.38).

We emphasize once more that the behavior of the L-R-C series circuit is completely analogous to that of the damped harmonic oscillator (see 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

WITH VARIATION PROBLEMS

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?

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 Ω to the circuit of Example 30.8 ($L = 10$ mH, $C = 25\ \mu$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.

KEYCONCEPT A series L-R-C circuit is analogous to a mass-spring system with damping. Increasing the resistance reduces the oscillation frequency of the charge and current; if the resistance is too great, the circuit is critically damped or overdamped and there is no oscillation.

TEST YOUR UNDERSTANDING OF SECTION 30.6 An L-R-C series circuit includes a 2.0 Ω resistor. At $t = 0$ the capacitor charge is 2.0 μC. For which of the following values of the inductance and capacitance will the charge on the capacitor *not* oscillate? (i) $L = 3.0\ \mu$H, $C = 6.0\ \mu$F; (ii) $L = 6.0\ \mu$H, $C = 3.0\ \mu$F; (iii) $L = 3.0\ \mu$H, $C = 3.0\ \mu$F.

ANSWER

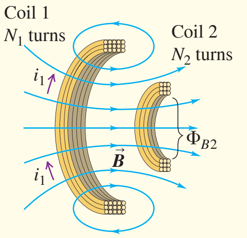
(i) and (iii) There are no oscillations if $R^2 \geq 4L/C$. In each case $R^2 = (2.0\ \Omega)^2 = 4.0\ \Omega^2$. In case (i) $4L/C = 4(3.0\ \mu\text{H})/(6.0\ \mu\text{F}) = 2.0\ \Omega^2$, so there are no oscillations (the system is overdamped); in case (ii) $4L/C = 4(6.0\ \mu\text{H})/(3.0\ \mu\text{F}) = 8.0\ \Omega^2$, 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).

CHAPTER 30 SUMMARY

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.)

$$\mathcal{E}_2 = -M\frac{di_1}{dt} \quad \text{and} \quad \mathcal{E}_1 = -M\frac{di_2}{dt} \tag{30.4}$$

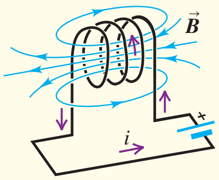
$$M = \frac{N_2\Phi_{B2}}{i_1} = \frac{N_1\Phi_{B1}}{i_2} \tag{30.5}$$



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.)

$$\mathcal{E} = -L\frac{di}{dt} \tag{30.7}$$

$$L = \frac{N\Phi_B}{i} \tag{30.6}$$

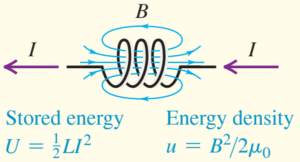


Magnetic-field energy: An inductor with inductance L carrying current I has energy U associated with the inductor's magnetic field. The magnetic energy density u (energy per unit volume) is proportional to the square of the magnetic-field magnitude. (See Example 30.5.)

$$U = \tfrac{1}{2}LI^2 \tag{30.9}$$

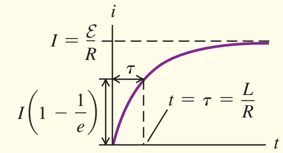
$$u = \frac{B^2}{2\mu_0} \quad \text{(in vacuum)} \tag{30.10}$$

$$u = \frac{B^2}{2\mu} \quad \text{(in a material with magnetic permeability } \mu) \tag{30.11}$$



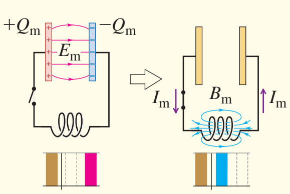
***R-L* circuits:** In a circuit containing a resistor R, an inductor L, and a source of emf, the growth and decay of current are exponential. The time constant τ is the time required for the current to approach within a fraction $1/e$ of its final value. (See Examples 30.6 and 30.7.)

$$\tau = \frac{L}{R} \tag{30.16}$$



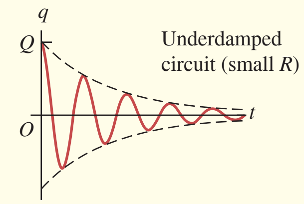
***L-C* circuits:** A circuit that contains inductance L and capacitance C undergoes electrical oscillations with an angular frequency ω that depends on L and C. This is analogous to a mechanical harmonic oscillator, with inductance L analogous to mass m, the reciprocal of capacitance $1/C$ to force 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.)

$$\omega' = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \tag{30.29}$$



Chapter 30 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 30.1 and 30.2 (Section 30.1) and EXAMPLES 30.3 and 30.4 (Section 30.2) before attempting these problems.

VP30.4.1 A long solenoid (coil 1) is surrounded at its center by a shorter coil (coil 2) as in Fig. 30.3. At an instant when the current in coil 1 is 24.0 A and increasing at a rate of 2.00 A/s, the induced emf in coil 2 is 1.30×10^{-4} V. (a) What is the mutual inductance of the two coils? (b) At this instant, what is the total magnetic flux through all the turns of coil 2 due to the current in coil 1?

VP30.4.2 The mutual inductance of two coils (1 and 2) is 48.0 μH. Find the magnitude of the emf induced in coil 2 if the current in coil 1 is (a) 35.0 A and increasing at 7.00 A/s; (b) 28.0 A and constant; (c) 26.0 A and decreasing at 3.00 A/s.

VP30.4.3 A toroidal solenoid with self-inductance 76.0 μH has 465 turns of wire. Find (a) the magnetic flux through each turn when the current in the solenoid is 12.0 A and (b) the magnitude of the induced emf in the solenoid when the current is 12.0 A and decreasing at 55.0 A/s.

VP30.4.4 When a coil is carrying a current of 25.0 A that is increasing at 145 A/s, the induced emf in the coil has magnitude 3.70 mV. Find (a) the self-inductance of the coil and (b) the magnitude and direction of the induced emf at $t = 2.00$ s if the current is $i(t) = (225\ \text{A/s}^2)t^2$.

Be sure to review EXAMPLES 30.6 and 30.7 (Section 30.4) before attempting these problems.

VP30.7.1 In the circuit shown in Fig. 30.11, the inductance is 4.90 mH and the emf is 24.0 V. Initially both switches are open. You want the current to reach 75.0% of its maximum value 80.0 μs after switch S_1 is closed. Find (a) the required resistance and (b) the current 80.0 μs after switch S_1 is closed.

VP30.7.2 In the circuit shown in Fig. 30.11, $\mathcal{E} = 12.0$ V, $R = 18.0\ \Omega$, and $L = 37.5$ mH. Initially both switches are open. You close switch S_1 at $t = 0$. At $t = 0.700$ ms, find (a) the current, (b) the rate of change of the current, and (c) the amount of energy stored in the inductor.

VP30.7.3 In the circuit shown in Fig. 30.11, $\mathcal{E} = 12.0$ V, $R = 18.0\ \Omega$, and $L = 37.5$ mH. Switch S_1 has been closed for a long time and switch S_2 is open. At $t = 0$ you open switch S_1 and close switch S_2. At $t = 0.700$ ms, find (a) the current, (b) the rate of change of the current, and (c) the amount of energy stored in the inductor.

VP30.7.4 For the R-L circuit with an increasing current shown in Fig. 30.12, find (a) the time when the voltage across the inductor has the same magnitude as the voltage across the resistor, (b) the current at this time, and (c) the rate of change of the current at this time.

Be sure to review EXAMPLES 30.8 and 30.9 (Section 30.5) and EXAMPLE 30.10 (Section 30.6) before attempting these problems.

VP30.10.1 The L-C circuit in Fig. 30.15 contains an inductor with $L = 25.0\ \mu$H. The charge and current in this circuit oscillate with frequency 445 Hz, and the maximum capacitor charge is 5.00 mC. Find (a) the capacitance of the capacitor and (b) the maximum current in the circuit.

VP30.10.2 You construct an L-C circuit that contains a 30.0 μH inductor and a 655 μF capacitor. At $t = 0$ the capacitor has its maximum charge of 0.400 mC. Find (a) the first time after $t = 0$ when the capacitor charge is zero and (b) the magnitude of the current in the circuit at this time.

VP30.10.3 The sum of the electrical and magnetic energies in an L-C circuit is 0.800 J. At a certain instant the energy is exactly half electrical and half magnetic, the capacitor charge is 5.30 mC, and the current is 8.00 A. Find (a) the capacitance, (b) the inductance, and (c) the angular frequency of oscillation.

VP30.10.4 An L-R-C series circuit has inductance 42.0 mH, capacitance C, and resistance R. Without the resistor, the angular frequency of oscillation is 624 rad/s. With the resistor, the angular frequency is 208 rad/s. Find the values of (a) C and (b) R.

BRIDGING PROBLEM Analyzing an L-C Circuit

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

•, ••, •••: 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.

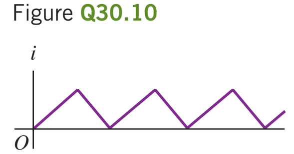
Q30.7 You are to make a resistor by winding a wire around a cylindrical form. To make the inductance as small as possible, it is proposed that you wind half the wire in one direction and the other half in the opposite direction. Would this achieve the desired result? Why or why not?

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 • 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?

30.4 • 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.

Section 30.2 Self-Inductance and Inductors

30.5 • 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?

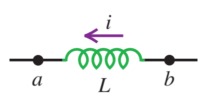
30.6 • A toroidal solenoid has 500 turns, cross-sectional area 6.25 cm^2, and mean radius 4.00 cm. (a) Calculate the coil's self-inductance. (b) If the current decreases uniformly from 5.00 A to 2.00 A in 3.00 ms, calculate the self-induced emf in the coil. (c) The current is directed from terminal a of the coil to terminal b. Is the direction of the induced emf from a to b or from b to a?

30.7 • 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?

30.8 •• 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?

30.9 • The inductor in **Fig. E30.9** 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 **E30.9**



30.10 • The inductor shown in Fig. E30.9 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?

30.11 •• **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 $L = \mu_0 A N^2/l$. 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?

30.12 • 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?

Section 30.3 Magnetic-Field Energy

30.13 •• When the current in a long, straight, air-filled solenoid is changing at the rate of 2000 A/s, the voltage across the solenoid is 0.600 V. The solenoid has 1200 turns and uniform cross-sectional area 25.0 mm^2. Assume that the magnetic field is uniform inside the solenoid and zero outside, so the result $L = \mu_0 A N^2/l$ (see Exercise 30.11) applies. What is the magnitude B of the magnetic field in the interior of the solenoid when the current in the solenoid is 3.00 A?

30.14 • 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.

30.15 • 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?

30.16 • 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.

30.17 •• 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.

30.18 • 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?

30.19 •• 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?

30.20 • A region of vacuum contains both a uniform electric field with magnitude E and a uniform magnetic field with magnitude B. (a) What is the ratio E/B if the energy density for the magnetic field equals the energy density for the electric field? (b) If $E = 500$ V/m, what is B, in teslas, if the magnetic-field and electric-field energy densities are equal?

Section 30.4 The *R-L* Circuit

30.21 • 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.

30.22 • In Fig. 30.11, $R = 15.0$ Ω 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?

30.23 • 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?

30.24 •• A resistor and an inductor are connected in series to a battery with emf 240 V and negligible internal resistance. The circuit is completed at time $t = 0$. At a later time $t = T$ the current is 5.00 A and is increasing at a rate of 20.0 A/s. After a long time the current in the circuit is 15.0 A. What is the value of T, the time when the current is 5.00 A?

30.25 • A resistor with $R = 30.0$ Ω and an inductor with $L = 0.600$ H are connected in series to a battery that has emf 50.0 V and negligible internal resistance. At time t after the circuit is completed, the energy stored in the inductor is 0.400 J. At this instant, what is the voltage across the inductor?

30.26 • In the circuit shown in Fig. 30.11 switch S_1 has been closed a long time while switch S_2 has been left open. Then S_2 is closed at the same instant when S_1 is opened. Just after S_2 is closed, the current through the resistor is 12.0 A and its rate of decrease is $di/dt = -36.0$ A/s. How long does it take the current to decrease to 6.00 A, one-half its initial value?

30.27 • In Fig. 30.11, suppose that $\mathcal{E} = 60.0$ V, $R = 240$ Ω, 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 30.21. (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?

30.31 • Consider the rate P_L at which energy is being stored in the R-L circuit of Fig. 30.12. Answer these questions, in terms of $\mathcal{E}$, R, and L as needed: (a) What is P_L at $t = 0$, just after the circuit is completed? (b) What is P_L at $t \rightarrow \infty$, a long time after the circuit is completed? (c) What is P_L at the instant when $i = \mathcal{E}/2R$, one-half the current's final value?

Section 30.5 The *L-C* Circuit

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 • 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?

30.34 • A 7.50 nF capacitor is charged 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.

30.35 • ***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 • 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?

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?

PROBLEMS

30.41 •• It is possible to make your own inductor by winding wire around a cylinder, such as a pencil. Assume you have a spool of AWG 20 copper wire, which has a diameter of 0.812 mm. (a) Estimate the diameter of a pencil. (b) Estimate how many times can you tightly wrap AWG 20 copper wire around a pencil to form a solenoid with a length of 4.0 cm. (c) Estimate the inductance of this solenoid by assuming the magnetic field inside is constant. (d) If a current of 1.0 A flows through this solenoid, how much magnetic energy will be stored inside?

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 •• **CP** Consider a coil of wire that has radius 3.00 cm and carries a sinusoidal current given by $i(t) = I_0 \sin(2\pi f t)$, where the frequency $f = 60.0$ Hz and the initial current $I_0 = 1.20$ A. (a) Estimate the magnetic flux through this coil as the product of the magnetic field at the center of the coil and the area of the coil. Use this magnetic flux to estimate the self-inductance L of the coil. (b) Use the value of L that you estimated in part (a) to calculate the magnitude of the maximum emf induced in the coil.

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.

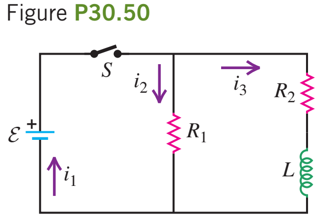
30.47 • (a) What would have to be the self-inductance of a solenoid for it to store 10.0 J of energy when a 2.00 A current runs through it? (b) If this solenoid's cross-sectional diameter is 4.00 cm, and if you could wrap its coils to a density of 10 coils/mm, how long would the solenoid be? (See Exercise 30.11.) Is this a realistic length for ordinary laboratory use?

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 • **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?

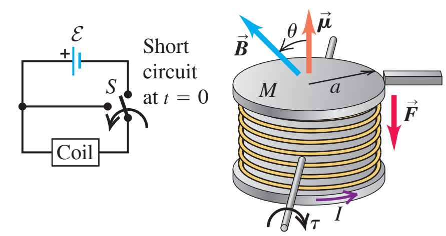
30.50 •• **CALC** An inductor with inductance $L = 0.300$ H and negligible resistance is connected to a battery, a switch S, and two resistors, $R_1 = 12.0\ \Omega$ and $R_2 = 16.0\ \Omega$ (**Fig. P30.50**). The battery has emf 96.0 V and negligible internal resistance. S is closed at $t = 0$. (a) What 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**



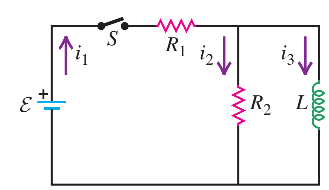
30.51 ••• **CP** An alternating-current electric motor includes a thin, hollow, cylindrical spool (similar to a ring) with mass $M = 1.11$ kg and radius $a = 5.00$ cm wrapped $N = 500$ times with a copper wire with resistance $R = 5.00\ \Omega$ and inductance $L = 77.0$ mH. Within the spool is a battery that supplies current $I = 1.00$ A, which makes the spool a magnetic dipole with dipole moment $\vec{\mu}$ parallel to the cylinder axis. A constant magnetic field with magnitude $B = 2.00$ T is supplied by an external stator magnet, while the spool turns freely on an axis perpendicular to its own axis. At a certain time, a bar is inserted, stopping the spool's motion (**Fig. P30.51**). At that instant the angle between the spool axis and the magnetic field is $\theta = 45°$. (a) What is the magnitude of the downward force $\vec{F}$ applied by the bar onto the spool immediately after the bar is inserted? (b) Later, at time $t = 0$ with spool still at rest, the coil is short-circuited and a constant counter-torque $\tau = 0.500$ N·m is applied. The current subsides, and the magnetic torque decreases exponentially. At what time t does the force applied by the bar vanish? (*Hint:* Determine when the magnetic torque balances the counter-torque.) (c) After the spool rotates 180° it becomes stuck on the top side of the bar. The counter-torque is no longer applied, and the switch is returned to its original position. After a long time, the bar is removed. What is the angular acceleration of the spool immediately after the bar is removed? The moment of inertia of the spool for an axis along its diameter is $I = \frac{1}{2}Ma^2$.

Figure **P30.51**



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 ••• **CP CALC** A cylindrical solenoid with radius 1.00 cm and length 10.0 cm consists of 300 windings of AWG 20 copper wire, which has a resistance per length of 0.0333 Ω/m. This solenoid is connected in series with a 10.0 μF capacitor, which is initially uncharged. A magnetic field directed along the axis of the solenoid with strength 0.100 T is switched on abruptly. (a) The solenoid may be considered an inductor and a resistor in series. Use Faraday's law to determine the average emf across the solenoid during the brief switch-on interval, and determine the net charge initially deposited on the capacitor. (See Exercise 29.4.) (b) At time $t = 0$ the capacitor is fully charged and there is no current. How much time does it take for the capacitor to fully discharge the first time? (c) What is the frequency with which the current oscillates? (d) How much energy is stored in the capacitor at $t = 0$? (e) How long does it take for the total energy stored in the circuit to drop to 10% of that value?

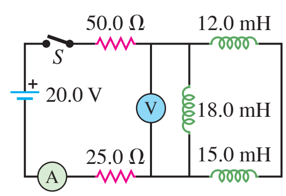
30.54 •• A 6.40 nF capacitor is charged to 24.0 V and then disconnected from the battery in the circuit and connected in series with a coil that has $L = 0.0660$ H and negligible resistance. After the circuit has been completed, there are current oscillations. (a) At an instant when the charge of the 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?

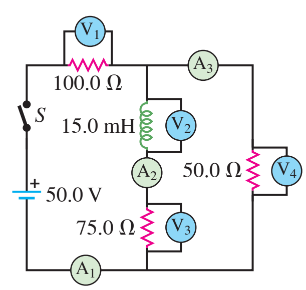
30.57 ••• **CP** In the circuit shown in **Fig. P30.57**, the switch S has been open for a long time and is suddenly closed. Neither the battery nor the inductors have any appreciable resistance. What do the ammeter and the voltmeter read (a) just after S is closed; (b) after S has been closed a very 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**



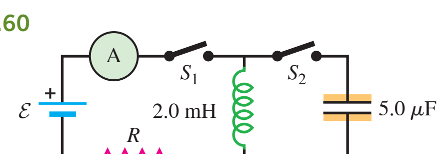
30.59 •• **DATA** To investigate the properties of a large industrial solenoid, you connect the solenoid and a resistor in series with a battery. Switches allow the battery to be replaced by a short circuit across the solenoid and resistor. Therefore Fig. 30.11 applies, with $R = R_{ext} + R_L$, where 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?

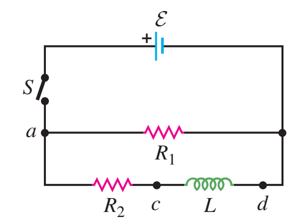
30.60 •• In the circuit shown in **Fig. P30.60**, switch S_1 has been closed for a long enough time so that the current reads a steady 3.50 A. Suddenly, switch S_2 is closed and S_1 is opened at the same instant. (a) What is the maximum charge that the capacitor will receive? (b) What is the current 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?

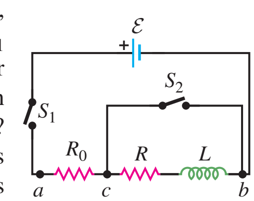
Figure **P30.61**



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.

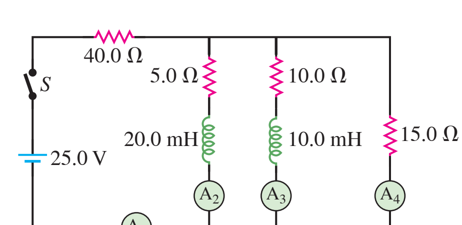
Figure **P30.63**



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.

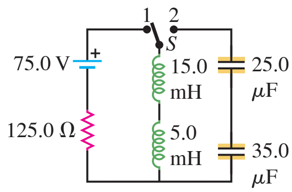
30.65 •• **CP** In the circuit shown in **Fig. P30.65**, switch S is closed at time $t = 0$. (a) Find the reading of each meter just after S is closed. (b) What does each meter read long after S is closed?

Figure **P30.65**



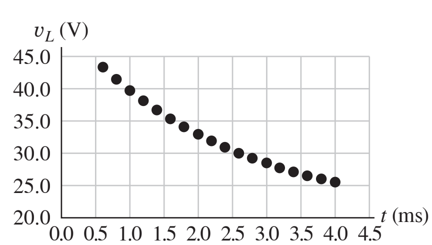
30.66 ••• **CP** In the circuit shown in **Fig. P30.66**, neither the battery nor the inductors have any appreciable resistance, the capacitors are initially uncharged, and the switch S has been in position 1 for a very long time. (a) What is the current in the circuit? (b) The switch is now suddenly 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**



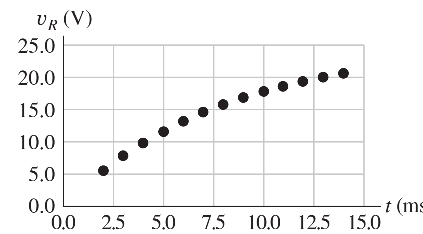
30.67 •• **DATA** During a summer internship as an electronics technician, you are asked to measure the self-inductance L of a solenoid. You connect the solenoid in series with a 10.0 Ω resistor, a battery that has negligible internal resistance, and a switch. Using an ideal voltmeter, you measure and 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 $t \rightarrow \infty$? Use this result to calculate the current in the circuit as $t \rightarrow \infty$. (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?

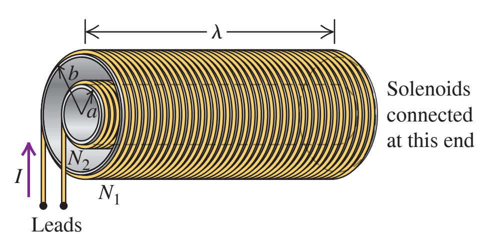
Figure **P30.68**



CHALLENGE PROBLEMS

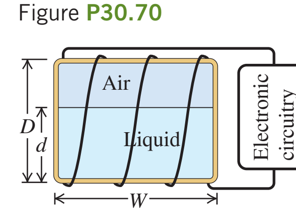
30.69 ••• A long solenoid with N_1 windings and radius b surrounds a coaxial, narrower solenoid with N_2 windings and radius $a < b$, as shown in **Fig. P30.69**. At the far end, the outer solenoid is attached to the inner solenoid by a wire, so that current flows down the outer coil and then back through the inner coil as shown. (a) The two leads are attached to a supply circuit that includes a battery, supplying current I as indicated. What is the magnetic flux through each turn of the inner coil, taking rightward as the positive direction? (b) What is the magnetic flux through each turn of the outer coil? (c) What is the inductance as seen by the two leads? (d) What would be the inductance if the sense of the inner coil were reversed? (e) In the original configuration, what would be the inductance if $\lambda = 20.0$ cm, $a = 1.00$ cm, $b = 2.00$ cm, $N_1 = 1200$, and $N_2 = 750$? (f) Using the values from part (e), what would be the inductance in the configuration of part (d)?

Figure **P30.69**



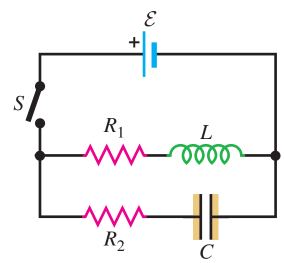
30.70 ••• **CP** **A Volume Gauge.** A tank containing a liquid has turns of wire wrapped around it, causing it to act like an inductor. The liquid content of the tank can be measured by using its inductance to determine the height of the liquid in the tank. The inductance of the tank changes from a 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, $R_1 = 25\ \Omega$, and $R_2 = 5000\ \Omega$. (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



MCAT-STYLE 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.

Key Example VARIATION Problems

VP30.4.1 (a) 6.50×10^{-5} H (b) $1.56 \times 10^{-3}\ \text{T} \cdot \text{m}^2$
VP30.4.2 (a) 3.36×10^{-4} V (b) zero (c) 1.44×10^{-4} V
VP30.4.3 (a) $1.96 \times 10^{-6}\ \text{T} \cdot \text{m}^2$ (b) 4.18×10^{-3} V
VP30.4.4 (a) 25.5 μH (b) 23.0 mV, opposite to the current
VP30.7.1 (a) 84.9 Ω (b) 0.212 A
VP30.7.2 (a) 0.190 A (b) +229 A/s (c) 6.79×10^{-4} J
VP30.7.3 (a) 0.476 A (b) −229 A/s (c) 4.26×10^{-3} J
VP30.7.4 (a) $(L/R)\ln 2 = 0.693(L/R)$ (b) $\mathcal{E}/2R$ (c) $\mathcal{E}/2L$
VP30.10.1 (a) 5.12×10^{-3} F (b) 14.0 A
VP30.10.2 (a) 2.20×10^{-4} s (b) 2.85 A
VP30.10.3 (a) 35.1 μF (b) 0.0125 H (c) 1.51×10^3 rad/s
VP30.10.4 (a) 61.1 μF (b) 49.4 Ω

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 OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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'll 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.

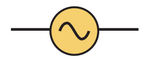
Eventually, Westinghouse prevailed, and most present-day household and industrial power distribution systems operate with alternating current. Any appliance that you plug into a wall outlet uses ac. Circuits in modern communication equipment also make extensive use of ac.

In this chapter we'll learn how resistors, inductors, and capacitors behave in **ac circuits**—that is, 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**).

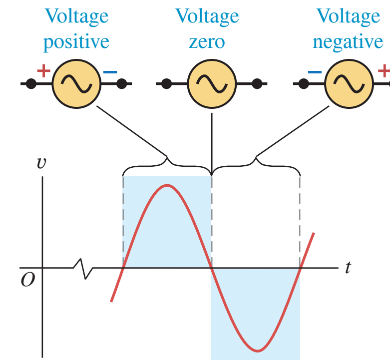


Figure 31.1 The voltage across a sinusoidal ac source.

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:

Instantaneous current · Angular frequency · Time

$$i = I\cos\omega t \tag{31.2}$$

Current amplitude (maximum current)

Phasor Diagrams

To represent sinusoidally varying voltages and currents in ac circuits, we'll 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).

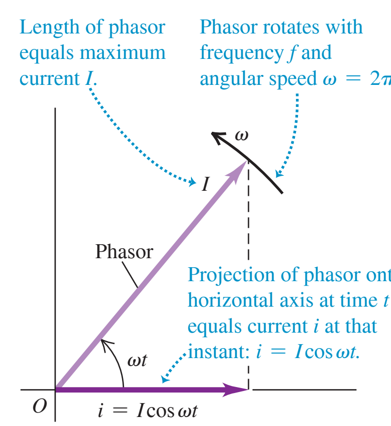


Figure 31.2 A phasor diagram.

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.

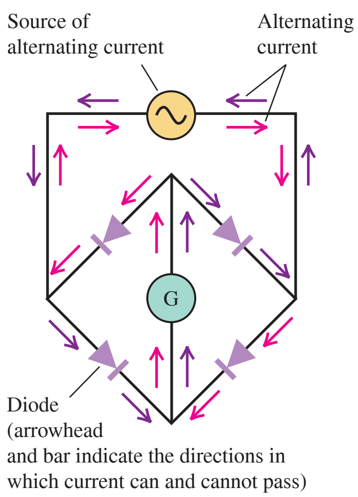
Rectified Alternating Current

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.

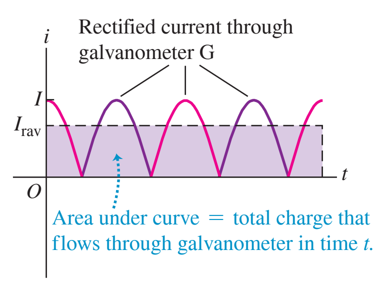
To get a measurable one-way current through the meter, we can use *diodes,* which we described in Section 25.3. A diode is a device that conducts better in one direction than in the other; an ideal diode has zero resistance for one direction of current and infinite resistance for the other. **Figure 31.3a** (next page) shows one possible arrangement, called

Figure **31.3** (a) A full-wave rectifier circuit. (b) Graph of the resulting current through the galvanometer G.

(a) A full-wave rectifier circuit



(b) Graph of the full-wave rectified current and its average value, the rectified average current I_{rav}



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 $\quad I_{rav} = \dfrac{2}{\pi}I = 0.637I \quad$ Current amplitude $\qquad$ (31.3)

(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 $\quad I_{rms} = \dfrac{I}{\sqrt{2}} \quad$ Current amplitude $\qquad$ (31.4)

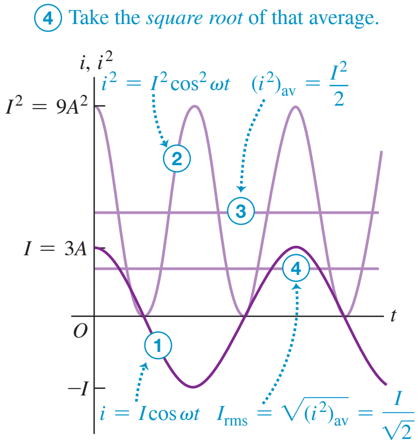
In the same way, the root-mean-square value of a sinusoidal voltage is

Root-mean-square (rms) value of a sinusoidal voltage $\quad V_{rms} = \dfrac{V}{\sqrt{2}} \quad$ Voltage amplitude (maximum value) $\qquad$ (31.5)

Figure **31.4** Calculating the root-mean-square (rms) value of an alternating current.

Meaning of the rms value of a sinusoidal quantity (here, ac current with $I = 3$ A):

1. Graph current i versus time.
2. *Square* the instantaneous current i.
3. Take the *average* (mean) value of i^2.
4. Take the *square root* of that average.



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

$$V = \sqrt{2}V_{\text{rms}}$$
$$= \sqrt{2}(120\ \text{V}) = 170\ \text{V}$$

Sixty times per second, the instantaneous voltage across a socket's terminals varies from 170 V to −170 V and back again.

Figure **31.5** This wall socket delivers a root-mean-square (rms) voltage of 120 V.

EXAMPLE 31.1 Current in a desktop computer

The plate on the back of a desktop 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?

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,

$$I_{\text{rms}} = \sqrt{(i^2)_{\text{av}}} \text{ so } (i^2)_{\text{av}} = (I_{\text{rms}})^2 = (2.7\ \text{A})^2 = 7.3\ \text{A}^2$$

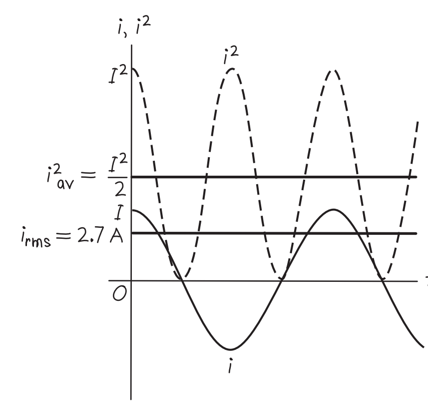
(c) From Eq. (31.4), the current amplitude I is

$$I = \sqrt{2}I_{\text{rms}} = \sqrt{2}(2.7\ \text{A}) = 3.8\ \text{A}$$

Figure 31.6 shows graphs of i and i^2 versus time t.

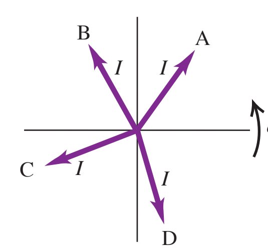
EVALUATE Why would we be interested in the average of the square of the current? Recall that the rate at which energy is dissipated in a resistor R equals i^2R. This rate varies if the current is alternating, so it is best described by its average value $(i^2)_{\text{av}}R = I_{\text{rms}}{}^2R$. We'll use this idea in Section 31.4.

Figure **31.6** Our graphs of the current i and the square of the current i^2 versus time t.



KEYCONCEPT The root-mean-square (rms) value of a time-varying quantity is the square root of the average value of the square of that quantity. To find the rms value of a quantity that varies sinusoidally with time, like the current or voltage in an ac circuit, divide its amplitude by $\sqrt{2}$.

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?



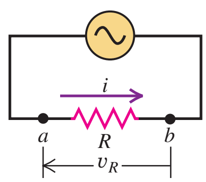
ANSWER

I (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.

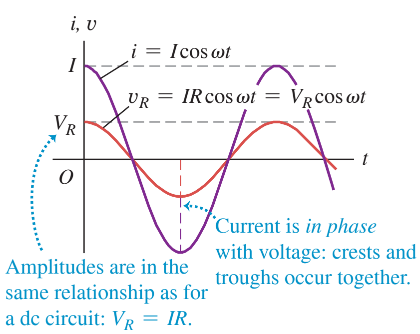
31.2 RESISTANCE AND REACTANCE

Figure 31.7 Resistance R connected across an ac source.

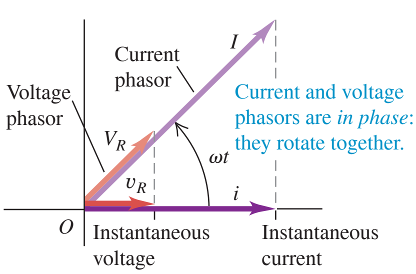
(a) Circuit with ac source and resistor



(b) Graphs of current and voltage versus time



(c) Phasor diagram



In this section we'll 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 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 $\quad V_R = IR$ $\quad$ Current amplitude (I), Resistance (R) $\quad$ (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.

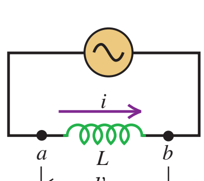
Inductor in an ac Circuit

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.

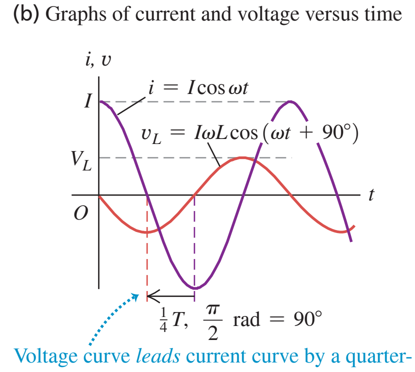
Although there is no resistance, there is a potential difference v_L between the inductor terminals a and b because the current varies with time, giving rise to a self-induced emf. The induced emf in the direction of i is given by Eq. (30.7), $\mathcal{E} = -L\,di/dt$; however, the voltage v_L is *not* simply equal to $\mathcal{E}$. To see why, notice that if the current in the inductor is in the

Figure 31.8 Inductance L connected across an ac source.

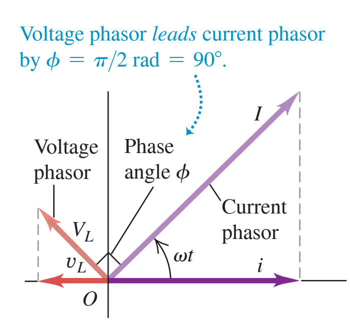
(a) Circuit with ac source and inductor



(b) Graphs of current and voltage versus time



(c) Phasor diagram



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

$$v_L = L\frac{di}{dt} = L\frac{d}{dt}(I\cos\omega t) = -I\omega L\sin\omega t \qquad (31.9)$$

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 $\cos(A + 90°) = -\sin A$:

$$v_L = I\omega L\cos(\omega t + 90°) \qquad (31.10)$$

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'll 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

$$V_L = I\omega L \qquad (31.11)$$

We define the **inductive reactance** X_L of an inductor as

$$X_L = \omega L \qquad \text{(inductive reactance)} \qquad (31.12)$$

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

$$V_L = IX_L \qquad (31.13)$$

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).

EXAMPLE 31.2 An inductor in an ac circuit

WITH VARIATION PROBLEMS

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?

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 $I = V_L/X_L = V_L/\omega L = V_L/2\pi fL$. 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.

KEYCONCEPT For any component in an ac circuit, the reactance is the ratio of the amplitude of the voltage across the component to the amplitude of the current through the component. If the component is an inductor, the reactance equals the product of the inductance and the angular frequency of the current.

Capacitor in an ac Circuit

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.

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 $\cos(A - 90^\circ) = \sin A$:

$$v_C = \frac{I}{\omega C}\cos(\omega t - 90^\circ) \tag{31.16}$$

This corresponds to a phase angle $\phi = -90^\circ$. 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

$$X_C = \frac{1}{\omega C} \qquad \text{(capacitive reactance)} \tag{31.18}$$

Then

Amplitude of voltage across a capacitor, ac circuit $V_C = IX_C$ — I: Current amplitude; X_C: Capacitive reactance (31.19)

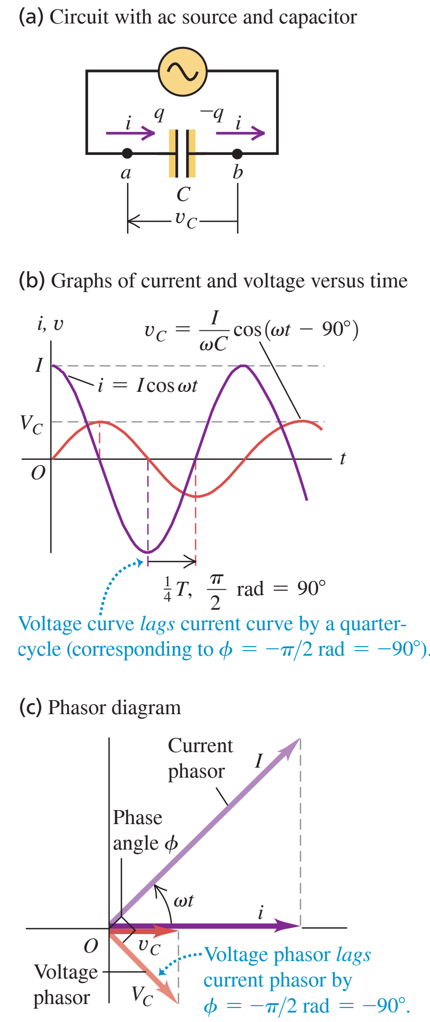
The SI unit of X_C is the ohm, the same as for resistance and inductive reactance, because X_C is the ratio of a voltage and a current.

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).

Figure 31.9 Capacitor C connected across an ac source.



EXAMPLE 31.3 A resistor and a capacitor in an ac circuit — WITH VARIATION PROBLEMS

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.

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.

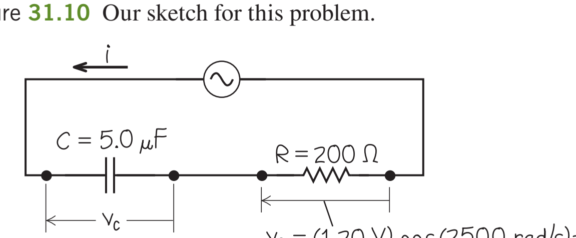
EXECUTE (a) From Eq. (31.6), $v_R = iR$, we find

$$i = \frac{v_R}{R} = \frac{(1.20\text{ V})\cos[(2500\text{ rad/s})t]}{200\ \Omega}$$
$$= (6.0 \times 10^{-3}\text{ A})\cos[(2500\text{ rad/s})t]$$

(b) From Eq. (31.18), the capacitive reactance at $\omega = 2500$ rad/s is

$$X_C = \frac{1}{\omega C} = \frac{1}{(2500\text{ rad/s})(5.0 \times 10^{-6}\text{ F})} = 80\ \Omega$$

Figure 31.10 Our sketch for this problem.



(c) From Eq. (31.19), the capacitor voltage amplitude is

$$V_C = IX_C = (6.0 \times 10^{-3}\text{ A})(80\ \Omega) = 0.48\text{ V}$$

(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):

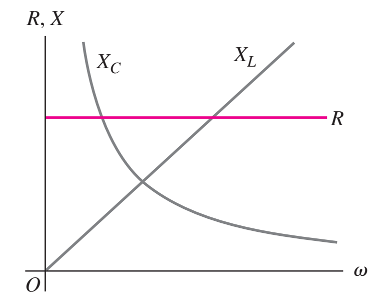
$$v_C = V_C\cos(\omega t - 90^\circ) = (0.48\text{ V})\cos[(2500\text{ rad/s})t - \pi/2\text{ rad}]$$

Continued

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.

KEYCONCEPT The reactance (ratio of voltage amplitude to current amplitude) for a resistor in an ac circuit is equal to its resistance. For a capacitor, the reactance equals the reciprocal of the product of the capacitance and the angular frequency of the current.

Figure 31.11 Graphs of R, X_L, and X_C as functions of angular frequency ω.



Comparing ac Circuit Elements

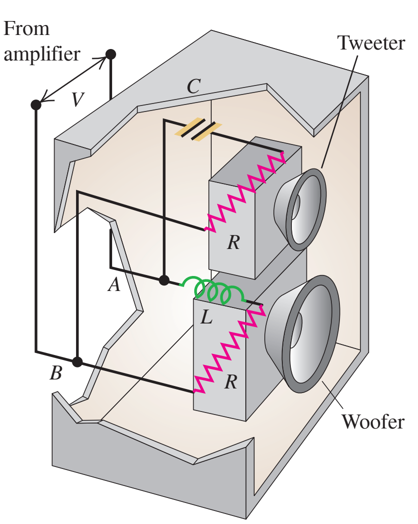
Table 31.1 summarizes the relationships of voltage and current amplitudes for the three circuit elements we have discussed. Note again that *instantaneous* voltage and current are proportional in a resistor, where there is zero phase difference between v_R and i (see Fig. 31.7b). The instantaneous voltage 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 31.12 (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

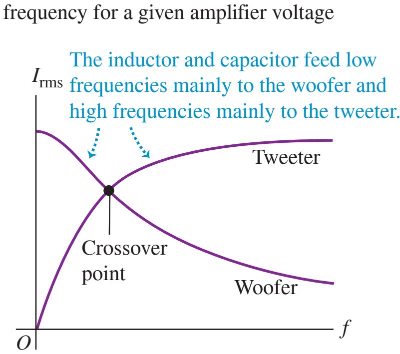


Figure 31.11 shows how the resistance of a resistor and the reactances of an inductor and a capacitor vary with angular frequency ω. Resistance R is independent of frequency, while the reactances X_L and X_C are not. If $\omega = 0$, corresponding to a dc circuit, there is *no* current through a capacitor because $X_C \to \infty$, and there is no inductive effect because $X_L = 0$. In the limit $\omega \to \infty$, 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?

ANSWER

(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 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'll 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

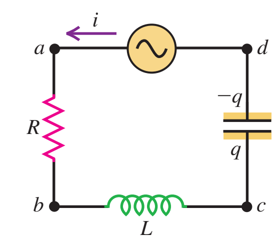
$$V = \sqrt{V_R^2 + (V_L - V_C)^2} = \sqrt{(IR)^2 + (IX_L - IX_C)^2} \quad \text{or}$$
$$V = I\sqrt{R^2 + (X_L - X_C)^2} \tag{31.20}$$

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

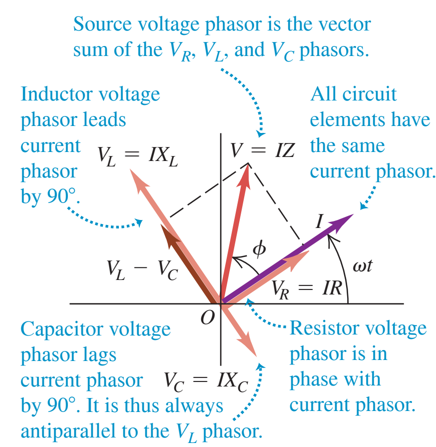
$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{31.21}$$

Figure **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 $X_L < X_C$

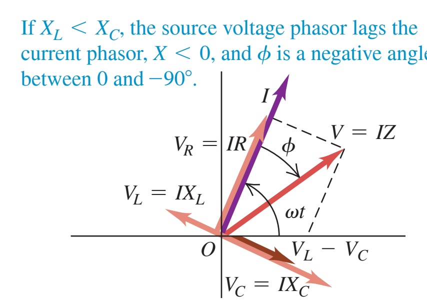


Figure 31.14 This gas-filled glass sphere has an alternating voltage between its surface and the electrode at its center. The glowing streamers show the resulting alternating current that passes through the gas. When you touch the outside of the sphere, your fingertips and the inner surface of the sphere 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.

so we can rewrite Eq. (31.20) as

$$\text{Amplitude of voltage across an ac circuit} \rightarrow V = I Z \quad (I: \text{Current amplitude};\ Z: \text{Impedance of circuit}) \tag{31.22}$$

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:

$$\text{Impedance of an } L\text{-}R\text{-}C \text{ series circuit} \rightarrow Z = \sqrt{\underbrace{R^2}_{\text{Resistance}} + [\omega \underbrace{L}_{\text{Inductance}} - (1/\omega \underbrace{C}_{\text{Capacitance}})]^2} \quad (\omega: \text{Angular frequency}) \tag{31.23}$$

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,

$$\tan\phi = \frac{V_L - V_C}{V_R} = \frac{I(X_L - X_C)}{IR} = \frac{X_L - X_C}{R}$$

$$\text{Phase angle of voltage with respect to current in an } L\text{-}R\text{-}C \text{ series circuit} \rightarrow \tan\phi = \frac{\omega L - 1/\omega C}{R} \quad (\omega: \text{Angular frequency};\ L: \text{Inductance};\ C: \text{Capacitance};\ R: \text{Resistance}) \tag{31.24}$$

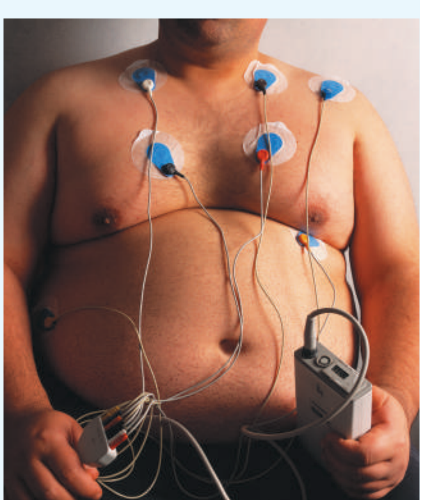
If the current is $i = I\cos\omega t$, then the source voltage v is

$$v = V\cos(\omega t + \phi) \tag{31.25}$$

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 ($X_C = 1/\omega C = 0$).

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.



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 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

WITH VARIATION PROBLEMS

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.

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.

KEYCONCEPT For an ac circuit that includes an inductor L, resistor R, and capacitor C in series, the impedance—the ratio of the amplitude of the voltage across the L-R-C combination to the amplitude of the current—depends on the values of L, R, C, and the angular frequency of the current.

EXAMPLE 31.5 An *L-R-C* series circuit II

WITH VARIATION PROBLEMS

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).

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

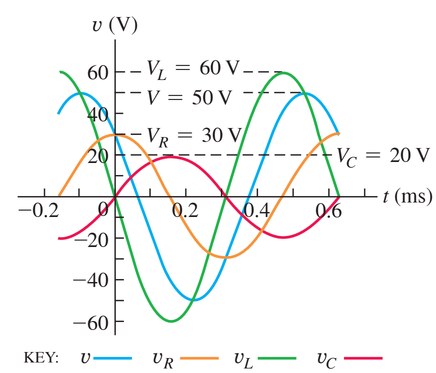
$$v_C = V_C\cos(\omega t - 90°) = V_C\sin\omega t = (20\text{ V})\sin[(10{,}000\text{ rad/s})t]$$

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

$$\begin{aligned} v &= V\cos(\omega t + \phi) \\ &= (50\text{ V})\cos\left[(10{,}000\text{ rad/s})t + \left(\frac{2\pi\text{ rad}}{360°}\right)(53°)\right] \\ &= (50\text{ V})\cos[(10{,}000\text{ rad/s})t + 0.93\text{ rad}] \end{aligned}$$

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.

Figure 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.



KEYCONCEPT In an L-R-C series circuit that includes an ac source, the instantaneous voltage across the L-R-C combination is the sum of the instantaneous voltages across each of the three components. Because these voltages are not in phase with one another, the amplitude of the voltage across the L-R-C combination is *not* equal to the sum of the individual voltage amplitudes.

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.

ANSWER

(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 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

$$P_{av} = \tfrac{1}{2}VI \qquad \text{(for a pure resistor)} \tag{31.27}$$

An equivalent expression is

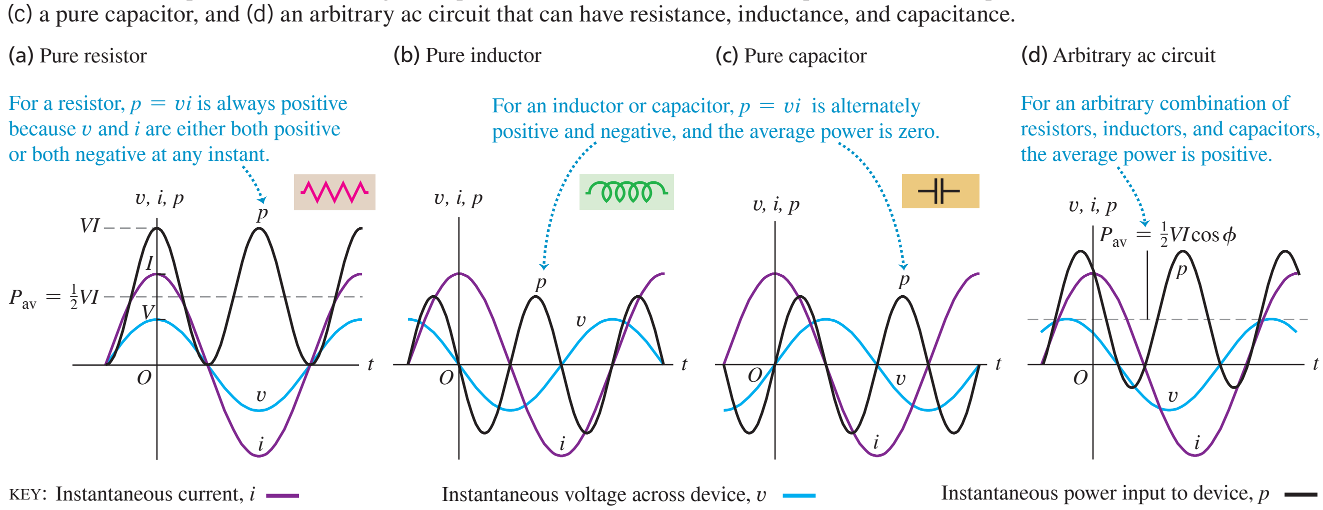
$$P_{av} = \frac{V}{\sqrt{2}}\frac{I}{\sqrt{2}} = V_{rms}I_{rms} \qquad \text{(for a pure resistor)} \tag{31.28}$$

Also, $V_{rms} = I_{rms}R$, so we can express P_{av} by any of the equivalent forms

$$P_{av} = I_{rms}{}^2R = \frac{V_{rms}{}^2}{R} = V_{rms}I_{rms} \qquad \text{(for a pure resistor)} \tag{31.29}$$

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.

Figure **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

$$\underbrace{P_{av}}_{\text{Average power into a general ac circuit}} = \frac{1}{2}\underbrace{V}_{\text{Voltage amplitude}}\underbrace{I}_{\text{Current amplitude}}\cos\overbrace{\phi}^{\text{Phase angle of voltage with respect to current}} = \underbrace{V_{rms}}_{\text{rms voltage}}\underbrace{I_{rms}}_{\text{rms current}}\cos\phi \tag{31.31}$$

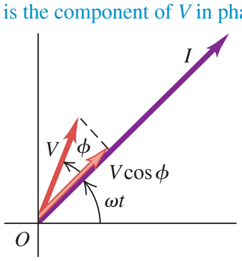
Figure 31.17 shows the general relationship of the current and voltage phasors. When v and i are in phase, so $\phi = 0$, the average power equals $\frac{1}{2}VI = V_{rms}I_{rms}$; when v and i are 90° out of phase, the average power is zero. In the general case, when v has a phase angle ϕ with respect 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.

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

Figure **31.17** Using phasors to calculate the average power for an arbitrary ac circuit.

Average power $= \frac{1}{2}I(V\cos\phi)$, where $V\cos\phi$ is the component of V in phase with I.



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

WITH VARIATION PROBLEMS

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.)

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.

KEYCONCEPT The power delivered to a resistor in an ac circuit is not constant, but oscillates between zero and a maximum value. The average power delivered to the resistor is one-half of this maximum, and equals the rms current through the resistor times the rms voltage across the resistor.

EXAMPLE 31.7 Power in an *L-R-C* series circuit

WITH VARIATION PROBLEMS

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.

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.

KEYCONCEPT The power factor of an ac circuit equals the cosine of the phase angle ϕ between the oscillating voltage across the circuit and the oscillating current in the circuit. The greater the power factor, the less current is required to supply a given average power for a given voltage amplitude.

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.

ANSWER

(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 RESONANCE IN ALTERNATING-CURRENT CIRCUITS

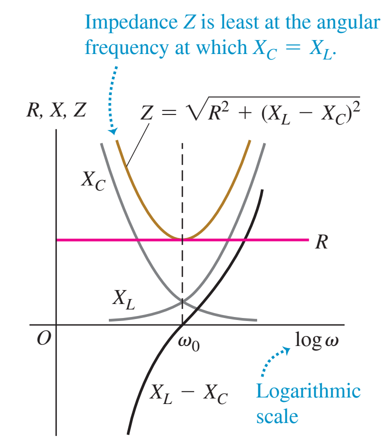
?

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.

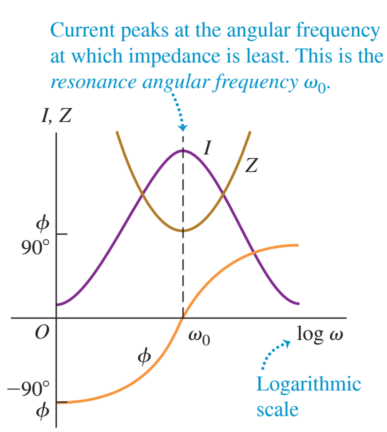
To see how an *L-R-C* series circuit can be used in this way, suppose we connect an ac source with constant voltage amplitude V but adjustable angular frequency ω across an *L-R-C* series circuit. The current that appears in the circuit has the same angular frequency as the source and a current amplitude $I = V/Z$, 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.

Figure 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



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

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{31.32}$$

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 0° 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.

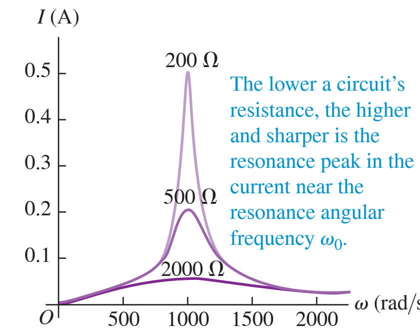
The resonance frequency is determined by L and C; what happens when we change R? Figure 31.19 also shows graphs of I as a function of ω for $R = 200\ \Omega$ and for $R = 2000\ \Omega$. The curves are similar for frequencies far away from resonance, where the impedance is dominated by X_L or X_C. But 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 Problems 31.54 and 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.

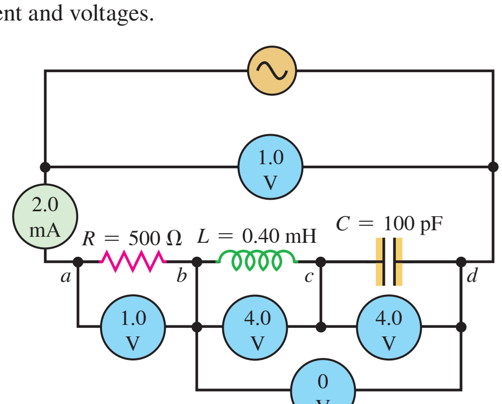
Figure **31.19** Graph of current amplitude I as a function of angular frequency ω for an L-R-C series circuit with $V = 100$ V, $L = 2.0$ H, $C = 0.50\ \mu$F, and three different values of the resistance R.



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.

Figure **31.20** A radio tuning circuit at resonance. The circles denote rms current and voltages.



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.

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, $Z = R = 500\ \Omega$.

Continued

(c) From Eq. (31.26) the rms current at resonance is

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{Z} = \frac{V_{\text{rms}}}{R} = \frac{1.0\text{ V}}{500\ \Omega} = 0.0020\text{ A} = 2.0\text{ mA}$$

(d) The rms potential difference across the resistor is

$$V_{R\text{-rms}} = I_{\text{rms}}R = (0.0020\text{ A})(500\ \Omega) = 1.0\text{ V}$$

The rms potential differences across the inductor and capacitor are

$$V_{L\text{-rms}} = I_{\text{rms}}X_L = (0.0020\text{ A})(2000\ \Omega) = 4.0\text{ V}$$

$$V_{C\text{-rms}} = I_{\text{rms}}X_C = (0.0020\text{ A})(2000\ \Omega) = 4.0\text{ V}$$

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}.

KEYCONCEPT In an ac circuit, the current amplitude is maximum for a given source voltage amplitude when the source frequency equals the circuit's resonance frequency. For a series *L-R-C* circuit the resonance frequency is the same as the circuit's natural oscillation frequency without the resistor or the source.

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.

ANSWER

(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 TRANSFORMERS

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.

One 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 North and Central America, and 220 V to 240 V in most of the rest of the world. The necessary voltage conversion is accomplished by using **transformers.**

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

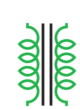
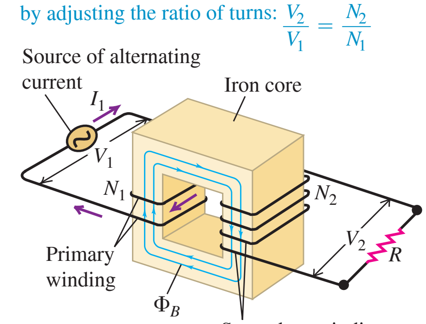


Figure 31.21 Schematic diagram of an idealized step-up transformer. The primary is connected to an ac source; the secondary is connected to a device with resistance R.

The induced emf *per turn* is the same in both coils, so we adjust the ratio of terminal voltages by adjusting the ratio of turns: $\dfrac{V_2}{V_1} = \dfrac{N_2}{N_1}$



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. 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**).

Figure **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.

Figure **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:

Primary voltage amplitude or rms value

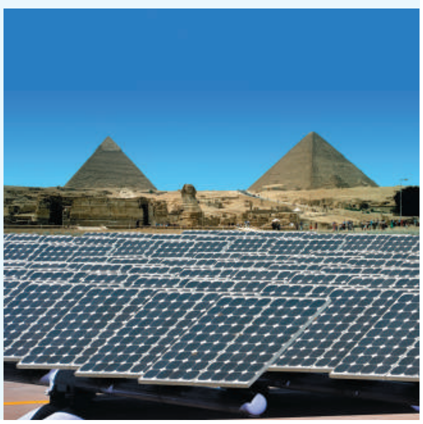
Secondary voltage amplitude or rms value

$$V_1 I_1 = V_2 I_2 \tag{31.36}$$

Current in primary

Current in secondary

APPLICATION When dc Power Transmission Is Better Than ac In the cables used for transmitting ac power, eddy currents oppose and cancel the current near the central axis of the cable. Hence current flows only in a reduced area near the cable's surface, increasing the cable's effective resistance and causing the loss of more power. For this and other reasons, high-voltage dc can be more efficient than high-voltage ac for transmitting power over long distances, such as in a planned application to take solar power generated in North Africa to Europe through dc lines at the bottom of the Mediterranean.



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.

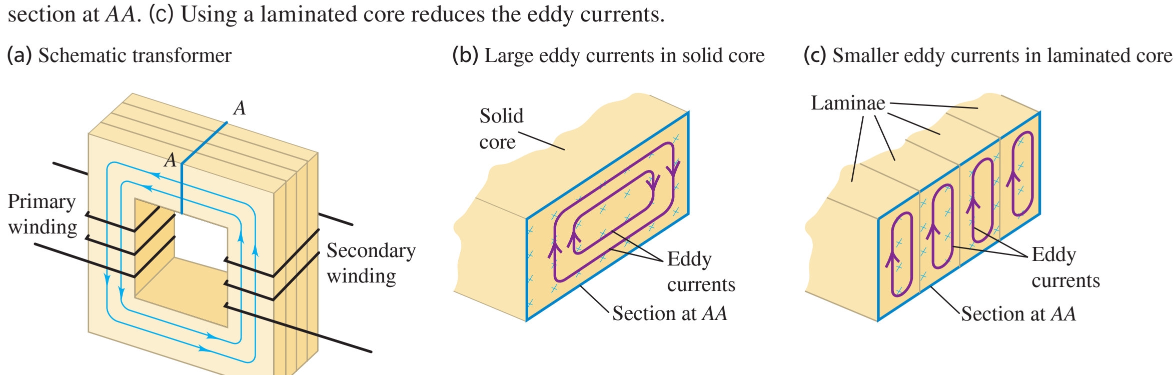
Real transformers always have some energy losses. (That's why an ac adapter like the one shown in Fig. 31.23 feels warm to the touch after it's been in use for a while; the transformer is heated by the dissipated energy.) The windings have some resistance, leading to i^2R losses. There are also energy 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 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%.

Figure 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.



EXAMPLE 31.9 "Wake up and smell the (transformer)!"

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.)

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.

KEYCONCEPT A transformer can be used to raise or lower the voltage amplitude of an alternating current. If there are more turns in the secondary coil than in the primary coil (step-up transformer), the transformer increases the voltage amplitude and decreases the current amplitude; the reverse is true if there are fewer turns in the secondary than in the primary (step-down transformer).

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.

ANSWER

(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.

CHAPTER 31 SUMMARY

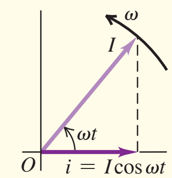
Phasors and alternating current: An alternator or ac source produces an emf that varies sinusoidally with time. A sinusoidal voltage or current can be represented by a phasor, a vector that rotates counterclockwise with constant angular velocity ω equal to the angular frequency of the sinusoidal quantity. 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.)

$$I_{\text{rav}} = \frac{2}{\pi} I = 0.637I \quad (31.3)$$

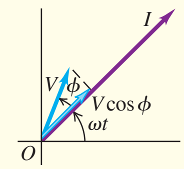
$$I_{\text{rms}} = \frac{I}{\sqrt{2}} \quad (31.4)$$

$$V_{\text{rms}} = \frac{V}{\sqrt{2}} \quad (31.5)$$



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.

$$i = I\cos\omega t \quad (31.2)$$

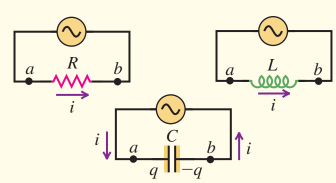


Resistance and reactance: The voltage across a resistor R is in phase with the current. The voltage across an inductor L leads the current by 90° ($\phi = +90°$), while the voltage across a capacitor C lags the current by 90° ($\phi = -90°$). The voltage amplitude across each type of device is proportional 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.)

$$V_R = IR \quad (31.7)$$

$$V_L = IX_L \quad (31.13)$$

$$V_C = IX_C \quad (31.19)$$

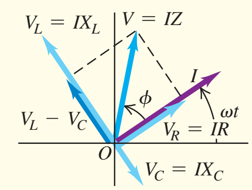


Impedance and the *L-R-C* series circuit: In a general ac circuit, the voltage and current amplitudes are related by the circuit impedance Z. In an L-R-C series circuit, the values of L, R, C, and the angular frequency ω determine the impedance and the phase angle ϕ of the voltage relative to the current. (See Examples 31.4 and 31.5.)

$$V = IZ \quad (31.22)$$

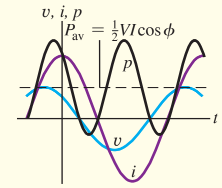
$$Z = \sqrt{R^2 + [\omega L - (1/\omega C)]^2} \quad (31.23)$$

$$\tan\phi = \frac{\omega L - 1/\omega C}{R} \quad (31.24)$$



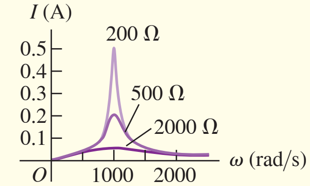
Power in ac circuits: The average power input P_{av} to an ac circuit depends on the voltage and current amplitudes (or, equivalently, their rms values) and the phase angle ϕ of the voltage relative to the current. The quantity $\cos\phi$ is called the power factor. (See Examples 31.6 and 31.7.)

$$P_{\text{av}} = \tfrac{1}{2}VI\cos\phi = V_{\text{rms}}I_{\text{rms}}\cos\phi \quad (31.31)$$



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.)

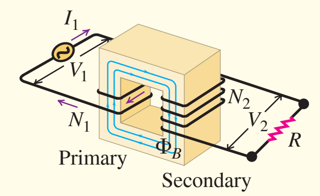
$$\omega_0 = \frac{1}{\sqrt{LC}} \quad (31.32)$$



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.)

$$\frac{V_2}{V_1} = \frac{N_2}{N_1} \quad (31.35)$$

$$V_1 I_1 = V_2 I_2 \quad (31.36)$$



Chapter 31 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 31.2 and 31.3 (Section 31.2) before attempting these problems.

VP31.3.1 An inductor with $L = 2.50$ mH is part of the circuit for an aviation radio beacon that operates at 108 MHz. Find (a) the inductive reactance of the inductor and (b) the amplitude of the current in the inductor if the amplitude of the voltage across the inductor is 4.20 kV.

VP31.3.2 An ac circuit includes a resistor with $R = 125\ \Omega$ connected in series with a capacitor with $C = 7.00\ \mu$F. The current through the resistor and capacitor is $i = (2.40 \times 10^{-3}\text{ A})\cos[(1.75 \times 10^3\text{ rad/s})t]$. Find (a) the voltage across the resistor as a function of time, (b) the capacitive reactance of the capacitor, and (c) the voltage across the capacitor as a function of time.

VP31.3.3 An ac circuit includes a 155 Ω resistor in series with a 8.00 μF capacitor. The current in the circuit has amplitude 4.00×10^{-3} A. (a) Find the frequency for which the capacitive reactance equals the resistance. (b) At this frequency, what are the amplitudes of the voltages across the resistor and capacitor? (c) At this frequency, when the voltage across the resistor is maximum, what is the voltage across the capacitor? When the voltage across the capacitor is maximum, what is the voltage across the resistor?

VP31.3.4 A 116 Ω resistor and an unknown capacitor are connected in series to an ac source with angular frequency 5.10×10^3 rad/s. The amplitude of the resistor voltage is 2.45 V, and the current in the circuit has its maximum positive value at $t = 0$. Find (a) the current amplitude, (b) the current at $t = 3.50$ ms, (c) the maximum voltage across the capacitor, and (d) the capacitance.

Be sure to review EXAMPLES 31.4 and 31.5 (Section 31.3) before attempting these problems.

VP31.5.1 In the series circuit of Fig. 31.13a the resistance is 255 Ω, the inductance is 45.0 mH, the capacitance is 0.400 μF, and the source has voltage amplitude 55.0 V. The inductive reactance is equal to the resistance. Find (a) the frequency of the source, (b) the impedance, (c) the current amplitude, (d) the phase angle, and (e) whether the voltage leads or lags the current.

VP31.5.2 In the series circuit of Fig. 31.13a the source has voltage amplitude 20.0 V and angular frequency 5.40×10^3 rad/s, the inductance is 6.50 mH, the capacitance is 0.600 μF, and the impedance is 474 Ω. Find (a) the resistance, (b) the current amplitude, (c) the resistor voltage amplitude, (d) the inductor voltage amplitude, and (e) the capacitor voltage amplitude.

VP31.5.3 The source in an *L-R-C* series circuit has angular frequency 1.20×10^4 rad/s. The capacitance is 0.965 μF and the resistance is 65.0 Ω. If you want the source voltage to lead the current by 15.0°, find (a) the required value of $X_L - X_C$ and (b) the required inductance.

VP31.5.4 The current in an *L-R-C* series circuit has amplitude 0.120 A and angular frequency 8.00×10^3 rad/s, and it has its maximum positive value at $t = 0$. The resistance is 95.0 Ω, the inductance is 6.50 mH, and the capacitance is 0.440 μF. For the resistor, inductor, and capacitor, find (a) the voltage amplitudes and (b) the instantaneous voltages at $t = 0.305$ ms.

Be sure to review EXAMPLES 31.6 and 31.7 (Section 31.4) before attempting these problems.

VP31.7.1 In operation an electric toaster has an rms current of 3.95 A when plugged into a wall socket with rms voltage 1.20×10^2 V. Find the toaster's (a) average power, (b) maximum instantaneous power, and (c) resistance.

VP31.7.2 An *L-R-C* series circuit has a source with voltage amplitude 35.0 V and angular frequency 1.30×10^3 rad/s. The resistance is 275 Ω, the inductance is 82.3 mH, and the capacitance is 1.10 μF. Find (a) the inductive and capacitive reactances, (b) the phase angle, and (c) the power factor.

VP31.7.3 For the circuit in the preceding problem, find (a) the impedance, (b) the current amplitude, and (c) the average power delivered to the resistor.

VP31.7.4 The power factor of an *L-R-C* series circuit is 0.800. The angular frequency is 1.30×10^4 rad/s, the inductance is 7.50 mH, and the capacitance is 0.440 μF. Find (a) the inductive and capacitive reactances, (b) the phase angle, and (c) the resistance.

BRIDGING PROBLEM An Alternating-Current Circuit

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'll 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

•, ••, •••: 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 • 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

31.3 • 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?

31.4 • 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 • (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?

31.6 • 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 $C = 3.50$ μF, what is the numerical value of the angular frequency in part (a), and what is the reactance of each element?

31.7 • (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?

31.8 • **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?

31.9 •• 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?

31.10 •• 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.

31.11 •• 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

31.12 • 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.

31.13 • The resistor, inductor, capacitor, and voltage source described in Exercise 31.12 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.

31.14 •• 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.

31.15 • 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?

31.16 •• An *L-R-C* series circuit has voltage amplitudes $V_L = 180$ V, $V_C = 120$ V, and $V_R = 160$ V. At time t the instantaneous voltage across the inductor is 80.0 V. At this instant, what is the voltage across the capacitor and across the resistor?

31.17 • An *L-R-C* series circuit has source voltage amplitude $V = 240$ V, and the voltage amplitudes for the inductor and capacitor are $V_L = 310$ V and $V_C = 180$ V. What is the phase angle ϕ?

Section 31.4 Power 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) Use Figs. 31.13b and 31.13c to show that for a series *L-R-C* circuit, $V\cos\phi = IR$. Use this result in Eq. (31.31) to show that the average power delivered by the source is $P_{\text{av}} = \frac{1}{2}I^2R$. Therefore the average power delivered by the source is equal to the average power consumed by the resistor. This means the average power for the capacitor and for the inductor is zero. These circuit elements consume electrical energy during the part of the current oscillations when they are storing energy, but all the stored energy in each is then released during another part of the current oscillations. (b) In an *L-R-C* series circuit the amplitude of the source voltage is 120 V, the source voltage leads the current by 53.1°, and the average power supplied by the source is 80.0 W. What is the resistance R of the resistor in the circuit?

31.22 •• A circuit has an ac voltage source and a resistor and capacitor connected in series. There is no inductor. The ac voltage source has voltage amplitude 900 V and angular frequency $\omega = 20.0$ rad/s. The voltage amplitude across the capacitor is 500 V. The resistor has resistance $R = 300\ \Omega$. (a) What is the voltage amplitude across the resistor? (b) What is the capacitance C of the capacitor? (c) Does the source voltage lag or lead the current? (d) What is the average rate at which the ac source supplies electrical energy to the circuit?

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 $V_{\text{rms}} = 80.0$ 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 • An *L-R-C* series circuit has $R = 400\ \Omega$, $L = 0.600$ H, and $C = 5.00 \times 10^{-8}$ F. The voltage amplitude of the source is 80.0 V. When the ac source operates at the resonance frequency of the circuit, what is the average power delivered by the source?

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

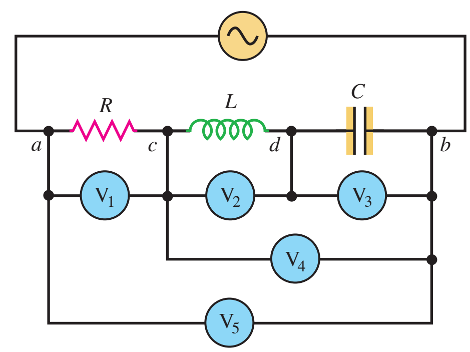
31.37 • A coil has a resistance of 48.0 Ω. At a frequency of 80.0 Hz the voltage across the coil leads the current in it by 52.3°. Determine the inductance of the coil.

31.38 •• In an *L-R-C* series circuit the phase angle is 53.0° and the source voltage lags the current. The resistance of the resistor is 300 Ω and the reactance of the capacitor is 500 Ω. The average power delivered by the source is 80.0 W. (a) What is the reactance of the inductor? (b) What is the current amplitude in the circuit? (c) What is the voltage amplitude of the source?

31.39 •• An *L-R-C* series circuit has $C = 4.80\ \mu$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, $C = 6.00\ \mu$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**

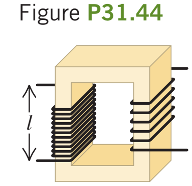


31.41 •• **CP** A parallel-plate capacitor having square plates 4.50 cm on each side and 8.00 mm apart is placed in series with the following: an ac source of angular frequency 650 rad/s and voltage amplitude 22.5 V; a 75.0 Ω resistor; and an ideal solenoid that is 9.00 cm long, has a circular cross section 0.500 cm in diameter, and carries 125 coils per centimeter. What is the resonance angular frequency of this circuit? (See Exercise 30.11.)

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?

Figure **P31.44**



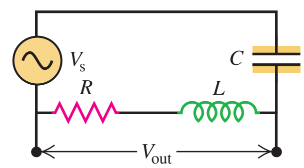
31.44 •• Consider the ac adapter that bisects the power cord to your laptop computer, cell phone charger, or another electronic device. (a) What dc output voltage V and current I are listed on your device? (b) What power P is listed on your device? Does $P = IV$? (c) Your ac adapter is most likely not a simple transformer; however, it is instructive to analyze the device as if it were: Assume your device is a transformer with a 120 V rms voltage on its primary coil and a full-wave rectifier following its secondary coil. If an *R-C* circuit with a sufficient time constant followed, the ultimate output would maintain the dc value of the voltage amplitude on the rectified secondary coil. If the primary coil has 200 turns, how many turns are needed on the secondary coil to produce the stated dc output according to this scheme? (Give the closest integer value.) (d) Estimate the current amplitude in the primary coil. (e) Based on the size of your device and assuming the transformer geometry shown in **Fig. P31.44**, estimate the length l of the primary coil. (f) Use Ampere's law to estimate the average strength of the magnetic field inside the core, assuming a relative permeability of 5000.

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?

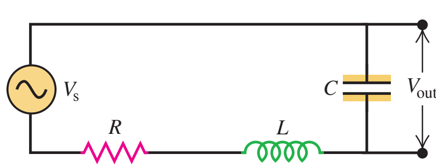
31.47 •• **A High-Pass Filter.** One application of L-R-C series circuits is to high-pass or low-pass filters, which filter out either the low- or high-frequency components of a signal. A high-pass filter is shown in **Fig. P31.47**, where the output voltage is taken across the L-R combination. (The 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 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.

Figure **P31.48**



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 $V = 100$ V, $R = 200\ \Omega$, $L = 2.0$ H, and $C = 0.50\ \mu$F. Compare to the light purple curve in Fig. 31.19. Discuss the behavior of I and P in the limits $\omega = 0$ and $\omega \rightarrow \infty$.

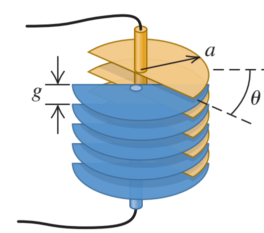
31.50 ••• An L-R-C series circuit is connected to an ac source of constant voltage amplitude V and variable angular frequency ω. Using the results of Problem 31.49, find an expression for (a) the amplitude V_L of the voltage across the inductor as a function of ω and (b) the amplitude V_C of the voltage across the capacitor as a function of ω. (c) Graph V_L and V_C as functions of ω for $V = 100$ V, $R = 200\ \Omega$, $L = 2.0$ H, and $C = 0.50\ \mu$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 •• **CP** Cell phones that use 4G technology receive signals broadcast between 2 GHz and 8 GHz. (a) If you want to create a simple L-R-C series circuit to detect a 4.0 GHz cell phone signal, what is the relevant value of the product LC, where L is the inductance and C is the capacitance? (b) If you choose a capacitor that has $C = 1.0 \times 10^{-15}$ F, what inductance do you need? (c) Suppose you want to wind your own toroidal inductor and fit it inside a box as thin as your cell phone. Based on the size of your phone, estimate the largest cross-sectional area possible for this. (d) Assume the largest allowable radius of the toroid is 1.0 cm and estimate the lowest number of windings needed to create your inductor, assuming the material inside has a relative permeability of 1.

31.53 ••• **CP** Five semicircular conducting plates, each with radius a, are attached to a conducting rod along their common axis such that the plates are parallel with a separation gap g between adjacent plates. A similar arrangement involving five additional semicircular conducting plates is attached to a second conducting rod symmetrically intertwined, upside down, with the first arrangement such that the plates can rotate freely, as shown in **Fig. P31.53**. The two five-plate constructions, with an adjacent plate separation of $g/2$, are insulated from each other, and wire leads are attached to the separate axles. This is a common type of variable capacitor, with a capacitance proportional to the overlap angle θ. (a) Determine a formula for the capacitance C in terms of a, g, and θ, where θ is in radians. (b) The plate radius is $a = 6.00$ cm while the gap distance is $g = 2.00$ mm. This variable capacitor is placed in series with a 100 μH inductor and a 10.0 Ω resistor. A voltage source is supplied by an amplified signal from an antenna, and the output voltage is taken across the capacitor. The circuit is used as a radio receiver that receives a signal frequency equal to the resonance frequency of the L-R-C circuit. To receive an AM radio signal at 1180 kHz, what angle θ, in degrees, should be selected? (c) If the input signal is a sinusoidal voltage with amplitude 100.0 mV and frequency 1180 kHz, what is the amplitude of the output voltage? (d) If the knob on the capacitor is turned so that $\theta = 120°$, to what frequency is the circuit tuned? (e) If the input signal has the frequency resonant with this setting and voltage amplitude 100 mV, what is the voltage amplitude of the output signal?

Figure **P31.53**



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 $R = 60.0\ \Omega$, $L = 0.800$ H, and $C = 3.00 \times 10^{-4}$ 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?

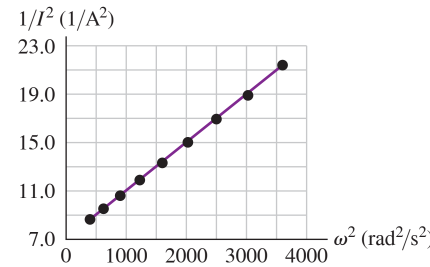
31.59 • In an *L-R-C* series circuit, the source has a voltage amplitude of 120 V, $R = 80.0\ \Omega$, and the reactance of the capacitor is 480 Ω. The voltage amplitude across the capacitor is 360 V. (a) What is the current amplitude in the circuit? (b) What is the impedance? (c) What two values can 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.

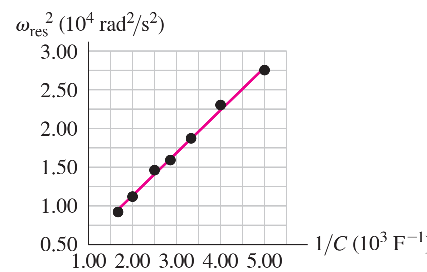
31.62 •• **DATA** A coworker of yours was making measurements of a large solenoid that is connected to an ac voltage source. Unfortunately, she left for vacation before she completed the analysis, and your boss has asked you to finish it. You are given a graph of $1/I^2$ versus ω^2 (**Fig. P31.62**), 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 P31.62



31.63 •• **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 ω_{res}^2 versus $1/C$ (**Fig. P31.63**). 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 P31.63



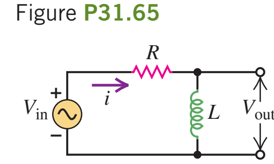
31.64 •• **DATA** You are given this table of data recorded for a circuit that has a resistor, an inductor with negligible resistance, and a capacitor, all in series with an ac voltage source:

f **(Hz)**	80	160
Z **(Ω)**	15	13
ϕ **(deg)**	−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. 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?

31.65 ••• **CP** The frequency response of a filter circuit is commonly specified, in decibels, as the attenuation $G = (20\text{ dB})\log_{10}(V_{out}/V_{in})$, where V_{in} is the input voltage amplitude and V_{out} is the output voltage amplitude. This is analogous to the definition of decibels given in Eq. (16.15). For the *R-L* circuit shown in **Fig. P31.65**, the input signal is $v_{in}(t) = V_{in}\cos(\omega t)$. (a) The current is $i(t) = I\cos(\omega t - \phi)$. What is the current amplitude I? (b) What is the phase angle ϕ? (c) What is the ratio V_{out}/V_{in}? (d) For what frequency $f = f_{3db}$ does this circuit have an attenuation of −3.0 dB? (The frequency f_{3db} is commonly cited as the nominal boundary demarcating the "pass band.") (e) If $R = 100\ \Omega$, what inductance L will "pass" frequencies higher than 10.0 kHz and block frequencies lower than 10.0 kHz?

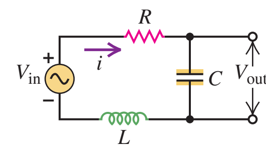
Figure P31.65



CHALLENGE PROBLEMS

31.66 ••• CALC The L-R-C series circuit shown in **Fig. P31.66** has an ac voltage source $v_{in}(t) = V_{in}\cos(\omega t)$, where V_{in} is the input voltage amplitude and ω is its angular frequency. The values of the resistance, inductance, and capacitance of the components are R, L, and C. The capacitance is variable. The output voltage is $v_{out}(t) = V_{out}\cos(\omega t + \theta)$. (a) What is the output voltage amplitude V_{out}? (b) What is the phase angle θ of the output voltage? (c) What is the value of θ at resonance? (d) If $R = 100\ \Omega$, $L = 1.00\ \mu\text{H}$, and $V_{in} = 10.0$ V, what value of C would result in a resonance frequency of 100 kHz? (e) In that case what would be the output voltage amplitude at resonance?

Figure P31.66



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.)

MCAT-STYLE 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.)

Key Example ✓ARIATION Problems

VP31.3.1 (a) $1.70 \times 10^6\ \Omega$ (b) 2.48×10^{-3} A
VP31.3.2 (a) $(0.300\ \text{V})\cos[(1.75 \times 10^3\ \text{rad/s})t]$ (b) 81.6 Ω (c) $(0.196\ \text{V})\cos[(1.75 \times 10^3\ \text{rad/s})t - \pi/2\ \text{rad}]$
VP31.3.3 (a) 128 Hz (b) both are 0.620 V (c) zero; zero
VP31.3.4 (a) 0.0215 A (b) 0.0116 A (c) 2.45 V (d) 1.72 μF
VP31.5.1 (a) 902 Hz (b) 316 Ω (c) 0.174 A (d) –36.1° (e) lags
VP31.5.2 (a) 387 Ω (b) 0.0422 A (c) 16.3 V (d) 1.48 V (e) 13.0 V
VP31.5.3 (a) 17.4 Ω (b) 8.65 mH
VP31.5.4 (a) $V_R = 11.4$ V, $V_L = 6.24$ V, $V_C = 34.1$ V (b) $v_R = -8.71$ V, $v_L = -4.03$ V, $v_C = +22.0$ V
VP31.7.1 (a) 474 W (b) 948 W (c) 30.4 Ω
VP31.7.2 (a) $X_L = 107\ \Omega$, $X_C = 699\ \Omega$ (b) –65.1° (c) 0.421
VP31.7.3 (a) 653 Ω (b) 0.0536 A (c) 0.395 W
VP31.7.4 (a) $X_L = 97.5\ \Omega$, $X_C = 85.5\ \Omega$ (b) 36.9° (c) 16.0 Ω

Bridging Problem

(a) 8.35×10^4 rad/s and 3.19×10^5 rad/s
(b) At 8.35×10^4 rad/s: $V_{source} = 49.5$ 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: $V_{source} = 49.5$ 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

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.

LEARNING OUTCOMES

In this chapter, you'll learn...

- **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.

You'll need to review...

- **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.

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:

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q_{\text{encl}}}{\epsilon_0} \qquad \text{(Gauss's law)} \tag{29.18}$$

$$\oint \vec{B} \cdot d\vec{A} = 0 \qquad \text{(Gauss's law for magnetism)} \tag{29.19}$$

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \qquad \text{(Faraday's law)} \tag{29.20}$$

$$\oint \vec{B} \cdot d\vec{l} = \mu_0\left(i_C + \epsilon_0 \frac{d\Phi_E}{dt}\right)_{\text{encl}} \qquad \text{(Ampere's law)} \tag{29.21}$$

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 $d\Phi_E/dt$.

Figure 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."

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**).

Figure 32.2 (a) When your mobile phone sends a text message or photo, the information is transmitted in the form of electromagnetic waves produced by electrons accelerating within the phone's circuits. (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

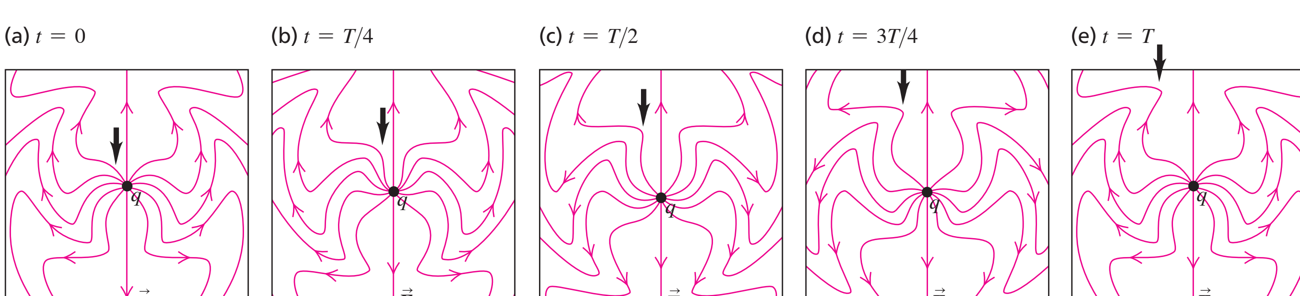
One way in which a point charge can be made to emit electromagnetic waves is by making it oscillate in simple harmonic motion, so that it has an acceleration at almost every instant (the exception is when the charge is passing through its equilibrium position). **Figure 32.3** shows some of the electric 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.

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,

Figure 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 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) $c = 299{,}792{,}458\ \text{m/s}$. 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 to 750×10^{-9} m), with corresponding frequencies from about 790 to 400 THz (7.9 to 4.0×10^{14} 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.

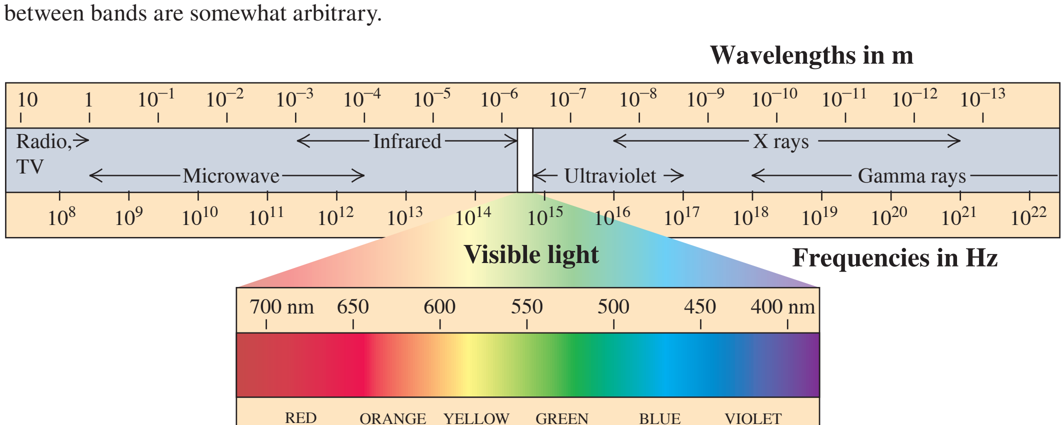
TABLE 32.1 Wavelengths of Visible Light

380–450 nm	Violet
450–495 nm	Blue
495–570 nm	Green
570–590 nm	Yellow
590–620 nm	Orange
620–750 nm	Red

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 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

Figure **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.

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?

ANSWER

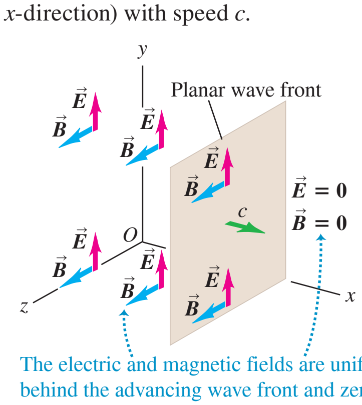
(a) **no,** (b) **no** A purely electric wave would have a varying electric field. Such a field necessarily generates a magnetic field through Ampere's law, Eq. (29.21), so a purely electric wave is impossible. In the same way, a purely magnetic wave is impossible: The varying magnetic field in such a wave would automatically give rise to an electric field through Faraday's law, Eq. (29.20).

32.2 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'll 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

Figure 32.5 An electromagnetic wave front. The plane representing the wave front moves to the right (in the positive x-direction) with speed c.



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'll consider more complex plane waves.

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:

$$\oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \tag{32.1}$$

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:

$$\oint \vec{E} \cdot d\vec{l} = -Ea \tag{32.2}$$

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

$$\frac{d\Phi_B}{dt} = Bac \tag{32.3}$$

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:

$$\underbrace{E}_{\text{Electric-field magnitude}} = \underbrace{c}_{\text{Speed of light in vacuum}}\ \underbrace{B}_{\text{Magnetic-field magnitude}} \tag{32.4}$$

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.

Figure **32.6** Gaussian surface for a transverse plane electromagnetic wave.

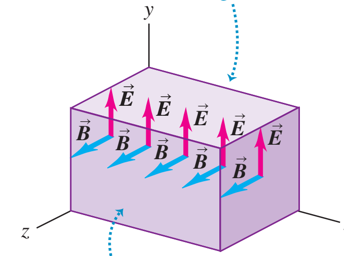
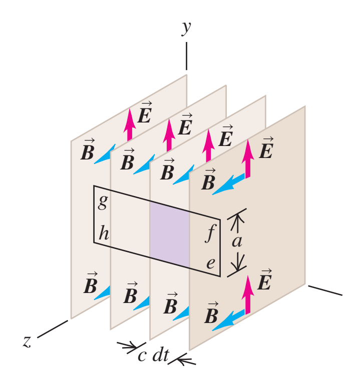
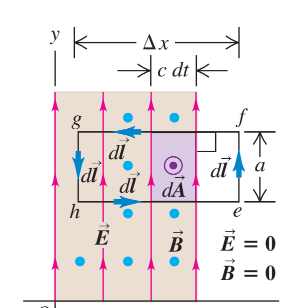


Figure **32.7** (a) Applying Faraday's law to a plane wave. (b) In a time dt, the magnetic flux through the rectangle in the xy-plane increases by an amount $d\Phi_B$. This increase equals the flux through the shaded rectangle with area $ac\,dt$; that is, $d\Phi_B = Bac\,dt$. Thus $d\Phi_B/dt = Bac$.



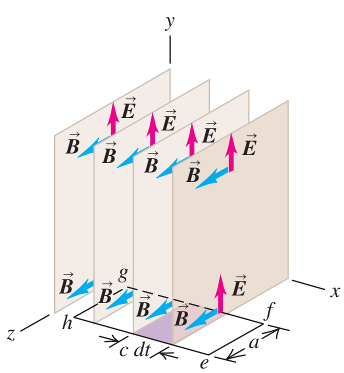


Finally, let's do a similar calculation with Ampere's law, the last of Maxwell's equations. There is no conduction current ($i_C = 0$), so Ampere's law is

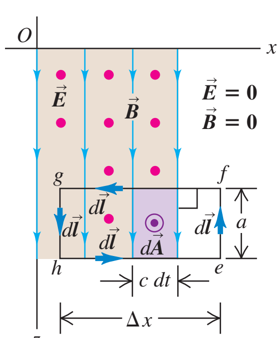
$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt} \tag{32.5}$$

Figure **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 to four significant figures, we find

$$\begin{aligned} c &= \frac{1}{\sqrt{(8.854 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2})(1.257 \times 10^{-6}\ \mathrm{N/A^2})}} \\ &= 2.998 \times 10^8\ \mathrm{m/s} \end{aligned}$$

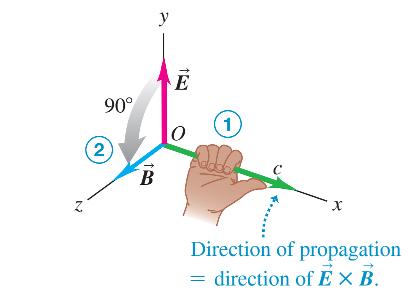
Our assumed wave is consistent with all of Maxwell's equations, provided that the wave front moves with the speed given above, which is the speed of light! Recall that the *exact* value of c is defined to be 299,792,458 m/s; both ϵ_0 and μ_0 have small uncertainties (see Sections 21.3 and 28.4), but their product in Eq. (32.9) has *zero* uncertainty.

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.

Figure **32.9** A right-hand rule for electromagnetic waves relates the directions of $\vec{E}$ and $\vec{B}$ and the direction of propagation.



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'll 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'll study polarization in greater detail in Chapter 33.

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.

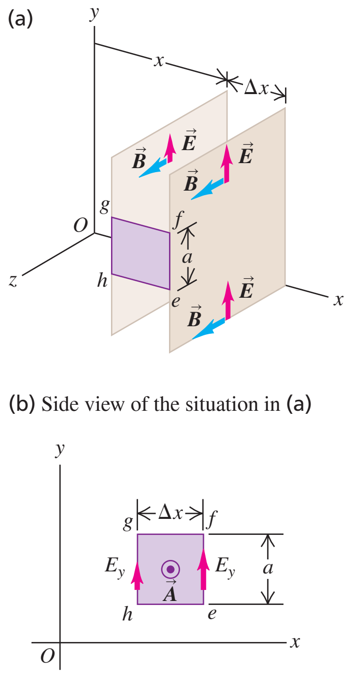
During our discussion of mechanical waves in Section 15.3, we showed that a function $y(x, t)$ that represents the displacement of any point in a mechanical wave traveling along the x-axis must satisfy a differential equation, Eq. (15.12):

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x, t)}{\partial t^2} \qquad (32.10)$$

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

Figure **32.10** Faraday's law applied to a rectangle with height a and width Δx parallel to the xy-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 $x + \Delta x$.

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

$$\oint \vec{E} \cdot d\vec{l} = -E_y(x, t)a + E_y(x + \Delta x, t)a$$
$$= a[E_y(x + \Delta x, t) - E_y(x, t)] \tag{32.11}$$

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

$$\frac{d\Phi_B}{dt} = \frac{\partial B_z(x, t)}{\partial t} a\,\Delta x$$

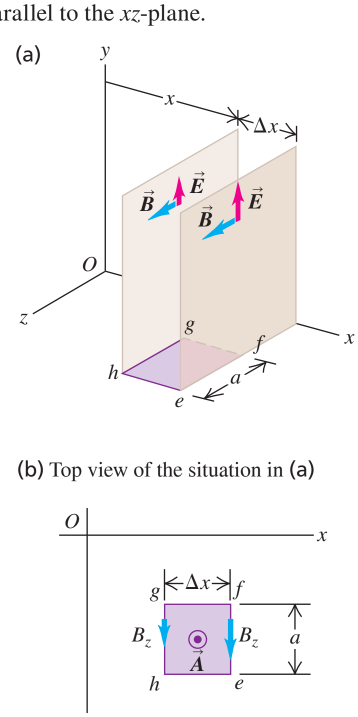
We use partial-derivative notation because B_z is a function of both x and t. When we substitute this expression and Eq. (32.11) into Faraday's law, Eq. (32.1), we get

$$a[E_y(x + \Delta x, t) - E_y(x, t)] = -\frac{\partial B_z}{\partial t} a\,\Delta x$$
$$\frac{E_y(x + \Delta x, t) - E_y(x, t)}{\Delta x} = -\frac{\partial B_z}{\partial t}$$

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

$$\frac{\partial E_y(x, t)}{\partial x} = -\frac{\partial B_z(x, t)}{\partial t} \tag{32.12}$$

Figure **32.11** Ampere's law applied to a rectangle with height a and width Δx parallel to the xz-plane.



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

$$\frac{\partial^2 E_y(x,t)}{\partial x^2} = \epsilon_0\mu_0\frac{\partial^2 E_y(x,t)}{\partial t^2} \qquad \text{(electromagnetic wave equation in vacuum)} \qquad (32.15)$$

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

$$\frac{1}{v^2} = \epsilon_0\mu_0 \qquad \text{or} \qquad v = \frac{1}{\sqrt{\epsilon_0\mu_0}}$$

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.

ANSWER

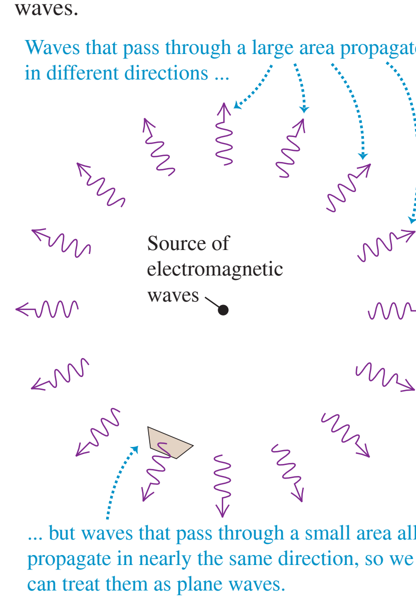
(a) **positive y-direction**, (b) **negative x-direction**, (c) **positive y-direction** You can verify these answers by using the right-hand rule to show that $\vec{E} \times \vec{B}$ in each case is in the direction of propagation, or by using the rule shown in Fig. 32.9.

32.3 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.

Figure 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

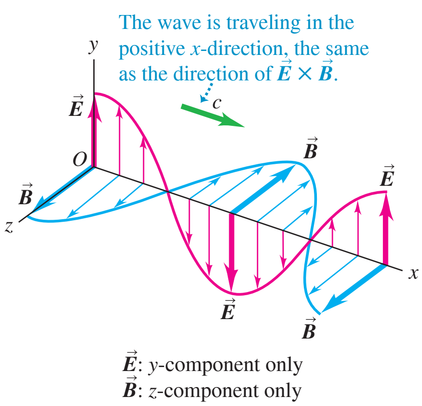
$$\lambda = \frac{3 \times 10^8 \text{ m/s}}{10^8 \text{ Hz}} = 3 \text{ m}$$

Figure 32.4 shows the inverse proportionality between wavelength and frequency.

Fields of a Sinusoidal Wave

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.

Figure **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.



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.

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

$$E_y(x, t) = E_{max}\cos(kx - \omega t) \qquad B_z(x, t) = B_{max}\cos(kx - \omega t) \qquad (32.16)$$

We can also write the wave functions in vector form:

Sinusoidal electromagnetic plane wave, propagating in $+x$-direction:

$$\underbrace{\vec{E}}_{\text{Electric field}}(x, t) = \hat{\jmath}\,\underbrace{E_{max}}_{\text{Electric-field magnitude}}\cos(\underbrace{k}_{\text{Wave number}}x - \underbrace{\omega}_{\text{Angular frequency}} t)$$
$$\underbrace{\vec{B}}_{\text{Magnetic field}}(x, t) = \hat{k}\,\underbrace{B_{max}}_{\text{Magnetic-field magnitude}}\cos(kx - \omega t) \qquad (32.17)$$

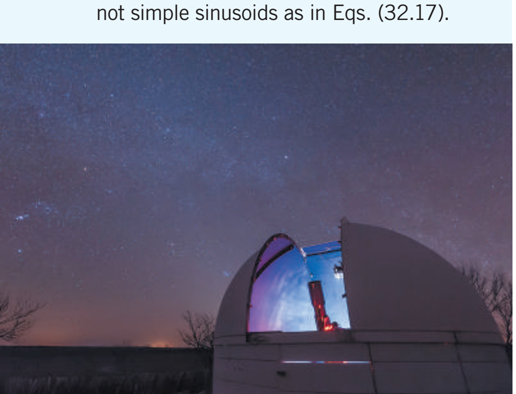
CAUTION **Two meanings of the symbol *k*** Note the two different k's in Eqs. (32.17): the unit vector $\hat{k}$ in the z-direction and the wave number k. Don't get these confused!

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:

$$\underbrace{E_{max}}_{\text{Electric-field amplitude}} = \underbrace{c}_{\text{Speed of light in vacuum}}\underbrace{B_{max}}_{\text{Magnetic-field amplitude}} \qquad (32.18)$$

APPLICATION Electromagnetic Plane Waves from Space The moon, sun, planets, and stars are so distant that the light we receive from them is well approximated by plane waves. This simplifies the problem of designing the optics of astronomical telescopes to collect this light (see Chapter 34). The light from objects in space is not all of a single frequency, however, so the waves are not simple sinusoids as in Eqs. (32.17).



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 32.30.)

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

$$\begin{aligned} \vec{E}(x, t) &= \hat{\jmath} E_{\max} \cos(kx + \omega t) \\ \vec{B}(x, t) &= -\hat{k} B_{\max} \cos(kx + \omega t) \end{aligned} \tag{32.19}$$

(sinusoidal electromagnetic plane wave, propagating in $-x$-direction)

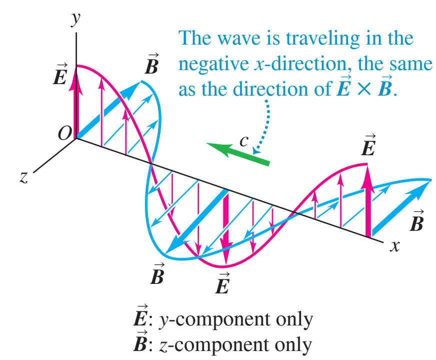


Figure 32.14 Representation of one wavelength of a linearly polarized sinusoidal plane electromagnetic wave traveling in the negative x-direction at $t = 0$. The fields are shown only for points along the x-axis. (Compare with Fig. 32.13.)

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.

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.)

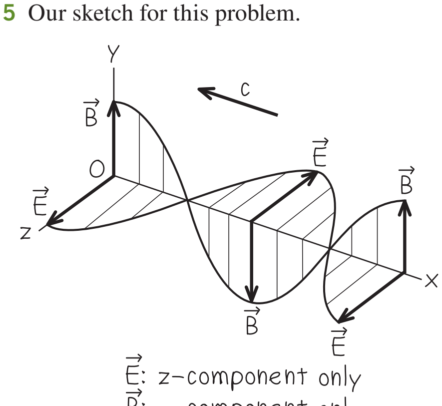
EXAMPLE 32.1 Electric and magnetic fields of a laser beam

WITH VARIATION PROBLEMS

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.

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.

Figure 32.15 Our sketch for this problem.



EXECUTE A possible pair of wave functions that describe the wave shown in Fig. 32.15 are

$$\begin{aligned} \vec{E}(x, t) &= \hat{k} E_{\max} \cos(kx + \omega t) \\ \vec{B}(x, t) &= \hat{\jmath} B_{\max} \cos(kx + \omega t) \end{aligned}$$

Continued

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.

KEYCONCEPT In a sinusoidal electromagnetic wave in vacuum, the electric field $\vec{E}$ and magnetic field $\vec{B}$ oscillate in phase with each other, are always perpendicular to each other, and are both perpendicular to the direction of propagation. The magnetic-field amplitude $B_{\max}$ equals the electric-field amplitude $E_{\max}$ divided by the speed of light in vacuum.

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

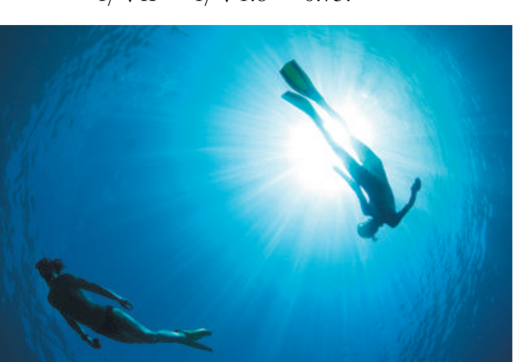
$$E = vB \qquad \text{and} \qquad B = \epsilon\mu v E \tag{32.20}$$

Following the same procedure as for waves in vacuum, we find that

Speed of electromagnetic waves in a dielectric

$$v = \frac{1}{\sqrt{\overset{\text{Permeability}}{\underset{\text{Permittivity}}{\epsilon\mu}}}} = \frac{1}{\sqrt{\underset{\substack{\text{Dielectric}\\ \text{constant}}}{K}\underset{\substack{\text{Relative}\\ \text{permeability}}}{K_m}}}\,\frac{1}{\sqrt{\underset{\substack{\text{Electric}\\ \text{constant}}}{\epsilon_0}\underset{\substack{\text{Magnetic}\\ \text{constant}}}{\mu_0}}} = \frac{\overset{\text{Speed of light in vacuum}}{c}}{\sqrt{KK_m}} \tag{32.21}$$

Figure 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*).

EXAMPLE 32.2 Electromagnetic waves in different materials

WITH VARIATION PROBLEMS

(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.

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.

KEYCONCEPT The speed of light in a transparent medium is slower than in vacuum. The greater the dielectric constant K of the medium and the greater the relative permeability K_m of the medium, the slower the speed. For waves of a given frequency, the slower the wave speed, the shorter the wavelength.

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.

ANSWER

(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 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

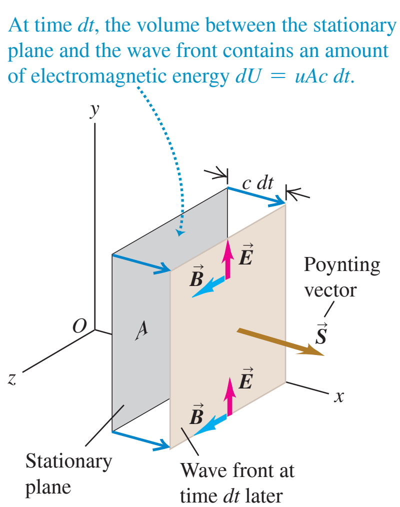
$$B = \frac{E}{c} = \sqrt{\epsilon_0 \mu_0}\, E \tag{32.24}$$

Combining Eqs. (32.23) and (32.24), we can also express the energy density u in a simple electromagnetic wave in vacuum as

$$u = \tfrac{1}{2}\epsilon_0 E^2 + \frac{1}{2\mu_0}\left(\sqrt{\epsilon_0\mu_0}\,E\right)^2 = \epsilon_0 E^2 \tag{32.25}$$

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.

Figure **32.17** A wave front at a time dt after it passes through the stationary plane with area A.



Electromagnetic Energy Flow and the Poynting Vector

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'll call S, is

$$S = \frac{1}{A}\frac{dU}{dt} = \epsilon_0 c E^2 \qquad \text{(in vacuum)} \tag{32.26}$$

Using Eqs. (32.4) and (32.9), you can derive the alternative forms

$$S = \frac{\epsilon_0}{\sqrt{\epsilon_0\mu_0}}E^2 = \sqrt{\frac{\epsilon_0}{\mu_0}}E^2 = \frac{EB}{\mu_0} \qquad \text{(in vacuum)} \tag{32.27}$$

The units of S are energy per unit time per unit area, or power per unit area. The SI unit of S is 1 J/s $\cdot$ m^2 or 1 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:**

Poynting vector in vacuum

$$\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B} \tag{32.28}$$

Electric field ($\vec{E}$); Magnetic field ($\vec{B}$); Magnetic constant (μ_0)

Figure **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 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 W/m^2.

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 $\vec{S}_{\text{av}} = \hat{\imath}S_{\text{av}}$, where

$$S_{\text{av}} = \frac{E_{\max}B_{\max}}{2\mu_0}$$

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

$$I = S_{\text{av}} = \frac{E_{\max}B_{\max}}{2\mu_0} = \frac{E_{\max}{}^2}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_{\max}{}^2 = \frac{1}{2}\epsilon_0 c E_{\max}{}^2 \qquad (32.29)$$

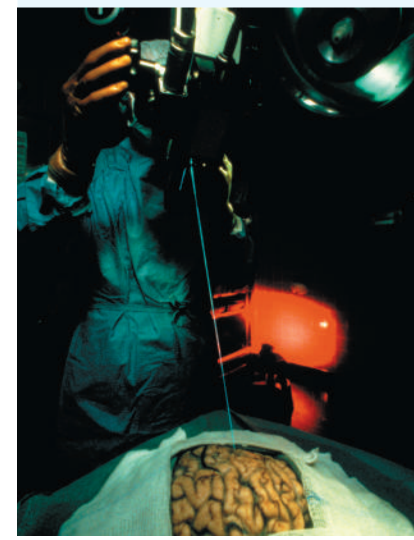
Magnitude of average Poynting vector · Magnetic constant · Speed of light in vacuum

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.

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.

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.



EXAMPLE 32.3 Energy in a nonsinusoidal wave

WITH VARIATION PROBLEMS

For the nonsinusoidal wave described in Section 32.2, suppose that $E = 100\ \text{V/m} = 100\ \text{N/C}$. Find the value of B, the energy density u, and the rate of energy flow per unit area S.

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),

$$\begin{aligned} u &= \epsilon_0 E^2 = (8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2)(100\ \text{N/C})^2 \\ &= 8.85 \times 10^{-8}\ \text{N/m}^2 = 8.85 \times 10^{-8}\ \text{J/m}^3 \end{aligned}$$

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):

$$\begin{aligned} S &= \epsilon_0 c E^2 = (8.85 \times 10^{-12}\ \text{C}^2/\text{N} \cdot \text{m}^2)(3.00 \times 10^8\ \text{m/s})(100\ \text{N/C})^2 \\ &= 26.5\ \text{W/m}^2 \end{aligned}$$

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.

KEYCONCEPT The Poynting vector $\vec{S} = (1/\mu_0)\vec{E} \times \vec{B}$ for an electromagnetic wave points in the direction of wave propagation. The average magnitude of $\vec{S}$ equals the intensity (average power per unit area) of the wave.

EXAMPLE 32.4 Energy in a sinusoidal wave

WITH VARIATION PROBLEMS

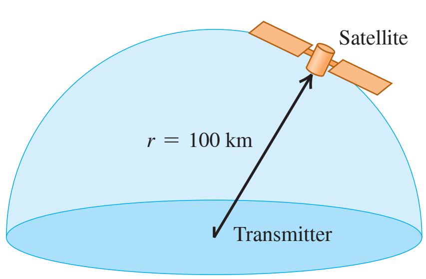
A radio station on the earth's surface emits a sinusoidal wave with average total power 50 kW (**Fig. 32.19**). Assuming that the transmitter radiates equally in all directions above the ground (which is unlikely in real situations), find the electric-field and magnetic-field amplitudes E_{max} and B_{max} detected by a satellite 100 km from the antenna.

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}.

EXECUTE The surface area of a hemisphere of radius $r = 100\ \text{km} = 1.00 \times 10^5\ \text{m}$ is

$$A = 2\pi R^2 = 2\pi(1.00 \times 10^5\ \text{m})^2 = 6.28 \times 10^{10}\ \text{m}^2$$

Figure 32.19 A radio station radiates waves into the hemisphere shown.



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

$$\begin{aligned} E_{max} &= \sqrt{2\mu_0 c S_{av}} \\ &= \sqrt{2(4\pi \times 10^{-7}\ \text{T} \cdot \text{m/A})(3.00 \times 10^8\ \text{m/s})(7.96 \times 10^{-7}\ \text{W/m}^2)} \\ &= 2.45 \times 10^{-2}\ \text{V/m} \end{aligned}$$

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).

KEYCONCEPT The magnitude S_{av} of the average Poynting vector of a sinusoidal electromagnetic wave depends on the field amplitudes E_{max} and B_{max}. The value of S_{av} is proportional to the product of E_{max} and B_{max}.

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

$$\frac{dp}{dV} = \frac{EB}{\mu_0 c^2} = \frac{S}{c^2} \qquad (32.30)$$

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

Poynting vector magnitude
Electric-field magnitude
Flow rate of electromagnetic momentum

$$\frac{1}{A}\frac{dp}{dt} = \frac{S}{c} = \frac{EB}{\mu_0 c} \tag{32.31}$$

Magnetic-field magnitude
Speed of light in vacuum
Momentum transferred per unit surface area per unit time
Magnetic constant

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 32.35). 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**).

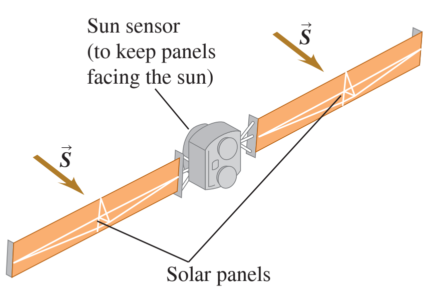
Figure **32.20** At the center of this interstellar gas cloud is a group of intensely luminous stars that exert tremendous radiation pressure on their surroundings. Aided by a "wind" of particles emanating from the stars, over the past million years the radiation pressure has carved out a bubble within the cloud 70 light-years across.

EXAMPLE 32.5 Power and pressure from sunlight

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.

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 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.

Figure **32.21** Solar panels on a satellite.



Continued

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 $p_{\mathrm{rad}} = 4.7 \times 10^{-6}\ \mathrm{Pa} = 4.7 \times 10^{-6}\ \mathrm{N/m^2}$. The total force F is the pressure p_{rad} times the area A:

$$F = p_{\mathrm{rad}}A = (4.7 \times 10^{-6}\ \mathrm{N/m^2})(4.0\ \mathrm{m^2}) = 1.9 \times 10^{-5}\ \mathrm{N}$$

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.

KEYCONCEPT An electromagnetic wave carries both energy and momentum. As a result an electromagnetic wave exerts pressure on any surface that either absorbs or reflects the wave.

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) $x = 3\lambda/4$.

ANSWER

(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 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.

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

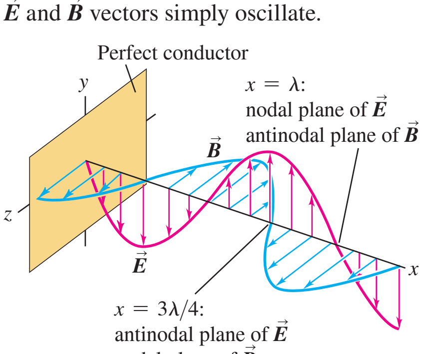
$$E_y(x, t) = E_{\max}[\cos(kx + \omega t) - \cos(kx - \omega t)]$$

$$B_z(x, t) = B_{\max}[-\cos(kx + \omega t) - \cos(kx - \omega t)]$$

We can expand and simplify these expressions by using the identities

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$$

Figure **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 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

$$x = 0, \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, \ldots \qquad \text{(nodal planes of } \vec{\boldsymbol{E}}\text{)} \tag{32.36}$$

These planes are called the **nodal planes** of the $\vec{\boldsymbol{E}}$ 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 $\vec{\boldsymbol{E}}$, 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 $\vec{\boldsymbol{B}}$, and they occur where

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots \qquad \text{(nodal planes of } \vec{\boldsymbol{B}}\text{)} \tag{32.37}$$

There is an antinodal plane of $\vec{\boldsymbol{B}}$ 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 $\vec{\boldsymbol{E}}$ 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 $\vec{\boldsymbol{E}}$ are midway between those of $\vec{\boldsymbol{B}}$, and vice versa; hence the nodes of $\vec{\boldsymbol{E}}$ coincide with the antinodes of $\vec{\boldsymbol{B}}$, 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 $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{B}}$ 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 $\vec{\boldsymbol{E}}$; 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

$$\lambda_n = \frac{2L}{n} \qquad (n = 1, 2, 3, \ldots) \tag{32.38}$$

Figure 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.

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.

EXAMPLE 32.6 Intensity in a standing wave

WITH VARIATION PROBLEMS

Calculate the intensity of the standing wave represented by Eqs. (32.34) and (32.35).

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.

KEYCONCEPT While there is energy flow in an electromagnetic standing wave, the intensity (the magnitude of the average Poynting vector) is zero at any point.

EXAMPLE 32.7 Standing waves in a cavity

WITH VARIATION PROBLEMS

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?

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

$$\lambda_1 = 2L = 2(1.50\text{ cm}) = 3.00\text{ cm}$$

$$f_1 = \frac{c}{2L} = \frac{3.00\times10^8\text{ m/s}}{2(1.50\times10^{-2}\text{ m})} = 1.00\times10^{10}\text{ Hz} = 10\text{ GHz}$$

(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 $\vec{E}$.

KEYCONCEPT In a sinusoidal electromagnetic standing wave, the electric field $\vec{E}$ and magnetic field $\vec{B}$ are 90° out of phase with each other. The antinodal planes of $\vec{E}$ are the nodal planes of $\vec{B}$, and the antinodal planes of $\vec{B}$ are the nodal planes 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?

ANSWER

| no There are places where $\vec{E} = 0$ 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 $\vec{B} = 0$ 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.

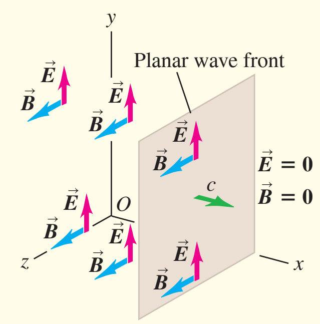
CHAPTER 32 SUMMARY

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 $\vec{E} \times \vec{B}$.

$$E = cB \qquad (32.4)$$

$$B = \epsilon_0\mu_0 cE \qquad (32.8)$$

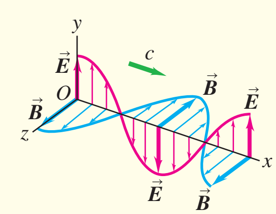
$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}} \qquad (32.9)$$



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.)

$$\vec{E}(x,t) = \hat{\jmath}E_{\max}\cos(kx - \omega t)$$
$$\vec{B}(x,t) = \hat{k}B_{\max}\cos(kx - \omega t) \qquad (32.17)$$

$$E_{\max} = cB_{\max} \qquad (32.18)$$



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.)

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{1}{\sqrt{KK_m}}\frac{1}{\sqrt{\epsilon_0\mu_0}} = \frac{c}{\sqrt{KK_m}} \qquad (32.21)$$

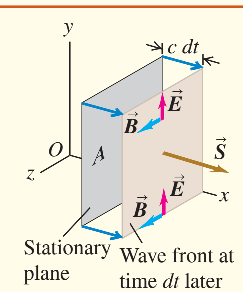
Energy and momentum in electromagnetic waves: The energy flow rate (power per unit area) in an electromagnetic wave in vacuum is given by the Poynting vector $\vec{S}$. The magnitude of the time-averaged value of the Poynting vector is called the intensity I of the wave. Electromagnetic waves also carry 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, $p_{rad} = I/c$; if the surface is a perfect reflector, $p_{rad} = 2I/c$. (See Examples 32.3–32.5.)

$$\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B} \qquad (32.28)$$

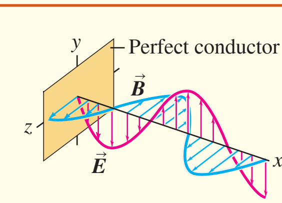
$$I = S_{av} = \frac{E_{\max}B_{\max}}{2\mu_0} = \frac{E_{\max}{}^2}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_{\max}{}^2 = \frac{1}{2}\epsilon_0 cE_{\max}{}^2 \qquad (32.29)$$

$$\frac{1}{A}\frac{dp}{dt} = \frac{S}{c} = \frac{EB}{\mu_0 c} \qquad (32.31)$$

(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.)



Chapter 32 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 32.1 and 32.2 (Section 32.3) before attempting these problems.

VP32.2.1 A sinusoidal electromagnetic wave in vacuum has magnetic-field amplitude 4.30×10^{-3} T and wave number 2.50×10^{6} rad/m. At a certain position and time the electric field $\vec{E}$ points in the $-y$-direction and the magnetic field $\vec{B}$ points in the $+x$-direction. Find (a) the amplitude of $\vec{E}$, (b) the wavelength, (c) the frequency, and (d) the direction of wave propagation.

VP32.2.2 The electric field in a sinusoidal electromagnetic wave in vacuum is given by $\vec{E}(y,t) = \hat{\imath}(2.45 \times 10^{6}\ \text{V/m}) \cos[(6.50 \times 10^{6}\,\text{rad/m})y - \omega t]$. Find (a) the wavelength, (b) the angular frequency ω, and (c) the direction of wave propagation. (d) Write the wave function for the magnetic field.

VP32.2.3 A dental laser emits a sinusoidal electromagnetic wave that propagates in vacuum in the positive x-direction with wavelength $2.94\ \mu$m. At $x = 0$ and $t = 0$ the electric field $\vec{E}$ points in the $+y$-direction and its magnitude equals the amplitude $E_{max} = 4.20$ MV/m. Find the magnitude and direction of (a) the magnetic field $\vec{B}$ at $x = 0$ and $t = 0$, (b) $\vec{E}$ at $x = 1.85\ \mu$m and $t = 0$, and (c) $\vec{B}$ at $x = 1.85\ \mu$m and $t = 0$.

VP32.2.4 The light from a red laser pointer has wavelength 6.35×10^{-7} m in vacuum. When this light is sent into a transparent material with relative permeability 1.00, its wavelength decreases to 3.47×10^{-7} m. Find (a) the frequency of the light in vacuum, (b) the frequency of the light in the material, (c) the speed of the light in the material, and (d) the dielectric constant of the material.

Be sure to review EXAMPLES 32.3 and 32.4 (Section 32.4) before attempting these problems.

VP32.4.1 A nonsinusoidal electromagnetic wave like that described in Section 32.2 has uniform electric and magnetic fields. The magnitude of the Poynting vector for this wave is 11.0 W/m^2. Find (a) the magnitudes of the electric and magnetic fields and (b) the energy density in the wave.

VP32.4.2 A sinusoidal electromagnetic wave in vacuum is given by the wave functions in Eqs. (32.17). Find the Poynting vector at (a) $x = 0$, $t = 0$; (b) $x = \lambda/4$, $t = 0$; (c) $x = \lambda/4$, $t = \pi/4\omega$.

VP32.4.3 A transmitter on the earth's surface radiates sinusoidal radio waves equally in all directions above the ground. An airplane flying directly over the radio station at an altitude of 12.5 km measures the wave from the station to have electric-field amplitude 0.360 V/m. Find (a) the magnetic-field amplitude and intensity measured by the airplane and (b) the total power emitted by the transmitter.

VP32.4.4 The sun emits light (which we can regard as a sinusoidal wave) equally in all directions. The distance from the sun to the earth is 1.50×10^{11} m, and the intensity of the sunlight that reaches the top of earth's atmosphere is 1.36 kW/m^2. Find (a) the amplitudes of the electric field and magnetic field at the top of the atmosphere and (b) the total power radiated by the sun.

Be sure to review EXAMPLES 32.6 and 32.7 (Section 32.5) before attempting these problems.

VP32.7.1 For the electromagnetic standing wave represented by Eqs. (32.34) and (32.35), find (a) the maximum magnitude of the Poynting vector anywhere in the standing wave; (b) the Poynting vector at $x = \lambda/8$, $t = 0$; (c) the Poynting vector at $x = \lambda/8$, $t = \pi/3\omega$; (d) the Poynting vector at $x = \lambda/8$, $t = 3\pi/4\omega$.

VP32.7.2 In a linearly polarized electromagnetic standing wave like that shown in Fig. 32.22, the amplitude of the magnetic field in the plane of the conductor is 1.20×10^{-7} T. The nodal plane of the magnetic field closest to the conductor is 3.60 mm from the conductor. Find (a) the wavelength, (b) the frequency, and (c) the amplitude of the electric field in the first nodal plane of the magnetic field.

VP32.7.3 You set up electromagnetic standing waves in a cavity that has two parallel conducting walls, one at $x = 0$ and one at $x = 4.56$ cm. Find (a) the three lowest standing-wave frequencies and their corresponding wavelengths, and (b) the positions of the nodal planes of the electric field for each of these frequencies.

VP32.7.4 A microwave oven sets up a standing wave of wavelength 12.2 cm between two parallel conducting walls 48.8 cm apart. Find (a) the wave frequency and (b) the number of antinodal planes of the electric field between the walls.

BRIDGING PROBLEM Detecting Electromagnetic Waves

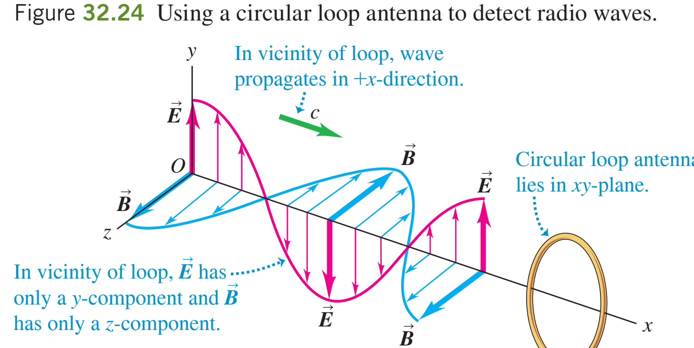
A circular loop of wire can be used as a radio antenna. If an 18.0-cm-diameter antenna is located 2.50 km from a 95.0 MHz source with a total power of 55.0 kW, what is the maximum emf induced in the loop? The orientation of the antenna loop and the polarization of the wave are as shown in **Fig. 32.24**. 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.

Figure 32.24 Using a circular loop antenna to detect radio waves.



2. Select the equations that you'll 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.
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

•, ••, •••: 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 • 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?

32.8 • 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 • 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 $\vec{B}(y, t)$.

32.10 • 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 • 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.

32.12 • 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?

32.13 • 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

32.14 • **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?

32.15 •• **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?

32.16 •• 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?

32.17 • A space probe 2.0×10^{10} m from a star measures the total intensity of electromagnetic radiation from the star to be 5.0×10^3 W/m^2. If the star radiates uniformly in all directions, what is its total average power output?

32.18 •• The energy flow to the earth from sunlight is about 1.4 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.

32.19 •• A point source emits monochromatic electromagnetic waves into air uniformly in all directions. You measure the amplitude E_{max} of the electric field at several distances from the source. After graphing your results as E_{max} versus $1/r$, you find that the data lie close to a straight line that has slope 75.0 N · m/C. What is the average power output of the source?

32.20 • A sinusoidal electromagnetic wave emitted by a mobile 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.

32.21 • 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.

32.22 • **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?

32.23 •• **BIO** **Laser Safety.** If the eye receives an average intensity greater than 1.0×10^2 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 W/cm^2.

32.24 • 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 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.

32.25 • **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?

32.26 • **CP** A totally reflecting disk has radius 8.00 μm and average density 600 kg/m^3. A laser has an average power output P_{av} spread uniformly over a cylindrical beam of radius 2.00 mm. When the laser beam shines upward on the disk in a direction perpendicular to its flat surface, the radiation pressure produces a force equal to the weight of the disk. (a) What value of P_{av} is required? (b) What average laser power is required if the radius of the disk is doubled?

Section 32.5 Standing Electromagnetic Waves

32.27 • **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?

32.28 • 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 $\vec{B}$?

32.29 •• An air-filled cavity for producing electromagnetic standing waves has two parallel, highly conducting walls separated by a distance L. One standing-wave pattern in the cavity produces nodal planes of the electric field with a spacing of 1.50 cm. The next-higher-frequency standing wave in the cavity produces nodal planes with a spacing of 1.25 cm. What is the distance L between the walls of the cavity?

PROBLEMS

32.30 •• **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.)

32.31 •• **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?

32.32 •• **CP** (a) Estimate how long it takes to bring 1 cup (237 mL) of room-temperature water to 100°C in a microwave oven. (b) How much heat is required to heat this water? (c) Divide the required heat by the time in part (a) to estimate the power supplied to the water by the microwaves. (d) If this energy were supplied by a plane electromagnetic wave incident from one direction, and if half of the wave energy was absorbed by the water, then the average wave intensity would be the twice power estimate divided by the cross-sectional area of the cup seen by the wave. Use this model to estimate the average intensity of the microwaves in the oven. (e) Use Eqs. (32.29) to estimate the amplitude of the electric field in a microwave oven.

32.33 • 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?

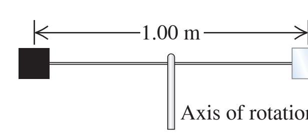
32.34 •• **CP** If we move a refrigerator magnet back and forth, we generate an electromagnetic wave that propagates away from us. Assume we move the magnet along the z-axis centered on the origin with an amplitude of 10 cm and with a frequency of 2.0 Hz. Consider a loop in the xy-plane centered on the origin with a radius of 2.0 cm. Assume that when the magnet is closest to the loop, the magnetic field within the loop has a spatially uniform value of 0.010 T. When the magnet has moved 10 cm in the direction away from the loop, the field in the loop becomes negligible. (a) Calculate the average rate at which the magnetic flux through the loop changes in time during each half-cycle of the motion of the magnet. (b) Use Faraday's law to estimate the average magnitude of the induced electric field at points on the loop. (c) Use Eq. (32.28) to estimate the average intensity of the electromagnetic wave in the xy-plane as it propagates radially away from the z-axis a distance of 2.0 cm. (d) Assume the intensity has the same value along the surface of a cylinder centered on the z-axis that extends 5.0 cm above and 5.0 cm below the xy-plane, and is negligible above or below this. (We are crudely postulating that the wave is emitted perpendicular to the axis of the motion in a narrow band.) Use this assumption to estimate the total power dissipated by the wave.

32.35 • 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?

32.36 • 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?

32.37 •• **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. P32.37**). 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 **P32.37**



32.38 •• 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?

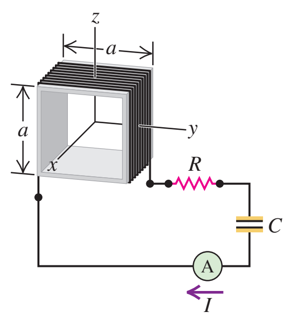
32.39 • **CP CALC** A cylindrical conductor with a circular cross section has a radius a and a resistivity ρ and carries a constant current I. (a) What are the magnitude and direction of the electric-field vector $\vec{E}$ at a point just inside the wire at a distance a from the axis? (b) What are the 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.

32.40 •• **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 W/m^2, and the wavelength of the wave is 6.90 m. What is the maximum emf induced in the loop?

32.41 • 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?

32.42 ••• **CP CALC** An antenna is created by wrapping a square frame with side length a by a wire N times and then connecting the leads to a resistor R and a capacitor C, as shown in **Fig. P32.42**. The loop itself has self-inductance L. A plane electromagnetic wave with electric field $\vec{E} = E_{max}\cos(kz - \omega t)\hat{j}$ propagates in the $+z$-direction. The origin is at the center of the frame. (a) What is the magnetic flux through the coil in the direction of the $+x$-axis? (*Hint:* The identity $\sin(A + B) + \sin(A - B) = 2\sin A \cos B$ may prove helpful.) (b) What is the emf generated in the coil? (c) The electromagnetic wave has frequency 4.00 Mz and intensity 100 W/m^2. The coil has $N = 50$ windings and side length a = 10.0 cm. It follows that its self-inductance is L = 78.0 μH. If the resistance in the circuit is R = 100 Ω, what value of the capacitance C results in the resonance frequency of the L-R-C circuit being equal to the frequency of the wave? (d) What rms value of the current I_{rms} flows in that case?

Figure **P32.42**



32.43 •• **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?

32.44 •• **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.

32.45 •• **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).]

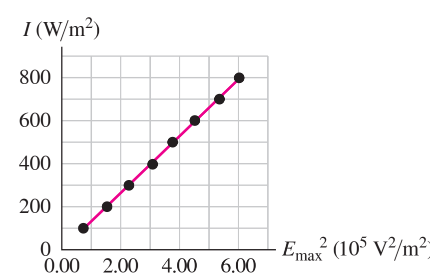
32.46 •• **DATA** The company where you work has obtained and stored five lasers in a supply room. You have been asked to determine the intensity of the electromagnetic radiation produced by each laser. The lasers are marked with specifications, but unfortunately different information is given for each 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 = 8.7×10^{-6} T
Laser D: diameter of cylindrical beam = 1.8 mm; force on totally reflecting surface = 6.0×10^{-8} N
Laser E: average energy density in beam = 3.0×10^{-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.

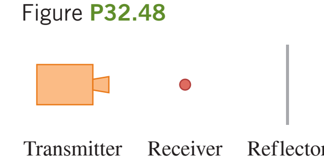
32.47 •• **DATA** Because the speed of light in vacuum (or air) has such a large value, it is very difficult to measure directly. To measure this speed, you conduct an experiment in which you measure the amplitude of the electric field in a laser beam as you change the intensity of the beam. **Figure P32.47** 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 **P32.47**



32.48 •• **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. P32.48**). 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.

Figure **P32.48**



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.

32.49 ••• CP When electromagnetic radiation strikes perpendicular to a flat surface, a totally absorbing surface feels radiation pressure I_0/c, where I_0 is the intensity of incident electromagnetic radiation. A totally reflecting surface feels twice that pressure. More generally, a surface absorbs a proportion e of the incident radiation and reflects a complementary proportion, $1 - e$, where e is the emissivity of the surface, as introduced in Chapter 17. Note that $0 \leq e \leq 1$. (a) Determine the radiation pressure p_{rad} in terms of I_0 and e. (b) Consider cosmic dust particles in outer space at a distance of 1.5×10^{11} m from the sun, where $I_{sun} = 1.4\ \text{kW/m}^2$. We can model these particles as tiny disks with $e = 0.61$, diameter 8.0 μm and mass 1.0×10^{-10} grams, all oriented perpendicular to the sun's rays. What is the force on one of these particles that is exerted by the radiation from the sun? (c) What is the ratio of this force to the attractive force of gravity exerted by the sun on the particle?

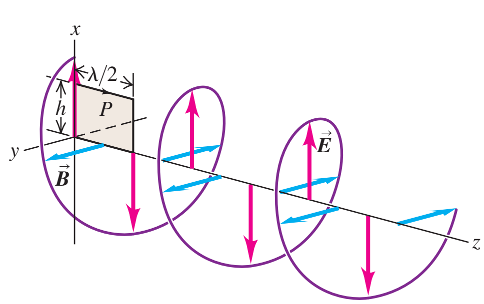
CHALLENGE PROBLEMS

32.50 ••• CALC An electromagnetic wave is specified by the following electric and magnetic fields:

$$\vec{E}(z, t) = E[\cos(kz - \omega t)\hat{\imath} + \sin(kz - \omega t)\hat{\jmath}]$$
$$\vec{B}(z, t) = B[-\sin(kz - \omega t)\hat{\imath} + \cos(kz - \omega t\hat{\jmath}]$$

where E and B are constant. (a) What is the value of the line integral $\oint \vec{E} \cdot d\vec{l}$ over the rectangular loop P shown in **Fig. P32.50**, which extends a distance h in the x-direction and half a wavelength in the z-direction, and passes through the origin? (b) What is the magnetic flux Φ_B through the loop P? (c) Use Faraday's law and the relationship $\omega/k = c$ to determine B in terms of E and c. (d) The wave specified above is "right circularly polarized." We can describe "left circularly polarized" light by swapping the signs on the two terms that involve sine functions. Superpose the above wave with such a reverse polarized analog. Write expressions for the electric and magnetic fields in that sum. Note that the result is a linearly polarized wave. (e) Consider right circularly polarized green light with wavelength 500 nm and intensity 100 W/m^2. What are the values of E and B? (*Hint:* Use Eq. (32.28) to determine the Poynting vector. Note that this is constant; its magnitude is the intensity of the light.)

Figure **P32.50**



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

$$\frac{dE}{dt} = \frac{q^2a^2}{6\pi\epsilon_0 c^3}$$

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 $1.72 \times 10^{-8}\ \Omega \cdot \text{m}$; 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.

MCAT-STYLE 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.

Key Example VARIATION Problems

VP32.2.1 (a) 1.29×10^6 V/m (b) 2.51×10^{-6} m (c) 1.19×10^{14} Hz (d) $+z$-direction

VP32.2.2 (a) 9.67×10^{-7} m (b) 1.95×10^{15} rad/s (c) $+y$-direction (d) $\vec{B}(y, t) = -\hat{k}(8.17 \times 10^{-3}\text{ T})\cos[(6.50 \times 10^6\text{ rad/m})y - (1.95 \times 10^{15}\text{ rad/s})t]$

VP32.2.3 (a) 1.40×10^{-2} T, $+z$-direction (b) 2.89 MV/m, $-y$-direction (c) 9.64×10^{-3} T, $-z$-direction

VP32.2.4 (a) 4.72×10^{14} Hz (b) 4.72×10^{14} Hz (c) 1.64×10^8 m/s (d) 3.35

VP32.4.1 (a) $E = 64.4$ V/m, $B = 2.15 \times 10^{-7}$ T (b) 3.67×10^{-8} J/m^3

VP32.4.2 (a) $\hat{\imath}(E_{max}B_{max}/\mu_0)$ (b) zero (c) $\hat{\imath}(E_{max}B_{max}/2\mu_0)$

VP32.4.3 (a) $B_{max} = 1.20 \times 10^{-9}$ T, $I = 1.72 \times 10^{-4}$ W/m^2 (b) 1.69×10^5 W

VP32.4.4 (a) $E_{max} = 1.01 \times 10^3$ V/m, $B_{max} = 3.38 \times 10^{-6}$ T (b) 3.85×10^{26} W

VP32.7.1 (a) $E_{max}B_{max}/\mu_0$ (b) zero (c) $\hat{\imath}\sqrt{3}\,E_{max}B_{max}/2\mu_0$ (d) $-\hat{\imath}\,E_{max}B_{max}/\mu_0$

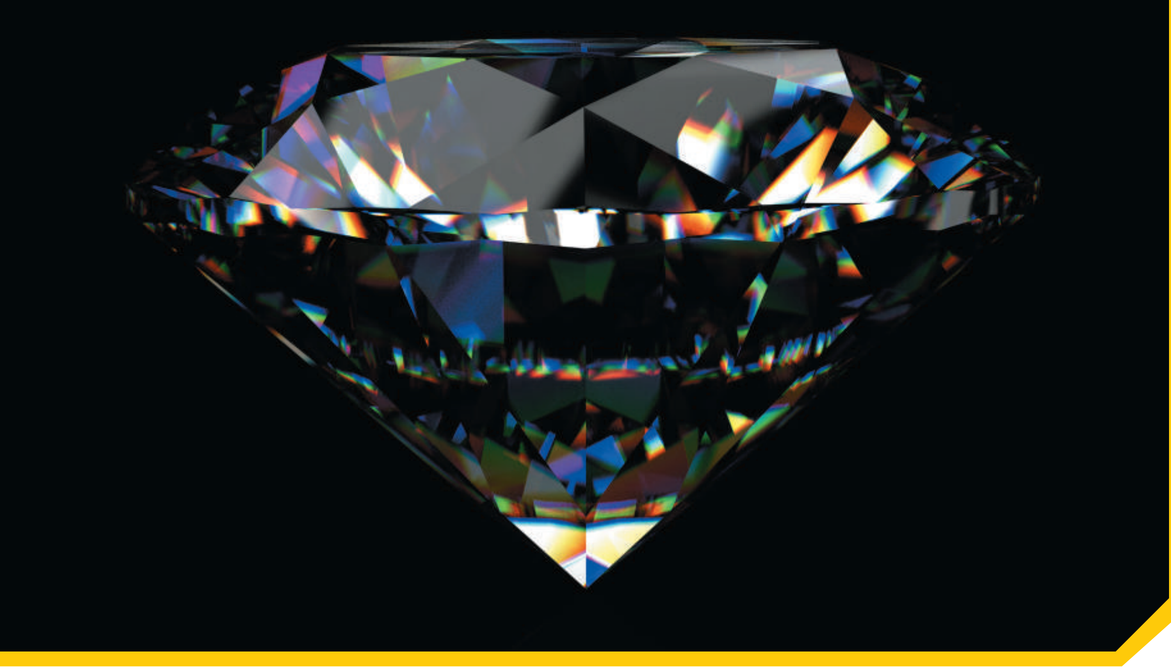
VP32.7.2 (a) 1.44 cm (b) 2.08×10^{10} Hz (c) 36.0 V/m

VP32.7.3 (a) $f_1 = 3.29 \times 10^9$ Hz, $\lambda_1 = 9.12$ cm; $f_2 = 6.58 \times 10^9$ Hz, $\lambda_2 = 4.56$ cm; $f_3 = 9.87 \times 10^9$ Hz, $\lambda_3 = 3.04$ cm (b) $x = 0$ and 4.56 cm; $x = 0$, 2.28 cm, and 4.56 cm; $x = 0$, 1.52 cm, 3.04 cm, and 4.56 cm

VP32.7.4 (a) 2.46×10^9 Hz (b) 8

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

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'll 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.

LEARNING OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

1.3 Speed of light in vacuum.
21.2 Polarization of an object 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.

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

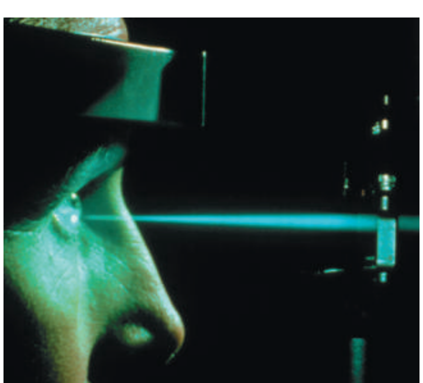
Figure 33.1 An electric heating element emits primarily infrared radiation. But if its temperature is high enough, it also emits a discernible amount of visible light.

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 objects 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 object 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.

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.

Figure 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.



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**).

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 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'll 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'll study wave phenomena and physical optics.

Figure **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.

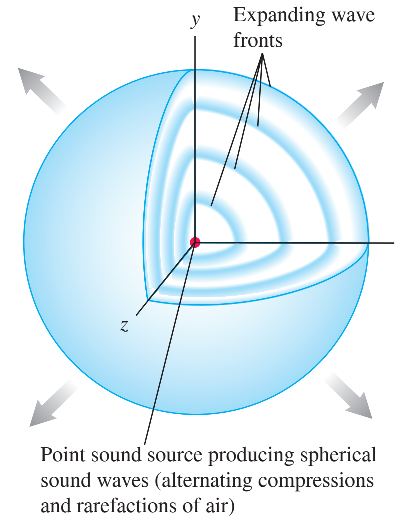
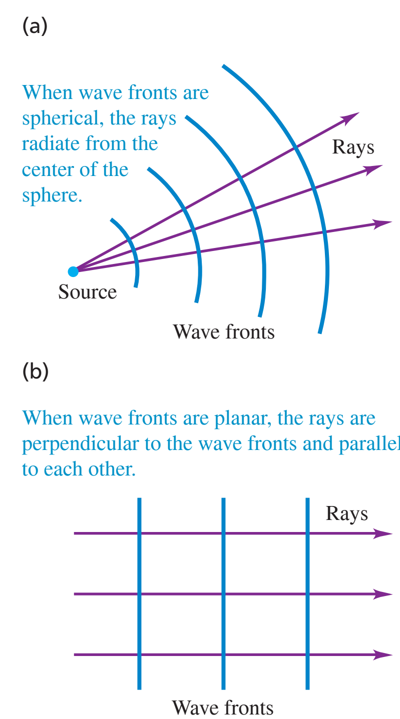


Figure **33.4** Wave fronts (blue) and rays (purple).



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.

ANSWER

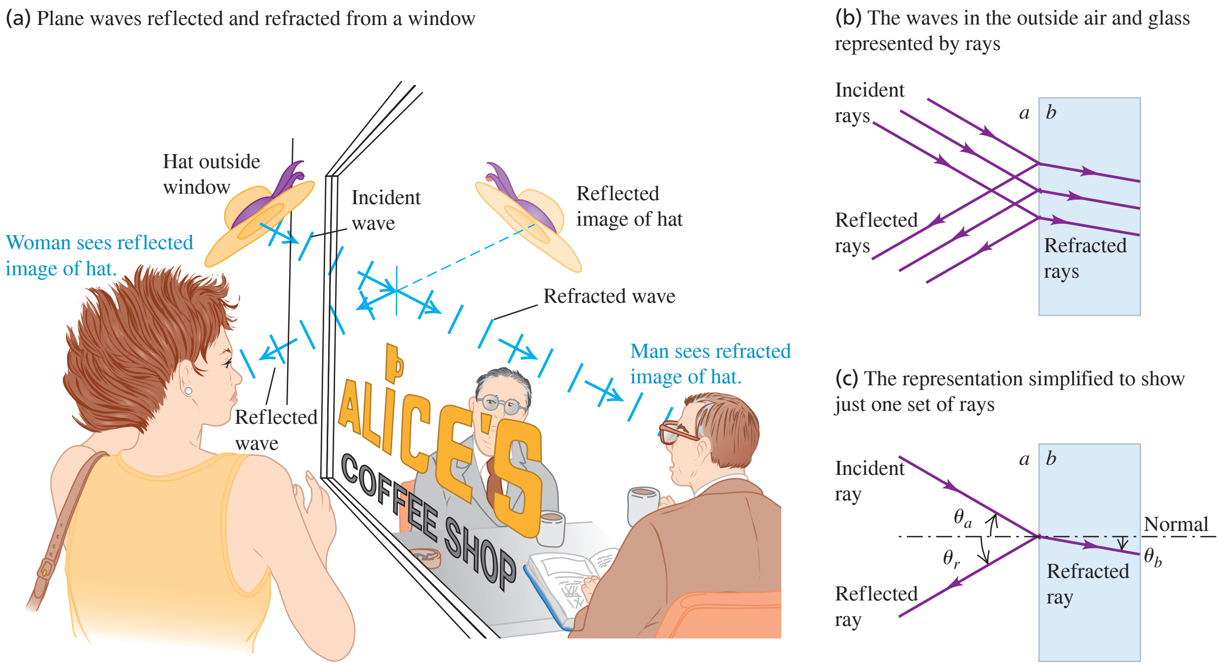
(iii) The waves go farther in the y-direction in a given amount of time than in the other directions, so the wave fronts are elongated in the y-direction.

33.2 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** (next page). 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.

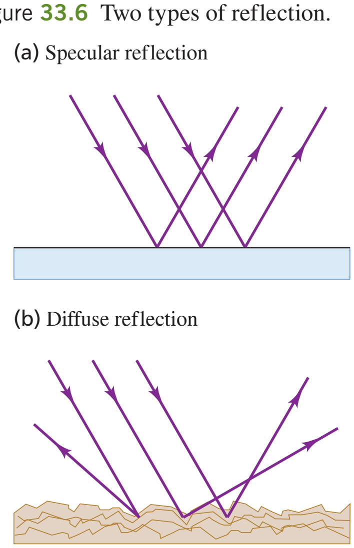
The segments of plane waves shown in Fig. 33.5a can be represented by bundles of rays forming *beams* of light (Fig. 33.5b). For simplicity we often draw only one ray in each beam (Fig. 33.5c). Representing these waves in terms of rays is the basis of geometric optics. We begin our study with the behavior of an individual ray.

Figure **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'll always mean *specular* reflection.

Figure **33.6** Two types of reflection.



The **index of refraction** of an optical material (also called the **refractive index**), denoted by n, plays a central role in geometric optics:

Index of refraction of an optical material

$$n = \frac{c}{v} \qquad (33.1)$$

c: Speed of light in vacuum
v: Speed of light in the material

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 larger the index of refraction in a material, the *slower* the wave speed in that material.

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

Figure 33.7 The laws of reflection and refraction.

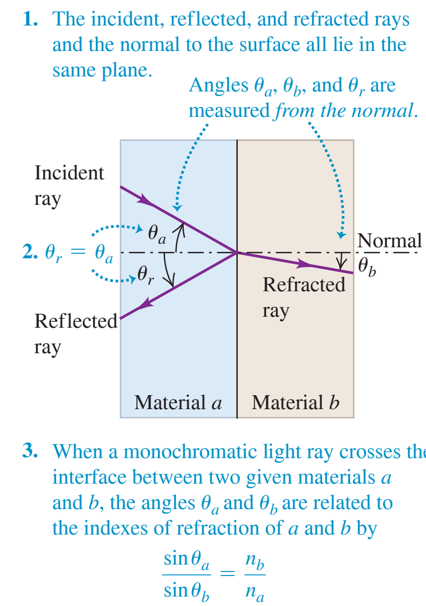
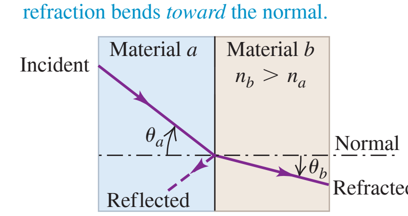
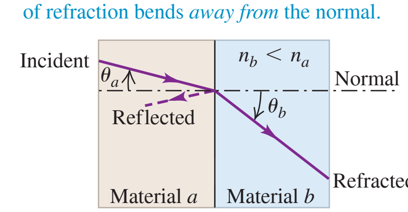


Figure 33.8 Refraction and reflection in three cases. (a) Material b has a larger index of refraction than material a. (b) Material b has a smaller index of refraction than material a. (c) The incident light ray is normal to the interface between the materials.

(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.

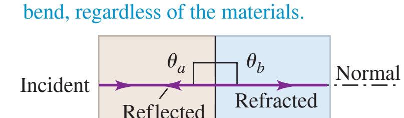
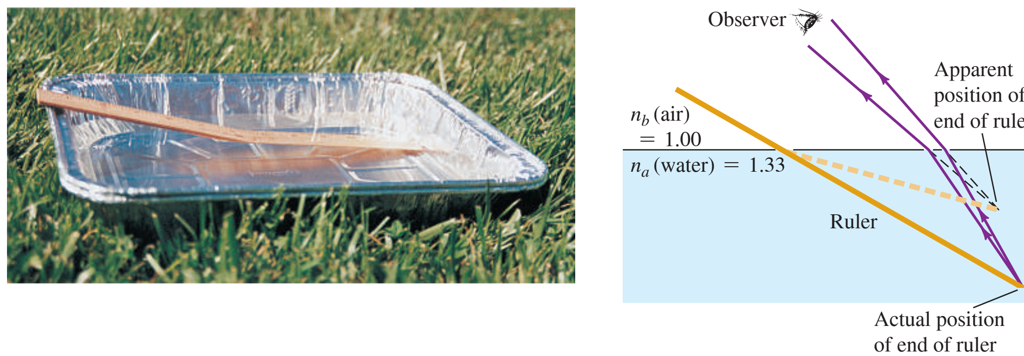


Figure 33.9 (a) This ruler is actually straight, but it appears to bend at the surface of the water. (b) Light rays from any submerged object bend away from the normal when they emerge into the air. As seen by an observer above the surface of the water, the object appears to be much closer to the surface than it actually is.

(a) A straight ruler half-immersed in water

(b) Why the ruler appears bent



air–water interface, so the rays appear to be coming from a position above 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, that person 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.

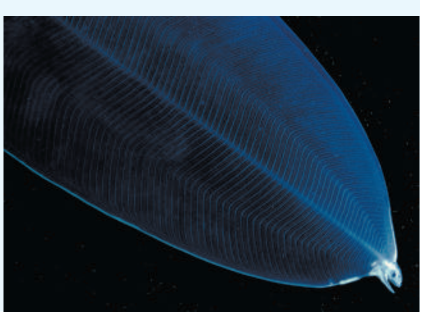
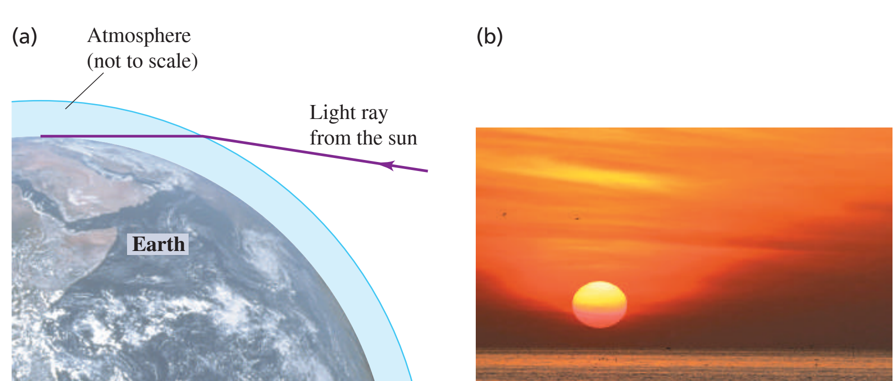


Figure 33.10 (a) The index of refraction of air is slightly greater than 1, so light rays from the setting sun bend downward when they enter our atmosphere. (The effect is exaggerated in this figure.) (b) Stronger refraction occurs for light coming from the lower limb of the sun (the part that appears 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.)



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'll 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'll 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).

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

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

$$\underbrace{\lambda}_{\text{Wavelength of light in a material}} = \frac{\overbrace{\lambda_0}^{\text{Wavelength of light in vacuum}}}{\underbrace{n}_{\text{Index of refraction of the material}}} \tag{33.5}$$

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.

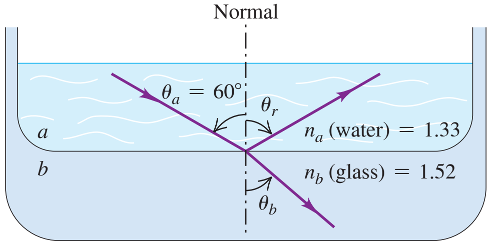
EXAMPLE 33.1 Reflection and refraction

WITH VARIATION PROBLEMS

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.

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 larger than n_a, so by Snell's law [Eq. (33.4)] θ_b is slightly smaller than θ_a.

Figure 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 $\theta_b < \theta_a$.

KEYCONCEPT When light strikes a smooth interface between two transparent materials, it can be both *reflected* back into the first material and *refracted* into the second material. The angles of incidence and reflection are always equal. The angle of refraction is determined by the angle of incidence and the indexes of refraction of the two materials (Snell's law). Always measure these angles from the normal to the interface.

EXAMPLE 33.2 Index of refraction in the eye

WITH VARIATION PROBLEMS

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.

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:

$$f_0 = \frac{c}{\lambda_0} = \frac{3.00 \times 10^8\text{ m/s}}{633 \times 10^{-9}\text{ m}} = 4.74 \times 10^{14}\text{ Hz}$$

KEYCONCEPT When a light wave travels from one material into another with a different index of refraction n, the wave speed and wavelength both change but the wave frequency remains the same. A larger n means a slower speed and a shorter wavelength; a smaller n means a faster speed and a longer wavelength.

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**. What is the ray's final direction relative to its original direction?

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.

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.

KEYCONCEPT When light undergoes multiple reflections, multiple refractions, or a combination of reflections and refractions, apply the law of reflection or the law of refraction separately at each interface.

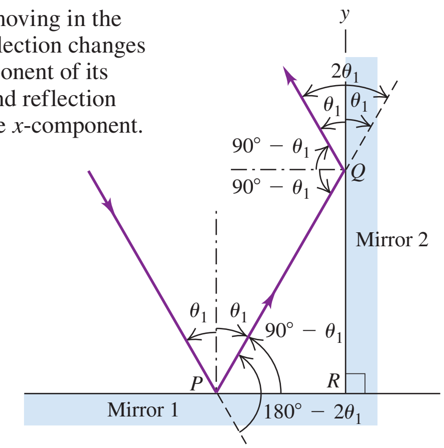
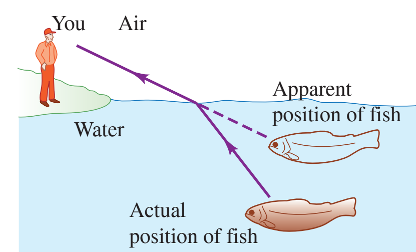


Figure 33.12 A ray moving in the *xy*-plane. The first reflection changes the sign of the *y*-component of its velocity, and the second reflection changes the sign of the *x*-component.

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?



ANSWER

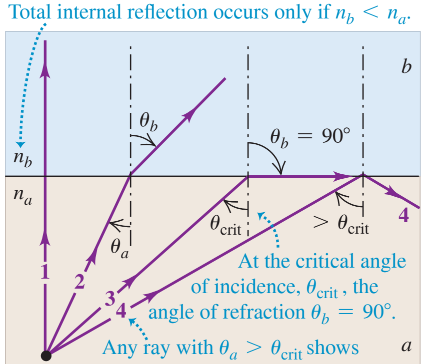
❙ (a) **(ii)**, (b) **(iii)** As shown in the figure, light rays coming from the fish bend away from the normal when they pass from the water ($n = 1.33$) into the air ($n = 1.00$). As a result, the fish appears to be higher in the water than it actually is. Hence you should aim a spear *below* the apparent 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 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,

$$\sin\theta_b = \frac{n_a}{n_b}\sin\theta_a$$

(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.

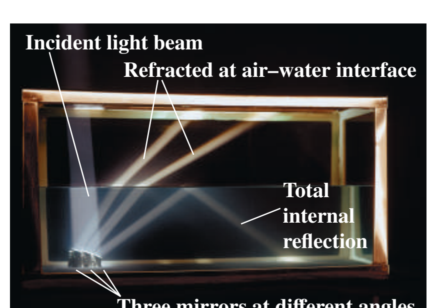


Figure 33.13 (a) Total internal reflection. The angle of incidence for which the angle of refraction is 90° is called the critical angle: This is the case for ray 3. The reflected portions of rays 1, 2, and 3 are omitted for clarity. (b) Rays of laser light enter the water in the fishbowl from above; they are reflected at the bottom by mirrors tilted at slightly different angles. One ray undergoes total internal reflection at the air–water interface.

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.

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

$$\underbrace{\sin\theta_{\text{crit}}}_{\text{Critical angle for total internal reflection}} = \frac{\overset{\text{Index of refraction of second material}}{n_b}}{\underset{\text{Index of refraction of first material}}{n_a}} \tag{33.6}$$

Total internal reflection will occur if the angle of incidence θ_a is larger than or equal to θ_{crit}.

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

$$\sin\theta_{\text{crit}} = \frac{1}{1.52} = 0.658 \qquad \theta_{\text{crit}} = 41.1°$$

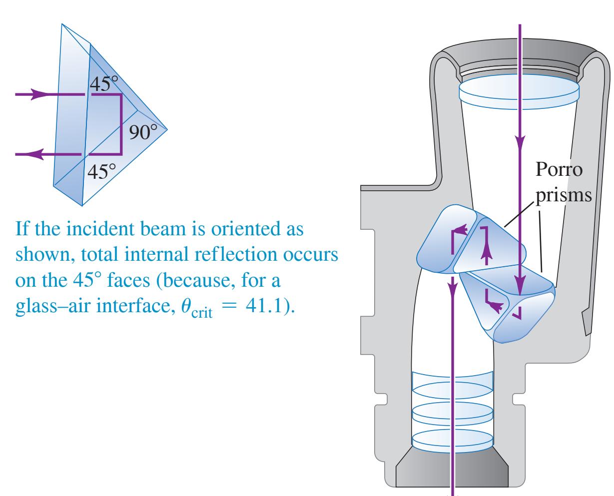
The light will be *totally reflected* if it strikes the glass–air surface at an angle of 41.1° or larger. Because the critical angle is slightly smaller than 45°, it is possible to use a prism with angles of 45°−45°−90° as a totally reflecting surface. As reflectors, totally reflecting prisms have some 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.

Figure 33.14 (a) Total internal reflection in a Porro prism. (b) A combination of two Porro prisms in binoculars.

(a) Total internal reflection in a Porro prism

(b) Binoculars use Porro prisms to reflect the light to each eyepiece.



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.

Figure 33.15 A transparent rod with refractive index greater than that of the surrounding material.

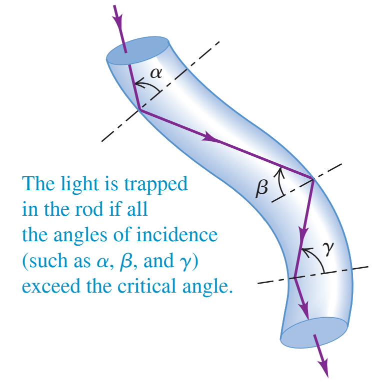
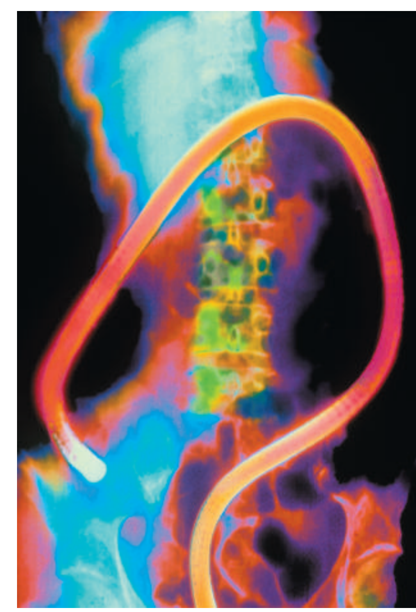


Figure 33.16 This colored x-ray image of a patient's abdomen shows an endoscope winding through the colon.



CONCEPTUAL EXAMPLE 33.4 A leaky periscope

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.

KEYCONCEPT For total internal reflection to occur when light in one medium strikes an interface with a second medium, two conditions must be met: The index of refraction n_b of the second medium must be smaller than the index of refraction n_a of the first medium, and the angle of incidence must be larger than the critical angle (whose value depends on the ratio n_b/n_a).

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°.

ANSWER **(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 $\sin\theta_{\text{crit}} = n_b/n_a$) must be smaller than the angle of incidence θ_a. In the first two cases both conditions are met: The critical angles are (i) $\theta_{\text{crit}} = \sin^{-1}(1/1.33) = 48.8°$ and (ii) $\theta_{\text{crit}} = \sin^{-1}(1.33/1.52) = 61.0°$, 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.

33.4 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 33.17 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 $\lambda = \lambda_0/n$.

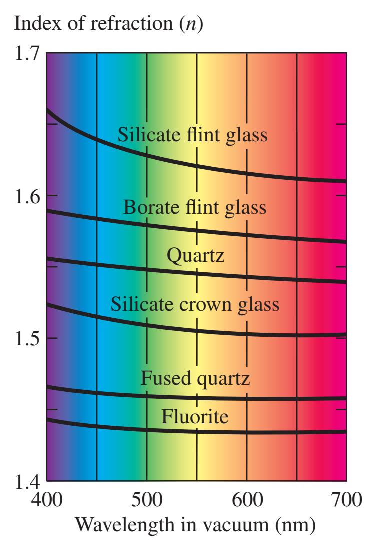
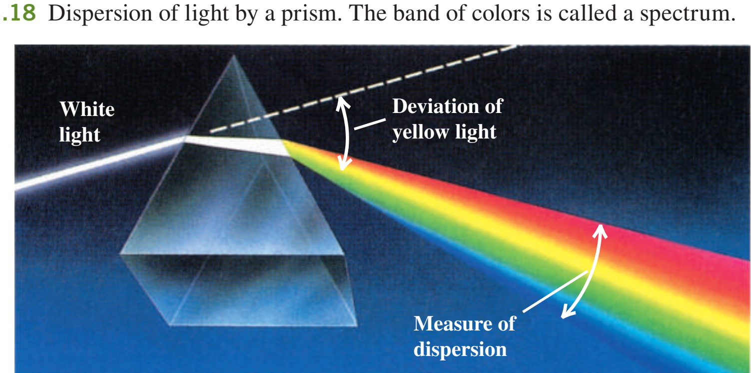


Figure 33.18 shows a ray of white light incident on a prism. The deviation (change of direction) produced by the prism increases with increasing index of refraction and frequency and decreasing wavelength. So violet light is deviated most, and red is deviated least. When it comes out of the prism, the 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. ?

Figure 33.18 Dispersion of light by a prism. The band of colors is called a spectrum.



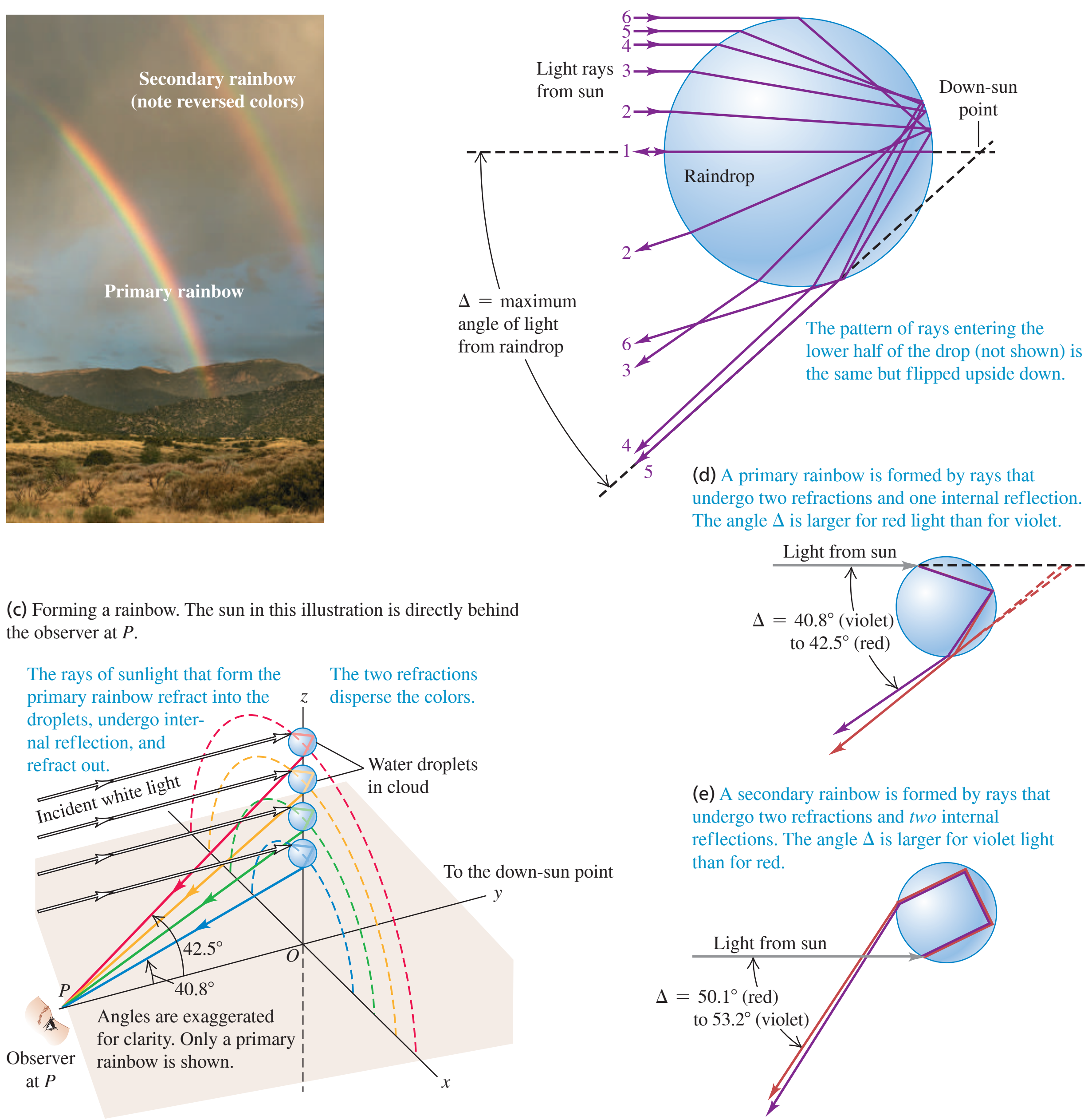
Rainbows

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

Figure **33.19** How rainbows form.

(a) A double rainbow

(b) The paths of light rays entering the upper half of a raindrop



(c) Forming a rainbow. The sun in this illustration is directly behind the observer at P.

(d) A primary rainbow is formed by rays that undergo two refractions and one internal reflection. The angle Δ is larger for red light than for violet.

(e) A secondary rainbow is formed by rays that undergo two refractions and *two* internal reflections. The angle Δ is larger for violet light than for red.

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),

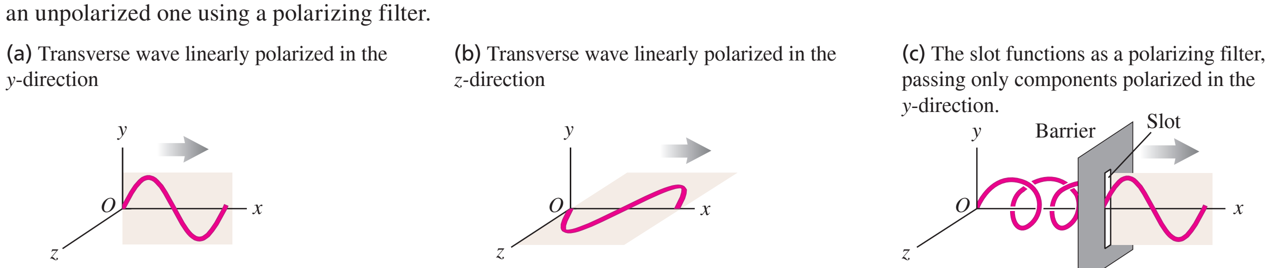
$$\vec{E}(x,t) = \hat{\jmath}E_{\max}\cos(kx - \omega t)$$

$$\vec{B}(x,t) = \hat{k}B_{\max}\cos(kx - \omega t)$$

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 an object, such as in response to a nearby charged object; we described this latter kind of polarization in Section 21.2 (see Fig. 21.7). Don't confuse these two concepts!

Figure 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.



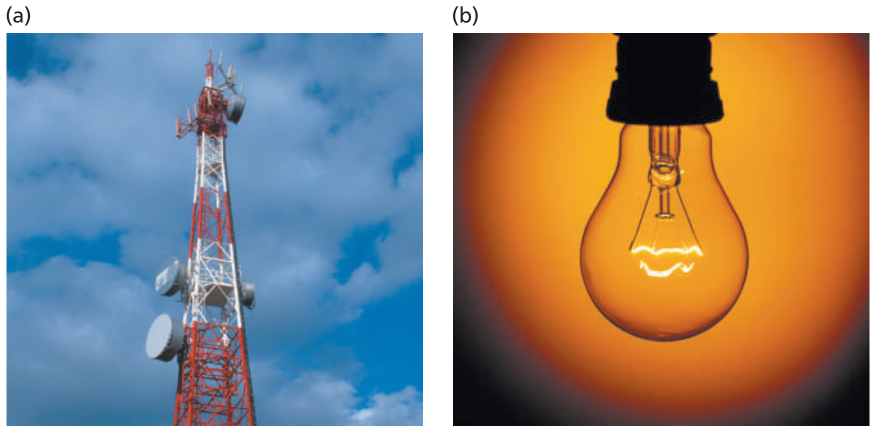


Figure 33.21 (a) Electrons in the red and white broadcast antenna oscillate vertically, producing vertically polarized electromagnetic waves that propagate away from the antenna in the horizontal direction. (The small gray antennas are for relaying cellular phone signals.) (b) No matter how this light bulb is oriented, the random motion of electrons in the filament produces unpolarized light waves.

Polarizing Filters

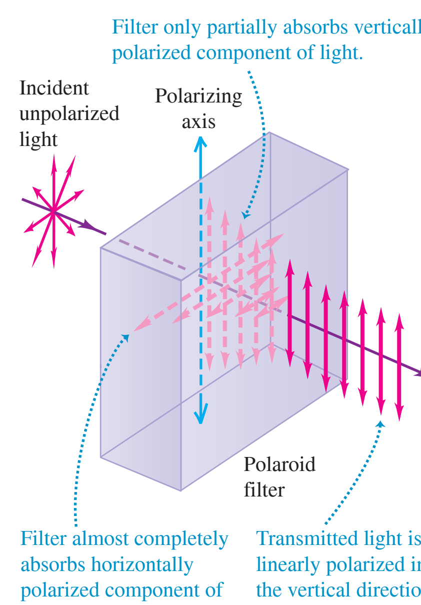
Waves emitted by a radio transmitter are usually linearly polarized. The vertical antennas that are used for radio broadcasting emit waves that, in a horizontal plane around the antenna, are polarized in the vertical direction (parallel to the antenna) (**Fig. 33.21a**).

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.

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.

Figure 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).



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 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'll assume that all

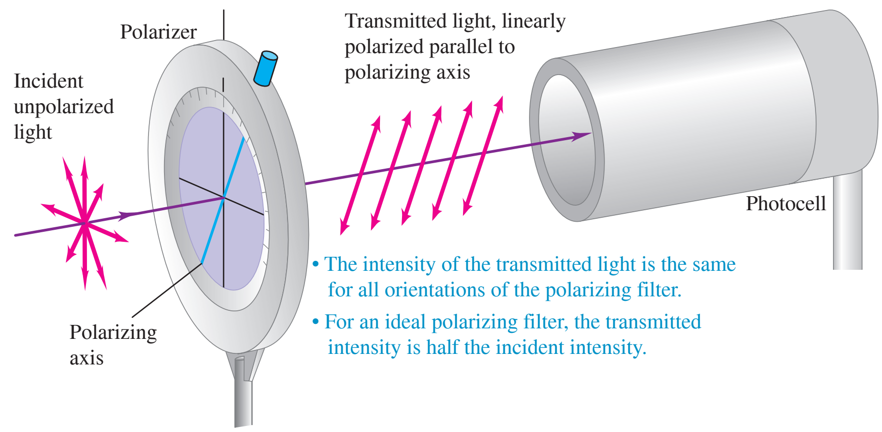


Figure 33.23 Unpolarized natural light is incident on the polarizing filter. The photocell measures the intensity of the transmitted linearly polarized light.

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* 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

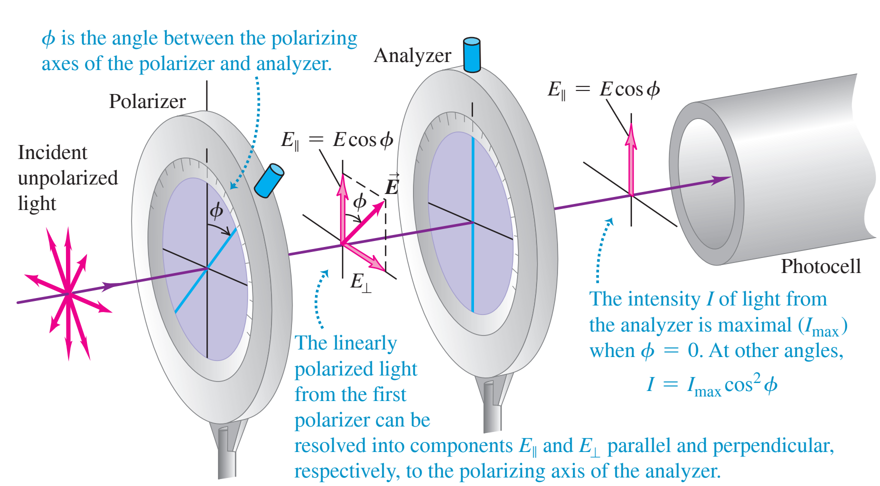


Figure 33.24 An ideal analyzer transmits only the electric field component parallel to its transmission direction (that is, its polarizing axis).

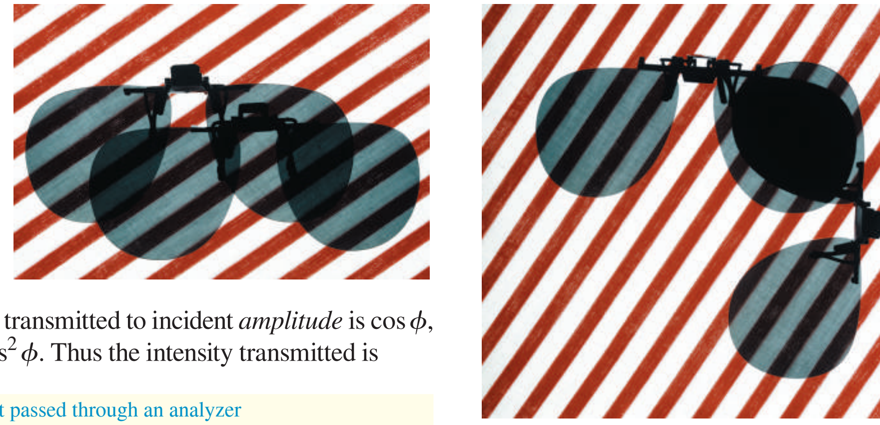


Figure **33.25** These photos show the view through Polaroid sunglasses whose polarizing axes are (left) perpendicular to each other ($\phi = 90°$) and (right) aligned with each other ($\phi = 0$). The transmitted intensity is greatest when the axes are aligned; it is zero when the axes are perpendicular.

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

Malus's law:

$$\underbrace{I}_{\text{Intensity of polarized light passed through an analyzer}} = \underbrace{I_{\max}}_{\text{Maximum transmitted intensity}} \cos^2 \underbrace{\phi}_{\text{Angle between polarization axis of light and polarizing axis of analyzer}} \qquad (33.7)$$

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

WITH VARIATION PROBLEMS

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°.

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.

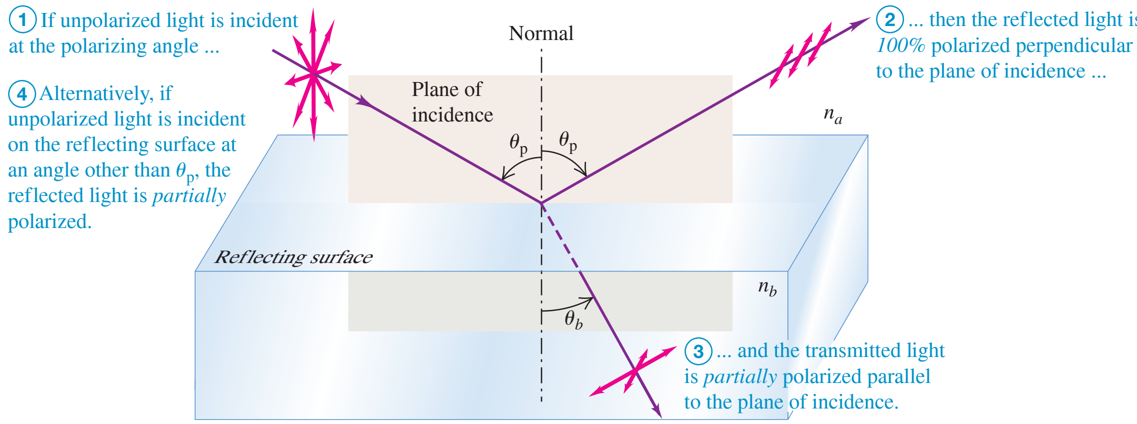
EXECUTE The incident light is unpolarized, so the intensity of the linearly polarized light transmitted by the first polarizer is $I_0/2$. From Eq. (33.7) with $\phi = 30°$, the second polarizer reduces the intensity by a further factor of $\cos^2 30° = \frac{3}{4}$. Thus the intensity transmitted by the second polarizer is

$$\left(\frac{I_0}{2}\right)\left(\frac{3}{4}\right) = \tfrac{3}{8}I_0$$

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 $\phi = 0$.

KEYCONCEPT When light enters a polarizing filter, the light that emerges is polarized along the polarizing axis of the filter. If the incident light is unpolarized, the intensity of the light that emerges is one-half that of the incident light; if the incident light is polarized at an angle θ to the polarization of the filter, the intensity of the light that emerges is $\cos^2\theta$ times that of the incident light.

Figure **33.26** When light is incident on a reflecting surface at the polarizing angle, the reflected light is linearly polarized.



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 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:

$$\tan\theta_p = \frac{n_b}{n_a} \qquad (33.8)$$

Polarizing angle (angle of incidence for which reflected light is 100% polarized)

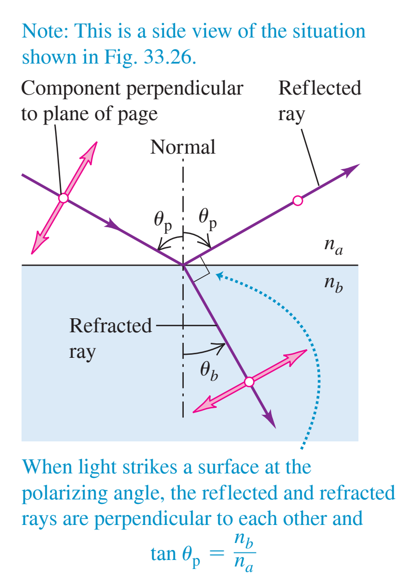
n_b: Index of refraction of second material

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.

Figure **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.



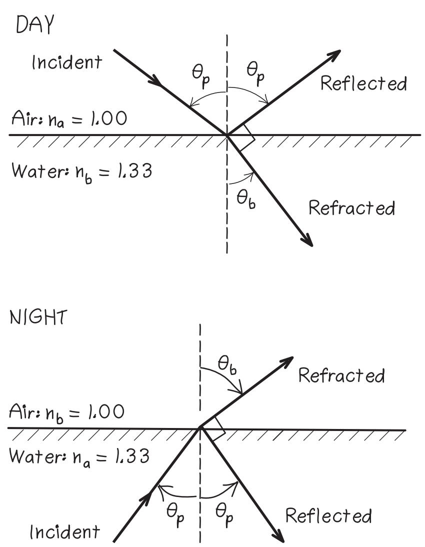
EXAMPLE 33.6 Reflection from a swimming pool's surface

WITH VARIATION PROBLEMS

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.

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. 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, $\theta_b = 90° - \theta_p$).

Figure **33.28** Our sketches for this problem.



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:

$$\sin\theta_b = \frac{n_a \sin\theta_p}{n_b} = \frac{1.00 \sin 53.1°}{1.33} = 0.601$$

$$\theta_b = \arcsin(0.601) = 36.9°$$

Note that the two polarizing angles found in parts (a) and (c) add to 90°. This is *not* an accident; can you see why?

KEYCONCEPT When light strikes an interface between two materials, how much is reflected depends on the polarization of the incident light. If the angle of incidence equals the polarizing angle, light polarized in the plane of incidence is completely refracted into the second material; none is reflected. Light polarized perpendicular to the plane of incidence is partially refracted and partially reflected, so the reflected light is completely polarized in the perpendicular direction.

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.**

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.

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 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 **33.29** 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.

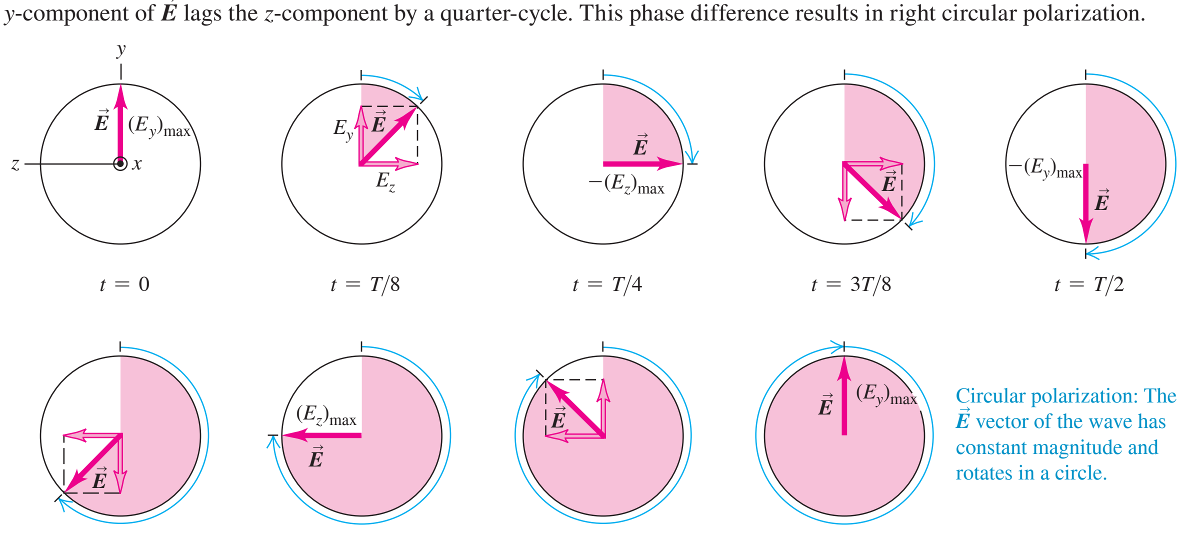


Figure **33.30** This plastic model of an artificial hip joint was photographed between two polarizing filters (a polarizer and an analyzer) with perpendicular polarizing axes. The colored interference pattern reveals the direction and magnitude of stresses in the model. Engineers use these results to help design the actual hip joint (used in hip replacement surgery), which is made of metal.

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.

ANSWER

(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.

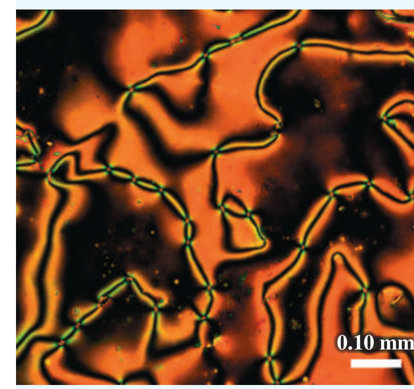
33.6 SCATTERING OF LIGHT

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'll 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** (next page)

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

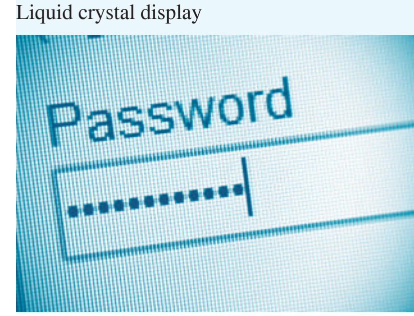
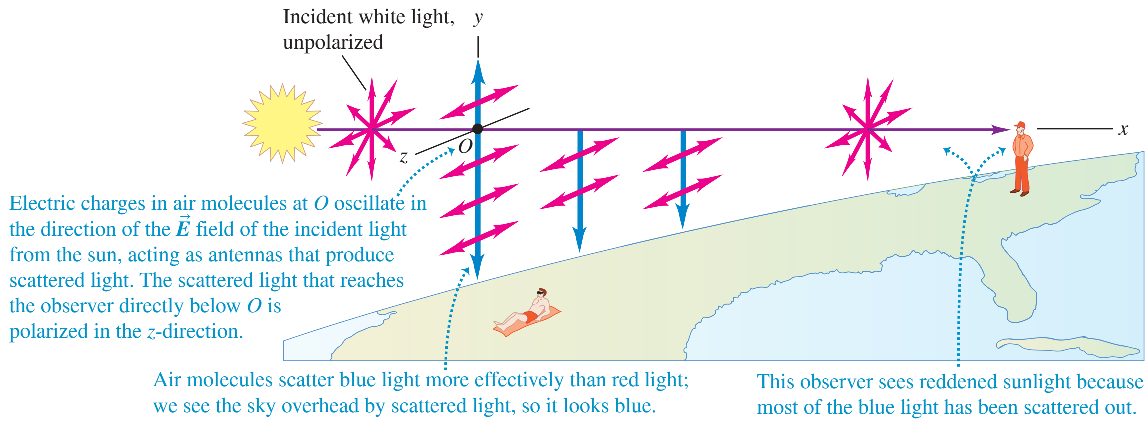


Figure **33.31** When the sunbathing observer on the left looks up, he sees blue, polarized sunlight that has been scattered by air molecules. The observer on the right sees reddened, unpolarized light when he looks at the sun.



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.

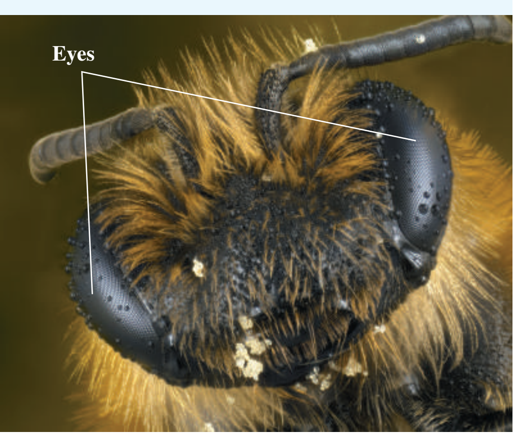
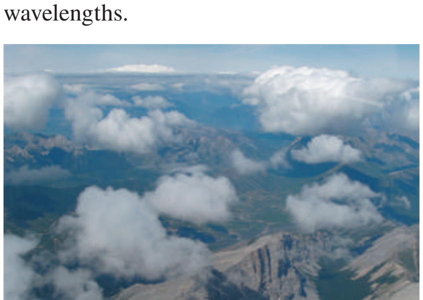


Figure **33.32** Clouds are white because they efficiently scatter sunlight of all wavelengths.



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.

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).

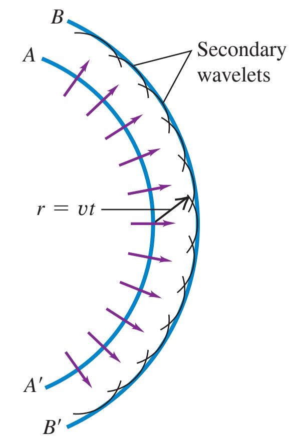
33.7 HUYGENS'S PRINCIPLE

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.

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'.

Figure 33.33 Applying Huygens's principle to wave front AA' to construct a new wave front BB'.



Reflection and Huygens's Principle

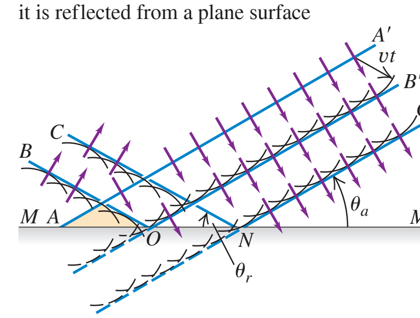
To derive the law of reflection from Huygens's principle, we consider a plane wave approaching a plane reflecting surface. In **Fig. 33.34a** the lines AA', OB', and NC' represent successive positions of a wave front approaching the surface MM'. Point A on the wave front AA' has just arrived at the reflecting 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.

Figure 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)

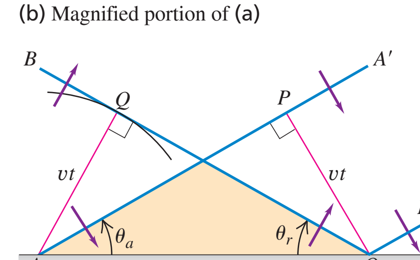
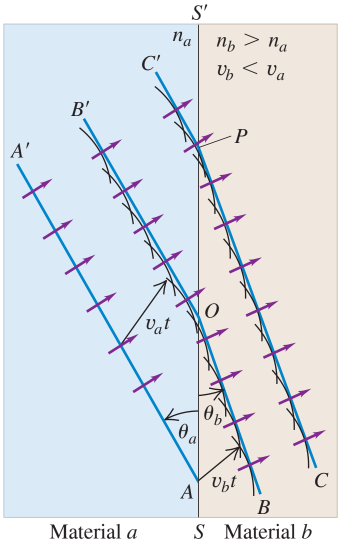
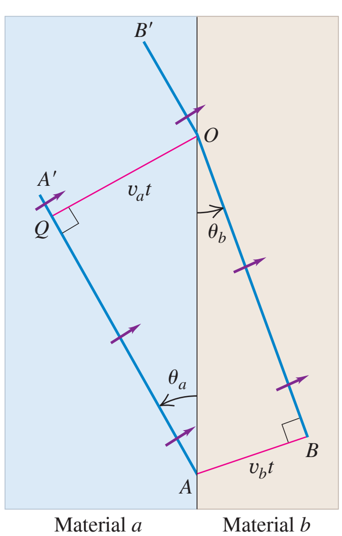


Figure 33.35 Using Huygens's principle to derive the law of refraction. The case $v_b < v_a$ $(n_b > n_a)$ is shown.

(a) Successive positions of a plane wave AA' as it is refracted by a plane surface



(b) Magnified portion of (a)



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

$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{v_a}{v_b} \tag{33.9}$$

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

$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{n_b}{n_a} \quad \text{or}$$

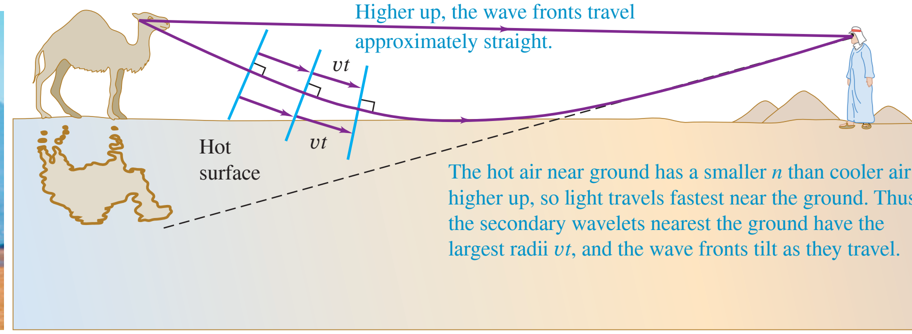
$$n_a \sin\theta_a = n_b \sin\theta_b$$

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.

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.

Figure 33.36 How mirages are formed.



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.

ANSWER

(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.

CHAPTER 33 SUMMARY

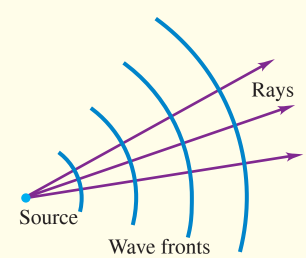
Light and its properties: Light is an electromagnetic wave. When emitted or absorbed, it also shows particle properties. It is emitted by accelerated electric charges.

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.)

$$n = \frac{c}{v} \quad (33.1)$$

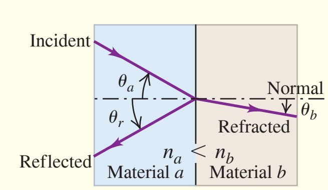
$$\lambda = \frac{\lambda_0}{n} \quad (33.5)$$



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.)

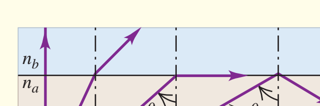
$$\theta_r = \theta_a \quad (33.2)$$
(law of reflection)

$$n_a \sin\theta_a = n_b \sin\theta_b \quad (33.4)$$
(law of refraction)



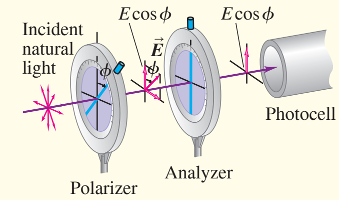
Total internal reflection: When a ray travels in a material of index of refraction n_a toward a material of index $n_b < n_a$, total internal reflection occurs at the interface when the angle of incidence equals or exceeds a critical angle θ_{crit}. (See Example 33.4.)

$$\sin\theta_{crit} = \frac{n_b}{n_a} \quad (33.6)$$



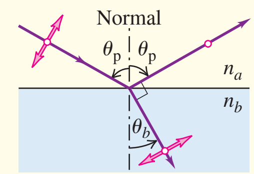
Polarization of light: The direction of polarization of a linearly polarized electromagnetic wave is the direction of the $\vec{E}$ field. A polarizing filter passes waves that are linearly polarized along its polarizing axis and blocks waves polarized perpendicularly to that axis. When polarized light 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)

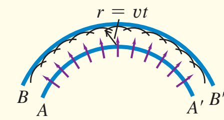


Polarization by reflection: When unpolarized light strikes an interface between two materials, Brewster's law states that the reflected light is completely polarized perpendicular to the plane of incidence (parallel to the interface) if the angle of incidence equals the polarizing angle θ_p. (See Example 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.



Chapter 33 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 33.1 and 33.2 (Section 33.2) before attempting these problems.

VP33.2.1 A block of glass with index of refraction 1.80 has a smooth surface. Light in air strikes this surface at an angle of incidence of 70.0° measured from the normal to the surface of the glass. Find the angles measured relative to this normal of (a) the reflected ray and (b) the refracted ray.

VP33.2.2 A ray of light in water $(n = 1.33)$ strikes a submerged glass block at an angle of incidence of 55.0°. The angle of refraction for the light that enters the glass is 37.0°. Find (a) the index of refraction of the glass and (b) the speed of light in the glass.

VP33.2.3 The light from a red laser pointer has wavelength 635 nm in air and 508 nm in a transparent liquid. You point the laser in air so that the beam strikes the surface of the liquid at an angle of 35.0° from the normal. Find (a) the index of refraction of the liquid, (b) the angle of refraction, (c) the frequency of the light in air, and (d) the frequency of the light in the liquid.

VP33.2.4 A glass of ethanol $(n = 1.36)$ has an ice cube $(n = 1.309)$ floating in it. A light beam in the ethanol goes into the ice cube at an angle of refraction of 85.0°. Find (a) the angle of incidence in the ethanol and (b) the ratio of the wavelength of the light in ice to its wavelength in ethanol.

Be sure to review EXAMPLE 33.5 (Section 33.5) before attempting these problems.

VP33.5.1 A polarized laser beam of intensity 255 W/m^2 shines on an ideal polarizer. The angle between the polarization direction of the laser beam and the polarizing axis of the polarizer is 15.0°. What is the intensity of the light that emerges from the polarizer?

VP33.5.2 You shine unpolarized light with intensity 54.0 W/m^2 on an ideal polarizer, and then the light that emerges from this polarizer falls on a second ideal polarizer. The light that emerges from the second polarizer has intensity 19.0 W/m^2. Find (a) the intensity of the light that emerges from the first polarizer and (b) the angle between the polarizing axes of the two polarizers.

VP33.5.3 A beam of polarized light of intensity 60.0 W/m^2 propagates in the $+x$-direction. The light is polarized in the $+y$-direction. The beam strikes an ideal polarizer whose plane is parallel to the yz-plane and has its polarizing axis at 25.0° clockwise from the y-direction. Then the beam that emerges from this polarizer strikes a second ideal polarizer whose plane is also parallel to the yz-plane but has its polarizing axis at 50.0° clockwise from the y-direction. Find the intensity of the light that emerges (a) from the first polarizer, (b) from the second polarizer, and (b) from the second polarizer if the first polarizer is removed.

VP33.5.4 You simultaneously shine two light beams, each of intensity I_0, on an ideal polarizer. One beam is unpolarized, and the other beam is polarized at an angle of exactly 30° to the polarizing axis of the polarizer. Find the intensity of the light that emerges from the polarizer.

Be sure to review EXAMPLE 33.6 (Section 33.5) before attempting these problems.

VP33.6.1 Unpolarized sunlight in air shines on a block of a transparent solid with index of refraction 1.73. (a) For what angle of incidence is the reflected light completely polarized? (b) For this angle of incidence, is the light refracted into the solid completely polarized, partially polarized, or unpolarized?

VP33.6.2 You shine a beam of unpolarized light in air on a block of glass. You find that if the angle of incidence is 57.0°, the reflected light is completely polarized. Find (a) the index of refraction of the glass and (b) the angle of refraction.

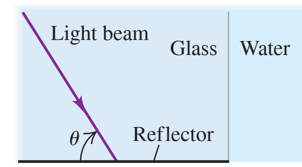
VP33.6.3 You shine a beam of *polarized* light in air on a piece of dense flint glass $(n = 1.66)$. (a) If the polarization direction is perpendicular to the plane of incidence, is there an angle of incidence for which no light is reflected from the glass? If so, what is this angle? (b) Repeat part (a) if the polarization direction is in the plane of incidence.

VP33.6.4 A glass container holds water $(n = 1.33)$. If unpolarized light propagating in the glass strikes the glass–water interface, the light reflected back into the glass will be completely polarized if the angle of refraction is 53.5°. Find (a) the polarizing angle in this situation and (b) the index of refraction of the glass.

BRIDGING PROBLEM Reflection and Refraction

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?

Figure 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'll 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

•, ••, •••: 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.

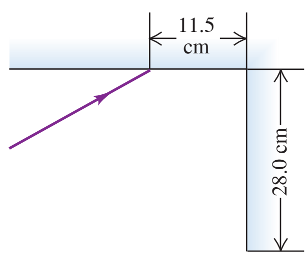
Q33.23 The explanation given in Section 33.6 for the color of the setting sun should apply equally well to the *rising* sun, since sunlight travels the same distance through the atmosphere to reach your eyes at either sunrise or sunset. Typically, however, sunsets are redder than sunrises. Why? (*Hint:* 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

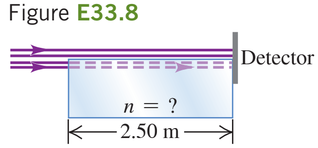
33.1 • Two plane mirrors intersect at right angles. A laser beam strikes the first of them at a point 11.5 cm from their point of intersection, as shown in **Fig. E33.1**. For what angle of incidence at the first mirror will this ray strike the midpoint of the second mirror (which is 28.0 cm long) after reflecting from the first mirror?

Figure E33.1



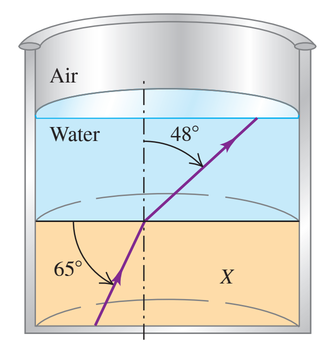
33.2 • **BIO Light Inside the Eye.** The vitreous humor, a transparent, gelatinous fluid that fills most of the eyeball, has an index of refraction of 1.34. Visible light ranges in wavelength from 380 nm (violet) to 750 nm (red), as measured in air. This light travels through the vitreous humor and strikes 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 liquid lies on top of the horizontal surface of a block of glass. A ray of light traveling in the glass has speed 1.85×10^8 m/s, wavelength 365 nm, and frequency 5.07×10^{14} Hz. The ray is incident on the surface of the glass at an angle of 38.0° with respect to the normal to the surface. The ray that refracts into the liquid makes an angle of 44.7° with the normal to the interface between the two materials. What are the speed, wavelength, and frequency of the light when it is traveling in the liquid?
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

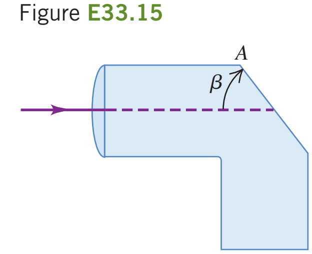


33.12 •• A horizontal, parallel-sided plate of glass having a refractive index of 1.52 is in contact with the surface of water in a tank. A ray coming from above in air makes an angle of incidence of 35.0° with the normal to the top surface of the glass. (a) What angle does the ray refracted into the 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 block of a transparent solid sits on top of the horizontal surface of a block of glass. A ray of light traveling in the glass is incident on the top surface of the glass at an angle of 62.0° with respect to the normal to the surface. The light has wavelength 447 nm in the glass and 315 nm in the transparent solid. What angle does the ray that refracts into the transparent solid make with the normal to the surface?
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 β 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 β 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?

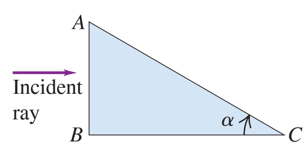
33.18 • A beam of light is traveling inside a solid glass cube that has index of refraction 1.62. It strikes the surface of the cube from the inside. (a) If the cube is in air, at what minimum angle with the normal inside the glass will this light *not* enter the air at this surface? (b) What would be 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**

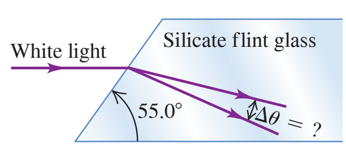


33.22 • A thick sheet of transparent plastic sits on top of the horizontal surface of a block of glass. For a ray traveling in the glass, the critical angle at the glass–plastic interface is 48.6° and the wavelength of the light in the glass is 350 nm. For an incident angle of 40.0° in the glass, what is the wavelength of the light after it refracts into the plastic?

Section 33.4 Dispersion

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 • 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.

33.25 • 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

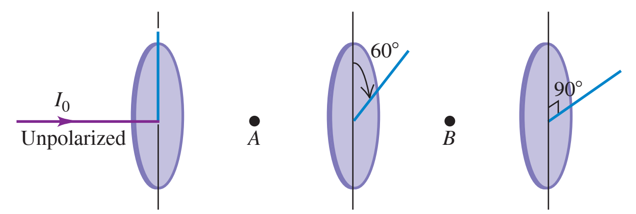
33.26 •• Birefringence is discussed in Section 33.5 and the refractive indexes for the two perpendicular polarization directions in calcite are given. A crystal of calcite serves as a quarter-wave plate; it converts linearly polarized light to circularly polarized light if the numbers of wavelengths within the crystal differ by one-fourth for the two polarization components. For light with wavelength 589 nm in air, what is the minimum thickness of a quarter-wave plate made of calcite?

33.27 • 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?

33.28 • A layer of liquid sits on top of the horizontal surface of a transparent solid. For a ray traveling in the solid and incident on the interface of the two materials, the critical angle is 38.7°. (a) For a ray traveling in the solid and reflecting at the interface with the liquid, for what incident angle with respect to the normal is the reflected ray 100% polarized? (b) What is the polarizing angle if the ray is traveling in the liquid?

33.29 •• 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. E33.29**. (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 **E33.29**



33.30 •• Light of original intensity I_0 passes through two ideal polarizing filters having their polarizing axes oriented as shown in **Fig. E33.30**. 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 **E33.30**

33.31 • 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?

33.32 • 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?

33.33 •• 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 are stacked, with the polarizing axis of the second and third filters at 23.0° and 62.0°, respectively, clockwise 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?

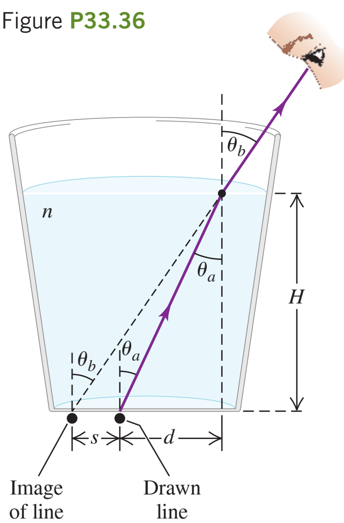
Section 33.6 Scattering of Light

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 •• Draw a straight line on a piece of paper. Fill a transparent drinking glass with water and place it on top of the line you have drawn, making sure the line extends outward from both sides of the glass. Position your head above the glass, but not directly above, such that you look downward at an angle to see the image of the line through the water, as shown in Fig. **P33.36**. The image of the line beneath the glass will appear shifted away from the line you have drawn. (a) Mark a spot on the paper beside the glass where the line seen through the water would extend. Estimate the distance s between that mark and the original line. (b) Keeping your eye in the same position, hold a pen along the line of sight from your eye to the surface of the water, where the image of the line appears. The tip of the pen should just touch the surface of the water. Now move your head above the glass and sight downward from directly above the tip of the pen. Mark a spot on the paper beside the glass indicating the horizontal distance between the drawn line and the tip of the pen. Estimate that distance d. (c) Let θ_a and θ_b represent the angle of incidence and the angle of refraction, respectively, of the upward light ray traveling toward your eye, as shown in the figure, and let H be the depth of the water. Write expressions for $\tan\theta_a$ and $\tan\theta_b$ in terms of s, d, and H. (d) Assume H is large so the angles are small and we can approximate $\tan\theta_a \approx \sin\theta_a$ and $\tan\theta_b \approx \sin\theta_b$. Using Snell's law and these approximations, write an expression for the index of refraction n of the water in terms of s and d. (e) Use your values of s and d to estimate the index of refraction of the water.

Figure P33.36

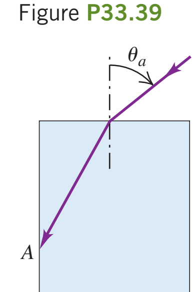


33.37 •• **BIO** **Heart Sonogram.** Physicians use high-frequency (f = 1–5 MHz) sound waves, called ultrasound, to image internal organs. The speed of these ultrasound waves is 1480 m/s in muscle and 344 m/s in air. We define the index of refraction of a material for sound waves to be the ratio of the 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?

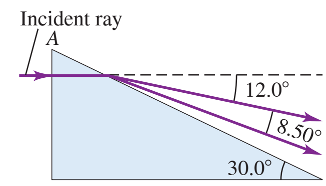
33.39 •• A ray of light is incident in air on a block of a transparent solid whose index of refraction is n. If n = 1.38, what is the *largest* angle of incidence θ_a for which total internal reflection will occur at the vertical face (point A shown in **Fig. P33.39**)?

Figure 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.

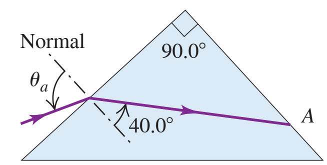
Figure P33.40



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?

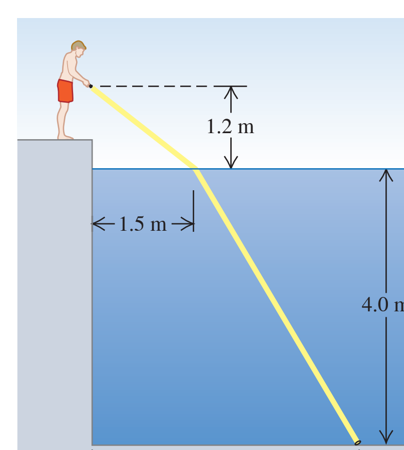
Figure P33.42



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?

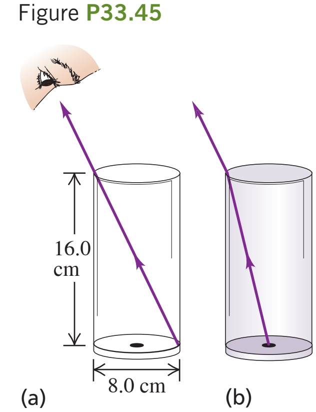
33.44 • After a long day of driving you take a late-night swim in a motel swimming pool. When you go to your room, you realize that you have lost your room key in the pool. You borrow a powerful flashlight and walk around the pool, shining the light into it. The light shines on the key, which is lying 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



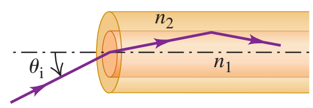
33.45 • You sight along the rim of a glass with vertical sides so that the top rim is lined up with the opposite edge of the bottom (**Fig. P33.45a**). The glass is a thin-walled, hollow cylinder 16.0 cm high. The diameter of the top and bottom of the glass is 8.0 cm. While you keep your eye in the same 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**

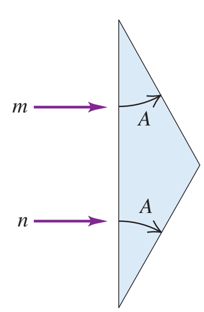


33.47 • A thin layer of ice ($n = 1.309$) floats on the surface of water ($n = 1.333$) in a bucket. A ray of light from the bottom of the bucket travels upward through the water. (a) What is the largest angle with respect to the normal that the ray can make at the ice–water interface and still pass out into the air above the ice? (b) What is this angle after the ice melts?

33.48 •• When linearly polarized light passes through a polarizer, its polarizing axis may be rotated by any angle $\phi < 90°$ at the expense of a loss of intensity, as determined by Malus's law. By using sequential polarizers, you can achieve a similar axis rotation but retain greater intensity. In fact, if you use many intermediate polarizers, the polarization axis can be rotated by 90° with virtually undiminished intensity. (a) Derive an equation for the resulting intensity if linearly polarized light passes through successive N polarizers, each with the polarizing axis rotated by an angle $90°/2N$ larger than the preceding polarizer. (b) By making a table of the resulting intensity for various values of N, estimate the minimum number N of polarizers needed so that the light will have its polarization axis rotated by 90° while maintaining more than 90% of its intensity. (c) Estimate the minimum number of polarizers needed to maintain more than 95% and 99% intensity.

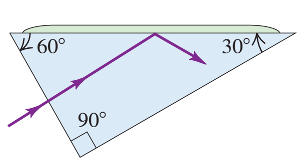
33.49 •• The prism shown in **Fig. P33.49** has a refractive index of 1.66, and the angles A are 25.0°. Two light rays m and n are parallel as they enter the prism. What is the angle between them after they emerge?

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**

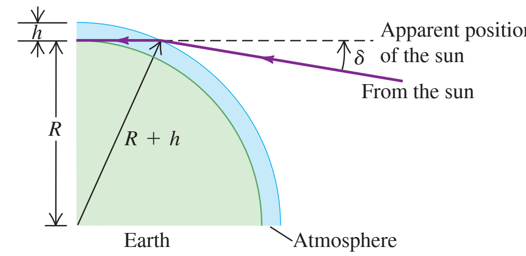


33.51 •• When the sun is either rising or setting and appears to be just on the horizon, it is in fact *below* the horizon. The explanation for this seeming paradox is that light from the sun bends slightly when entering the earth's atmosphere, as shown in **Fig. P33.51**. Since our perception is based 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**

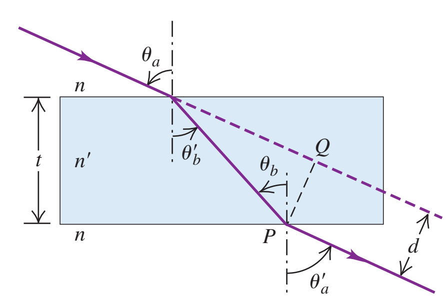


33.52 •• Light is incident in air at an angle θ_a (**Fig. P33.52**) on the upper surface of a transparent plate, the surfaces of the plate being plane and parallel to each other. (a) Prove that $\theta_a = \theta'_a$. (b) Show that this is true for any number of different parallel plates. (c) Prove that the lateral displacement d of the emergent beam is given by the relationship

$$d = t\frac{\sin(\theta_a - \theta'_b)}{\cos\theta'}$$

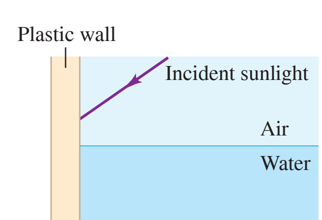
where t is the thickness of the plate. (d) A ray of light is incident at an angle of 66.0° on one surface of a glass plate 2.40 cm thick with an index of refraction of 1.80. The medium on either side of the plate is air. Find the lateral displacement between the incident and emergent rays.

Figure **P33.52**



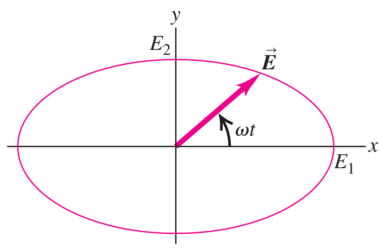
33.53 •• A beam of unpolarized sunlight strikes the vertical plastic wall of a water tank at an unknown angle. Some of the light reflects from the wall and enters the water (**Fig. P33.53**). The refractive index of the plastic wall is 1.61. If the light that has been reflected from the wall into the water is observed to be completely polarized, what angle does this beam make with the normal inside the water?

Figure **P33.53**



33.54 ••• **CP** A circularly polarized electromagnetic wave propagating in air has an electric field given by $\vec{E} = E[\cos(kz - \omega t)\,\hat{\imath} + \sin(kz - \omega t)\,\hat{\jmath}]$. This wave is incident with an intensity of 150 W/m^2 at the polarizing angle θ_p onto a flat interface perpendicular to the xz-plane with a material that has index of refraction $n = 1.62$. (a) What is the angle of refraction θ_b? (b) The ratio between the electric-field amplitude in the reflected wave and the electric-field amplitude of the component of the incident wave polarized parallel to the interface is $\sin(\theta_p - \theta_b)/\sin(\theta_p + \theta_b)$. (This result, known as the Fresnel equation, may be derived using deeper analysis.) From this relationship, determine the intensity $I_{\text{reflected}}$ of the reflected wave. (c) Determine the intensity $I_{\parallel}$ of the component of the refracted wave polarized parallel to the interface. (d) Determine the intensity $I_{\perp}$ of the component of the refracted wave polarized in the x-direction. (e) The reflected wave is linearly polarized and the refracted wave is elliptically polarized, such that its electric field is characterized as shown in **Fig. P33.54**. Determine the elliptical eccentricity $e = \sqrt{1 - (E_1/E_2)^2}$.

Figure **P33.54**



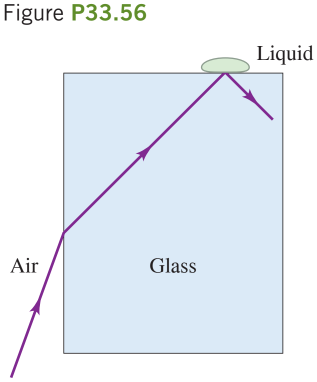
33.55 •• **DATA** In physics lab, you are studying the properties of four transparent liquids. You shine a ray of light (in air) onto the surface of each liquid—A, B, C, and D—one at a time, at a 60.0° angle of incidence; you then measure the angle of refraction. The table gives your data:

Liquid	A	B	C	D
$\boldsymbol{\theta_a}$ **(°)**	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?

33.56 •• **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. P33.56**). Your results are given in the table:

Figure **P33.56**

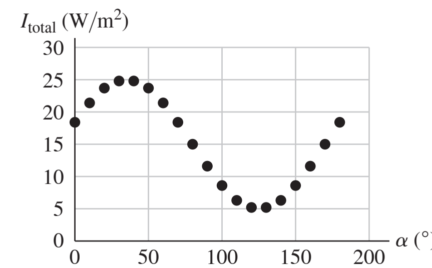


Liquid	A	B	C
$\boldsymbol{\theta_a}$ **(°)**	52.0	44.3	36.3

What is the refractive index of each liquid at this wavelength?

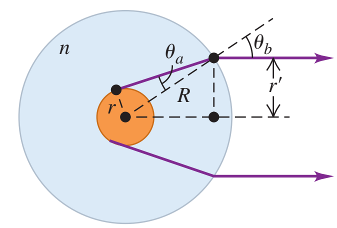
33.57 •• **DATA** A beam of light traveling horizontally is made of an unpolarized component with intensity I_0 and a polarized component with intensity I_p. The plane of polarization of the polarized component is oriented at an angle θ with respect to the vertical. **Figure P33.57** is a graph of the 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 **P33.57**



33.58 •• **CP** An object surrounded by a translucent material with index of refraction greater than unity appears enlarged. This includes items such as plants or shells inside plastic or resin paperweights, as well as ancient bugs trapped in amber. Consider a spherical object in air with radius r fixed at the center of a sphere with radius $R > r$ with refractive index n, as shown in **Fig. P33.58**. When viewed from outside, the spherical object appears to have radius $r' > r$. (a) Using right triangles found in the figure, write an expression for $\sin\theta_a$ in terms of r and R. (b) Similarly, write an expression for $\sin\theta_b$ in terms of r' and R. (c) Using Snell's law, derive an expression for r' in terms of r, R, and n. (d) If the object is a spherical dandelion seed head with diameter 45.0 mm fixed at the center of a solidified resin sphere with radius 80.0 mm and index of refraction 1.53, what diameter will the seed head appear to have when viewed from outside?

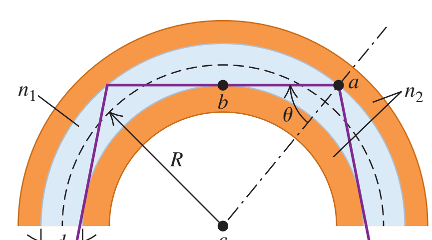
Figure **P33.58**



CHALLENGE PROBLEMS

33.59 ••• A fiber-optic cable consists of a thin cylindrical core with thickness d made of a material with index of refraction n_1, surrounded by cladding made of a material with index of refraction $n_2 < n_1$. Light rays traveling within the core remain trapped in the core provided they do not strike the core–cladding interface at an angle larger than the critical angle for total internal reflection. (a) **Figure P33.59** shows a light ray traveling within a cable bent to its critical extent, beyond which light will leak out of the core. The radius of curvature of the center of the core is R. Derive an expression for the angle of incidence θ at which the indicated light ray strikes the outer edge of the core. (*Hint:* Construct a right triangle with vertices a, b, and c as in the figure.) (b) Determine the

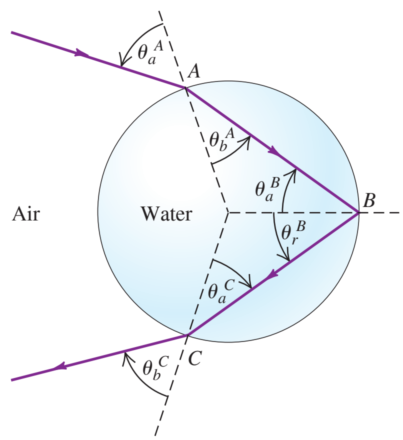
Figure **P33.59**



critical radius R in terms of n_1, n_2, and d. (c) A typical fiber-optic cable has a 50.0-μm-thick core with index of refraction 1.4475 and has a cladding with index of refraction 1.4440. What is the minimum radius of curvature for this cable? (d) If a light ray traversed a 1.00 km length of such a cable bent to its critical extent, and if the ray were maximally angled within the core as shown in the figure, what would be the extra distance, beyond 1.00 km, traveled by the ray owing to its multiple reflections? Keep extra digits in intermediate steps to reduce rounding errors. (e) How much longer would it take light to travel 1.00 km through the cable than 1.00 km in air?

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 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?

Figure P33.60



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.

MCAT-STYLE 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).

Key Example VARIATION Problems

VP33.2.1 (a) 70.0° (b) 31.5°
VP33.2.2 (a) 1.81 (b) 1.66×10^8 m/s
VP33.2.3 (a) 1.25 (b) 27.3° (c) 4.72×10^{14} Hz (d) 4.72×10^{14} Hz
VP33.2.4 (a) 73.5° (b) 1.04
VP33.5.1 238 W/m^2
VP33.5.2 (a) 27.0 W/m^2 (b) 33.0°
VP33.5.3 (a) 49.3 W/m^2 (b) 40.5 W/m^2 (c) 24.8 W/m^2
VP33.5.4 $5I_0/4$
VP33.6.1 (a) 60.0° (b) partially polarized
VP33.6.2 (a) 1.54 (b) 33.0°
VP33.6.3 (a) no (b) yes, 58.9°
VP33.6.4 (a) 36.5° (b) 1.80

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 OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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; 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'll 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'll see cases in which the outgoing rays really *do* pass through an image point, and we'll 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** (next page) 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.

Figure 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.

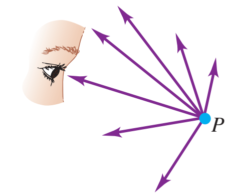


Figure 34.2 Light rays from the object at point P are reflected from a plane mirror. The reflected rays entering the eye look as though they had come from image point P'.

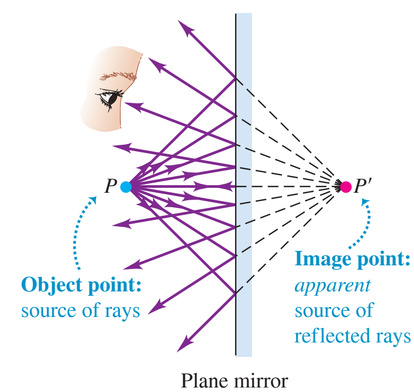


Figure 34.3 Light rays from the object at point P are refracted at the plane interface. The refracted rays entering the eye look as though they had come from image point P'.

When $n_a > n_b$, P' is closer to the surface than P; for $n_a < n_b$, the reverse is true.

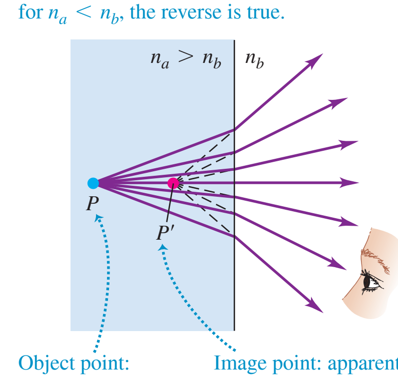
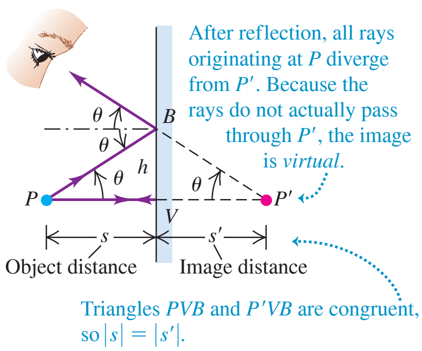


Figure 34.4 Construction for determining the location of the image formed by a plane mirror. The image point P' is as far behind the mirror as the object point P is in front of it.



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'.

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'll 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.

Figure 34.5 For both of these situations, the object distance s is positive (rule 1) and the image distance s' is negative (rule 2).

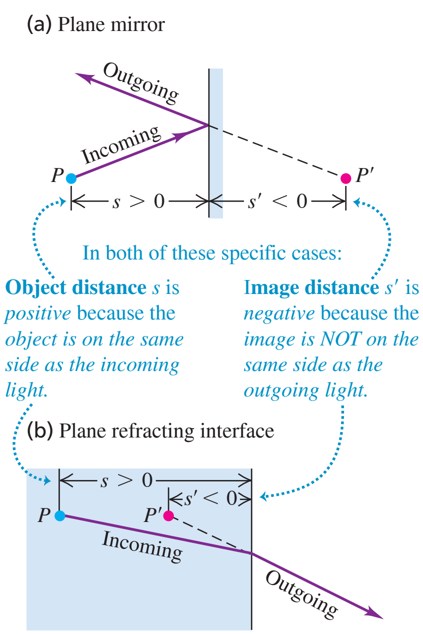


Figure 34.5 illustrates rules 1 and 2 for two different situations. For a mirror the incoming and outgoing sides are always the same; for example, in Figs. 34.2, 34.4, and 34.5a they are both on the left side. For the refracting surfaces in Figs. 34.3 and 34.5b the incoming and outgoing sides are on 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.

Figure 34.6 Construction for determining the height of an image formed by reflection at a plane reflecting surface.

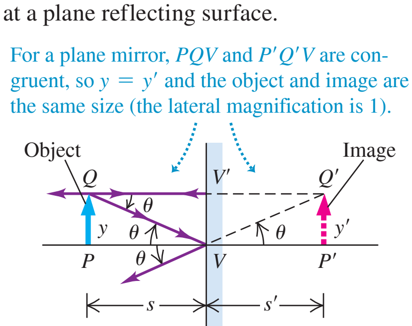


Image of an Extended Object: Plane Mirror

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

$$m = \frac{y'}{y} \qquad (34.2)$$

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'll 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.

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**).

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.

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 $2v$.

ANSWER

(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 $2v$.

Figure 34.7 The image formed by a plane mirror is virtual, erect, and reversed. It is the same size as the object.

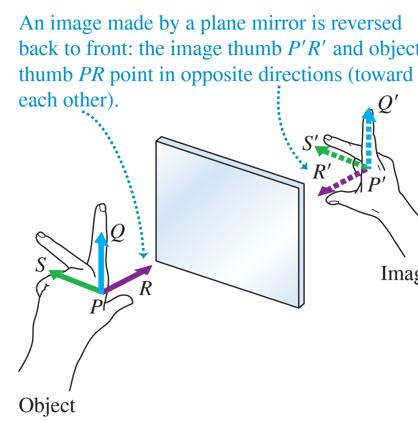


Figure 34.8 The image formed by a plane mirror is reversed; the image of a right hand is a left hand, and so on. (The hand is resting on a horizontal mirror.) Are images of the letters I, H, and T reversed?

Figure 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.

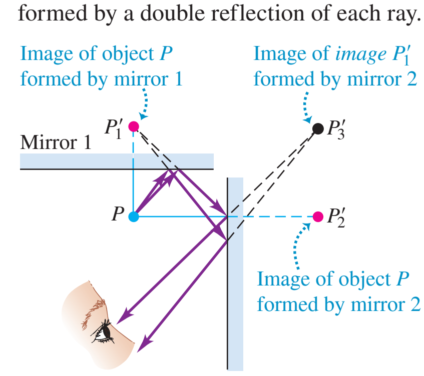
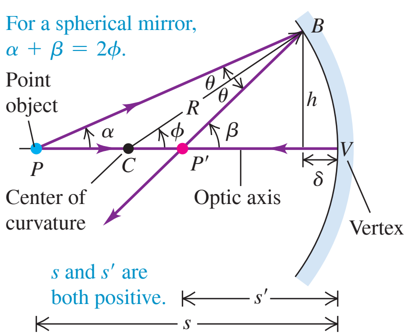


Figure 34.10 (a) A concave spherical mirror forms a real image of a point object P on the mirror's optic axis. (b) The eye sees some of the outgoing rays and perceives them as having come from P'.

(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

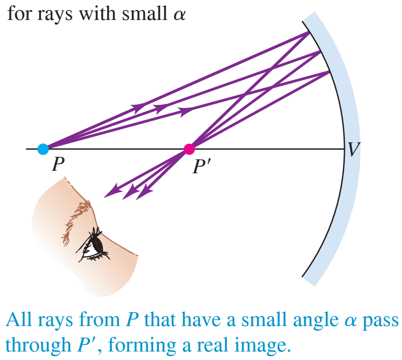
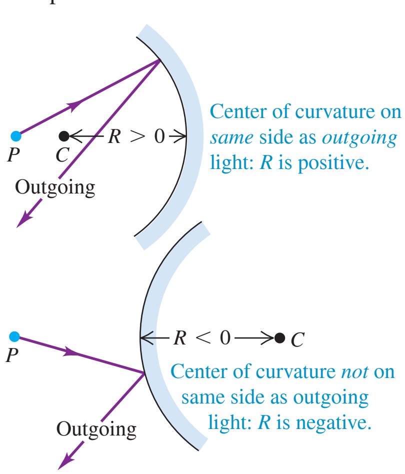


Figure 34.11 The sign rule for the radius of a spherical mirror.



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'll 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

$$\alpha + \beta = 2\phi \tag{34.3}$$

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'll 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

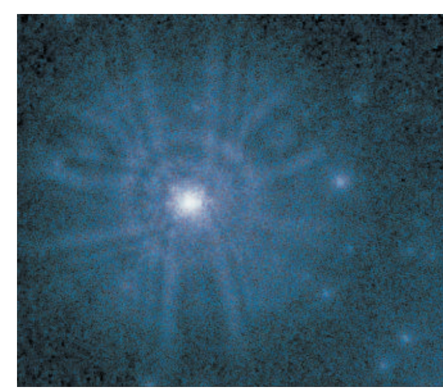
When the object point P is very far from the spherical mirror ($s = \infty$), the incoming rays are parallel. (The star shown in Fig. 34.12c is an example of such a distant object.) From Eq. (34.4) the image distance s' in this case is given by

$$\frac{1}{\infty} + \frac{1}{s'} = \frac{2}{R} \qquad s' = \frac{R}{2}$$

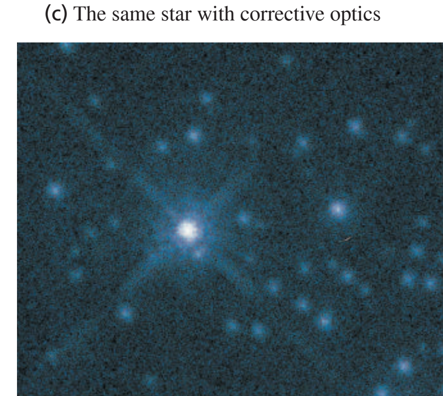
Figure **34.12** (a), (b) Soon after the Hubble Space Telescope (HST) was placed in orbit in 1990, it was discovered that the concave primary mirror (also called the *objective mirror*) was too shallow by about $\frac{1}{50}$ the width of a human hair, leading to spherical aberration of the star's image. (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



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

$$f = \frac{R}{2} \qquad \text{(focal length of a spherical mirror)} \qquad (34.5)$$

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.

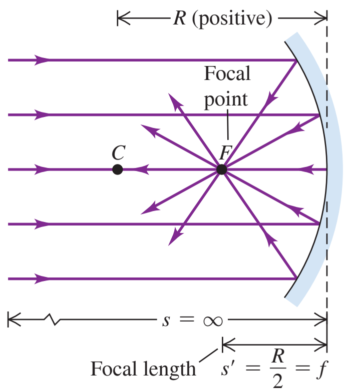
Thus the focal point F of a spherical mirror has the properties that (1) any incoming ray parallel to the optic axis is reflected through the focal point and (2) any incoming ray that passes through the focal point is reflected parallel to the optic axis. For spherical mirrors these statements are true 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'll usually express the relationship between object and image distances for a mirror, Eq. (34.4), in terms of the focal length f:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad \text{(object–image relationship, spherical mirror)} \qquad (34.6)$$

Figure 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.

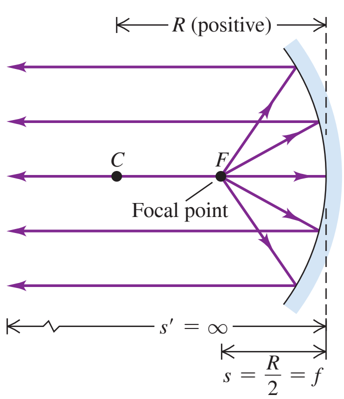
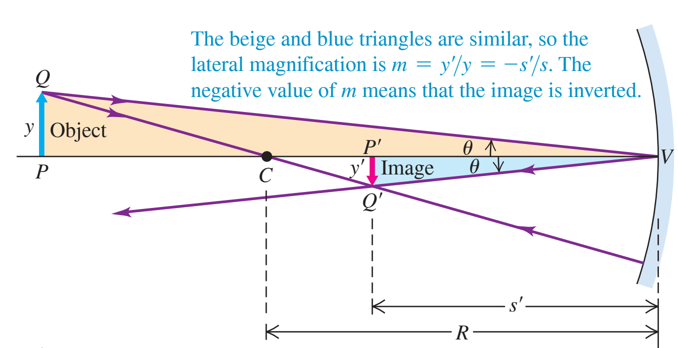


Image of an Extended Object: Spherical Mirror

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. Note that the object and image arrows have different sizes, y and y', respectively, and that they have opposite orientation.

Figure 34.14 Construction for determining the position, orientation, and height of an image formed by a concave spherical mirror.



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

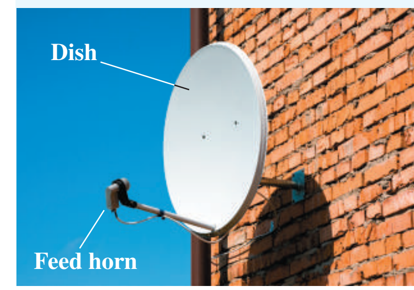
$$m = \frac{y'}{y} = -\frac{s'}{s} \qquad \text{(lateral magnification, spherical mirror)} \qquad (34.7)$$

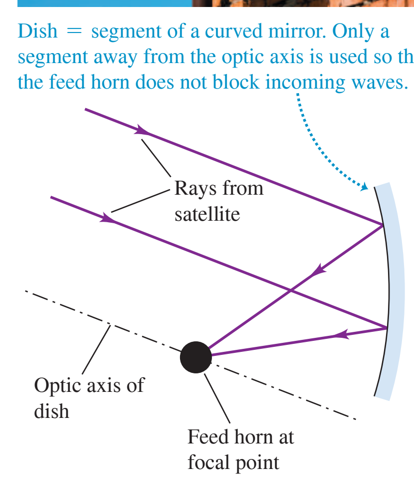
If m is positive, the image is erect in comparison to the object; if m is negative, the image is *inverted* relative to the object, as in Fig. 34.14. For a *plane* mirror, $s = -s'$, so $y' = y$ and $m = +1$; since m is positive, the image is erect, and since $|m| = 1$, the image is the same size as 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.

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.





EXAMPLE 34.1 Image formation by a concave mirror I

WITH VARIATION PROBLEMS

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?

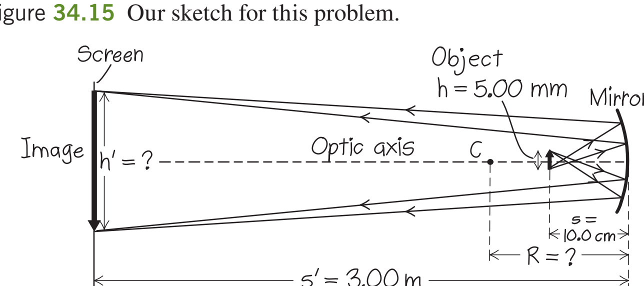
IDENTIFY and SET UP **Figure 34.15** shows our sketch. Our target variables are the radius of curvature R, focal length f, lateral magnification m, and image height y'. We are given the distances from the mirror to the object (s) and from the mirror to the image (s'). We solve Eq. (34.4) for R, and then use Eq. (34.5) to find f. Equation (34.7) yields both m and y'.

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:

$$\frac{1}{10.0\text{ cm}} + \frac{1}{300\text{ cm}} = \frac{2}{R}$$

$$R = 2\left(\frac{1}{10.0\text{ cm}} + \frac{1}{300\text{ cm}}\right)^{-1} = 19.4\text{ cm}$$

Figure **34.15** Our sketch for this problem.



The focal length of the mirror is $f = R/2 = 9.68$ cm.

(b) From Eq. (34.7) the lateral magnification is

$$m = -\frac{s'}{s} = -\frac{300\text{ cm}}{10.0\text{ cm}} = -30.0$$

Continued

Because m is negative, the image is inverted. The height of the image is 30.0 times the height of the object, or $(30.0)(5.00\text{ mm}) = 150\text{ mm}$.

EVALUATE Our sketch indicates that the image is inverted; our calculations agree. Note that the object (at $s = 10.0\text{ cm}$) is just outside the focal point ($f = 9.68\text{ cm}$). 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.

KEYCONCEPT For a concave spherical mirror, the focal length—the distance from the mirror to the point where incoming parallel light rays come to a focus—equals one-half the radius of the mirror. The distance from the mirror to an object, the distance from the mirror to the resulting image, and the focal length are related by Eq. (34.6). The magnification of the image (which is negative if the image is inverted) equals the ratio of the image height to the object height.

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).

KEYCONCEPT A concave mirror of any size makes a complete image of an object. The larger the mirror, the more light is incorporated into the image and the brighter the image is.

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

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} \quad \text{and} \quad m = \frac{y'}{y} = -\frac{s'}{s}$$

Figure 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

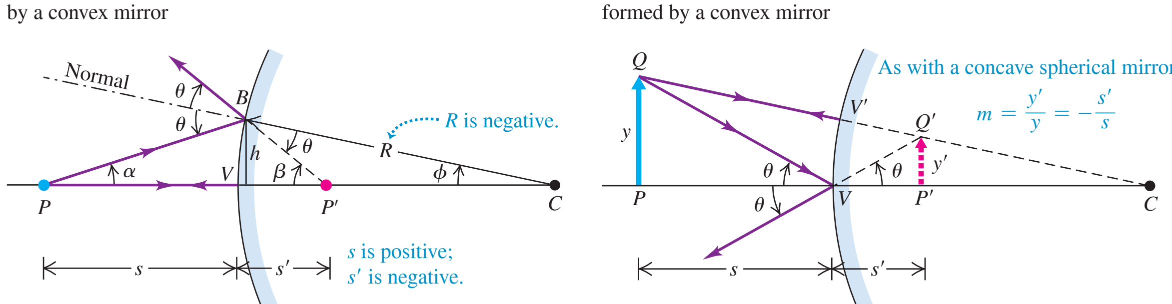
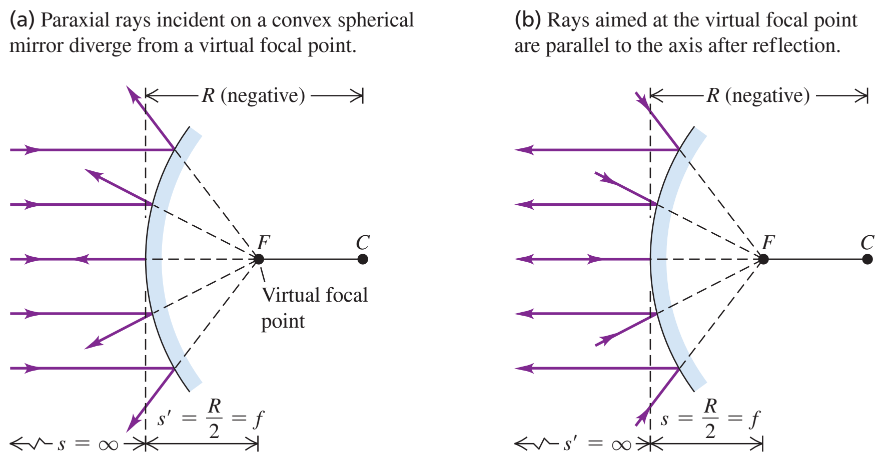


Figure 34.17 The focal point and focal length of a convex mirror.



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.

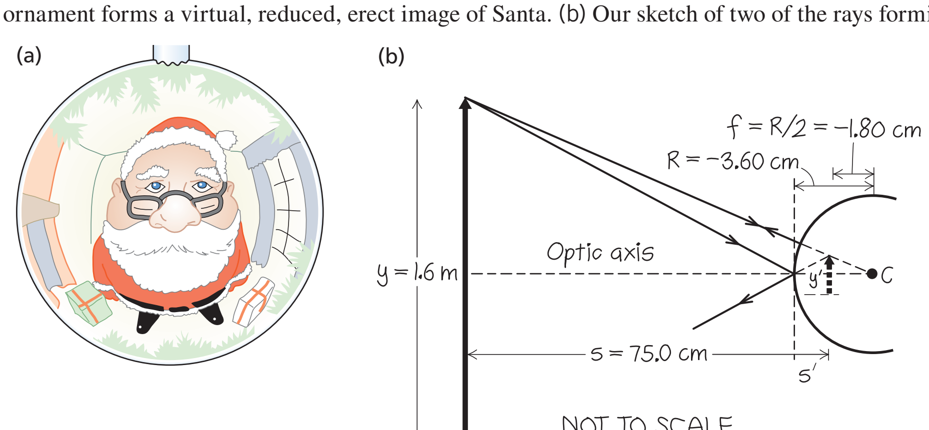
EXAMPLE 34.3 Santa's image problem

WITH VARIATION PROBLEMS

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?

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 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.

Figure 34.18 (a) The ornament forms a virtual, reduced, erect image of Santa. (b) Our sketch of two of the rays forming the image.



Continued

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.")

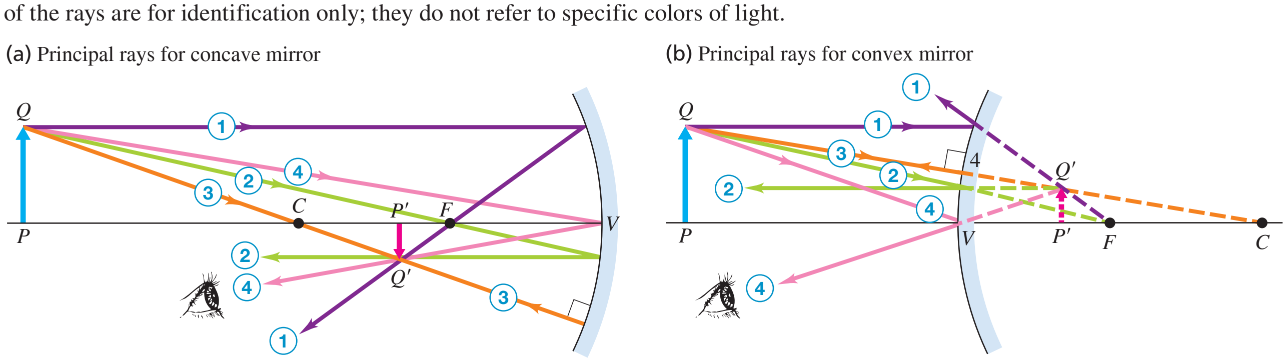
KEYCONCEPT A convex spherical mirror has a negative focal length: Incoming parallel light rays reflect from the mirror as though they were coming from a focal point behind the mirror. The relationships among object distance, image distance, and focal length, and among magnification, object distance, and image distance are the same as for a concave mirror.

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.

Figure 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)
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.

(b)
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

WITH VARIATION PROBLEMS

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.

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** (next page) 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:

$$\text{(a)}\quad \frac{1}{30\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = 15\text{ cm}$$

$$\text{(b)}\quad \frac{1}{20\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = 20\text{ cm}$$

$$\text{(c)}\quad \frac{1}{10\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = \infty \quad (\text{or } -\infty)$$

$$\text{(d)}\quad \frac{1}{5\text{ cm}} + \frac{1}{s'} = \frac{1}{10\text{ cm}} \qquad s' = -10\text{ cm}$$

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),

$$\text{(a)}\ m = -\frac{15\text{ cm}}{30\text{ cm}} = -\frac{1}{2}$$

$$\text{(b)}\ m = -\frac{20\text{ cm}}{20\text{ cm}} = -1$$

$$\text{(c)}\ m = -\frac{\infty\text{ cm}}{10\text{ cm}} = -\infty \ (\text{or} +\infty)$$

Continued

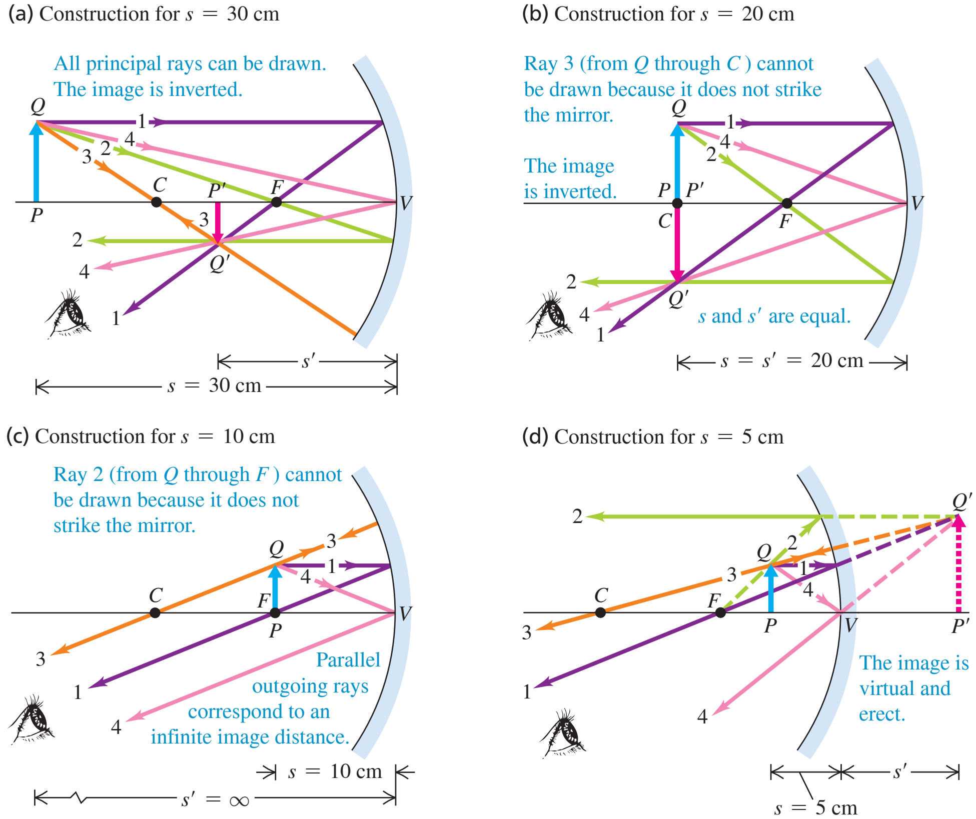
(d) $m = -\dfrac{-10\text{ cm}}{5\text{ cm}} = +2$

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.

KEYCONCEPT If an object is *outside* the focal point of a concave mirror, the reflected rays converge to a real, inverted image. If an object is *at* the focal point, the reflected rays are parallel and form an image at infinity. If an object is *inside* the focal point, the reflected rays appear to be coming from behind the mirror and so form a virtual, erect image.

Figure **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 $|f|$.

ANSWER

(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 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.

Figure **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, $n_a < n_b$.

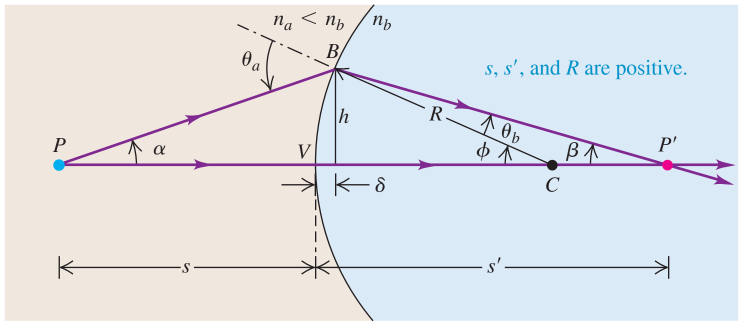


Image of a Point Object: Spherical Refracting Surface

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'll 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

$$\theta_b = \frac{n_a}{n_b}(\alpha + \phi)$$

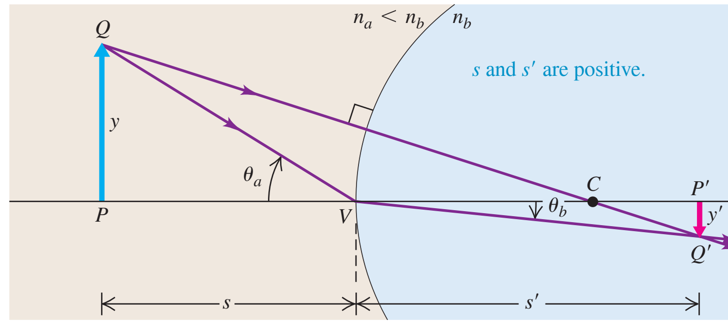
When we substitute this into the second of Eqs. (34.8), we get

$$n_a\alpha + n_b\beta = (n_b - n_a)\phi \tag{34.10}$$

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

$$\alpha = \frac{h}{s} \qquad \beta = \frac{h}{s'} \qquad \phi = \frac{h}{R}$$

Figure 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

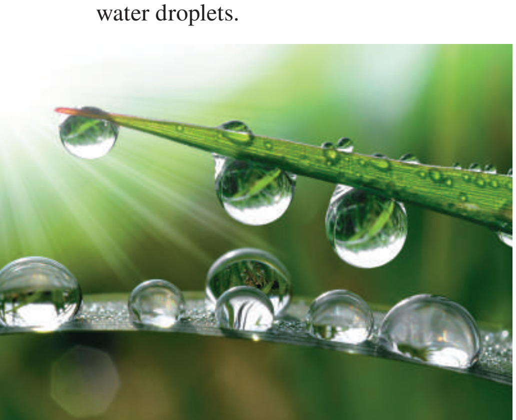
$$\frac{n_a y}{s} = -\frac{n_b y'}{s'} \qquad \text{or}$$

$$m = \frac{y'}{y} = -\frac{n_a s'}{n_b s} \qquad \text{(lateral magnification, spherical refracting surface)} \tag{34.12}$$

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.

Figure 34.23 Light rays refract as they pass through the curved surfaces of these water droplets.



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.

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

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.

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),

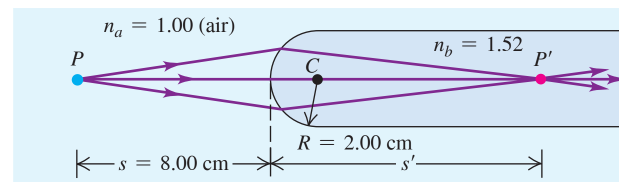
$$\frac{1.00}{8.00\text{ cm}} + \frac{1.52}{s'} = \frac{1.52 - 1.00}{+2.00\text{ cm}}$$

$$s' = +11.3\text{ cm}$$

(b) From Eq. (34.12),

$$m = -\frac{n_a s'}{n_b s} = -\frac{(1.00)(11.3\text{ cm})}{(1.52)(8.00\text{ cm})} = -0.929$$

Figure **34.24** The glass rod in air forms a real image.



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.

KEYCONCEPT The position of the image formed by a spherical refracting surface depends on the position of the object, the radius of curvature of the surface, and the indexes of refraction of the materials on either side of the surface.

EXAMPLE 34.6 Image formation by refraction II

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.

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

$$\frac{1.33}{8.00\text{ cm}} + \frac{1.52}{s'} = \frac{1.52 - 1.33}{+2.00\text{ cm}}$$

$$s' = -21.3\text{ cm}$$

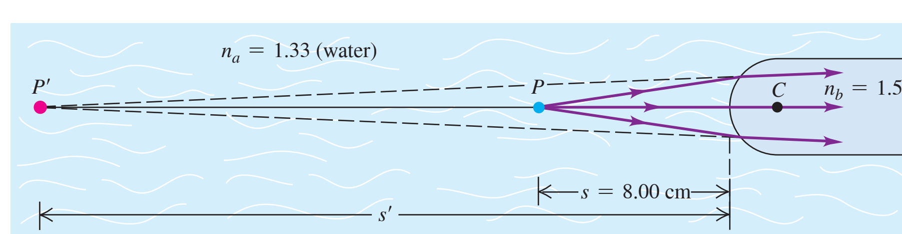
The lateral magnification in this case is

$$m = \frac{n_a s'}{n_b s} = -\frac{(1.33)(-21.3\text{ cm})}{(1.52)(8.00\text{ cm})} = +2.33$$

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.

KEYCONCEPT A spherical refracting surface forms a real image if the refracted rays converge but forms a virtual image if the refracted rays diverge.

Figure **34.25** When immersed in water, the glass rod forms a virtual image.



EXAMPLE 34.7 Apparent depth of a swimming pool

If you look straight down into a swimming pool where it is 2.00 m deep, how deep does it appear to be?

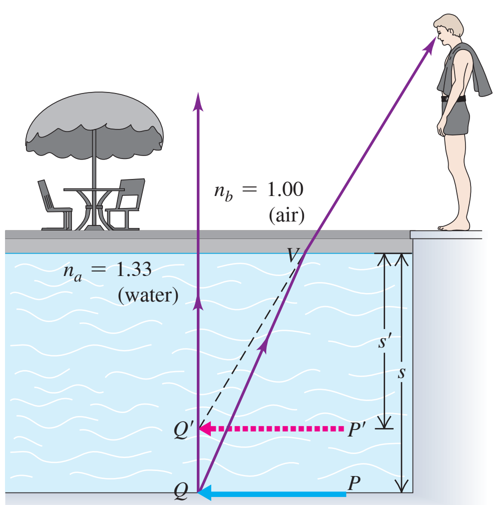
IDENTIFY and SET UP **Figure 34.26** shows the situation. The surface of the water acts as a plane refracting surface. To determine the pool's apparent depth, we imagine an arrow PQ painted on the bottom. The pool's refracting surface forms a virtual image $P'Q'$ of this arrow. We solve Eq. (34.13) to 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),

$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{1.33}{2.00\text{ m}} + \frac{1.00}{s'} = 0$$

$$s' = -1.50\text{ m}$$

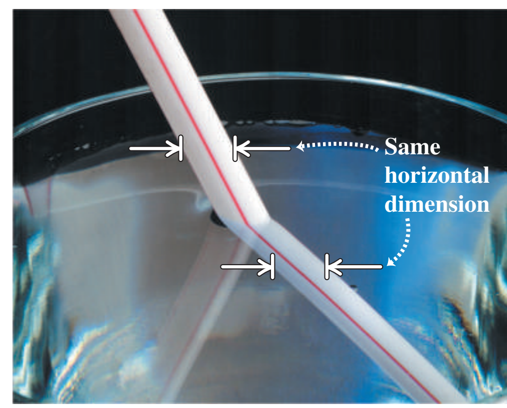
Figure 34.26 Arrow $P'Q'$ is the virtual image of underwater arrow PQ. The angles of the ray with the vertical are exaggerated for clarity.



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.

Figure 34.27 The submerged portion of this straw appears to be at a shallower depth (closer to the surface) than it actually is.



KEYCONCEPT A plane refracting surface always forms a virtual image. An object on the other side of the refracting surface appears to be at a different distance from the surface than its actual distance.

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?

ANSWER

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 THIN LENSES

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

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'll 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.

Figure 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.

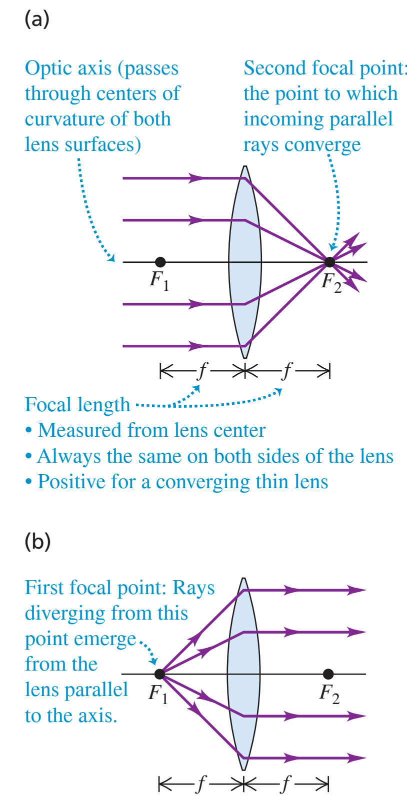


Image of an Extended Object: Converging Lens

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

$$\frac{y}{s} = -\frac{y'}{s'} \quad \text{or} \quad \frac{y'}{y} = -\frac{s'}{s} \tag{34.14}$$

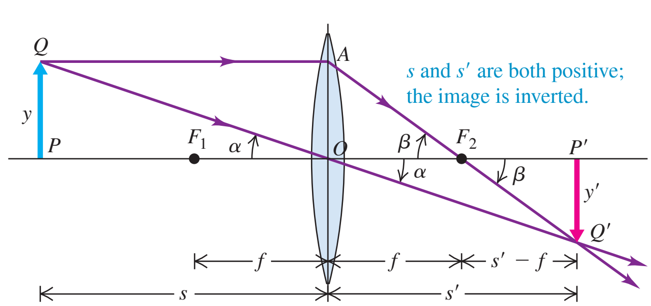


Figure 34.29 Construction used to find image position for a thin lens. To emphasize that the lens is assumed to be very thin, the ray QAQ' is shown as bent at the midplane of the lens rather than at the two surfaces and ray QOQ' is shown as a straight line.

(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

$$\frac{y}{f} = -\frac{y'}{s' - f} \qquad \text{or}$$

$$\frac{y'}{y} = -\frac{s' - f}{f} \tag{34.15}$$

We now equate Eqs. (34.14) and (34.15), divide by s', and rearrange to obtain

Object–image relationship, thin lens:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \tag{34.16}$$

Object distance: s; Image distance: s'; Focal length of lens: f

Equation (34.14) also gives the lateral magnification for the lens:

$$m = y'/y = -\frac{s'}{s} \qquad \text{(lateral magnification, thin lens)} \tag{34.17}$$

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'll 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, your left forefinger along PQ,

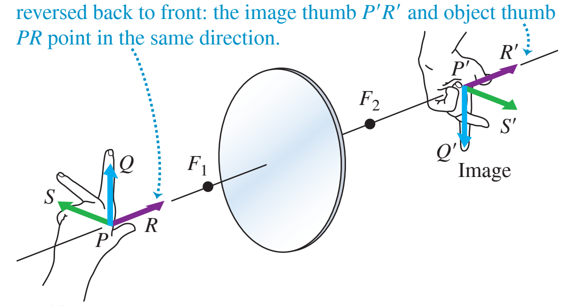


Figure **34.30** The image $S'P'Q'R'$ of a three-dimensional object $SPQR$ is not reversed by a lens.

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** (next page). 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

$$\frac{1}{s_1} + \frac{n}{s'_1} = \frac{n - 1}{R_1}$$
$$-\frac{n}{s'_1} + \frac{1}{s'_2} = \frac{1 - n}{R_2}$$

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:

$$\frac{1}{s_1} + \frac{1}{s'_2} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

Figure **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.

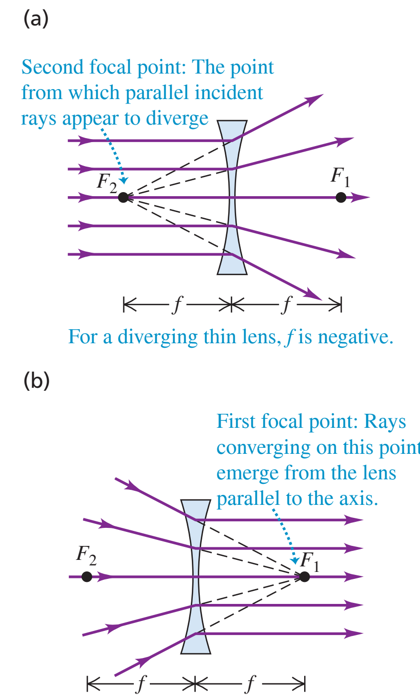


Figure **34.32** Various types of lenses.

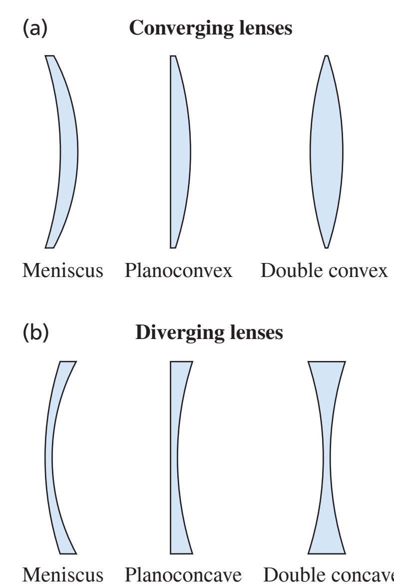


Figure **34.33** The image formed by the first surface of a lens serves as the object for the second surface. The distances s_1' and s_2 are taken to be equal; this is a good approximation if the lens thickness t is small.

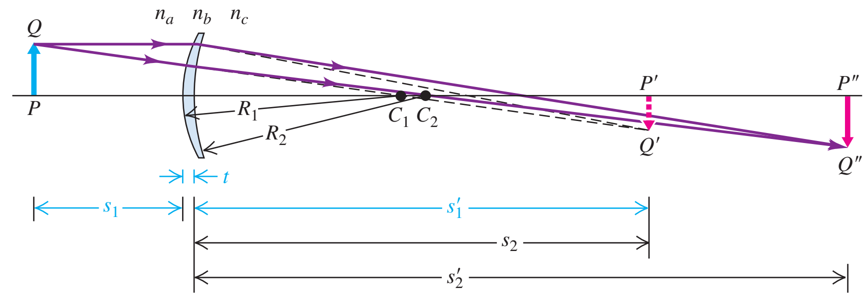
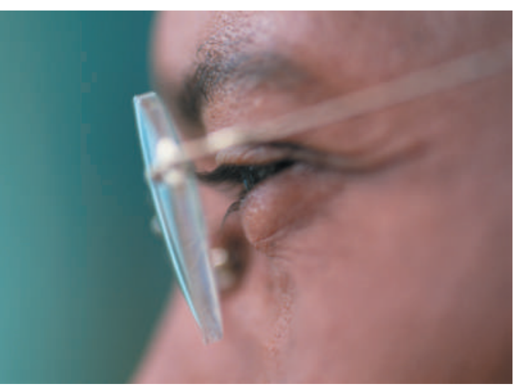


Figure **34.34** These eyeglass lenses satisfy the thin-lens approximation: Their thickness is small compared to the object and image distances.



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':

$$\frac{1}{s} + \frac{1}{s'} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad (34.18)$$

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:

Index of refraction of lens material

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad (34.19)$$

Focal length | Radius of curvature of first surface | Radius of curvature of second surface

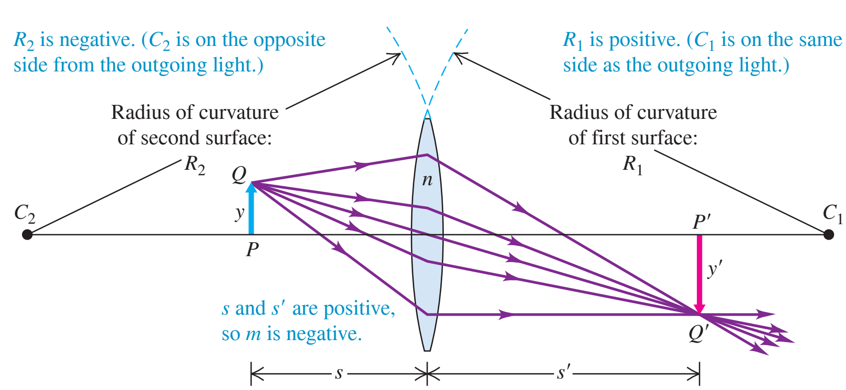
In the process of rederiving the relationship among object distance, image distance, and focal length for a thin lens, we have also derived Eq. (34.19), an expression for the focal length f of a lens in terms of its index of refraction n and the radii of curvature R_1 and R_2 of its surfaces. This can 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.

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.

Figure **34.35** A converging thin lens with a positive focal length f.



EXAMPLE 34.8 Determining the focal length of a lens

WITH VARIATION PROBLEMS

(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?

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),

$$\frac{1}{f} = (1.52 - 1)\left(\frac{1}{+10\text{ cm}} - \frac{1}{-10\text{ cm}}\right) \quad \text{or} \quad f = 9.6\text{ cm}$$

(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

$$\frac{1}{f} = (1.52 - 1)\left(\frac{1}{-10\text{ cm}} - \frac{1}{+10\text{ cm}}\right) \quad \text{or} \quad f = -9.6\text{ cm}$$

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.

KEYCONCEPT The lensmaker's equation, Eq. (34.19), describes how the focal length of a thin lens surrounded by air depends on the radii of curvature of its surfaces and the index of refraction of the material of which it is made. A radius is positive if the center of curvature is on the side of the outgoing light, and negative if the center is on the side of the incoming light.

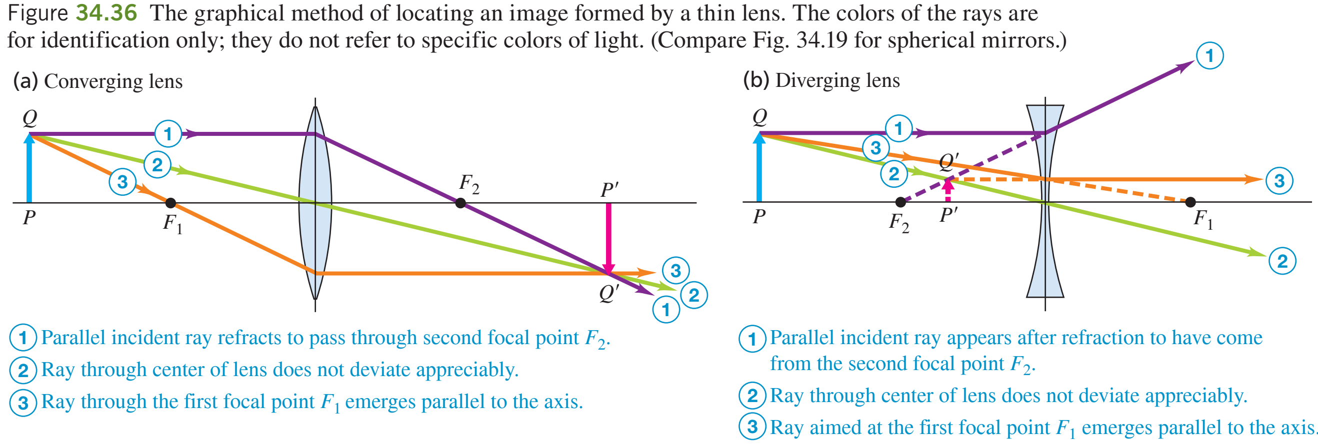
Graphical Methods for Lenses

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'll 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.

Figure 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.)



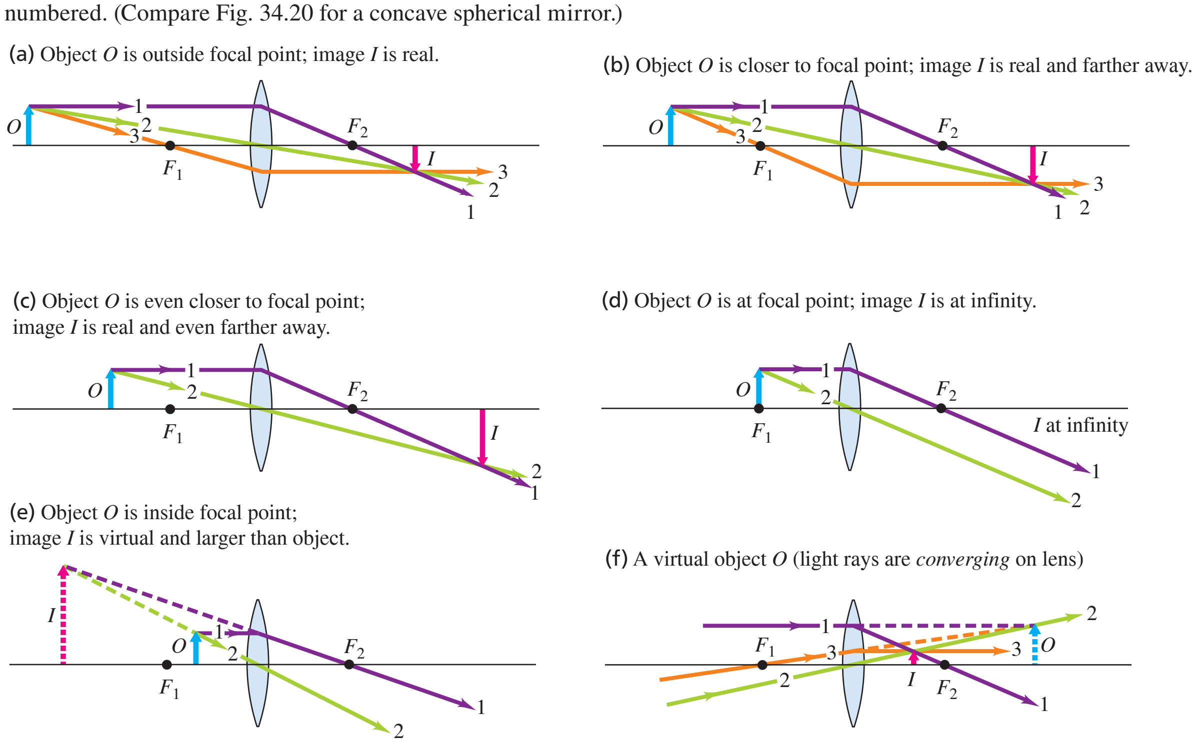
When the image is real, the position of the image point is determined by the intersection of any two rays 1, 2, and 3 (Fig. 34.36a). When the image is virtual, we extend the diverging outgoing rays backward to their intersection point to find the image point (Fig. 34.36b).

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.

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). 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.

Figure 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.)



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

WITH VARIATION PROBLEMS

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.

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),

$$\text{(a)}\ m = -\frac{33.3\ \text{cm}}{50\ \text{cm}} = -\tfrac{2}{3} \qquad \text{(b)}\ m = -\frac{\pm\infty\ \text{cm}}{20\ \text{cm}} = \pm\infty$$

$$\text{(c)}\ m = -\frac{-60\ \text{cm}}{15\ \text{cm}} = +4 \qquad \text{(d)}\ m = -\frac{13.3\ \text{cm}}{-40\ \text{cm}} = +\tfrac{1}{3}$$

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.

KEYCONCEPT If an object is *outside* the focal point of a converging lens, the outgoing rays converge to a real, inverted image. If an object is *at* the focal point, the outgoing rays are parallel and form an image at infinity. If an object is *inside* the focal point, the outgoing rays appear to be coming from in front of the lens and so form a virtual, erect image.

EXAMPLE 34.10 Image formation by a diverging lens

WITH VARIATION PROBLEMS

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.

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.

Continued

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

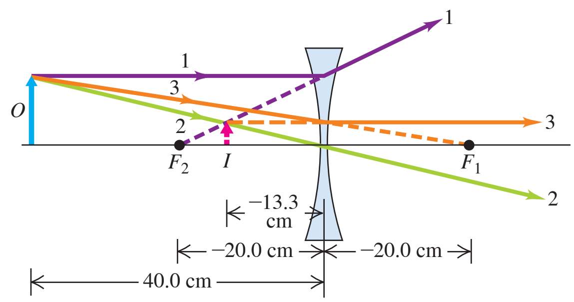
$$s' = -\frac{s}{3} = -\frac{40.0\text{ cm}}{3} = -13.3\text{ cm}$$

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).

Figure **34.38** Principal-ray diagram for an image formed by a thin diverging lens.



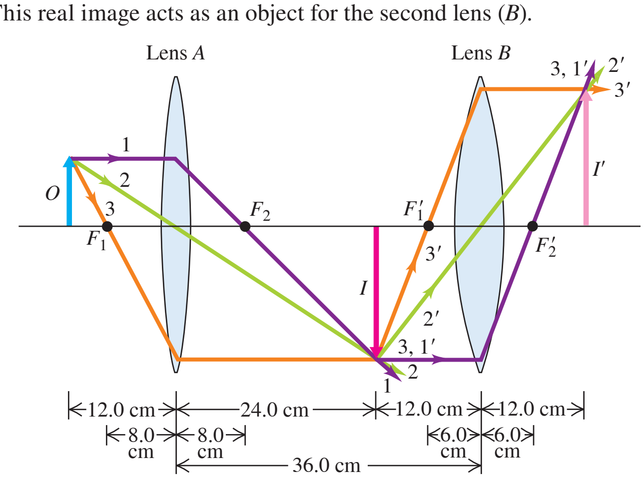
KEYCONCEPT No matter where an object is placed in front of a diverging lens, the outgoing rays appear to be coming from in front of the lens and so form a virtual, erect image. This image is always smaller than the object.

EXAMPLE 34.11 An image of an image

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, which we'll discuss in Section 34.7.)

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.

Figure **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).



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 $36.0\text{ cm} - 24.0\text{ cm} = 12.0\text{ cm}$ 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

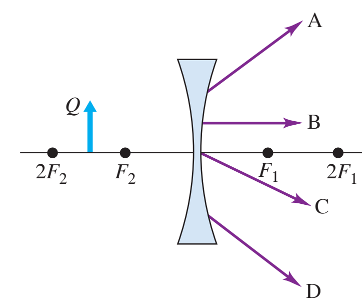
$$\frac{1}{12.0\text{ cm}} + \frac{1}{s'_{I',B}} = \frac{1}{6.0\text{ cm}} \qquad s'_{I',B} = +12.0\text{ cm}$$

The final image I' is 12.0 cm to the right of lens B. The magnification produced by lens B is $m_B = -(12.0\text{ cm})/(12.0\text{ cm}) = -1.00$.

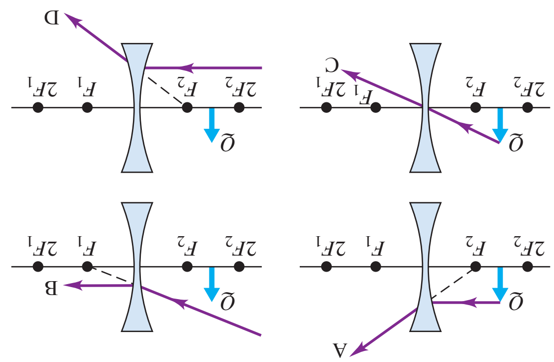
EVALUATE The value of m_B means that the final image I' is just as large as the first image I but has the opposite orientation. The *overall* magnification is $m_A m_B = (-2.00)(-1.00) = +2.00$. Hence the final image I' is $(2.00)(8.0\text{ cm}) = 16$ cm tall and has the same orientation as the original object O, just as Fig. 34.39 shows.

KEYCONCEPT To analyze an optical system that has more than one lens, first find the position of the image made by the first lens. This first image then acts as an object for the second lens, which in turn makes an image of this object.

TEST YOUR UNDERSTANDING OF SECTION 34.4 A diverging lens and an object are positioned as shown in the figure at right. Which of the rays A, B, C, and D could emanate from point Q at the top of the object?



ANSWER



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 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

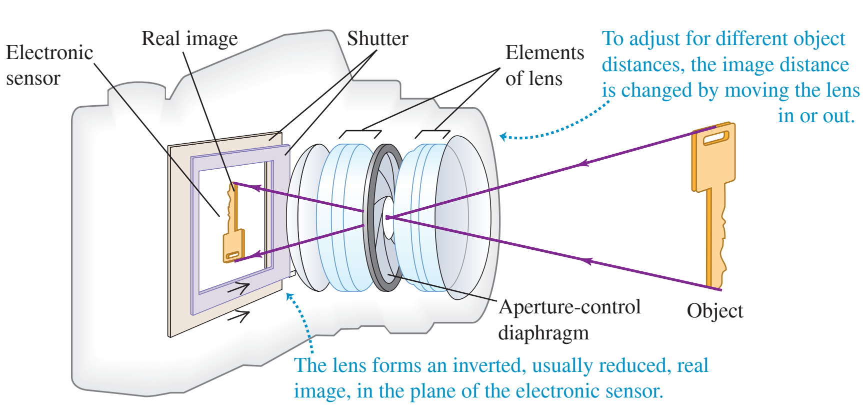
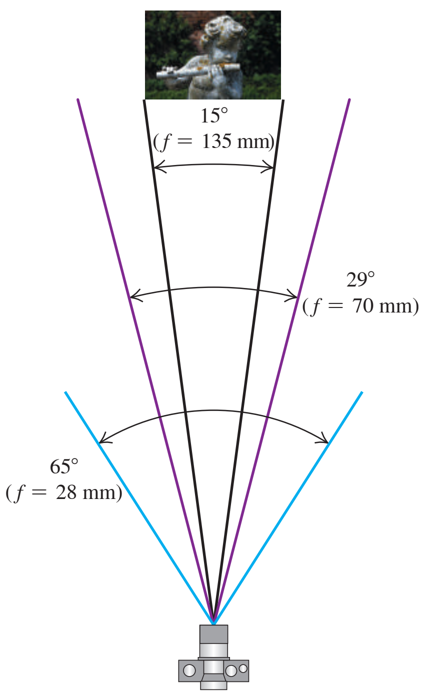


Figure **34.40** Key elements of a digital camera.

Figure 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 × 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)



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.

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.

Photographers commonly express the light-gathering capability of a lens in terms of the ratio f/D, called the ***f*-number** of the lens:

$$f\text{-number of a lens} = \frac{f \leftarrow \text{Focal length of lens}}{D \leftarrow \text{Aperture diameter}} \qquad (34.20)$$

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.

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.

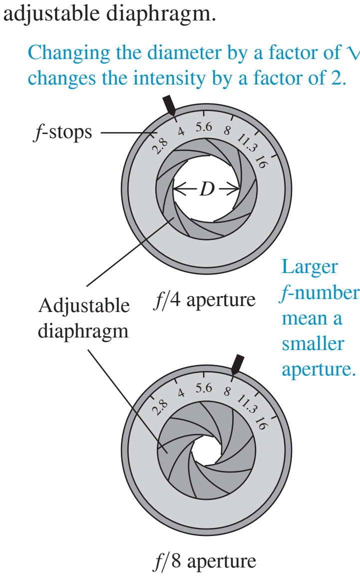


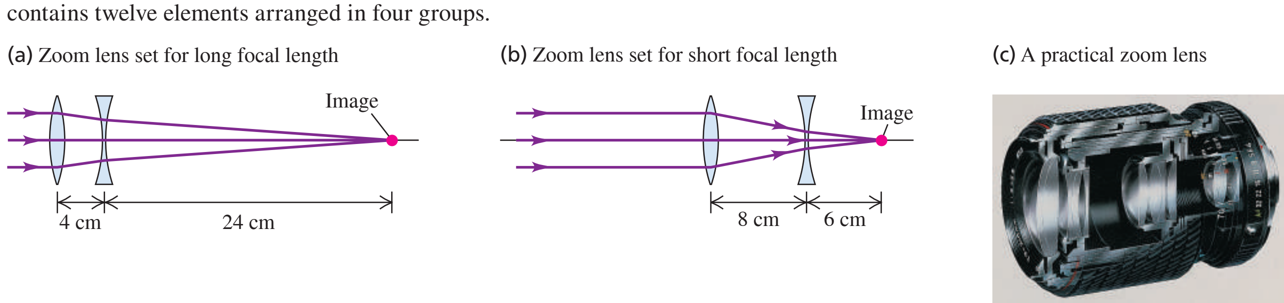
Figure 34.42 A camera lens with an adjustable diaphragm.

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 15 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.

Figure 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?

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.

KEYCONCEPT The f-number of a photographic lens is the ratio of the focal length to the aperture diameter. The larger the f-number, the lower the intensity of the image produced by the lens and the longer the exposure time needed to record the 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.

ANSWER | **(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.

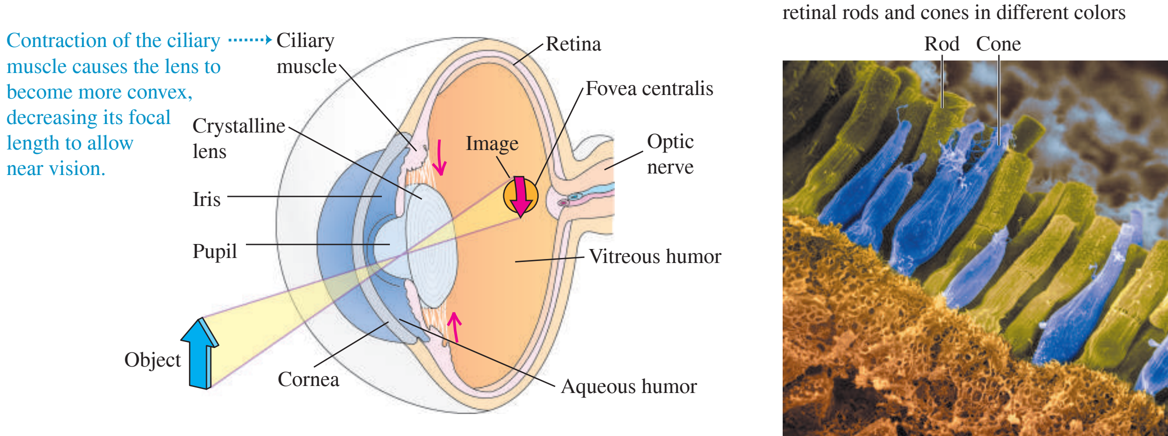
34.6 THE EYE

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

Figure 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



the cornea contains a liquid called the *aqueous humor*. Next comes the *crystalline lens,* a capsule containing a fibrous jelly, hard 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 becomes less flexible throughout a person's life and the ciliary muscles are less able to distort the 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.

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**, next page). In effect the lens moves the object farther away from the eye to a point where a sharp retinal image

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.

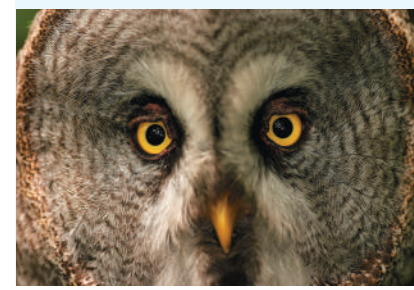


Figure **34.45** Refractive errors for (a) a normal eye, (b) a myopic (nearsighted) eye, and (c) a hyperopic (farsighted) eye viewing a very distant object. In each case, the eye is shown with the ciliary muscle relaxed. The dashed blue curve indicates the required position of the retina.

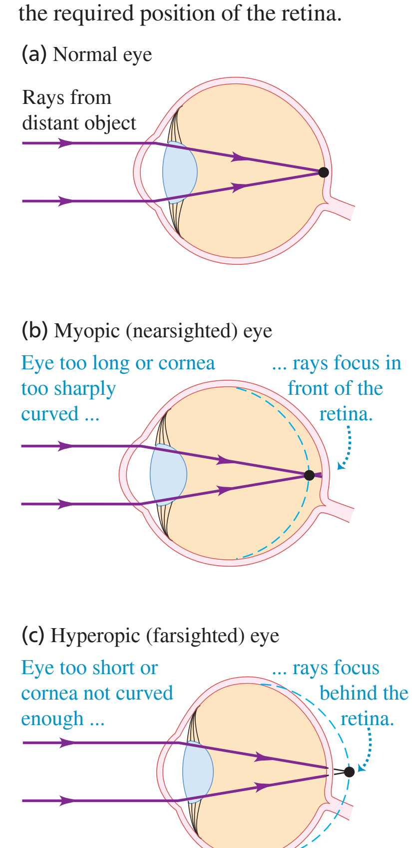


Figure **34.46** (a) An uncorrected hyperopic (farsighted) eye. (b) A positive (converging) lens gives the extra convergence needed for a hyperopic eye to focus the image on the retina.

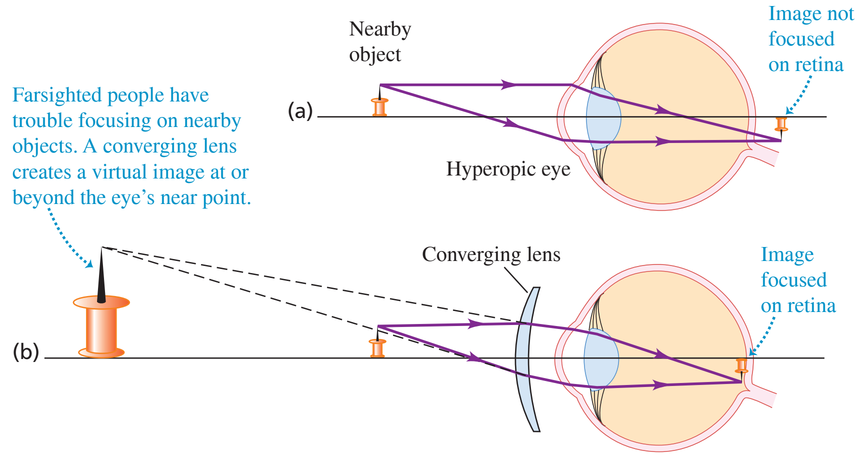


Figure **34.47** (a) An uncorrected myopic (nearsighted) eye. (b) A negative (diverging) lens spreads the rays farther apart to compensate for the excessive convergence of the myopic eye.

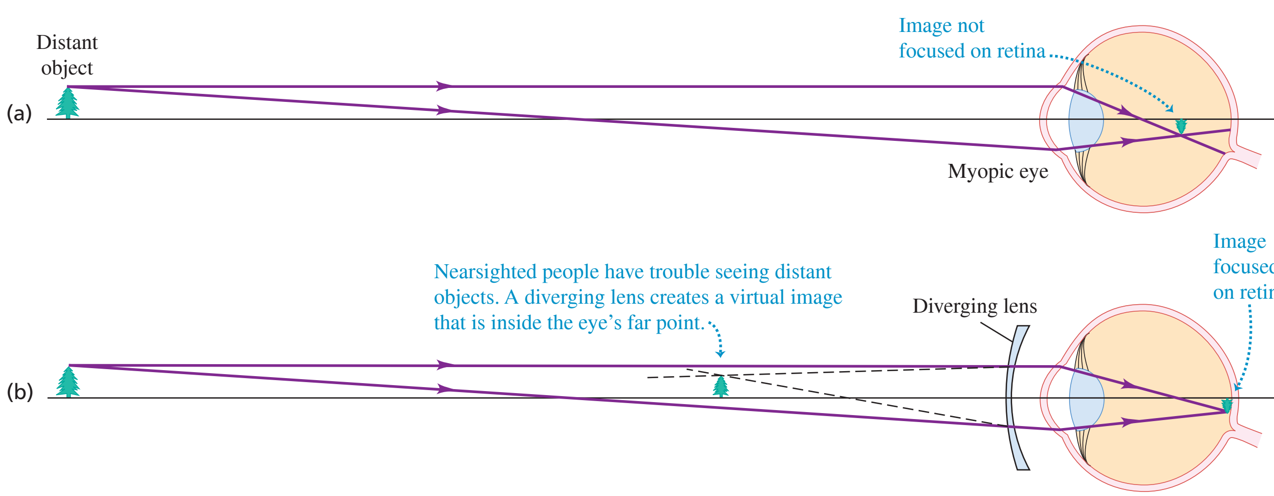


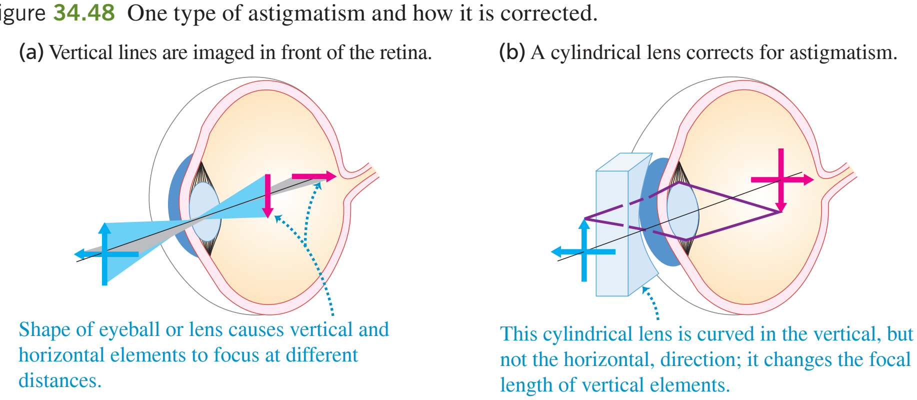
TABLE 34.1 Receding of Near Point with Age

Age (years)	Near Point (cm)
10	7
20	10
30	14
40	22
50	40
60	200

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.

Figure **34.48** One type of astigmatism and how it is corrected.



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 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

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.

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 $1/f$.

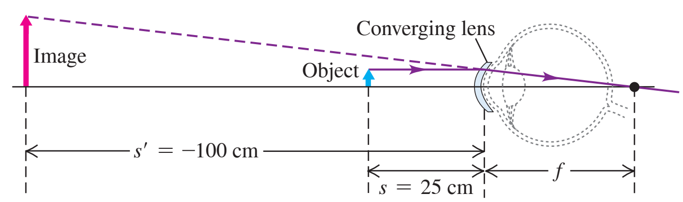
EXECUTE From Eq. (34.16),

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} = \frac{1}{+25\text{ cm}} + \frac{1}{-100\text{ cm}}$$
$$f = +33\text{ cm}$$

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).

Figure 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.



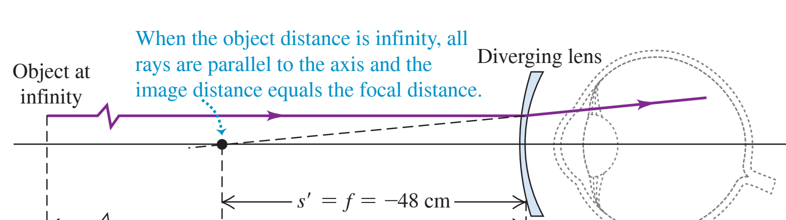
KEYCONCEPT In a hyperopic (farsighted) eye, the cornea and lens cause too little convergence of light rays. Distant objects can form sharp images on the retina, but nearby objects cannot. Hyperopia can be corrected by placing a converging lens in front of the eye; the virtual, erect image made by this lens acts as a more distant object on which the eye can focus.

EXAMPLE 34.14 Correcting for nearsightedness

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.

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.

Figure 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.



Continued

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?

KEYCONCEPT In a myopic (nearsighted) eye, the cornea and lens cause too much convergence of light rays. Nearby objects can form sharp images on the retina, but distant objects cannot. Myopia can be corrected by placing a diverging lens in front of the eye; the virtual, erect image made by this lens acts as a closer object on which the eye can focus.

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.

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.

ANSWER

| **(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 THE MAGNIFIER

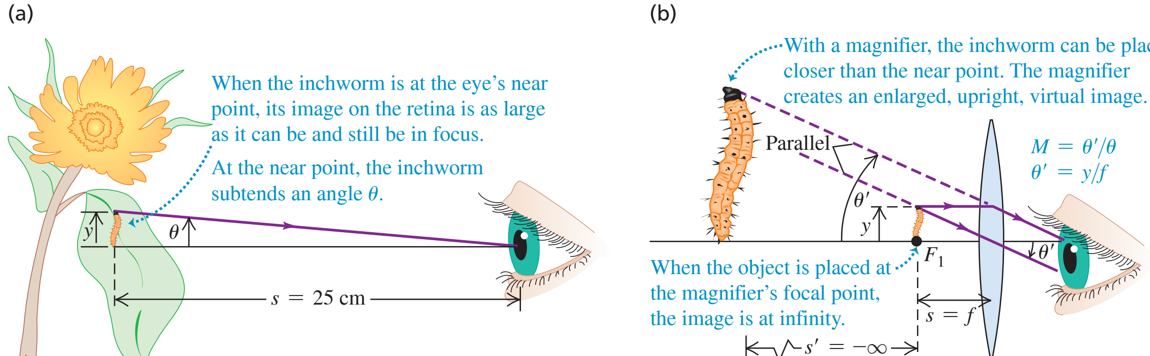
The apparent size of an object is determined by the size of its image on the retina. If the eye is unaided, this size depends on the *angle* θ subtended by the object at the eye, called its **angular size** (**Fig. 34.51a**).

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'll 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, and the angle

Figure 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.



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 you deal 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'll 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.

ANSWER

(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.

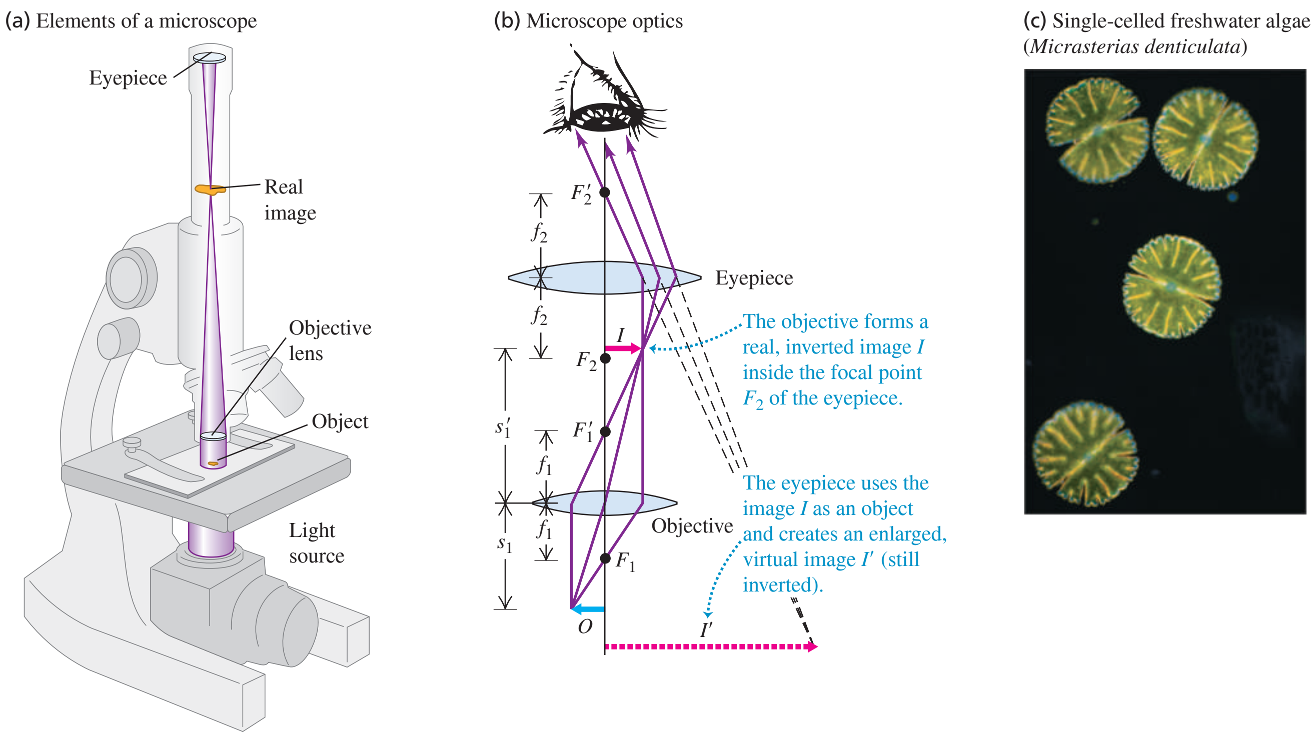
34.8 MICROSCOPES AND TELESCOPES

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 (next page) 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 for a second

Figure 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.



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

$$m_1 = -\frac{s_1'}{s_1} \qquad (34.23)$$

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:

$$M = m_1 M_2 = \frac{(25\text{ cm})s_1'}{f_1 f_2} \qquad \text{(angular magnification for a microscope)} \tag{34.24}$$

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.

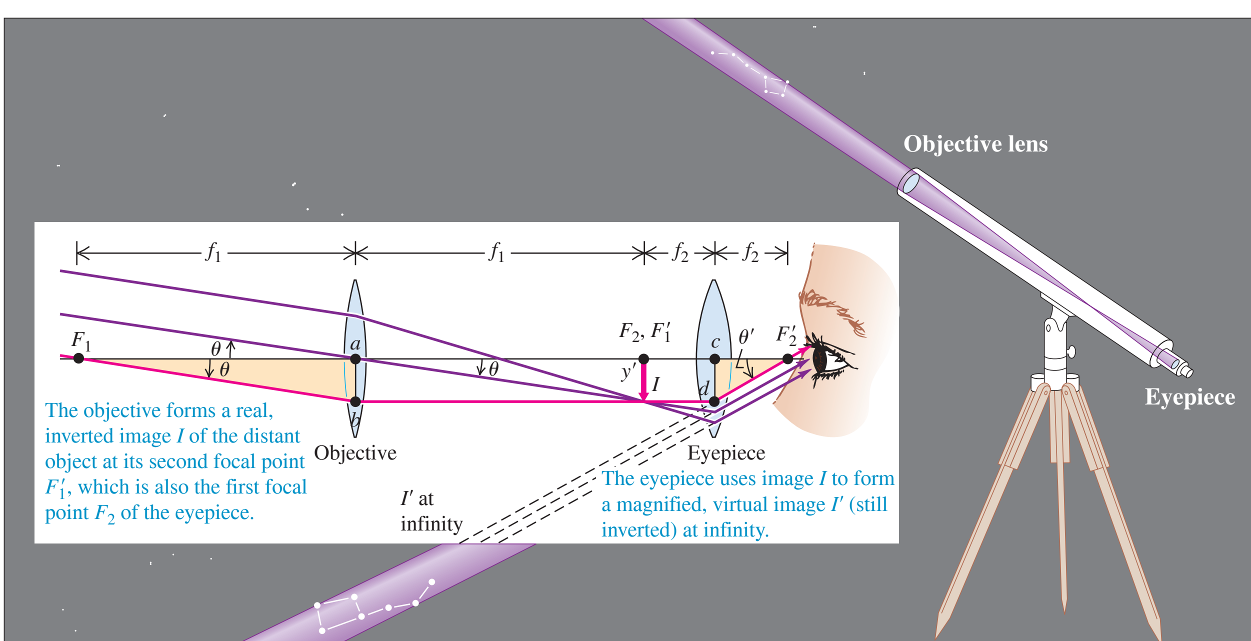
To use a microscope to take a photograph (called a *photomicrograph* or *micrograph*), the eyepiece is removed and a camera placed so that the real image I falls on the camera's electronic sensor or film. Figure 34.52c shows such a photograph. In this case what matters is the *lateral* magnification 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 lens forms a

Figure 34.53 Optical system of an astronomical refracting telescope.



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 $F_2'cd$,

$$\theta = \frac{-y'}{f_1}$$

$$\theta' = \frac{y'}{f_2}$$

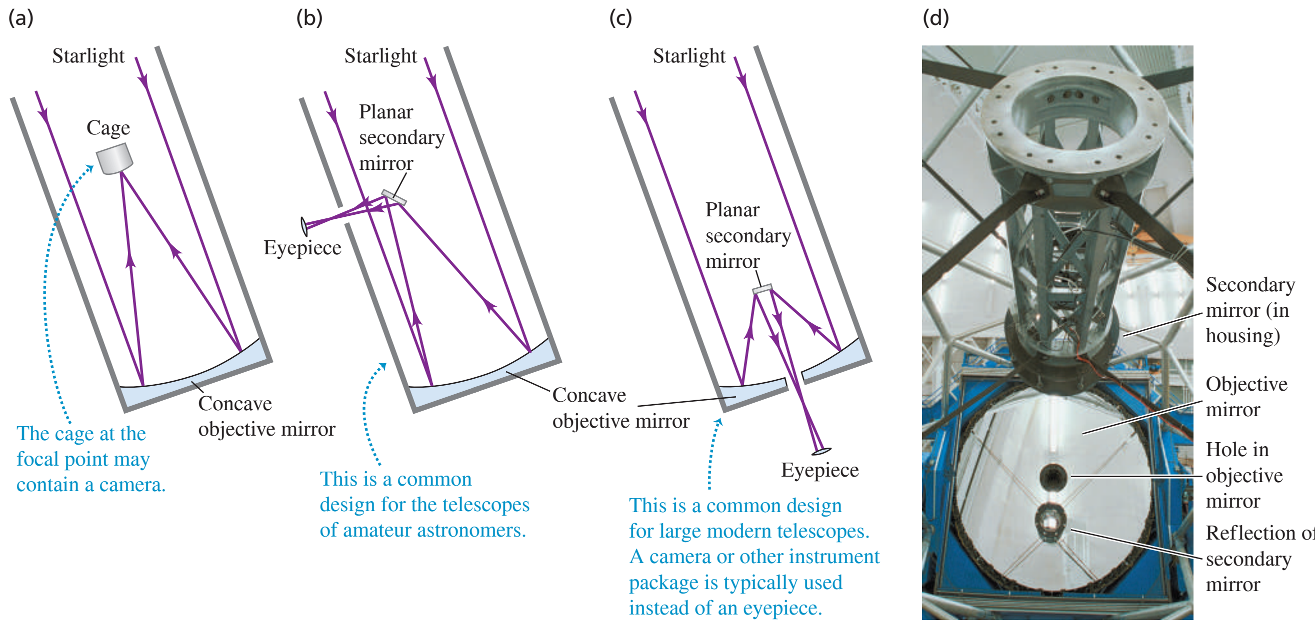
and the angular magnification M is

$$M = \frac{\theta'}{\theta} = -\frac{y'/f_2}{y'/f_1} = -\frac{f_1}{f_2} \qquad \text{(angular magnification for a telescope)} \qquad (34.25)$$

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.

Figure **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.

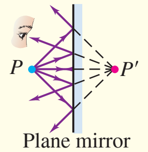
TEST YOUR UNDERSTANDING OF SECTION 34.8 Which gives a lateral magnification of greater absolute value? (i) The objective lens in a microscope (Fig. 34.52); (ii) the objective lens in a refracting telescope (Fig. 34.53); or (iii) not enough information is given to decide.

ANSWER

(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.)

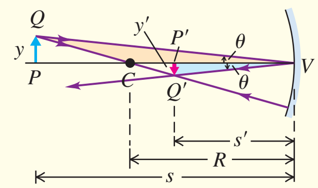
CHAPTER 34 SUMMARY

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.

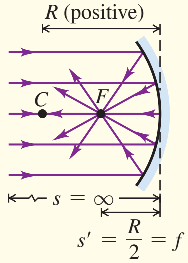


Lateral magnification: The lateral magnification m in any reflecting or refracting situation is defined as the ratio of image height y' to object height y. When m is positive, the image is erect; when m is negative, the image is inverted.

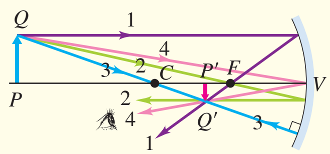
$$m = \frac{y'}{y} \qquad (34.2)$$



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.)



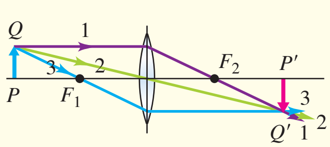
	Plane Mirror	Spherical Mirror	Plane Refracting Surface	Spherical Refracting Surface
Object and image distances	$\frac{1}{s} + \frac{1}{s'} = 0$	$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} = \frac{1}{f}$	$\frac{n_a}{s} + \frac{n_b}{s'} = 0$	$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R}$
Lateral 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.)

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad (34.16)$$

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad (34.19)$$

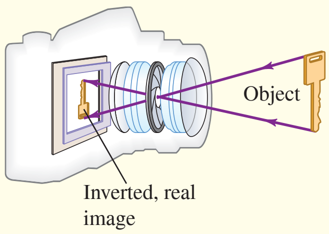


Sign rules: The following sign rules are used with all plane and spherical reflecting and refracting surfaces:

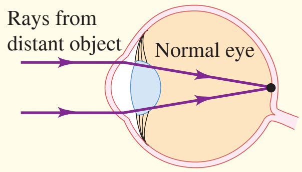
- $s > 0$ when the object is on the incoming side of the surface (a real object); $s < 0$ otherwise.
- $s' > 0$ when the image is on the outgoing side of the surface (a real image); $s' < 0$ otherwise.
- $R > 0$ when the center of curvature is on the outgoing side of the surface; $R < 0$ otherwise.
- $m > 0$ when the image is erect; $m < 0$ when inverted.

Cameras: A camera forms a real, inverted 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.)

$$f\text{-number} = \frac{\text{Focal length}}{\text{Aperture diameter}} = \frac{f}{D} \qquad (34.20)$$

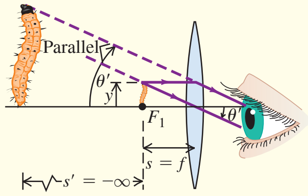


The eye: In the eye, refraction at the surface of the cornea forms a real image on the retina. Adjustment for various object distances is made by squeezing the lens, thereby making it bulge and decreasing its focal length. A nearsighted eye is too long for its lens; a farsighted eye is too short. The power of a corrective lens, in diopters, is the reciprocal of the focal length in meters. (See Examples 34.13 and 34.14.)

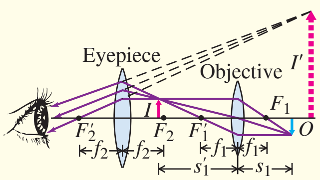


The simple magnifier: The simple magnifier creates a virtual image whose angular size θ' is larger than the angular size θ of the object itself at a distance of 25 cm, the nominal closest distance for comfortable viewing. The angular magnification M of a simple magnifier is the ratio of the angular size of the virtual image to that of the object at this distance.

$$M = \frac{\theta'}{\theta} = \frac{25\text{ cm}}{f} \qquad (34.22)$$



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.



Chapter 34 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 34.1, 34.3, and 34.4 (Section 34.2) before attempting these problems.

VP34.4.1 You place a light bulb 15.0 cm in front of a concave spherical mirror. The mirror forms an image of the bulb on a wall 4.50 m in front of the mirror. Find (a) the mirror's focal length, (b) the mirror's radius of curvature, and (c) the lateral magnification of the image.

VP34.4.2 A concave spherical mirror has radius of curvature 37.0 cm. Find the image distance and lateral magnification for each of the following object distances. In each case state whether the image is real or virtual, whether it is erect or inverted, and whether it is larger or smaller than the object. (a) 11.0 cm; (b) 31.0 cm; (c) 55.0 cm.

VP34.4.3 A spherical mirror has radius of curvature −44.0 cm. When you look at your eye in the mirror, your eye's reflection appears to be 18.0 cm behind the mirror's surface. (a) Is the mirror concave or convex? (b) How far is your eye from the mirror? (c) What is the lateral magnification of the image of your eye? Is the image real or virtual? Erect or inverted? Larger or smaller than your eye?

VP34.4.4 A convex spherical mirror has radius of curvature −37.0 cm. Find the image distance and lateral magnification for each of the following object distances. In each case state whether the image is real or virtual, whether it is erect or inverted, and whether it is larger or smaller than the object. (a) 11.0 cm; (b) 31.0 cm; (c) 55.0 cm.

Be sure to review EXAMPLE 34.8 (Section 34.4) before attempting these problems.

VP34.8.1 Both sides of a double convex thin lens have radii of curvature of the same magnitude. The lens is made of glass with index of refraction 1.65, and the focal length of the lens is +30.0 cm. Find the radius of curvature of (a) the front surface and (b) the back surface.

VP34.8.2 The side of a thin lens that faces the object is convex and has radius of curvature 15.0 cm. The other side is concave and has radius of curvature 25.0 cm. The lens is made of glass with index of refraction 1.55. (a) Is this lens thicker at its center or at its edges? (b) What is the focal length of the lens? (c) Is the lens converging or diverging?

VP34.8.3 The side of a thin lens that faces the object is convex and has radius of curvature 25.0 cm. The other side is concave and has radius of curvature 15.0 cm. The lens is made of glass with index of refraction 1.55. (a) Is this lens thicker at its center or at its edges? (b) What is the focal length of the lens? (c) Is the lens converging or diverging?

VP34.8.4 You are designing a lens to be made of glass with index of refraction 1.70. The first surface (the surface toward the object) is to be convex with radius of curvature 28.0 cm, and the focal length of the lens is to be 14.0 cm. (a) What must be the radius of curvature of the second surface (the surface away from the object)? (b) Will the second surface be concave or convex?

Be sure to review EXAMPLES 34.9 and 34.10 (Section 34.4) before attempting these problems.

VP34.10.1 You place a strawberry 15.0 cm in front of a converging lens with focal length +25.0 cm. (a) What is the image position? Is it on the same side of the lens as the strawberry or on the opposite side? (b) What is the lateral magnification of the image? (c) Is the image real or virtual? Erect or inverted? Larger or smaller than the strawberry?

VP34.10.2 If you put an eraser 28.0 cm in front of a lens, an image of the eraser is formed 42.0 cm behind the lens. (a) Find the focal length of the lens. Is the lens converging or diverging? (b) What is the lateral magnification of the image? (c) Is the image real or virtual? Erect or inverted? Larger or smaller than the eraser?

VP34.10.3 When you place a thumbtack 48.0 cm in front of a lens, the resulting image is real, inverted, and the same size as the thumbtack. (a) Is the lens converging or diverging? (b) What is the image distance? (c) What is the focal length of the lens?

VP34.10.4 The front door of many houses has a small "peephole" lens that gives a wide-angle view of the outside. To provide this wide-angle view, the lens must produce an image of an object outside that is erect and smaller than the object. For one lens, if an object 20.0 cm tall is 2.50 m in front of the door, its erect image is only 4.00 cm tall. (a) What is the lateral magnification of the image? (b) What is the image distance? Is the image on the same side of the lens as the object or on the opposite side? (c) What is the focal length of the lens? Is the lens converging or diverging?

BRIDGING PROBLEM Image Formation by a Wine Goblet

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

•, ••, •••: 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 $s = 10\,\text{cm}$ 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.

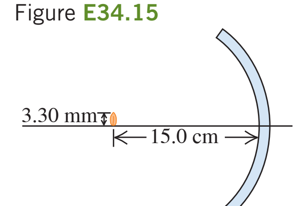
34.12 • For a concave spherical mirror that has focal length $f = +18.0$ cm, what is the distance of an object from the mirror's vertex if the image is real and has the same height as the object?

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**



34.16 • An object 0.600 cm tall is placed 24.0 cm to the left of the vertex of a concave spherical mirror. The image of the object is inverted and is 2.50 cm tall. What is the radius of curvature of the mirror?

34.17 • A shiny spoon provides both a concave mirror and a convex mirror, one on each side. (a) Hold a shiny spoon about 25 cm from your face. Observe your image in the concave side. Is your image up-side down or right-side up? (b) Does that mean that your image is real or virtual? (c) Estimate the height of your head. (d) Estimate the height of the image of your head. (Use the size of the spoon as a guide.) (e) Using these estimates, determine the magnification. (f) Use Eqs. (34.6) and (34.5) to estimate the radius of curvature of the spoon. (g) Now observe your image in the convex side of the spoon. Is it up-side down or right-side up? (h) Is your image real or virtual?

Section 34.3 Refraction at a Spherical Surface

34.18 •• 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?

34.19 • 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?

34.20 • Parallel rays from a distant object are traveling in air and then are incident on the concave end of a glass rod with a radius of curvature of 15.0 cm. The refractive index of the glass is 1.50. What is the distance between the vertex of the glass surface and the image formed by the refraction at the concave surface of the rod? Is the image in the air or in the glass?

34.21 • 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?

34.22 • 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?

34.23 •• **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?

34.24 • 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.

34.25 •• The glass rod of Exercise 34.24 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?

34.26 •• 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?

34.27 • A 3.20-mm-tall object is 20.0 cm from the vertex of a spherical mirror. The mirror forms an image 60.0 cm from the mirror. (a) If the image is real, what is the radius of curvature of the mirror? What is the height of the image? Is it upright or inverted? (b) Repeat part (a) for the case where the image is virtual.

Section 34.4 Thin Lenses

34.28 •• To determine the focal length f of a converging thin lens, you place a 4.00-mm-tall object a distance s to the left of the lens and measure the height h' of the real image that is formed to the right of the lens. You repeat this process for several values of s that produce a real image. After graphing your results as $1/h'$ versus s, both in cm, you find that they lie close to a straight line that has slope 0.208 cm^{-2}. What is the focal length of the lens?

34.29 • 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.

34.30 • 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.

34.31 • 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?

34.32 • 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?

34.33 •• 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?

34.34 • 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?

34.35 •• 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?

34.36 • 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?

34.37 • A thin lens is made of glass that has refractive index $n = 1.50$. The lens is surrounded by air. The left-hand surface of the lens is flat and the right-hand spherical surface is convex with radius of curvature 20.0 cm, so the lens is thicker in the middle than at its edges. What is the height of the image formed by the lens for a 6.00-mm-tall object that is placed 20.0 cm to the left of the center of the lens?

34.38 •• 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?

34.39 •• A thin lens is made of glass that has refractive index $n = 1.50$. The lens is surrounded by air. The left-hand spherical surface of the lens is concave with radius of curvature R and the right-hand side is flat, so the lens is thinner in the middle than at its edges. An object is placed 12.0 cm to the left of the center of the lens. What is the value of R if the image formed by the lens is 8.00 cm to the left of the center of the lens?

34.40 • 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.

34.41 • Repeat Exercise 34.40 for the case in which the lens is diverging, with a focal length of −48.0 cm.

34.42 • 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.

34.43 •• **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.

34.44 •• **Combination of Lenses II.** Repeat Exercise 34.43 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 34.43.

34.45 •• **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 exercises.)

34.46 • The focal points of a thin diverging lens are 25.0 cm from the center of the lens. An object is placed to the left of the lens, and the lens forms an image of the object that is 18.0 cm from the lens. (a) Is the image to the left or right of the lens? (b) How far is the object from the center of the lens? (c) Is the height of the image less than, greater than, or the same as the height of the object?

34.47 • Figure 34.37e shows the principal-ray diagram of the image formation by a converging lens with focal length f and object distance $s = 2f/3$ to the left of the lens. (a) Use Eq. (34.17) to find the image distance s' in terms of f. Based on your result, is the image real or virtual? Is it to the left or right of the lens? (b) If the height of the object is h, use Eq. (34.17) to find the height of the image. Does that mean the image is upright or inverted? (c) Do your results agree with what is shown in Fig. 34.37e?

34.48 •• An object is to the left of a thin lens. The lens forms an image on a screen that is 2.60 m to the right of the object. The height of the image is 2.50 times the height of the object. (a) Is the image upright or inverted? (b) What is the focal length of the lens? Is the lens converging or diverging?

Section 34.5 Cameras

34.49 • 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 × 36 mm?

34.50 • 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?

34.51 • **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.

Section 34.6 The Eye

34.52 •• 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?

34.53 •• 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?

34.54 • 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?

34.55 •• 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 34.54 prefers ordinary glasses to contact lenses. What focal length lenses are needed to correct his vision, and what is their power in diopters?

34.56 • 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?

34.57 •• A woman wears bifocal glasses with the lenses 2.0 cm in front of her eyes. The upper half of each lens has power −0.500 diopter and corrects her far vision so that she can focus clearly on distant objects when looking through that half. The lower half of each lens has power +2.00 diopters and corrects her near vision when she looks through that half. (a) What are the far point and near point of her eyes? (b) While the woman is repairing a leaky pipe under her kitchen sink, she looks at close objects through the upper half of her bifocal lenses. What is the closest object that she can see clearly?

Section 34.7 The Magnifier

34.58 •• 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.

34.59 • 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?

34.60 • 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?

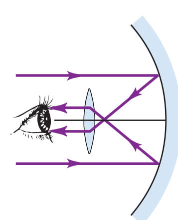
Section 34.8 Microscopes and Telescopes

34.61 •• 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**



34.64 • A compound microscope has an objective lens with focal length 14.0 mm and an eyepiece with focal length 20.0 mm. The final image is at infinity. The object to be viewed is placed 2.0 mm beyond the focal point of the objective lens. (a) What is the distance between the two lenses? (b) Without making the approximation $s_1 \approx f_1$, use $M = m_1 M_2$ with $m_1 = -s_1'/s_1$ to find the overall angular magnification of the microscope. (c) What is the percentage difference between your result and the result obtained if the approximation $s_1 \approx f_1$ is used to find M?

34.65 •• The overall angular magnification of a microscope is $M = -178$. The eyepiece has focal length 15.0 mm and the final image is at infinity. The separation between the two lenses is 202 mm. What is the focal length of the objective? Do not use the approximation $s_1 \approx f_1$ in the expression for M.

PROBLEMS

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?

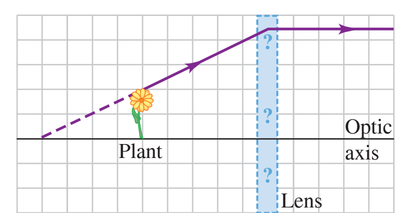
34.69 •• **CP CALC** You are in your car driving on a highway at 25 m/s when you glance in the passenger-side mirror (a convex mirror with radius of curvature 150 cm) and notice a truck approaching. If the image of the truck is approaching the vertex of the mirror at a speed of 1.9 m/s when the truck 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 •• An object with height 4.00 mm is placed 28.0 cm to the left of a converging lens that has focal length 8.40 cm. A second lens is placed 8.00 cm to the right of the converging lens. (a) What is the focal length of the second lens if the final image is inverted relative to the 4.00-mm-tall object and has height 5.60 mm? (b) What is the distance between the original object and the final image?

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 •• Fill a transparent cylindrical drinking glass with water. Hold a small coin between two fingers submerged in the water immediately inside the glass, and then view the coin from outside the glass. Watch the coin as you move it steadily toward the back of the glass. (a) Does the coin appear to widen as you move it backward in the glass? (b) Estimate the apparent horizontal magnification m_{back} when the coin is flush with the back side of the glass. (c) Since the curvature is in a horizontal plane, this horizontal magnification is the same as the lateral magnification analyzed in Section 34.3. Use Eqs. (34.11) and (34.12) to derive an expression for the horizontal magnification m as a function of the object position s, the magnitude of the radius of curvature of the glass $|R|$, and the index of refraction n of the water. (d) Use your result to derive an expression for n as a function of m_{back}. (e) Substitute your estimate for m_{back} into your equation for n to estimate the index of refraction of the water.

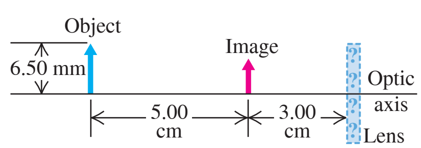
34.77 •• (a) You want to use a lens with a focal length of 35.0 cm to produce a real image of an object, with the height of the image twice the height of the object. What kind of lens do you need, and where should the object be placed? (b) Suppose you want a virtual image of the same object, with the 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?

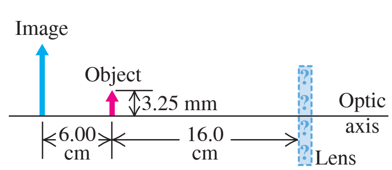
34.80 • **Figure P34.80** shows an object and its image formed by a thin lens. (a) What is the focal length of the lens, and what type of lens (converging or diverging) is it? (b) What is the height of the image? Is it real or virtual?

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**



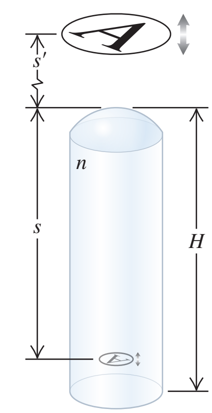
34.82 ••• A transparent rod 30.0 cm long is cut flat at one end and rounded to a hemispherical surface of radius 10.0 cm at the other end. A small object is embedded within the rod along its axis and halfway between its ends, 15.0 cm from the flat end and 15.0 cm from the vertex of the curved end. When 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 ••• CP CALC A transparent cylindrical tube with radius $r = 1.50$ cm has a flat circular bottom and a top that is convex as seen from above, with radius of curvature of magnitude 2.50 cm. The cylinder is filled with quinoline, a colorless highly refractive liquid with index of refraction $n = 1.627$. Near the bottom of the tube, immersed in the liquid, is a luminescent LED display mounted on a platform whose height may be varied. The display is the letter A inside a circle that has a diameter of 1.00 cm. A real image of this display is formed at a height s' above the top of the tube, as shown in **Fig. P34.85**. (a) What is the minimum tube height H for which this display apparatus can function? (b) The luminescent object is moved up and down periodically so that the real image moves up and down in the air above the tube. A mist in the air renders this display visible and dramatic. If we want the image to move from 50.0 cm above the top of the cylinder to 1.00 m above the top of the cylinder during this motion, what is the corresponding range of motion for the object distance s? (c) What is the height s' of the image when the object is halfway through its motion? (d) The object in part (b) is moved sinusoidally according to $s = A\sin(\omega t)$, where A is the amplitude and ω is the angular frequency. The frequency of the motion is 1.00 Hz. What is the velocity of the image when the object is at its midpoint and traveling upward? (e) What is the diameter of the image when it is at its largest size?

Figure **P34.85**



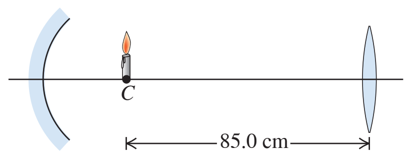
34.86 • An object is placed 22.0 cm from a screen. (a) At what two points between object and screen may a converging lens with a 3.00 cm focal length be placed to obtain an image on the screen? (b) What is the magnification of the image for each position of the lens?

34.87 •• 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 with the same power instead?

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** (next page), 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 image, 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 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.

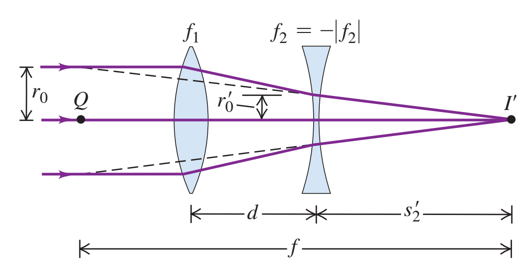
34.96 •• A camera with a 90-mm-focal-length lens is focused on an object 1.30 m from the lens. To refocus on an object 6.50 m from the lens, by how much must the distance between the lens and the sensor be changed? To refocus on the more distant object, is the lens moved toward or away from the sensor?

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 34.52. 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 34.52, what is the distance from the cornea vertex to the retina? What does this tell you about the shape of the nearsighted eye?

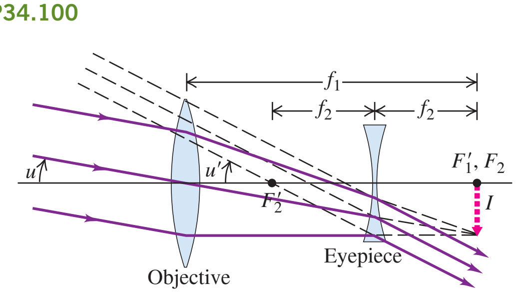
34.99 ••• Focal Length of a Zoom Lens. Figure P34.99 shows a simple version of a zoom lens. The converging lens has focal length f_1 and the diverging lens has focal length $f_2 = -|f_2|$. The two lenses are separated by a variable distance d that is always less than f_1. Also, the magnitude of the focal 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 **P34.99**



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.

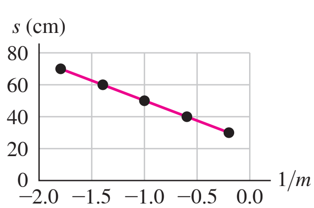
Figure **P34.100**



34.101 •• 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?

34.102 •• 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 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:

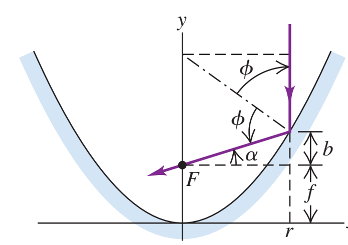
s (cm)	22.5	25.0	30.0	35.0	40.0	45.0
s' (cm)	271.6	148.3	89.4	71.1	60.8	53.2

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

34.104 ••• In a spherical mirror, rays parallel to but relatively distant from the optic axis do not reflect precisely to the focal point, and this causes spherical aberration of images. In parabolic mirrors, by contrast, all paraxial rays that enter the mirror, arbitrarily distant from the optic axis, converge to a focal point on the optic axis without any approximation whatsoever. We can prove this and determine the location of the focal point by considering **Fig. P34.104**. The mirror is a parabola defined by $y = ax^2$, with a in units of $(\text{distance})^{-1}$, rotated around the y-axis. Consider the ray that enters the mirror a distance r from the axis. (a) What is the slope of the parabola at the point where the ray strikes the mirror, in terms of a and r? (b) What is the slope of the dashed line that is normal to the curve at that point? (c) What is the angle ϕ? (d) Using trigonometry, find the angle α in terms of ϕ. (e) Using trigonometry, find the distance b in terms of a and r. [*Hint:* The identities $\tan(\theta - \pi/2) = -\cot\theta$ and $\cot(2\theta) = (\cot^2\theta - 1)/2\cot\theta$ may be useful.] (f) Find the distance f in terms of a. Note that your answer does not depend on the distance r; this proves our assertion.

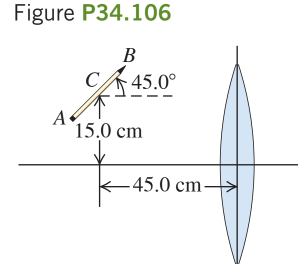
Figure **P34.104**



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 34.52, 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.)

MCAT-STYLE 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.

Key Example √ARIATION Problems

VP34.4.1 (a) 14.5 cm (b) 29.0 cm (c) −30.0
VP34.4.2 (a) $s' = -27.1$ cm, $m = +2.47$, virtual, erect, larger
(b) $s' = +45.9$ cm, $m = -1.48$, real, inverted, larger
(c) $s' = +27.9$ cm, $m = -0.507$, real, inverted, smaller
VP34.4.3 (a) convex (b) 99.0 cm (c) +0.182, virtual, erect, smaller
VP34.4.4 (a) $s' = -6.90$ cm, $m = +0.627$, virtual, erect, smaller
(b) $s' = -11.6$ cm, $m = +0.374$, virtual, erect, smaller
(c) $s' = -13.8$ cm, $m = +0.252$, virtual, erect, smaller
VP34.8.1 (a) +39.0 cm (b) −39.0 cm
VP34.8.2 (a) center (b) +68 cm (c) converging
VP34.8.3 (a) edges (b) −68 cm (c) diverging
VP34.8.4 (a) −15 cm (b) convex
VP34.10.1 (a) −37.5 cm, same side (b) +2.50 (c) virtual, erect, larger
VP34.10.2 (a) +16.8 cm, converging (b) −1.50 (c) real, inverted, larger
VP34.10.3 (a) converging (b) +48.0 cm (c) +24.0 cm
VP34.10.4 (a) +0.200 (b) −50.0 cm, same side (c) −62.5 cm, diverging

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 OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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.

PRINCIPLE OF SUPERPOSITION 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.

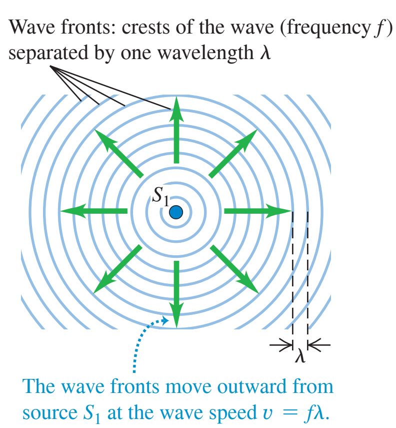
(In some special situations, such as electromagnetic waves propagating in a crystal, this principle may not apply. A discussion of these situations is beyond our scope.)

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.

Figure 35.1 A "snapshot" of sinusoidal waves of frequency f and wavelength λ spreading out from source S_1 in all directions.



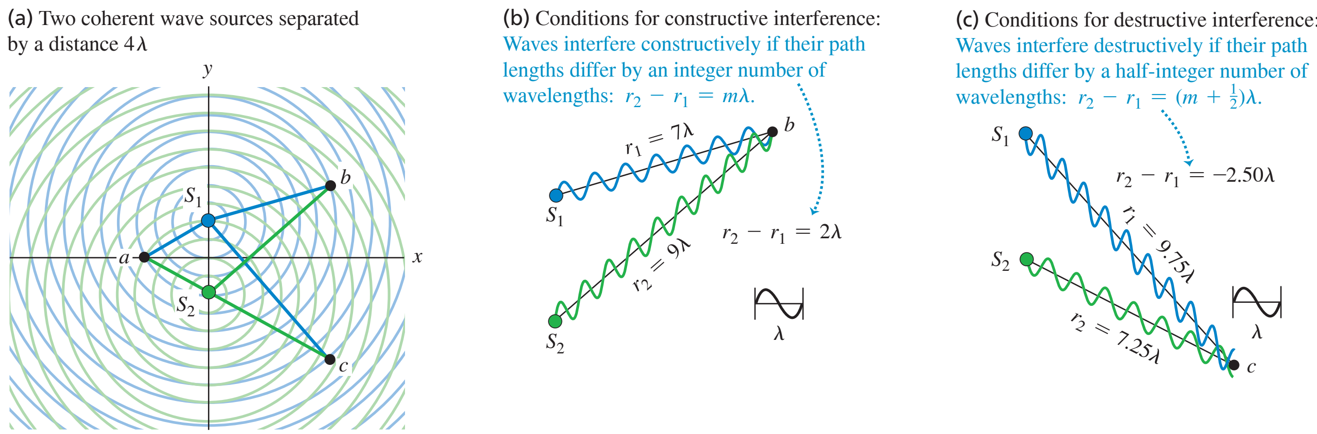
Interference effects are most easily seen when we combine *sinusoidal* waves with a single frequency f and wavelength λ. **Figure 35.1** shows a "snapshot" of a *single* source S_1 of sinusoidal waves and some of the wave fronts produced by this source. The figure shows only the wave fronts corresponding 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.

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'll assume that we are working with monochromatic waves (unless we explicitly state otherwise).

Constructive and Destructive Interference

Figure 35.2a 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'll see that if 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.

Figure **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.



If the waves emitted by the two coherent sources are *transverse,* like electromagnetic waves, then we’ll 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 z-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 integer multiple of the wavelength λ:

$$r_2 - r_1 = m\lambda \qquad (m = 0, \pm 1, \pm 2, \pm 3, \ldots) \qquad \text{(constructive interference, sources in phase)} \tag{35.1}$$

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-integer* 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 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.

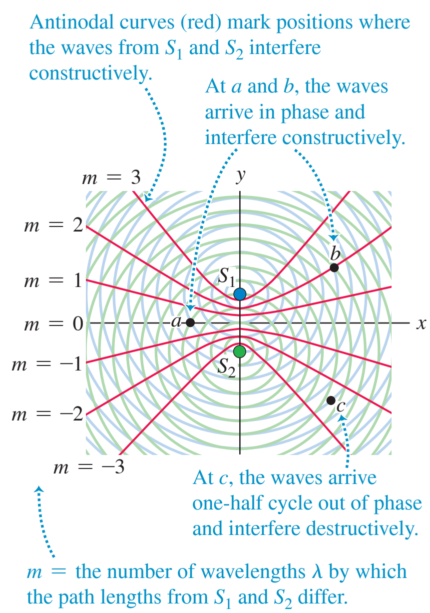


Figure 35.3 shows the same situation as in Fig. 35.2a, but with red curves that show all positions where *constructive* interference occurs. On each curve, the path difference $r_2 - r_1$ is equal to an integer m times the wavelength, as in Eq. (35.1). These curves are called **antinodal curves.** They 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 be monochromatic, 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.)

ANSWER (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.

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.

35.2 TWO-SOURCE INTERFERENCE OF LIGHT

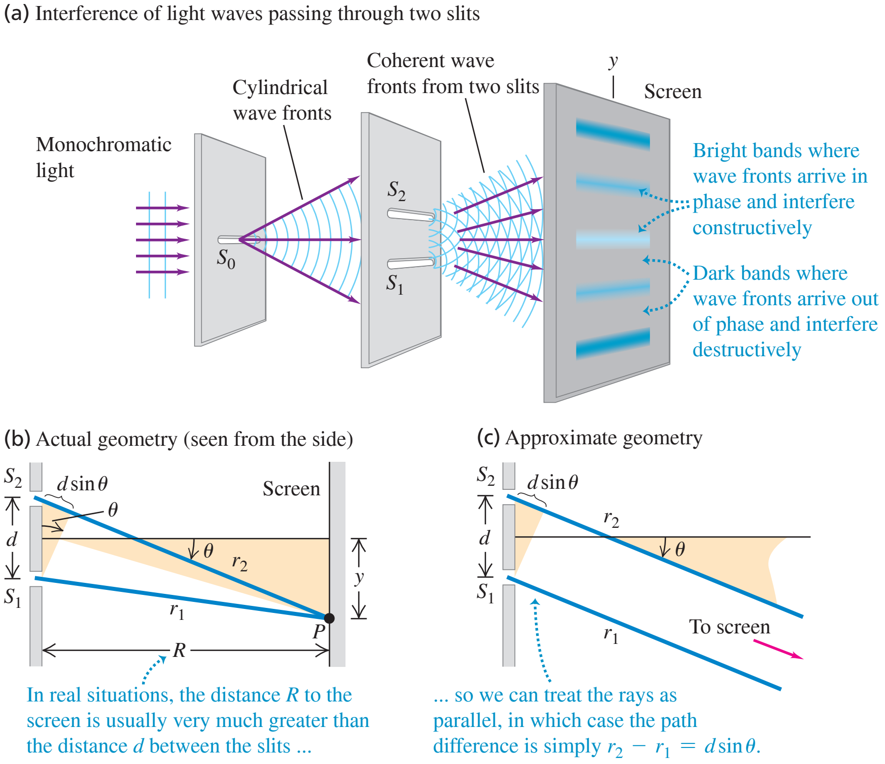
The interference pattern produced by two coherent sources of *water* waves of the same wavelength can be readily seen in a ripple tank with a shallow layer of water (**Fig. 35.4**). This pattern is not directly visible when the interference is between *light* waves, since light traveling in a uniform 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.

Figure 35.4 The concepts of constructive interference and destructive interference apply to these water waves as well as to light waves and sound waves.

Figure 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.



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. 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

$$r_2 - r_1 = d\sin\theta \qquad (35.3)$$

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 integer 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

$$d\sin\theta = m\lambda \qquad (m = 0, \pm1, \pm2, \ldots) \qquad (35.4)$$

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-integer number of wavelengths, $\left(m + \frac{1}{2}\right)\lambda$:

Destructive interference, two slits:

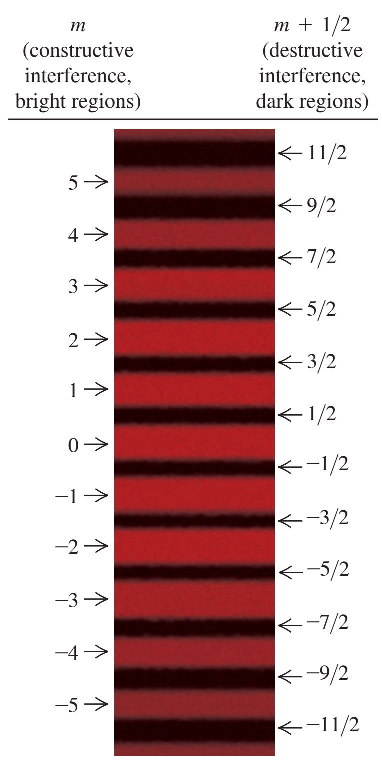
Distance between slits; Wavelength

$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \qquad (m = 0, \pm1, \pm2, \ldots) \qquad (35.5)$$

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**.

Figure 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.



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

$$y_m = R\frac{m\lambda}{d} \qquad (m = 0, \pm1, \pm2, \ldots) \qquad (35.6)$$

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.

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.

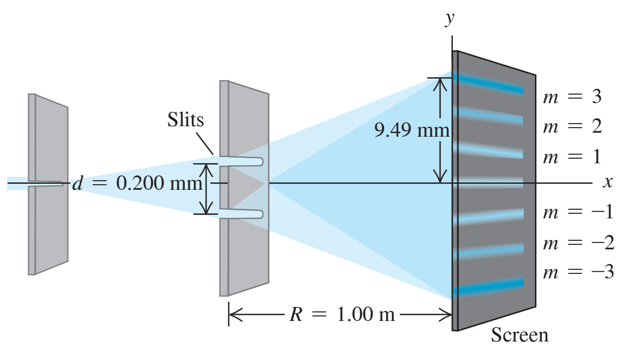
While we have described the experiment that Young performed with visible light, the results given in Eqs. (35.4) and (35.5) are valid for *any* type of wave, provided that the resultant wave from two coherent sources is detected at a point that is far away in comparison to the separation d.

EXAMPLE 35.1 Two-slit interference

WITH VARIATION PROBLEMS

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 bright fringe. Find the wavelength of the light.

Figure **35.7** Using a two-slit interference experiment to measure the wavelength of light.



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 value 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?

KEYCONCEPT In two-slit interference, each wave front of a monochromatic wave enters the two slits simultaneously. Hence the waves that emerge from the slits always have the same phase relationship and so produce a steady interference pattern. Equation (35.6) gives the positions of the bright fringes where constructive interference occurs.

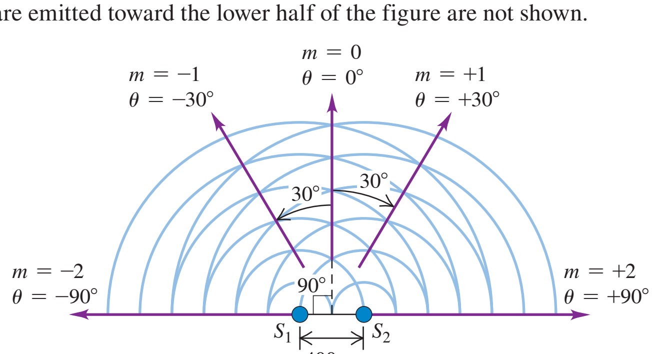
EXAMPLE 35.2 Broadcast pattern of a radio station

WITH VARIATION PROBLEMS

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?

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.

Figure **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.



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.

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.

KEYCONCEPT Equations (35.4) and (35.5) give the angles in two-slit interference for which constructive interference (bright regions) and destructive interference (dark regions) occur. These equations are also valid for any two monochromatic sources that emit in phase.

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.

ANSWER

(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 INTENSITY IN INTERFERENCE PATTERNS

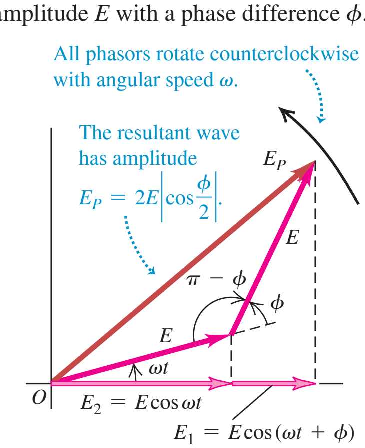
In Section 35.2 we found the positions of maximum and minimum intensity in a two-source interference pattern. Let's now see how to find the intensity at *any* point in the pattern. To do this, we have to combine the two sinusoidally varying fields (from the two sources) at a point P in the radiation 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'll 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.

Figure 35.9 Phasor diagram for the superposition at a point P of two waves of equal amplitude E with a phase difference ϕ.



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 $\cos(\pi - \phi) = -\cos\phi$:

$$\begin{aligned} E_P{}^2 &= E^2 + E^2 - 2E^2\cos(\pi - \phi) \\ &= E^2 + E^2 + 2E^2\cos\phi \end{aligned}$$

Then, using the identity $1 + \cos\phi = 2\cos^2(\phi/2)$, we obtain

$$E_P{}^2 = 2E^2(1 + \cos\phi) = 4E^2\cos^2\left(\frac{\phi}{2}\right)$$

Amplitude of wave from one source

Electric-field amplitude in two-source interference

$$E_P = 2E\left|\cos\frac{\phi}{2}\right| \qquad (35.7)$$

Phase difference between waves

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:

$$I = S_{av} = \frac{E_P{}^2}{2\mu_0 c} = \tfrac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_P{}^2 = \tfrac{1}{2}\epsilon_0 c E_P{}^2 \qquad (35.8)$$

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

$$I = \tfrac{1}{2}\epsilon_0 c E_P{}^2 = 2\epsilon_0 c E^2 \cos^2\frac{\phi}{2} \qquad (35.9)$$

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

Maximum intensity

Intensity in two-source interference

$$I = I_0\cos^2\frac{\phi}{2} \qquad (35.10)$$

Phase difference between waves

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 $\phi = 2\pi \text{ rad} = 360°$. When the path difference is $\lambda/2$, $\phi = \pi \text{ rad} = 180°$, 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:

$$\frac{\phi}{2\pi} = \frac{r_2 - r_1}{\lambda}$$

Phase difference in two-source interference

$$\phi = \frac{2\pi}{\underbrace{\lambda}_{\text{Wavelength}}}\overbrace{(\underbrace{r_2}_{\text{Distance from source 2}} - \underbrace{r_1}_{\text{Distance from source 1}})}^{\text{Path difference}} = \overbrace{k}^{\text{Wave number} = 2\pi/\lambda}(r_2 - r_1) \tag{35.11}$$

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

$$\lambda = \frac{\lambda_0}{n} \qquad \text{and} \qquad k = nk_0 \tag{35.12}$$

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:

$$I = I_0\cos^2\left(\frac{kdy}{2R}\right) = I_0\cos^2\left(\frac{\pi dy}{\lambda R}\right) \quad \text{(intensity in two-slit interference)} \tag{35.15}$$

Figure **35.10** Intensity distribution in the interference pattern from two identical slits.

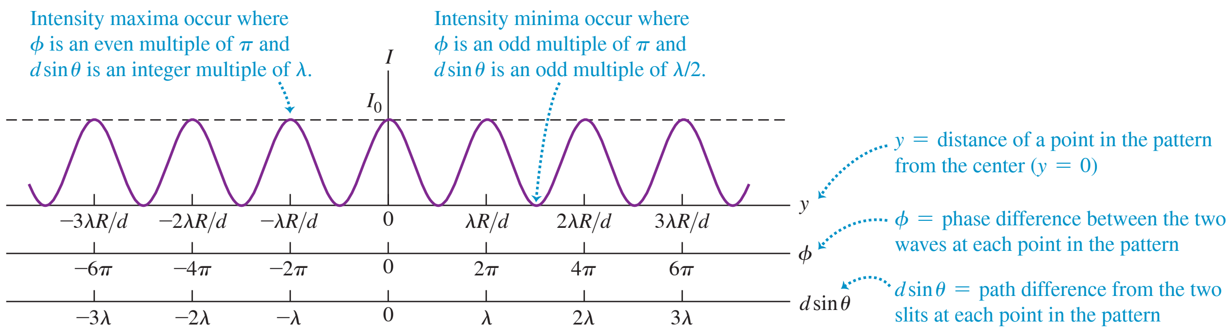


Figure 35.10 shows a graph of Eq. (35.15); we can compare this with the photographically recorded pattern of Fig. 35.6. All peaks in Fig. 35.10 have the same intensity, while those in Fig. 35.6 fade off as we go away from the center. We'll explore the reasons for this variation in peak intensity in Chapter 36.

EXAMPLE 35.3 A directional transmitting antenna array

WITH VARIATION PROBLEMS

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.

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 $\sin\theta = \pm 0.375$, ± 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.

KEYCONCEPT The intensity at a given point in a two-source interference pattern depends on the phase difference ϕ between the waves that arrive at that point from the two sources [Eqs. (35.10) and (35.11)]. The intensity is *maximum* at points where ϕ is an *even* multiple of π ($0, \pm 2\pi, \pm 4\pi, \ldots$) and *minimum* at point where ϕ is an *odd* multiple of π ($\pm\pi, \pm 3\pi, \ldots$).

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.

ANSWER

(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 INTERFERENCE IN THIN FILMS

?

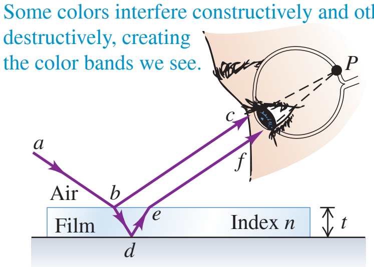
You often see bright bands of color when light reflects from a thin layer of oil floating on water or from a soap bubble (see the photograph that opens this chapter). These are the results of interference. Light waves are reflected from the front and back surfaces of such thin films, and constructive 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.

Figure **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

$$E_{\mathrm{r}} = \frac{n_a - n_b}{n_a + n_b} E_{\mathrm{i}} \qquad \text{(normal incidence)} \tag{35.16}$$

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 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.

Figure **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.

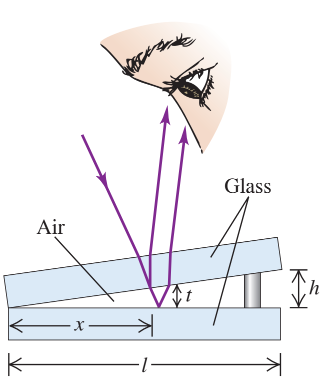
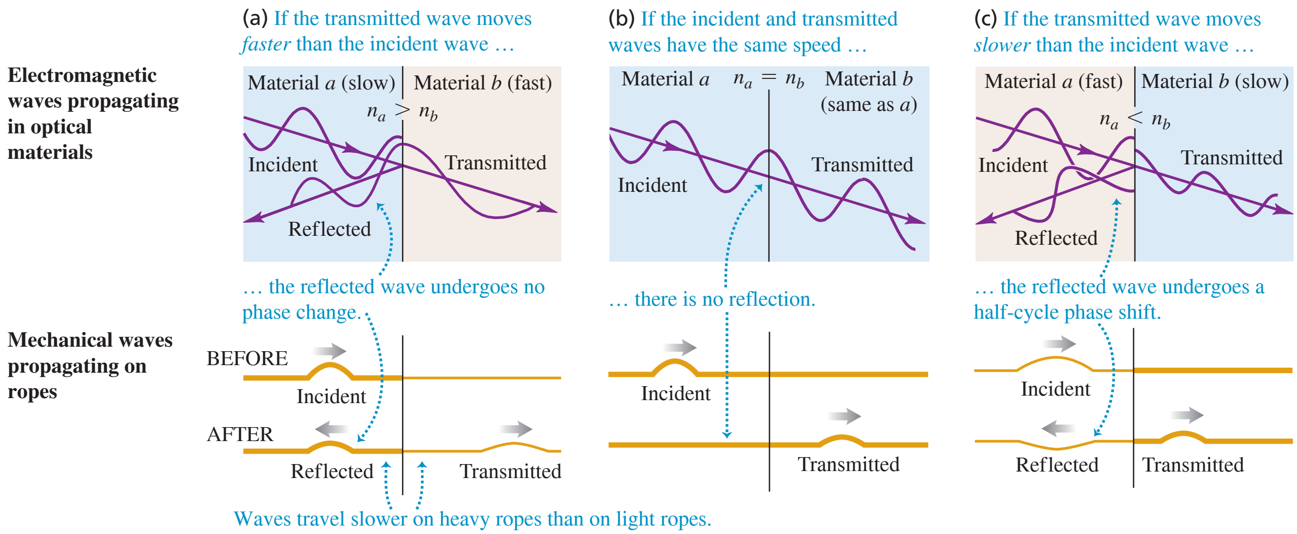


Figure **35.13** Upper figures: electromagnetic waves striking an interface between optical materials at normal incidence (shown as a small angle for clarity). Lower figures: mechanical wave pulses on ropes.



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 $$2t = m\lambda \quad (m = 0, 1, 2, \dots) \tag{35.17a}$$

(From thin film, no relative phase shift) Thickness of film, Wavelength

Destructive reflection $$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \dots) \tag{35.17b}$$

If *one* of the two waves has a half-cycle reflection phase shift, the conditions for constructive and destructive interference are reversed:

Constructive reflection $$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \dots) \tag{35.18a}$$

(From thin film, half-cycle phase shift) Thickness of film, Wavelength

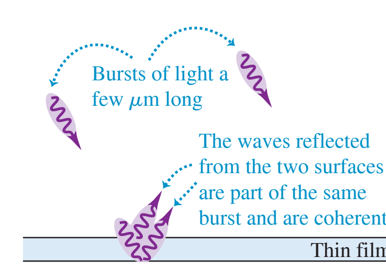
Destructive reflection $$2t = m\lambda \quad (m = 0, 1, 2, \dots) \tag{35.18b}$$

Thin and Thick Films

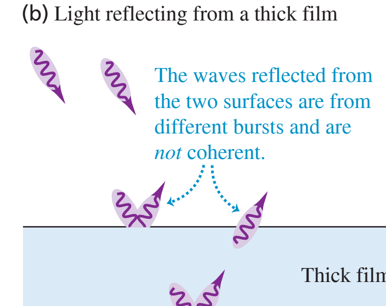
We have emphasized *thin* films in our discussion because of a principle we introduced in Section 35.1: In order for two waves to cause a steady interference pattern, the waves must be *coherent,* with a definite and constant phase relationship. The sun and light bulbs emit light in a stream of short 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,

Figure **35.14** (a) Light reflecting from a thin film produces a steady interference pattern, but (b) light reflecting from a thick film does not.

(a) Light reflecting from a thin film



(b) Light reflecting from a thick film



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).

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'll 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

WITH VARIATION PROBLEMS

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.

IDENTIFY and SET UP **Figure 35.15** depicts the situation. We'll consider only interference between the light reflected from the 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:

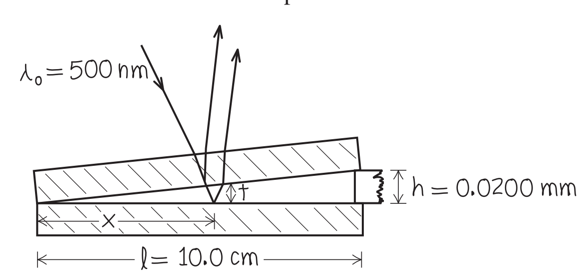
$$\frac{t}{x} = \frac{h}{l}$$

Combining this with Eq. (35.18b), we find

$$\frac{2xh}{l} = m\lambda_0$$

$$x = m\frac{l\lambda_0}{2h} = m\frac{(0.100\ \text{m})(500 \times 10^{-9}\ \text{m})}{(2)(0.0200 \times 10^{-3}\ \text{m})} = m(1.25\ \text{mm})$$

Figure **35.15** Our sketch for this problem.



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.)

KEYCONCEPT In thin-film interference, light waves reflected from the front surface of a thin film interfere with light waves reflected from the back surface. The phase difference between these waves depends on the thickness of the film, the wavelength of the light in the film, and whether the light undergoes a phase change on reflecting from each surface.

EXAMPLE 35.5 Thin-film interference II

WITH VARIATION PROBLEMS

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?

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 $\lambda = \lambda_0/n = (500\ \text{nm})/(1.33) = 376\ \text{nm}$. 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 $(0.0200\ \text{mm})/1.33 = 0.0150\ \text{mm}$? 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).]

KEYCONCEPT The wavelength of interest in thin-film interference is the wavelength of light in the material of which the film is composed. This equals the wavelength in vacuum divided by the index of refraction of the material.

EXAMPLE 35.6 Thin-film interference III

WITH VARIATION PROBLEMS

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?

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, $\lambda = \lambda_0/n = (500\ \text{nm})/1.50 = 333\ \text{nm}$. 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.

KEYCONCEPT When a light wave in one medium reflects at the surface of a second medium, it undergoes a half-cycle phase shift only if the speed of light is slower in the second medium than in the first; otherwise, there is no phase shift. If both reflected waves in thin-film interference undergo the same phase shift on reflection, there is no *relative* phase shift.

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.**

Figure 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.

(a) A convex lens in contact with a glass plane

(b) Newton's rings: circular interference fringes

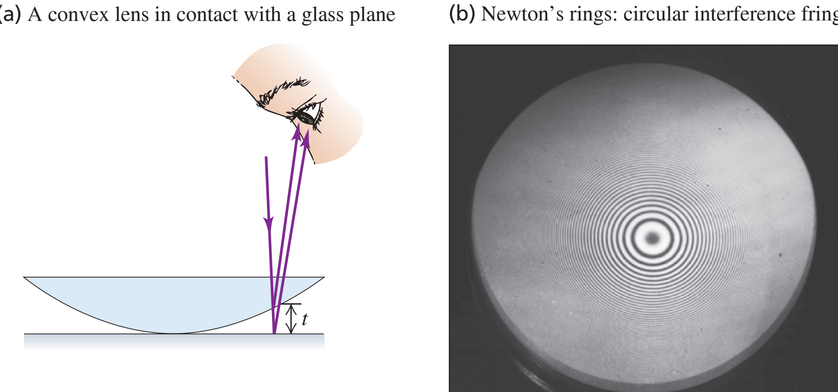
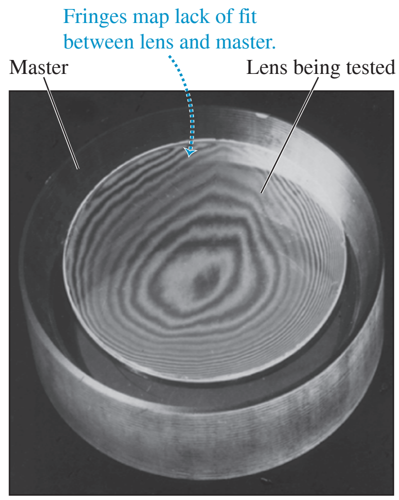


Figure **35.17** The surface of a telescope objective lens under inspection during manufacture.



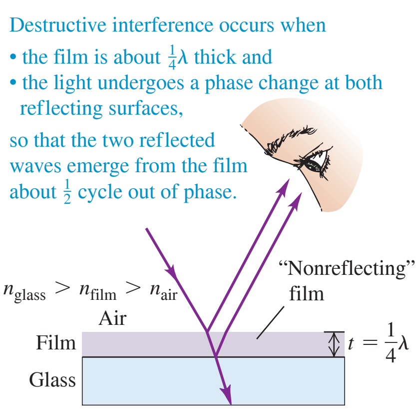
We can use interference fringes to compare the surfaces of two optical parts by placing the two in contact and observing the interference fringes. **Figure 35.17** is a photograph made during the grinding of a telescope objective lens. The lower, larger-diameter, thicker disk is the correctly shaped 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).

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.

Figure **35.18** A nonreflective coating has an index of refraction intermediate between those of glass and air.



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. 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% reflection (or, if desired, 100% transmission) for particular wavelengths. Some practical applications of reflective coatings are for color separation in television cameras and for infrared "heat reflectors" in motion-picture projectors, mirrored sunglasses, and astronauts' visors.

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 $n = 1.52$?

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 that isn't visible to the eye (see Section 32.1), so a lens designer need not worry about such enhanced reflection.

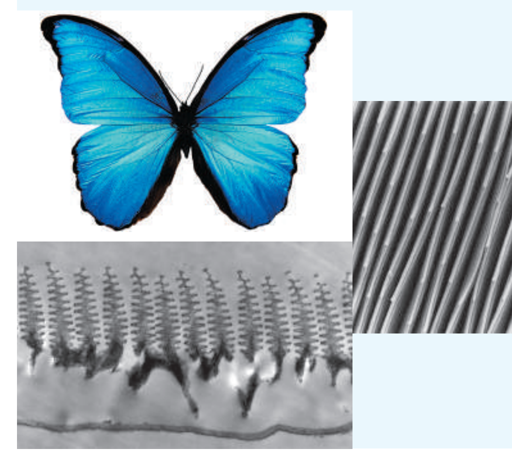
KEYCONCEPT A thin-film coating *minimizes* reflection for certain wavelengths by causing *destructive* interference of the reflected waves. The same coating *maximizes* reflection for other wavelengths by causing *constructive* interference of the reflected waves.

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.

ANSWER

(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 $t = \left(m + \frac{1}{2}\right)\lambda/2 = \left(m + \frac{1}{2}\right)(400\text{ nm})/2 = 100\text{ nm}, 300\text{ nm}, 500\text{ nm}, \ldots$.

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.)



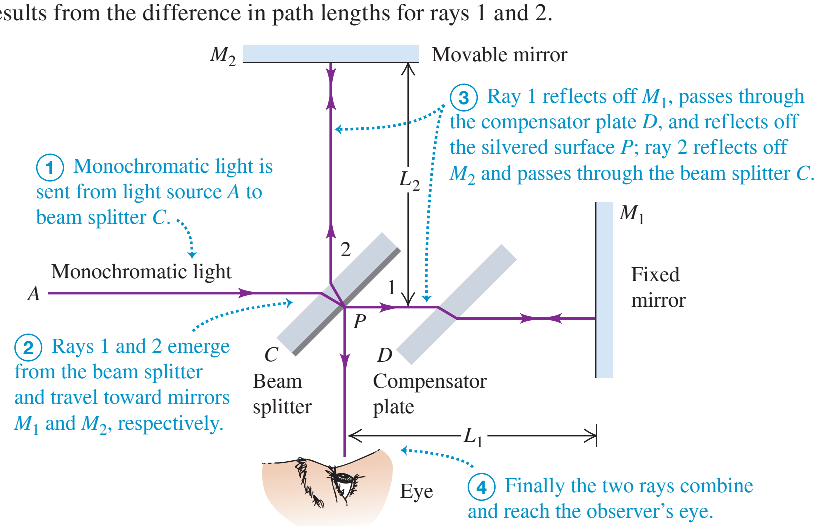
35.5 THE MICHELSON INTERFEROMETER

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.

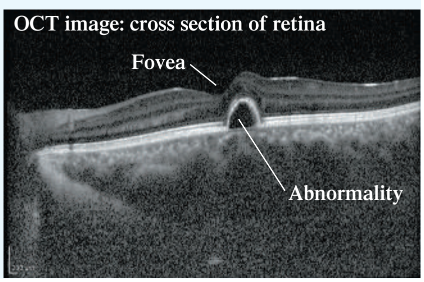
How a Michelson Interferometer Works

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.

Figure **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.



BIO APPLICATION Seeing Below the Surface with Interferometry Visible light can penetrate up to half a millimeter below the surface of biological tissue. However, only a very small fraction of the light is reflected back. To capture this light and image the subsurface features, a technique called *optical coherence tomography* (OCT) is used. OCT uses the same setup as a Michelson interferometer (Fig. 35.19), but with one of the mirrors replaced by the tissue to be studied. Only the light that reflects directly off the subsurface tissue gives constructive interference, which makes it possible to isolate that reflected light and form an image. OCT is widely used by ophthalmologists to get a cross-sectional view of the subsurface structure of the retina, which makes it easier to diagnose and monitor abnormalities of the eye.



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.

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'll 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.

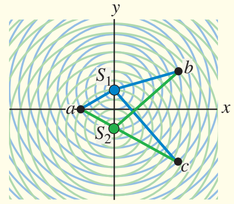
TEST YOUR UNDERSTANDING OF SECTION 35.5 You are observing the pattern of fringes in a Michelson interferometer like that shown in Fig. 35.19. If you change the index of refraction (but not the thickness) of the compensator plate, will the pattern change?

ANSWER

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.

CHAPTER 35 SUMMARY

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.

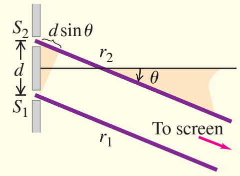


Two-source interference of light: When two sources are in phase, constructive interference occurs where the difference in path length from the two sources is zero or an integer number of wavelengths; destructive interference occurs where the path difference is a half-integer number of wavelengths. If 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.)

$$d\sin\theta = m\lambda \quad (m = 0, \pm1, \pm2, \ldots) \tag{35.4}$$
(constructive interference)

$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, \pm1, \pm2, \ldots) \tag{35.5}$$
(destructive interference)

$$y_m = R\frac{m\lambda}{d} \quad (m = 0, \pm1, \pm2, \ldots) \tag{35.6}$$
(bright fringes)

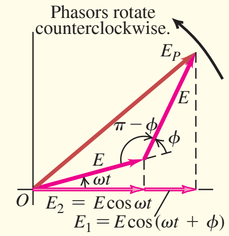


Intensity in interference patterns: When two sinusoidal waves with equal amplitude E and phase difference ϕ are superimposed, the resultant amplitude E_P and intensity I are given by Eqs. (35.7) and (35.10), respectively. If the two sources emit in phase, the phase difference ϕ at a point P (located 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.)

$$E_P = 2E\left|\cos\frac{\phi}{2}\right| \tag{35.7}$$

$$I = I_0\cos^2\frac{\phi}{2} \tag{35.10}$$

$$\phi = \frac{2\pi}{\lambda}(r_2 - r_1) = k(r_2 - r_1) \tag{35.11}$$



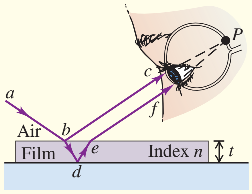
Interference in thin films: When light is reflected from both sides of a thin film of thickness t and no phase shift occurs at either surface, constructive interference between the reflected waves occurs when $2t$ is equal to an integer number of wavelengths. If a half-cycle phase shift occurs at one 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.)

$$2t = m\lambda \quad (m = 0, 1, 2, \ldots) \tag{35.17a}$$
(constructive reflection from thin film, no relative phase shift)

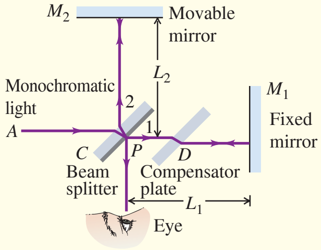
$$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \ldots) \tag{35.17b}$$
(destructive reflection from thin film, no relative phase shift)

$$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \ldots) \tag{35.18a}$$
(constructive reflection from thin film, half-cycle phase shift)

$$2t = m\lambda \quad (m = 0, 1, 2, \ldots) \tag{35.18b}$$
(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.



Chapter 35 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE √ARIATION PROBLEMS

Be sure to review Examples EXAMPLES 35.1 and 35.2 (Section 35.2) before attempting these problems.

VP35.2.1 In a two-slit interference experiment like that shown in Fig. 35.7, the slits are 0.180 mm apart and the screen is 1.50 m from the slits. The $m = 2$ bright fringe is 11.4 mm from the central bright fringe. Find (a) the wavelength of the light and (b) the distance from the central bright fringe to the $m = -3$ bright fringe.

VP35.2.2 Two identical vertical radio antennas are 0.810 km apart. Both emit in phase at a frequency of 1.05 MHz $= 1.05 \times 10^6$ Hz. Find the angles in the range from 0° and 90° at which the intensity measured several kilometers away from these antennas is (a) maximum and (b) minimum.

VP35.2.3 You shine red laser light of wavelength 685 nm on two closely spaced slits. This creates a pattern on a screen that is 2.10 m from the double slit. The $m = 3$ bright fringe in this pattern is 9.15 mm from the central bright fringe. (a) What is the distance between the slits? (b) What would be the distance between the central and $m = 3$ bright fringes if you used green laser light of wavelength 515 nm?

VP35.2.4 In a two-slit interference experiment the distance between the slits is 0.370 mm and the angle of the line from the slits to the $m = 4$ bright region on the screen is 0.407°. Find (a) the wavelength of the light used and (b) the angle of the line from the slits to the $m = 2$ dark region on the screen.

Be sure to review EXAMPLE 35.3 (Section 35.3) before attempting these problems.

VP35.3.1 Two identical radio antennas are 13.0 m apart and transmit in phase at frequency 59.3 MHz. At a distance of 0.850 km from the midpoint between the antennas in the direction $\theta = 0$, the intensity is $I_0 = 0.0330\ \mathrm{W/m^2}$. Find (a) the intensity at this same distance in the direction $\theta = 5.00°$ and (b) the smallest angle for which the intensity is $I_0/2$.

VP35.3.2 A two-slit interference experiment uses laser light of wavelength 655 nm. The slits are 0.230 mm apart and 1.75 m from the screen on which the interference pattern appears. The intensity of the light at the central bright fringe is $I_0 = 0.0520\ \mathrm{W/m^2}$. Find (a) the intensity at a point on the screen 6.50 mm from the central bright fringe and (b) the distance on the screen from the central bright fringe to the nearest point where the intensity is $I_0/4$.

VP35.3.3 You walk in a circle of radius 1.20 km around the midpoint of a pair of identical radio antennas that are 8.00 m apart. The intensity decreases continuously from $0.0540\ \mathrm{W/m^2}$ in the direction $\theta = 0°$ to $0.0303\ \mathrm{W/m^2}$ in the direction $\theta = 6.00°$. Find (a) the wavelength of the radio waves and (b) the intensity in the direction $\theta = 12.0°$.

VP35.3.4 A laser of unknown wavelength illuminates a pair of slits separated by 0.110 mm, producing an interference pattern on a wall 5.00 m from the slits. A point on the wall 10.5 mm from the central bright fringe is between that fringe and the $m = 1$ bright fringe, and the intensity there is 9.00% as great as at the central bright fringe. Find (a) the wavelength of the laser and (b) the distance from the central bright fringe to the nearest point on the wall where the intensity is 30.0% as great as at the central bright fringe.

Be sure to review EXAMPLES 35.4, 35.5, and 35.6 (Section 35.4) before attempting these problems.

VP35.6.1 Two identical microscope slides made of glass ($n = 1.50$) are stacked on top of each other. The slides are 8.00 cm long and are in contact at one end; at the other end they are separated by a piece of paper. You illuminate the slides from above with monochromatic light that has a wavelength in air of 565 nm. The spacing between the interference fringes seen by reflection is 1.30 mm. (a) Find the thickness of the piece of paper. (b) What would be the spacing between fringes if the space between the slides were filled with ethanol ($n = 1.36$) rather than air?

VP35.6.2 A thin, flat slide of glass ($n = 1.50$) lies on a table, and a second, identical glass slide is suspended 146 nm above the first one. The surfaces of the slides are parallel to each other. You illuminate the two slides from above. Find the wavelengths in air, if any, in the visible range (380 to 750 nm) for which there will be constructive reflection and destructive reflection if the space between the slides is filled with (a) air or (b) carbon tetrachloride ($n = 1.46$).

VP35.6.3 A thin, flat sheet of plastic ($n = 1.40$) lies on a table, and a second sheet of the same dimensions but made of flint glass ($n = 1.60$) is suspended 172 nm above the first sheet. The surfaces of the sheets are parallel to each other. You illuminate the two sheets from above. Find the wavelengths in air, if any, in the visible range (380 to 750 nm) for which there will be constructive reflection and destructive reflection if the space between the sheets is filled with (a) air or (b) ethanol ($n = 1.36$).

VP35.6.4 A bowl of water ($n = 1.33$) has a thin layer of grease ($n = 1.60$) floating on top of the water. When you illuminate the grease from above with monochromatic light, you get constructive reflection if the wavelength in air is 565 nm. Find (a) the smallest possible thickness of the grease layer and (b) the second smallest possible thickness.

BRIDGING PROBLEM Thin-Film Interference in an Oil Slick

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

•, ••, •••: 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

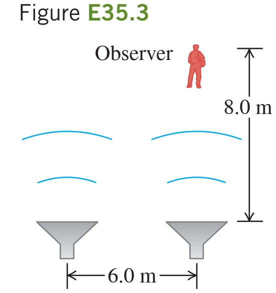
Section 35.1 Interference and Coherent Sources

35.1 •• 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?

35.2 • **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?

35.3 • 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. E35.3**. (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 **E35.3**



35.4 • 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?

35.5 •• Antenna B is 40.0 m to the right of antenna A. The two antennas emit electromagnetic waves that are in phase and have wavelength 7.00 m. (a) At how many points along the line connecting A and B is the interference constructive? (b) What is the smallest distance to the right of antenna A for which is there a point of constructive interference?

Section 35.2 Two-Source Interference of Light

35.6 •• Coherent light of wavelength 500 nm is incident on two very narrow and closely spaced slits. The interference pattern is observed on a very tall screen that is 2.00 m from the slits. Near the center of the screen the separation between two adjacent interference maxima is 3.53 cm. What is the distance on the screen between the $m = 49$ and $m = 50$ maxima?

35.7 • Young's experiment is performed with light from excited helium atoms $(\lambda = 502\ \text{nm})$. 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 •• Coherent light of frequency f travels in air and is incident on two narrow slits. The interference pattern is observed on a distant screen that is directly opposite the slits. The frequency of light f can be varied. For $f = 5.60 \times 10^{12}$ Hz there is an interference maximum for $\theta = 60.0°$. The next higher frequency for which there is an interference maximum at this angle is 7.47×10^{12} Hz. What is the separation d between the two slits?

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 • 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 $I_0/2$?

35.19 • 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 0.0720 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?

35.21 • Consider two antennas separated by 9.00 m that radiate in phase at 120 MHz, as described in Exercise 35.1. 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?

Section 35.4 Interference in Thin Films

35.22 •• Two identical horizontal sheets of glass have a thin film of air of thickness t between them. The glass has refractive index 1.40. The thickness t of the air layer can be varied. Light with wavelength λ in air is at normal incidence onto the top of the air film. There is constructive interference between the light reflected at the top and bottom surfaces of the air film when its thickness is 650 nm. For the same wavelength of light the next larger thickness for which there is constructive interference is 910 nm. (a) What is the wavelength λ of the light when it is traveling in air? (b) What is the smallest thickness t of the air film for which there is constructive interference for this wavelength of light?

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 •• 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.

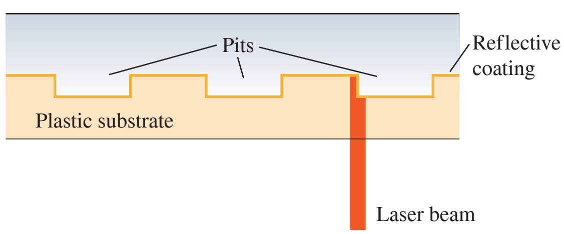
35.26 • A plastic film with index of refraction 1.70 is applied to the surface of a car window to increase the reflectivity and thus to keep the car's interior cooler. The window glass has index of refraction 1.52. (a) What minimum thickness is required if light of wavelength 550 nm in air reflected 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?

35.27 • 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.

35.28 •• 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?

35.29 •• **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. E35.29**). 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 **E35.29**



Section 35.5 The Michelson Interferometer

35.30 • 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?

35.31 • 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?

PROBLEMS

35.32 •• The LIGO experiment, which historically detected gravitational waves for the first time in September 2015, uses a pair of highly sensitive Michelson interferometers. These have arms that are 4.00 km long and use powerful Nd:Yag lasers with 1064 nm wavelength. The beams traverse the arms both ways 280 times before recombining, which effectively lengthens the arm length to 1120 km. The devices are tuned so that the beams destructively interfere when they recombine if no gravitational wave is present. (a) The beam has a power of 100 kW, concentrated into an area of a square centimeter. Calculate the amplitude of the electric field in the beam. (b) LIGO can detect a gravitational wave that temporarily lengthens one arm by the minuscule amount of 10^{-18} m! When this happens, the beams combine with a phase difference of $\pi + \delta$. Estimate the shift δ in radians. Note that the phase difference accumulates during both traversals of each round trip. (c) Use Eq. (35.7) to estimate the sensitivity of the photodetector in terms of the minimal electric field strength needed to detect a gravitational wave.

35.33 •• 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?

35.34 •• 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.

35.35 • 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?

35.36 •• 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?

35.37 •• 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?

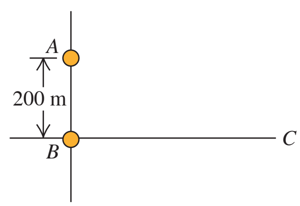
35.38 •• A typical red laser pointer has a wavelength of 650 nm. Suppose we wanted to test the wave nature of light by carefully cutting two parallel slits in a dark plastic sheet. We would then shine the laser through the slits onto a wall located 1 m beyond the sheet. (a) Determine the slit spacing needed so that the bright spots on the wall would be discernible with 1 cm spacing. (b) Is this a feasible "home experiment"? Is it possible to cut slits with that separation using typical household tools? (c) Suppose we wanted to test the wave nature of sound in a similar manner, by placing two small speakers driven in-phase near each other, separated by 40 cm, both facing a wall 2 m distant. If we used a 1.0 kHz tone, determine the distance between points along the wall that would exhibit enhanced sound. (d) Suppose we wanted the sound-enhanced points to be separated by only 1.75 m to render this experiment feasible. Estimate an audible frequency f and use it to determine a speaker separation distance d that would accomplish this. (e) Is this a feasible "home experiment"?

35.39 ••• 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?

35.40 •• 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.)

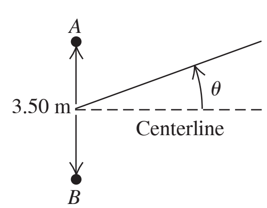
35.41 •• Two radio antennas radiating in phase are located at points A and B, 200 m apart (**Fig. P35.41**). The radio waves have a frequency of 5.80 MHz. A radio receiver is moved out from point B along a line perpendicular to the line connecting A and B (line BC shown in Fig. P35.41). At what distances 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 **P35.41**



35.42 •• 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. P35.42**) 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 **P35.42**



35.43 •• 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.)

35.44 ••• 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?

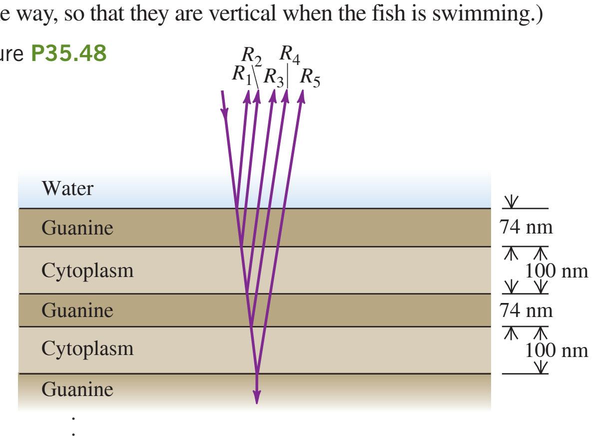
35.45 •• White light reflects at normal incidence from the top and bottom surfaces of a glass plate ($n = 1.52$). There is air above and below the plate. Constructive interference is observed for light whose wavelength in air is 477.0 nm. What is the thickness of the plate if the next longer wavelength for which there is constructive interference is 540.6 nm?

35.46 •• 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?

35.47 •• 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?

35.48 •• 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. P35.48**). In one typical platelet, the guanine 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. P35.48, 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.)

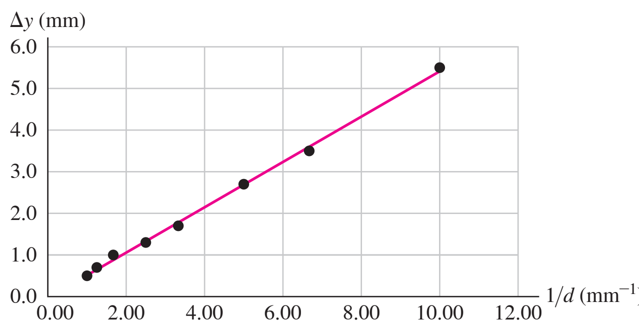
Figure **P35.48**



35.49 •• 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?

35.50 •• 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. P35.50**, 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. P35.50 to calculate λ.

Figure **P35.50**

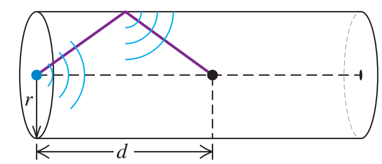


35.51 •• DATA Short-wave radio antennas A and B are connected to the same transmitter and emit coherent waves in phase and with the same frequency f. You must determine the value of f and the placement of the antennas that produce a maximum intensity through constructive interference at a receiving antenna 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?

35.52 •• 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?

35.53 •• CP An acoustic waveguide consists of a long cylindrical tube with radius r designed to channel sound waves, as shown in **Fig. P35.53**. A tone with frequency f is emitted from a small source at the center of one end of this tube. Depending on the radius of the tube and the frequency of the tone, pressure nodes can develop along the tube axis where rays reflected from the periphery constructively interfere with direct rays. (a) For sound waves with wavelength λ, what is the minimum tube radius for which at least one such node exists? (b) The tube has radius 25.0 cm and the temperature is 20°C. If the tone has frequency 2.50 kHz, how many nodes exist? (c) At what distance d are these nodes located? (d) If the tube were filled with helium rather than air, how many nodes would exist, and at what value of d?

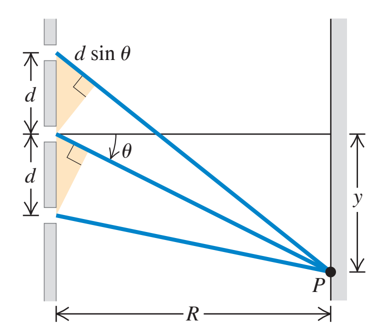
Figure **P35.53**



CHALLENGE PROBLEMS

35.54 ••• Three-Slit Interference. Monochromatic light with wavelength λ is incident on a screen with three narrow slits with separation d, as shown in **Fig. P35.54**. Light from the middle slit reaches point P with electric field $E\cos(\omega t)$. From the small-angle approximation, light from the upper and lower slits reaches point P with electric fields $E\cos(\omega t + \phi)$ and $E\cos(\omega t - \phi)$, respectively, where $\phi = (2\pi d \sin\theta)/\lambda$ is the phase lag and phase lead associated with the different path lengths. (a) Using either a phasor analysis similar to Fig. 35.9 or trigonometric identities, determine the electric-field amplitude E_P associated with the net field at point P, in terms of ϕ. (b) Determine the intensity I at point P in terms of the maximum net intensity I_0 and the phase angle ϕ. (c) There are two sets of relative maxima: one with intensity I_0 when $\phi = 2\pi m$, where m is an integer, and another with a smaller intensity at other values of ϕ. What values of ϕ exhibit the "lesser" maxima? (d) What is the intensity at the lesser maxima, in terms of I_0? (e) What values of ϕ correspond to the dark fringes closest to the center? (f) If the incident light has wavelength 650 nm, the slits are separated by $d = 0.200$ mm, and the distance to the far screen is $R = 1.00$ m, what is the distance from the central maximum to the first lesser maximum? (g) What is the distance from the central maximum to the closest absolute maximum?

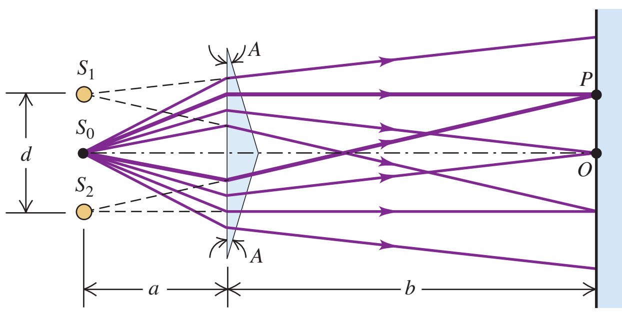
Figure **P35.54**



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 2.50×10^{-5}/C°. The coefficient of linear expansion of the glass is 5.00×10^{-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**



MCAT-STYLE 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 ?

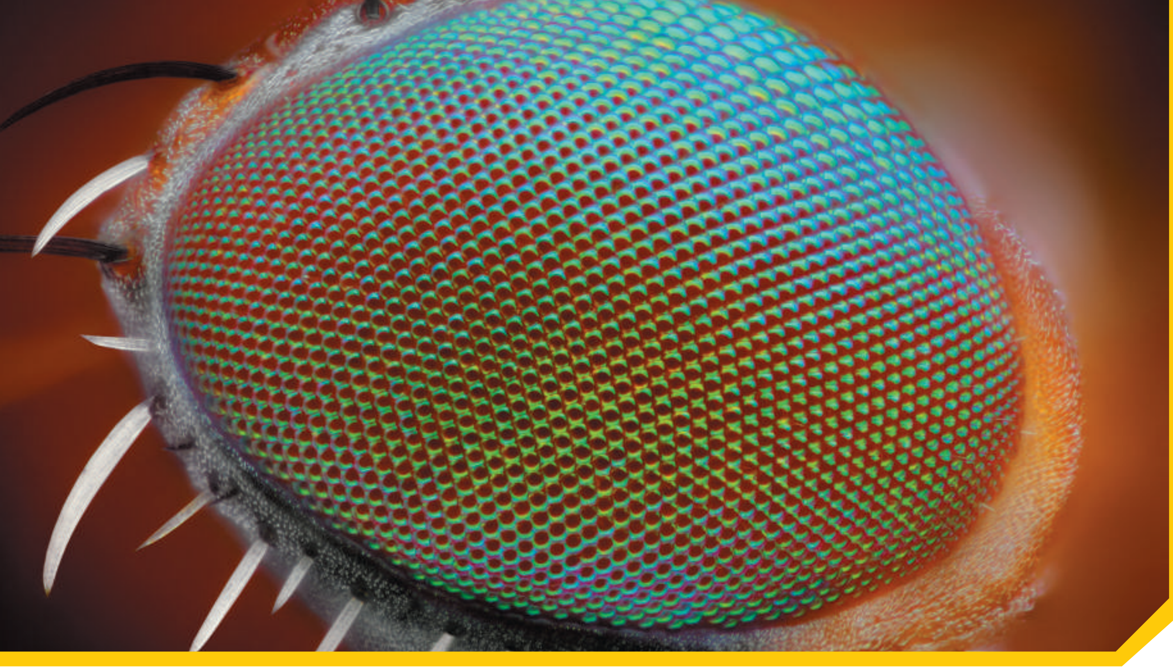
(v) The colors appear as a result of constructive interference between light waves reflected from the upper and lower surfaces of the oil film. The wavelength of light for which the most constructive interference occurs at a point, and hence the color that appears the brightest at that point, depends 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.)

Key Example VARIATION Problems

VP35.2.1 (a) 684 nm (b) 17.1 mm
VP35.2.2 (a) 0°, 20.7°, 44.9° (b) 10.2°, 31.9°, 61.9°
VP35.2.3 (a) 0.472 mm (b) 6.88 mm
VP35.2.4 (a) 657 nm (b) 0.254°
VP35.3.1 (a) 0.0192 W/m^2 (b) 5.58°
VP35.3.2 (a) 0.0173 W/m^2 (b) 1.66 mm
VP35.3.3 (a) 3.63 m (b) 9.14×10^{-4} W/m^2
VP35.3.4 (a) 573 nm (b) 8.22 mm
VP35.6.1 (a) 0.0174 mm (b) 0.956 mm
VP35.6.2 (a) constructive: 584 nm; destructive: none (b) constructive: none; destructive: 426 nm
VP35.6.3 (a) constructive: none; destructive: 688 nm (b) constructive: 468 nm; destructive: none
VP35.6.4 (a) 88.3 nm (b) 265 nm

Bridging Problem

(a) 441 nm; violet
(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

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.

LEARNING OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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.

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**, next page) 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 36.1 A point source of light illuminates a straightedge.

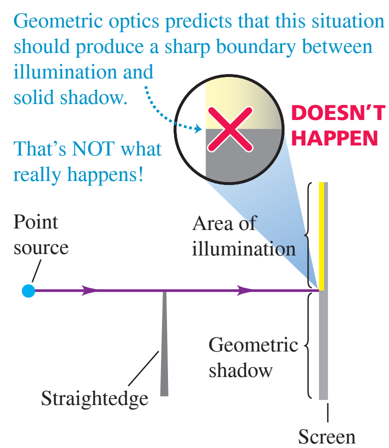


Figure 36.2 shows an example of diffraction. The photograph in Fig. 36.2a was made by placing a razor blade halfway between a pinhole, illuminated by monochromatic light, and a photographic film. The film recorded the shadow cast by the blade. Figure 36.2b is an enlargement of a region near the shadow 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.

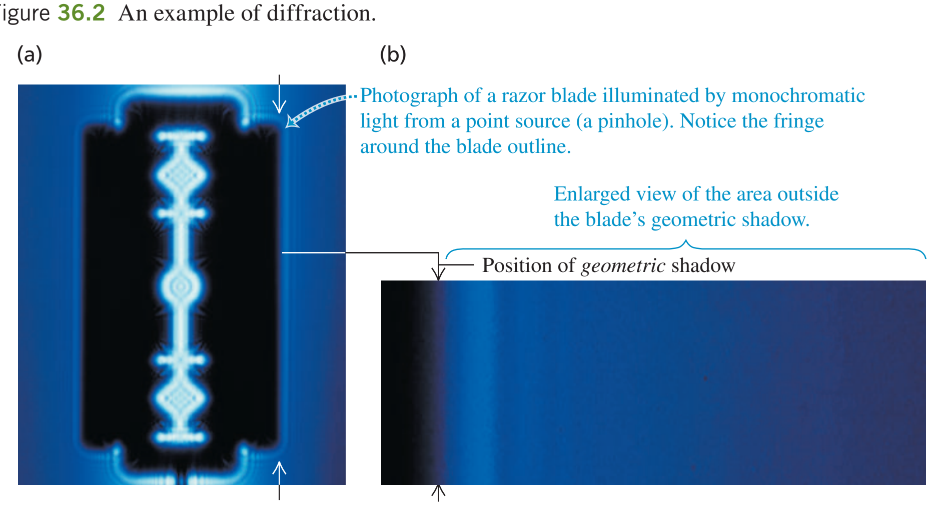
Diffraction and Huygens's Principle

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'll restrict the following discussion 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.

Figure 36.2 An example of diffraction.



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?

ANSWER

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 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.

Single-Slit Diffraction: Locating the Dark Fringes

Figure 36.4 (next page) 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.

Figure **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.

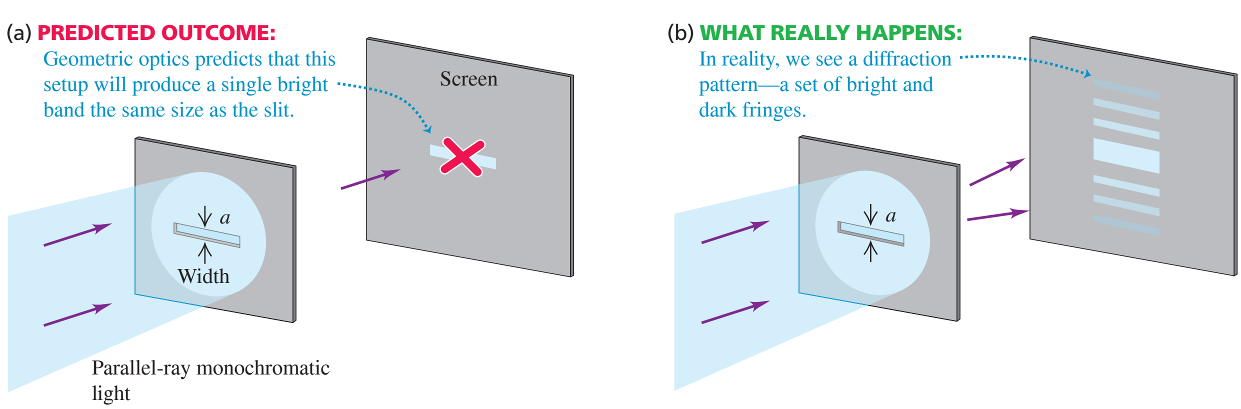
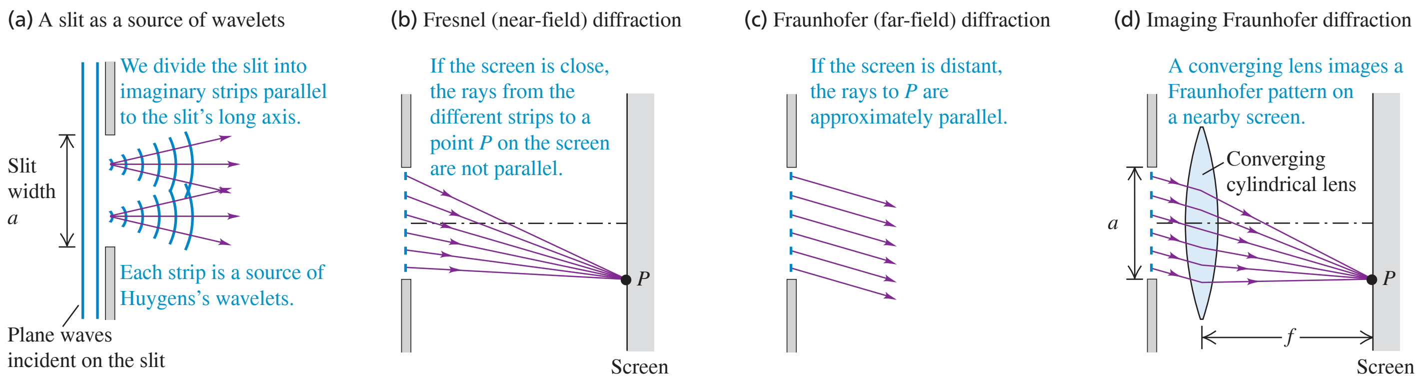


Figure **36.4** Diffraction by a single rectangular slit. The long sides of the slit are perpendicular to the figure.

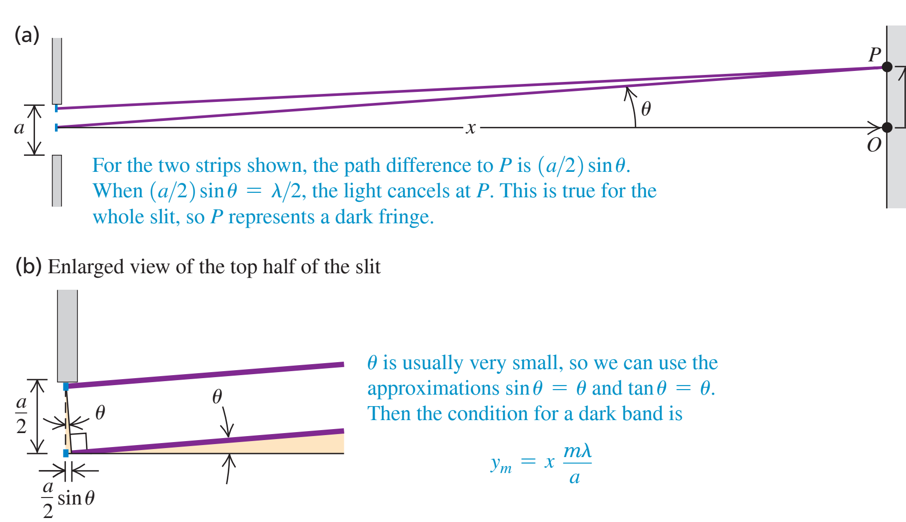


In Fig. 36.4b a screen is placed to the right of the slit. We can calculate the resultant intensity at a point P on the screen by adding the contributions from the individual wavelets, taking proper account of their various phases and amplitudes. It's easiest to do this calculation if we assume that 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 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

Figure **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.



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:

$$\sin\theta = \frac{m\lambda}{a} \quad (m = \pm 1, \pm 2, \pm 3, \dots) \tag{36.2}$$

Angle of line from center of slit to mth dark fringe on screen

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 $500\text{ nm} = 5 \times 10^{-7}\text{ m}$. This is often much smaller than the slit width a; a typical slit width is $10^{-2}\text{ cm} = 10^{-4}\text{ m}$. 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

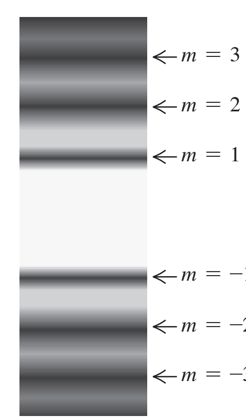
$$\theta = \frac{m\lambda}{a} \qquad (m = \pm 1, \pm 2, \pm 3, \dots) \qquad \text{(for small angles } \theta \text{ in radians)}$$

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.

Figure **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. Finally, note that Eq. (36.2) gives the angle θ 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. Be careful!

EXAMPLE 36.1 Single-slit diffraction

WITH VARIATION PROBLEMS

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**, next page). How wide is the slit?

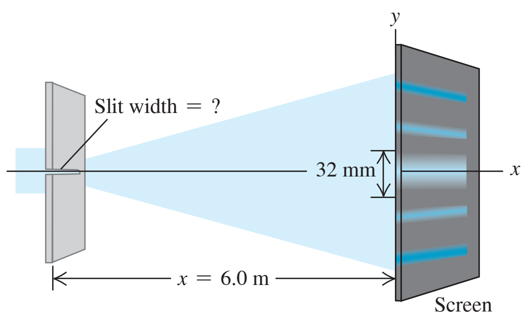
IDENTIFY and SET UP This problem involves the relationship between the positions of dark fringes in a single-slit diffraction pattern and the slit width a (our target variable). The distances between fringes on the screen are much smaller than the slit-to-screen distance, so the angle θ shown in Fig. 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 $y_1 = (32\text{ mm})/2 = 16\text{ mm}$. Solving Eq. (36.3) for a, we find

$$a = \frac{x\lambda}{y_1} = \frac{(6.0\text{ m})(633 \times 10^{-9}\text{ m})}{16 \times 10^{-3}\text{ m}} = 2.4 \times 10^{-4}\text{ m} = 0.24\text{ mm}$$

Continued

Figure 36.7 A single-slit diffraction experiment.

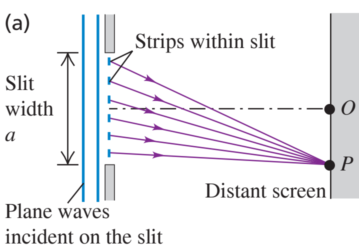


EVALUATE The angle θ is small only if the wavelength is small compared to the slit width. Since $\lambda = 633\text{ nm} = 6.33 \times 10^{-7}\text{ m}$ and we have found $a = 0.24\text{ mm} = 2.4 \times 10^{-4}\text{ m}$, our result is consistent with this: The wavelength is $(6.33 \times 10^{-7}\text{ m})/(2.4 \times 10^{-4}\text{ m}) = 0.0026$ as large as the slit width. Can you show that the distance between the *second* minima on either side is $2(32\text{ mm}) = 64\text{ mm}$, and so on?

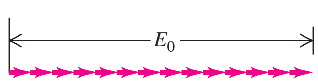
KEYCONCEPT In single-slit diffraction, wavelets that emerge from different parts of the slit interfere and produce a pattern of bright and dark fringes. Equations (36.2) and (36.3) give the positions of the dark fringes where destructive interference occurs.

Figure 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.

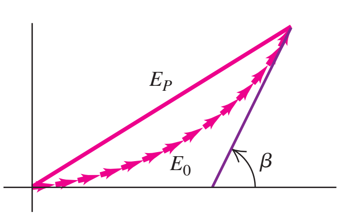
(a)



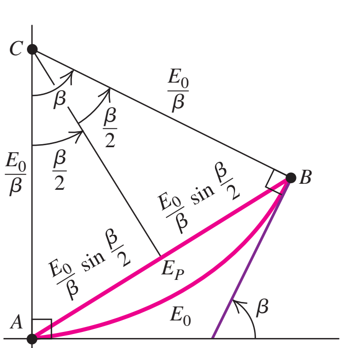
(b) At the center of the diffraction pattern (point O), the phasors from all strips within the slit are in phase.



(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



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.

ANSWER

(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 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**). 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.

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

$$E_P = E_0\frac{\sin(\beta/2)}{\beta/2} \qquad \text{(amplitude in single-slit diffraction)} \tag{36.4}$$

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

$$I = I_0\left[\frac{\sin(\beta/2)}{\beta/2}\right]^2 \qquad \text{(intensity in single-slit diffraction)} \tag{36.5}$$

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

$$\beta = \frac{2\pi}{\lambda}a\sin\theta \tag{36.6}$$

and Eq. (36.5) becomes

Angle of line from center of slit to position on screen

Intensity in single-slit diffraction

$$I = I_0\left\{\frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda}\right\}^2 \tag{36.7}$$

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

$$\frac{a\sin\theta}{\lambda} = m \qquad (m = \pm 1, \pm 2, \dots)$$

$$\sin\theta = \frac{m\lambda}{a} \qquad (m = \pm 1, \pm 2, \dots) \tag{36.8}$$

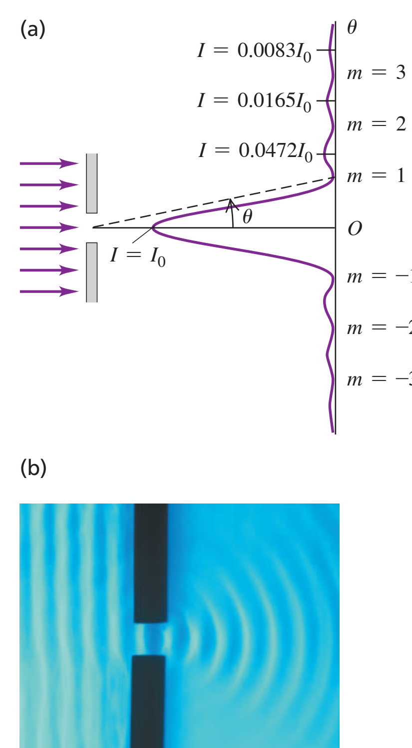
This agrees with our previous result, Eq. (36.2). Note again that $\beta = 0$ (corresponding to $\theta = 0$) is *not* a minimum. Equation (36.5) is indeterminate at $\beta = 0$, but we can evaluate the limit as $\beta \to 0$ by using L'Hôpital's rule. We find that at $\beta = 0$, $I = I_0$, as we should 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,

$$\beta \approx \pm(2m + 1)\pi \qquad (m = 0, 1, 2, \dots) \tag{36.9}$$

Figure **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.



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

$$0.0450 I_0 \qquad 0.0162 I_0 \qquad 0.0083 I_0$$

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

$$0.0472 I_0 \qquad 0.0165 I_0 \qquad 0.0083 I_0 \qquad \ldots$$

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.

Width of the Single-Slit Pattern

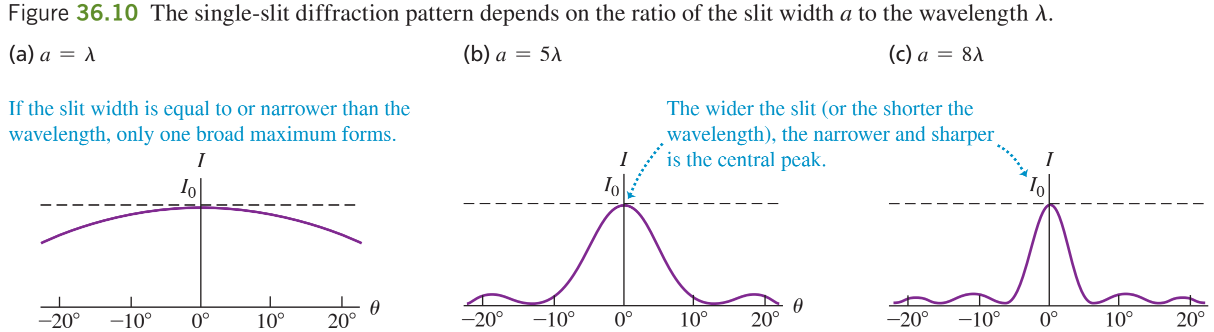
For small angles the angular spread of the diffraction pattern is inversely proportional to the ratio of the slit width a to the wavelength λ. **Figure 36.10** shows graphs of intensity I as a function of the angle θ for three values of the ratio a/λ.

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.

Figure **36.10** The single-slit diffraction pattern depends on the ratio of the slit width a to the wavelength λ.



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.

Figure 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

WITH VARIATION PROBLEMS

(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?

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 $a = (86)(550\text{ nm}) = 4.7 \times 10^{-5}\text{ m} = 0.047$ 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 $m = \pm 3$.

KEYCONCEPT The intensity at a given point in a single-slit diffraction pattern depends on the phase difference β between the wavelets that arrive at that point from the two edges of the slit [Eqs. (36.6) and (36.7)]. The intensity is greatest at the center of the pattern where $\beta = 0$; the intensity is zero at points where $\beta = \pm 2\pi, \pm 4\pi, \pm 6\pi, \ldots$.

EXAMPLE 36.3 Single-slit diffraction: Intensity II

WITH VARIATION PROBLEMS

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?

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, from Eq. (36.7),

$$\frac{\pi a\sin\theta}{\lambda} = \frac{\pi(2.4 \times 10^{-4}\text{ m})(5.0 \times 10^{-4})}{6.33 \times 10^{-7}\text{ m}} = 0.60$$

$$I = I_0\left(\frac{\sin 0.60}{0.60}\right)^2 = 0.89I_0$$

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.

KEYCONCEPT When light of wavelength λ passes through a single slit of width a, you can write the intensity of the diffracted light at angle θ [Eq. (36.7)] in terms of the ratio $(\pi a\sin\theta)/\lambda$, which is one-half of the phase difference β between wavelets coming from the two edges of the slit. The intensity is greatest at the center of the pattern ($\theta = 0$); the intensity is zero where $(\pi a\sin\theta)/\lambda = \pm\pi, \pm 2\pi, \pm 3\pi, \ldots$.

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.

ANSWER

(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 mm $= 1.00 \times 10^{-5}$ m, so this condition is satisfied for (ii) ($\lambda = 10.6\ \mu\text{m} = 1.06 \times 10^{-5}$ m) and (iii) ($\lambda = 1.00\text{ mm} = 1.00 \times 10^{-3}$ m) but not for (i) ($\lambda = 500\text{ nm} = 5.00 \times 10^{-7}$ m) or (iv) ($\lambda = 50.0\text{ nm} = 5.00 \times 10^{-8}$ m).

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.

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):

$$I = I_0 \cos^2 \frac{\phi}{2} \left[\frac{\sin(\beta/2)}{\beta/2} \right]^2 \qquad \text{(two slits of finite width)} \qquad (36.12)$$

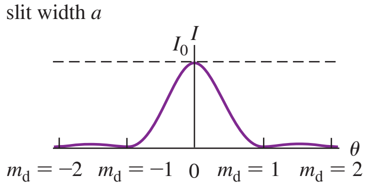
where, as before,

$$\phi = \frac{2\pi d}{\lambda} \sin\theta \qquad \beta = \frac{2\pi a}{\lambda} \sin\theta$$

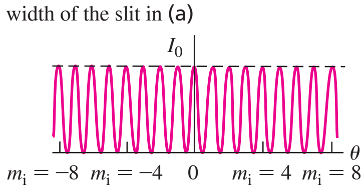
In Fig. 36.12c, every fourth interference maximum at the sides is *missing* because these interference maxima ($m_i = \pm 4, \pm 8, \ldots$) coincide with diffraction minima ($m_d = \pm 1, \pm 2, \ldots$). This can also be seen in Fig. 36.12d, which is a photograph of an actual pattern with $d = 4a$. You should be able to convince yourself that there will be "missing" maxima whenever d is an integer multiple of a.

Figure 36.12 Finding the intensity pattern for two slits of finite width.

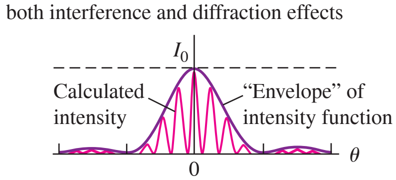
(a) Single-slit diffraction pattern for a slit width a



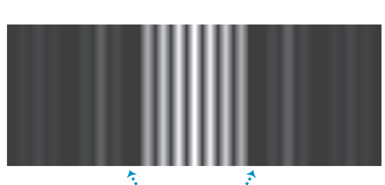
(b) Two-slit interference pattern for narrow slits whose separation d is four times the width of the slit in (a)



(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.

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.

Several Slits

Next let's consider patterns produced by *several* very narrow slits. As we'll 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:

$$d \sin\theta = m\lambda \qquad (m = 0, \pm 1, \pm 2, \dots)$$

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 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 **36.13** Multiple-slit diffraction. Here a lens is used to give a Fraunhofer pattern on a nearby screen, as in Fig. 36.4d.

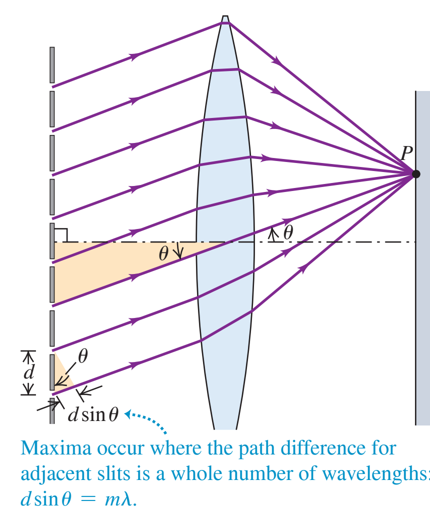


Figure **36.14** Phasor diagrams for light passing through eight narrow slits. Intensity maxima occur when the phase difference $\phi = 0, 2\pi, 4\pi, \dots$. Between the maxima at $\phi = 0$ and $\phi = 2\pi$ are seven minima, corresponding to $\phi = \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2$, and $7\pi/4$. Can you draw phasor diagrams for the other minima?

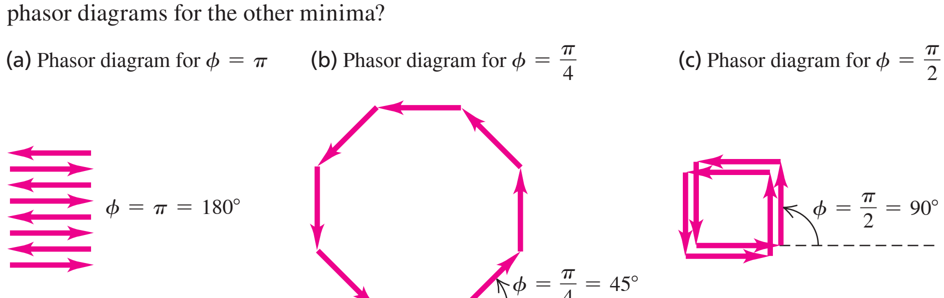
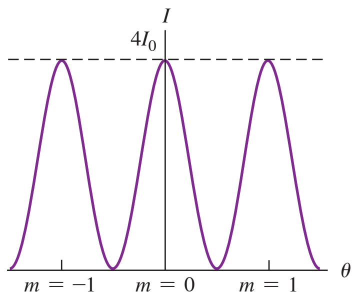
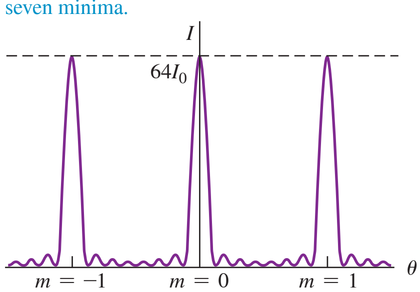


Figure 36.15 Interference patterns for N equally spaced, very narrow slits. (a) Two slits. (b) Eight slits. (c) Sixteen slits. The vertical scales are different for each graph; the maximum intensity is I_0 for a single slit and $N^2 I_0$ for N slits. The width of each peak is proportional to $1/N$.

(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.

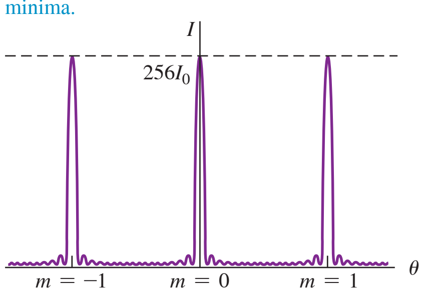


Figure 36.15b shows the result of a detailed calculation of the eight-slit pattern. The large maxima, called *principal maxima,* are in the same positions as for the two-slit pattern of Fig. 36.15a but are much narrower. If the phase difference ϕ between adjacent slits is slightly different from a multiple 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 integer multiple of $2\pi/N$ (except when ϕ is an integer 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?

ANSWER

❙ **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 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'll 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.*

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:

$$\underbrace{d}_{\text{Distance between slits}}\sin\underbrace{\theta}_{\text{Angle of line from center of slit array to } m\text{th bright region on screen}} = m\underbrace{\lambda}_{\text{Wavelength}} \quad (m = 0, \pm 1, \pm 2, \ldots) \qquad (36.13)$$

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 $\frac{1}{1000}\text{ mm} = 1000\text{ nm}$.

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 integer 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 $\mu\text{m} = 740\text{ nm}$. 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.

Figure **36.16** A portion of a transmission diffraction grating. The separation between the centers of adjacent slits is d.

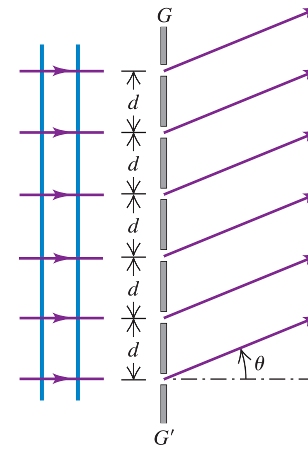


Figure **36.17** Microscopic pits on the surface of this DVD act as a reflection grating, splitting white light into its component colors.

EXAMPLE 36.4 Width of a grating spectrum

WITH VARIATION PROBLEMS

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?

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).

Continued

(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.

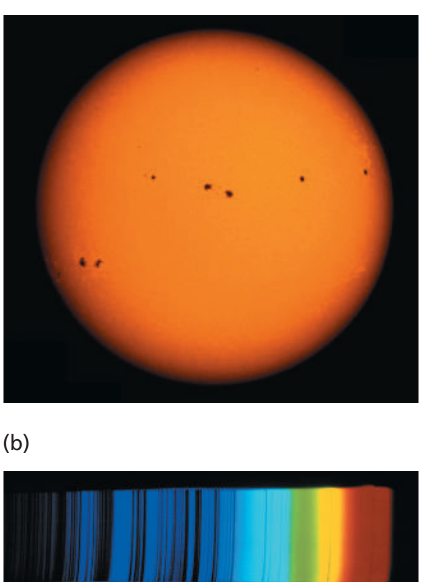
EVALUATE The fundamental reason the first-order and second-order visible spectra don't overlap is that the human eye is sensitive to only a narrow range of wavelengths. Can you show that if the eye could detect wavelengths from 380 nm to 900 nm (in the near-infrared range), the first and second orders *would* overlap?

KEYCONCEPT In the interference pattern produced by a transmission or reflection grating, the intensity maxima occur at angles that are determined by the grating spacing and the wavelength of the light used. These angles are the same as in the interference pattern produced in the double-slit experiment; the difference is that the maxima are much narrower for a grating.

Figure 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)



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 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.

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

$$R = \frac{\lambda}{\Delta\lambda} \qquad \text{(chromatic resolving power)} \qquad (36.14)$$

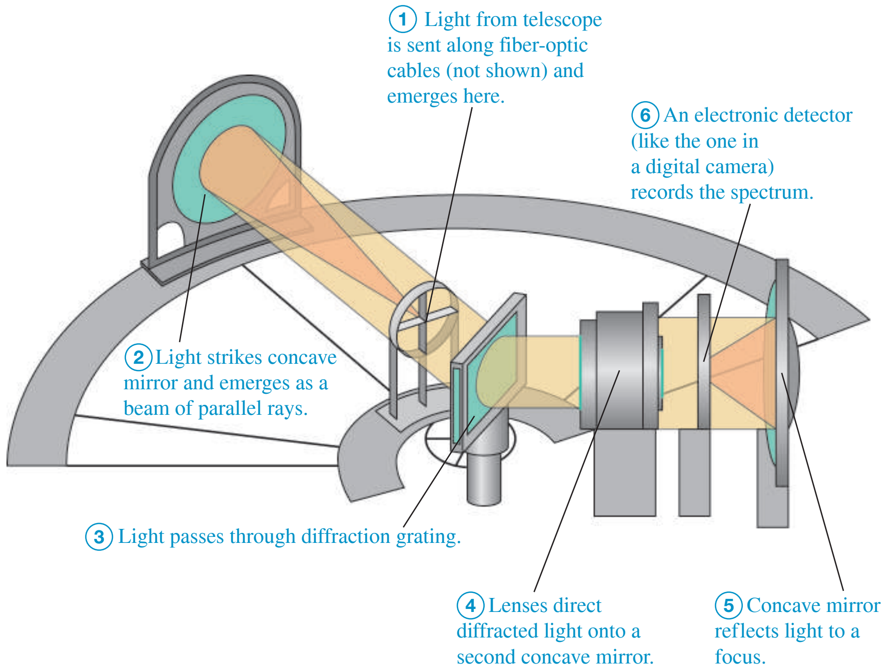


Figure **36.19** A schematic diagram of a diffraction-grating spectrograph for use in astronomy. Note that the light does not strike the grating normal to its surface, so the intensity maxima are given by a somewhat different expression than Eq. (36.13).

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 with table salt in it on a gas range. If the water boils over onto the flame, dissolved sodium from the 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

$$\frac{2\pi}{N} = \frac{2\pi d \cos\theta \, d\theta}{\lambda} \quad \text{or} \quad d\cos\theta \, d\theta = \frac{\lambda}{N}$$

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

$$\frac{\lambda}{N} = m \, d\lambda \quad \text{and} \quad \frac{\lambda}{d\lambda} = Nm$$

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.

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 $d\phi$!

If $\Delta\lambda$ is small, we can replace $d\lambda$ by $\Delta\lambda$, and the resolving power R is

$$R = \frac{\lambda}{\Delta\lambda} = Nm \tag{36.15}$$

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.

ANSWER

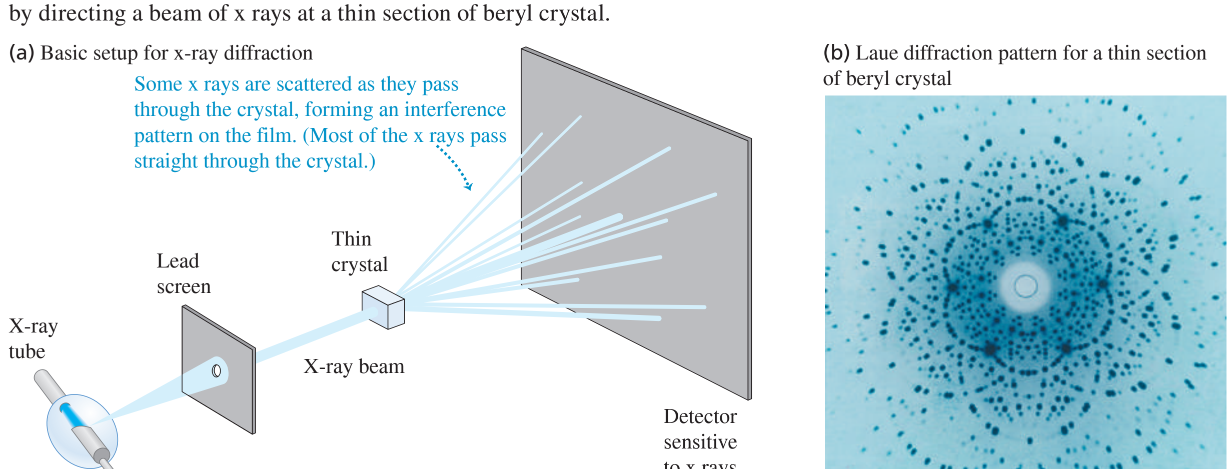
(i) As described in the text, the resolving power needed is $R = Nm = 1000$. In the first order $(m = 1)$ we need $N = 1000$ slits, but in the fourth order $(m = 4)$ we need only $N = R/m = 1000/4 = 250$ slits. (These numbers are only approximate because of the arbitrary nature of our criterion for resolution and because real gratings always have slight imperfections in the shapes and spacings of the slits.)

36.6 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 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.

Figure 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 beryl crystal.



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:

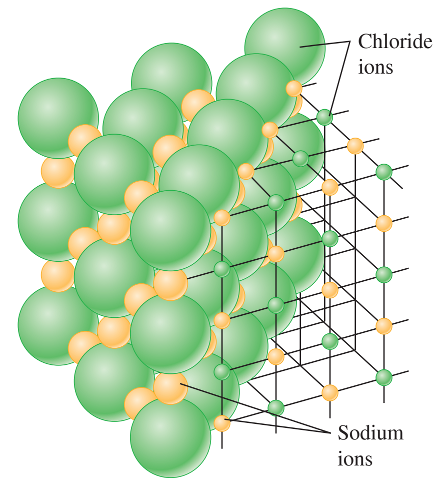
Distance between adjacent rows in array; Wavelength

$$2d\sin\theta = m\lambda \quad (m = 1, 2, 3, \dots) \tag{36.16}$$

Angle of line from surface of array to mth bright region on screen

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.

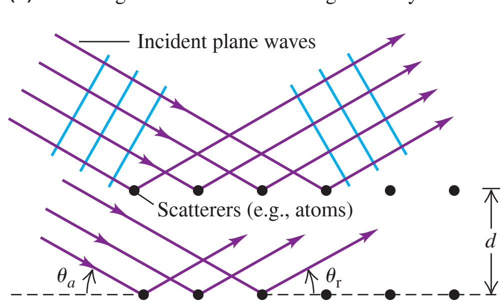
Figure 36.21 Model of the arrangement of ions in a crystal of NaCl (table salt). The spacing of adjacent atoms is 0.282 nm. (The electron clouds of the atoms actually overlap slightly.)



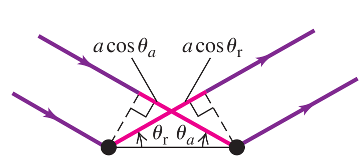
CAUTION Scattering from an array In Eq. (36.16) the angle θ is measured with respect to the *surface* of the crystal rather than with respect to the *normal* to the plane of an array of slits or a grating. Also, note that the path difference in Eq. (36.16) is $2d\sin\theta$, not $d\sin\theta$ as in Eq. (36.13) for a diffraction grating.

Figure 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.

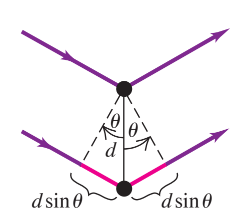
(a) Scattering of waves from a rectangular array



(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 integer number of wavelengths, as in Eq. (36.16).



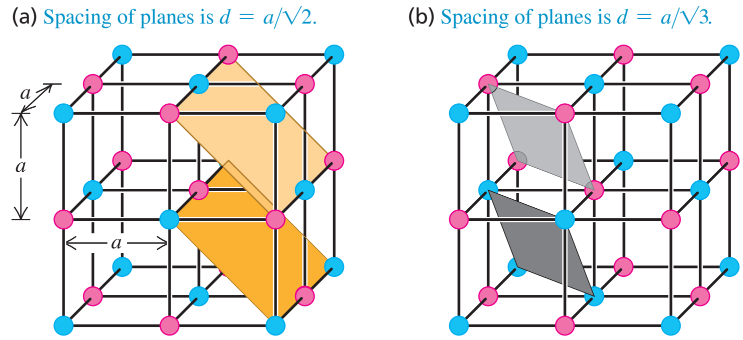


Figure 36.23 A cubic crystal and two different families of crystal planes. There are also three sets of planes parallel to the cube faces, with spacing a.

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).

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.**

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.

Figure 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.

EXAMPLE 36.5 X-ray diffraction

WITH VARIATION PROBLEMS

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 larger angles of incidence?

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?

KEYCONCEPT In x-ray diffraction, waves that pass through or reflect from a crystal are scattered by rows of atoms within the crystal. The pattern of scattered light has maxima at points where waves that interact with different rows arrive in phase and interfere constructively [Eq. (36.16)].

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.

ANSWER

(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 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'll 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. 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

$$\sin\theta_1 = 1.22\frac{\lambda}{D} \qquad (36.17)$$

λ: Wavelength; D: Aperture diameter

The angular radii of the next two dark rings are given by

$$\sin\theta_2 = 2.23\frac{\lambda}{D} \qquad \sin\theta_3 = 3.24\frac{\lambda}{D} \qquad (36.18)$$

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

$$\sin\theta = 1.63\frac{\lambda}{D}, \qquad 2.68\frac{\lambda}{D}, \qquad 3.70\frac{\lambda}{D} \qquad (36.19)$$

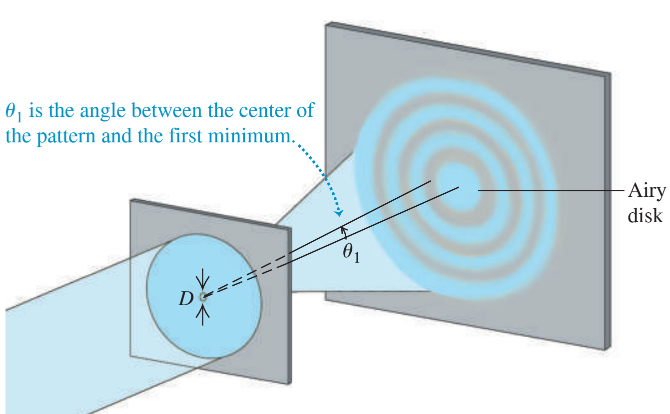
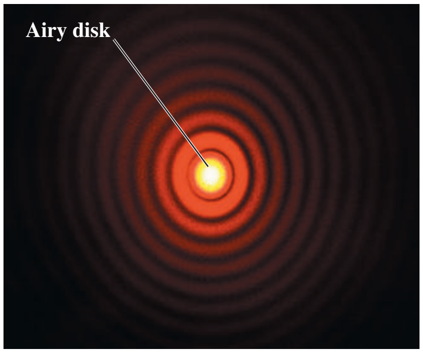


Figure **36.25** Diffraction pattern formed by a circular aperture of diameter D. The pattern consists of a central bright spot and alternating dark and bright rings. The angular radius θ_1 of the first dark ring is shown. (This diagram is not drawn to scale.)

APPLICATION Bigger Telescope, Better Resolving Power 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 resolving power, you could read the "20/20" line on an eye chart more than 30 km away!

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.

Figure 36.26 Photograph of the diffraction pattern formed by a circular aperture.

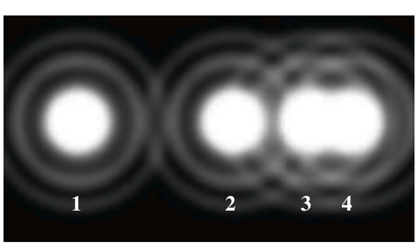


Diffraction and Image Formation

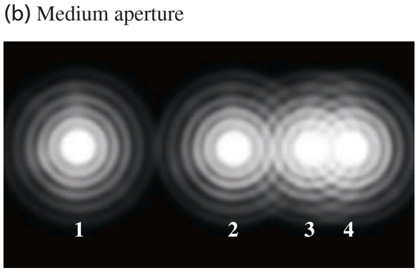
Diffraction has far-reaching implications for image formation by lenses and mirrors. In our study of optical instruments in Chapter 34 we assumed that a lens with focal length f focuses a parallel beam (plane wave) to a *point* at a distance f from the lens. We now see that what we get is *not* a point 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 sources 3 and 4 have merged. In Fig. 36.27b, with a larger aperture diameter and hence smaller Airy disks, the images of sources 3 and 4 are better resolved. In Fig. 36.27c, with a still larger aperture, they are well resolved.

Figure 36.27 These images show the diffraction patterns produced when four very small ("point") sources of light are viewed through a lens with a circular aperture placed in front of the lens. 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).

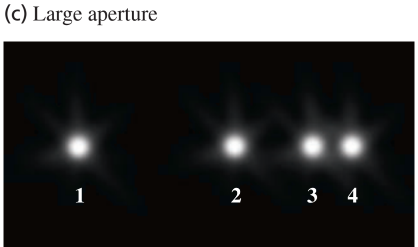
(a) Small aperture



(b) Medium aperture



(c) Large aperture



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.

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).

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 $\frac{5}{9}$ 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.

EXAMPLE 36.6 Resolving power of a camera lens

WITH VARIATION PROBLEMS

vA 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.

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 $D = f/(f\text{-number}) = (50\ \text{mm})/2 = 25\ \text{mm} = 25 \times 10^{-3}$ 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.

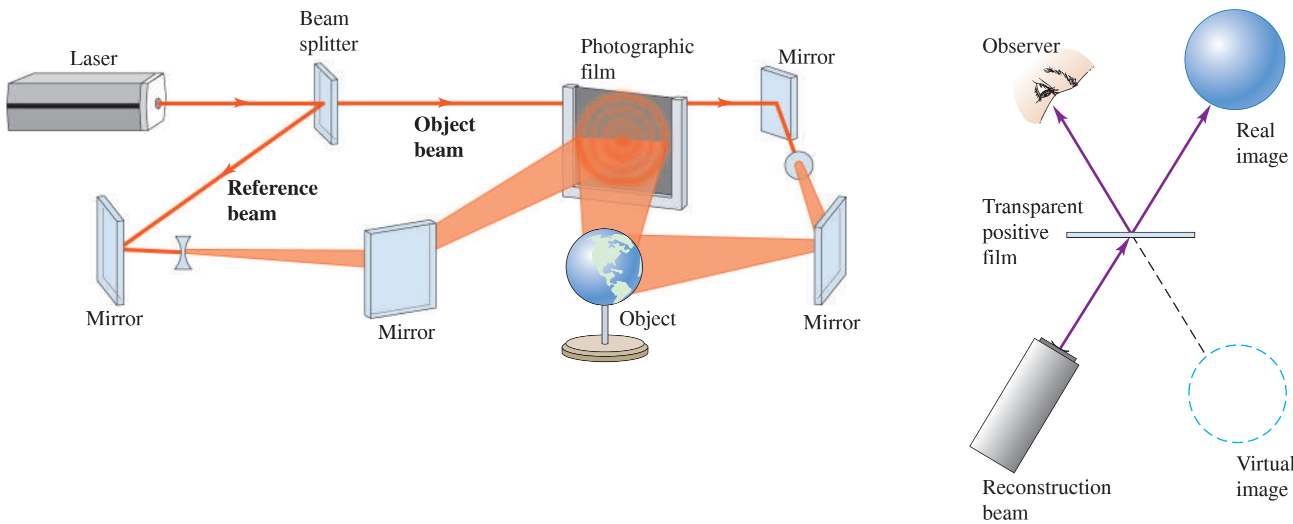
KEYCONCEPT Like waves that pass through a narrow slit, waves that pass through a circular aperture undergo diffraction and spread out. This sets a limit on the ability of a lens or telescope to resolve objects that are close to each other.

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.

ANSWER

❙ **(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) $(10\ \mu\text{m})/(2.0\ \text{m}) = (1.0 \times 10^{-5}\ \text{m})/(2.0\ \text{m}) = 5.0 \times 10^{-6}$.

Figure 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.



36.8 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'll 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

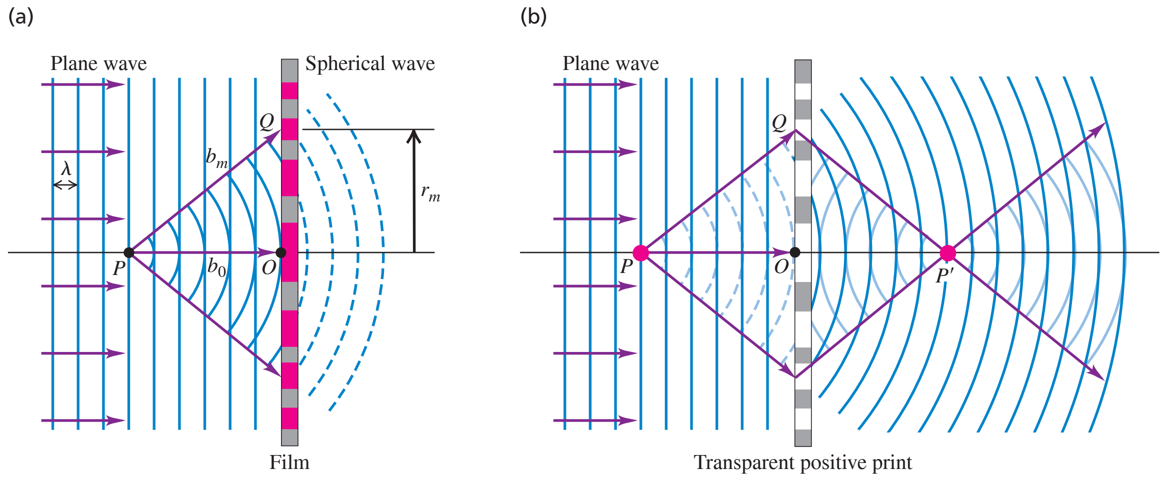
A complete analysis of holography is beyond our scope, but we can gain some insight into the process by looking at how a single point is holographed and imaged. Consider the interference pattern that is formed on a sheet of photographic negative film by the superposition of an incident plane wave and a spherical wave, as shown in **Fig. 36.29a**. 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 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

$$r_m^2 = \lambda(2mb_0 + m^2\lambda)$$

Figure **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 integer 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.



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.

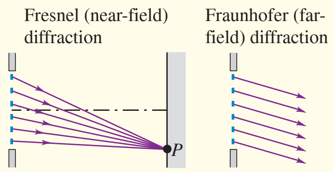
Because of the principle of superposition for waves, what is true for the imaging of a single point is also true for the imaging of any number of points. The film records the superposed interference pattern from the various points, and when light is projected through the film, the various image points 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.

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.

Figure **36.30** Two views of the same hologram seen from different angles.

CHAPTER 36 SUMMARY

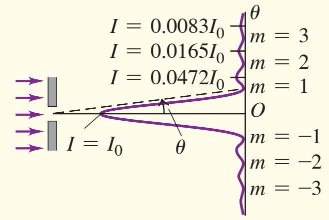
Fresnel and Fraunhofer diffraction: Diffraction occurs when light passes through an aperture or around an edge. When the source and the observer are so far away from the obstructing surface that the outgoing rays can be considered parallel, it is called Fraunhofer diffraction. When the source or the 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.)

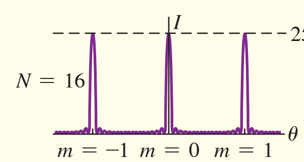
$$\sin\theta = \frac{m\lambda}{a} \qquad (36.2)$$
$$(m = \pm 1, \pm 2, \pm 3, \ldots)$$

$$I = I_0 \left\{ \frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda} \right\}^2 \qquad (36.7)$$



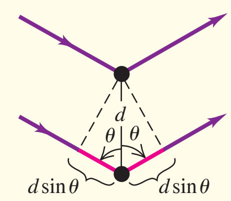
Diffraction gratings: A diffraction grating consists of a large number of thin parallel slits, spaced a distance d apart. The condition for maximum intensity in the interference pattern is the same as for the two-source pattern, but the maxima for the grating are very sharp and narrow. (See Example 36.4.)

$$d\sin\theta = m\lambda \qquad (36.13)$$
$$(m = 0, \pm 1, \pm 2, \pm 3, \ldots)$$



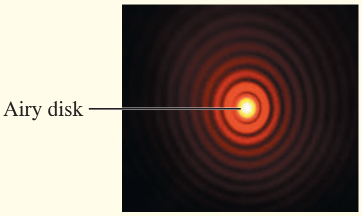
X-ray diffraction: A crystal serves as a three-dimensional diffraction grating for x rays with wavelengths of the same order of magnitude as the spacing between atoms in the crystal. For a set of crystal planes spaced a distance d apart, constructive interference occurs when the angles of incidence and scattering (measured from the crystal planes) are equal and when the Bragg condition [Eq. (36.16)] is satisfied. (See Example 36.5.)

$$2d\sin\theta = m\lambda \qquad (36.16)$$
$$(m = 1, 2, 3, \ldots)$$



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.)

$$\sin\theta_1 = 1.22\frac{\lambda}{D} \qquad (36.17)$$



Chapter 36 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLE 36.1 (Section 36.2) and EXAMPLES 36.2 and 36.3 (Section 36.3) before attempting these problems.

VP36.3.1 You pass laser light of wavelength 645 nm through a slit 0.250 mm in width and observe the diffraction pattern on a screen a large distance away. On the screen, the centers of the second minima on either side of the central bright fringe are 28.0 mm apart. (a) How far away is the screen? (b) What would the distance be between these minima if the wavelength were 525 nm?

VP36.3.2 You shine a laser on a narrow slit 0.221 mm in width. In the diffraction pattern that appears on a screen 5.00 m from the slit, the third minimum is 45.7 mm from the middle of the central bright fringe. Find (a) the wavelength of the laser light and (b) the angle of a line from the center of the slit to the second dark fringe on the screen.

VP36.3.3 At a certain point in a single-slit diffraction pattern there is a phase difference of 35.0 radians between wavelets arriving at the point from the two edges of the slit. The slit is 0.250 mm wide, and the light used has wavelength 545 nm. (a) What is the angle of a line from the center of the slit to this point in the diffraction pattern? (b) If the intensity at the center of the diffraction pattern is I_0, what is the intensity at this point?

VP36.3.4 In a single-slit diffraction experiment, the slit width is 0.185 mm and the wavelength of the light used is 576 nm. (a) Find the angle of a line from the center of the slit to the first dark fringe. (b) You might expect the intensity at one-half the angle found in part (a), which is halfway between the middle of the central bright fringe (where the intensity is I_0) and the first dark fringe (where the intensity is zero), to be $I_0/2$. What is the actual intensity at this angle?

Be sure to review EXAMPLE 36.4 (Section 36.5) and EXAMPLE 36.5 (Section 36.6) before attempting these problems.

VP36.5.1 You shine laser light so that it falls normally on a transmission grating with 825 slits per millimeter. The second-order bright region occurs at an angle of 41.0° from the central maximum. Find (a) the wavelength of the light and (b) the angles for the first-order and third-order bright regions.

VP36.5.2 The first-order bright region created by a diffraction grating occurs at an angle of 24.0° from the central maximum. The wavelength of light that falls normally onto the grating is 625 nm. Find (a) the number of slits per mm on the grating and (b) the angles of the second-order bright region and the third-order bright region (if any).

VP36.5.3 In a certain crystal the spacing between crystal planes is 0.165 nm. (a) If you shine a beam of x rays of wavelength 0.124 nm on this crystal, for what angle between the beam and the crystal planes does the first strong interference maximum occur? (b) For what larger angles (if any) do strong interference maxima occur?

VP36.5.4 In an x-ray diffraction experiment there is only one strong interference maximum, and this occurs when the x-ray beam makes an angle of 36.0° with the crystal planes. The spacing between the crystal planes is 0.158 nm. (a) What is the wavelength of the x rays? (b) What would the wavelength have to be to have three strong interference maxima, with the third at an angle of 88.0°?

Be sure to review EXAMPLE 36.6 (Section 36.7) before attempting these problems.

VP36.6.1 You wish to study the radio emission from the sun at wavelength 1.70 cm. In order to see details on the sun's surface no larger than the diameter of the earth, the limit of resolution of the telescope must be 9.00×10^{-5} rad (about 0.005°). Using the Rayleigh criterion, find (a) the minimum diameter your radio telescope must have and (b) the limit of resolution of this telescope at wavelength 21.1 cm.

VP36.6.2 A telephoto lens for a camera has focal length 0.250 m and maximum aperture $f/4.0$. Take the wavelength of visible light to be 5.50×10^{-7} m. (a) Assuming that the resolution is limited by diffraction, how far away can an object be for you to be able to resolve two points on the object that are 5.00 mm apart? (b) How far apart are the corresponding points on the image made by the lens?

VP36.6.3 A distant point source of light (like a star) emits light of wavelength 575 nm. When this light enters a camera whose lens has a focal length of 135 mm, the diffraction pattern formed on the camera's detector has an Airy disk of radius 0.0112 mm. Find (a) the diameter of the lens aperture and (b) the f-number.

VP36.6.4 One important goal of astronomers is to have a telescope in space that can resolve planets like the earth orbiting other stars. If a planet orbits its star at a distance of 1.5×10^{11} m (the radius of the earth's orbit around the sun) and the telescope has a mirror of diameter 8.0 m, how far from the telescope could the star and its planet be if the wavelength used was (a) 690 nm and (b) 1400 nm? Use the Rayleigh criterion and give your answers in light-years (1 ly $= 9.46 \times 10^{15}$ m).

BRIDGING PROBLEM Observing the Expanding Universe

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

•, ••, •••: 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 integer multiple of $2\pi/N$ (except when ϕ is an integer 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 • Coherent electromagnetic waves with wavelength λ pass through a narrow slit of width a. The diffraction pattern is observed on a tall screen that is 2.00 m from the slit. When $\lambda = 500$ nm, the width on the screen of the central maximum in the diffraction pattern is 8.00 mm. For the same slit and screen, what is the width of the central maximum when $\lambda = 0.125$ mm?

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?

36.12 • When coherent electromagnetic waves with wavelength $\lambda = 120\ \mu$m are incident on a single slit of width a, the width of the central maximum on a tall screen 1.50 m from the slit is 90.0 cm. For the same slit and screen, for what wavelength of the incident waves is the width of the central maximum 180.0 cm, double the value when $\lambda = 120\ \mu$m?

Section 36.3 Intensity in the Single-Slit Pattern

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 •• 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?

36.16 • 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 $\theta = 1.20°$?

Section 36.4 Multiple Slits

36.17 • Nearly monochromatic coherent light waves leave two rectangular slits in phase and at an angle of $\theta = 22.0°$ with the normal. When the light reaches a distant screen, the waves from the center of one slit are 344° out of phase with the waves from the center of the other slit, and the waves from the top of either slit are 172° out of phase with the waves from the bottom of that slit. (a) How is the center-to-center distance between the slits related to the width of either slit? (b) Calculate the intensity at the screen for $\theta = 22.0°$ if the intensity at $\theta = 0°$ is 0.234 W/m^2.

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.

36.23 • Coherent electromagnetic waves with wavelength $\lambda = 500$ nm pass through two identical slits. The width of each slit is a, and the distance between the centers of the slits is $d = 9.00$ mm. (a) What is the smallest possible width a of the slits if the $m = 3$ maximum in the interference pattern is not present? (b) What is the next larger value of the slit width for which the $m = 3$ maximum is absent?

36.24 •• Coherent light with wavelength 200 nm passes through two identical slits. The width of each slit is a, and the distance between the centers of the slits is $d = 1.00$ mm. The $m = 5$ maximum in the two-slit interference pattern is absent, but the maxima for $m = 0$ through $m = 4$ are present. What is the ratio of the intensities for the $m = 1$ and $m = 2$ maxima in the two-slit pattern?

Section 36.5 The Diffraction Grating

36.25 • 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?

36.26 •• 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?

36.27 • You send coherent 550 nm light through a diffraction grating that has slits of equal widths and constant separation between adjacent slits. You expect to see the fourth-order interference maximum at an angle of 66.6° with respect to the normal to the grating. However, that order is missing because 66.6° is also the angle for the third diffraction minimum (as measured from the central diffraction maximum) for each slit. (a) Find the center-to-center distance between adjacent slits. (b) Find the number of slits per mm. (c) Find the width of each slit.

36.28 • 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)?

36.29 • Two Fraunhofer lines in the solar absorption spectrum have wavelengths of 430.790 nm and 430.774 nm. A diffraction grating has 12,800 slits. (a) What is the minimum chromatic resolving power needed to resolve these two spectral lines? (b) What is the lowest order required to resolve these two lines?

36.30 • 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.)

36.31 • (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.

36.32 •• **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.

36.33 • **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.

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 • **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?

36.43 •• **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 • **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

36.45 •• **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?

36.46 •• **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?

36.47 ••• 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.

36.48 • **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?

36.49 • **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.

36.50 •• **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?

36.51 ••• **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?

36.52 •• 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 $I_0/2$?

36.53 •• **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?

36.54 •• **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.)

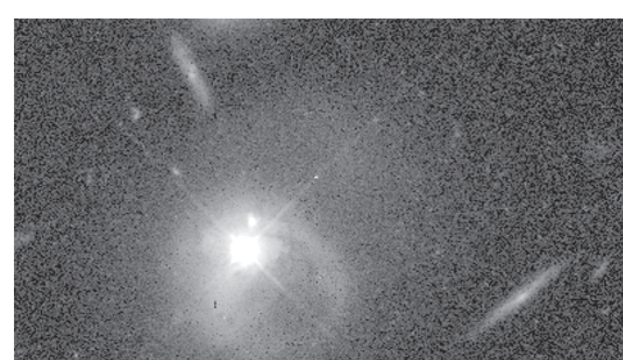
36.55 • What is the longest wavelength that can be observed in the third order for a transmission grating having 9200 slits/cm? Assume normal incidence.

36.56 •• Humans perceive sound with frequencies in the range 20 to 20,000 Hz. Speech lies in the middle of this range, 400 to 1000 Hz, while bells, sirens, and the knock of silverware on plates all extend above the range for speech. Musical instruments that keep rhythm, such as drums and bass guitars, supply sounds with lower frequencies. Doorways act as natural filters, so noises sound different when they come from around corners. (a) Estimate the width of a typical doorway in your house. (b) Determine the sound frequency that corresponds to a wavelength comparable to the width of a doorway. (Such waves freely diffract around the corner of the door.) (c) Determine the frequency of sound that would diffract only $\pm 20°$ from directly forward when passing through a doorway. (d) Based on your estimates, does a doorway act as a noticeable filter for sounds we can hear? (e) If so, which frequency range (low, middle, high) is harder to hear from around the corner of a doorway? (f) Experiment to see if you can notice this effect. It is more readily perceived from outside a house as you listen to sounds from inside. Why is this so?

36.57 • 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.)

36.58 •• *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. P36.58**; 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. P36.58 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 has placed a radio telescope in a large orbit around the earth. When this telescope is 77,000 km from earth and the signals it receives are combined with signals from the ground-based telescopes of the VLBA, the resolution is that of a single radio telescope 77,000 km in diameter. What is the size of the smallest detail that this arrangement can 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 P36.58

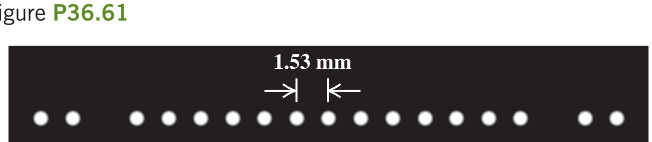


36.59 •• In the 1920s Clinton Davisson and Lester Germer accidentally observed diffraction when electrons with 54 eV of energy were scattered off crystalline nickel. The diffraction peak occurred when the angle between the incident beam and the scattered beam was 50°. (a) What is the corresponding angle θ relevant for Eq. (36.16)? (b) The planes in crystalline nickel are separated by 0.091 nm, as determined by x-ray scattering experiments. According to the Bragg condition, what wavelength do the electrons in these experiments have? (c) Given the mass of an electron as 9.11×10^{-31} kg, what is the corresponding classical speed v_{cl} of the diffracted electrons? (d) Assuming the electrons correspond to a wave with speed v_{cl} and wavelength λ, what is the frequency f of the diffracted waves? (e) Quantum mechanics postulates that the energy E and the frequency f of a particle are related by $E = hf$, where h is known as Planck's constant. Estimate h from these observations. (f) Our analysis has a small flaw: The relevant wave velocity, known as a quantum phase velocity, is *half* the classical particle velocity, for reasons explained by deeper aspects of quantum physics. Re-estimate the value of h using this modification. (g) The established value of Planck's constant is $6.6 \times 10^{-34}\ \mathrm{J \cdot s}$. Does this agree with your estimate?

36.60 •• 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?

36.61 •• 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. P36.61**. (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 P36.61



36.62 •• 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) $a = 0.78\ \mu$m and (ii) $a = 15.60\ \mu$m?

36.63 •• 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

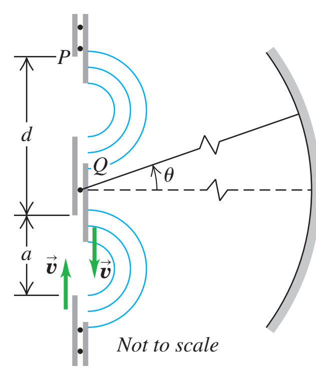
(a) Use each set of measurements to calculate D. Because the measurements contain some error, calculate the average of the four values of D and take that to be your reported result. (b) For $x = 1.00$ m, what are the radii of the second and third dark rings in the diffraction pattern?

36.64 •• 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?

CHALLENGE PROBLEMS

36.65 ••• An opaque barrier has an inner membrane and an outer membrane that slide past each other, as shown in **Fig. P36.65**. Each membrane includes parallel slits of width a separated by a distance d. A screen forms a circular arc subtending 60° at the fixed midpoint between the slits. A green 532 nm laser impinges on the slits from the left. The outer membrane moves upward with speed v while the inner membrane moves downward with the same speed, propelled by nanomotors. At time $t = 0$, point P on the outer membrane is adjacent to point Q on the inner membrane so that the effective aperture width is zero. The aperture is fully closed again at $t = 3.00$ s. (a) At $t = 1.00$ s, there are 19 evenly spaced bright spots on the screen, each of approximately the same intensity. At the edges of the screen the first diffraction minimum and a two-slit interference maximum coincide. What is the slit distance d? (Note: The screen does not encompass the entire diffraction pattern.) (b) What is the speed v? (c) What is the maximum aperture width a? (d) At a certain time, the outermost spots (the $m = \pm 9$ spots) disappear. What is that time? (e) At $t = 1.50$ s, what is the intensity of the $m = \pm 1$ spots in terms of the $m = 0$ central spot? (f) What are the angular positions of these spots?

Figure P36.65



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 integer multiple of $2\pi/N$, except when ϕ is an integer 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).

MCAT-STYLE 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.

Key Example √ARIATION Problems

VP36.3.1 (a) 2.30 m (b) 22.8 mm
VP36.3.2 (a) 673 nm (b) 0.349°
VP36.3.3 (a) 0.696° (b) $(3.11 \times 10^{-3})I_0$
VP36.3.4 (a) 0.178° (b) $0.405I_0$
VP36.5.1 (a) 398 nm (a) 19.1° for $m = 1$, 79.8° for $m = 3$
VP36.5.2 (a) 651 slits/mm (b) 54.4° for $m = 2$; there is no $m = 3$ bright region
VP36.5.3 (a) 22.1° (b) 48.7°
VP36.5.4 (a) 0.186 nm (b) 0.105 nm
VP36.6.1 (a) 2.50×10^2 m (b) 1.11×10^{-3} rad $= 0.0640°$
VP36.6.2 (a) 466 m (b) 0.00268 mm
VP36.6.3 (a) 8.46 mm (b) $f/16.0$
VP36.6.4 (a) 150 ly (b) 74 ly

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

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.

LEARNING OUTCOMES

In this chapter, you'll learn...

- **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.

You'll need to review...

- **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.

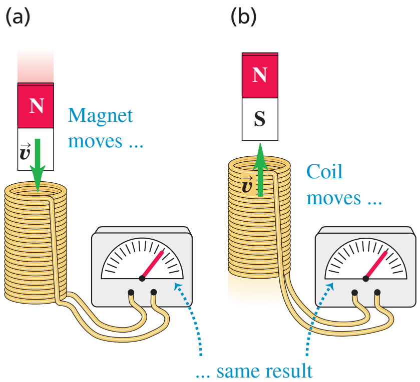
37.1 INVARIANCE OF PHYSICAL LAWS

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

Figure 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.



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.

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'll see, the speed of light in vacuum plays a central role in the theory of relativity.

Einstein's Second Postulate

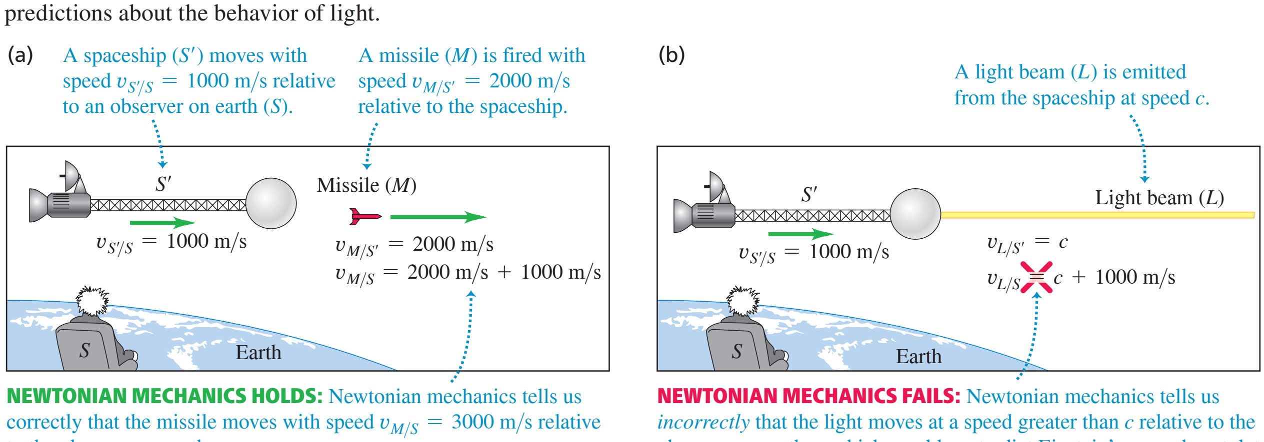
During the 19th century, most physicists believed that light traveled through a hypothetical medium called the *ether,* just as sound waves travel through air. If so, the speed of light measured by observers would depend on their motion relative to the ether and would therefore be different in different 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**). 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,

Figure 37.2 (a) Newtonian mechanics makes correct predictions about relatively slow-moving objects; (b) it makes incorrect predictions about the behavior of light.



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.

The Ultimate Speed Limit

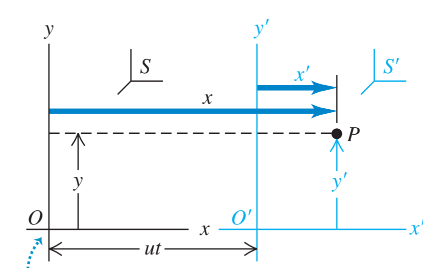
Einstein's second postulate immediately implies the following result: **It is impossible for an inertial observer to travel at c, the speed of light in vacuum.**

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

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.

Figure **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'.



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. When solving problems with reference frames, you should *always* draw a diagram showing which reference frame goes with which observer.

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

$$x = x' + ut \qquad y = y' \qquad z = z' \quad \text{(Galilean coordinate transformation)} \tag{37.1}$$

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?

ANSWER

(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.

Figure 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.

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.

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.

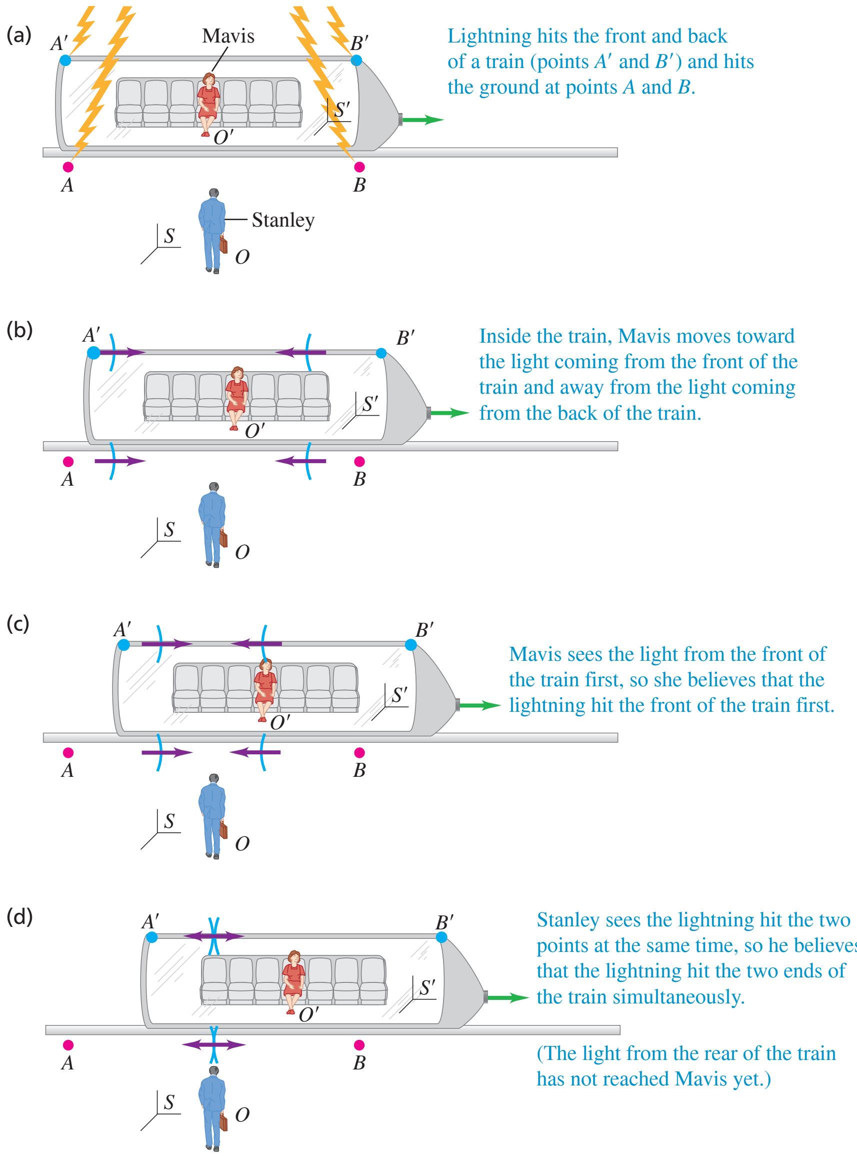


Figure **37.5** A thought experiment in simultaneity.

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 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.

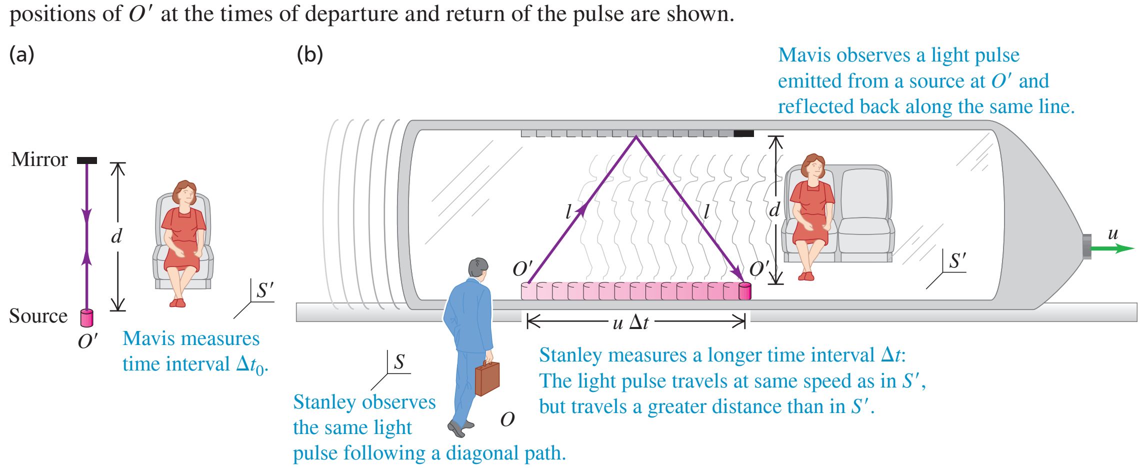
ANSWER

(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 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**. We label the time interval Δt_0, using the subscript zero as a reminder

Figure **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.



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. 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'll 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

Time dilation:

Proper time between two events (measured in rest frame)

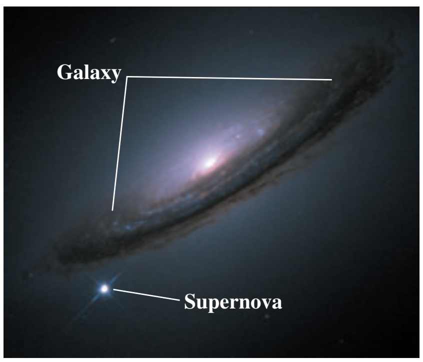
$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} \tag{37.6}$$

Speed of light in vacuum

Speed of second frame relative to rest frame

Time interval between same events measured in second frame of reference

Figure **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.



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)

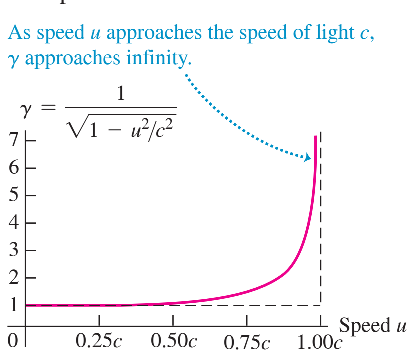
Time dilation: $$\Delta t = \gamma \Delta t_0 \qquad (37.8)$$
Lorentz factor relating the two frames
Time interval between same events measured in second frame of reference

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.

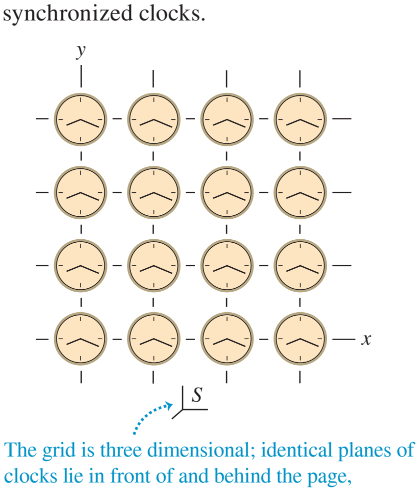
If the relative speed u is great enough that γ is appreciably greater than 1, the speed is said to be *relativistic;* if the difference between γ and 1 is negligibly small, the speed u is called *nonrelativistic*. Thus $u = 6.00 \times 10^7$ m/s $= 0.200c$ (for which $\gamma = 1.02$) is a relativistic speed, but $u = 6.00 \times 10^4$ m/s $= 0.000200c$ (for which $\gamma = 1.00000002$) is a nonrelativistic speed.

Figure **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.



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.)

Figure **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'll 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'll 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'll 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 0.990c

WITH VARIATION PROBLEMS

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 $2.20\ \mu\text{s} = 2.20 \times 10^{-6}$ 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?

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, $\Delta t_0 = 2.20\ \mu\text{s}$. 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),

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} = \frac{2.20\ \mu\text{s}}{\sqrt{1 - (0.990)^2}} = 15.6\ \mu\text{s}$$

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).

KEYCONCEPT The time interval between two events depends on the frame of reference of the observer. The time interval is shortest in an inertial frame S where the two events occur at the same point; this is the proper time Δt_0. In any other inertial frame that is moving with respect to S, the time interval is longer (it is *dilated*) [Eq. (37.6)].

EXAMPLE 37.2 Time dilation at airliner speeds

WITH VARIATION PROBLEMS

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?

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):

$$(1 - 1.00 \times 10^{-12})^{1/2} = 1 - \left(\tfrac{1}{2}\right)(1.00 \times 10^{-12}) + \cdots$$

Continued

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.

KEYCONCEPT If two events take place at the same point in an inertial frame S, an observer in a frame S' moving relative to S observes a time interval between these events that is greater than in S by the Lorentz factor γ [Eqs. (37.7) and (37.8)]. This factor is substantially different from 1 only if the relative speed of S' and S is a substantial fraction of the speed of light.

EXAMPLE 37.3 Just when is it proper?

WITH VARIATION PROBLEMS

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?

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 $0.600c = 0.600(3.00 \times 10^8 \text{ m/s}) = 1.80 \times 10^8 \text{ m/s}$ 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

$$\Delta t_0 = \Delta t \sqrt{1 - u^2/c^2} = 0.500 \text{ s} \sqrt{1 - (0.600)^2} = 0.400 \text{ s}$$

Since Stanley measures 0.500 s elapsed on his timer but only 0.400 s elapsed on Mavis's timer, Stanley concludes that Mavis's timer runs slow.

(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:

$$\Delta t_0 = \Delta t \sqrt{1 - u^2/c^2} = 0.400 \text{ s} \sqrt{1 - (0.600)^2} = 0.320 \text{ s}$$

Since Mavis measures 0.400 s elapsed on her timer for the duration of Stanley's blink but only 0.320 s elapsed on Stanley's timer, Mavis concludes that Stanley's timer runs slow. *Both* observers measure that a moving clock runs slow!

EVALUATE This example illustrates both time dilation and 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.

KEYCONCEPT No matter which inertial frame of reference you are in, you observe dilated time intervals for timekeepers (clocks) of any kind that move relative to you. If S and S' are two inertial frames in relative motion, an observer in S measures clocks in S' to run slow, and an observer in S' measures clocks in S to run slow.

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.

ANSWER (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.

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 for her, 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!

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'.

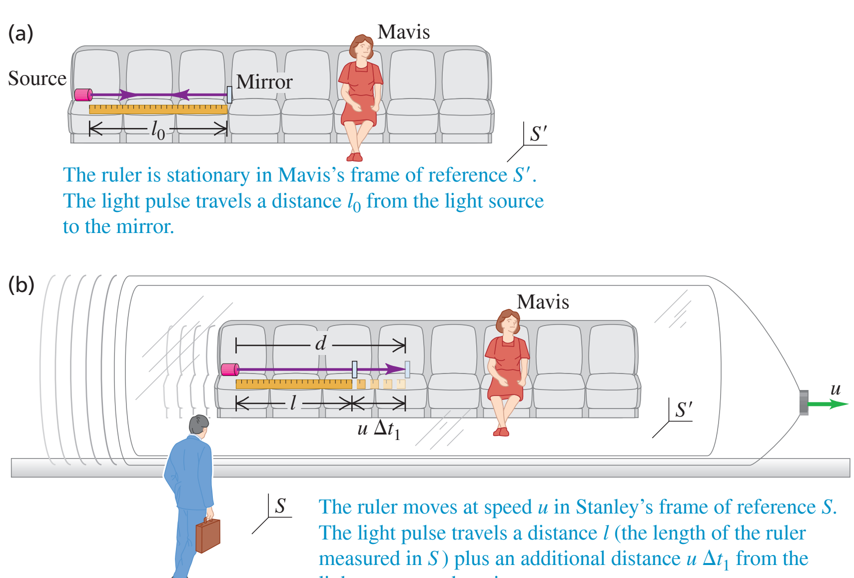


Figure **37.10** (a) A ruler is at rest in Mavis's frame S'. A light pulse is emitted from a source at one end of the ruler, reflected by a mirror at the other end, and returned to the source position. (b) Motion of the light pulse as measured in Stanley's frame S.

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

$$d = l + u\,\Delta t_1 \tag{37.10}$$

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'.

A length measured in the frame in which the object is at rest (the rest frame of the object) 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**).

Figure 37.11 The speed at which electrons traverse the 3 km beam line of the SLAC National Accelerator Laboratory is slower than c by only 10 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 8 cm long!

Beam line

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'.

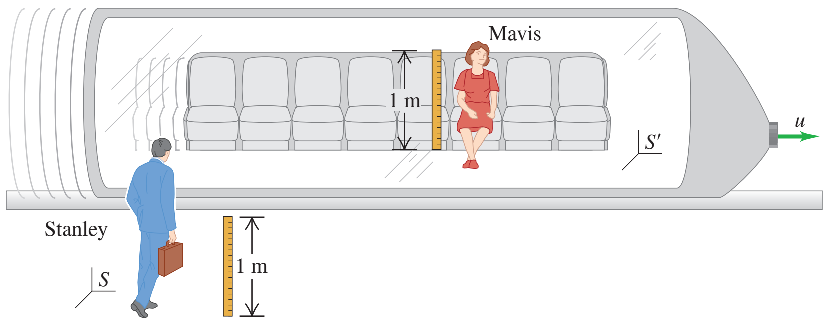


Figure **37.12** The meter sticks are perpendicular to the relative velocity. For any value of u, both Stanley and Mavis measure either meter stick to have a length of 1 meter.

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.

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?

WITH VARIATION PROBLEMS

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?

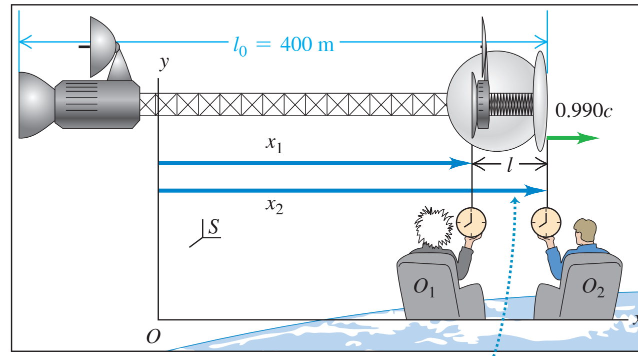
IDENTIFY and SET UP This problem is about the nose-to-tail length of the spaceship as measured on the spaceship and on earth. This length is along the direction of relative motion (**Fig. 37.13**), so there will be length contraction. The spaceship's 400 m length is the *proper* length l_0 because it 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.)

Figure **37.13** Measuring the length of a moving spaceship.



KEYCONCEPT The dimensions of an object depend on the frame of reference of the observer. If the object is moving at constant velocity relative to your inertial frame, its length is contracted *along* the direction of relative motion [Eq. (37.16)]. There is no change in length *perpendicular* to the direction of relative motion.

EXAMPLE 37.5 How far apart are the observers?

WITH VARIATION PROBLEMS

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?

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.

KEYCONCEPT No matter which inertial frame of reference you are in, you observe objects that move relative to you to be contracted along the direction of relative motion. If S and S' are two inertial frames in relative motion, an observer in S measures objects in S' to have contracted lengths, and an observer in S' measures objects in S to have contracted lengths.

How an Object Moving Near *c* Would Appear

Let's think a little about the visual appearance of a moving three-dimensional object. If we could see the positions of all points of the object 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.

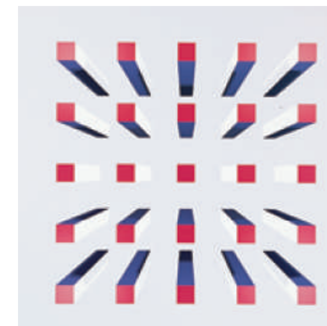
TEST YOUR UNDERSTANDING OF SECTION 37.4 A miniature spaceship is flying past you, moving horizontally at a substantial fraction of the speed of light. At a certain instant, you observe that the nose and tail of the spaceship align exactly with the two ends of a meter stick that you hold in your hands. 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.

ANSWER

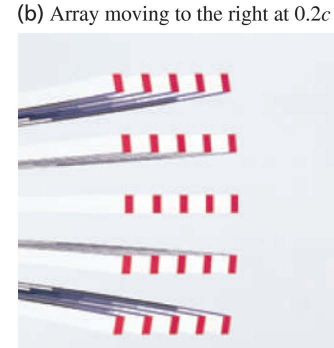
| (ii), (i) and (iii) (tie), (iv) You measure both the rest length of the stationary meter stick and the contracted length of the moving spaceship to be 1 meter. The rest length of the spaceship is greater than the contracted length that you measure, and so must be greater than 1 meter. A miniature observer 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.

Figure **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).

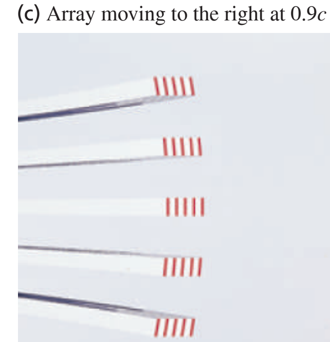
(a) Array at rest



(b) Array moving to the right at $0.2c$



(c) Array moving to the right at $0.9c$



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**, (next page) 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

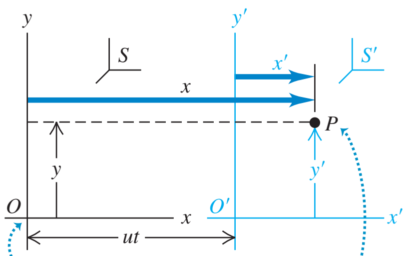
$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}} \tag{37.18}$$

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

$$x' = -ut' + x\sqrt{1 - \frac{u^2}{c^2}} \tag{37.19}$$

Figure 37.15 As measured in frame of reference S, x' is contracted to x'/γ, so $x = ut + (x'/\gamma)$ and $x' = \gamma(x - ut)$.

Frame S' moves relative to frame S with constant velocity u along the common x-x'-axis.



Origins O and O' coincide at time $t = 0 = t'$.

The Lorentz coordinate transformation relates the spacetime coordinates of an event as measured in the two frames: (x, y, z, t) in frame S and (x', y', z', t') in frame S'.

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

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} \tag{37.20}$$

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

$$\begin{aligned} x' &= \frac{x - ut}{\sqrt{1 - u^2/c^2}} = \gamma(x - ut) \\ y' &= y \\ z' &= z \\ t' &= \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} = \gamma(t - ux/c^2) \end{aligned} \tag{37.21}$$

Lorentz factor relating the two frames

Speed of light in vacuum

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):

$$\begin{aligned} dx' &= \gamma(dx - u\,dt) \\ dt' &= \gamma(dt - u\,dx/c^2) \end{aligned}$$

We divide the first equation by the second and then divide the numerator and denominator of the result by dt to obtain

$$\frac{dx'}{dt'} = \frac{\dfrac{dx}{dt} - u}{1 - \dfrac{u}{c^2}\dfrac{dx}{dt}}$$

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):

x-velocity of object in frame S'

x-velocity of object in frame S

$$v'_x = \frac{v_x - u}{1 - uv_x/c^2} \tag{37.22}$$

Velocity of S' relative to S in positive direction along x-x'-axis

Speed of light in vacuum

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'):

$$\underbrace{v_x}_{x\text{-velocity of object in frame } S} = \frac{\overbrace{v'_x}^{x\text{-velocity of object in frame } S'} + \overbrace{u}^{\text{Velocity of } S' \text{ relative to } S \text{ in positive direction along } x\text{-}x'\text{-axis}}}{1 + uv'_x/\underbrace{c^2}_{\text{Speed of light in vacuum}}} \qquad (37.23)$$

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 an object 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 object 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.

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.

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.

Reference frame S: Our rest frame as we stand next to the track

Reference frame S': Rest frame of runner in red. She moves with speed u relative to us.

Runner in purple has velocity v_x in S and velocity v'_x in S'.

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.

EXAMPLE 37.6 Was it received before it was sent?

WITH VARIATION PROBLEMS

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?

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.

KEYCONCEPT Given the position and time of an event as measured in one inertial frame of reference, you can use the Lorentz coordinate transformation to find the position and time of that same event as measured in a second inertial frame moving relative to the first one.

EXAMPLE 37.7 Relative velocities

WITH VARIATION PROBLEMS

(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?

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'.

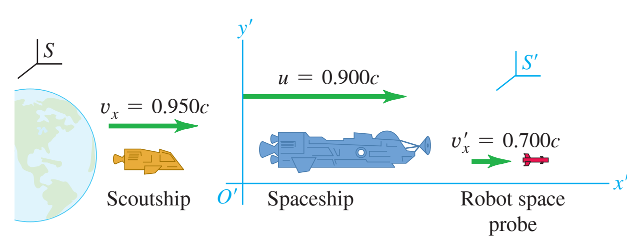
EXECUTE (a) We use Eq. (37.23) to find the probe velocity relative to the earth:

$$v_x = \frac{v'_x + u}{1 + uv'_x/c^2} = \frac{0.700c + 0.900c}{1 + (0.900c)(0.700c)/c^2} = 0.982c$$

(b) We use Eq. (37.22) to find the scoutship velocity relative to the spaceship:

$$v'_x = \frac{v_x - u}{1 - uv_x/c^2} = \frac{0.950c - 0.900c}{1 - (0.900c)(0.950c)/c^2} = 0.345c$$

Figure 37.16 The spaceship, robot space probe, and scoutship.



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.

KEYCONCEPT Given the velocity of an object as measured in one inertial frame of reference, you can use the Lorentz velocity transformation to find the velocity of that same object as measured in a second inertial frame moving relative to the first one. If an object is moving slower than c as measured in one inertial frame, its speed is less than c as measured in any other inertial frame.

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?

ANSWER

(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.

37.6 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. He measures the frequency 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:

$$f = \frac{c}{c-u}\frac{\sqrt{c^2 - u^2}}{c}f_0$$

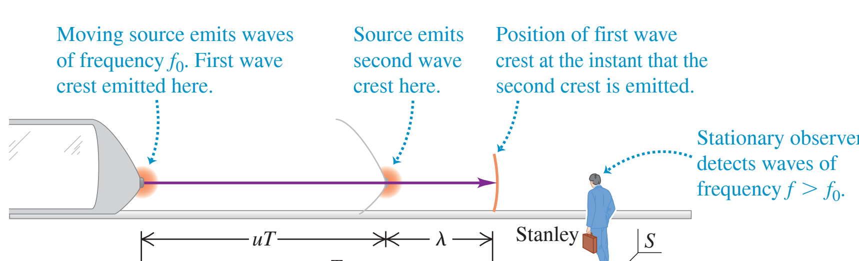


Figure **37.17** The Doppler effect for light. A light source moving at speed u relative to Stanley emits a wave crest, then travels a distance uT toward an observer and emits the next crest. In Stanley's reference frame S, the second crest is a distance λ behind the first crest.

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.

Figure 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

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?

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

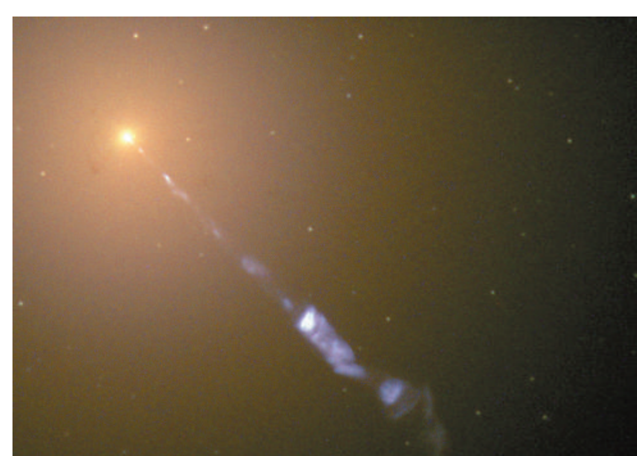
$$u = \frac{(f/f_0)^2 - 1}{(f/f_0)^2 + 1}c$$

We have $f/f_0 = (6.66 \times 10^{14}\ \text{Hz})/(5.55 \times 10^{13}\ \text{Hz}) = 12.0$, so

$$u = \frac{(12.0)^2 - 1}{(12.0)^2 + 1}c = 0.986c$$

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 that $u = c(\Delta f/f_0) = c(6.66 \times 10^{14}\ \text{Hz} - 5.55 \times 10^{13}\ \text{Hz})/(5.55 \times 10^{13}\ \text{Hz}) = 11.0c$. This result cannot be correct because the jet material cannot travel faster than light.

Figure 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).



KEYCONCEPT The frequency of electromagnetic waves as measured by an observer moving relative to their source is different from the frequency measured in the rest frame of the source. The measured frequency is greater than the source frequency if the observer and source are approaching, and less than the source frequency if they are receding [Eq. (37.25)].

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'll 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 objects interact, the total momentum is constant,* provided that the net external force acting on the objects 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'll 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

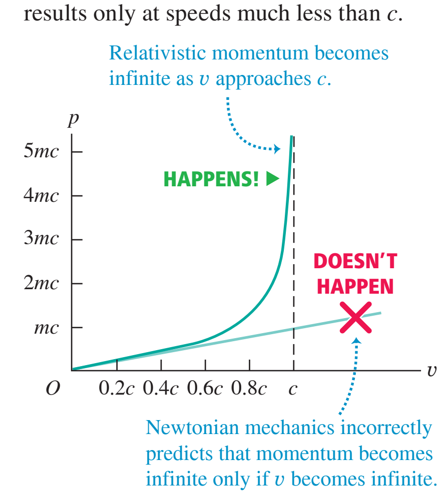
Relativistic momentum

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} \tag{37.27}$$

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.

Figure **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 material 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 material 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'll 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'll mostly deal with individual particles, so we'll 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

WITH VARIATION PROBLEMS

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.

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 $p = \gamma mv$ are

$$\begin{aligned} p_1 &= (1.00)(9.11 \times 10^{-31}\text{ kg})(0.010)(3.00 \times 10^8\text{ m/s}) \\ &= 2.7 \times 10^{-24}\text{ kg}\cdot\text{m/s at } v_1 = 0.010c \\ p_2 &= (2.29)(9.11 \times 10^{-31}\text{ kg})(0.90)(3.00 \times 10^8\text{ m/s}) \\ &= 5.6 \times 10^{-22}\text{ kg}\cdot\text{m/s at } v_2 = 0.90c \\ p_3 &= (7.09)(9.11 \times 10^{-31}\text{ kg})(0.99)(3.00 \times 10^8\text{ m/s}) \\ &= 1.9 \times 10^{-21}\text{ kg}\cdot\text{m/s at } v_3 = 0.99c \end{aligned}$$

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.

KEYCONCEPT The expression for momentum [Eq. (37.31)] and the relationship between net force and acceleration [Newton's second law, Eqs. (37.32) and (37.33)] are both modified in the special theory of relativity. As the speed of a material particle approaches c, its momentum approaches infinity and the acceleration caused by a net force approaches zero.

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.

ANSWER

(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 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,

$$W = \int_{x_1}^{x_2} F\,dx = \int_{x_1}^{x_2} \frac{ma\,dx}{(1 - v_x^2/c^2)^{3/2}} \tag{37.34}$$

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

Relativistic kinetic energy

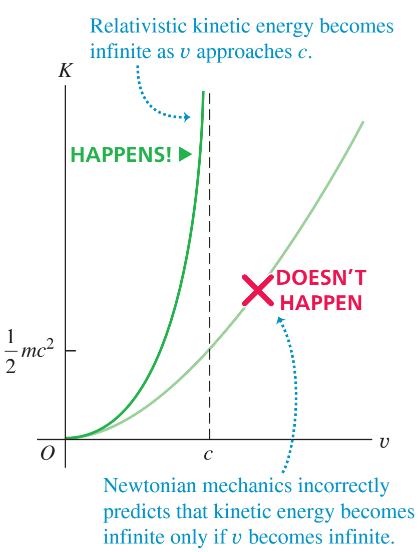
$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2 \tag{37.36}$$

Rest mass of particle: m; Speed of light in vacuum: c; Speed of particle: v; Lorentz factor relating rest frame of particle and frame of observer: γ

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 **?**

$$(1 + x)^n = 1 + nx + n(n - 1)x^2/2 + \cdots$$

Figure **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.



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 $\frac{1}{2}mv^2$.

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

Total energy of a particle

$$E = K + mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \gamma mc^2 \tag{37.38}$$

Kinetic energy: K; Rest energy: mc^2; Rest mass of particle: m; Speed of light in vacuum: c; Speed of particle: v; Lorentz factor relating rest frame of particle and frame of observer: γ

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.

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:

$$\overset{\text{Total energy}}{E^2} = \overset{\text{Rest energy}}{(\underset{\text{Rest mass}}{m}\underset{\text{Speed of light in vacuum}}{c^2})^2} + (\overset{\text{Magnitude of momentum}}{p}c)^2 \qquad (37.39)$$

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.

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.

EXAMPLE 37.10 Energetic electrons

WITH VARIATION PROBLEMS

(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).

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, $1\ \text{eV} = 1.602 \times 10^{-19}$ J. Using this, we find

$$mc^2 = (8.187 \times 10^{-14}\ \text{J})\frac{1\ \text{eV}}{1.602 \times 10^{-19}\ \text{J}}$$
$$= 5.11 \times 10^5\ \text{eV} = 0.511\ \text{MeV}$$

Continued

(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:

$$E = \gamma mc^2 = mc^2 + eV_{ba} \qquad \text{or} \qquad \gamma = 1 + \frac{\text{eV}_{\text{ba}}}{\text{mc}^2}$$

An electron accelerated through a potential increase of $V_{ba} = 20.0\ \text{kV}$ 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 $V_{ba} = 5.00\ \text{MV}$, we find $eV_{ba}/mc^2 = 9.78$, $\gamma = 10.78$, and $v = 0.996c$.

EVALUATE With $V_{ba} = 20.0\ \text{kV}$, 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 $V_{ba} = 5.00\ \text{MV}$, 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.

KEYCONCEPT The expression for kinetic energy [Eq. (37.36)] is modified in the special theory of relativity. As the speed of a material particle approaches c, its kinetic energy approaches infinity. The total energy of a particle of rest mass m is the sum of its kinetic energy and its rest energy mc^2.

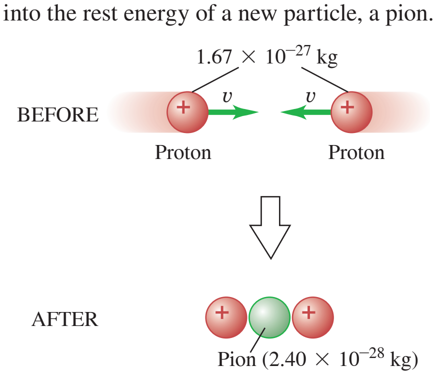
EXAMPLE 37.11 A relativistic collision

WITH VARIATION PROBLEMS

Two protons (each with mass $m_p = 1.67 \times 10^{-27}\ \text{kg}$) 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 $m_\pi = 2.40 \times 10^{-28}\ \text{kg}$ (**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.

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.

Figure 37.22 In this collision the kinetic energy of two protons is transformed into the rest energy of a new particle, a pion.



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 $(\gamma - 1)m_p c^2 = 0.072 m_p c^2 = (0.072)(938\ \text{MeV}) = 67.5\ \text{MeV}$. 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.

KEYCONCEPT In a relativistic collision, momentum and total energy are both conserved. Particles can be created or destroyed in such collisions by converting kinetic energy to rest energy, or vice versa.

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.

ANSWER

(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. (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 asks, however.)

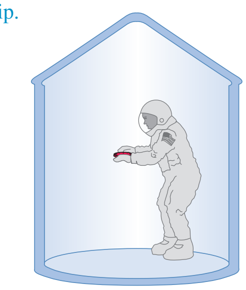
37.9 NEWTONIAN MECHANICS AND RELATIVITY

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.

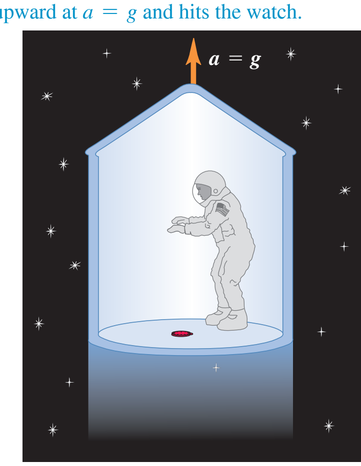
The laws of Newtonian mechanics are not *wrong;* they are *incomplete*. They are a limiting case of relativistic mechanics. They are *approximately* correct when all speeds are small in comparison to c, and they become exactly correct in the limit when all speeds approach zero. Thus relativity does not 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.**

Figure **37.23** Without information from outside the spaceship, the astronaut cannot distinguish situation (b) from situation (c).

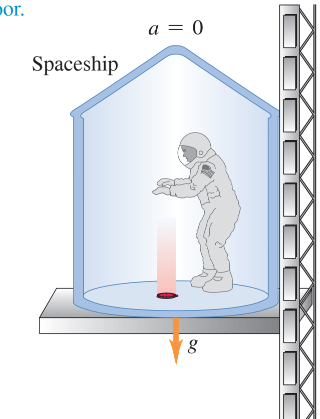
(a) An astronaut is about to drop her watch in a spaceship.



(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 $9.8\ \text{m/s}^2$. 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 spacetime 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**, next page).

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 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.

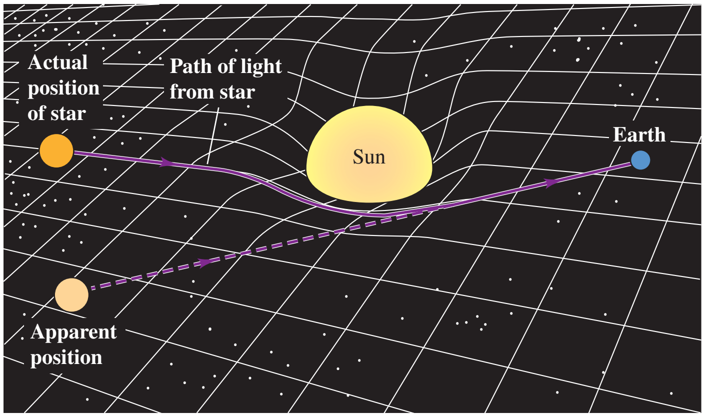


Figure 37.24 A two-dimensional representation of curved space. We imagine the space (a plane) as being distorted as shown by a massive object (the sun). Light from a distant star (solid line) follows the distorted surface on its way to the earth. The dashed line shows the direction from which the light *appears* to be coming. The effect is greatly exaggerated; for the sun, the maximum deviation is only 0.00048°.

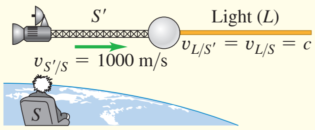
Figure 37.25 A GPS receiver uses radio signals from the orbiting GPS satellites to determine its position. To account for the effects of relativity, the receiver must be tuned to a slightly higher frequency (10.23 MHz) than the frequency emitted by the satellites (10.22999999543 MHz).

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.

CHAPTER 37 SUMMARY

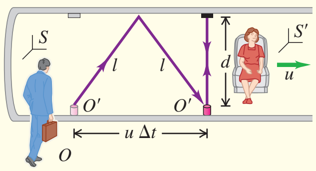
Invariance of physical laws, simultaneity: All of the fundamental laws of physics have the same form in all inertial frames of reference. The speed of light in vacuum is the same in all inertial frames and is independent of the motion of the source. Simultaneity is not an absolute concept; events that 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.)

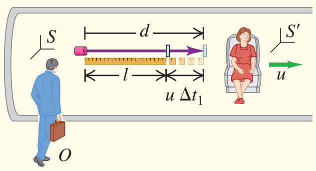
$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} = \gamma\,\Delta t_0 \qquad (37.6), (37.8)$$

$$\gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \qquad (37.7)$$



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.)

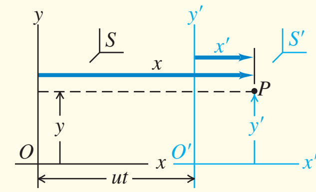
$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}} = \gamma(x - ut)$$

$$y' = y \qquad z' = z \tag{37.21}$$

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} = \gamma(t - ux/c^2)$$

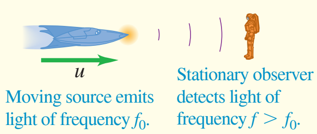
$$v'_x = \frac{v_x - u}{1 - uv_x/c^2} \tag{37.22}$$

$$v_x = \frac{v'_x + u}{1 + uv'_x/c^2} \tag{37.23}$$



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.)

$$f = \sqrt{\frac{c + u}{c - u}}\, f_0 \tag{37.25}$$



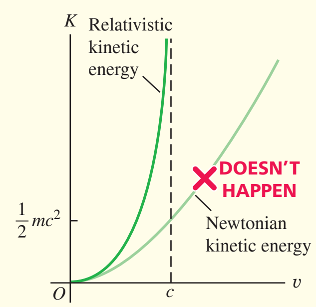
Relativistic momentum and energy: For a particle of rest mass m moving with velocity $\vec{v}$, the relativistic momentum $\vec{p}$ is given by Eq. (37.27) or (37.31) and the relativistic kinetic energy K is given by Eq. (37.36). The total energy E is the sum of the kinetic energy and the rest energy 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.)

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} = \gamma m\vec{v} \qquad (37.27), (37.31)$$

$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2 \tag{37.36}$$

$$E = K + mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \gamma mc^2 \tag{37.38}$$

$$E^2 = (mc^2)^2 + (pc)^2 \tag{37.39}$$



Chapter 37 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 37.1, 37.2, and 37.3 (Section 37.3) and EXAMPLES 37.4 and 37.5 (Section 37.4) before attempting these problems.

VP37.5.1 A positive K meson (K^+) is an unstable subatomic particle that decays into other particles. At rest, its mean lifetime is 1.23×10^{-8} s. (a) What do you measure the mean lifetime of a K^+ to be if it is moving at $0.800c$ relative to you? (b) As measured from the reference frame of this K^+, what is the mean lifetime of a second K^+ that is at rest relative to you?

VP37.5.2 Alia flies in her spacecraft at $0.600c$ relative to the planet Arrakis. As she passes Paul, at rest on Arrakis, they both start timers. (a) According to Alia's timer, 20.0 s elapses from when Paul starts his timer to when he stops his timer. What does Paul's timer read when he stops it? (b) Alia stops her timer when it reads 24.0 s. As measured by Paul, how much time elapses from when Alia starts her timer to when she stops it?

VP37.5.3 At one point in its orbit the earth is 1.50×10^{11} m from the sun. A spacecraft flies along a line from the earth to the sun at $0.950c$. (a) As measured by you on the earth, how much time does it take the spacecraft to travel the distance from the earth to the sun? (b) As measured by an astronaut on the spacecraft, what is the distance from the earth to the sun, and how much time does it take her to travel that distance?

VP37.5.4 A spacecraft flies past the earth in the direction toward the moon, a distance of 3.84×10^5 km. As measured by your clock on the earth, it takes 6.00 s for the clock on the spacecraft to tick off 2.00 s. You measure the length of the spacecraft in the direction of its motion to be 24.0 m. As measured by an astronaut at rest on the spacecraft, what are (a) the distance from the earth to the moon and (b) the length of the spacecraft?

Be sure to review EXAMPLES 37.6 and 37.7 (Section 37.5) before attempting these problems.

VP37.7.1 Gamora flies her spacecraft in the $+x$-direction past the planet Xandar at $0.750c$ while her sister Nebula is at rest on the surface of the planet. As Gamora passes Nebula, both sisters set their clocks to zero. Each chooses the zero of the x-axis to be at her position. (a) Nebula sets off fireworks next to her 2.50 s after Gamora passes her. What are the coordinates of this event as measured by Gamora? (b) Gamora detects an explosion in space that occurs 4.00×10^8 m in front of her 2.50 s after she passes Nebula. What are the coordinates of this event as measured by Nebula?

VP37.7.2 Two events occur at different x-coordinates but at the same time as measured by Kamala. Doreen, who is moving at $0.800c$ relative to Kamala in the $+x$-direction, measures that one event takes place 0.600 s before the other. What is the distance between the two events as measured by (a) Kamala and (b) Doreen?

VP37.7.3 The spaceship *Nostromo* flies away from the earth in the $+x$-direction at $0.800c$. It fires a probe in the $-x$-direction at $0.600c$ relative to *Nostromo*. A scoutship is sent from the earth to recover the probe; the scoutship travels in the $+x$-direction at $0.700c$ relative to the earth. Find the velocity (in terms of c) of (a) the probe relative to the earth and (b) the scoutship relative to *Nostromo*.

VP37.7.4 As measured from the earth, the spaceship *Macross* is flying directly toward the earth at $0.750c$ while on the other side of the earth the spaceship *Yamato* is flying at $0.650c$ along the same line as *Macross*. Find the speed (in terms of c) of *Yamato* relative to *Macross* if *Yamato* is flying (a) toward the earth and (b) away from the earth.

Be sure to review EXAMPLE 37.9 (Section 37.7) and EXAMPLES 37.10 and 37.11 (Section 37.8) before attempting these problems.

VP37.11.1 A proton (rest mass 1.67×10^{-27} kg) is moving at $0.925c$. Find (a) the momentum of the proton, (b) its acceleration if a force of 9.00×10^{-14} N acts on the proton in the direction of its motion, and (c) its acceleration if a force of 9.00×10^{-14} N acts on the proton perpendicular to the direction of its motion.

VP37.11.2 A moving electron (rest mass 9.11×10^{-31} kg) has total energy 4.00×10^{-13} J. Find (a) the kinetic energy of the electron, (b) its Lorentz factor γ, and (c) its speed in terms of c.

VP37.11.3 A moving electron (rest mass 9.11×10^{-31} kg) has momentum $2.60 \times 10^{-22}\ \text{kg} \cdot \text{m/s}$. Find (a) the total energy of the electron, (b) its Lorentz factor γ, and (c) its speed in terms of c.

VP37.11.4 The $\psi(2S)$ meson is an unstable particle with rest energy 3686 MeV. One way in which a $\psi(2S)$ can decay is into a positive K meson (K^+) and a negative K meson (K^-), each with rest energy 495 MeV. Assume the $\psi(2S)$ is at rest before it decays. Find (a) the kinetic energy in MeV, (b) the Lorentz factor γ, and (c) the speed (in terms of c) of each K meson.

BRIDGING PROBLEM Colliding Protons

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

•, ••, •••: 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.

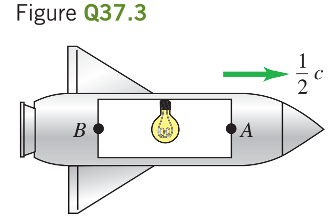
DISCUSSION QUESTIONS

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 • 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

37.8 • At $x = x' = 0$ and $t = t' = 0$ a clock ticks aboard an extremely fast spaceship moving past us in the $+x$-direction with a Lorentz factor of 100 so $v \approx c$. The captain hears the clock tick again 1.00 s later. Where and when do we measure the second tick to occur?

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 ft = 0.3048 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 • 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?

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?

Section 37.5 The Lorentz Transformations

37.14 • The Lorentz coordinate transformation assumes that $t = t'$ at $x = x' = 0$. At what other values of x and x' does $t = t'$?

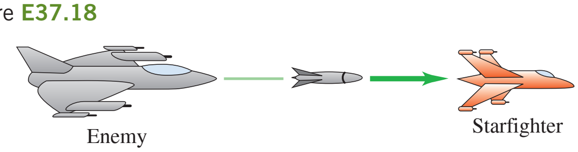
37.15 • An observer in frame S' is moving to the right ($+x$-direction) at speed $u = 0.600c$ away from a stationary observer in frame S. The observer in S' measures the speed v' of a particle moving to the right away from her. What speed v does the observer in S measure for the particle if (a) $v' = 0.400c$; (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 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?

Figure **E37.18**



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?

Section 37.6 The Doppler Effect for Electromagnetic Waves

37.22 • 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?

37.23 • **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 • 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

37.25 • A particle zips by us with a Lorentz factor of 1.12. Then another particle zips by us moving at twice the speed of the first particle. (a) What is the Lorentz factor of the second particle? (b) If the particles were moving with a speed much less than c, the magnitude of the momentum of the second particle would be twice that of the first. However, what is the ratio of the magnitudes of momentum for these relativistic particles?

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 • A spaceship has length 120 m, diameter 25 m, and mass 4.0×10^3 kg as measured by its crew. As the spaceship moves parallel to its cylindrical axis and passes us, we measure its length to be 90 m. (a) What do we measure its diameter to be? (b) What do we measure the magnitude of its momentum to be?

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.

Section 37.8 Relativistic Work and Energy

37.30 • 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?

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 proton's speed?

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 • A particle has rest mass 6.64×10^{-27} kg and momentum 2.10×10^{-18} kg · 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?

37.36 •• 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 • 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.

37.38 • **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).

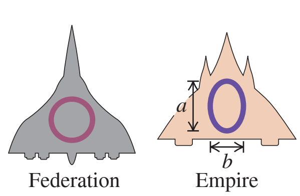
37.39 • (a) Through what potential difference does an electron have to be accelerated, starting from rest, to achieve a speed of $0.980c$? (b) What is the kinetic energy of the electron at this speed? Express your answer in joules and in electron volts.

PROBLEMS

37.40 • 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?

37.41 ••• 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. P37.41**). 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 **P37.41**



37.42 •• 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?

37.43 •• 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?

37.44 •• A $\Sigma+$ particle has a mean lifetime of 80.2 ps. A physicist measures that mean lifetime to be 403 ps as the particle moves in his lab. The rest mass of the particle is 2.12×10^{-27} kg. (a) How fast is the particle moving? (b) How far does it travel, as measured in the lab frame, over one mean lifetime? (c) What are its rest, kinetic, and total energies in the lab frame of reference? (d) What are its rest, kinetic, and total energies in the particle's frame?

37.45 •• Spaceship A moves past the earth at $0.80c$ to the west. Spaceship B approaches A, moving to the east. Both spaceship crews measure their relative speed of approach to be $0.98c$. What mass would the crews of both spaceships measure for the standard kilogram, kept at rest on the earth, (a) according to classical physics and (b) according to the special theory of relativity?

37.46 •• One way to strictly enforce a speed limit would be to alter the laws of nature. Suppose the speed of light were 65 mph and your workplace was 30 miles from your home. Assume you travel to work at a typical driving speed of 60 mph. (a) If you drove at that speed for the round trip to and from work, light, how much would your wristwatch lag your kitchen clock each day? (b) Estimate the length of your car. (c) If you were driving at your estimated driving speed, how long would your car be when viewed from the roadside? (d) What would be the speed relative to you of similar cars traveling toward you in the opposite lane with the same ground speed as you? (e) How long would you measure those cars to be? (f) If the total mass of you and your car was 2000 kg, how much work would be required to get you up to speed? (Note: Your rest mass energy in this world is mc^2, where $c = 65$ mph.) (g) How much work would be required in the real world, where the speed of light is 3.0×10^8 m/s, to get you up to speed?

37.47 • **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.

37.48 •• 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?

37.49 •• **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.)

37.50 •• 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?

37.51 • 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?

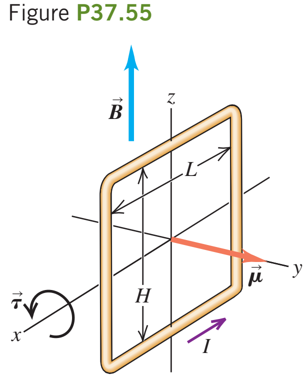
37.52 •• Quarks and gluons are fundamental particles that will be discussed in Chapter 44. A proton, which is a bound state of two up quarks and a down quark, has a rest mass of $m_p = 1.67 \times 10^{-27}$ kg. This is significantly greater than the sum of the rest mass of the up quarks, which is $m_u = 4.12 \times 10^{-30}$ kg each, and the rest mass of the down quark, which is $m_d = 8.59 \times 10^{-30}$ kg. Suppose we (incorrectly) model the rest energy of the proton $m_p c^2$ as derived from the kinetic energy of the three quarks, and we split that energy equally among them. (a) Estimate the Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$ for each of the up quarks using Eq. (37.36). (b) Similarly estimate the Lorentz factor γ for the down quark. (c) Are the corresponding speeds v_u and v_d greater than 99% of the speed of light? (d) More realistically, the quarks are held together by massless gluons, which mediate the strong nuclear interaction. Suppose we model the proton as the three quarks, each with a speed of $0.90c$, with the remainder of the proton rest energy supplied by gluons. In this case, estimate the percentage of the proton rest energy associated with gluons. (e) Model a quark as oscillating with an average speed of $0.90c$ across the diameter of a proton, 1.7×10^{-15} m. Estimate the frequency of that motion.

37.53 • 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?

37.54 •• 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?

37.55 •• CP In a laboratory, a rectangular loop of wire surrounds the origin in the xz-plane, with extent H in the z-direction and extent L in the x-direction (**Fig. P37.55**). The loop carries current I in the counterclockwise direction as viewed from the positive y-axis. A magnetic field $\vec{B} = B\hat{k}$ is present. (a) What is the magnetic dipole moment $\vec{\mu}$ as seen in the frame S of the laboratory? (b) What is the torque $\vec{\tau}$ felt by the current loop? (c) The same loop is viewed from a passing alien spaceship moving with velocity $\vec{v} = v\hat{\imath}$ as seen from the laboratory. From the point of view of the aliens, the loop is length contracted in the direction of this motion. What is the magnetic dipole moment $\vec{\mu}'$ according to the aliens if their coordinate axes are aligned with those of the humans? (d) The torque on the loop is the same when viewed from the laboratory frame and from the spaceship frame. Accordingly, the magnetic field must be frame-dependent. What is the magnetic field $\vec{B}'$ in the spaceship frame S'? (e) If the electromagnetic field is $(\vec{E}, \vec{B}) = (0, \vec{B})$ in an inertial frame S, then in another frame S' moving at velocity $\vec{v}$ relative to S, what is the component of the magnetic field $\vec{B}_\perp$ perpendicular to $\vec{v}$?

Figure **P37.55**



37.56 •• CP A small sphere with charge Q is resting motionless on an insulated pedestal in a laboratory. In the frame of reference S of the laboratory, the z-axis points upward, there is a magnetic field $\vec{B} = -B\hat{\jmath}$, and there is no electric field. (a) What is the net force on the sphere? (b) The same sphere is viewed from a passing spacecraft moving with velocity $\vec{v} = v\hat{\imath}$ as seen from the laboratory. Using the result of Problem 37.55, what is the magnetic field $\vec{B}'$ seen by observers in the spacecraft if their coordinate axes are aligned with those of the laboratory? (c) From the perspective of the spacecraft, the sphere has a nonzero velocity and therefore feels a nonzero magnetic force. What is that magnetic force? (d) The net electromagnetic force on the sphere is necessarily the same when viewed from the laboratory frame and from the spaceship frame. This can be true only if there is a nonzero electric field $\vec{E}'$ in the spaceship frame. Determine $\vec{E}'$. (e) Generalize your conclusion: If the electromagnetic field is $(\vec{E}, \vec{B}) = (0, \vec{B})$ in inertial frame S, then in another frame S' moving at velocity $\vec{v}$ relative to S, what is the component of the electric field $\vec{E}_\perp$ perpendicular to $\vec{v}$?

37.57 • 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?

37.58 •• 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 •• 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?)

37.60 •• Albert in Wonderland. Einstein and Lorentz, avid tennis players, play a fast-paced game on a court where they stand 20.0 m from each other. They play without a net. The tennis ball has mass 0.0580 kg. 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 uses his pocket 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 • CP In a particle accelerator a proton moves at constant speed $0.750c$ in a circle of radius 628 m. What is the net force on the proton?

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 •• **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?

37.64 •• **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** 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?

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.

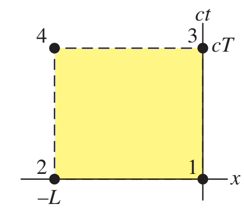
Source	*A*	*B*	*C*	*D*
f (THz)	7.1	5.4	6.1	8.1
f_0 (THz)	9.2	8.6	7.9	8.9

(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?

CHALLENGE PROBLEMS

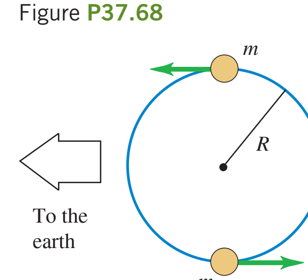
37.67 ••• A fluorescent tube with length L is fixed horizontally above a ticket window in a train station. At time $t = 0$ the tube lights up, changing color from white to bright yellow. After a duration T the tube turns off, reverting to its white color. The tube lies along the x-axis of frame S fixed with respect to the station, with its right end at the origin. We can represent the history of the tube using a "spacetime" diagram, as shown in **Fig. P37.67**. Events 1 and 2 correspond, respectively, to the right and left ends of the tube at time $t = 0$, while events 3 and 4 correspond to the right and left ends of the tube at time $t = T$. The totality of events corresponding to positions on the tube when it is lit correspond to the yellow region on the diagram. The tube is observed from the window of a rocket train passing the station from left to right at speed v. The frame S' is fixed with respect to the train such that the x'-axis coincides with the x-axis of frame S', and such that event 1 occurs at the origin $(x', t') = (0, 0)$. (a) Use the Lorentz transformations to determine the spacetime coordinates (x', t') for events 1, 2, 3, and 4 in the train frame S'. (b) Construct a spacetime diagram analogous to Figure P37.67 showing the tube in the train frame S. Carefully label relevant points, using the economical definition $\gamma = (1 - v^2/c^2)^{-1/2}$ when needed. (c) We use the product ct rather than t on the vertical axis of our figure so that both axes have the dimension of distance. Thus the yellow region has a well-defined area. What is the area of the yellow region in the S frame plot? (d) Compute the area of the yellow region in the S' frame plot you constructed in part (b). (*Hint:* The area of a parallelogram is given by the magnitude of the vector product of two vectors describing adjacent edges.) (e) Is the area of the yellow region the same when we change the frame of reference from S to S'?

Figure **P37.67**



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'.

MCAT-STYLE 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).

Key Example √ARIATION Problems

VP37.5.1 (a) 2.05×10^{-8} s (b) 2.05×10^{-8} s
VP37.5.2 (a) 16.0 s (b) 30.0 s
VP37.5.3 (a) 526 s (b) 4.68×10^{10} m, 164 s
VP37.5.4 (a) 1.28×10^5 km (b) 72.0 m
VP37.7.1 (a) $x' = -8.50 \times 10^8$ m, $t' = 3.78$ s (b) $x = 1.46 \times 10^9$ m, $t = 5.29$ s
VP37.7.2 (a) 1.35×10^8 m (b) 2.25×10^8 m
VP37.7.3 (a) $+0.385c$ (b) $-0.227c$
VP37.7.4 (a) $0.941c$ (b) $0.195c$
VP37.11.1 (a) 1.22×10^{-18} kg · m/s (b) 2.96×10^{12} m/s^2 (c) 2.05×10^{13} m/s^2
VP37.11.2 (a) 3.18×10^{-13} J (b) 4.88 (c) $0.979c$
VP37.11.3 (a) 1.13×10^{-13} J (b) 1.38 (c) $0.689c$
VP37.11.4 (a) 1348 MeV (b) 3.72 (c) $0.963c$

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 individual 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

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 exhibit *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.

LEARNING OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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.

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**, next page). 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

Figure 38.1 The photoelectric effect.

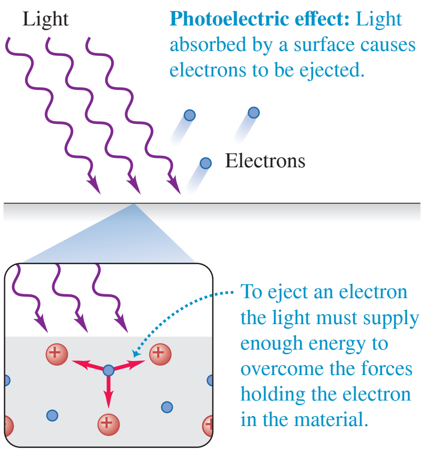
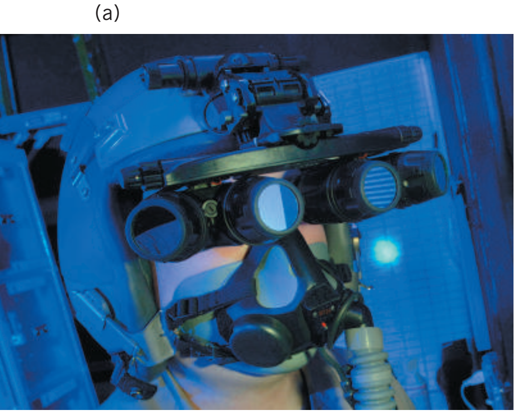


Figure 38.2 (a) A night-vision scope makes use of the photoelectric effect. Photons entering the scope strike a plate, ejecting electrons that pass through a thin disk in which there are millions of tiny channels. The current through each channel is amplified electronically and then directed toward a 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)

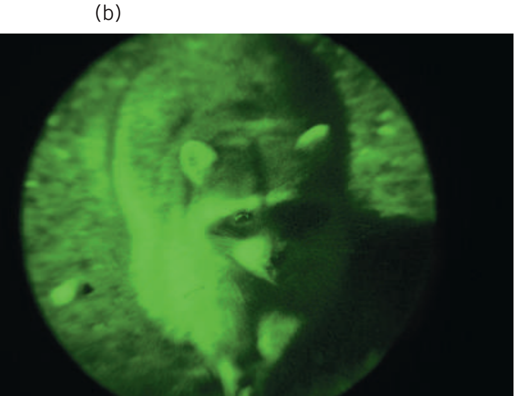


image (**Fig. 38.2**). Sunlight striking the moon causes surface dust to eject electrons, leaving the dust particles with a positive charge. The mutual electric repulsion of these charged dust particles causes them to rise above the moon's surface, a phenomenon that was observed from lunar orbit by the *Apollo* astronauts.

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** 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

$$\begin{aligned} W_{tot} &= -eV_0 = \Delta K = 0 - K_{max} \\ K_{max} &= \tfrac{1}{2}mv_{max}{}^2 = eV_0 \end{aligned} \qquad \text{(maximum kinetic energy of photoelectrons)} \qquad (38.1)$$

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.*

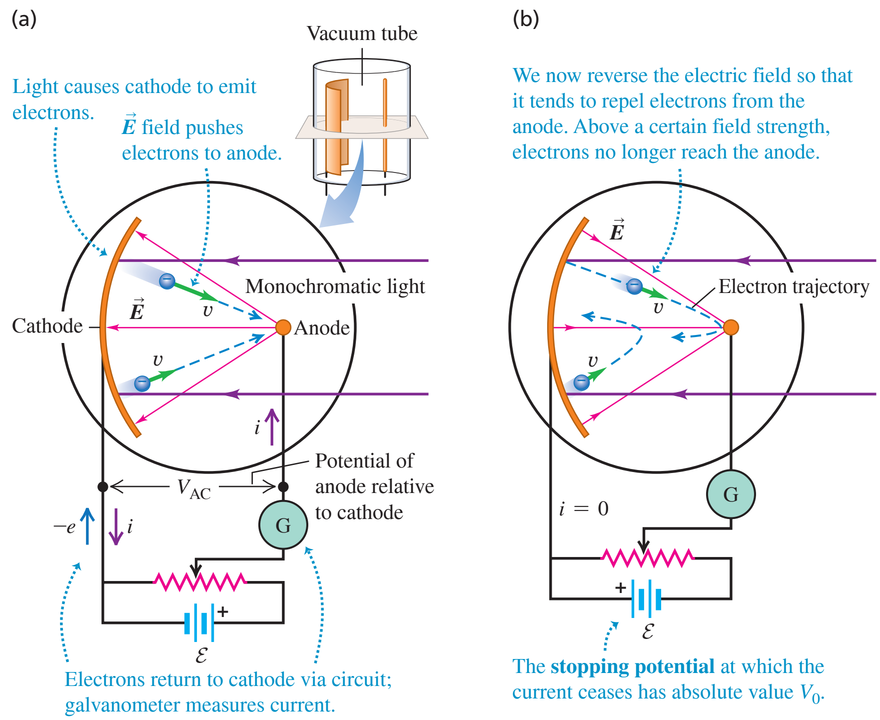


Figure **38.3** An experiment testing whether the photoelectric effect is consistent with the wave model of light.

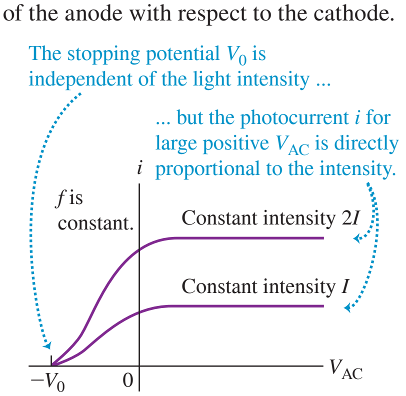
The experimental results proved to be *very* different from these predictions. Here is what was found in the years between 1877 and 1905:

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.

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.

Figure **38.4** Photocurrent i for light frequency f as a function of the potential V_{AC} of the anode with respect to the cathode.



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

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, unlike billiard balls, 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.

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

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. Here **Planck's constant,** h, is a universal constant. In the modern SI system of units (Section 1.3), its value is *defined* to be

$$h = 6.62607015 \times 10^{-34}\ \text{J} \cdot \text{s}$$

In Einstein's picture, all the energy of an individual photon arriving at the surface in Fig. 38.1 is absorbed by a single electron. 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 more photons absorbed per second, and thus more 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:

CAUTION **Greater work function means smaller photoelectron energy** Note what Eq. (38.3) tells us: 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.

$$K_{max} = \tfrac{1}{2}mv_{max}{}^2 = hf - \phi \qquad (38.3)$$

Substituting $K_{max} = eV_0$ from Eq. (38.1), we find

Maximum kinetic energy of photoelectron
Energy of absorbed photon

Photoelectric effect:
$$eV_0 = hf - \phi \qquad (38.4)$$

Work function
Light frequency
Magnitude of electron charge
Stopping potential
Planck's constant

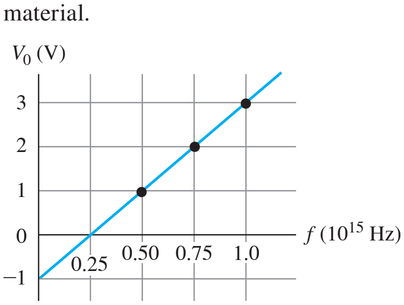


Figure **38.5** Stopping potential as a function of frequency for a particular cathode material.

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,

$$1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}$$

To this accuracy, Planck's constant is

$$h = 6.626 \times 10^{-34}\ \text{J} \cdot \text{s} = 4.136 \times 10^{-15}\ \text{eV} \cdot \text{s}$$

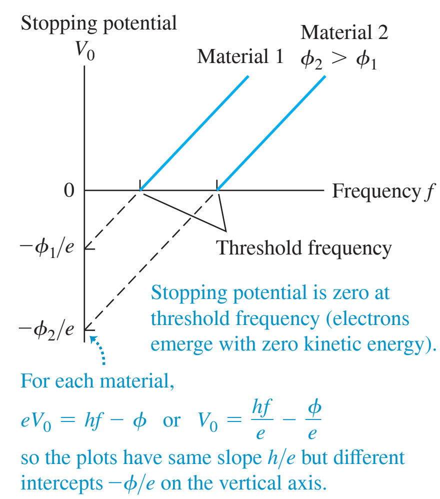
Table 38.1 lists the work functions of several elements. These values are approximate because they are very sensitive to surface impurities. The greater the work function, the higher the minimum frequency needed to emit photoelectrons (**Fig. 38.6**).

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.

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

Figure **38.6** Stopping potential as a function of frequency for two cathode materials having different work functions ϕ.



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

$$\underset{\text{Momentum of a photon}}{p} = \frac{\overset{\text{Photon energy}}{E}}{\underset{\text{Speed of light in vacuum}}{c}} = \frac{\overset{\text{Planck's constant}}{h}\underset{\text{Frequency}}{f}}{c} = \frac{h}{\underset{\text{Wavelength}}{\lambda}} \qquad (38.5)$$

The greater the frequency and the shorter the wavelength of a photon, the greater the magnitude of the photon's momentum. 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: $1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}$. If the photon energy E is given in electron volts, use $h = 4.136 \times 10^{-15}\ \text{eV} \cdot \text{s}$; if E is in joules, use $h = 6.626 \times 10^{-34}\ \text{J} \cdot \text{s}$.

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 $\lambda = 600\ \text{nm}$ and $f = 5 \times 10^{14}\ \text{Hz}$ has an energy E of about 2 eV, or about $3 \times 10^{-19}\ \text{J}$.

EXAMPLE 38.1 Laser-pointer photons

WITH VARIATION PROBLEMS

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?

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

$$\begin{aligned} p &= \frac{h}{\lambda} = \frac{6.626 \times 10^{-34}\text{ J}\cdot\text{s}}{6.50 \times 10^{-7}\text{ m}} \\ &= 1.02 \times 10^{-27}\text{ kg}\cdot\text{m/s} \end{aligned}$$

(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

$$\begin{aligned} E &= pc = (1.02 \times 10^{-27}\text{ kg}\cdot\text{m/s})(3.00 \times 10^{8}\text{ m/s}) \\ &= 3.06 \times 10^{-19}\text{ J} = 1.91\text{ eV} \end{aligned}$$

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):

$$\begin{aligned} E &= hf = \frac{hc}{\lambda} = \frac{(6.626 \times 10^{-34}\text{ J}\cdot\text{s})(3.00 \times 10^{8}\text{ m/s})}{6.50 \times 10^{-7}\text{ m}} \\ &= 3.06 \times 10^{-19}\text{ J} = 1.91\text{ eV} \end{aligned}$$

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.

KEYCONCEPT Photons have both particle properties (energy and momentum) and wave properties (frequency and wavelength). The energy E and the magnitude of momentum p of a photon are both proportional to the frequency and inversely proportional to the wavelength; furthermore, E is directly proportional to p.

EXAMPLE 38.2 A photoelectric-effect experiment

WITH VARIATION PROBLEMS

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.

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

$$\begin{aligned} v_{max} &= \sqrt{\frac{2K_{max}}{m}} = \sqrt{\frac{2(2.00 \times 10^{-19}\text{ J})}{9.11 \times 10^{-31}\text{ kg}}} \\ &= 6.63 \times 10^{5}\text{ m/s} \end{aligned}$$

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.)

KEYCONCEPT In an experiment with the photoelectric effect, not all photoelectrons have the same kinetic energy. The *stopping voltage*—the voltage required to stop photoelectrons emitted at the cathode from reaching the anode—tells you the *maximum* kinetic energy of the photoelectrons.

EXAMPLE 38.3 Determining ϕ and h experimentally

WITH VARIATION PROBLEMS

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.

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'll 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.

KEYCONCEPT To analyze the results of a photoelectric-effect experiment, graph the stopping potential as a function of the frequency of the light used. You can calculate the work function of the material used from the vertical intercept of this graph, and you can calculate the value of Planck's constant from the slope of this graph.

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).

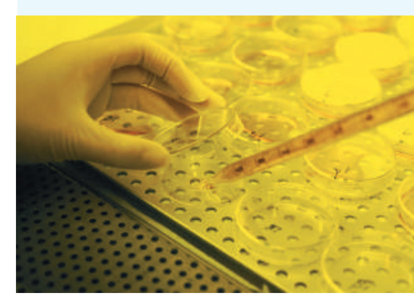
ANSWER

(iv) From Eq. (38.2), a photon of energy $E = 1.14$ eV has wavelength

$$\lambda = hc/E$$
$$= (4.136 \times 10^{-15}\text{ eV}\cdot\text{s})(3.00 \times 10^8\text{ m/s})/(1.14\text{ eV})$$
$$= 1.09 \times 10^{-6}\text{ m} = 1090\text{ nm}$$

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.

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.



38.2 LIGHT EMITTED AS PHOTONS: X-RAY PRODUCTION

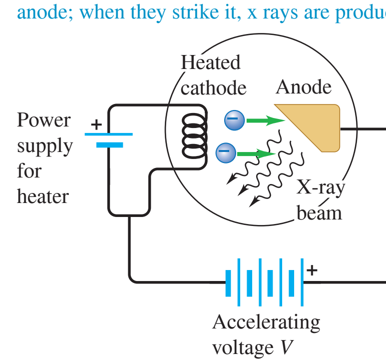
The photoelectric effect provides solid evidence that light is *absorbed* in the form of photons. For physicists to accept Einstein's radical photon concept, however, it was also necessary to show that light is *emitted* as photons. An experiment that demonstrates this convincingly is the inverse of 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.

Figure **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.



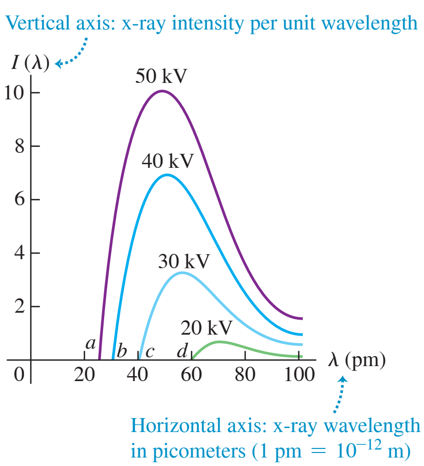
The anode produces x rays in part simply by slowing the electrons abruptly. (Recall from Section 32.1 that accelerated charges emit electromagnetic waves.) This process is called *bremsstrahlung* (German for "braking radiation"). Because the electrons undergo accelerations of very great magnitude, they 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.

Figure **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 λ_{min} for each voltage.



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,

Bremsstrahlung:

$$\overbrace{e}^{\text{Magnitude of electron charge}}\;\overbrace{V_{AC}}^{\text{Accelerating voltage}}\;\text{(Kinetic energy lost by electron)} = \underbrace{h f_{max}}_{\text{Maximum energy of an emitted photon}} = \frac{hc}{\lambda_{min}} \qquad (38.6)$$

Kinetic energy lost by electron; Maximum energy of an emitted photon; Planck's constant (h); Speed of light in vacuum (c); Magnitude of electron charge (e); Accelerating voltage (V_{AC}); Maximum photon frequency (f_{max}); Minimum photon wavelength (λ_{min})

(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 longer 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 41.7.

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.

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.

KEYCONCEPT In bremsstrahlung, a photon of maximum energy and minimum wavelength $\lambda_{\min}$ is produced when all of the kinetic energy of an electron goes into producing that photon. The value of $\lambda_{\min}$ is inversely proportional to the voltage used to accelerate the electron.

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.

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.

Figure 38.9 This radiologist is operating a CT scanner (seen through the window) from a separate room to avoid repeated exposure to x rays.

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.

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.

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.

ANSWER

(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 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 would predict for how light behaves when it undergoes scattering by a single electron, such as an individual electron within an atom.

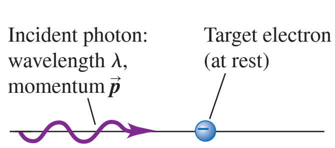
Wave-Model Prediction: In the wave description, scattering would be a process of absorption and re-radiation. Part of the energy of the light wave would be absorbed by the electron, which would oscillate in response to the oscillating electric field of the wave. The oscillating electron would act like 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

Figure 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.

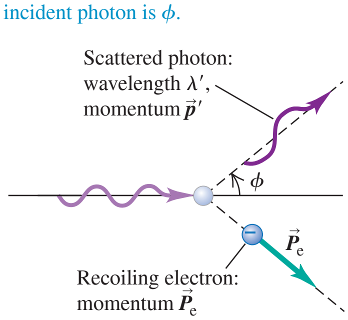
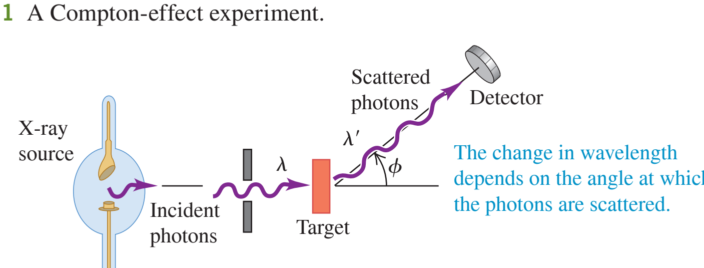


Figure 38.11 A Compton-effect experiment.



discovered that some of the scattered radiation has 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:

$$\underbrace{\lambda'}_{\text{Wavelength of scattered radiation}} - \underbrace{\lambda}_{\text{Wavelength of incident radiation}} = \frac{\overset{\text{Planck's constant}}{h}}{\underset{\text{Electron rest mass}}{m}\,\underset{\text{Speed of light in vacuum}}{c}}(1 - \cos \underbrace{\phi}_{\text{Scattering angle}}) \tag{38.7}$$

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

$$\frac{h}{mc} = \frac{6.626 \times 10^{-34}\ \text{J} \cdot \text{s}}{(9.109 \times 10^{-31}\ \text{kg})(2.998 \times 10^{8}\ \text{m/s})} = 2.426 \times 10^{-12}\ \text{m}$$

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 $\vec{\boldsymbol{p}}$, with magnitude p and energy pc. The scattered photon has momentum $\vec{\boldsymbol{p}}\,'$, 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 $\vec{\boldsymbol{P}}_e$ 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

$$pc + mc^2 = p'c + E_e$$

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 $\vec{\boldsymbol{P}}_e$ from Eq. (38.8) by using momentum conservation. From **Fig. 38.12** we see that $\vec{\boldsymbol{p}} = \vec{\boldsymbol{p}}\,' + \vec{\boldsymbol{P}}_e$, or

$$\vec{\boldsymbol{P}}_e = \vec{\boldsymbol{p}} - \vec{\boldsymbol{p}}\,' \tag{38.9}$$

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).

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.

Figure **38.12** Vector diagram showing conservation of momentum in Compton scattering.

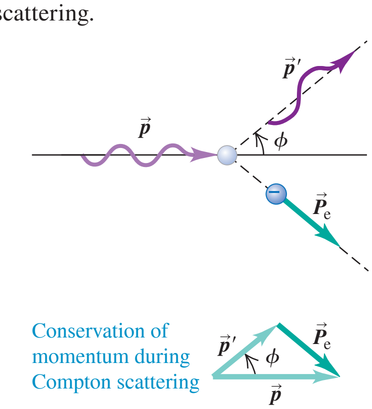
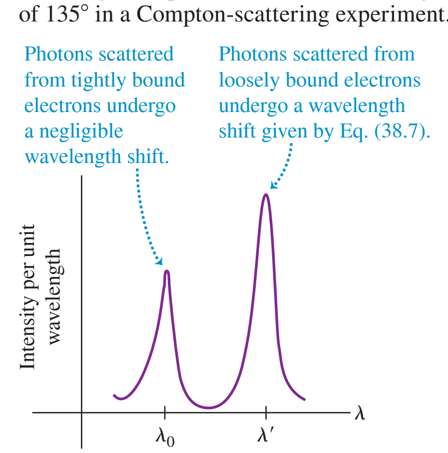


Figure **38.13** Intensity as a function of wavelength for photons scattered at an angle of 135° in a Compton-scattering experiment.



EXAMPLE 38.5 Compton scattering

WITH VARIATION PROBLEMS

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?

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,

$$\cos\phi = 1 - \frac{6.2 \times 10^{-14} \text{ m}}{2.426 \times 10^{-12} \text{ m}} = 0.9744$$

$$\phi = 13.0°$$

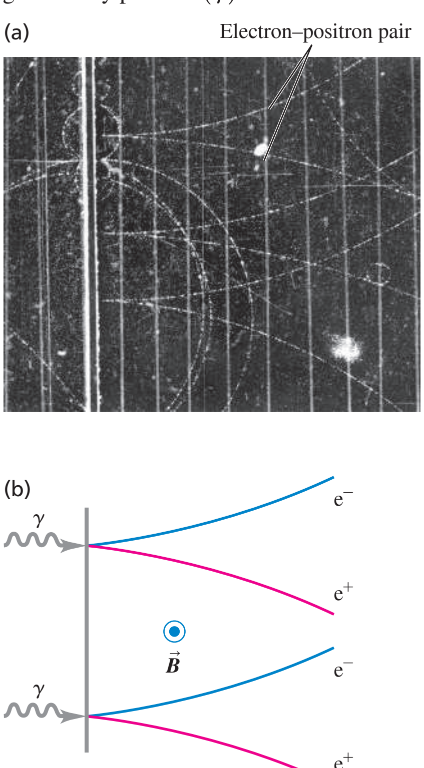
EVALUATE Our results show that smaller scattering angles give smaller wavelength shifts. Thus in a grazing collision the photon energy loss and the electron recoil energy are smaller than when the scattering angle is larger. This is just what we would expect for an elastic collision, whether between a photon and an electron or between two billiard balls.

KEYCONCEPT In Compton scattering, a photon collides with an electron; both energy and momentum are conserved in the collision. If the electron is initially at rest, the photon loses both energy and momentum to the electron. The scattered photon wavelength is longer than the wavelength of the incident photon by an amount $\Delta\lambda$ that depends on the photon scattering angle.

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

Figure 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).



$$E_{\text{min}} = 2mc^2 = 2(9.109 \times 10^{-31} \text{ kg})(2.998 \times 10^8 \text{ m/s})^2$$
$$= 1.637 \times 10^{-13} \text{ J} = 1.022 \text{ MeV}$$

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.

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 $2m_ec^2 = 1.022$ 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

WITH VARIATION PROBLEMS

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 rest energy of an electron or a positron is 0.511 MeV.

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 electron and the positron each have energy $K + mc^2$, 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 or positron 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 0.225 pm and the frequency is 1.33×10^{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.

KEYCONCEPT In *pair annihilation,* an electron and a positron collide and disappear, and they are replaced by two or more photons. Energy and momentum are conserved in pair annihilation.

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?

ANSWER

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 $h/mc = 2.426 \times 10^{-12}\ \text{m} = 0.002426$ 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).

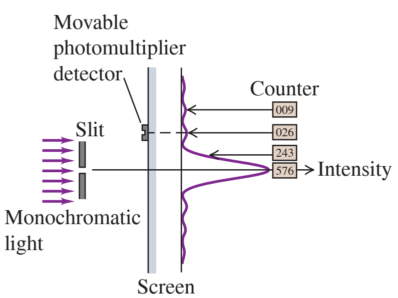
38.4 WAVE–PARTICLE DUALITY, PROBABILITY, AND UNCERTAINTY

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'll never need to use both at the same time to describe a single part of an occurrence.

Diffraction and Interference in the Photon Picture

Figure 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.

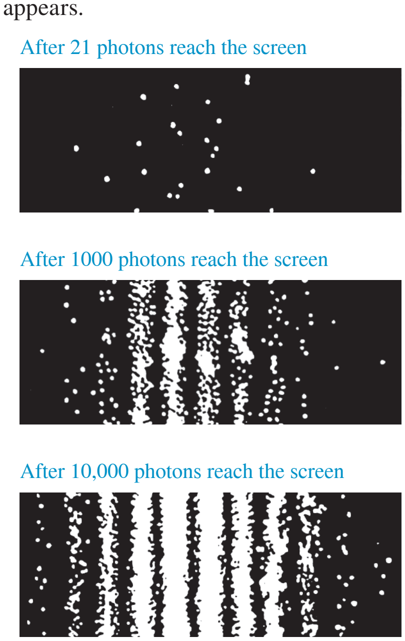


Let's start by considering again the diffraction pattern for a single slit, which we analyzed in Sections 36.2 and 36.3. Instead of recording the pattern on a digital camera chip or photographic film, we use a detector called a *photomultiplier* that can actually detect individual photons. Using the 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.

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.

Figure 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'll 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. 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 $\sin\theta_1 = \lambda/a$.

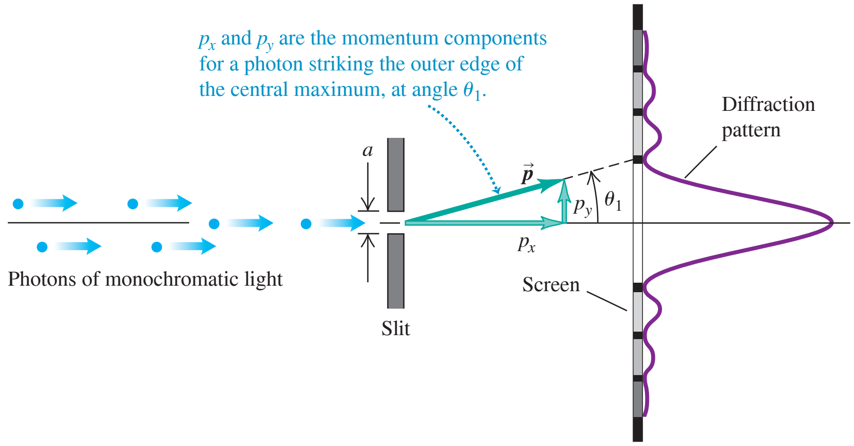


Figure **38.17** Interpreting single-slit diffraction in terms of photon momentum.

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

$$\theta_1 = \frac{\lambda}{a} \qquad (38.12)$$

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

$$p_y = p_x\theta_1 \qquad (38.13)$$

Substituting Eq. (38.12), $\theta_1 = \lambda/a$, into Eq. (38.13) gives

$$p_y = p_x\frac{\lambda}{a} \qquad (38.14)$$

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 $(p_y)_{av} = 0$. There will be an *uncertainty* Δp_y in the y-component of momentum at least as great as $p_x\lambda/a$. That is,

$$\Delta p_y \geq p_x\frac{\lambda}{a} \qquad (38.15)$$

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

$$\Delta p_y \geq p_x\frac{h}{p_x a} = \frac{h}{a}$$

$$\Delta p_y a \geq h \qquad (38.16)$$

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

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!

Heisenberg uncertainty principle for position and momentum:

Uncertainty in coordinate x — Planck's constant divided by 2π

$$\Delta x \, \Delta p_x \geq \hbar/2 \tag{38.17}$$

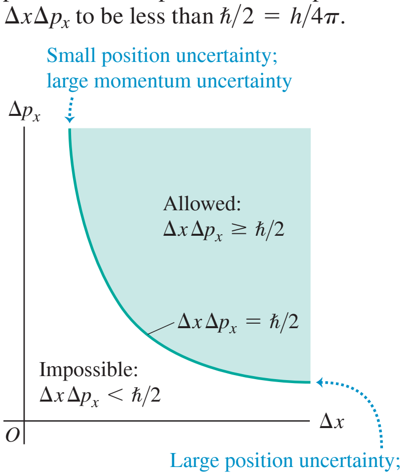
Uncertainty in corresponding momentum component p_x

The quantity $\hbar$ (pronounced "h-bar") is Planck's constant divided by 2π. To four significant figures,

$$\hbar = \frac{h}{2\pi} = 1.055 \times 10^{-34}\ \mathrm{J \cdot s} = 6.582 \times 10^{-16}\ \mathrm{eV \cdot s}$$

We'll use $\hbar$ frequently to avoid writing a lot of factors of 2π in later equations.

Figure 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 $\hbar/2 = h/4\pi$.



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'll 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 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 different values E_1 and E_2 of energy. The total wave function is

$$E_y(x, t) = A_1 \sin[(p_{1x}x - E_1 t)/\hbar] + A_2 \sin[(p_{2x}x - E_2 t)/\hbar] \tag{38.21}$$

Consider what this wave function looks like at a particular instant of time, say, $t = 0$. At this instant Eq. (38.21) becomes

$$E_y(x, t = 0) = A_1 \sin(p_{1x}x/\hbar) + A_2 \sin(p_{2x}x/\hbar) \tag{38.22}$$

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!

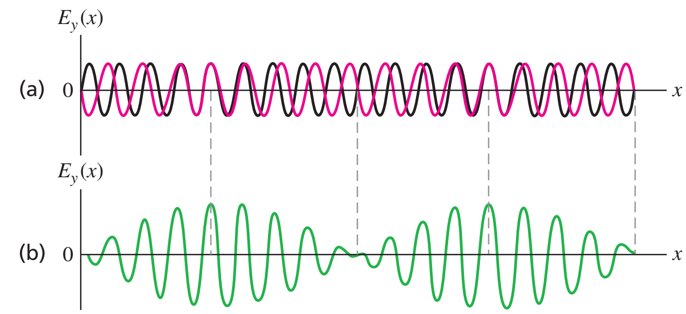


Figure **38.19** (a) Two sinusoidal waves with slightly different wave numbers k and hence slightly different values of momentum $p_x = \hbar k$ shown at one instant of time. (b) The superposition of these waves has a momentum equal to the average of the two individual values of momentum. The amplitude varies, giving the total wave a lumpy character not possessed by either individual wave.

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

$$\begin{aligned} E_y(x, t) &= A_1 \sin(-E_1 t/\hbar) + A_2 \sin(-E_2 t/\hbar) \\ &= -A_1 \sin(E_1 t/\hbar) - A_2 \sin(E_2 t/\hbar) \end{aligned} \tag{38.23}$$

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'll 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 relationship, 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:

$$\overbrace{\Delta t}^{\text{Time uncertainty of a phenomenon}} \underbrace{\Delta E}_{\text{Energy uncertainty of same phenomenon}} \geq \overbrace{\hbar/2}^{\text{Planck's constant divided by } 2\pi} \tag{38.24}$$

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

WITH VARIATION PROBLEMS

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.

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

$$\begin{aligned} E &= hf = (6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(3.75 \times 10^{14}\ \text{Hz}) \\ &= 2.48 \times 10^{-19}\ \text{J} \end{aligned}$$

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:

$$\begin{aligned} \Delta x &= c\Delta t = (3.00 \times 10^{8}\ \text{m/s})(4.00 \times 10^{-15}\ \text{s}) \\ &= 1.20 \times 10^{-6}\ \text{m} \\ \Delta x &= \frac{1.20 \times 10^{-6}\ \text{m}}{8.00 \times 10^{-7}\ \text{m/wavelength}} = 1.50\ \text{wavelengths} \end{aligned}$$

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) the 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.

KEYCONCEPT The Heisenberg uncertainty principle takes two forms. First, the smaller the uncertainty in the *location* of a physical system, the greater the uncertainty in the *momentum* of the system. Second, the smaller the uncertainty in the *time* of a physical phenomenon, the greater the uncertainty in the *energy* of the phenomenon.

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.

ANSWER

(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.

CHAPTER 38 SUMMARY

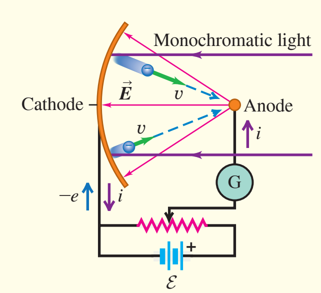
Photons: Electromagnetic radiation behaves as both waves and particles. The energy in an electromagnetic wave is carried in units called photons. The energy E of one photon is proportional to the wave frequency f and inversely proportional to the wavelength λ, and is proportional to a universal quantity h called Planck's constant. The momentum of a photon has magnitude E/c. (See Example 38.1.)

$$E = hf = \frac{hc}{\lambda} \qquad (38.2)$$

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda} \qquad (38.5)$$

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.)

$$eV_0 = hf - \phi \qquad (38.4)$$



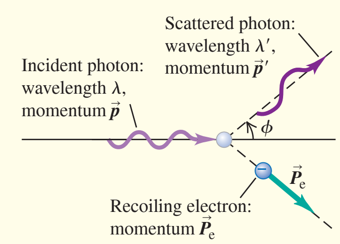
Photon production, photon scattering, and pair production: X rays can be produced when electrons accelerated to high kinetic energy across a potential increase V_{AC} strike a target. The photon model explains why the maximum frequency and minimum wavelength produced are given by Eq. (38.6). (See Example 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.)

$$eV_{AC} = hf_{max} = \frac{hc}{\lambda_{min}} \qquad (38.6)$$

(bremsstrahlung)

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\phi) \qquad (38.7)$$

(Compton scattering)



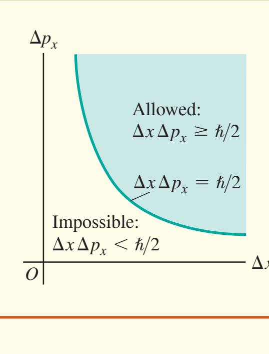
The Heisenberg uncertainty principle: It is impossible to determine both a photon's position and its momentum at the same time to arbitrarily high precision. The precision of such measurements for the x-components is limited by the Heisenberg uncertainty principle, Eq. (38.17); there are corresponding 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.)

$$\Delta x\,\Delta p_x \geq \hbar/2 \qquad (38.17)$$

(Heisenberg uncertainty principle for position and momentum)

$$\Delta t\,\Delta E \geq \hbar/2 \qquad (38.24)$$

(Heisenberg uncertainty principle for energy and time)



GUIDED PRACTICE

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 38.1, 38.2, and 38.3 (Section 38.1) before attempting these problems.

VP38.3.1 A laser emits 2.13×10^{16} photons per second, each of which has wavelength 625 nm. Find (a) the energy and momentum of a single photon and (b) the power output of the laser.

VP38.3.2 A certain metal has a work function of 4.55 eV. For this metal, find (a) the minimum frequency that light must have to produce photoelectrons and (b) the frequency that light must have to produce photoelectrons with maximum kinetic energy 1.53 eV.

VP38.3.3 In a photoelectric-effect experiment, the stopping potential is 1.37 V if the light used to illuminate the cathode has wavelength 475 nm. Find (a) the work function (in eV) of the cathode material and (b) the stopping potential if the wavelength is decreased to 425 nm.

VP38.3.4 Your results from a photoelectric-effect experiment show that the maximum speed of the emitted photoelectrons is 6.95×10^5 m/s when the cathode is illuminated with ultraviolet light of wavelength 306 nm. Find (a) the maximum photoelectron kinetic energy in eV, (b) the photon energy in eV, and (c) the work function.

Be sure to review EXAMPLES 38.5 and 38.6 (Section 38.3) before attempting these problems.

VP38.6.1 You use x-ray photons of wavelength 0.251 nm in a Compton-scattering experiment. At a scattering angle of 14.5°, what is the increase in the wavelength of the scattered x rays compared to the incident x rays? Express your results (a) in nm and (b) as a percentage of the wavelength of the incident x rays.

VP38.6.2 You use x rays of *frequency* 1.260×10^{18} Hz in a Compton-scattering experiment. Find the frequency of the scattered x rays if the scattering angle is (a) 90.00° and (b) 180.0°.

VP38.6.3 Compton scattering can also occur if a photon collides with a proton (mass 1.673×10^{-27} kg) at rest. If you do an experiment of this kind using gamma-ray photons with wavelength 2.50×10^{-12} m = 2.50 pm, for what scattering angle is the wavelength of the scattered gamma rays longer than the wavelength of the incident gamma rays by (a) 0.0100% and (b) 0.0800%?

VP38.6.4 An antiproton has the same mass and rest energy (938.3 MeV) as a proton, but has a negative change $-e$ instead of the positive charge $+e$ of the proton. A proton and an antiproton, initially far apart, move toward each other with the same speed and collide head-on, annihilating each other and producing two photons. Find the energies and wavelengths of the photons if the initial kinetic energies of the proton and antiproton are (a) both negligible and (b) both 545 MeV.

Be sure to review EXAMPLE 38.7 (Section 38.4) before attempting these problems.

VP38.7.1 A laser produces light of wavelength 633 nm in pulses that propagate in the $+x$-direction and that are 7.00×10^{-6} m in length. For an average photon in this pulse, find (a) the momentum p, (b) the minimum momentum uncertainty in kg · m/s and as a percentage of p, (c) the energy E, and (d) the minimum energy uncertainty in J and as a percentage of E.

VP38.7.2 Weather radar systems emit radio waves in pulses. For a typical system the frequency used is 3.00 GHz (1 GHz = 10^9 Hz) and the minimum energy uncertainty of the radio photons is 5.50×10^{-29} J. Find (a) the time duration of each pulse, (b) the length of each pulse, (c) the energy of an average photon, and (d) the minimum frequency uncertainty.

VP38.7.3 In a pulsed laser the minimum energy uncertainty of a photon is 2.33×10^{-20} J, which is 6.70% of the average energy of a photon in the pulse. Find (a) the wavelength and frequency of the laser light, (b) the minimum frequency uncertainty, and (c) the time duration of the pulse.

VP38.7.4 The average energy of a photon in a pulsed laser beam is 2.39 eV, with a minimum uncertainty of 0.0155 eV. Each pulse has an average of 5.00×10^{12} photons. Find (a) the time duration of each pulse, (b) the wavelength of the light, and (c) the energy per pulse in J.

BRIDGING PROBLEM Compton Scattering and Electron Recoil

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

•, ••, •••: 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 $(3 \times 10^8 \text{ m/s})(10^{-12} \text{ s}) = 3 \times 10^{-4} \text{ m} = 0.3$ 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 $8.24 \times 10^{-28} \text{ kg} \cdot \text{m/s}$. (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 •• 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.

38.7 •• 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?

38.8 •• 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.

38.9 • 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

38.10 • 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?

38.11 • An electron accelerates through a potential difference of 50.0 kV in an x-ray tube. When the electron strikes the target, 70.0% of its kinetic energy is imparted to a single photon. Find the photon's frequency, wavelength, and magnitude of momentum.

38.12 •• (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

38.13 • 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?

38.14 • 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)?

38.15 •• 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?

38.16 •• 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?

38.17 •• 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.

38.18 •• 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?

38.19 • A beam of photons with just barely enough individual photon energy to create electron–positron pairs undergoes Compton scattering from free electrons before any pair production occurs. (a) Can the Compton-scattered photons create electron–positron pairs? (b) Calculate the momentum magnitude of the photons found at a 20.0° angle from the direction of the original beam.

38.20 • 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

38.21 • 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?

38.22 • 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.

38.23 • 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

38.24 •• (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?

38.25 •• 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 J/kg·K, 2.256×10^6 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?

38.26 • 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)?

38.27 •• CP We can estimate the number of photons in a room using the following reasoning: The intensity of the sun's rays at the earth's surface is roughly 1000 W/m^2. A typical room is illuminated by indirect light or by light bulbs so that its light intensity is some fraction of that value. (a) Estimate the intensity of the light that enters your room. (b) Model the light as entering uniformly at the ceiling and exiting uniformly at the floor. Estimate the area A of the floor and ceiling, and the height H of your room. (c) Estimate how long it takes light to travel from the ceiling to the floor of your room by dividing H by the speed of light. (d) Estimate the total power of the light that enters your room P by multiplying your estimated intensity I by the area of your ceiling A. (e) Estimate the total light energy in your room E_{room} by multiplying the power P by the length of time it takes light to travel from ceiling to floor. (f) An average wavelength for light is in the middle of the visible spectrum, at roughly 500 nm. What is the energy of a 500 nm photon? (g) The total number of photons in your room N is the ratio of the energy E_{room} to the energy per photon. What is your estimate for N?

38.28 •• 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?

38.29 •• 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?

38.30 •• CP All of the nutritional energy contained in fruits and vegetables is supplied by solar photons through photosynthesis reactions. (a) Estimate the amount of energy stored in a tomato by considering the number of Calories in that fruit. (*Note:* 1 Cal = 1 kcal = 4186 J, and you can look up the Calorie content of a tomato.) (b) Estimate the total leaf area on an average tomato plant. (c) If the solar intensity in your garden is 800 W/m^2, then what is the power supplied by solar photons to those leaves? (d) Fruit leaves have approximately 5% efficiency. The majority, 95%, of photons do not land on chloroplasts or lie outside the frequency range for photosynthesis. Use this information and an average wavelength of 600 nm to determine the number of solar photons involved in photosynthesis reactions each second on a tomato plant. (e) If an average tomato plant has ten tomatoes, and if half of the photon energy enters the fruit while the rest is used to grow roots and maintain plant cells, then how many photons supply energy to each tomato each second? (f) How many photons are needed to supply the energy content of each tomato? (g) Using these results and assuming the sun shines on the tomatoes 12 hours per day, estimate how long it should take to grow a tomato.

38.31 •• 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.)

38.32 ••• CP A positron with speed v impinges horizontally upon an electron at rest. These particles annihilate and produce two photons with the same wavelength λ, each of which travels at angle ϕ with respect to the horizontal. (a) What is the condition for energy conservation in terms of m, γ, c, and h, where m is the electron mass and $\gamma = (1 - v^2/c^2)^{-1/2}$? (b) What is the condition for momentum conservation in terms of the same parameters and the angle ϕ? (c) Solve these equations and determine the energy of each photon E_{photon} and the angle ϕ, the latter as a function of only γ. (d) If the positron had 5.11 MeV of kinetic energy, then what is the energy of each photon, and what is the angle ϕ? (e) What positron speed would result in the photons traveling off perpendicular to each other? (f) If the positron with speed v was traveling in the $+x$-direction, then with what speed u should we perform a Lorentz velocity transformation to bring us to the "center of momentum" frame, where the total momentum is zero?

38.33 •• 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?

38.34 •• 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.

38.35 •• 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?

38.36 •• 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?

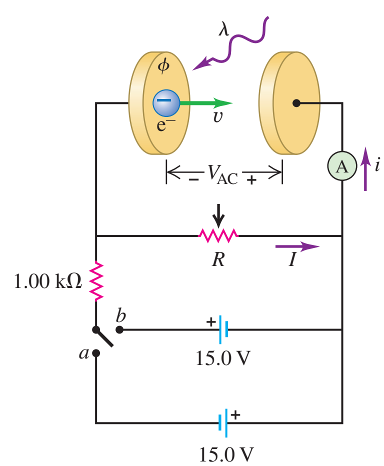
38.37 •• DATA To test the photon concept, you perform a Compton-scattering experiment in a research lab. Using photons of very short wavelength, you measure the wavelength λ' of scattered photons as a function of the scattering angle ϕ, the angle between the direction of a scattered photon and the incident 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.

38.38 ••• **CP** The work function of a material is examined using the apparatus shown schematically in **Fig. P38.38**. The ammeter measures the photocurrent. The voltage of the anode relative to the cathode is V_{AC}. (a) The switch is in position a while radiation with wavelength 140 nm is incident on the cathode. The potentiometer is set at $R = 3.20\ \text{k}\Omega$, and there is no photocurrent. What is the potential difference V_{AC}? (b) The resistance R is slowly adjusted to lower values until the first hint of a photocurrent is detected in the ammeter. That happens when $R = 334\ \Omega$. What is the value of V_{AC} at that point? (c) What is the work function ϕ of the cathode material? (d) The switch is then thrown to position b and the ammeter measures a photocurrent $i = 3.71$ mA. What is the current I in this case? (e) What is the value of V_{AC}? (f) The switch is returned to position a, and the wavelength of the light is reduced to 65.0 nm. The photocurrent increases. What is the minimum value of R for which the photocurrent will cease?

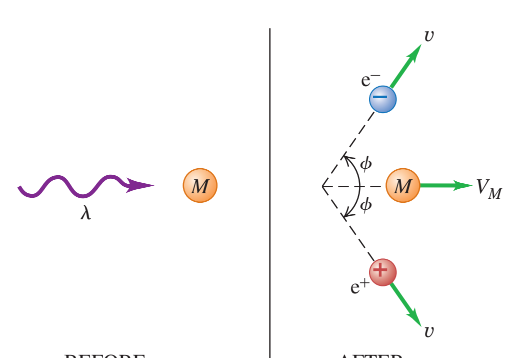
Figure **P38.38**



CHALLENGE PROBLEMS

38.39 ••• **CP** A photon with wavelength λ is incident on a stationary particle with mass M, as shown in **Fig. P38.39**. The photon is annihilated while an electron–positron pair is produced. The target particle moves off in the original direction of the photon with speed V_M. The electron travels with speed v at angle ϕ with respect to that direction. Owing to momentum conservation, the positron has the same speed as the electron. (a) Using the notation $\gamma = (1 - v^2/c^2)^{-1/2}$ and $\gamma_M = (1 - V_M^2/c^2)^{-1/2}$, write a relativistic expression for energy conservation in this process. Use the symbol m for the electron mass. (b) Write an analogous expression for momentum conservation.

Figure **P38.39**



(c) Eliminate the ratio h/λ between your energy and momentum equations to derive a relationship between v/c, V_M/c, and ϕ in terms of the ratio m/M, keeping γ and γ_M as a useful shorthand notation. (d) Consider the case where $V_M \ll c$, and use a binomial expansion to derive an expression for γ_M to the first order in V_M. Use that result to rewrite your previous result as an expression for V_M in terms of c, v, m/M, and ϕ. (e) Are there choices of v and ϕ for which $V_M = 0$? (f) Suppose the target particle is a proton. If the electron and positron remain stationary, so that $v = 0$, then with what speed does the proton move, in km/s? (g) If the electron and positron each have total energy 5.00 MeV and move with $\phi = 60°$, then what is the speed of the proton? (*Hint:* First solve for γ and v.) (h) What is the energy of the incident photon in this case?

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 ($\lambda = 10.6\ \mu\text{m}$) collides head-on with a beam of electrons, each of total energy $E = 10.0$ GeV ($1\ \text{GeV} = 10^9$ 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?

MCAT-STYLE 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.

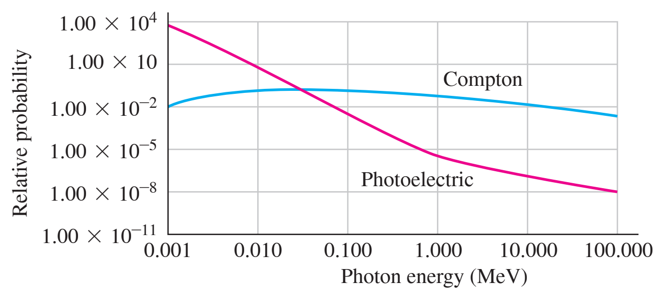
38.41 How much energy is imparted to one cell during one day's treatment? Assume that the specific gravity of the tumor is 1 and that $1\ \text{J} = 6 \times 10^{18}$ eV. (a) 120 keV; (b) 12 MeV; (c) 120 MeV; (d) 120×10^3 MeV.

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 ?

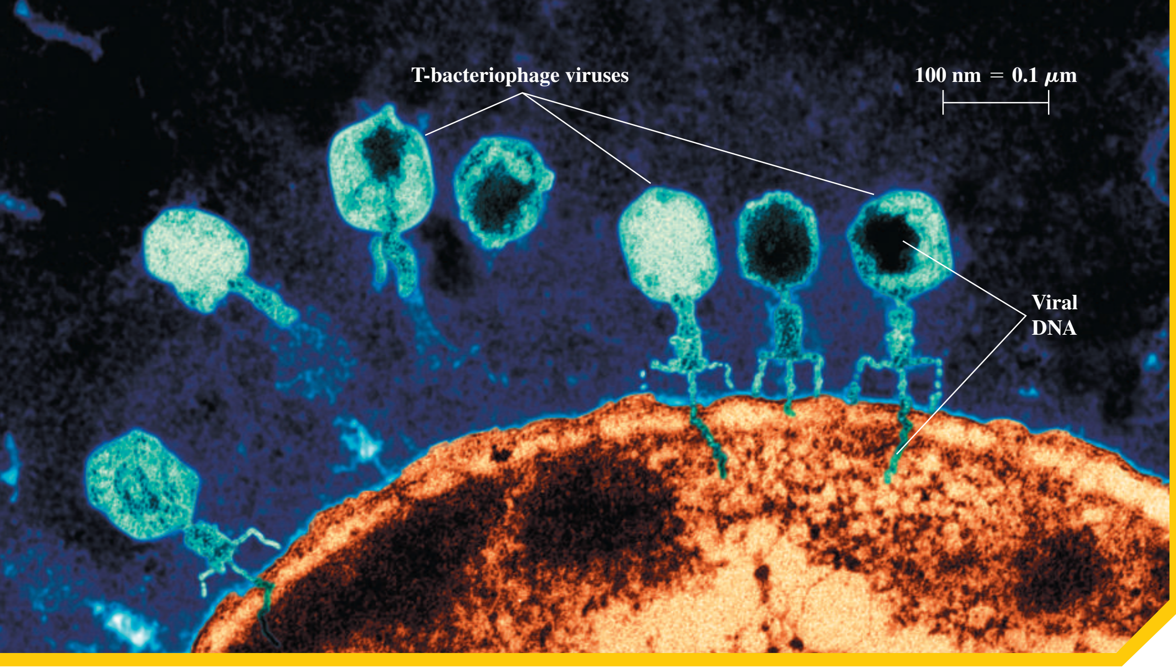
(i) The energy of a photon E is inversely proportional to its wavelength λ: The shorter the wavelength, the more energetic is the photon. Since visible light has shorter wavelengths than infrared light, the headlamp emits photons of greater energy. However, the light from the infrared laser is far more *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.

Key Example √ARIATION Problems

VP38.3.1 (a) $E = 3.18 \times 10^{-19}$ J, $p = 1.06 \times 10^{-27}$ kg · m/s (b) 6.77×10^{-3} W
VP38.3.2 (a) 1.10×10^{15} Hz (b) 1.47×10^{15} Hz
VP38.3.3 (a) 1.24 eV (b) 1.68 V
VP38.3.4 (a) 1.37 eV (b) 4.05 eV (c) 2.68 eV
VP38.6.1 (a) 7.73×10^{-5} nm (b) 0.0308%
VP38.6.2 (a) 1.247×10^{18} Hz (b) 1.235×10^{18} Hz
VP38.6.3 (a) 35.8° (b) 121°
VP38.6.4 (a) $E = 938.3$ MeV, $\lambda = 1.32 \times 10^{-15}$ m (b) $E = 1483.3$ MeV, $\lambda = 8.36 \times 10^{-16}$ m
VP38.7.1 (a) 1.05×10^{-27} kg · m/s (b) 7.54×10^{-30} kg · m/s, 0.720% (c) 3.14×10^{-19} J (d) 2.26×10^{-21} J, 0.720%
VP38.7.2 (a) 9.59×10^{-7} s (b) 288 m (c) 1.99×10^{-24} J (d) 8.30×10^4 Hz = 8.30×10^{-5} GHz
VP38.7.3 (a) $\lambda = 572$ nm, $f = 5.25 \times 10^{14}$ Hz (b) 3.52×10^{13} Hz (c) 2.26×10^{-15} s
VP38.7.4 (a) 2.12×10^{-14} s (b) 519 nm (c) 1.91×10^{-6} J = 1.91 μJ

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

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 $p = mv$ in exactly the same way

LEARNING OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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.

Figure 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.

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

$$\underbrace{\lambda}_{\text{De Broglie wavelength of a particle}} = \frac{\overbrace{h}^{\text{Planck's constant}}}{\underbrace{p}_{\text{Particle's momentum}}} = \frac{h}{\underbrace{m}_{\text{Particle's mass}}\overbrace{v}^{\text{Particle's speed}}} \tag{39.1}$$

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:

$$\underbrace{E}_{\text{Energy of a particle}} = \overbrace{h}^{\text{Planck's constant}}\underbrace{f}_{\text{Frequency}} \tag{39.2}$$

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.

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 planes that were continuous

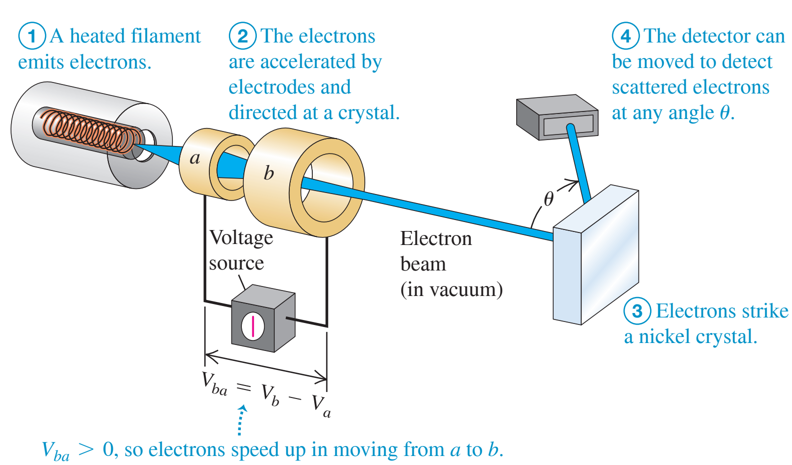


Figure 39.2 An apparatus similar to that used by Davisson and Germer to discover electron diffraction.

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:

$$\underbrace{\lambda}_{\text{De Broglie wavelength of an electron}} = \frac{\overbrace{h}^{\text{Planck's constant}}}{\underbrace{p}_{\text{Electron momentum}}} = \frac{h}{\sqrt{2\underbrace{m}_{\text{Electron mass}}\underbrace{e}_{\text{Magnitude of electron charge}}\underbrace{V_{ba}}_{\text{Accelerating voltage}}}} \qquad (39.3)$$

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

$$d\sin\theta = m\lambda \qquad (m = 1, 2, 3, \dots) \qquad (39.4)$$

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 electron and x-ray diffraction patterns for polycrystalline metal foils. (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.)

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.

Figure **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 $m = 1$.

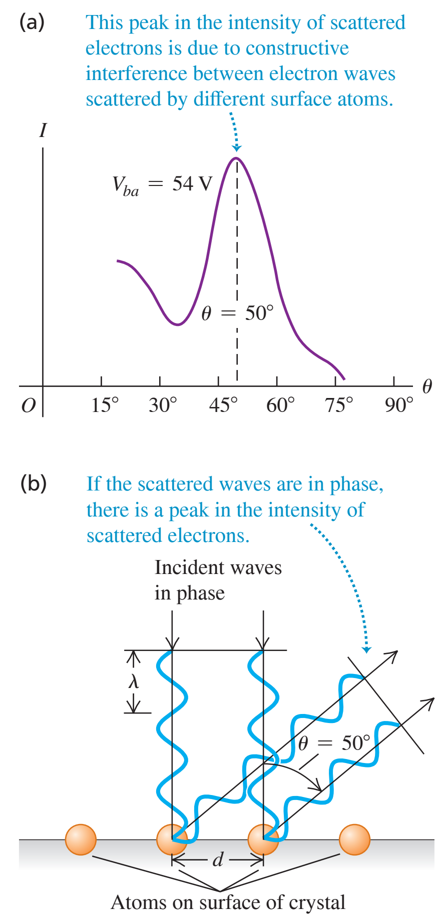
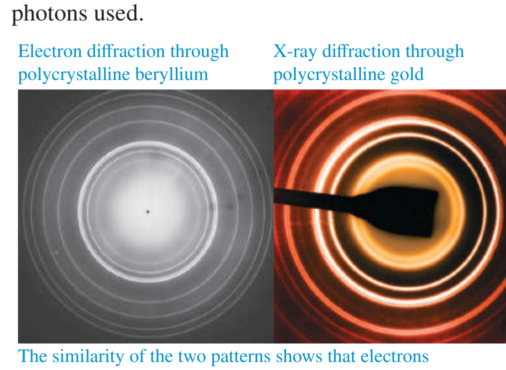


Figure **39.4** When electrons of a single momentum are sent through a polycrystalline metal foil, they produce the same kind of diffraction pattern as do x rays (shown here in enhanced color) sent through the same kind of foil. For both kinds of diffraction, the angles of the maxima depend on the wavelengths of the electrons or photons used.



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'll 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

WITH VARIATION PROBLEMS

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.

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.

KEYCONCEPT Electrons and other subatomic particles have both particle properties (kinetic energy and momentum) and wave properties (frequency and wavelength). Like a photon, the wavelength of a particle is inversely proportional to its momentum; unlike a photon, the wavelength of a nonrelativistic particle is inversely proportional to the square root of its kinetic energy.

EXAMPLE 39.2 Energy of a thermal neutron

WITH VARIATION PROBLEMS

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 ($T = 20°\text{C} = 293\ \text{K}$).

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 $T = 293\ \text{K}$ 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).

KEYCONCEPT For a nonrelativistic particle, the speed is proportional to the momentum and hence inversely proportional to the particle's wavelength.

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 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

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)?

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}:

$$\begin{aligned} V_{ba} &= \frac{h^2}{2me\lambda^2} \\ &= \frac{(6.626 \times 10^{-34}\,\mathrm{J \cdot s})^2}{2(9.109 \times 10^{-31}\,\mathrm{kg})(1.602 \times 10^{-19}\,\mathrm{C})(10 \times 10^{-12}\,\mathrm{m})^2} \\ &= 1.5 \times 10^4\ \mathrm{V} = 15{,}000\ \mathrm{V} \end{aligned}$$

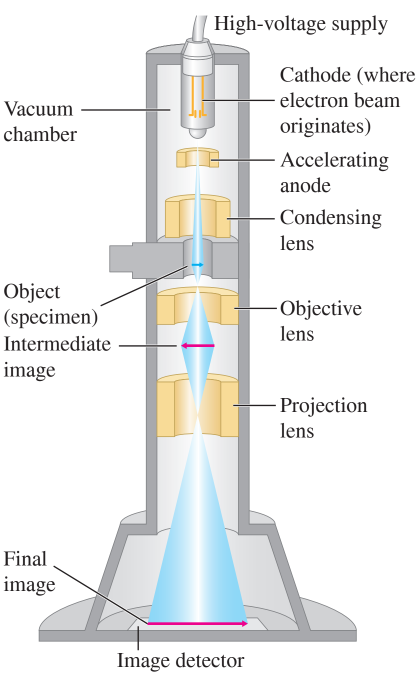
EVALUATE It is easy to attain 15 kV accelerating voltages from 120 V or 240 V line voltage by using a step-up transformer (Section 31.6) and a rectifier (Section 31.1). The accelerated electrons have kinetic energy 15 keV; since the electron rest energy is 0.511 MeV = 511 keV, these electrons are indeed nonrelativistic.

KEYCONCEPT The kinetic energy K of an electron equals the magnitude e of its charge multiplied by the voltage V_{ba} required to accelerate it from rest. If the electron is nonrelativistic, both K and V_{ba} are inversely proportional to the square of the particle's wavelength.

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.

Figure 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?

ANSWER

(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.

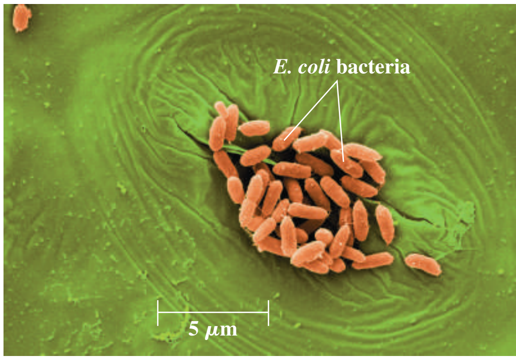


Figure 39.6 This scanning electron microscope image shows *Escherichia coli* bacteria crowded into a stoma, or respiration opening, on the surface of a lettuce leaf. (False color has been added.) If not washed off before the lettuce is eaten, these bacteria can be a health hazard. The *transmission* 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'll 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 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.

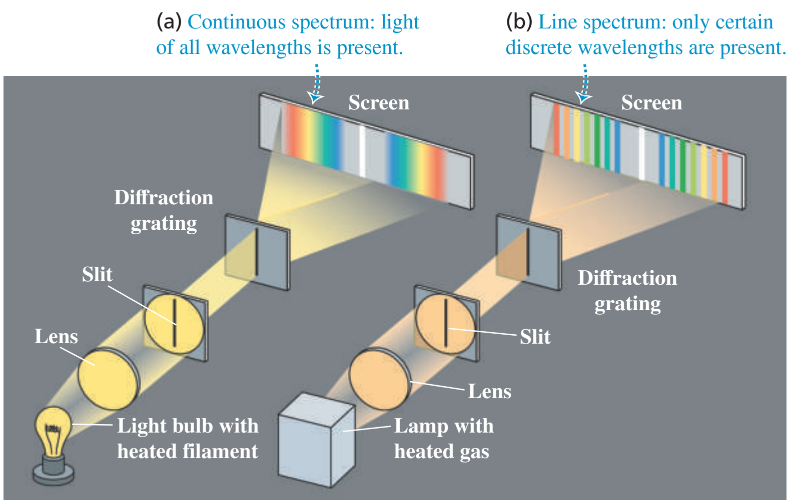
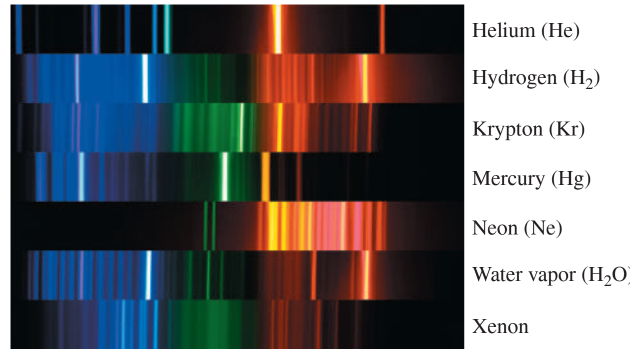


Figure 39.7 (a) Continuous spectrum produced by a glowing light bulb filament. (b) Emission line spectrum emitted by a lamp containing a heated gas.

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.

Figure 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.



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.

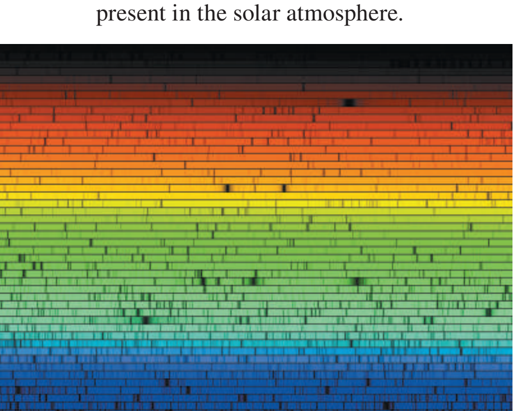
While a *heated* gas selectively *emits* only certain wavelengths, a *cool* gas selectively *absorbs* certain wavelengths. If we pass white (continuous-spectrum) light through a gas and look at the *transmitted* light with a spectrometer, we find a series of dark lines corresponding to the wavelengths 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 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.)

Figure 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.



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 foils 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.

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**, next page). The results of the Rutherford experiments were *very* different from the Thomson prediction.

Figure 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. Rutherford is shown here on the New Zealand $100 banknote.

Figure 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.

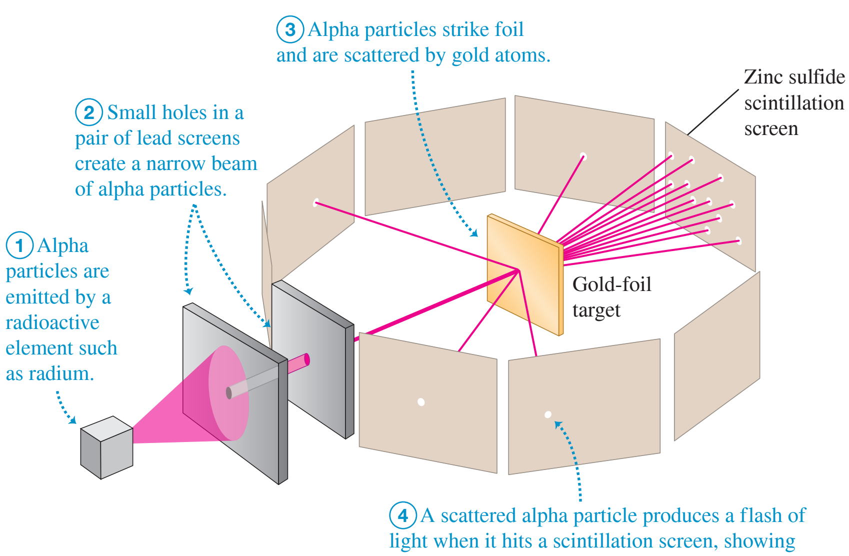
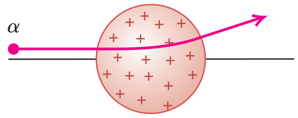
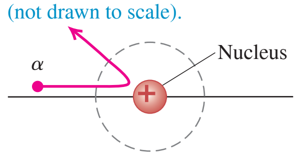


Figure 39.12 A comparison of Thomson's and Rutherford's models of the atom.

(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).



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.

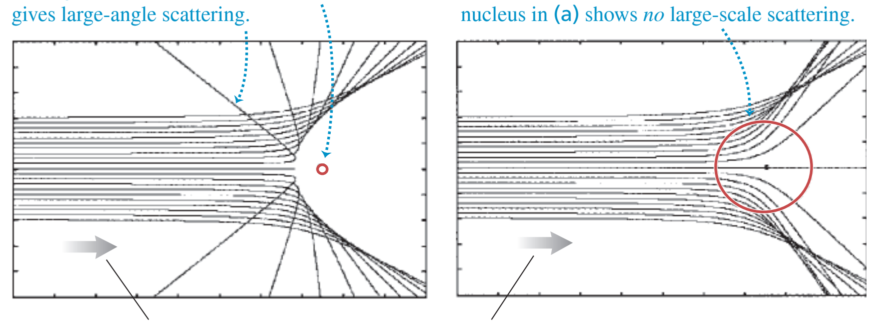


Figure 39.13 Computer simulation of scattering of 5.0 MeV alpha particles from a gold nucleus. Each curve shows a possible alpha-particle trajectory. (a) The scattering curves match Rutherford's experimental data if a radius of 7.0×10^{-15} m is assumed for a gold nucleus. (b) A model with a much larger radius for the gold nucleus does not match the data.

EXAMPLE 39.4 A Rutherford experiment

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.

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}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 $m_{Au} = 3.27 \times 10^{-25}$ kg) with this much momentum has a much slower speed $v_{Au} = 3.0 \times 10^5$ m/s and kinetic energy $K_{Au} = \frac{1}{2}mv_{Au}{}^2 = 1.5 \times 10^{-14}\ \mathrm{J} = 0.092$ MeV. This *recoil kinetic energy* of the gold nucleus is only 2% of the total mechanical energy in this situation, so we are justified in ignoring it.

KEYCONCEPT In Rutherford scattering, a positively charged alpha particle is scattered by a positively charged atomic nucleus. You can find the distance of closest approach using conservation of total mechanical energy (the kinetic energy plus the electric potential energy of the interaction between the alpha particle and the nucleus).

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.

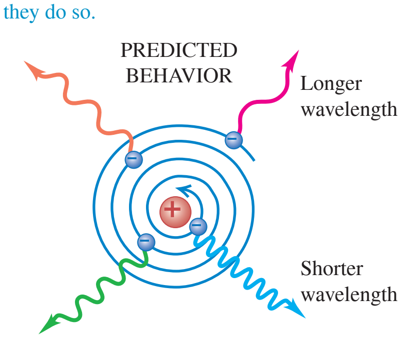
Thus Rutherford's model of electrons orbiting the nucleus, which is based on Newtonian mechanics and classical electromagnetic theory, makes three entirely *wrong* predictions about atoms: They should emit light continuously, they should be unstable, and the light they emit should have a continuous spectrum. Clearly a radical reappraisal of physics on the scale of the atom was needed. In the next section we'll 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.

Figure **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).

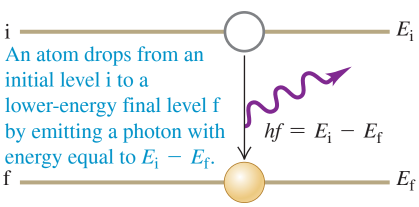
TEST YOUR UNDERSTANDING OF SECTION 39.2 Suppose you repeated Rutherford's scattering experiment with a thin sheet of solid hydrogen in place of the gold foil. (Hydrogen is a solid at temperatures below 14.0 K.) The nucleus of a hydrogen atom is a single proton, with about one-fourth the mass of an alpha 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?

ANSWER

(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).

Figure 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.

Figure 39.16 An excited atom emitting a photon.



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.

39.3 ENERGY LEVELS AND THE BOHR MODEL OF THE ATOM

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**):

Energy of emitted photon
Speed of light in vacuum
Final energy of atom after transition

$$hf = \frac{hc}{\lambda} = E_i - E_f \tag{39.5}$$

Planck's constant
Photon frequency
Photon wavelength
Initial energy of atom before transition

For example, an excited lithium atom emits red light with wavelength $\lambda = 671$ nm. The corresponding photon energy is

$$E = \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34}\ \text{J}\cdot\text{s})(3.00 \times 10^{8}\ \text{m/s})}{671 \times 10^{-9}\ \text{m}}$$

$$= 2.96 \times 10^{-19}\ \text{J} = 1.85\ \text{eV}$$

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.

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.

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 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.

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** (next page) 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).

Figure **39.17** An atom absorbing a photon. (Compare with Fig. 39.16.)

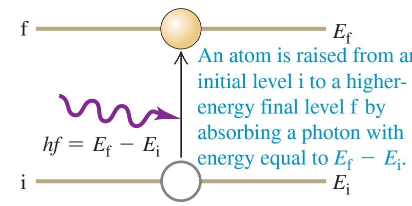
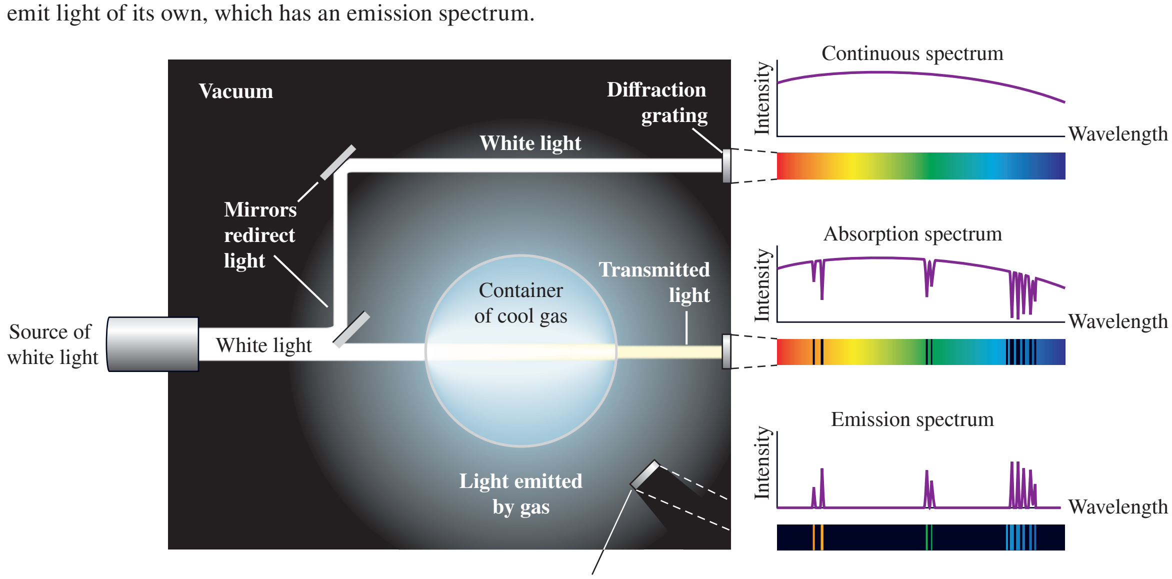


Figure **39.18** When a beam of white light with a continuous spectrum passes through a cool gas, the transmitted light has an absorption spectrum. The absorbed light energy excites the gas and causes it to emit light of its own, which has an emission spectrum.



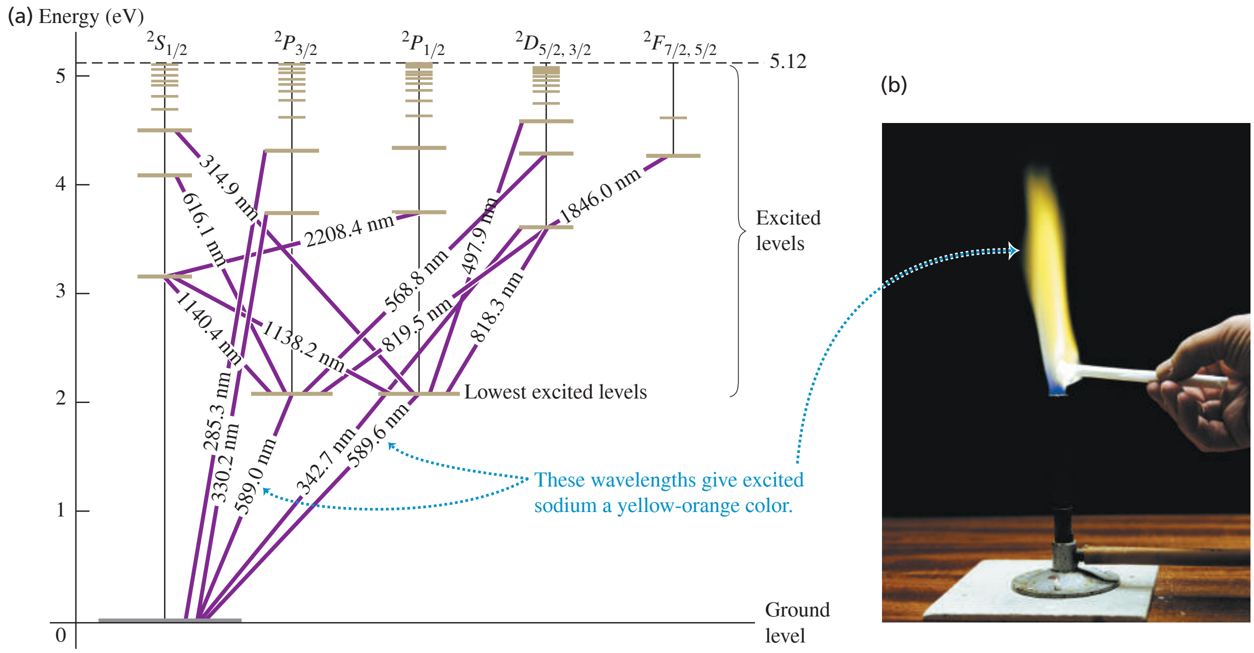


Figure **39.19** (a) Energy levels of the sodium atom relative to the ground level. Numbers on the lines between levels are wavelengths of the light emitted or absorbed during transitions between those levels. The column labels, such as $^2S_{1/2}$, refer to certain quantum states of the atom. (b) When 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.

EXAMPLE 39.5 Emission and absorption spectra

WITH VARIATION PROBLEMS

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?

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.

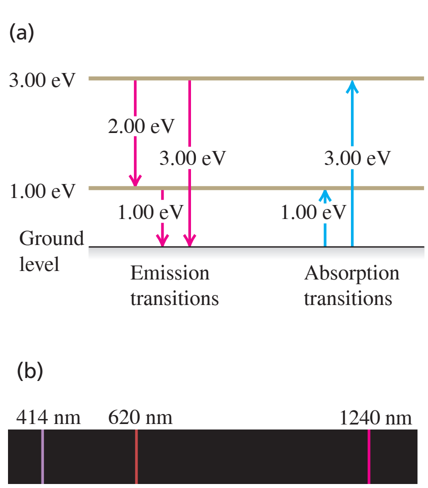


Figure **39.20** (a) Energy-level diagram for the hypothetical atom, showing the possible transitions for emission from excited levels and for absorption from the ground level. (b) Emission spectrum of this hypothetical atom.

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,

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^{8}\ \text{m/s}}{2.42 \times 10^{14}\ \text{Hz}} = 1.24 \times 10^{-6}\ \text{m} = 1240\ \text{nm}$$

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.

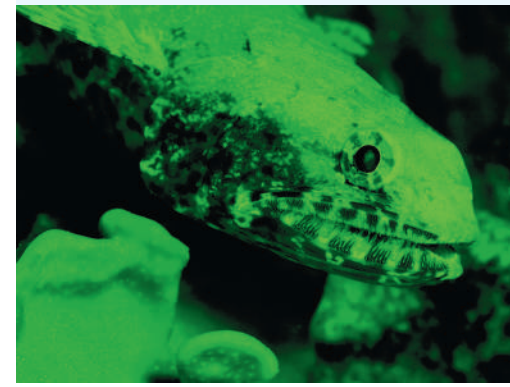
EVALUATE Note that if a gas of these atoms were at a sufficiently high temperature, collisions would excite a number of atoms into the 1.00 eV energy level. Such excited atoms *can* absorb 2.00 eV photons, as Fig. 39.20a shows, and an absorption line at 620 nm would appear in the spectrum. Thus the observed spectrum of a given substance depends on its energy levels and its temperature.

KEYCONCEPT When an atom makes a transition between two energy levels, it either emits a photon (if the atom loses energy) or absorbs a photon (if the atom gains energy). The energy of the photon that is emitted or absorbed equals the difference in energy between the two levels. The greater the photon energy, the shorter its wavelength.

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.

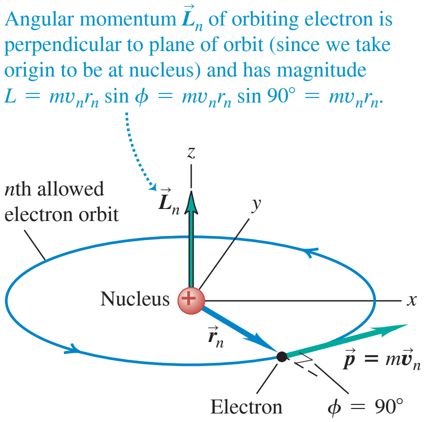
Electron Waves and the Bohr Model of Hydrogen

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.

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

Figure 39.21 Calculating the angular momentum of an electron in a circular orbit around an atomic nucleus.



written as kg · m^2/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 $L_n = mv_n r_n$ (**Fig. 39.21**). So Bohr's argument led to

Quantization of angular momentum:

Orbital angular momentum; Principal quantum number ($n = 1, 2, 3, \ldots$); Planck's constant

$$L_n = mv_n r_n = n\frac{h}{2\pi} \tag{39.6}$$

Electron mass; Electron speed; Electron orbital radius

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

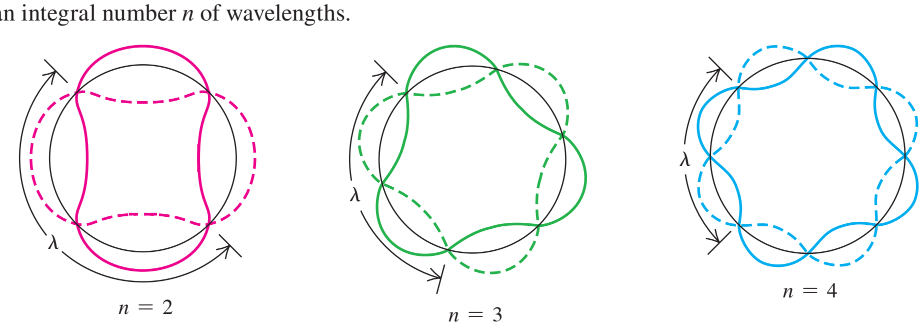
$$mv_n r_n = n\frac{h}{2\pi}$$

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),

$$F = \frac{1}{4\pi\epsilon_0}\frac{|(+e)(-e)|}{r_n^2} = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n^2}$$

Figure 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.



Hence Newton's second law states that

$$\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n^2} = \frac{mv_n^2}{r_n} \tag{39.7}$$

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

$$r_n = \epsilon_0 \frac{n^2h^2}{\pi m e^2} \tag{39.8}$$

Principal quantum number ($n = 1, 2, 3, \ldots$); Planck's constant; Magnitude of electron charge; Electron mass; Electric constant

Orbital speed in nth orbit in the Bohr model

$$v_n = \frac{1}{\epsilon_0}\frac{e^2}{2nh} \tag{39.9}$$

Magnitude of electron charge; Planck's constant; Electric constant; Principal quantum number ($n = 1, 2, 3, \ldots$)

Equation (39.8) shows that the orbital radius r_n is proportional to n^2, so the smallest orbital radius corresponds to $n = 1$. We'll denote this minimum radius, called the *Bohr radius,* as a_0:

$$a_0 = \epsilon_0 \frac{h^2}{\pi m e^2} \qquad \text{(Bohr radius)} \tag{39.10}$$

Then we can rewrite Eq. (39.8) as

Radius of nth orbit in the Bohr model

$$r_n = n^2 a_0 \tag{39.11}$$

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

$$a_0 = \frac{(8.854 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2})(6.626 \times 10^{-34}\ \mathrm{J \cdot s})^2}{\pi(9.109 \times 10^{-31}\ \mathrm{kg})(1.602 \times 10^{-19}\ \mathrm{C})^2} = 5.29 \times 10^{-11}\ \mathrm{m}$$

This gives an atomic diameter of about 10^{-10} m $= 0.1$ nm, which is consistent with atomic dimensions estimated by other methods.

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

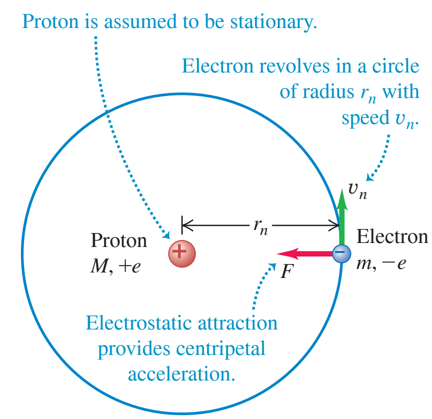
We can now use Eqs. (39.8) and (39.9) to find the kinetic and potential energies K_n and U_n when the electron is in the orbit with quantum number n:

$$K_n = \tfrac{1}{2}mv_n^2 = \frac{1}{\epsilon_0^2}\frac{me^4}{8n^2h^2} \qquad \text{(kinetic energies in the Bohr model)} \tag{39.12}$$

$$U_n = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n} = -\frac{1}{\epsilon_0^2}\frac{me^4}{4n^2h^2} \qquad \text{(potential energies in the Bohr model)} \tag{39.13}$$

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

Figure 39.23 The Bohr model of the hydrogen atom.



energy between orbits, so the reference position doesn't matter. The total mechanical energy E_n is the sum of the kinetic and potential energies:

$$E_n = K_n + U_n = -\frac{1}{\epsilon_0^2}\frac{me^4}{8n^2h^2} \qquad \text{(total energies in the Bohr model)} \tag{39.14}$$

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 orbital radius r_n, speed v_n, angular momentum $L_n = nh/2\pi$, and total mechanical 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 total mechanical 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. (39.15) for the energies as

Planck's constant
Speed of light in vacuum
Electron mass
Magnitude of electron charge
Total energy for nth orbit in the Bohr model

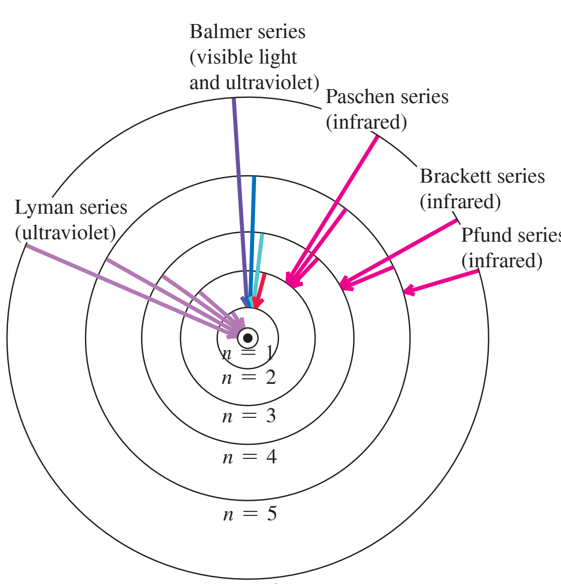
$$E_n = -\frac{hcR}{n^2}, \quad \text{where} \quad R = \frac{me^4}{8\epsilon_0^2h^3c} \tag{39.15}$$

Principal quantum number ($n = 1, 2, 3, \ldots$)
Rydberg constant
Electric constant

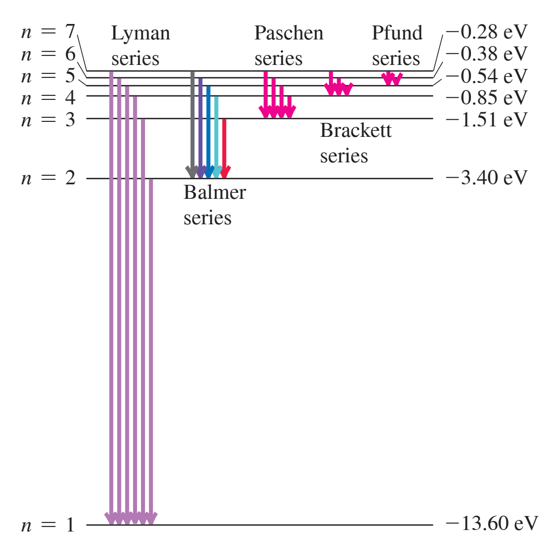
The quantity R in Eq. (39.15) is called the **Rydberg constant** (named for the Swedish physicist Johannes Rydberg, who did pioneering work on the hydrogen spectrum). When

Figure **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



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.

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:

$$\frac{1}{\lambda} = (1.097 \times 10^7\ \text{m}^{-1})\left(\tfrac{1}{4} - \tfrac{1}{9}\right) \qquad \text{or} \qquad \lambda = 656.3\ \text{nm}$$

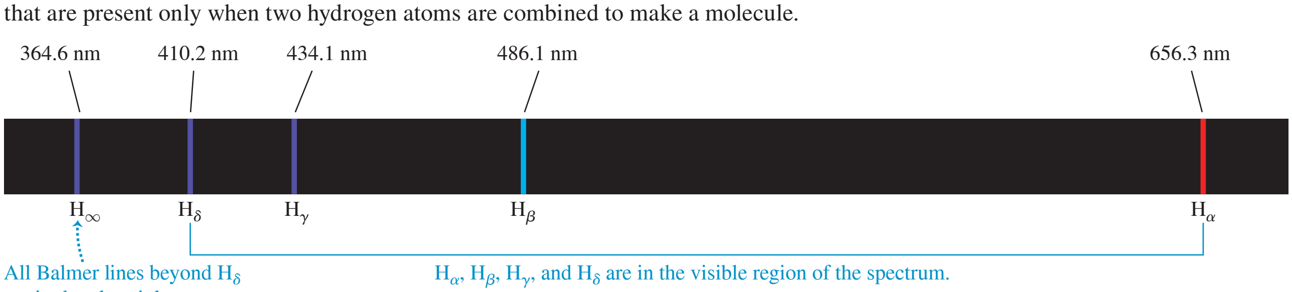
With $n_L = 2$ and $n_U = 4$ we obtain the wavelength of the H_β line, and so on. With $n_L = 2$ and $n_U = \infty$ we obtain the shortest wavelength in the series, $\lambda = 364.6$ nm. These theoretical predictions are within 0.1% of the observed hydrogen wavelengths! This close agreement provides 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

CAUTION **"Skipping over" energy levels is allowed** An atom can make transitions between nonconsecutive energy levels, such as from $n = 5$ to $n = 3$ or from $n = 1$ to $n = 4$, by emitting or absorbing a single photon. (Figure 39.24 shows several transitions of this kind.) The atom does *not* have to move only one level at a time (e.g., emit a photon and transition from $n = 5$ to $n = 4$, then emit another photon and transition from $n = 4$ to $n = 3$).

Figure **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.



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 orbital 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

WITH VARIATION PROBLEMS

Find the kinetic, potential, and total mechanical 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.

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:

$$\begin{aligned}\frac{me^4}{8\epsilon_0^2h^2} &= hcR\\ &= (6.626\times10^{-34}\ \mathrm{J\cdot s})(2.998\times10^{8}\ \mathrm{m/s})(1.097\times10^{7}\ \mathrm{m^{-1}})\\ &= 2.179\times10^{-18}\ \mathrm{J} = 13.60\ \mathrm{eV}\end{aligned}$$

This allows us to rewrite Eqs. (39.12), (39.13), and (39.14) as

$$K_n = \frac{13.60\ \mathrm{eV}}{n^2} \qquad U_n = \frac{-27.20\ \mathrm{eV}}{n^2} \qquad E_n = \frac{-13.60\ \mathrm{eV}}{n^2}$$

For the first excited level $(n = 2)$, we have $K_2 = 3.40\ \mathrm{eV}$, $U_2 = -6.80\ \mathrm{eV}$, and $E_2 = -3.40\ \mathrm{eV}$. For the ground level $(n = 1)$, $E_1 = -13.60\ \mathrm{eV}$. The energy of the emitted photon is then $E_2 - E_1 = -3.40\ \mathrm{eV} - (-13.60\ \mathrm{eV}) = 10.20\ \mathrm{eV}$, and

$$\begin{aligned}\lambda &= \frac{hc}{E_2 - E_1} = \frac{(4.136\times10^{-15}\ \mathrm{eV\cdot s})(3.00\times10^{8}\ \mathrm{m/s})}{10.20\ \mathrm{eV}}\\ &= 1.22\times10^{-7}\ \mathrm{m} = 122\ \mathrm{nm}\end{aligned}$$

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 $1/r^2$.

KEYCONCEPT In the Bohr model of the hydrogen atom, the electron moves around the nucleus in a circular Newtonian orbit. However, only certain orbital radii are allowed. The nth allowed orbit has a distinct kinetic energy K_n, electric potential energy U_n, and total mechanical energy $E_n = K_n + U_n$. When the electron makes a transition between two allowed orbits, a photon is emitted or absorbed with an energy equal to the difference between the values of E_n for the two orbits.

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 objects of masses m_1 and m_2, the reduced mass is

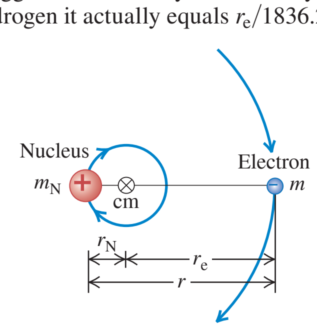


Figure 39.26 The nucleus and the electron both orbit around their common center of mass. The distance r_N has been exaggerated for clarity; for ordinary hydrogen it actually equals $r_e/1836.2$.

$$m_r = \frac{m_1m_2}{m_1 + m_2} \tag{39.17}$$

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 object 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.

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 $Z = 2$.

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'll find that an even more radical departure from classical concepts was needed before the understanding of atomic structure could progress further.

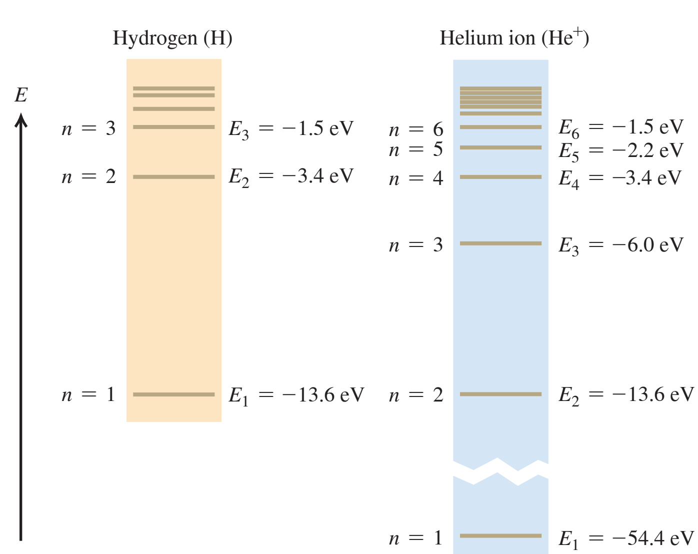


Figure **39.27** Energy levels of H and He^+. The energy expression, Eq. (39.14), is multiplied by $Z^2 = 4$ for He^+, so the energy of an He^+ ion with a given n is almost exactly four times that of an H atom with the same n. (There are small differences of the order of 0.05% because the reduced masses are slightly different.)

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.

ANSWER

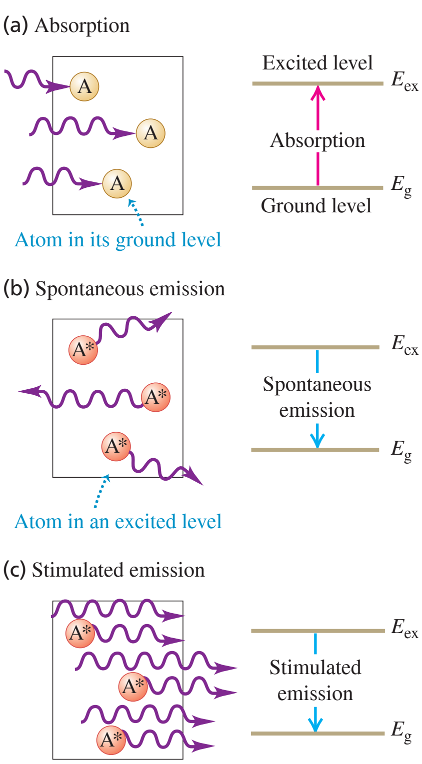
(iv) Figure 39.27 shows that many (though *not* all) of the energy levels of He^+ are the same as those of H. Hence photons emitted during transitions between corresponding pairs of levels in He^+ and H have the same energy E and the same wavelength $\lambda = hc/E$. An H atom that drops from the $n = 2$ 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 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

Figure 39.28 Three processes in which atoms interact with light.



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'll 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 $E_{ex} - E_g = 2.0 \text{ eV} = 3.2 \times 10^{-19}$ J, the energy of a 620 nm visible-light photon. At $T = 3000$ 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.

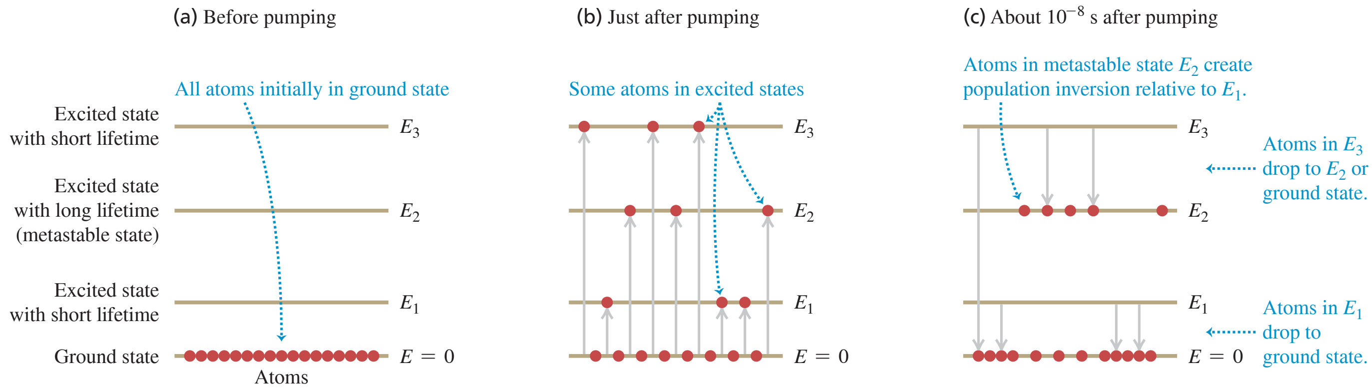
Enhancing Stimulated Emission: Population Inversions

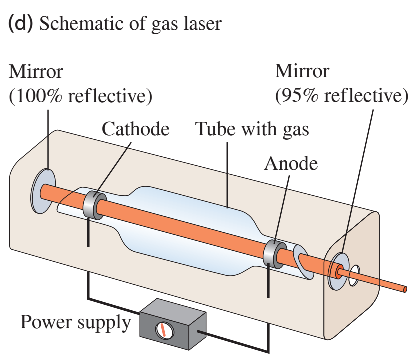
To make a laser, we need to promote stimulated emission by increasing the number of atoms in excited states. Can we do that simply by illuminating the container with radiation of frequency $f = E/h$ corresponding to the energy difference $E = E_{ex} - E_g$, as in Fig. 39.28a? Some of the atoms absorb 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 starting with atoms that have the right kinds of excited states. **Figure 39.29a** (next page) 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

Figure **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.





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.

ANSWER

❙ (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 CONTINUOUS SPECTRA

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 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. To nine significant figures, the value of the Stefan–Boltzmann constant σ is

$$\sigma = 5.67037442 \times 10^{-8}\ \mathrm{W/m^2 \cdot K^4}$$

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:

$$I = \int_0^\infty I(\lambda)\,d\lambda \qquad (39.20)$$

Figure **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.

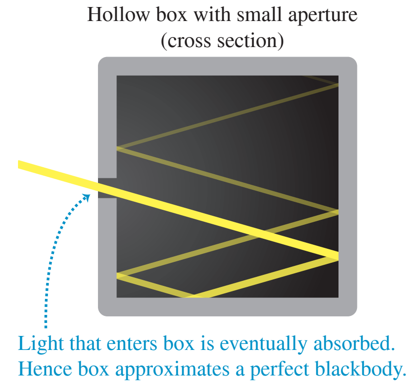


Figure **39.31** This close-up view of the sun's surface shows dark sunspots. Their temperature is about 4000 K, while the surrounding solar material is at T = 5800 K. From the Stefan–Boltzmann law, the intensity from a given area of sunspot is only $(4000\ \mathrm{K}/5800\ \mathrm{K})^4 = 0.23$ as great as the intensity from the same area of the surrounding material—which is why sunspots appear dark.

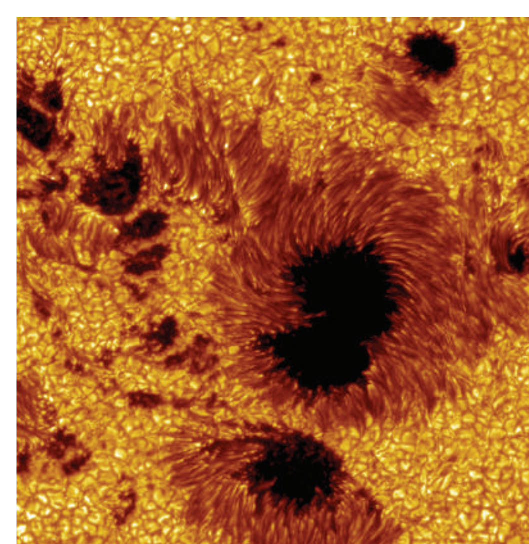
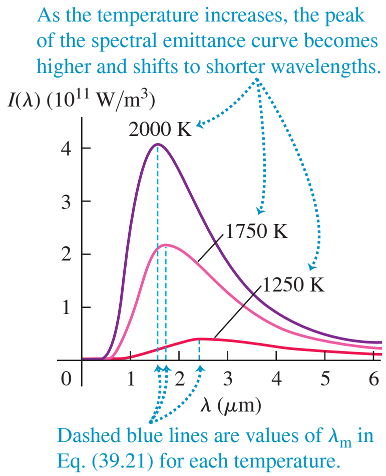


Figure 39.32 These graphs show the spectral emittance $I(\lambda)$ for radiation from a blackbody at three different temperatures.



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.

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: λ_m; Absolute temperature of blackbody: T)

$$\lambda_m T = 2.90 \times 10^{-3}\ \text{m} \cdot \text{K} \tag{39.21}$$

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 an object 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:

$$I(\lambda) = \frac{2\pi ckT}{\lambda^4} \quad \text{(Rayleigh's calculation)} \tag{39.22}$$

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.

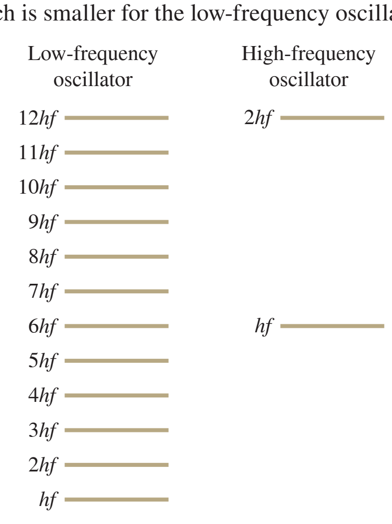
Figure 39.33 shows energy-level diagrams for two of the oscillators that Planck envisioned in the walls of the rectangular box, one with a low frequency and the other with a high frequency. The spacing in energy between adjacent levels is hf. This spacing is small for the low-frequency oscillator that 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

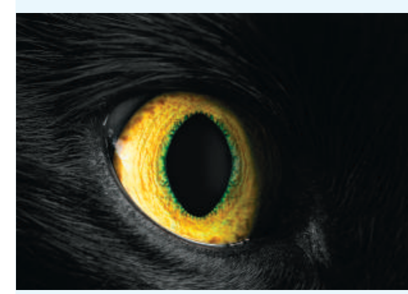
Figure **39.33** Energy levels for two of the oscillators that Planck envisioned in the walls of a blackbody like that shown in Fig. 39.30. The spacing between adjacent energy levels for each oscillator is hf, which is smaller for the low-frequency oscillator.



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.

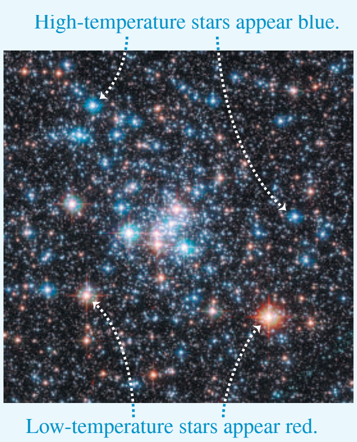
(a)

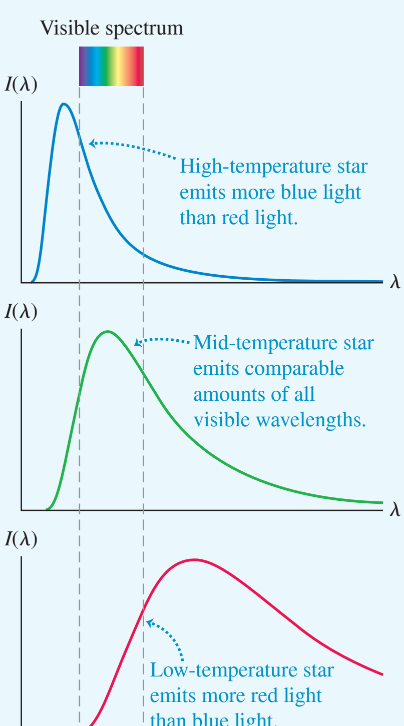
(b)



(c)

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.





of wavelength $\lambda = 3.00\ \mu\text{m}$, $hf = hc/\lambda = 0.413\ \text{eV}$; for a higher-frequency oscillator that emits photons of wavelength $\lambda = 0.500\ \mu\text{m}$, $hf = hc/\lambda = 2.48\ \text{eV}$. 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 $\lambda = 3.00\ \mu\text{m}$ 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 $\lambda = 0.500\ \mu\text{m}$ 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 $\lambda = 0.500\ \mu\text{m}$ 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 — Absolute temperature of blackbody — Wavelength — Boltzmann constant

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5\left(e^{hc/\lambda kT} - 1\right)} \tag{39.24}$$

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

$$\lambda_m = \frac{hc}{4.965kT} \tag{39.25}$$

To obtain this result, you have to solve the equation

$$5 - x = 5e^{-x} \tag{39.26}$$

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 $2.90 \times 10^{-3}\ \text{m} \cdot \text{K}$ 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 39.59). This is not a simple integral; the result is

$$I = \int_0^\infty I(\lambda)\, d\lambda = \frac{2\pi^5 k^4}{15c^2h^3}T^4 = \sigma T^4 \tag{39.27}$$

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:

$$\sigma = \frac{2\pi^5 k^4}{15c^2h^3} \tag{39.28}$$

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

WITH VARIATION PROBLEMS

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?

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 $\mathrm{kW/m^2}$.

KEYCONCEPT The continuous spectrum of light emitted by a blackbody is maximum at a wavelength λ_{max} given by the Wien displacement law: The greater the absolute temperature T of the blackbody, the smaller the value of λ_{max}. The total power radiated from a unit surface area of the blackbody is proportional to T^4.

EXAMPLE 39.8 A slice of sunlight

WITH VARIATION PROBLEMS

Find the power per unit area radiated from the sun's surface in the wavelength range 600.0 to 605.0 nm.

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 $\lambda = 602.5\ \mathrm{nm} = 6.025 \times 10^{-7}$ m, we first evaluate the quantity $hc/\lambda kT$ in Eq. (39.24) and then substitute the result into Eq. (39.24):

$$\frac{hc}{\lambda kT} = \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})(2.998 \times 10^8\ \mathrm{m/s})}{(6.025 \times 10^{-7}\ \mathrm{m})(1.381 \times 10^{-23}\ \mathrm{J/K})(5800\ \mathrm{K})} = 4.116$$

$$I(\lambda) = \frac{2\pi(6.626 \times 10^{-34}\ \mathrm{J \cdot s})(2.998 \times 10^8\ \mathrm{m/s})^2}{(6.025 \times 10^{-7}\ \mathrm{m})^5(e^{4.116} - 1)}$$
$$= 7.81 \times 10^{13}\ \mathrm{W/m^3}$$

The intensity in the 5.0 nm range from 600.0 to 605.0 nm is then approximately

$$I(\lambda)\Delta\lambda = (7.81 \times 10^{13}\ \mathrm{W/m^3})(5.0 \times 10^{-9}\ \mathrm{m})$$
$$= 3.9 \times 10^5\ \mathrm{W/m^2} = 0.39\ \mathrm{MW/m^2}$$

EVALUATE In part (b) of Example 39.7, we found the power radiated per unit area by the sun at *all* wavelengths to be $I = 64.2\ \mathrm{MW/m^2}$; here we have found that the power radiated per unit area in the wavelength range from 600 to 605 nm is $I(\lambda)\Delta\lambda = 0.39\ \mathrm{MW/m^2}$, about 0.6% of the total.

KEYCONCEPT The Planck radiation law gives the spectral emittance (power per unit area per unit wavelength interval) $I(\lambda)$ for an ideal blackbody. To find the power per unit area emitted over a range of wavelengths, integrate $I(\lambda)$ over that range.

TEST YOUR UNDERSTANDING OF SECTION 39.5 (a) Does a blackbody at 2000 K emit x rays? (b) Does it emit radio waves?

ANSWER

(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.

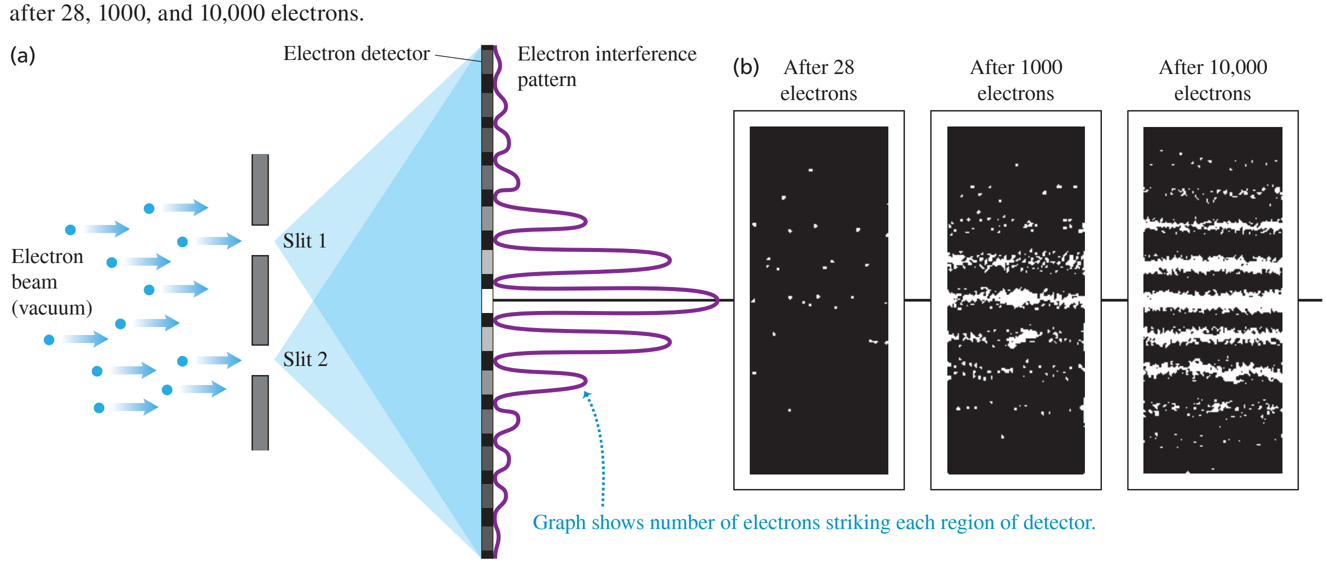
39.6 THE UNCERTAINTY PRINCIPLE REVISITED

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).

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.

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**). 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.

Figure **39.34** (a) A two-slit interference experiment for electrons. (b) The interference pattern after 28, 1000, and 10,000 electrons.



The Heisenberg Uncertainty Principles for Matter

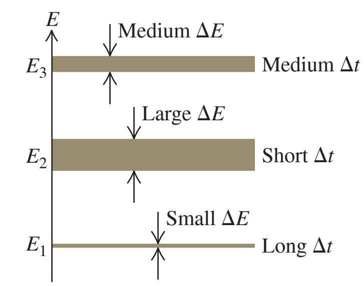
Just as electrons and photons show the same behavior in a two-slit interference experiment, electrons and other forms of matter obey the same Heisenberg uncertainty principles as photons do:

$$\begin{aligned}\Delta x \Delta p_x &\geq \hbar/2 \\ \Delta y \Delta p_y &\geq \hbar/2 \\ \Delta z \Delta p_z &\geq \hbar/2\end{aligned} \qquad \text{(Heisenberg uncertainty principle for position and momentum)} \qquad (39.29)$$

$$\Delta t\, \Delta E \geq \hbar/2 \qquad \text{(Heisenberg uncertainty principle for energy and time interval)} \qquad (39.30)$$

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.

Figure 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.

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

$$\begin{aligned}\Delta p_x &= \frac{\hbar}{2\Delta x} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(5.000 \times 10^{-11}\ \text{m})} = 1.055 \times 10^{-24}\ \text{J}\cdot\text{s/m} \\ &= 1.055 \times 10^{-24}\ \text{kg}\cdot\text{m/s}\end{aligned}$$

(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

$$\begin{aligned}K &= \frac{p^2}{2m} = \frac{(1.055 \times 10^{-24}\ \text{kg}\cdot\text{m/s})^2}{2(9.11 \times 10^{-31}\ \text{kg})} \\ &= 6.11 \times 10^{-19}\ \text{J} = 3.81\ \text{eV}\end{aligned}$$

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.

KEYCONCEPT The Heisenberg uncertainty principle for position and momentum applies to particles as well as to photons. The smaller the uncertainty in the position of a particle, the greater the uncertainty in the particle's momentum. Consequently, a particle confined to a small region of space could have a very large magnitude of momentum and a very large kinetic energy.

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?

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),

$$\begin{aligned}\Delta E &= \frac{\hbar}{2\Delta t} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(1.6 \times 10^{-8}\ \text{s})} \\ &= 3.3 \times 10^{-27}\ \text{J} = 2.1 \times 10^{-8}\ \text{eV}\end{aligned}$$

Continued

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.

KEYCONCEPT The Heisenberg uncertainty principle for energy and time interval states that the shorter the duration of an excited state of a system, the greater the minimum uncertainty in the energy of that state.

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'll 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.

ANSWER

(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.

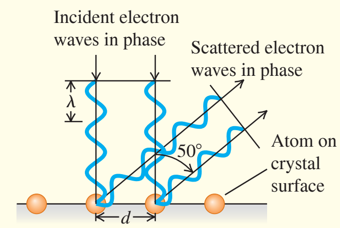
CHAPTER 39 SUMMARY

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.)

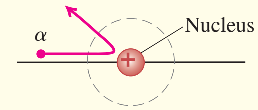
$$\lambda = \frac{h}{p} = \frac{h}{mv} \qquad (39.1)$$

$$E = hf \qquad (39.2)$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2meV_{ba}}} \qquad (39.3)$$

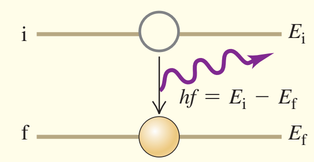


The nuclear atom: The Rutherford scattering experiments show that most of an atom's mass and all of its positive charge are concentrated in a tiny, dense nucleus at the center of the atom. (See Example 39.4.)



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 $E_{\mathrm{i}} - E_{\mathrm{f}}$. 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.)

$$hf = \frac{hc}{\lambda} = E_{\mathrm{i}} - E_{\mathrm{f}} \qquad (39.5)$$



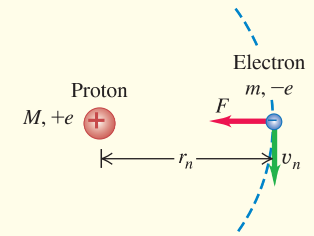
The Bohr model: In the Bohr model of the hydrogen atom, the permitted values of angular momentum are integral multiples of $h/2\pi$. The integer multiplier n is called the principal quantum number for the level. The orbital radii are proportional to n^2. The energy levels of the hydrogen atom are given by Eq. (39.15), where R is the Rydberg constant. (See Example 39.6.)

$$L_n = mv_nr_n = n\frac{h}{2\pi} \qquad (39.6)$$
$$(n = 1, 2, 3, \ldots)$$

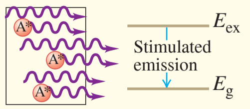
$$r_n = \epsilon_0\frac{n^2h^2}{\pi me^2} = n^2a_0 \qquad (39.8), (39.11)$$

$$v_n = \frac{1}{\epsilon_0}\frac{e^2}{2nh} \qquad (39.9)$$

$$E_n = -\frac{hcR}{n^2} = -\frac{13.60\text{ eV}}{n^2} \qquad (39.15)$$
$$(n = 1, 2, 3, \ldots)$$



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.

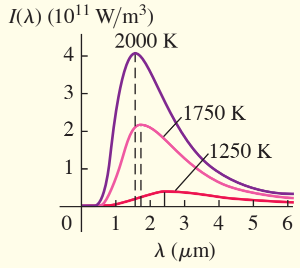


Blackbody radiation: The total radiated intensity (average power radiated per area) from a blackbody surface is proportional to the fourth power of the absolute temperature T. The quantity $\sigma = 5.67 \times 10^{-8}\ \mathrm{W/m^2 \cdot K^4}$ is called the Stefan–Boltzmann constant. The wavelength 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.)

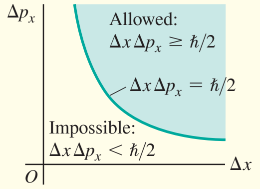
$$I = \sigma T^4$$
(Stefan–Boltzmann law) (39.19)

$$\lambda_{\mathrm{m}}T = 2.90 \times 10^{-3}\ \mathrm{m \cdot K}$$
(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.)



Chapter 39 Media Assets

For assigned homework and other learning materials, go to Mastering Physics.

GUIDED PRACTICE

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 39.1 and 39.2 (Section 39.1) before attempting these problems.

VP39.2.1 In an electron-diffraction experiment, the spacing between atoms in the target is 0.172 nm and the electrons have negligible kinetic energy before being accelerated by an accelerating voltage of 69.0 V. Find (a) the de Broglie wavelength of each electron and (b) the smallest scattering angle for which there is a maximum in the diffraction pattern.

VP39.2.2 The $m = 1$ intensity maximum in an electron-diffraction experiment occurs at an angle of 48.0°. The accelerating voltage of the electrons is 36.5 V. Find (a) the kinetic energy of each electron in joules, (b) the electron de Broglie wavelength, and (c) the spacing between atoms in the target.

VP39.2.3 Using a target with a spacing between atoms of 0.218 nm, you want to produce an electron-diffraction pattern for which the $m = 2$ intensity maximum occurs at an angle of 75.0°. Find (a) the required electron de Broglie wavelength and (b) the required accelerating voltage.

VP39.2.4 A proton (charge e, mass 1.67×10^{-27} kg) is accelerated from rest to 2.38×10^{5} m/s. Find (a) the proton de Broglie wavelength and (b) the accelerating voltage.

Be sure to review EXAMPLES 39.5 and 39.6 (Section 39.3) before attempting these problems.

VP39.6.1 A hypothetical atom has energy levels at 0.00 eV, 2.00 eV, and 5.00 eV. If the atom starts in the 5.00 eV level, find all the wavelengths that the atom could possibly emit in the process of returning to the ground level.

VP39.6.2 A hypothetical atom emits a photon of 674 nm when it transitions from the second excited level to the first excited level, and a photon of 385 nm when it transitions from the first excited level to the ground level. Find (a) the energies of the first and second excited levels relative to the ground level and (b) the wavelength of the photon emitted when the atom transitions from the second excited energy level to the ground level.

VP39.6.3 Calculate the energy and the wavelength of the photon emitted by a hydrogen atom when it makes a transition from (a) the $n = 5$ level to the $n = 3$ level, (b) the $n = 4$ level to the $n = 2$ level, and (c) the $n = 3$ level to the $n = 1$ level.

VP39.6.4 A hydrogen atom makes a transition from the $n = 2$ level to the $n = 6$ level. In the Bohr model, find (a) the change in electron kinetic energy, (b) the change in electric potential energy, and (c) the wavelength of the photon that the atom absorbs to cause the transition.

Be sure to review EXAMPLES 39.7 and 39.8 (Section 39.5) before attempting these problems.

VP39.8.1 The star Betelgeuse has surface temperature 3590 K and can be regarded as a blackbody. (a) Find the wavelength at which Betelgeuse emits most strongly. Is this visible, ultraviolet, or infrared? (b) Find the amount of power radiated per unit area of the surface of Betelgeuse.

VP39.8.2 The total power radiated per unit area from a blackbody is 78.0 MW/m^2. Find (a) the temperature of the blackbody and (b) the wavelength at which the blackbody emits most strongly. (c) Is the wavelength in part (b) visible, ultraviolet, or infrared?

VP39.8.3 The spectral emittance curve for the star Rigel A is a good approximation of the curve for a blackbody. The maximum of this curve is at 239 nm. Find (a) the surface temperature of Rigel A and (b) the power that Rigel A radiates per unit surface area.

VP39.8.4 Proxima Centauri, the nearest star to our sun, has a surface temperature of 3040 K. Find (a) the wavelength at which this star emits most strongly and (b) the power that this star radiates per unit surface area in a range of wavelengths 12.0 nm wide centered on the wavelength in part (a).

BRIDGING PROBLEM Hot Stars and Hydrogen Clouds

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?

Figure 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'll 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

•, ••, •••: 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 objects always emit a continuous spectrum—that is, a continuous smear of wavelengths. Can you account for this difference?

Q39.10 As an object 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 $(m = 6.64 \times 10^{-27}\ \text{kg})$.

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 $(m = 6.64 \times 10^{-27}\ \text{kg})$ 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 • Calculate the de Broglie wavelength of (a) a 50.0 kg woman jogging leisurely at 2.0 m/s, (b) a free electron with kinetic energy 2.0 MeV, and (c) a free electron with kinetic energy 20 eV. Use the proper relativistic expression when necessary.

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 photon and a free electron each have an energy of 6.00 eV. (a) What is the wavelength of the photon if it is traveling in air? (b) What is the de Broglie wavelength of the electron? (c) Which wavelength is longer?

39.16 • A photon traveling in air has a wavelength of 500 nm, and a free electron has a de Broglie wavelength of 500 nm. (a) What is the energy of each, in eV? (b) Which energy is greater?

Section 39.2 The Nuclear Atom and Atomic Spectra

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?

39.18 •• **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?

Section 39.3 Energy Levels and the Bohr Model of the Atom

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?

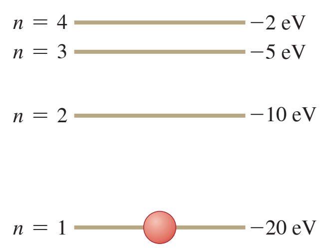
39.22 •• Consider the Bohr-model description of a hydrogen atom. (a) Calculate $E_2 - E_1$ and $E_{10} - E_9$. As n increases, does the energy separation between adjacent energy levels increase, decrease, or stay the same? (b) Show that $E_{n+1} - E_n$ approaches $(27.2\ \text{eV})/n^3$ as n becomes 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 • An electron is in a bound state of a hydrogen atom. The energy state of the atom is labeled with principal quantum number n. In the Bohr model description of this bound state, the electron has linear momentum $p = 6.65 \times 10^{-25}\ \text{kg} \cdot \text{m/s}$. In the Bohr model description, what are (a) the kinetic energy of the electron, (b) the angular momentum of the electron, and (c) the quantum number 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 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?

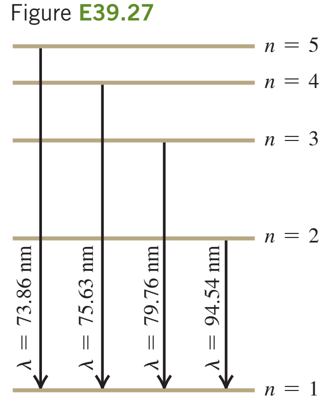
Figure **E39.25**



39.26 •• A positronium atom consists of a positron and an electron. In a Bohr-like model, the two particles rotate in circles about their common center of mass. (a) Calculate the reduced mass of a positronium atom in terms of the mass of an electron. (b) Determine the orbital radius of its ground-state electron. (c) Find its ground-state energy. (d) The longest visible-light emission wavelength for ordinary hydrogen is 656.3 nm in air and is for the $n = 3$ to $n = 2$ transition. Calculate the wavelength for the same transition in positronium.

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 • 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

39.36 • The temperature of a blackbody is changed so that the intensity I of radiation from the blackbody increases by a factor of 16. By what factor does the peak wavelength λ_m change?

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 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?

39.41 •• 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 • **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

39.43 • (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.

39.44 • 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 3.0×10^{-25} kg·m/s. Use the Heisenberg uncertainty principle to evaluate the validity of this claim.

PROBLEMS

39.48 • 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?

39.49 •• 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?

39.50 •• A hydrogen atom in an excited bound state labeled with principal quantum number $n = 3$ absorbs a photon that has wavelength λ. The atom is ionized and the electron has kinetic energy 8.00 eV after it has left the atom. What was the wavelength λ of the photon?

39.51 •• The wavelengths λ in the Pickering emission series are given by $\frac{1}{\lambda} = (1.097 \times 10^7\ \text{m}^{-1})\left[\frac{1}{4} - \frac{1}{(n/2)^2}\right]$ for $n = 5, 6, 7, \ldots$ and were attributed to hydrogen by some scientists. However, Bohr realized that this was *not* a hydrogen series, but rather belonged to another element, ionized so that it has only one electron. (a) What are the shortest and longest wavelengths in the Pickering series? (b) Which element gives rise to this series, and what is the common final-state quantum number n_L for each transition in the series?

39.52 •• 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.

39.53 ••• 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 photo-electrons? (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?

39.54 •• **CALC** Consider two energy levels, 1 and 2, of an atom that have nearly the same energies, so that their energies E_1 and E_2 relative to the ground state are close in value. When an atom undergoes a transition from one of these states to the ground state, the wavelength of the photon emitted is λ_1 or λ_2, respectively. Since $\Delta E = |E_1 - E_2|$ is small, $\Delta\lambda = |\lambda_1 - \lambda_2|$ is small. (a) Use the fact that $\Delta\lambda$ is small to derive an expression for ΔE in terms of $\Delta\lambda$ and $\lambda \approx \lambda_1 \approx \lambda_2$. (b) The transitions that produce the 589.0 nm and 589.6 nm wavelengths in the atomic emission spectrum of sodium are shown in Fig. 39.19. Use the result from part (a) to calculate the energy difference, in eV, between the two initial states for these two transitions.

39.55 •• **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).

39.56 •• **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?

39.57 •• **CP** The moon has a mass of 7.35×10^{22} kg, and the length of a sidereal day is 27.3 days. (a) Estimate the de Broglie wavelength of the moon in its orbit around the earth. (b) Using M_{earth} for the mass of the earth and M_{moon} for the mass of the moon, we can use Newton's law of gravitation to determine the radius of the moon's orbit in terms of an integer-valued quantum number m as $R_m = m^2 a_{\text{moon}}$, where a_{moon} is the analog of the Bohr radius for the earth–moon gravitational system. Determine a_{moon} in terms of Newton's constant G, Planck's constant h, and the masses M_{earth} and M_{moon}. (c) The mass of the earth is $M_{\text{earth}} = 5.97 \times 10^{24}$ kg. Estimate the numerical value of a_{moon}. (d) The radius of the moon's orbit is 3.84×10^8 m. Estimate the moon's quantum number m. (e) The quantized energy levels of the moon are given by $E = -E_0/m^2$. Estimate the quantum ground-state energy E_0 of the moon.

39.58 •• 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?

39.59 •• 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'll 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.

39.60 •• 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.

39.61 •• If you could keep utterly motionless, your de Broglie wavelength would be infinite. As soon as you make the slightest motion, however, your wavelength collapses. (a) Estimate the lowest speed you can perceive. (b) Estimate your wavelength if you moved with that slowest perceptible speed. (c) A grain of sand has a mass of about 0.5 mg. Estimate the wavelength of a grain of sand moving at your slowest perceptible speed. (It should be clear that the wave aspects of macroscopic material things are hidden from us by our size.) (d) If nature were to alter her laws so that Planck's constant became $h = 1\ \text{J}\cdot\text{s}$, then what would be the wavelength of a grain of sand moving at 1 m/s? (e) Under these same circumstances, estimate your own wavelength if you ran at 2.5 m/s. (f) A baseball has a mass of 145 g. Estimate the speed that a baseball would need to have a perceptible diffraction, meaning a central maximum subtending 10°, when thrown through a doorway, if h were 1 J·s.

39.62 • 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?

39.63 •• 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?

39.64 • 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?

39.65 •• 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?

39.66 • 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?

39.67 • 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.

39.68 • 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.

39.69 • 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.

39.70 • 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.")

39.71 • 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 39.70.)

39.72 • 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.

39.73 • 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?

39.74 • 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.

39.75 •• 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.

39.76 •• 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?

39.77 •• 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?

39.78 •• 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?

39.79 •• 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 $U(x) \rightarrow \infty$ as $x \rightarrow \infty$, 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 39.78, determine approximately the zero-point energy of the particle.

39.80 •• 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?

39.81 •• 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?

39.82 •• 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.

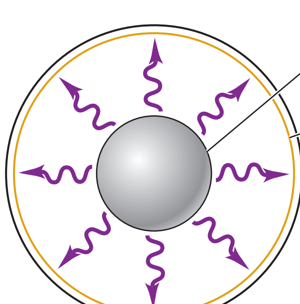
39.83 •• DATA As an amateur astronomer, you are studying the apparent brightness of stars. You know that a star's apparent brightness depends on its distance from the earth and also on the fraction of its radiated energy that is in the visible region of the electromagnetic spectrum. But, as a first step, 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?

39.84 ••• CP CALC An advanced civilization uses a fusion reactor to maintain a high temperature on a spherical blackbody whose radiation supplies energy for use on a spacecraft. The blackbody has radius $r = 1.23$ m and is surrounded by a larger sphere whose inside surface is densely lined with photocells (**Fig. P39.84**). These capture all photons that have energy within 1.00% of a central value E_0. The photocells are designed so that E_0 corresponds to the peak energy radiated from the blackbody. The range ΔE is sufficiently small that we can approximate the spectral emittance as constant over the range $E_0 \pm \Delta E$. This device generates an emf across a motor of resistance 1.00 kΩ. (a) The spectral emittance $I = \int I(\lambda)\,d\lambda$ can be structured as $\int I(E)\,dE$ by parameterizing the integral using $E = hc/\lambda$. Use Eq. (39.24) to determine $I(E)$. Be careful with the differential factor. (b) Determine the energy E_0 that maximizes $I(E)$, and determine $I_{max} = I(E_0)$ in terms of the temperature T. (*Hint:* You have to solve the equation $3 - x = 3e^{-x}$. Use the solution $x = 2.821$, which is accurate to four significant figures.) (c) If the device is designed for $T = 1000$ K, then how much power is supplied by the motor, and what is the current I? (*Hint:* Integrate the spectral emittance around the blackbody.) (d) If the device is designed for $T = 5000$ K, then how much power is supplied and what is the current?

Figure **P39.84**



39.85 ••• CP An alpha particle is incident with kinetic energy K on a gold nucleus at rest. The aim is direct. (a) If m is the mass of an alpha particle and M is the mass of a gold nucleus, solve the classical conditions for energy and momentum conservation to determine the recoil speed V of the nucleus after the collision. (b) Determine an expression for the fractional energy lost to the nucleus. (c) Is your result independent of the initial kinetic energy? (d) An alpha particle has mass $m = 6.64 \times 10^{-27}$ kg, and a gold nucleus has mass $M = 1.32 \times 10^{-25}$ kg. If $K = 5.00$ MeV, then what is the speed V as a fraction of c, and what proportion of the original energy is transferred to the gold nucleus? (e) According to the classical analysis, what speed of the incident alpha particle would result in a nuclear speed V of $0.10c$? (f) Is that possible?

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'll 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.)

MCAT-STYLE 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.)

Key Example VARIATION Problems

VP39.2.1 (a) 0.148 nm (b) 59.1°
VP39.2.2 (a) 5.85×10^{-18} J (b) 0.203 nm (c) 0.273 nm
VP39.2.3 (a) 0.105 nm (b) 136 V
VP39.2.4 (a) 1.67×10^{-12} m (b) 295 V
VP39.6.1 620 nm, 414 nm, 248 nm
VP39.6.2 (a) 3.22 eV for first excited level, 5.06 eV for second excited level (b) 245 nm
VP39.6.3 (a) $E = 0.967$ eV, $\lambda = 1.28\ \mu$m (b) $E = 2.55$ eV, $\lambda = 487$ nm (c) $E = 12.1$ eV, $\lambda = 103$ nm
VP39.6.4 (a) -3.02 eV (b) $+6.04$ eV (c) 411 nm
VP39.8.1 (a) 808 nm, infrared (b) 9.42 MW/m^2
VP39.8.2 (a) 6.09×10^3 K (b) 476 nm (c) visible
VP39.8.3 (a) 1.21×10^4 K (b) 1.23×10^9 W/m^2
VP39.8.4 (a) 954 nm (b) 40.1 kW/m^2

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

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.

LEARNING OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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.

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

Figure 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 $\Psi(x, y, z, t)$.

of the hydrogen atom (Section 39.3) we tried to have it both ways: We 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.

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.

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:*

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y(x, t)}{\partial t^2} \qquad \text{(wave equation for waves on a string)} \qquad (40.1)$$

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:

$$y(x, t) = A\cos(kx - \omega t) + B\sin(kx - \omega t) \qquad \text{(sinusoidal wave on a string)} \qquad (40.2)$$

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:

$$\frac{\partial y(x,t)}{\partial x} = -kA\sin(kx - \omega t) + kB\cos(kx - \omega t) \quad (40.3a)$$

$$\frac{\partial^2 y(x,t)}{\partial x^2} = -k^2A\cos(kx - \omega t) - k^2B\sin(kx - \omega t) \quad (40.3b)$$

$$\frac{\partial y(x,t)}{\partial t} = \omega A\sin(kx - \omega t) - \omega B\cos(kx - \omega t) \quad (40.3c)$$

$$\frac{\partial^2 y(x,t)}{\partial t^2} = -\omega^2A\cos(kx - \omega t) - \omega^2B\sin(kx - \omega t) \quad (40.3d)$$

If we substitute Eqs. (40.3b) and (40.3d) into the wave equation, Eq. (40.1), we get

$$-k^2A\cos(kx - \omega t) - k^2B\sin(kx - \omega t) = \frac{1}{v^2}\left[-\omega^2A\cos(kx - \omega t) - \omega^2B\sin(kx - \omega t)\right] \quad (40.4)$$

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

$$k^2 = \frac{\omega^2}{v^2} \quad \text{or} \quad \omega = vk \quad \text{(waves on a string)} \quad (40.5)$$

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:

$$E = \tfrac{1}{2}mv^2 = \frac{m^2v^2}{2m} = \frac{(mv)^2}{2m} = \frac{p^2}{2m} \quad \text{(energy of a free particle)} \quad (40.6)$$

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 k:

$$E = hf = \frac{h}{2\pi}2\pi f = \hbar\omega \quad (40.7a)$$

$$p = \frac{h}{\lambda} = \frac{h}{2\pi}\frac{2\pi}{\lambda} = \hbar k \quad (40.7b)$$

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 $A = -C^2 A$, 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

Figure **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).

$$\Psi(x,t) = A[\cos(kx - \omega t) + i\sin(kx - \omega t)] \quad \text{(sinusoidal wave function representing a free particle)} \tag{40.16}$$

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 $\Psi(x, 0) = A\cos kx + iA\sin kx$.

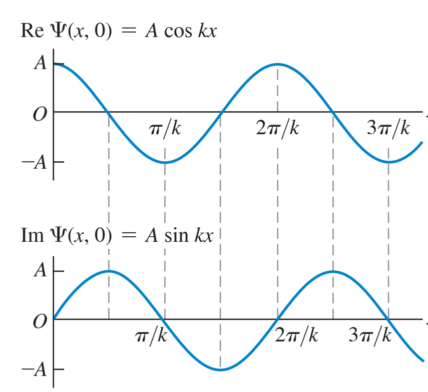


Figure **40.3** The spatial wave function $\Psi(x, t) = Ae^{i(kx-\omega t)}$ for a free particle of definite momentum $p = \hbar k$ is a complex function: It has both a real part and an imaginary part. These are graphed here as functions of x for $t = 0$.

We can rewrite Eq. (40.16) with *Euler's formula,* which states that for any angle θ,

$$\begin{aligned} e^{i\theta} &= \cos\theta + i\sin\theta \\ e^{-i\theta} &= \cos(-\theta) + i\sin(-\theta) = \cos\theta - i\sin\theta \end{aligned} \tag{40.17}$$

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 $E = \hbar\omega = \hbar^2 k^2/2m$. 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 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*.

Figure 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 superposing 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

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 $\Psi(x, t)$.

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

$$
\begin{aligned}
-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} &= -\frac{\hbar^2}{2m}\frac{\partial^2\left(Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}\right)}{\partial x^2} \\
&= -\frac{\hbar^2}{2m}\left[(ik_1)^2Ae^{i(k_1x-\omega_1t)} + (ik_2)^2Ae^{i(k_2x-\omega_2t)}\right] \\
&= \frac{\hbar^2k_1^2}{2m}Ae^{i(k_1x-\omega_1t)} + \frac{\hbar^2k_2^2}{2m}Ae^{i(k_2x-\omega_2t)}
\end{aligned}
$$

The right-hand side is

$$
\begin{aligned}
i\hbar\frac{\partial\Psi(x,t)}{\partial t} &= i\hbar\frac{\partial\left(Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}\right)}{\partial t} \\
&= i\hbar\left[(-i\omega_1)Ae^{i(k_1x-\omega_1t)} + (-i\omega_2)Ae^{i(k_2x-\omega_2t)}\right] \\
&= \hbar\omega_1Ae^{i(k_1x-\omega_1t)} + \hbar\omega_2Ae^{i(k_2x-\omega_2t)}
\end{aligned}
$$

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

$$\Psi^*(x,t) = A^*e^{-i(k_1x-\omega_1t)} + A^*e^{-i(k_2x-\omega_2t)}$$

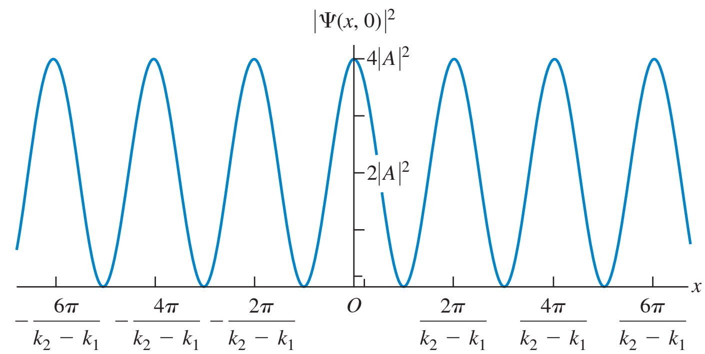
Hence

$$
\begin{aligned}
&|\Psi(x,t)|^2 \\
&= \Psi^*(x,t)\Psi(x,t) \\
&= \left(A^*e^{-i(k_1x-\omega_1t)} + A^*e^{-i(k_2x-\omega_2t)}\right)\left(Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}\right) \\
&= A^*A\left[\begin{aligned}&e^{-i(k_1x-\omega_1t)}e^{i(k_1x-\omega_1t)} + e^{-i(k_2x-\omega_2t)}e^{i(k_2x-\omega_2t)} \\ &+ e^{-i(k_1x-\omega_1t)}e^{i(k_2x-\omega_2t)} + e^{-i(k_2x-\omega_2t)}e^{i(k_1x-\omega_1t)}\end{aligned}\right] \\
&= |A|^2\left[e^0 + e^0 + e^{i[(k_2-k_1)x-(\omega_2-\omega_1)t]} + e^{-i[(k_2-k_1)x-(\omega_2-\omega_1)t]}\right]
\end{aligned}
$$

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

$$
\begin{aligned}
|\Psi(x,t)|^2 &= |A|^2\{2 + 2\cos[(k_2-k_1)x - (\omega_2-\omega_1)t]\} \\
&= 2|A|^2\{1 + \cos[(k_2-k_1)x - (\omega_2-\omega_1)t]\}
\end{aligned}
$$

Figure 40.5 The probability distribution function at $t = 0$ for $\Psi(x, t) = Ae^{i(k_1x-\omega_1t)} + Ae^{i(k_2x-\omega_2t)}$.



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).

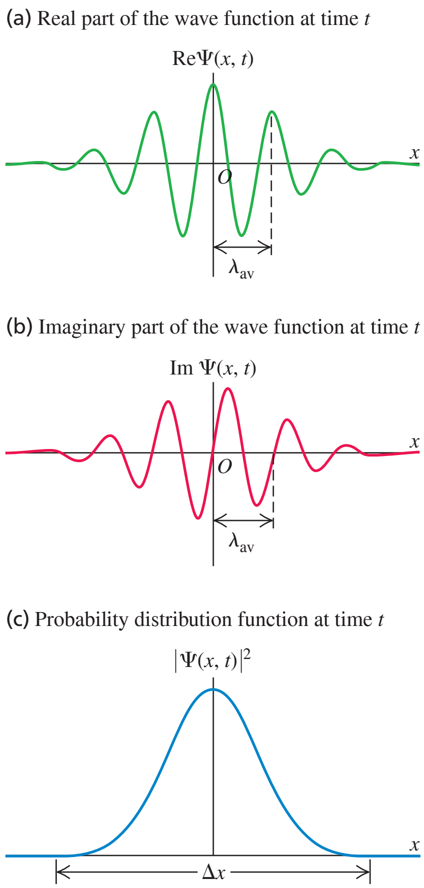
Note also that the probability distribution function is not stationary: It moves in the positive x-direction like the particle that it represents. To see this, recall from Section 15.3 that a sinusoidal wave given by $y(x, t) = A\cos(kx - \omega t)$ moves in the positive x-direction with speed $v = \omega/k$; 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?

KEYCONCEPT A free particle (one on which no forces act) that moves in the x-direction only is described by a wave function $\Psi(x, t)$ that obeys the free-particle Schrödinger equation [Eq. (40.15)]. A wave function with definite momentum and energy is completely delocalized; to write a wave function that is localized in space, you must superpose wave functions with different values of momentum and energy.

Figure 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.



Wave Packets

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. If we continue this process and superpose a very large number of waves with 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

$$\Psi(x, t) = \int_{-\infty}^{\infty} A(k) e^{i(kx-\omega t)}\, dk \tag{40.19}$$

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:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + U(x)\Psi(x,t) = i\hbar\frac{\partial\Psi(x,t)}{\partial t} \quad (40.20)$$

Planck's constant divided by 2π
Particle's wave function
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).

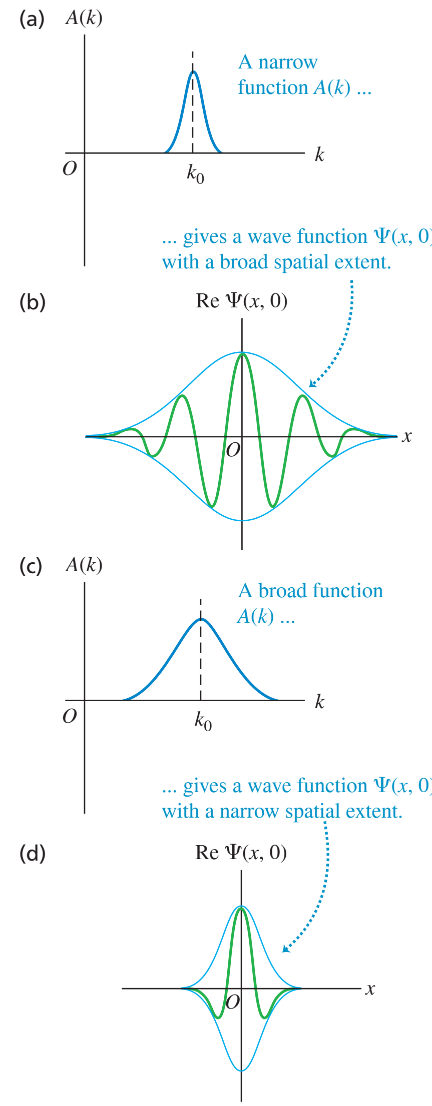
Here's the motivation behind Eq. (40.20). If $\Psi(x, t)$ is a sinusoidal wave function for a free particle, $\Psi(x, t) = Ae^{i(kx-\omega t)} = Ae^{ikx}e^{-i\omega t}$, the derivative terms in Eq. (40.20) become

$$\begin{aligned} -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} &= -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}(Ae^{ikx}e^{-i\omega t}) = -\frac{\hbar^2}{2m}(ik)^2(Ae^{ikx}e^{-i\omega t}) \\ &= \frac{\hbar^2k^2}{2m}\Psi(x,t) \\ i\hbar\frac{\partial\Psi(x,t)}{\partial t} &= i\hbar\frac{\partial}{\partial t}(Ae^{ikx}e^{-i\omega t}) = i\hbar(-i\omega)(Ae^{ikx}e^{-i\omega t}) = \hbar\omega\Psi(x,t) \end{aligned}$$

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 $U(x)$.

Figure 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 $t = 0$).



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

$$\Psi(x,t) = \psi(x)e^{-iEt/\hbar} \quad (40.21)$$

Time-independent wave function
Energy of state
Planck's constant divided by 2π

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 $|\Psi|^2$:

$$\begin{aligned}|\Psi(x,t)|^2 &= \Psi^*(x,t)\Psi(x,t) = [\psi^*(x)e^{+iEt/\hbar}][\psi(x)e^{-iEt/\hbar}]\\ &= \psi^*(x)\psi(x)e^{(+iEt/\hbar)+(-iEt/\hbar)} = |\psi(x)|^2e^0\\ &= |\psi(x)|^2\end{aligned} \tag{40.22}$$

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 $\psi(x) = A_1e^{ikx} + A_2e^{-ikx}$, 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?

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 $\psi(x) = A_1e^{ikx} + A_2e^{-ikx}$ and $U(x) = 0$ into Eq. (40.23), we obtain

$$\begin{aligned}-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} &= -\frac{\hbar^2}{2m}\frac{d^2(A_1e^{ikx} + A_2e^{-ikx})}{dx^2}\\ &= -\frac{\hbar^2}{2m}[(ik)^2A_1e^{ikx} + (-ik)^2A_2e^{-ikx}]\\ &= \frac{\hbar^2k^2}{2m}(A_1e^{ikx} + A_2e^{-ikx}) = \frac{\hbar^2k^2}{2m}\psi(x)\end{aligned}$$

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.

KEYCONCEPT A quantum-mechanical state of definite energy E has a wave function of the form $\Psi(x, t) = \psi(x)e^{-iEt/\hbar}$. It is called a stationary state because the associated probability distribution function $|\Psi(x, t)|^2 = |\psi(x)|^2$ does not depend on time.

TEST YOUR UNDERSTANDING OF SECTION 40.1 Does a wave packet given by Eq. (40.19) represent a stationary state?

ANSWER

no Equation (40.19) represents a superposition of wave functions with different values of wave number k and hence different values of energy $E = \hbar^2 k^2/2m$. 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 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.

Figure 40.8 The Newtonian view of a particle in a box.

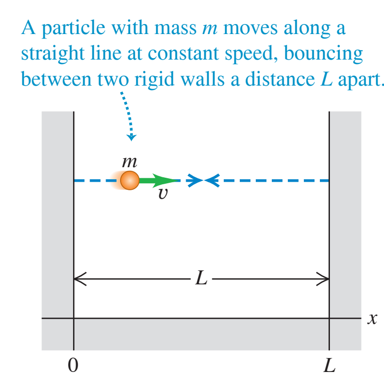
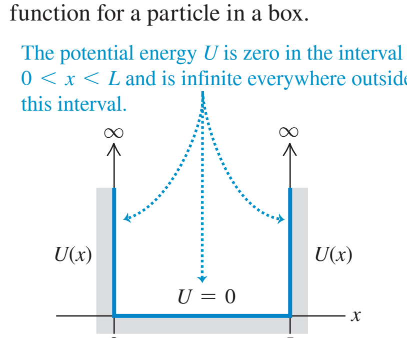


Figure 40.9 The potential-energy function for a particle in a box.

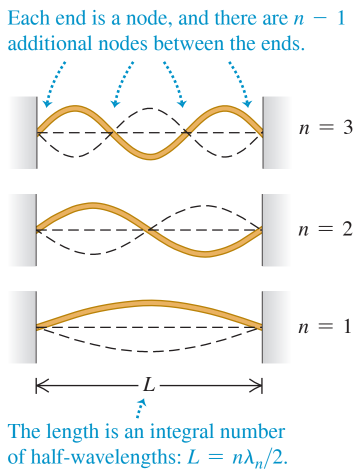


Wave Functions for a Particle in a Box

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

Figure **40.10** Normal modes of vibration for a string with length L, held at both ends.



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.

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 $0 \leq x \leq L$ subject to the above conditions. In this region $U(x) = 0$, so $\psi(x)$ in this region must satisfy

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x) \qquad \text{(particle in a box)} \tag{40.25}$$

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

$$\psi(x) = A_1e^{ikx} + A_2e^{-ikx} \tag{40.26}$$

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):

$$\begin{aligned}\psi(x) &= A_1(\cos kx + i\sin kx) + A_2[\cos(-kx) + i\sin(-kx)]\\ &= A_1(\cos kx + i\sin kx) + A_2(\cos kx - i\sin kx)\\ &= (A_1 + A_2)\cos kx + i(A_1 - A_2)\sin kx\end{aligned} \tag{40.27}$$

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

$$\psi(x) = 2iA_1\sin kx = C\sin kx \tag{40.28}$$

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 $kL = n\pi$ $(n = 1, 2, 3, \ldots)$. Hence Eq. (40.28) does indeed give the stationary-state wave functions for a particle in a box in the region $0 \leq x \leq L$. (Outside this region, $\psi(x) = 0$.) The possible values of k and the wavelength $\lambda = 2\pi/k$ are

$$k = \frac{n\pi}{L} \quad \text{and} \quad \lambda = \frac{2\pi}{k} = \frac{2L}{n} \quad (n = 1, 2, 3, \ldots) \tag{40.29}$$

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

$$\overset{\text{Energy levels for a particle in a box}}{E_n} = \frac{\overset{\text{Magnitude of momentum}}{p_n}^{2}}{\underset{\text{Particle's mass}}{2m}} = \frac{n^2\overset{\text{Planck's constant}}{h^2}}{8m\underset{\text{Width of box}}{L^2}} = \frac{n^2\pi^2\overset{\text{Planck's constant divided by }2\pi}{\hbar^2}}{2mL^2} \quad (\underset{\text{Quantum number}}{n} = 1, 2, 3, \ldots) \tag{40.31}$$

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

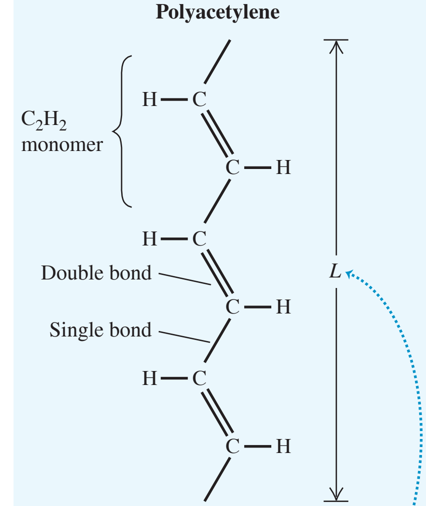
$$\psi_n(x) = C\sin\frac{n\pi x}{L} \quad (n = 1, 2, 3, \ldots) \tag{40.32}$$

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 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.

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



Electrons are confined to the length L of the molecule (like particles in a box).

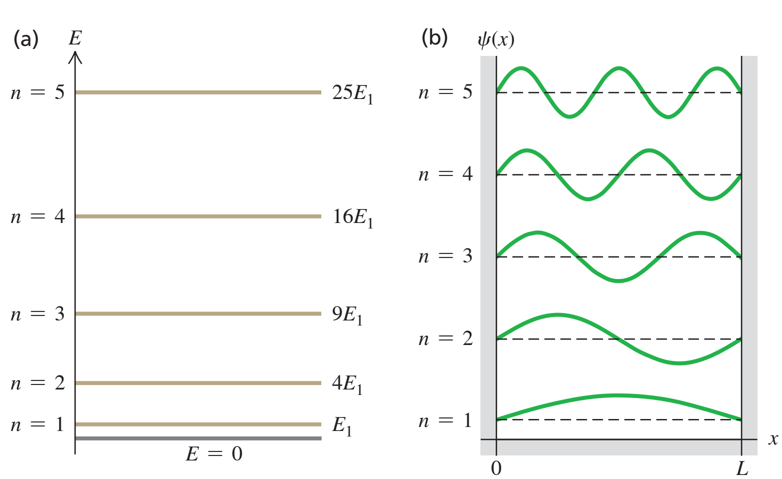


Figure **40.11** (a) Energy-level diagram for a particle in a box. Each energy is n^2E_1, where E_1 is the ground-level energy. (b) Wave functions for a particle in a box, with $n = 1, 2, 3, 4$, and 5. **CAUTION:** The five graphs have been displaced vertically for clarity, as in Fig. 40.10. Each of the horizontal dashed lines represents $\psi = 0$ for the respective wave function.

EXAMPLE 40.3 Electron in an atom-size box

WITH VARIATION PROBLEMS

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).

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.

KEYCONCEPT The wave functions for the stationary states of a particle confined to a one-dimensional box are sinusoidal standing waves inside the box with a node at each end; the wave function is zero outside the box. The nth stationary state has a wave function with n half-wavelengths within the length of the box, and has an energy equal to n^2 times the energy of the lowest-energy ($n = 1$) state. There are an infinite number of these stationary states.

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 $x = L$.

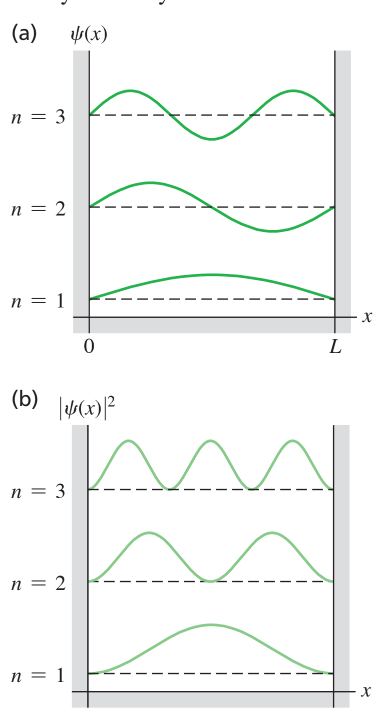
The particle must be *somewhere* on the x-axis—that is, somewhere between $x = -\infty$ and $x = +\infty$. So the *sum* of the probabilities for all the dx's everywhere (the *total* probability of finding the particle) must equal 1. That's the normalization condition that we discussed in Section 40.1:

Normalization condition, time-independent wave function:

$$\overbrace{\int_{-\infty}^{\infty}}^{\text{Integral over all } x} \underbrace{|\psi(x)|^2}_{\text{Probability distribution function}}\,dx = \underbrace{1}_{\text{Probability that particle is } somewhere \text{ on } x\text{-axis}} \qquad (40.33)$$

Time-independent wave function

Figure **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.



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

$$\int_0^L C^2 \sin^2 \frac{n\pi x}{L}\, dx = 1 \tag{40.34}$$

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

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \qquad (n = 1, 2, 3, \ldots) \tag{40.35}$$

Quantum number: n. Width of box: L.

EXAMPLE 40.4 A nonsinusoidal wave function?

WITH VARIATION PROBLEMS

(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?

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.

KEYCONCEPT To be physically reasonable, a stationary-state wave function for a given situation must satisfy both the Schrödinger equation and any boundary conditions that apply to that situation.

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

$$\Psi_n(x, t) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) e^{-iE_nt/\hbar} \qquad (n = 1, 2, 3, \ldots) \tag{40.36}$$

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.

ANSWER **(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.

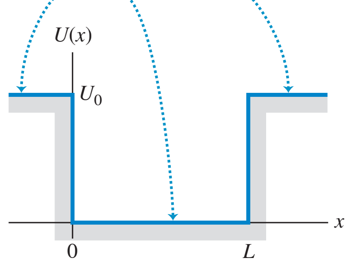
40.3 POTENTIAL WELLS

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.

Figure 40.13 A square-well potential.

The potential energy U is zero within the potential well (in the interval $0 \le x \le L$) and has the constant value U_0 outside this interval.



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

$$\psi(x) = A\cos\left(\frac{\sqrt{2mE}}{\hbar}x\right) + B\sin\left(\frac{\sqrt{2mE}}{\hbar}x\right) \qquad \text{(inside the well)} \quad (40.38)$$

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

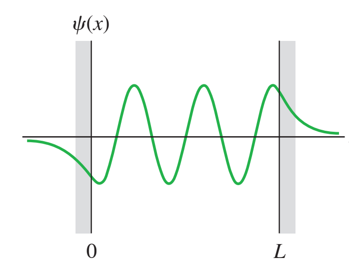
$$\psi(x) = Ce^{\kappa x} + De^{-\kappa x} \qquad \text{(outside the well)} \qquad (40.40)$$

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 $x \to +\infty$ or $x \to -\infty$. [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.

Figure **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

(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 $U_0 \to \infty$?

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

$$\begin{aligned}\frac{d^2\psi(x)}{dx^2} &= \frac{d^2}{dx^2}(Ce^{\kappa x}) + \frac{d^2}{dx^2}(De^{-\kappa x}) \\ &= C\kappa^2e^{\kappa x} + D(-\kappa)^2e^{-\kappa x} \\ &= \kappa^2(Ce^{\kappa x} + De^{-\kappa x}) = \kappa^2\psi(x)\end{aligned}$$

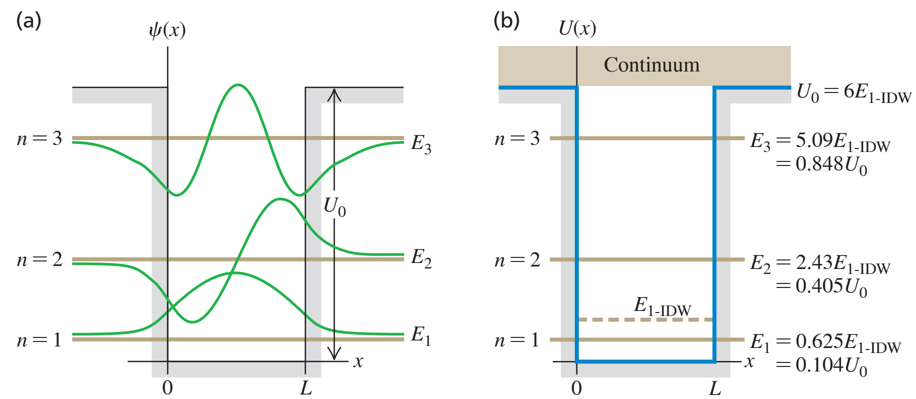
Since from Eq. (40.40) $\kappa^2 = 2m(U_0 - E)/\hbar^2$, this is equal to the right-hand side of the equation. The equation is satisfied, and $\psi(x)$ is a solution.

(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 (Section 5.2) and 5.13 (Section 5.3)]. Limiting cases are no less important in quantum mechanics.

KEYCONCEPT A finite potential well is a region where the potential energy is constant and less than the constant value outside that region. If a particle is in a bound state, so that in Newtonian mechanics it would not have enough energy to escape from the well, its stationary-state wave function is sinusoidal within the well and exponential outside the well.

Figure **40.15** (a) Wave functions for the three bound states for a particle in a finite potential well of depth $U_0 = 6E_{1-\mathrm{IDW}}$. (Here $E_{1-\mathrm{IDW}}$ 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.



CAUTION **Misconceptions about the finite potential well** Note that in our notation the energy of a particle in a finite potential well is given relative to the *bottom* of the well, which we take to be $E = 0$. (We did the same for the particle in a box, or infinite potential well, in Section 40.2.) If the depth of a finite square well is U_0, a bound state has an energy that is greater than zero but less than U_0. Note also that how far the exponential part of a bound state wave function extends outside a finite potential well depends on the energy of the bound state. The greater the bound state energy, the farther outside the well the wave function extends and the farther outside the particle is likely to be found (Fig. 40.15).

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),

$$E_{1-\mathrm{IDW}} = \frac{\pi^2\hbar^2}{2mL^2} \quad \text{(ground-level energy, infinitely deep well)} \tag{40.41}$$

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.

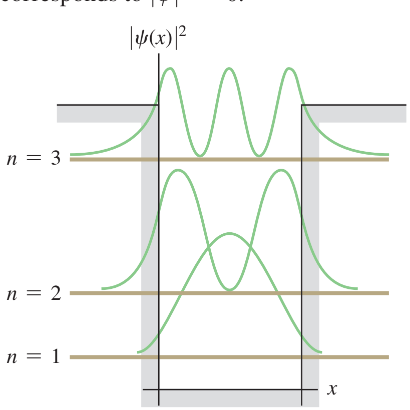
Figure 40.15 shows the case $U_0 = 6E_{1-\mathrm{IDW}}$, for which there are three bound states. In the figure, we express the energy levels both as fractions of the well depth U_0 and as multiples of $E_{1-\mathrm{IDW}}$. If the well were infinitely deep, the lowest three levels, as given by Eq. (40.31), 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.

It turns out that when U_0 is less than $E_{1-\mathrm{IDW}}$, there is only one bound state. In the limit when U_0 is *much smaller* than $E_{1-\mathrm{IDW}}$ (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

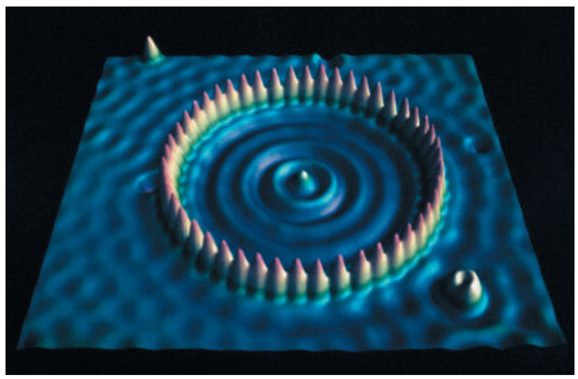
Figure **40.16** Probability distribution functions $|\psi(x)|^2$ for the square-well wave functions shown in Fig. 40.15. The horizontal brown line for each wave function corresponds to $|\psi|^2 = 0$.



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.

Figure **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.)



EXAMPLE 40.6 An electron in a finite well

WITH VARIATION PROBLEMS

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 $E_{1-\mathrm{IDW}}$ 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?

IDENTIFY and SET UP Equation (40.41) gives the ground-level energy $E_{1-\mathrm{IDW}}$ for an infinitely deep well, and Fig. 40.15b shows the energies for a square well with $U_0 = 6E_{1-\mathrm{IDW}}$. 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),

$$E_{1-\mathrm{IDW}} = \frac{\pi^2\hbar^2}{2mL^2} = \frac{\pi^2(1.055 \times 10^{-34}\ \mathrm{J \cdot s})^2}{2(9.11 \times 10^{-31}\ \mathrm{kg})(0.50 \times 10^{-9}\ \mathrm{m})^2}$$
$$= 2.4 \times 10^{-19}\ \mathrm{J} = 1.5\ \mathrm{eV}$$

(b) We have $U_0 = 6E_{1-\mathrm{IDW}} = 6(1.5\ \mathrm{eV}) = 9.0\ \mathrm{eV}$. We can read off the energy levels from Fig. 40.15b:

$$E_1 = 0.625E_{1-\mathrm{IDW}} = 0.625(1.5\ \mathrm{eV}) = 0.94\ \mathrm{eV}$$
$$E_2 = 2.43E_{1-\mathrm{IDW}} = 2.43(1.5\ \mathrm{eV}) = 3.6\ \mathrm{eV}$$
$$E_3 = 5.09E_{1-\mathrm{IDW}} = 5.09(1.5\ \mathrm{eV}) = 7.6\ \mathrm{eV}$$

(c) The photon energy and wavelength for the $n = 2$ to $n = 1$ transition are

$$E_2 - E_1 = 3.6\ \mathrm{eV} - 0.94\ \mathrm{eV} = 2.7\ \mathrm{eV}$$
$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\ \mathrm{eV \cdot s})(3.00 \times 10^8\ \mathrm{m/s})}{2.7\ \mathrm{eV}} = 460\ \mathrm{nm}$$

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 $U_0 - E_1 = 9.0\ \mathrm{eV} - 0.94\ \mathrm{eV} = 8.1\ \mathrm{eV}$, 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 $E_{1-\mathrm{IDW}} = 1.5\ \mathrm{eV}$, $E_{2-\mathrm{IDW}} = 4E_{1-\mathrm{IDW}} = 6.0\ \mathrm{eV}$, and $E_{3-\mathrm{IDW}} = 9E_{1-\mathrm{IDW}} = 13.5\ \mathrm{eV}$. 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,

Continued

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.

KEYCONCEPT A finite square well has a finite number of bound states. The energies of these bound states are determined by the boundary conditions that the wave function and its derivative are both continuous at the edges of the well.

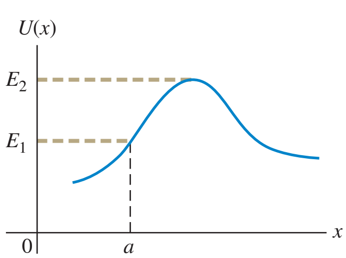
TEST YOUR UNDERSTANDING OF SECTION 40.3 Suppose that the width of the finite potential well shown in Fig. 40.15 is reduced by one-half. How must the value of U_0 change so that there are still just three bound energy levels whose energies are the fractions of U_0 shown in Fig. 40.15b? U_0 must (i) increase 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.

ANSWER

(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 POTENTIAL BARRIERS AND TUNNELING

Figure 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.



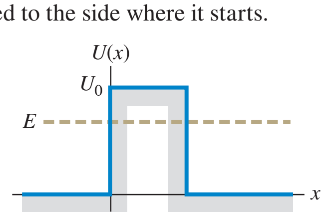
A **potential barrier** is the opposite of a potential well; it is a potential-energy function with a *maximum*. **Figure 40.18** shows an example. In classical Newtonian mechanics, if a particle (such as a roller coaster) is located to the left of the barrier (which might be a hill), and if the total 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 mechanical 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.

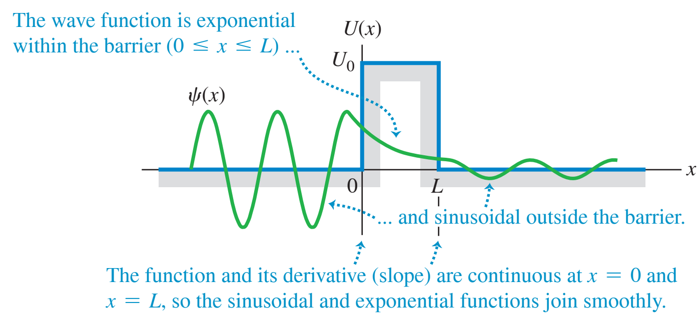
Figure 40.19 A rectangular potential-energy barrier with width L and height U_0. According to Newtonian mechanics, if the total energy E is less than U_0, a particle cannot pass over this barrier but is confined to the side where it starts.



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.

Figure **40.20** A possible wave function for a particle tunneling through the potential-energy barrier shown in Fig. 40.19.



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

$$T = Ge^{-2\kappa L} \text{ where } G = 16\frac{E}{U_0}\left(1 - \frac{E}{U_0}\right) \text{ and } \kappa = \frac{\sqrt{2m(U_0 - E)}}{\hbar} \quad (40.42)$$

(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.

EXAMPLE 40.7 Tunneling through a barrier

WITH VARIATION PROBLEMS

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?

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 $E = 2.0$ eV (electron energy), $U = 5.0$ eV (barrier height), $m = 9.11 \times 10^{-31}$ kg (mass of the electron), and $L = 1.00$ nm or 0.50 nm (barrier width).

EXECUTE First we evaluate G and κ in Eq. (40.42), using $E = 2.0$ 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 $L = 1.00\text{ nm} = 1.00 \times 10^{-9}$ 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 $L = 0.50$ 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. Note that our results for T in both part (a) and (b) are much less than 1, which is consistent with the assumptions that go into Eq. (40.42).

KEYCONCEPT Quantum-mechanical particles can be found in places where Newtonian mechanics would forbid them to be. An example is tunneling, in which a particle can pass through a barrier where its total mechanical energy is less than the potential energy. The probability that tunneling occurs depends on the width of the barrier and the additional energy the particle would need to "climb" over the barrier.

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.

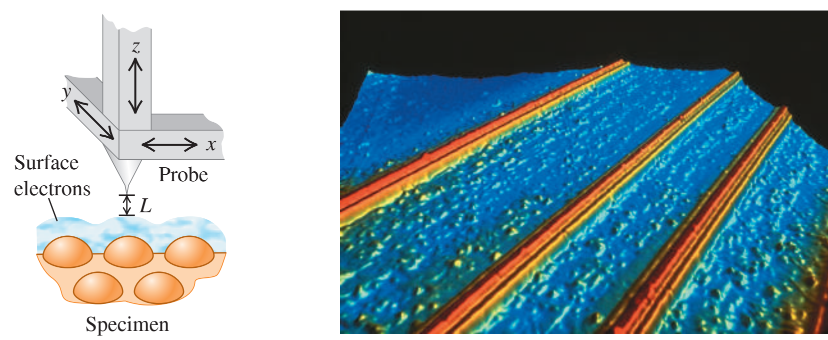
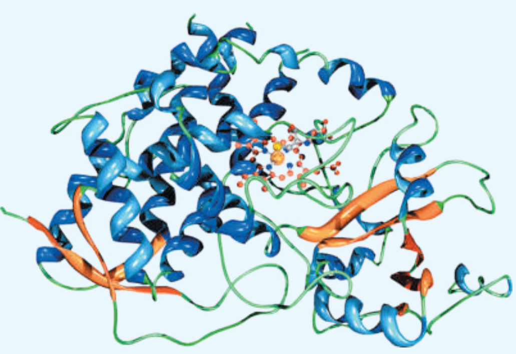


Figure **40.21** (a) Schematic diagram of the probe of a scanning tunneling microscope (STM). As the sharp conducting probe is scanned across the surface in the x- and y-directions, it is also moved in the z-direction to maintain a constant tunneling current. The changing position of the probe is recorded 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.

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!

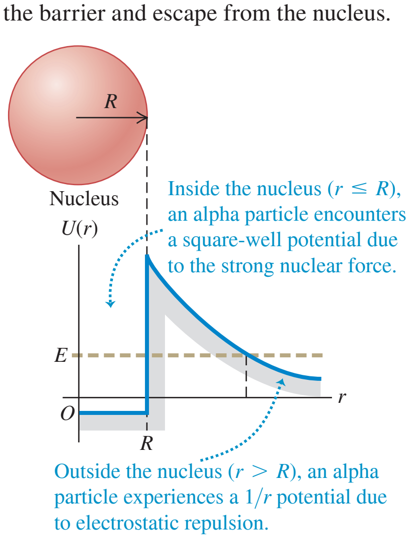


A *Josephson junction* consists of two superconductors separated by an oxide layer a few atoms (1 to 2 nm) thick. Electron pairs in the superconductors can tunnel through the barrier layer, giving such a device unusual circuit properties. Josephson junctions are useful for establishing precise voltage 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 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.

Figure **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?

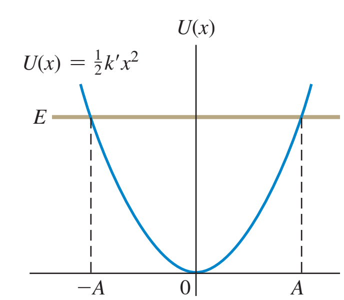
ANSWER **yes** Figure 40.20 shows a possible wave function $\psi(x)$ for tunneling. Since $\psi(x)$ is not zero within the barrier ($0 \le x \le L$), there is some probability that the particle can be found there.

40.5 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 $E = \frac{1}{2}k'A^2$.

Figure **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 $x < -A$.



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 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:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}k'x^2\psi(x) = E\psi(x) \qquad \text{(Schrödinger equation for the harmonic oscillator)} \tag{40.44}$$

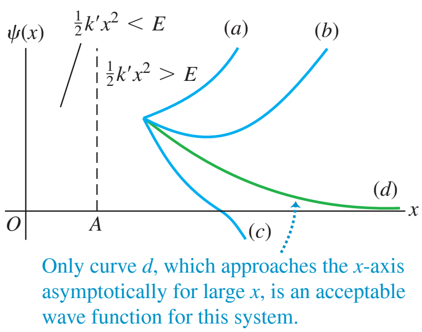
The solutions of this equation are wave functions for the physically possible states of the system.

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 $\psi(x) \to 0$ as $|x| \to \infty$ is not as trivial as it may seem. To see why this is, let's rewrite Eq. (40.44) in the form

$$\frac{d^2\psi(x)}{dx^2} = \frac{2m}{\hbar^2}\left(\tfrac{1}{2}k'x^2 - E\right)\psi(x) \tag{40.45}$$

Figure **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 $\psi(x) \to 0$ as $|x| \to \infty$, 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 $\psi(x) \to 0$ as $|x| \to \infty$ 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.

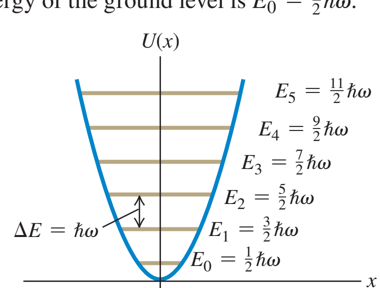
This qualitative discussion suggests how the boundary conditions as $|x| \to \infty$ determine the possible energy levels for the quantum-mechanical harmonic oscillator. It turns out that these boundary conditions are satisfied only if the energy E is equal to one of the values E_n:

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

Figure **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 $E_0 = \frac{1}{2}\hbar\omega$.



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.

EXAMPLE 40.8 Vibration in a crystal

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?

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'll explore this idea further in Chapter 42.

KEYCONCEPT As in Newtonian physics, in a quantum-mechanical harmonic oscillator a particle of mass m moves under the influence of the potential-energy function of an ideal spring with force constant k'. The lowest-energy stationary state of the harmonic oscillator has energy $\frac{1}{2}\hbar\omega$, where $\omega = \sqrt{k'/m}$ is the Newtonian angular frequency; each excited state has $\hbar\omega$ more energy than the state below it.

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

$$\psi(x) = Ce^{-\sqrt{mk'}\,x^2/2\hbar} \qquad (40.47)$$

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 $E_0 = \frac{1}{2}\hbar\omega$. It's a little messy, but the result is satisfying and worth the effort.

Figure 40.26 (next page) 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

$$\tfrac{1}{2}k'A^2 = \left(n + \tfrac{1}{2}\right)\hbar\omega \qquad (40.48)$$

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 **40.26** 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.

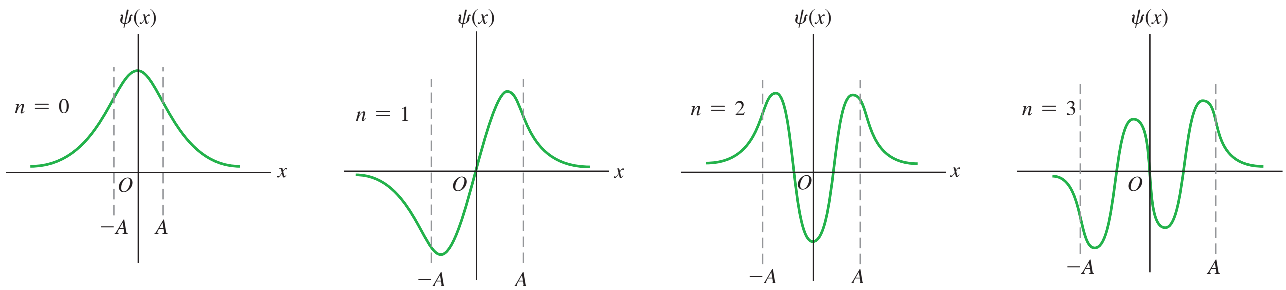


Figure **40.27** Probability distribution functions $|\psi(x)|^2$ for the harmonic-oscillator wave functions shown in Fig. 40.26. The amplitude A of the Newtonian motion with the same energy is shown for each. The blue lines show the corresponding probability distributions for the Newtonian motion. As n increases, the averaged-out quantum-mechanical functions resemble the Newtonian curves more and more.

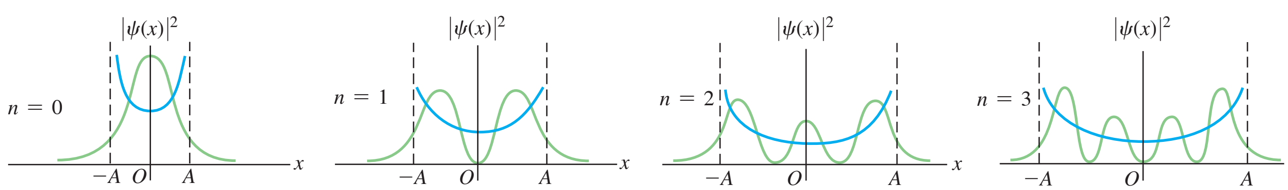


Figure **40.28** Newtonian and quantum-mechanical probability distribution functions for a harmonic oscillator for the state $n = 10$. The Newtonian amplitude A is also shown.

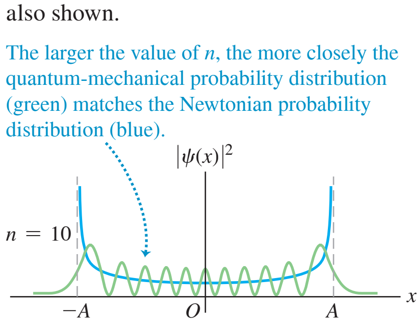


Figure 40.27 shows the probability distributions $|\psi(x)|^2$ for these states. Each graph also shows the probability distribution determined from Newtonian physics, in which the probability of finding the particle near a randomly chosen point is inversely proportional to the particle's speed at that 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.

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

$$E = \tfrac{1}{2}k'A^2 = \tfrac{1}{2}\hbar\omega = \tfrac{1}{2}\hbar\left(\frac{k'}{m}\right)^{1/2} \quad \text{so} \quad A = \frac{\hbar^{1/2}}{k'^{1/4}m^{1/4}}$$

$$E = \tfrac{1}{2}mv_{\text{max}}{}^2 = \tfrac{1}{2}k'A^2 \quad \text{so} \quad v_{\text{max}} = A\left(\frac{k'}{m}\right)^{1/2} = \frac{\hbar^{1/2}k'^{1/4}}{m^{3/4}}$$

The maximum *momentum* of the particle is

$$p_{\text{max}} = mv_{\text{max}} = \hbar^{1/2}k'^{1/4}m^{1/4}$$

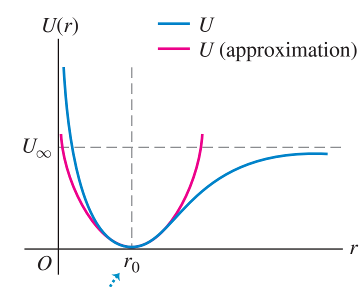
Here's where the Heisenberg uncertainty principle comes in. It turns out that the uncertainties in the particle's position and momentum (calculated as standard deviations) are, respectively, $\Delta x = A/\sqrt{2} = A/2^{1/2}$ and $\Delta p_x = p_{\max}/\sqrt{2} = p_{\max}/2^{1/2}$. Then the product 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), $\Delta x \, \Delta p_x \ge \hbar/2$, 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.

Figure **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.

ANSWER

(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 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'll 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:

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin k_nx$$

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 $e^{-i\omega_nt}$:

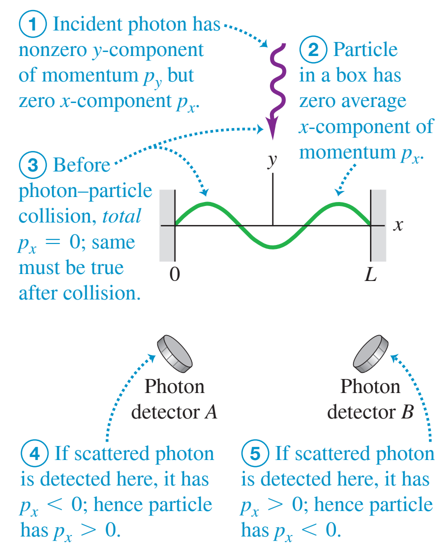
$$\begin{aligned}\Psi(x, t) &= \frac{1}{i\sqrt{2L}}(e^{ik_nx} - e^{-ik_nx})e^{-i\omega_nt} \\ &= \frac{1}{i\sqrt{2L}}e^{ik_nx}e^{-i\omega_nt} - \frac{1}{i\sqrt{2L}}e^{-ik_nx}e^{-i\omega_nt}\end{aligned} \tag{40.53}$$

In Eq. (40.53) the $e^{ik_nx}e^{-i\omega_nt}$ 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 $e^{-ik_nx}e^{-i\omega_nt}$ 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 $e^{ik_nx}e^{-i\omega_nt}$ and $e^{-ik_nx}e^{-i\omega_nt}$ 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 $p_x = -\hbar k_n$.

Figure **40.30** Using photon scattering to measure the x-component of momentum of a particle in a box.



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 $= 1.00$ μm, the electron momentum has a minimum magnitude of $p = 3.31 \times 10^{-28}$ kg $\cdot$ 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.

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 $e^{ik_nx}e^{-i\omega_nt}$ 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 $e^{-ik_nx}e^{-i\omega_nt}$ 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 $e^{ik_nx}e^{-i\omega_nt}$ nor $e^{-ik_nx}e^{-i\omega_nt}$ 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'll measure that value of energy.

You may ask, Does the wave function really collapse? Many 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'll 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.

ANSWER

(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 $|C| < |D|$.

CHAPTER 40 SUMMARY

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.)

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + U(x)\Psi(x,t) = i\hbar\frac{\partial\Psi(x,t)}{\partial t} \qquad (40.20)$$

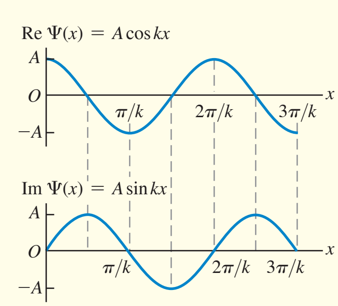
(general 1-D Schrödinger equation)

$$\Psi(x,t) = \psi(x)e^{-iEt/\hbar} \qquad (40.21)$$

(time-dependent wave function for a state of definite energy)

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x) \qquad (40.23)$$

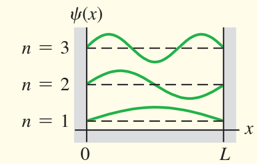
(time-independent 1-D Schrödinger equation)



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.)

$$E_n = \frac{p_n{}^2}{2m} = \frac{n^2h^2}{8mL^2} = \frac{n^2\pi^2\hbar^2}{2mL^2} \quad (n = 1, 2, 3, \ldots) \tag{40.31}$$

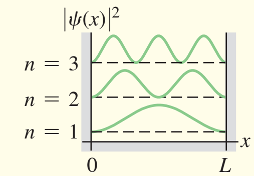
$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin\frac{n\pi x}{L} \quad (n = 1, 2, 3, \ldots) \tag{40.35}$$



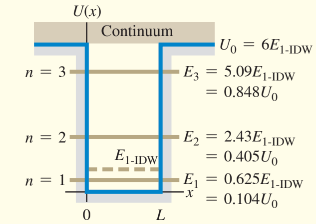
Wave functions and normalization: To be a solution of the Schrödinger equation, the wave function $\psi(x)$ and its derivative $d\psi(x)/dx$ must be continuous everywhere except where the potential-energy function $U(x)$ has an infinite discontinuity. Wave functions are usually normalized so that the total probability of finding the particle somewhere is unity.

$$\int_{-\infty}^{\infty} |\psi(x)|^2\, dx = 1 \tag{40.33}$$

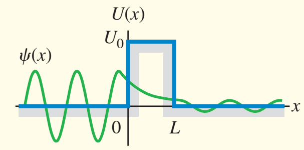
(normalization condition)



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.)

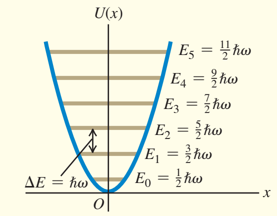


Potential barriers and tunneling: There is a certain probability that a particle will penetrate a potential-energy barrier even though its initial energy is less than the barrier height. This process is called tunneling. (See Example 40.7.)



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.)

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m}} = \left(n + \tfrac{1}{2}\right)\hbar\omega \quad (n = 0, 1, 2, 3, \ldots) \tag{40.46}$$



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.

Chapter 40 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE √ARIATION PROBLEMS

Be sure to review EXAMPLES 40.3 and 40.4 (Section 40.2) before attempting these problems.

VP40.4.1 The lowest energy level for an electron confined to a one-dimensional box is 2.00×10^{-19} J. Find (a) the width of the box and (b) the energies of the $n = 2$ and $n = 3$ energy levels.

VP40.4.2 A proton (mass 1.67×10^{-27} kg) is confined to a one-dimensional box of width 5.00×10^{-15} m. Find the energy difference between (a) the $n = 2$ and $n = 1$ energy levels and (b) the $n = 3$ and $n = 2$ energy levels.

VP40.4.3 A photon is emitted when an electron in a one-dimensional box transitions from the $n = 2$ energy level to the $n = 1$ energy level.

The wavelength of this photon is 655 nm. Find (a) the energy of this photon, (b) the width of the box, and (c) the wavelength of the photon emitted when the electron transitions from the $n = 3$ level to the $n = 2$ level.

VP40.4.4 A wave function for a particle in a box with energy E is

$$\psi(x) = A\cos\left(\frac{x\sqrt{2mE}}{\hbar} + \phi\right)$$

(a) Find the second derivative of $\psi(x)$ and show that this function satisfies the Schrödinger equation. (b) Use the boundary condition at $x = 0$ to determine the value of ϕ. (c) Use the boundary condition at $x = L$ to determine the value of E.

Be sure to review EXAMPLE 40.6 (Section 40.3) before attempting these problems.

VP40.6.1 An electron is confined to a potential well of width 0.350 nm. (a) Find the ground-level energy $E_{1-\text{IDW}}$ if the well is infinitely deep. If instead the depth U_0 of the well is six times the value of $E_{1-\text{IDW}}$, find (b) the ground-level energy and (c) the minimum energy required to free the electron from the well.

VP40.6.2 A finite potential well has a depth U_0 that is six times greater than the ground-level energy $E_{1-\text{IDW}}$ of an electron in an infinitely deep well of the same width. The photon emitted when an electron in the finite potential well transitions from the $n = 3$ level to the $n = 2$ level has energy 2.50×10^{-19} J. Find (a) the value of $E_{1-\text{IDW}}$, (b) the value of U_0, and (c) the width of the potential well.

VP40.6.3 An electron in a finite potential well can emit a photon of only one of three different wavelengths when it makes a transition from one energy level to a lower energy level. The depth of this potential well is six times greater than the ground-level energy $E_{1-\text{IDW}}$ that the electron would have if the well had the same width but was infinitely deep. (a) If the shortest of the three wavelengths is 355 nm, find the quantum numbers n for the initial and final electron energy levels for this transition. (b) Find the other two wavelengths and the quantum numbers n of the initial and final energy levels for each transition.

VP40.6.4 An electron is in a potential well that is 0.400 nm wide. (a) Find the ground-level energy of the electron $E_{1-\text{IDW}}$, in eV, if the potential well is infinitely deep. (b) If the depth of the potential is only 0.015 eV, much less than $E_{1-\text{IDW}}$, there is only one bound state. Find the energy, in eV, of this bound state.

Be sure to review EXAMPLE 40.7 (Section 40.4) before attempting these problems.

VP40.7.1 An electron of energy 3.75 eV encounters a potential barrier 6.10 eV high. Find the probability that the electron will tunnel through the barrier if the barrier width is (a) 0.750 nm and (b) 0.500 nm.

VP40.7.2 A 3.50 eV electron encounters a barrier with height 4.00 eV. Find (a) the probability that the electron will tunnel through the barrier if its width is 0.800 nm and (b) the barrier width that gives twice the tunneling probability you found in part (a).

VP40.7.3 A potential barrier is 5.00 eV high and 0.900 nm wide. Find the probability that an electron will tunnel through this barrier if its energy is (a) 4.00 eV, (b) 4.30 eV, and (c) 4.60 eV.

VP40.7.4 There is a 1-in-417 probability that an electron with energy 3.00 eV will tunnel through a barrier with height 5.00 eV. Find the width of the barrier.

BRIDGING PROBLEM A Packet in a Box

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 $T = 2\pi/\omega_{\text{osc}}$ 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

•, ••, •••: 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 objects 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.

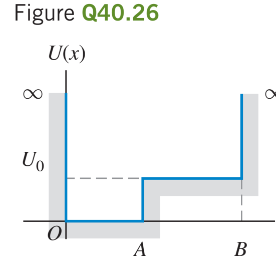
Q40.23 The probability distributions for the harmonic-oscillator wave functions (see Figs. 40.27 and 40.28) begin to resemble the classical (Newtonian) probability distribution when the quantum number n becomes large. Would the distributions become the same as in the classical case in the limit of very 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 • **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

40.7 • An electron is in a one-dimensional box. When the electron is in its ground state, the longest-wavelength photon it can absorb is 420 nm. What is the next longest-wavelength photon it can absorb, again starting in the ground state?

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 $(m = 0.20\ \text{kg})$ 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 • A particle with mass m is in a one-dimensional box with width L. If the energy of the particle is $9\pi^2\hbar^2/2mL^2$, (a) what is the linear momentum of the particle and (b) what is the ratio of the width of the box to the de Broglie wavelength λ of the particle?

40.16 • Recall that $|\psi|^2 dx$ 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 • Consider two different particles in two different boxes. Particle A has nine times the mass of particle B, while particle B's box has twice the width of particle A's box. What are the two lowest quantum numbers for the two systems for which particles A and B have equal energies?

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?

40.21 • A particle is in a box with infinitely rigid walls. The walls are at $x = -L/2$ and $x = +L/2$. (a) Show that $\psi_n = A \cos k_n x$ is a possible solution. What must k_n equal? (b) Show that $\psi_n = A \sin k_n x$ is a possible solution. What must k_n equal? (c) What are the allowed energy levels for $\psi_n = A \sin k_n x$ and for $\psi_n = A \cos k_n x$? (d) How does the set of energy levels you found in part (c) compare to the energy levels given by Eq. (40.31)?

Section 40.3 Potential Wells

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 depth $U_0 = 6E_{1-\text{IDW}}$. What is the width of the well if its ground-state energy is 2.00 eV?

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 •• 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?

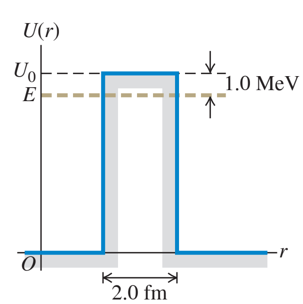
40.26 •• An electron is bound in a square well that has a depth equal to six times the ground-level energy $E_{1-\text{IDW}}$ of an infinite well of the same width. The longest-wavelength photon that is absorbed by this electron has a wavelength of 582 nm. Determine the width of the well.

40.27 •• 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.

Section 40.4 Potential Barriers and Tunneling

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 •• (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?

40.30 • 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 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?

40.32 •• 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 $0 < x < L$?

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 •• A harmonic oscillator with mass m and force constant k' is in an excited state that has quantum number n. (a) Let $p_{\max} = mv_{\max}$, where $v_{\max}$ is the maximum speed calculated in the Newtonian analysis of the oscillator. Derive an expression for $p_{\max}$ in terms of n, $\hbar$, k', and m. (b) Derive an expression for the classical amplitude A in terms of n, $\hbar$, k', and m. (c) If $\Delta x = A/\sqrt{2}$ and $\Delta p_x = p_{\max}/\sqrt{2}$, what is the uncertainty product $\Delta x \Delta p_x$? How does the uncertainty product depend on n?

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 •• 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.

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?

PROBLEMS

40.41 •• 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.

40.42 •• Hydrogen emits radiation with four prominent visible wavelengths—one red, one cyan, one blue, and one violet. The respective frequencies are 656 nm, 486 nm, 434 nm, and 410 nm. We can model the hydrogen atom as an electron in a one-dimensional box, and attempt to match four adjacent emission lines in the predicted spectrum to the visible part of the hydrogen spectrum. (a) Determine the photon energies associated with the visible part of the hydrogen spectrum. (b) The electron-in-a-box emission spectrum is $E_{n_i \to n_f} = (n_i^2 - n_f^2)\epsilon$, where n_i and n_f are the initial and final quantum numbers of the electron when it drops to a lower energy level and ϵ is the energy determined by the Schrödinger equation. What is the smallest possible value of n_i that can accommodate four emission lines? (c) Using the value from part (b) for n_i, estimate the order of magnitude of ϵ by dividing the four photon energies by the relevant differences $n_i^2 - n_f^2$ for transitions in the possible sets of (n_i, n_f) pairings. (d) Using Eq. (40.31) to identify ϵ, and using the mass of the electron, use your result from part (c) to estimate the length L of the box. (e) What is the ratio of your estimate of L to twice the Bohr radius? (*Note:* The hydrogen atom is better modeled using the Coulomb potential rather than as a particle in a box.)

40.43 •• 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.

40.44 • 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.

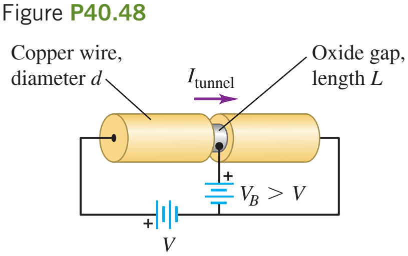
40.45 • Photon in a Dye Laser. An electron in a long, organic molecule used in a dye laser behaves approximately like a particle in a box with width 4.18 nm. What is the wavelength of the photon emitted when the electron undergoes a transition (a) from the first excited level to the ground level and (b) from the second excited level to the first excited level?

40.46 •• 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 $|\psi|^2 dx$ that the particle will be found in the interval x to $x + dx$ for (a) $x = L/4$; (b) $x = L/2$; (c) $x = 3L/4$.

40.47 •• Repeat Problem 40.46 for a particle in the first excited level.

40.48 •• Suppose we have two copper wires that meet on either side of an oxide gap with length $L = 1$ nm, as shown in **Fig. P40.48**. If the left side is at a potential V relative to the right side, then we can estimate the tunneling current as follows: The wires have diameter $d = 1$ mm, and copper has $n = 8.5 \times 10^{28}$ free electrons per m^3. (a) Assume that the oxide gap provides a rectangular potential barrier of height 12 eV and that the energy of the free electrons incident on this barrier is 5.0 eV. Assume the gap has length $L = 1$ nm. What is the probability T that an electron will tunnel through the barrier? (b) A length x of wire on the left of the barrier will have nAx free electrons, where $A = \pi(d/2)^2$ is the cross-sectional area of the wire. In a similar length of wire on the right side of the barrier, the number of tunneled electrons will be T times that number. If these electrons move with effective velocity $v = dx/dt$, then what will be the tunneling current, expressed symbolically in terms of T, n, A, e, and v? (c) Define their effective speed v by $\frac{1}{2}mv^2 = eV$. The tunneled electrons have energy eV. What is their effective speed v, expressed symbolically in terms of e, V, and m, where m is the mass of an electron? (f) Put these elements together to estimate the tunneling current.

Figure **P40.48**



40.49 •• 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.

40.50 •• The *penetration distance* η in a finite potential well is the distance at which the wave function has decreased to $1/e$ of the (b) 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.

40.51 •• 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

$$\psi(x) = \begin{cases} e^{+\kappa x}, & x < 0 \\ e^{-\kappa x}, & x \geq 0 \end{cases}$$

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.

40.52 • 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?

40.53 ••• CP Consider a one-dimensional lattice of carbon nuclei with interatomic spacing $d = 0.500$ nm. The potential energy of one nucleus due to electrostatic interaction with its two nearest neighbors is

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

where $q = 6e$ is the charge of each carbon nucleus and x is its position relative to its equilibrium position. (a) Sketch the potential $V(x)$ between $x = -d$ and $x = d$. (b) At low energy, the nucleus remains near its equilibrium position at $x = 0$ and the potential energy is well approximated as a harmonic oscillator. Use the power series in Appendix B to write $V(x)$ as a quadratic function in x, dropping contributions at higher order in x/d. (c) What is the spring constant k'? (d) What is the energy of the classical ground state, in eV? (e) What is the energy of the lowest-energy photons emitted when this system transitions between energy levels? (f) What is the wavelength of the corresponding radiation?

40.54 • 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?

40.55 • 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?

40.56 •• **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 $n = 0$.

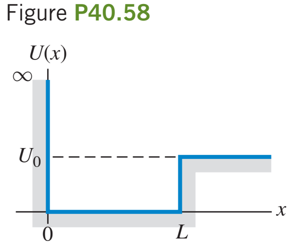
40.57 •• **CP** (a) The wave nature of particles results in the quantum-mechanical situation that a particle confined in a box can assume only wavelengths that result in standing waves in the box, with nodes at the box walls. Use this to show that an electron confined in a one-dimensional box of length 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?

40.58 ••• 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. P40.58**). 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 $x \rightarrow \infty$. 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 **P40.58**



40.59 •• **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?

40.60 •• **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.

40.61 •• **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 $T = 1$.

40.62 ••• **CP** A particle is confined to move on a circle with radius R but is otherwise free. We can parameterize points on this circle by using the distance x from a reference point or by using the angle $\theta = x/R$. Since $x = 0$ and $x = 2\pi R$ describe the same point, the wave function must satisfy $\psi(x) = \psi(x + 2\pi R)$ and $\psi'(x) = \psi'(x + 2\pi R)$. (a) Solve the free-particle time-independent Schrödinger equation subject to these boundary conditions. You should find solutions $\psi_n^{\pm}$, where n is a positive integer and where lower n corresponds to lower energy. Express your solutions in terms of θ using unspecified normalization constants A_n^+ and A_n^- corresponding, respectively, to modes that move "counterclockwise" toward higher x and "clockwise" toward lower x. (b) Normalize these functions to determine $A_n^{\pm}$. (c) What are the energy levels E_n? (d) Write the time-dependent wave functions $\Psi_n^{\pm}(x, t)$ corresponding to $\Psi_n^{\pm}(x)$. Use the symbol ω for $E_1/\hbar$. (e) Consider the nonstationary state defined at $t = 0$ by $\Psi(x, 0) = \frac{1}{\sqrt{2}}[\Psi_1^+(x) + \Psi_2^+(x)]$. Determine the probability density $|\Psi(x, t)|^2$ in terms of R, ω, and t. Simplify your result using the identity $1 + \cos\alpha = 2\cos^2(\alpha/2)$. (f) With what angular speed does the density peak move around the circle? (g) If the particle is an electron and the radius is the Bohr radius, then with what speed does its probability peak move?

CHALLENGE PROBLEMS

40.63 ••• **CALC** The Schrödinger equation for the quantum-mechanical harmonic oscillator in Eq. (40.44) may be written as

$$-\frac{\hbar^2}{2m}\psi'' + \frac{1}{2}m\omega^2 x^2 \psi = E\psi$$

where $\omega = \sqrt{k'/m}$ is the angular frequency for classical oscillations. Its solutions may be cleverly addressed using the following observations: The difference in applying two operations A and B, in opposite order, called a commutator, is denoted by $[A, B] = AB - BA$. For example, if A represents differentiation, so that $Af = \dfrac{df}{dx}$, where $f = f(x)$ is a function, and B represents multiplication by x, then $[A, B] = \left[\dfrac{d}{dx}, x\right]f = \dfrac{d}{dx}(xf) - x\dfrac{df}{dx}$. Applying the product rule for differentiation on the first term, we determine $\left[\dfrac{d}{dx}, x\right]f = f$, which we summarize by writing $\left[\frac{d}{dx}, x\right] = 1$. Similarly, consider the two operators a_+ and a_- defined by $a_{\pm} = 1/\sqrt{2m\hbar\omega}\left(\mp\hbar\frac{d}{dx} + m\omega x\right)$. (a) Determine the commutator $[a_-, a_+]$. (b) Compute $a_+a_-\psi$ in terms of ψ and ψ'', where $\psi = \psi(x)$ is a wave function. Be mindful that $\frac{d}{dx}(x\psi) = \psi + x\psi'$. (c) The Schrödinger equation above may be written as $H\psi = E\psi$, where H is a differential operator. Use your previous result to write H in terms of a_+ and a_-. (d) Determine the commutators $[H, a_{\pm}]$. (e) Consider a wave function ψ_n with definite energy E_n, whereby $H\psi_n = E_n\psi_n$. If we define a new function $\psi_{n+1} = a_+\psi_n$, then we can readily determine its energy by computing $H\psi_{n+1} = Ha_+\psi_n = (a_+H + [H, a_+])\psi_n$ and comparing the result with $E_{n+1}\psi_{n+1}$. Finish this calculation by inserting what we have determined to find E_{n+1} in terms of E_n and n. (f) We can prove that the lowest-energy wave function ψ_0 is annihilated by a_-, meaning that $a_-\psi_0 = 0$. Using your expression for H in terms of $a_{\pm}$, determine the ground-state energy E_0. (g) By applying a_+ repeatedly, we can generate all higher states. The energy of the nth excited state can be determined by considering $Ha_+^n\psi_0 = E_0\psi_0$. Using the above results, determine E_n. (*Note:* We did not have to solve any differential equations or take any derivatives to determine the spectrum of this theory.)

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

$$\int \sqrt{A^2 - x^2}\, dx = \tfrac{1}{2}\left[x\sqrt{A^2 - x^2} + A^2 \arcsin\left(\frac{x}{|A|}\right)\right]$$

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?

MCAT-STYLE 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.

Key Example √ARIATION Problems

VP40.4.1 (a) 0.549 nm (b) $E_2 = 8.00 \times 10^{-19}$ J, $E_3 = 1.80 \times 10^{-18}$ J

VP40.4.2 (a) 3.94×10^{-12} J (b) 6.57×10^{-12} J

VP40.4.3 (a) 3.03×10^{-19} J (b) 0.772 nm (c) 393 nm

VP40.4.4 (a) $\dfrac{d^2\psi(x)}{dx^2} = -\dfrac{2mE}{\hbar^2}A\cos\left(\dfrac{x\sqrt{2mE}}{\hbar} + \phi\right) = -\dfrac{2mE}{\hbar^2}\psi(x)$, so $-\dfrac{\hbar^2}{2m}\dfrac{d^2\psi(x)}{dx^2} = E\psi(x)$ (b) $\phi = \pm\pi/2$ (c) $\dfrac{L\sqrt{2mE}}{\hbar} = n\pi$, so $E = \dfrac{n^2\pi^2\hbar^2}{2mL^2}$, where $n = 1, 2, 3, \ldots$

VP40.6.1 (a) 4.92×10^{-19} J (b) 3.07×10^{-19} J (c) 2.64×10^{-18} J

VP40.6.2 (a) 9.40×10^{-20} J (b) 5.64×10^{-19} J (c) 0.800 nm

VP40.6.3 (a) initial $n = 3$, final $n = 1$ (b) 596 nm for initial $n = 3$, final $n = 2$; 878 nm for initial $n = 2$, final $n = 1$

VP40.6.4 (a) 2.35 eV (b) 0.010 eV

VP40.7.1 (a) 2.90×10^{-5} (b) 1.47×10^{-3}

VP40.7.2 (a) 5.33×10^{-3} (b) 0.704 nm

VP40.7.3 (a) 2.54×10^{-4} (b) 8.59×10^{-4} (c) 3.46×10^{-3}

VP40.7.4 0.509 nm

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 OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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:

$$\int |\Psi(x, y, z, t)|^2\, dV = 1 \tag{41.3}$$

(normalization condition in three dimensions)

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 $e^{-iEt/\hbar}$:

$$\Psi(x, y, z, t) = \psi(x, y, z)e^{-iEt/\hbar} \tag{41.4}$$

(time-dependent wave function for a state of definite energy)

(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

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi(x, y, z)}{\partial x^2} + \frac{\partial^2\psi(x, y, z)}{\partial y^2} + \frac{\partial^2\psi(x, y, z)}{\partial z^2}\right) + \underbrace{U(x, y, z)}_{\text{Potential-energy function}}\psi(x, y, z) = \underbrace{E}_{\text{Energy of state}}\psi(x, y, z) \tag{41.5}$$

Particle's mass

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

$$\int |\psi(x, y, z)|^2 \, dV = 1 \qquad \text{(normalization condition for a stationary state in three dimensions)} \tag{41.6}$$

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'll 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.

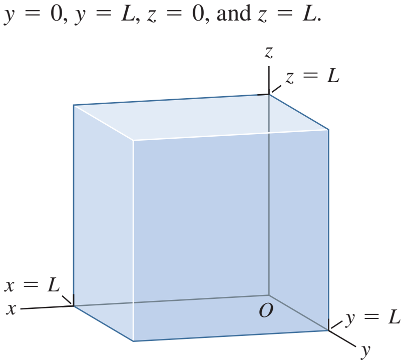
TEST YOUR UNDERSTANDING OF SECTION 41.1 In a certain region of space the potential-energy function for a quantum-mechanical particle is zero. In this region the wave function $\psi(x, y, z)$ for a certain stationary state is real and satisfies $\partial^2\psi/\partial x^2 > 0$, $\partial^2\psi/\partial y^2 > 0$, 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.

ANSWER

(ii) 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 PARTICLE IN A THREE-DIMENSIONAL BOX

Figure **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 $U(x, y, z) = 0$:

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi(x, y, z)}{\partial x^2} + \frac{\partial^2\psi(x, y, z)}{\partial y^2} + \frac{\partial^2\psi(x, y, z)}{\partial z^2}\right) = E\psi(x, y, z) \tag{41.7}$$

(particle in a three-dimensional box)

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 $z = L$.

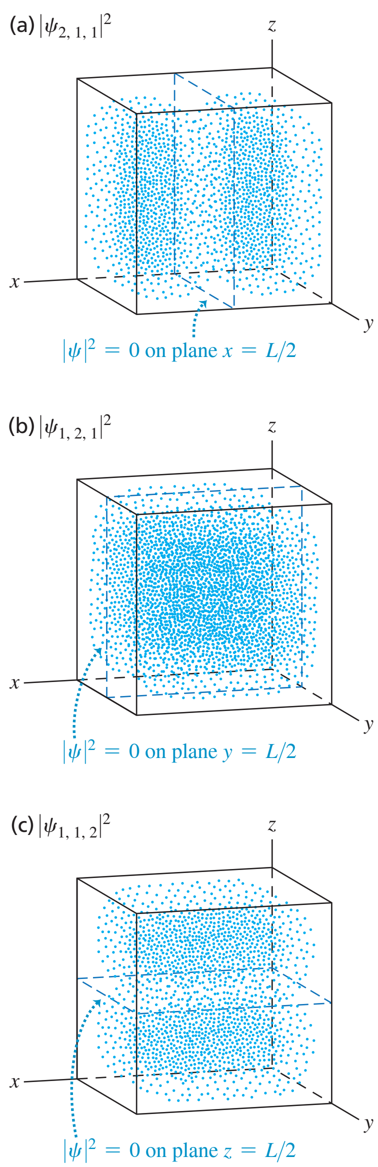
How can we interpret the three constants E_X, E_Y, and E_Z in Eqs. (41.11)? From Eq. (41.10), they are related to the energy E by

$$E_X + E_Y + E_Z = E \tag{41.12}$$

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.)

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

Figure **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.



$$X_{n_X}(x) = C_X \sin\frac{n_X \pi x}{L} \quad (n_X = 1, 2, 3, \ldots) \tag{41.13a}$$

$$Y_{n_Y}(y) = C_Y \sin\frac{n_Y \pi y}{L} \quad (n_Y = 1, 2, 3, \ldots) \tag{41.13b}$$

$$Z_{n_Z}(z) = C_Z \sin\frac{n_Z \pi z}{L} \quad (n_Z = 1, 2, 3, \ldots) \tag{41.13c}$$

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:

$$\psi_{n_X,n_Y,n_Z}(x, y, z) = C \sin\frac{n_X \pi x}{L} \sin\frac{n_Y \pi y}{L} \sin\frac{n_Z \pi z}{L} \tag{41.15}$$
$$(n_X = 1, 2, 3, \ldots;\ n_Y = 1, 2, 3, \ldots;\ n_Z = 1, 2, 3, \ldots)$$

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

WITH VARIATION PROBLEMS

(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) $(n_X, n_Y, n_Z) = (3, 1, 1)$.

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

$$\begin{aligned}
&\int |\psi_{n_X,n_Y,n_Z}(x, y, z)|^2 \, dV \\
&= |C|^2 \int_{x=0}^{x=L} \int_{y=0}^{y=L} \int_{z=0}^{z=L} \sin^2 \frac{n_X \pi x}{L} \sin^2 \frac{n_Y \pi y}{L} \sin^2 \frac{n_Z \pi z}{L} \, dx\, dy\, dz \\
&= |C|^2 \left(\int_{x=0}^{x=L} \sin^2 \frac{n_X \pi x}{L} \, dx \right) \left(\int_{y=0}^{y=L} \sin^2 \frac{n_Y \pi y}{L} \, dy \right) \\
&\quad \times \left(\int_{z=0}^{z=L} \sin^2 \frac{n_Z \pi z}{L} \, dz \right) \\
&= 1
\end{aligned}$$

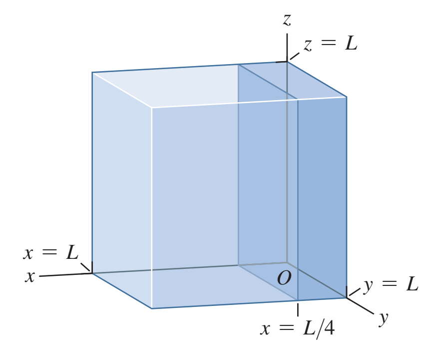
We can use the identity $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$ and the variable substitution $\theta = n_X \pi x / L$ to show that

$$\begin{aligned}
\int \sin^2 \frac{n_X \pi x}{L} \, dx &= \frac{L}{2 n_X \pi} \left[\frac{n_X \pi x}{L} - \frac{1}{2} \sin\left(\frac{2 n_X \pi x}{L} \right) \right] \\
&= \frac{x}{2} - \frac{L}{4 n_X \pi} \sin\left(\frac{2 n_X \pi x}{L} \right)
\end{aligned}$$

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

$$|C|^2 \left(\frac{L}{2}\right)\left(\frac{L}{2}\right)\left(\frac{L}{2}\right) = |C|^2 \left(\frac{L}{2}\right)^3 = 1$$

Figure **41.3** What is the probability that the particle is in the dark-colored quarter of the box?



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

$$\begin{aligned}
\int_{x=0}^{x=L/4} \sin^2 \frac{n_X \pi x}{L} dx &= \left[\frac{x}{2} - \frac{L}{4 n_X \pi} \sin\left(\frac{2 n_X \pi x}{L} \right) \right] \Bigg|_{x=0}^{x=L/4} \\
&= \frac{L}{8} - \frac{L}{4 n_X \pi} \sin\left(\frac{n_X \pi}{2} \right)
\end{aligned}$$

Hence the probability of finding the particle somewhere in the region $0 \le x \le L/4$ is

$$\begin{aligned}
P &= \left(\frac{2}{L}\right)^3 \left[\frac{L}{8} - \frac{L}{4 n_X \pi} \sin\left(\frac{n_X \pi}{2} \right) \right] \left(\frac{L}{2}\right)\left(\frac{L}{2}\right) \\
&= \frac{1}{4} - \frac{1}{2 n_X \pi} \sin\left(\frac{n_X \pi}{2} \right)
\end{aligned}$$

This depends only on the value of n_X, not on n_Y or n_Z. For each of the three cases we have

Continued

$$\text{(i) } n_X = 1\text{: } P = \tfrac{1}{4} - \frac{1}{2(1)\pi}\sin\left(\frac{\pi}{2}\right) = \tfrac{1}{4} - \frac{1}{2\pi}(1)$$

$$= \tfrac{1}{4} - \frac{1}{2\pi} = 0.091$$

$$\text{(ii) } n_X = 2\text{: } P = \tfrac{1}{4} - \frac{1}{2(2)\pi}\sin\left(\frac{2\pi}{2}\right) = \tfrac{1}{4} - \frac{1}{4\pi}\sin\pi$$

$$= \tfrac{1}{4} - 0 = 0.250$$

$$\text{(iii) } n_X = 3\text{: } P = \tfrac{1}{4} - \frac{1}{2(3)\pi}\sin\left(\frac{3\pi}{2}\right) = \tfrac{1}{4} - \frac{1}{6\pi}(-1)$$

$$= \tfrac{1}{4} + \frac{1}{6\pi} = 0.303$$

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.

KEYCONCEPT If a particle is in a three-dimensional stationary state described by wave function $\psi(x, y, z)$, the probability of finding that particle in an infinitesimal volume dV centered on (x, y, z) is $|\psi(x, y, z)|^2 dV$. The probability of finding the particle somewhere in a finite volume equals the integral $\int|\psi(x, y, z)|^2 dV$ evaluated over that volume.

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 $(n_X, n_Y, n_Z) = (2, 1, 1)$, $(1, 2, 1)$, and $(1, 1, 2)$. You can

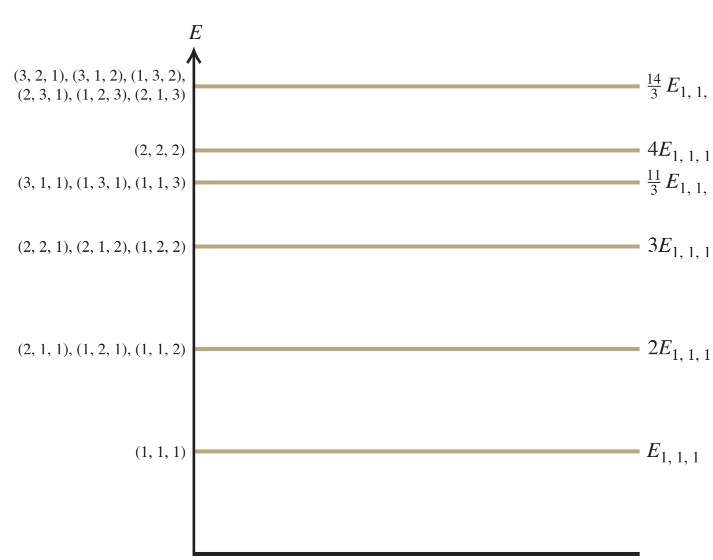


Figure **41.4** Energy-level diagram for a particle in a three-dimensional cubical box. We label each level with the quantum numbers of the states (n_X, n_Y, n_Z) with that energy. Several of the levels are degenerate (more than one state has the same energy). The lowest (ground) level, $(n_X, n_Y, n_Z) = (1, 1, 1)$, 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 $E_{1,1,1}$.

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

$$E_{n_X,n_Y,n_Z} = \left(\frac{n_X^2}{L_X^2} + \frac{n_Y^2}{L_Y^2} + \frac{n_Z^2}{L_Z^2}\right)\frac{\pi^2\hbar^2}{2m} \qquad \begin{pmatrix} n_X = 1, 2, 3, \ldots; \\ n_Y = 1, 2, 3, \ldots; \\ n_Z = 1, 2, 3, \ldots \end{pmatrix} \tag{41.17}$$

(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) $(n_X, n_Y, n_Z) = (1, 3, 3)$.

ANSWER

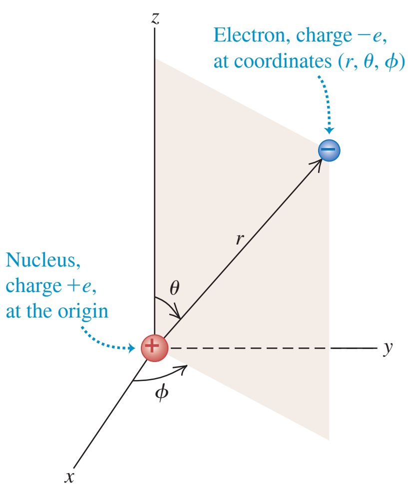
(iv), (ii), (i) and (iii) (tie) Equation (41.16) shows that the energy levels for a cubical box are proportional to the quantity $n_X^2 + n_Y^2 + n_Z^2$. Hence ranking in order of this quantity is the same as ranking in order of energy. For the four cases we are given, we have (i) $n_X^2 + n_Y^2 + n_Z^2 = 2^2 + 3^2 + 2^2 = 17$; (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 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

Figure 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:

$$U(r) = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r} \tag{41.18}$$

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:

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi) \tag{41.19}$$

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 $\Phi(\phi)$:

$$-\frac{\hbar^2}{2m_r r^2}\frac{d}{dr}\left(r^2\frac{dR(r)}{dr}\right) + \left(\frac{\hbar^2 l(l+1)}{2m_r r^2} + U(r)\right)R(r) = ER(r) \tag{41.20a}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta(\theta)}{d\theta}\right) + \left(l(l+1) - \frac{m_l{}^2}{\sin^2\theta}\right)\Theta(\theta) = 0 \tag{41.20b}$$

$$\frac{d^2\Phi(\phi)}{d\phi^2} + m_l{}^2\Phi(\phi) = 0 \tag{41.20c}$$

In Eqs. (41.20) E is the energy of the stationary state and l and m_l are constants that we'll discuss later.

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.

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'll 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

$$L = \sqrt{l(l+1)}\,\hbar \qquad (l = 0, 1, 2, \ldots, n-1) \qquad (41.22)$$

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

$$L = \sqrt{2(2+1)}\hbar = \sqrt{6}\hbar = 2.45\hbar$$

$$L_z = 2\hbar$$

Figure 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.

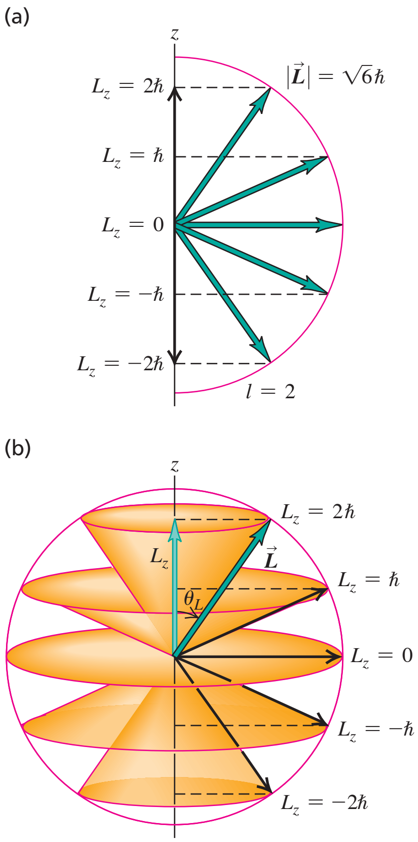


Figure 41.6 shows the situation. The minimum value of the angle θ_L between the vector $\vec{L}$ and the z-axis is

$$\theta_L = \arccos\frac{L_z}{L}$$

$$= \arccos\frac{2}{2.45} = 35.3°$$

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'll 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.

Figure **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	$3s$	
3	1	$-1, 0, 1$	$3p$	M
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.

EXAMPLE 41.2 Counting hydrogen states

WITH VARIATION PROBLEMS

How many distinct (n, l, m_l) states of the hydrogen atom with $n = 3$ are there? What are their energies?

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.

KEYCONCEPT Because the potential-energy function for the hydrogen atom is spherically symmetric, its energy levels are degenerate. Stationary states with the same principal quantum number n have the same energy, even though they have different values of the orbital quantum number l and the orbital magnetic quantum number m_l.

EXAMPLE 41.3 Angular momentum in an excited level of hydrogen

WITH VARIATION PROBLEMS

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 $\hbar$.

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.

KEYCONCEPT For each value of the principal quantum number $n = 1, 2, 3, \ldots$ for the stationary states of the hydrogen atom, there are n possible values of the orbital quantum number l that range from 0 to $n - 1$. For each value of l, there are $2l + 1$ possible values of the orbital magnetic quantum number m_l that range from $-l$ to 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:

$$dV = 4\pi r^2\,dr \tag{41.24}$$

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; Wave function; Volume of spherical shell with inner radius r, outer radius $r + dr$

$$\underbrace{P(r)\,dr}_{\text{Probability that the electron is between } r \text{ and } r+dr} = |\psi|^2\,dV = |\psi|^2 4\pi r^2\,dr \tag{41.25}$$

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:

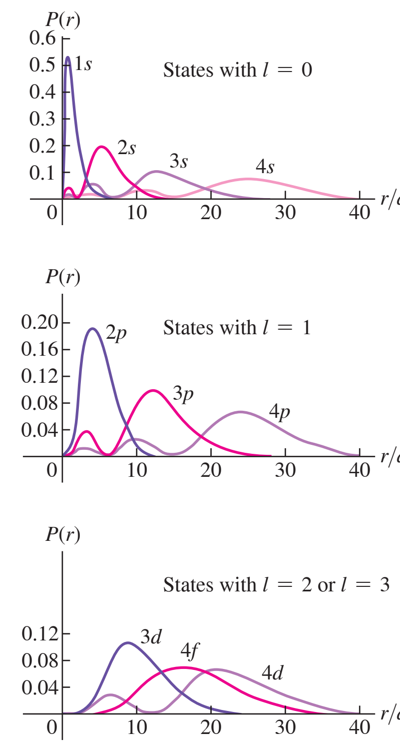
Radius of smallest orbit in Bohr model; Electric constant; Planck's constant; Planck's constant divided by 2π; Reduced mass; Magnitude of electron charge

$$a = \frac{\epsilon_0 h^2}{\pi m_r e^2} = \frac{4\pi\epsilon_0 \hbar^2}{m_r e^2} = 5.29 \times 10^{-11}\ \text{m} \tag{41.26}$$

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.

Figure **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, . . .) have only one maximum, located at $r = n^2a$.



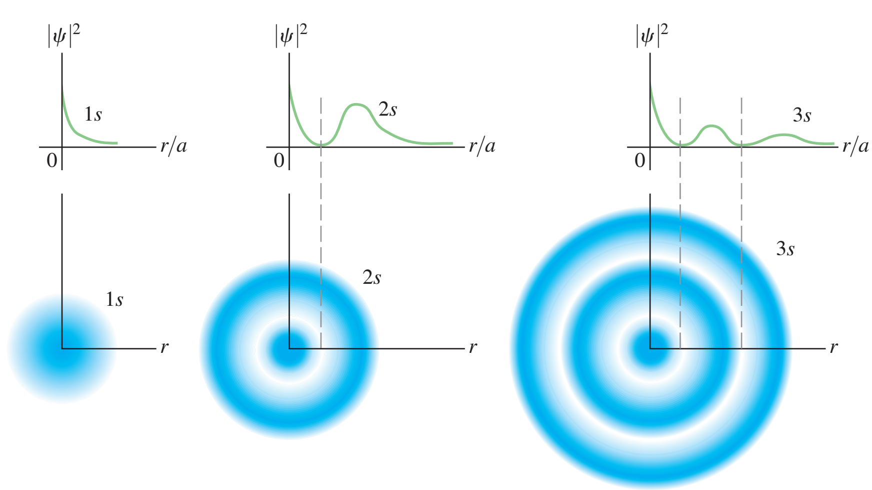
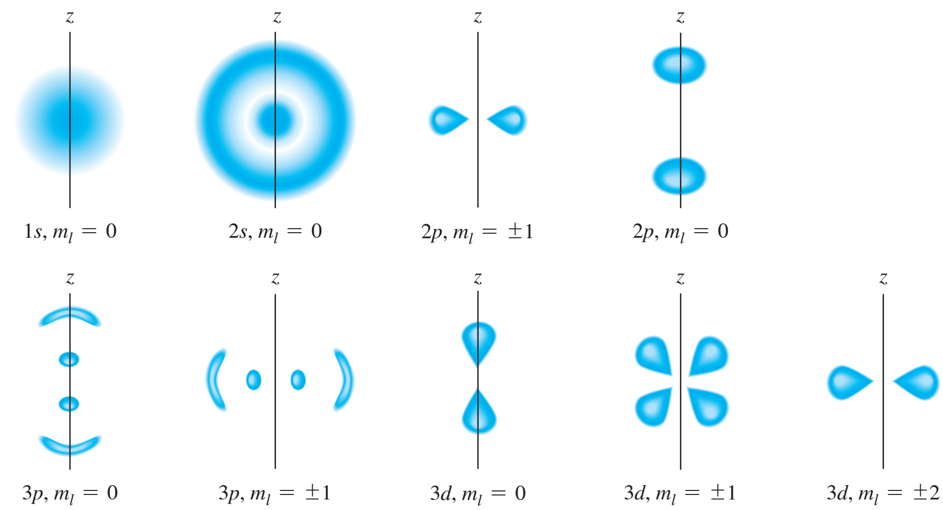


Figure **41.9** Three-dimensional probability distribution functions $|\psi|^2$ for the spherically symmetric 1s, 2s, and 3s hydrogen-atom wave functions.

Figure 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

WITH VARIATION PROBLEMS

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?

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, $r/a > 1$).

KEYCONCEPT If a particle is in a stationary state described by the wave function ψ, the probability of finding the particle at a radial coordinate between r and $r + dr$ is $P(r)dr$, where $P(r) = 4\pi r^2|\psi|^2$. If the wave function depends on the angle as well as on the radial coordinate, you must first average $|\psi|^2$ over all angles.

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) $n = 2, l = 1, m_l = 0$.

ANSWER

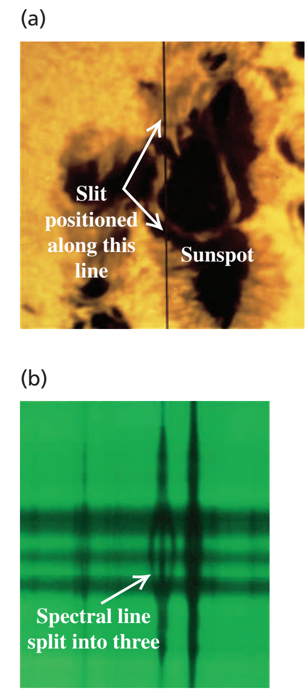
❙ **(ii) and (iii) (tie), (i)** An electron in a state with principal quantum number n is most likely to be found at $r = n^2a$. This result is independent of the values of the quantum numbers l and m_l. Hence an electron with $n = 2$ (most likely to be found at $r = 4a$) is more likely to be found near $r = 5a$ than an electron with $n = 1$ (most likely to be found at $r = a$).

41.4 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.

Figure **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.



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

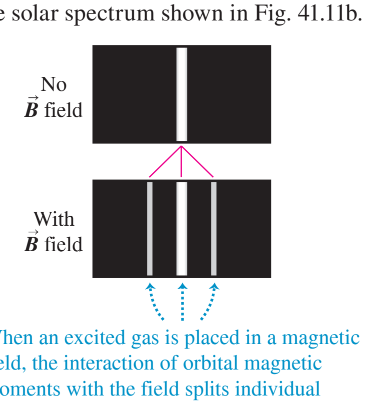
$$\mu = IA = \frac{ev}{2\pi r}\pi r^2 = \frac{evr}{2} \tag{41.29}$$

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.*

Figure **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 $\text{A}\cdot\text{m}^2$ 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 orbital 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

Magnetic dipole component in direction of $\vec{B}$

Magnitude of electron charge

Planck's constant divided by 2π

Bohr magneton

$$U = -\mu_z B = m_l \frac{e\hbar}{2m} B = m_l \mu_B B \tag{41.35}$$

Magnetic-field magnitude

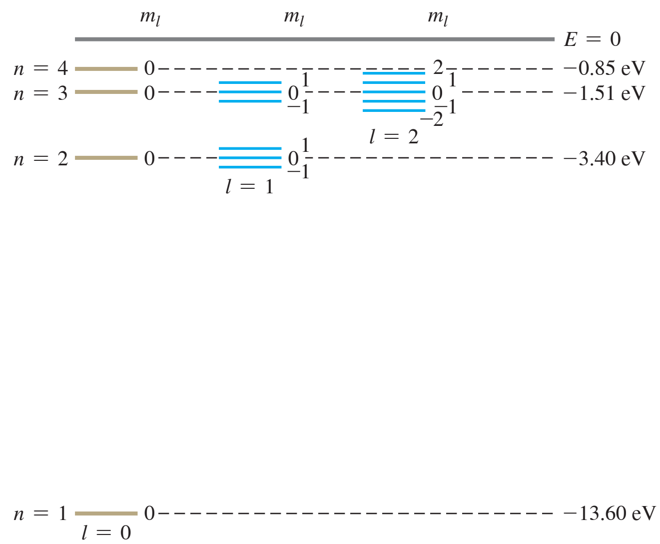
Electron mass

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; 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

Figure **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?



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

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?

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

$$\begin{aligned} U &= m_l \mu_B B = m_l (5.788 \times 10^{-5}\ \text{eV/T})(2.00\ \text{T}) \\ &= m_l (1.16 \times 10^{-4}\ \text{eV}) = m_l (1.85 \times 10^{-23}\ \text{J}) \end{aligned}$$

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).

KEYCONCEPT An electron in an atom has a nonzero magnetic moment if it is in a state with a nonzero orbital angular momentum. If this atom is in a magnetic field $\vec{B}$, the electron energy depends on the electron's orbital magnetic quantum number m_l, which describes how the magnetic moment vector is oriented with respect to $\vec{B}$.

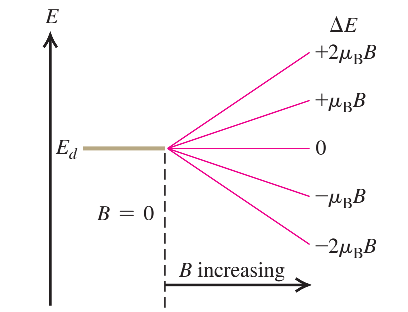


Figure **41.14** This figure shows how the splitting of the energy levels of a d state $(l = 2)$ depends on the magnitude B of an external magnetic field, assuming only an orbital magnetic moment.

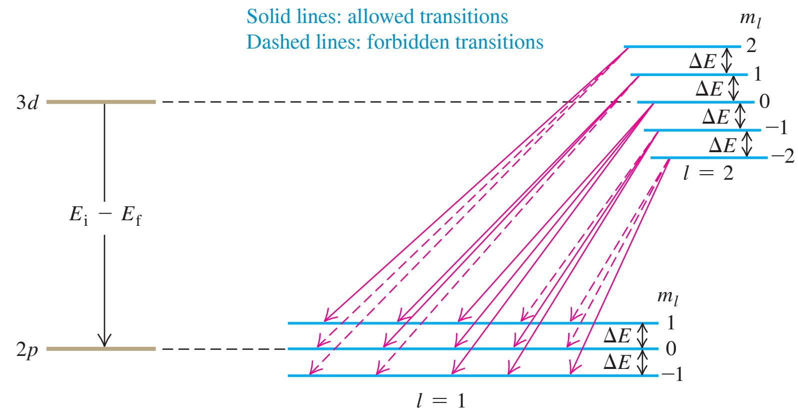


Figure **41.15** The cause of the normal Zeeman effect. The magnetic field splits the levels, but selection rules allow transitions with only three different energy changes, giving three different photon frequencies and wavelengths.

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?

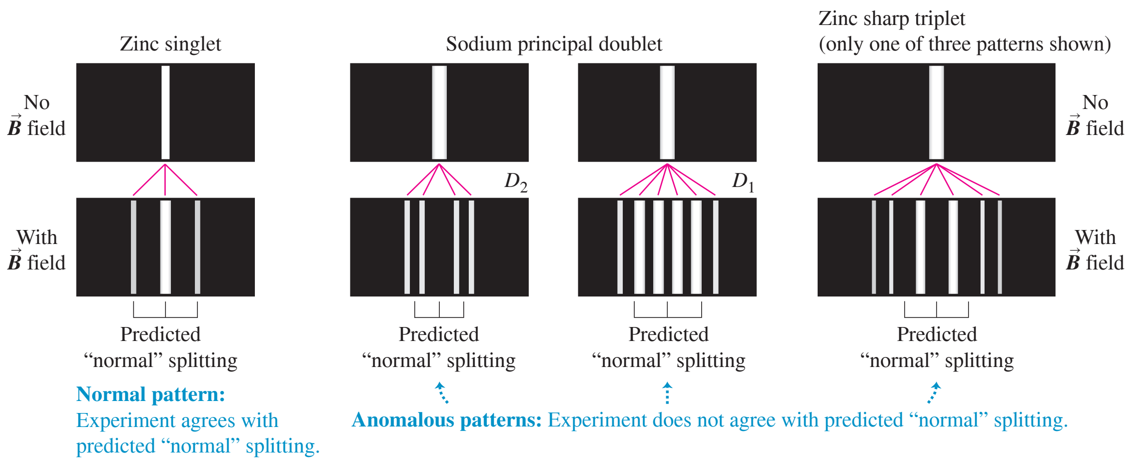
ANSWER

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 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

Figure **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 to distinguish it from the "normal" 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, setting

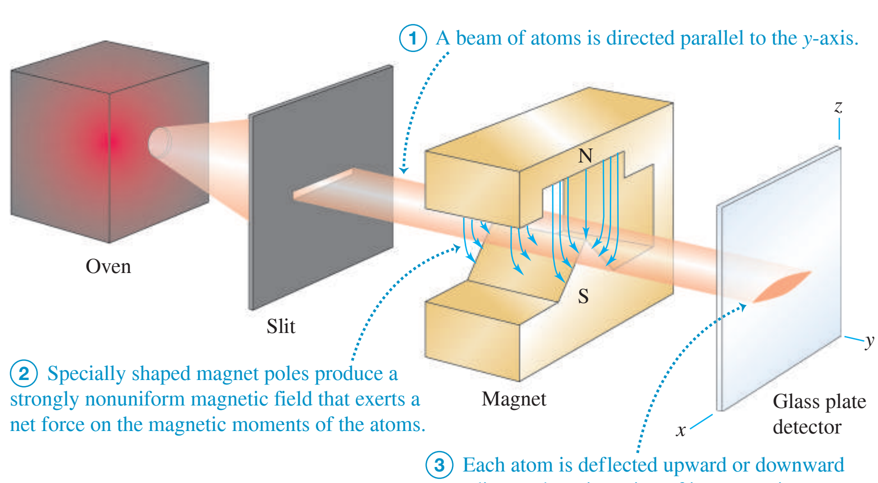


Figure **41.17** The Stern–Gerlach experiment.

$2j + 1$ 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 Goudsmit and George Uhlenbeck, proposed that the electron might have some additional motion. Using a semi-classical 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

z-component of spin angular momentum of electron

Spin magnetic quantum number $= \pm\frac{1}{2}$

$$S_z = m_s \hbar \qquad (41.36)$$

Planck's constant divided by 2π

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 2.00231930436182(52), 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

WITH VARIATION PROBLEMS

Calculate the interaction energy for an electron in an $l = 0$ state in a magnetic field with magnitude 2.00 T.

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

$$\mu_z = -(2.00232)\left(\frac{e}{2m}\right)\left(\pm\tfrac{1}{2}\hbar\right)$$

$$= \mp\tfrac{1}{2}(2.00232)\left(\frac{e\hbar}{2m}\right) = \mp(1.00116)\mu_B$$

$$= \mp(1.00116)(9.274\times10^{-24}\ \text{J/T})$$

$$= \mp9.285\times10^{-24}\ \text{J/T} = \mp5.795\times10^{-5}\ \text{eV/T}$$

Then from Eq. (41.32),

$$U = -\mu_z B = \pm(9.285\times10^{-24}\ \text{J/T})(2.00\ \text{T})$$

$$= \pm1.86\times10^{-23}\ \text{J} = \pm1.16\times10^{-4}\ \text{eV}$$

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)

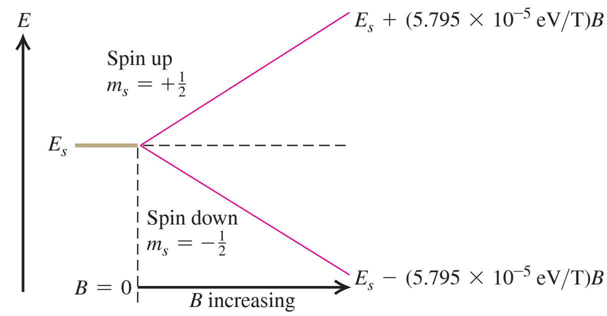
Continued

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.

KEYCONCEPT In addition to any orbital angular momentum, an electron has an intrinsic angular momentum $\vec{S}$ called spin. The component of $\vec{S}$ along a given axis can have only two possible values, $+\frac{1}{2}\hbar$ ("spin up") or $-\frac{1}{2}\hbar$ ("spin down"). An electron also has a spin magnetic moment $\vec{\mu}$ directed opposite to $\vec{S}$, so when it is placed in a magnetic field $\vec{B}$, the electron's energy depends on the component of $\vec{S}$ along the direction of $\vec{B}$.

Figure 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

WITH VARIATION PROBLEMS

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.

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

$$\begin{aligned}\Delta E &= \frac{hc}{\lambda_1} - \frac{hc}{\lambda_2} = hc\left(\frac{\lambda_2 - \lambda_1}{\lambda_2\lambda_1}\right)\\ &= (4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(2.998 \times 10^{8}\ \text{m/s})\\ &\quad\times \frac{(589.592 \times 10^{-9}\ \text{m}) - (588.995 \times 10^{-9}\ \text{m})}{(589.592 \times 10^{-9}\ \text{m})(588.995 \times 10^{-9}\ \text{m})}\\ &= 0.00213\ \text{eV} = 3.41 \times 10^{-22}\ \text{J}\end{aligned}$$

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.

KEYCONCEPT An atomic electron with an orbital angular momentum around the nucleus experiences an effective magnetic field. This field interacts with the spin magnetic moment of the electron, so the electron energy depends on the relative orientation of its spin angular momentum and orbital angular momentum. This is called spin-orbit coupling.

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 $j = \left|l \pm \frac{1}{2}\right|$. 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

$$\underbrace{E_{n,j}}_{\text{Energy levels of hydrogen, including fine structure}} = -\frac{13.60\text{ eV}}{\underbrace{n^2}_{\text{Principal quantum number }(n=1,2,3,\ldots)}}\left[1 + \frac{\overbrace{\alpha^2}^{\text{Fine-structure constant}}}{n^2}\left(\frac{n}{\underbrace{j}_{\text{Total angular momentum quantum number}} + \frac{1}{2}} - \frac{3}{4}\right)\right] \tag{41.41}$$

The *fine-structure constant* α that appears in Eq. (41.41) is a dimensionless number:

$$\alpha = \frac{1}{4\pi\epsilon_0}\frac{e^2}{\hbar c} = 7.2973525664(17) \times 10^{-3} \quad \text{(fine-structure constant)} \tag{41.42}$$

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

WITH VARIATION PROBLEMS

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?

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.

Continued

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:

$$\Delta\lambda = -\frac{121.6\text{ nm}}{10.20\text{ eV}}\left\{\left[10.20\text{ eV} + (3.40\text{ eV})\left(\frac{11\alpha^2}{16}\right)\right]\right.$$
$$\left. - \left[10.20\text{ eV} + (3.40\text{ eV})\left(\frac{15\alpha^2}{16}\right)\right]\right\}$$
$$= -\frac{121.6\text{ nm}}{10.20\text{ eV}}(3.40\text{ eV})\left(-\frac{4\alpha^2}{16}\right)$$
$$= \frac{121.6\text{ nm}}{10.20\text{ eV}}(3.40\text{ eV})\left(\frac{4}{16}\right)\left(\frac{1}{137.04}\right)^2$$
$$= 5.40 \times 10^{-4}\text{ nm}$$

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).

KEYCONCEPT When both magnetic and relativistic effects are included, the energy levels of the hydrogen atom [Eq. (41.41)] depend on both the principal quantum number $n = 1, 2, 3, \ldots$ and the quantum number j associated with the electron's total angular momentum (the vector sum of the orbital and spin angular momenta). If the electron has orbital quantum number l, the value of j is either $l + \frac{1}{2}$ or $\left|l - \frac{1}{2}\right|$. Transitions between states that involve emitting or absorbing a photon are allowed only if l changes by 1 in the transition.

(a) Galaxies in visible light (negative image; galaxies appear dark)

(b) Radio image of the same galaxies at wavelength 21 cm

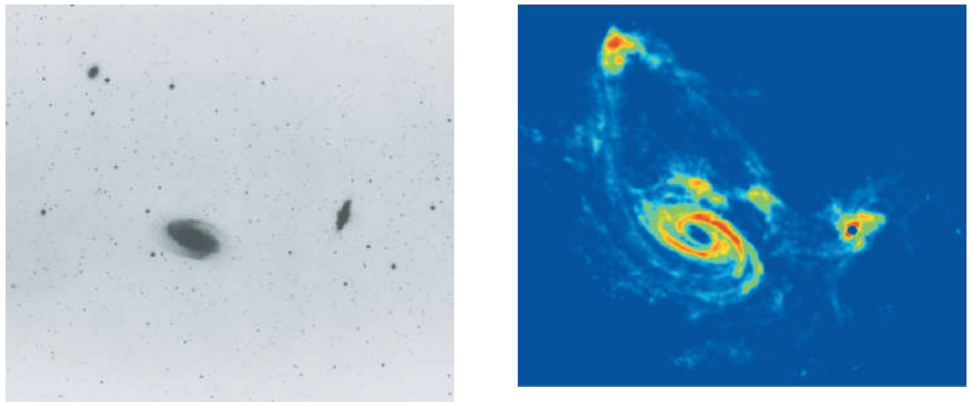


Figure **41.19** (a) In a visible-light image, these three distant galaxies appear to be unrelated. But in fact these galaxies are connected by immense streamers of hydrogen gas. This is revealed in (b), the false-color image made with a radio telescope tuned to the 21 cm wavelength emitted by hydrogen atoms.

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**).

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).

ANSWER

(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 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:

$$E_n = -\frac{1}{(4\pi\epsilon_0)^2}\frac{m_r Z^2 e^4}{2n^2\hbar^2} = -\frac{Z^2}{n^2}(13.6\ \text{eV}) \tag{41.43}$$

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

$$n \ge 1 \qquad 0 \le l \le n-1 \qquad |m_l| \le l \qquad m_s = \pm\tfrac{1}{2} \tag{41.44}$$

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 naturally occurring 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.)

Figure **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		

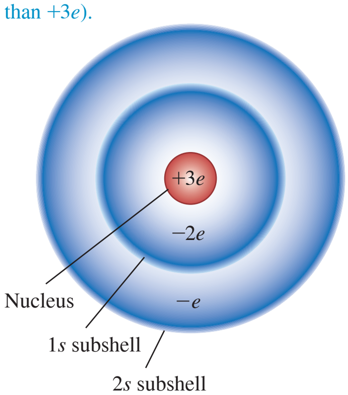
The Periodic Table

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 naturally occurring compounds.

Figure **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 $+3e$).

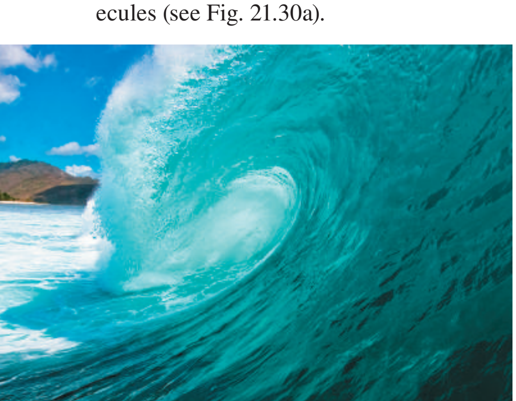


Lithium $(Z = 3)$ has three electrons. In its ground state, two are in 1s states and one is in a 2s state, so we denote the lithium ground state as $1s^2 2s$. On average, the 2s electron is considerably farther from the nucleus than are the 1s electrons (**Fig. 41.21**). According to Gauss's law, the *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 naturally occurring 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**).

Figure **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.

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).

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	$1s^2 2s$
Beryllium	Be	4	$1s^2 2s^2$
Boron	B	5	$1s^2 2s^2 2p$
Carbon	C	6	$1s^2 2s^2 2p^2$
Nitrogen	N	7	$1s^2 2s^2 2p^3$
Oxygen	O	8	$1s^2 2s^2 2p^4$
Fluorine	F	9	$1s^2 2s^2 2p^5$
Neon	Ne	10	$1s^2 2s^2 2p^6$
Sodium	Na	11	$1s^2 2s^2 2p^6 3s$
Magnesium	Mg	12	$1s^2 2s^2 2p^6 3s^2$
Aluminum	Al	13	$1s^2 2s^2 2p^6 3s^2 3p$
Silicon	Si	14	$1s^2 2s^2 2p^6 3s^2 3p^2$
Phosphorus	P	15	$1s^2 2s^2 2p^6 3s^2 3p^3$
Sulfur	S	16	$1s^2 2s^2 2p^6 3s^2 3p^4$
Chlorine	Cl	17	$1s^2 2s^2 2p^6 3s^2 3p^5$
Argon	Ar	18	$1s^2 2s^2 2p^6 3s^2 3p^6$
Potassium	K	19	$1s^2 2s^2 2p^6 3s^2 3p^6 4s$
Calcium	Ca	20	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$
Scandium	Sc	21	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d$
Titanium	Ti	22	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^2$
Vanadium	V	23	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^3$
Chromium	Cr	24	$1s^2 2s^2 2p^6 3s^2 3p^6 4s 3d^5$
Manganese	Mn	25	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^5$
Iron	Fe	26	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^6$
Cobalt	Co	27	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^7$
Nickel	Ni	28	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^8$
Copper	Cu	29	$1s^2 2s^2 2p^6 3s^2 3p^6 4s 3d^{10}$
Zinc	Zn	30	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10}$

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 $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2$ for strontium and $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$ 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 3*s* states are the lowest (most negative); one is the ground state for the 11th electron in sodium. The energy of the 3*d* states is quite close to the energy of the $n = 3$ state in hydrogen. The surprise is that the 4*s* state energy is 0.426 eV *below* the 3*d* state, even though the 4*s* 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 atom 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

$$\underbrace{E_n}_{\text{Energy levels of an electron with screening}} = -\frac{\overbrace{Z_{\text{eff}}}^{\text{Effective (screened) atomic number}}{}^2}{\underbrace{n^2}_{\text{Principal quantum number}}}(13.6\ \text{eV}) \qquad (41.45)$$

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 3*d* state is approximately the same as the $n = 3$ value of hydrogen, -1.51 eV. The distribution for the 3*d* 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 11*e* of the nucleus.) Thus in sodium a 3*d* 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 3*p* state (for which $l = n - 2$) has two peaks and that for the 3*s* state $(l = n - 3)$ has three peaks. For sodium the first small peak in the 3*p* distribution gives a 3*p* electron a higher probability (compared to the 3*d* state) of being *inside* the charge distributions for the electrons in the $n = 2$ states. That is, a 3*p* electron is less completely screened from the nucleus than is a 3*d* electron because it spends some of its time within the filled *K* and *L* shells. Thus for the 3*p* electrons, Z_{eff} is greater than unity. From Eq. (41.45) the 3*p* energy is lower (more negative) than the 3*d* energy of -1.521 eV. The actual value is -3.035 eV. A 3*s* electron spends even more time within the inner electron shells than a 3*p* 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

The measured energy of a 3s state of sodium is -5.138 eV. Calculate the value of Z_{eff}.

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

$$Z_{\text{eff}}{}^2 = -\frac{n^2 E_n}{13.6\text{ eV}} = -\frac{3^2(-5.138\text{ eV})}{13.6\text{ eV}} = 3.40$$

$$Z_{\text{eff}} = 1.84$$

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.

KEYCONCEPT If a many-electron atom has just one outer valence electron, the valence electron's energy levels are like those of a single-electron atom but with the atomic number Z of the nucleus replaced by an effective atomic number Z_{eff}. The value of Z_{eff} is less than Z because the inner electrons partially "screen" the electric field of the nucleus.

EXAMPLE 41.10 Energies for a valence electron

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.

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

$$E_4 = -\frac{Z_{\text{eff}}{}^2}{n^2}(13.6\text{ eV}) = -\frac{1}{4^2}(13.6\text{ eV}) = -0.85\text{ eV}$$

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, $Z_{\text{eff}} = 2$. A 3d state for Mg^+, for example, has an energy of about $-2^2(13.6\text{ eV})/3^2 = -6.0$ eV.

KEYCONCEPT For a many-electron atom with just one outer valence electron, if that electron is in a state of high orbital angular momentum it "sees" the effective atomic number Z_{eff} of the nucleus to be 1. In such a state the valence electron is almost always outside of the other $Z - 1$ electrons, which "screen" $(Z - 1)e$ of the nuclear charge $+Ze$.

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?

ANSWER

| **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.)

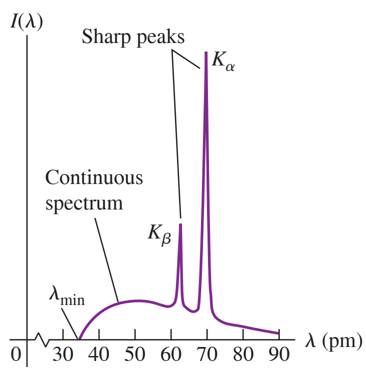
41.7 X-RAY SPECTRA

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:

Figure **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

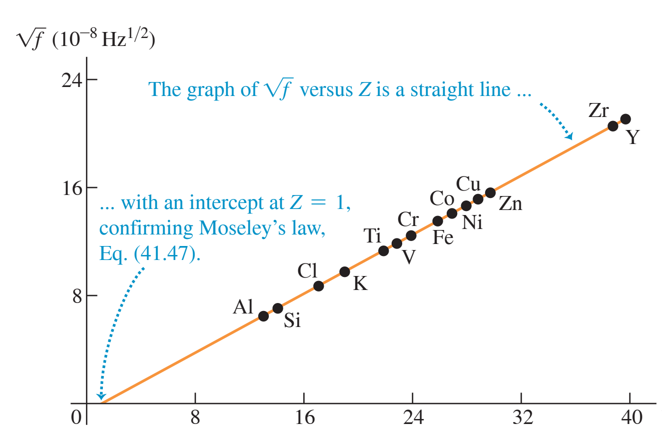
$$\lambda_{min} = \frac{hc}{eV_{AC}} \tag{41.46}$$

   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.

Figure **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: $$f = (2.48 \times 10^{15}\ \text{Hz})(Z-1)^2 \tag{41.47}$$

f: Frequency of K_α line in characteristic x-ray spectrum of an element

Z: 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'll 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,

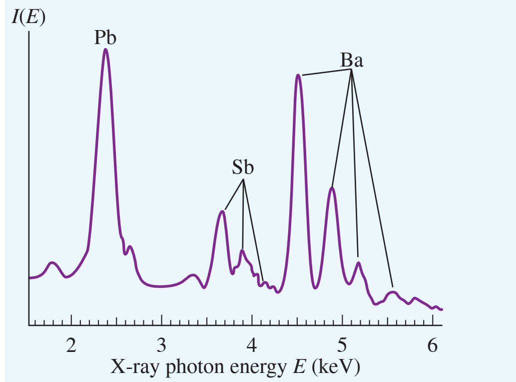
$$E_{K_\alpha} \approx (Z-1)^2(10.2\ \text{eV}) \tag{41.48}$$

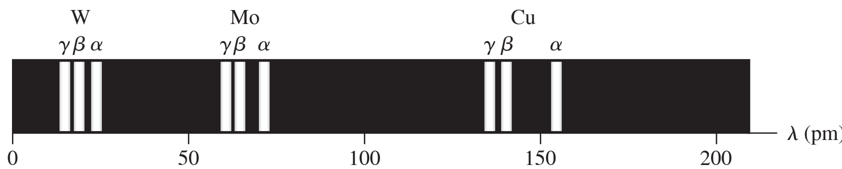
The frequency of the photon is its energy divided by Planck's constant:

$$f = \frac{E_{K_\alpha}}{h} \approx \frac{(Z-1)^2(10.2\ \text{eV})}{4.136 \times 10^{-15}\ \text{eV}\cdot\text{s}} = (2.47 \times 10^{15}\ \text{Hz})(Z-1)^2$$

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 X-ray spectra are an important tool in crime analysis. 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.





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.

Figure **41.25** Wavelengths of the K_α, K_β, and K_γ lines of tungsten (W), molybdenum (Mo), and copper (Cu).

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?

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.

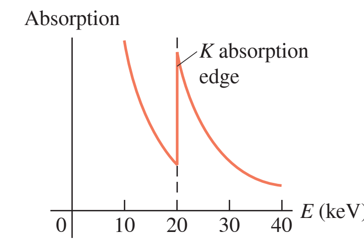
KEYCONCEPT If an electron is kicked out of the innermost, or K, shell of an atom of atomic number Z, then an electron from the next shell (L) drops down to the K shell to fill in the vacancy and emits an x-ray photon in the process. The frequency of this photon is proportional to $(Z - 1)^2$ (Moseley's law).

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.

Figure **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.

ANSWER

(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 $E_n = -(13.60 \text{ eV})/n^2 = -13.60$ eV, -3.40 eV, and -1.51 eV (see Fig. 39.24b).

41.8 QUANTUM ENTANGLEMENT

We've seen that quantum mechanics is very successful at correctly predicting the results of experiments. As we'll 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 helium 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

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_A(\vec{r}_1)\psi_B(\vec{r}_2) \pm \psi_B(\vec{r}_1)\psi_A(\vec{r}_2)] \tag{41.51}$$

(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:

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2}}[\psi_A(\vec{r}_1)\psi_B(\vec{r}_2) - \psi_B(\vec{r}_1)\psi_A(\vec{r}_2)] \tag{41.52}$$

(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 1200 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 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

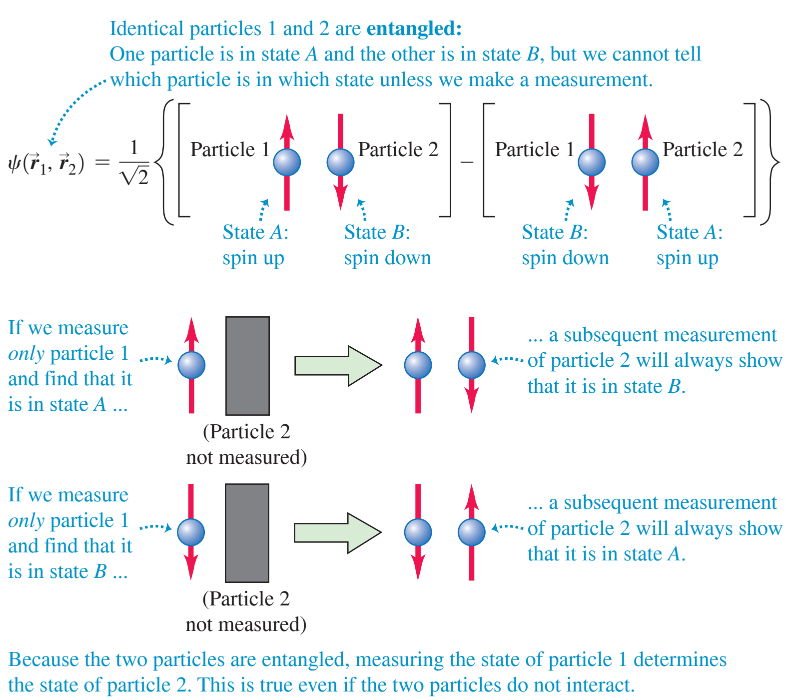


Figure **41.27** If two particles are in an entangled state, making a measurement of one particle determines the result of a subsequent measurement of the other particle.

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?

ANSWER

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).

CHAPTER 41 SUMMARY

Three-dimensional problems: The time-independent Schrödinger equation for three-dimensional problems is given by Eq. (41.5).

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi(x,y,z)}{\partial x^2}+\frac{\partial^2\psi(x,y,z)}{\partial y^2}+\frac{\partial^2\psi(x,y,z)}{\partial z^2}\right)+U(x,y,z)\psi(x,y,z) = E\psi(x,y,z) \qquad (41.5)$$

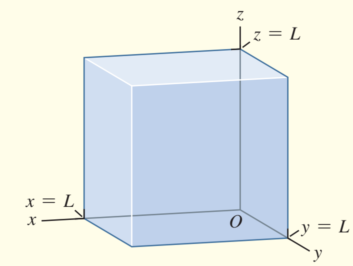
(three-dimensional time-independent Schrödinger equation)

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.)

$$E_{n_X,n_Y,n_Z} = \frac{(n_X{}^2+n_Y{}^2+n_Z{}^2)\pi^2\hbar^2}{2mL^2} \qquad (41.16)$$

$(n_X = 1, 2, 3, \ldots;$
$n_Y = 1, 2, 3, \ldots;$
$n_Z = 1, 2, 3, \ldots)$

(energy levels, particle in a three-dimensional cubical box)



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.)

$$E_n = -\frac{1}{(4\pi\epsilon_0)^2}\frac{m_r e^4}{2n^2\hbar^2} = -\frac{13.60\text{ eV}}{n^2} \qquad (41.21)$$

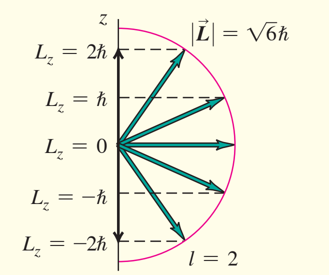
(energy levels of hydrogen)

$$L = \sqrt{l(l+1)}\,\hbar \qquad (l = 0, 1, 2, \ldots, n-1) \qquad (41.22)$$

$$L_z = m_l\hbar \qquad (m_l = 0, \pm1, \pm2, \ldots, \pm l) \qquad (41.23)$$

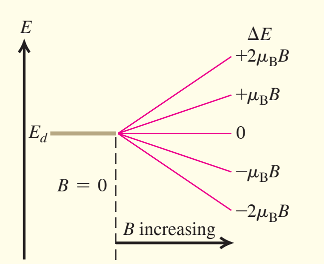
$$P(r)\,dr = |\psi|^2\,dV = |\psi|^2\,4\pi r^2\,dr \qquad (41.25)$$

$$a = \frac{\epsilon_0 h^2}{\pi m_r e^2} = \frac{4\pi\epsilon_0\hbar^2}{m_r e^2} = 5.29\times10^{-11}\text{ m} \qquad (41.26)$$



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.)

$$U = -\mu_z B = m_l\frac{e\hbar}{2m}B = m_l\mu_B B \qquad (m_l = 0, \pm1, \pm2, \ldots, \pm l) \qquad (41.35)$$



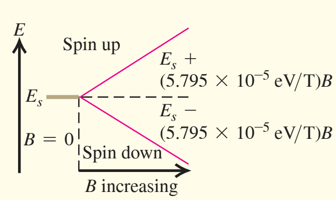
Electron spin: An electron has an intrinsic spin angular momentum of magnitude S, given by Eq. (41.37). The possible values of the z-component of the spin angular momentum are $S_z = m_s\hbar$, where $m_s = \pm\frac{1}{2}$. (See Examples 41.6 and 41.7.)

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.)

$$S = \sqrt{\tfrac{1}{2}\left(\tfrac{1}{2}+1\right)}\,\hbar = \sqrt{\tfrac{3}{4}}\,\hbar \qquad (41.37)$$

$$S_z = m_s\hbar \qquad \left(m_s = \pm\tfrac{1}{2}\right) \qquad (41.36)$$

$$E_{n,j} = -\frac{13.60\text{ eV}}{n^2}\left[1+\frac{\alpha^2}{n^2}\left(\frac{n}{j+\frac{1}{2}}-\frac{3}{4}\right)\right] \qquad (41.41)$$

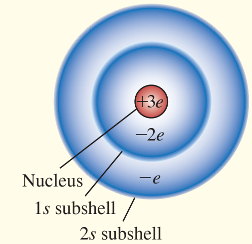


Continued

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.)

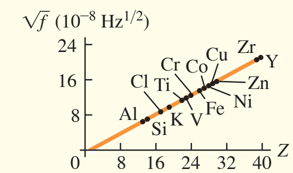
$$n \geq 1 \quad 0 \leq l \leq n-1$$
$$|m_l| \leq l \quad m_s = \pm\tfrac{1}{2} \tag{41.44}$$

$$E_n = -\frac{Z_{\text{eff}}^{\,2}}{n^2}(13.6\text{ eV}) \tag{41.45}$$

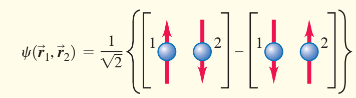


X-ray spectra: Moseley's law states that the frequency of a K_α x ray from a target with atomic number Z is given by Eq. (41.47). Characteristic x-ray spectra result from transitions to a hole in an inner energy level of an atom. (See Example 41.11.)

$$f = (2.48 \times 10^{15}\text{ Hz})(Z-1)^2 \tag{41.47}$$



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.



Chapter 41 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE √ARIATION PROBLEMS

Be sure to review EXAMPLE 41.1 (Section 41.2) before attempting these problems.

VP41.1.1 A particle in the three-dimensional box shown in Fig. 41.1 is in the state $n_X = 2$, $n_Y = 1$, $n_Z = 3$. Find (a) the planes (other than the walls of the box) on which the probability distribution function is zero and (b) the probability that the particle will be found somewhere in the region $0 \leq x \leq L/3$.

VP41.1.2 Half of the volume of the three-dimensional box shown in Fig. 41.1 is in the region $L/4 \leq x \leq 3L/4$. Find the probability that a particle in the box will be found in this region if the state of the particle is (a) $n_X = 1, n_Y = 1, n_Z = 1$; (b) $n_X = 2, n_Y = 1, n_Z = 2$; (c) $n_X = 3$, $n_Y = 2, n_Z = 3$; (d) $n_X = 4, n_Y = 1, n_Z = 1$.

VP41.1.3 The region $0 \leq x \leq L/4, 0 \leq y \leq L/4$ makes up $\frac{1}{16} = 0.0625$ of the volume of the three-dimensional box shown in Fig. 41.1. Find the probability that a particle in the box will be found in this region if the state of the particle is (a) $n_X = 1, n_Y = 1, n_Z = 1$; (b) $n_X = 2, n_Y = 1, n_Z = 2$; (c) $n_X = 3, n_Y = 2, n_Z = 3$; (d) $n_X = 4$, $n_Y = 1, n_Z = 1$.

VP41.1.4 Consider the cubical region given by $L/4 \leq x \leq 3L/4$, $L/4 \leq y \leq 3L/4$, $L/4 \leq z \leq 3L/4$ at the center of the three-dimensional box shown in Fig. 41.1. (a) What fraction of the total volume of the box is inside this cubical region? (b) If a particle in the box is in the state $n_X = 1, n_Y = 1, n_Z = 1$, find the probability that it will be found somewhere in this cubical region at the center of the box.

Be sure to review EXAMPLES 41.2, 41.3, and 41.4 (Section 41.3) before attempting these problems.

VP41.4.1 Consider the $n = 6$ states of the hydrogen atom. (a) How many distinct (l, m_l) states are there? (b) In terms of $\hbar$, what is the maximum magnitude of the orbital angular momentum L? (c) In terms of $\hbar$, what is the maximum value of the z-component of orbital angular momentum?

VP41.4.2 (a) List all the possible combinations of values of l and m_l for the $n = 3$ states of the hydrogen atom. (b) For which of these states is the angle between the orbital angular momentum vector and the negative z-axis a minimum, and what is that angle?

VP41.4.3 A photon is emitted when a hydrogen atom transitions from one energy level to a lower energy level. Find the energy of this photon, in eV, for each transition: (a) $n = 3, l = 2, m_l = -2$ to $n = 2, l = 1, m_l = -1$; (b) $n = 4, l = 2, m_l = 1$ to $n = 2, l = 1, m_l = 0$; (c) $n = 2, l = 1, m_l = 1$ to $n = 1, l = 0, m_l = 0$.

VP41.4.4 The wave function for an electron in a $1s$ state in a hydrogen atom is $\psi_{1s}(r) = 1/\sqrt{\pi a^3}\, e^{-r/a}$, where r is the distance from the nucleus. Find the probability that the electron will be found in the region (a) $0 \leq r \leq 2a$; (b) $a \leq r \leq 3a$; (c) $r \geq 4a$.

Be sure to review EXAMPLES 41.6, 41.7, and 41.8 (Section 41.5) before attempting these problems.

VP41.8.1 An isolated electron is placed in a magnetic field $\vec{B} = (3.14\text{ T})\hat{k}$. (a) Find the difference in energy between the $S_z = +\frac{1}{2}\hbar$ and $S_z = -\frac{1}{2}\hbar$ states of the electron. (b) Which state has the higher energy?

VP41.8.2 The outermost electron in a potassium atom is in an $l = 0$ state. If you place a potassium atom in a magnetic field of magnitude 2.36 T and illuminate it with monochromatic electromagnetic radiation, what must be the frequency and wavelength of the radiation to cause a transition between the spin-up and spin-down states of the outermost electron?

VP41.8.3 When the outermost electron in a potassium atom makes a transition from a $4p$ level to a $4s$ level, the wavelength of the photon it emits can be either 766.490 nm or 769.896 nm depending on the initial spin orientation of the electron. Find (a) the energy difference between the two $4p$ levels and (b) the effective magnetic field that the electron experiences in the $4p$ levels.

VP41.8.4 (a) Find the energy in terms of α of a state of the electron in a hydrogen atom with $n = 3$, $l = 1$, $j = \frac{3}{2}$ and the energy in terms of α of a state with $n = 3$, $l = 1$, $j = \frac{1}{2}$. (b) Find the difference between these energies, in eV. Which state has the higher energy? (c) Find the difference in wavelengths of the photons emitted in transitions from each of the states in part (a) to the state $n = 2$, $l = 0$, $j = \frac{1}{2}$. For which initial state is the wavelength longer?

BRIDGING PROBLEM A Many-Electron Atom in a Box

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

•, ••, •••: 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 an object 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 4*s* state. What does this tell you about the relative energies of the 3*d* and 4*s* 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 ($Z = 21$ 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 1*s* 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 photon with wavelength 8.00 nm is absorbed when an electron in a three-dimensional cubical box makes a transition from the ground state to the second excited state. What is the side length L of the box?
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?
41.7 • A particle is in a three-dimensional cubical box that has side length L. For the state $n_X = 3, n_Y = 2$, and $n_Z = 1$, for what planes (in addition to the walls of the box) is the probability distribution function zero?

Section 41.3 The Hydrogen Atom

41.8 • (a) A hydrogen atom is in a state with quantum number $n = 3$. In the quantum-mechanical description of the atom, what is the largest possible magnitude of the orbital angular momentum? What is the magnitude of the orbital angular momentum of the electron in the Bohr-model description of the atom? What is the percentage difference between these two values? (b) Answer the same questions as in part (a) for $n = 30$.
41.9 •• 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?
41.10 • 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?
41.11 • The orbital angular momentum of an electron has a magnitude of $4.716 \times 10^{-34}\ \text{kg} \cdot \text{m}^2/\text{s}$. What is the angular momentum quantum number l for this electron?
41.12 • 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?
41.13 •• 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?
41.14 •• 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?
41.15 • Consider the seventh excited level of the hydrogen atom. (a) What is the energy of this level? (b) What is the largest magnitude of the orbital angular momentum? (c) What is the largest angle between the orbital angular momentum and the *z*-axis?

41.16 • (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?

41.17 •• (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?

41.18 •• **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.

41.19 • 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 $e^{i\phi} = \cos\phi + i\sin\phi$.)

41.20 • (a) The radial probability distribution function for a hydrogen atom state has one peak, at $r = 0.476$ nm. What is the nl spectroscopic notation of this state? (b) What is r for the one peak of a $4f$ state?

Section 41.4 The Zeeman Effect

41.21 • 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?

41.22 • 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.

41.23 • 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?

41.24 •• **CP** A hydrogen atom undergoes a transition from a $2p$ state to the $1s$ ground state. In the absence of a magnetic field, the wavelength 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.

Section 41.5 Electron Spin

41.25 •• **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?

41.26 • **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.

41.27 • 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?

41.28 •• 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 $n \neq 1$?

Section 41.6 Many-Electron Atoms and the Exclusion Principle

41.29 • 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.

41.30 • For germanium (Ge, $Z = 32$), make a list of the number of electrons in each subshell ($1s, 2s, 2p, \ldots$). Use the allowed values of the quantum numbers along with the exclusion principle; do *not* refer to Table 41.3.

41.31 •• (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.

41.32 •• (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.

41.33 • The $5s$ electron in rubidium (Rb) sees an effective charge of $2.771e$. Calculate the ionization energy of this electron.

41.34 • 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?

41.35 • (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+}.

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 •• A particle is in a three-dimensional box. The y length of the box is twice the x length, and the z length is one-third of the y length. (a) What is the energy difference between the first excited level and the ground level? (b) Is the first excited level degenerate? (c) In terms of the x length, where is the probability distribution the greatest in the lowest-energy level?

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 •• When our sun exhausts its nuclear fuel, it will ultimately shrink due to gravity and become a white dwarf, with a radius of approximately 7000 km. (a) Using the mass of the sun, $M_{\text{sun}} = 2.0 \times 10^{30}$ kg, and the mass of a proton, $M_{\text{proton}} = 1.7 \times 10^{-27}$ kg, estimate the number of electrons in the sun. (b) From the radius given, estimate the average volume to be occupied by each electron in the eventual white dwarf. (c) The white dwarf will consist of mostly carbon. Since there are six electrons in each carbon atom, multiply the volume of an electron by 6 to obtain the volume of each carbon atom. (d) Model the atomic arrangement as a cubical lattice and use your earlier estimate to determine the distance L between adjacent carbon nuclei. (e) View each atom as an $L \times L \times L$ box filled with six electrons. Since no two electrons can be in the same quantum state, what quantum numbers (n_X, n_Y, n_Z, m_z) are associated with these six electrons, where m_z is the quantum number for the component of the electron spin along the z-axis? (f) Ignoring contributions from spin couplings and Coulomb interactions between the electrons, what would be the energy of the higher-energy electrons?

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 •• Interstellar hydrogen emits characteristic microwave radiation that can be understood as follows: (a) Model the magnetic moment of a proton using $\mu_p = (e/2m_p)\,S_z$, where $m_p = 1.7 \times 10^{-27}$ kg and $S_z = \pm\hbar/2$. Model the proton as a current loop with radius $R_p = 0.85$ fm. What current would produce the expected magnetic moment? (b) The magnetic field a distance x along the axis of a current loop is given by Eq. (28.15). Use this equation to estimate the magnetic field B at a distance $x = 0.40a_0$, where $a_0 \gg R_p$ is the Bohr radius. (c) The potential energy associated with the electron spin coupled to this magnetic field is $U_{\text{hyperfine}} = \pm\mu_z B$. Use Eq. (41.38) to show that $\mu_z = \mu_B$ where μ_B is the Bohr magneton. The sign depends on whether the electron spin is aligned or antialigned to that of the proton. Estimate the magnitude of this energy. (d) If an electron's spin were to flip, it would radiate a photon with energy $2|U_{\text{hyperfine}}|$. What would be the wavelength of such a photon? (e) Note that 21 cm radiation is observed in hydrogen clouds. Could electron spin flips explain this radiation?

41.53 •• CP The hydrogen spectrum includes four visible lines. Of these, the blue line corresponds to a transition from the $n = 5$ shell to the $n = 2$ shell and has a wavelength of 434 nm. If we look closer, this line is broadened by fine structure due to spin-orbit coupling and relativistic effects. (a) How many different sets of l and j quantum numbers are there for the $n = 5$ shell and for the $n = 2$ shell? (b) How many different energy levels are there for $n = 5$ and for $n = 2$? For each of these levels, what is their energy difference in eV from $-(13.6\text{ eV})/n^2$? (c) In a transition that emits a photon the quantum number l must change by ± 1. Which transition in the fine structure of the hydrogen blue line emits a photon of the shortest wavelength? For this photon what is the shift in wavelength due to the fine structure? (d) Which transition in the fine structure emits a photon of the longest wavelength? For this photon what is the shift in wavelength due to the fine structure? (e) By what total extent, in nm, is the wavelength of the blue line broadened around the 434 nm value?

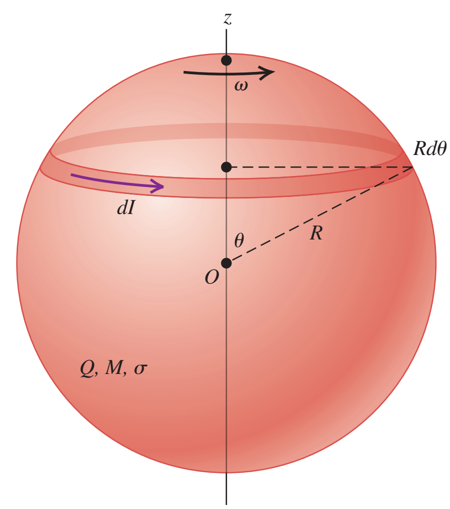
41.54 •• An atom in a $3d$ state emits a photon of wavelength 475.082 nm when it decays to a $2p$ state. (a) What is the energy (in electron volts) of the photon emitted in this transition? (b) Use the selection rules described in Section 41.4 to find the allowed transitions if the atom is now in an 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 CALC The magnetic moment $\vec{\mu}$ of a spinning object with charge Q and mass M is proportional to its angular momentum $\vec{L}$ according to $\vec{\mu} = g(Q/2M)\vec{L}$. The dimensionless coefficient g is known as the g-factor of the object. Consider a spherical shell with radius R and uniform charge density σ spinning with angular velocity ω around the z-axis. (a) What is the differential area da of the ring at latitude θ, with width $R\,d\theta$, as shown in **Fig. P41.57**? (b) The current dI carried by the ring is its charge $\sigma\,da$ divided by the period of rotation. Determine dI in terms of R, σ, ω, and θ. (c) Determine the magnetic moment $d\vec{\mu}$ of the ring by multiplying dI by the area enclosed by the ring. (d) Determine the magnetic moment of the spherical shell by integrating over the sphere. Express your result in terms of the total charge $Q = 4\pi R^2\sigma$. (e) Now consider a solid sphere of radius R with volume charge density $\rho = \rho(r)$, where r is the distance from the center, spinning with angular velocity ω about the z-axis. Determine its magnetic moment by integrating $\vec{\mu} = \int d\vec{\mu}$, where $d\vec{\mu}$ is now the magnetic moment associated with the shell at radius r with differential width dr. Express your answer in terms of an integral $\int_0^R r^4\rho(r)dr$. (f) The angular momentum of the sphere is $\vec{L} = I\vec{\omega}$, where $I = cMR^2$ is its moment of inertia, and c is a dimensionless factor determined by the mass distribution. Determine the g-factor in terms of the ρ integral and the value of c. (g) If the charge is uniformly distributed so that $\rho = Q/(\frac{4}{3}\pi R^3)$ and if the mass is uniformly distributed so that $c = \frac{2}{5}$, then what is the value of g?

Figure **P41.57**



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 $1s^22s$. The $1s^22p$ 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 $1s^22p$ 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 39.74(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 •• 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?

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)?

CHALLENGE PROBLEMS

41.67 ••• Spin-$\frac{3}{2}$ particles have four distinct spin states corresponding to $S_z = m_z\hbar$, where m_z may be $\pm\frac{1}{2}$ or $\pm\frac{3}{2}$. We can write the corresponding normalized wave functions as $\psi_{m_z}(\vec{r})$. (a) What is the magnitude of such a particle's spin $\vec{S}$? (b) Suppose we have three identical entangled spin-$\frac{3}{2}$ particles, each sharing an equal probability for exhibiting $m_z = \pm\frac{1}{2}$ or $m_z = -\frac{3}{2}$ but a zero probability for being in the state $m_z = +\frac{3}{2}$. The three-particle wave function is a sum of products such as $\psi_{-1/2}(\vec{r}_1)\psi_{+1/2}(\vec{r}_2)\psi_{-3/2}(\vec{r}_3)$, which we abbreviate as $\psi_{-1/2}\psi_{+1/2}\psi_{-3/2}$. (It is understood that the nth factor corresponds to the position $\vec{r}_n$.) Write the normalized three-particle wave function for the entangled state. Be mindful of the Pauli exclusion principal. (c) The three entangled particles are sent off in three different directions toward magnetic containment facilities A, B, and C located, respectively, at $\vec{r}_{1,2,3}$, where they are captured and retained in circular orbits. After an hour, the particle in facility A is sent through a Stern–Gerlach magnet, where it is determined that its state is $m_z = +\frac{1}{2}$. What is the three-particle wave function after this measurement? (d) At this time, what is the probability that a measurement in facility B will show that its particle is in each of the states $m_z = -\frac{1}{2}, +\frac{1}{2}, -\frac{3}{2}$, and $+\frac{3}{2}$? (e) Subsequently, the particle in facility C is measured to have $m_z = -\frac{3}{2}$. What is the wave function after this measurement? (f) At this time, what is the probability that a measurement in facility B will show that its particle is in each of the states $m_z = -\frac{1}{2}, +\frac{1}{2}, -\frac{3}{2}$, and $+\frac{3}{2}$?

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?

MCAT-STYLE 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 ?

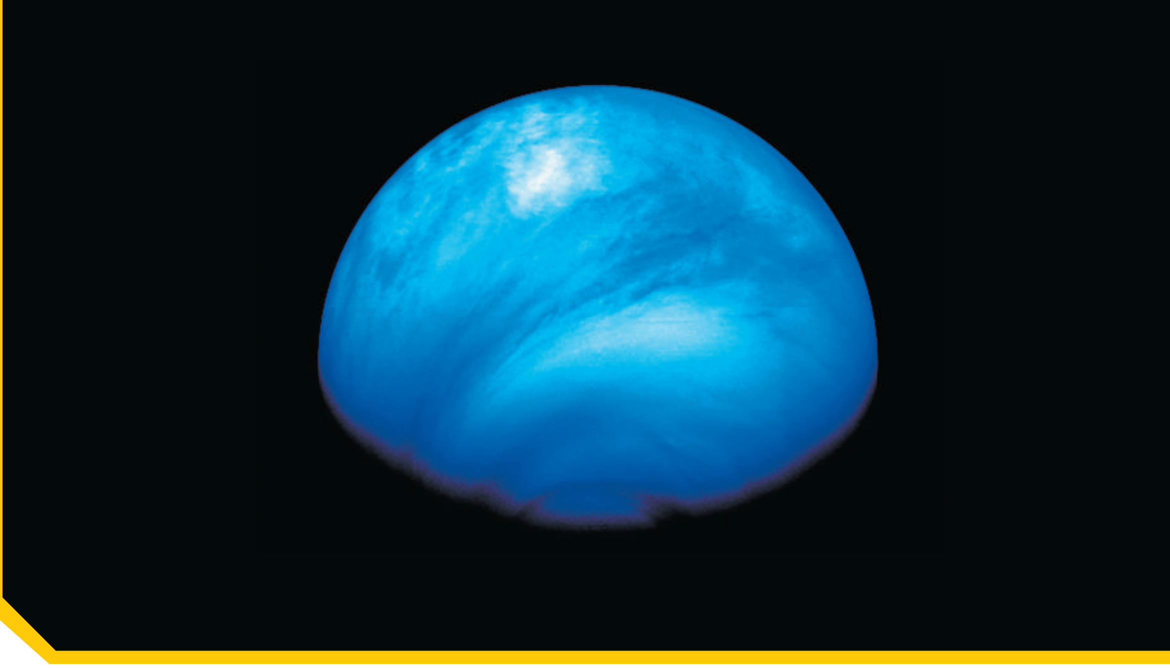
(iii) The Pauli exclusion principle is responsible. Helium is inert because its two electrons fill the K shell; lithium is very reactive because its third electron must go into the L shell and is loosely bound. See Section 41.6 for more details.

Key Example VARIATION Problems

VP41.1.1 (a) $x = L/2, z = L/3, z = 2L/3$ (b) 0.402
VP41.1.2 (a) 0.818 (b) 0.500 (c) 0.394 (d) 0.500
VP41.1.3 (a) 8.25×10^{-3} (b) 0.0227 (c) 0.0758 (d) 0.0227
VP41.1.4 (a) 0.125 (b) 0.548
VP41.4.1 (a) 36 (b) $\sqrt{30}\hbar = 5.477\hbar$ (c) $5\hbar$
VP41.4.2 (a) $(l, m_l) = (0, 0); (1, 1); (1, 0); (1, -1); (2, 2); (2, 1); (2, 0); (2, -1); (2, -2)$ (b) $(l, m_l) = (2, -2)$, $\theta_{\text{min}} = 35.3°$
VP41.4.3 (a) 1.89 eV (b) 2.55 eV (c) 10.2 eV
VP41.4.4 (a) 0.762 (b) 0.615 (c) 0.0138
VP41.8.1 (a) 3.64×10^{-4} eV (b) $S_z = +\frac{1}{2}\hbar$
VP41.8.2 $f = 6.61 \times 10^{10}$ Hz, $\lambda = 4.53$ mm
VP41.8.3 (a) 0.00716 eV (b) 61.8 T
VP41.8.4 (a) $E_{n=3,j=3/2} = -1.511\text{ eV}\left(1 + \dfrac{\alpha^2}{12}\right)$, $E_{n=3,j=1/2} = -1.511\text{ eV}\left(1 + \dfrac{\alpha^2}{4}\right)$ (b) $\Delta E = 1.34 \times 10^{-5}$ eV, the $j = \frac{3}{2}$ state (c) 4.66×10^{-3} nm, the $j = \frac{1}{2}$ state

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 OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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'll 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.

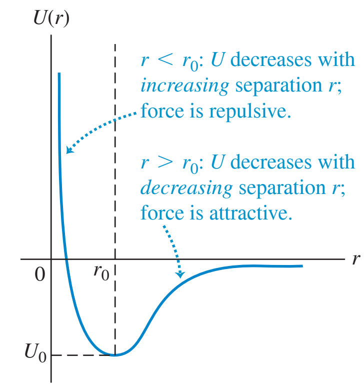
Let's look at the energy balance in this transaction. Removing the 3*s* electron from a neutral Na atom requires 5.138 eV of energy; this is called the *ionization energy* of Na. The neutral Cl atom can attract an extra electron into the vacancy in the 3*p* subshell, where it is incompletely screened 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 an Na^+ ion and a Cl^- ion that are well separated from each other (so they do not interact) requires a net investment of only 5.138 eV − 3.613 eV = 1.525 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 eV − 1.525 eV = 4.2 eV. Thus, if we ignore the kinetic energy of the ions, 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.

Figure 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.



EXAMPLE 42.1 Electric potential energy of the NaCl molecule

Find the electric potential energy of an Na^+ ion and a Cl^- ion separated by 0.24 nm. Consider the ions as point charges.

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 $r = 0.24\ \text{nm} = 0.24 \times 10^{-9}\ \text{m}$. 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.

KEYCONCEPT In an ionic bond, the electric potential energy of the two ions (one positive and one negative) is negative and roughly the same as if we consider the two ions as point charges.

Covalent Bonds

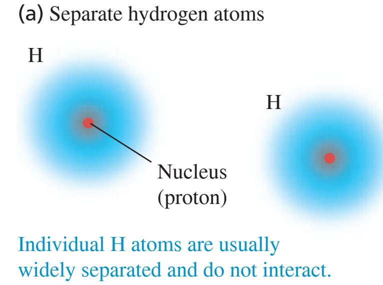
Unlike the transaction that occurs in an ionic bond, in a **covalent bond** there is no net transfer of electrons from one atom to another. The simplest covalent bond is found in the hydrogen molecule, a structure containing two protons and two electrons. As the separate atoms (**Fig. 42.2a**) come together, 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.

Figure 42.2 Covalent bond in a hydrogen molecule.

(a) Separate hydrogen atoms



(b) H_2 molecule

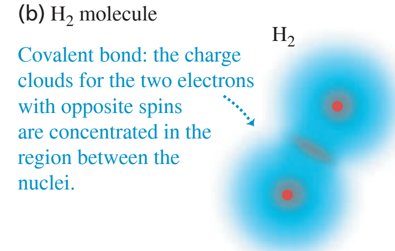
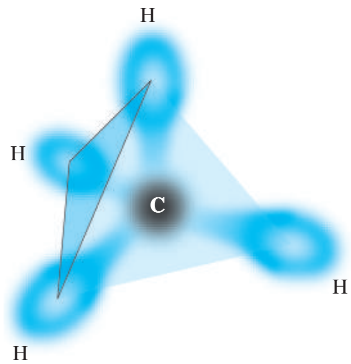


Figure 42.3 Schematic diagram of the methane (CH_4) molecule. The carbon atom is at the center of a regular tetrahedron and forms four covalent bonds with the hydrogen atoms at the corners. Each covalent bond includes two electrons with opposite spins, forming a charge cloud that is concentrated between the carbon atom and a hydrogen atom.



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.

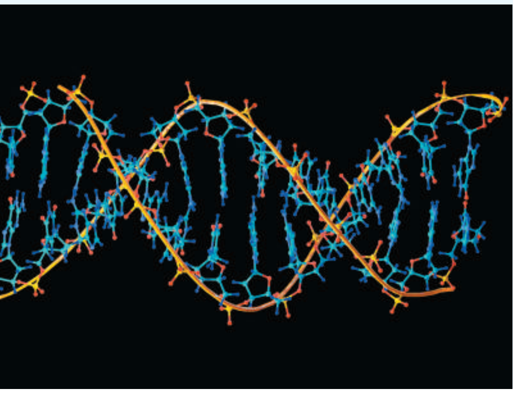
Ionic and covalent bonds represent two extremes in molecular bonding, but there is no sharp division between the two types. Often there is a *partial* transfer of one or more electrons from one atom to another. As a result, many molecules that have dissimilar atoms have electric dipole moments—that is, 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).

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.

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.



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 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.

ANSWER

(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'll 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 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?

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 quantum number ($l = 0, 1, 2, \ldots$)

Rotational energy levels of a diatomic molecule

$$E_l = l(l+1)\frac{\hbar^2}{2I} \tag{42.3}$$

$\hbar$: Planck's constant divided by 2π

I: Moment of inertia for axis through molecule's cm

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:

Figure **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 $r_1 + r_2 = r_0$.

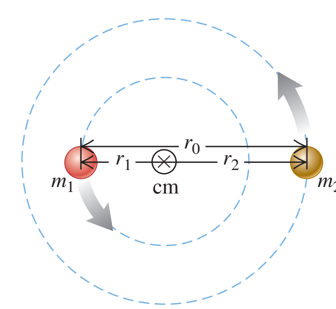
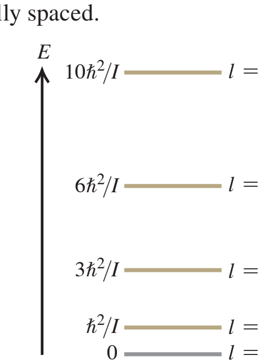


Figure **42.5** The ground level and first four excited rotational energy levels for a diatomic molecule. The levels are not equally spaced.



Reduced mass of a diatomic molecule

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{42.4}$$

Mass of atom 1, Mass of atom 2

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

$$I = m_r r_0^2 \tag{42.6}$$

Reduced mass; Distance between centers of molecule's two atoms

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, $\Delta l = \pm 1$.

EXAMPLE 42.2 Rotational spectrum of carbon monoxide

WITH VARIATION PROBLEMS

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.

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:

$$m_r = \frac{m_1 m_2}{m_1 + m_2}$$
$$= \frac{(1.993 \times 10^{-26}\text{ kg})(2.656 \times 10^{-26}\text{ kg})}{(1.993 \times 10^{-26}\text{ kg}) + (2.656 \times 10^{-26}\text{ kg})} = 1.139 \times 10^{-26}\text{ kg}$$

$$I = m_r r_0^2$$
$$= (1.139 \times 10^{-26}\text{ kg})(0.1128 \times 10^{-9}\text{ m})^2 = 1.449 \times 10^{-46}\text{ kg}\cdot\text{m}^2$$

The rotational levels are given by Eq. (42.3):

$$E_l = l(l+1)\frac{\hbar^2}{2I} = l(l+1)\frac{(1.0546 \times 10^{-34}\text{ J}\cdot\text{s})^2}{2(1.449 \times 10^{-46}\text{ kg}\cdot\text{m}^2)}$$
$$= l(l+1)(3.838 \times 10^{-23}\text{ J}) = l(l+1)0.2395\text{ meV}$$

($1 \text{ meV} = 10^{-3} \text{ eV}$.) 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

$$E = E_2 - E_1 = 0.958\text{ meV}$$

$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\text{ eV}\cdot\text{s})(3.00 \times 10^8\text{ m/s})}{0.958 \times 10^{-3}\text{ eV}}$$
$$= 1.29 \times 10^{-3}\text{ m} = 1.29\text{ mm}$$

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.

KEYCONCEPT Because angular momentum is quantized (quantum number l), the rotational kinetic energy of a diatomic molecule is quantized as well. This energy is proportional to $l(l+1)$ and inversely proportional to the moment of inertia of the molecule for an axis through the molecule's center of mass. Transitions between rotational states that involve photon emission or absorption are allowed only if l changes by 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 $1\ \mathrm{K} = 1$ 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.

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.

Figure **42.6** A diatomic molecule modeled as two point masses m_1 and m_2 connected by a spring with force constant k'.

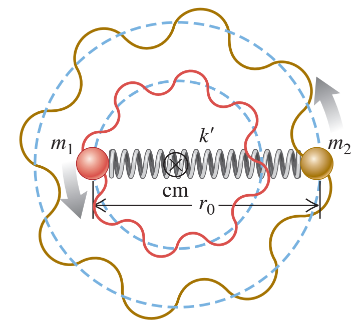


Figure **42.7** The ground level and first three excited vibrational levels for a diatomic molecule, assuming small displacements from equilibrium so we can treat the oscillations as simple harmonic. (Compare Fig. 40.25.)

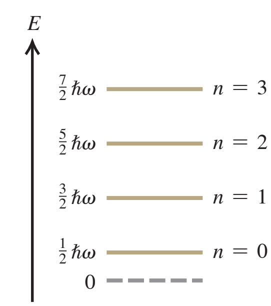


Figure **42.8** Energy-level diagram for vibrational and rotational energy levels of a diatomic molecule. For each vibrational level (n) there is a series of more closely spaced rotational levels (l). Several transitions corresponding to a single band in a band spectrum are shown. These transitions obey the selection rule $\Delta l = \pm 1$.

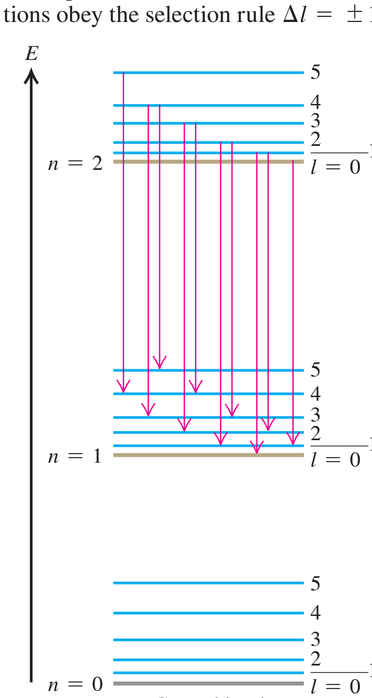


Figure 42.9 A typical molecular band spectrum.

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.

EXAMPLE 42.3 Vibration-rotation spectrum of carbon monoxide

WITH VARIATION PROBLEMS

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.

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.

KEYCONCEPT In addition to rotational motion around its center of mass (quantum number l), a diatomic molecule can have vibrational motion of its atoms. The vibrational energy levels (quantum number n) are evenly spaced in energy. Transitions between vibrational states that involve photon emission or absorption are allowed only if n and l both change by 1.

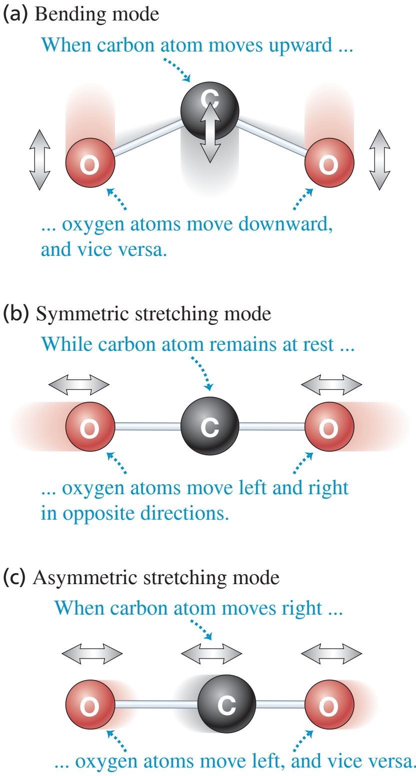
Complex Molecules

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.

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.

Figure **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.



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?

ANSWER

❙ (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 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.)

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.

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.

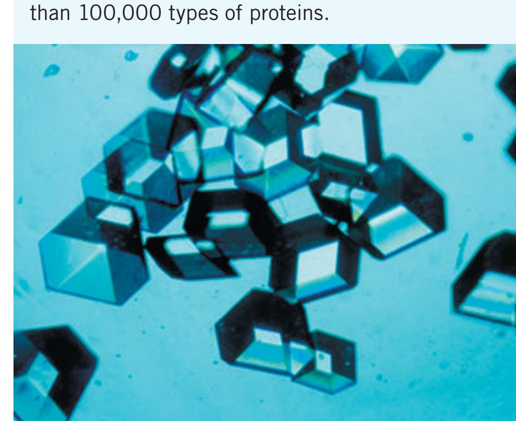
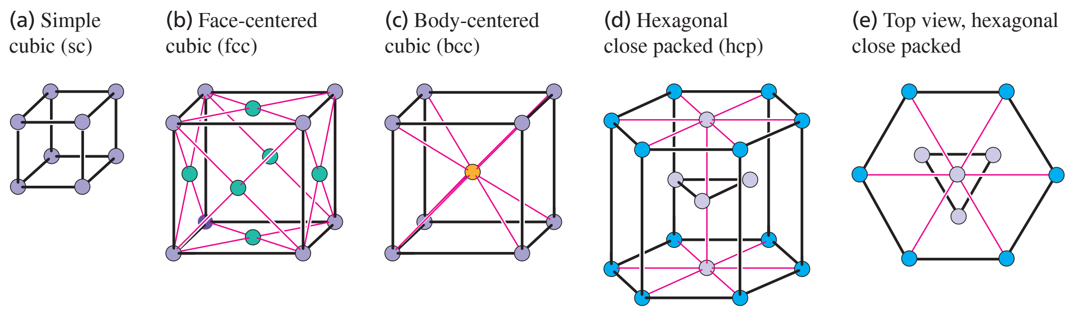


Figure **42.11** Portions of some common types of crystal lattices.



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* (hcp) 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.

Figure **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.

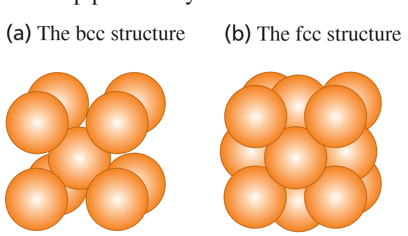
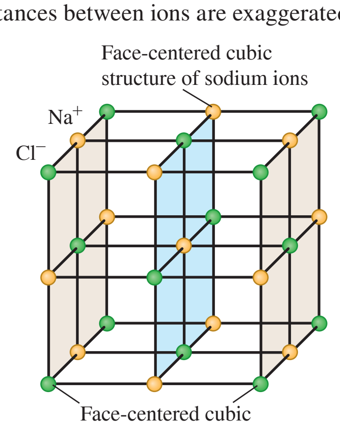


Figure **42.13** Representation of part of the sodium chloride crystal structure. The distances between ions are exaggerated.



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.

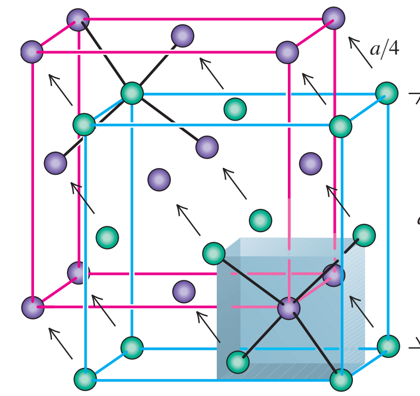
The bcc and fcc structures are two common simple crystal structures. The alkali metals have a bcc structure—that is, a bcc lattice with a basis of one atom at each lattice point. Each atom in a bcc structure has eight nearest neighbors (**Fig. 42.12a**). The elements Al, Ca, Cu, Ag, and Au have an fcc 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 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.

Figure **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 $a/4$.



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 discussed below. 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

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.

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

$$\begin{aligned}\Sigma U &= -\frac{e^2}{4\pi\epsilon_0}\frac{1}{a} + \frac{e^2}{4\pi\epsilon_0}\frac{1}{2a} - \frac{e^2}{4\pi\epsilon_0}\frac{1}{3a} + \cdots \\ &= -\frac{e^2}{4\pi\epsilon_0 a}\left(1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots\right)\end{aligned}$$

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

$$\Sigma U = -\frac{e^2}{4\pi\epsilon_0 a}\ln 2$$

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.

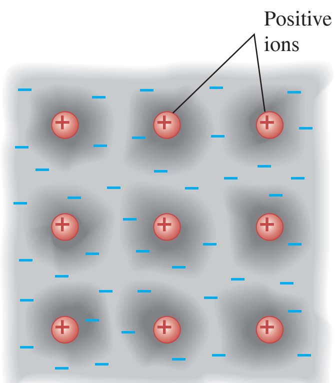
KEYCONCEPT An ionic crystal is stable because the net electric potential energy of its ions is less than if the crystal were taken apart and the ions were moved far from each other.

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

Figure **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.



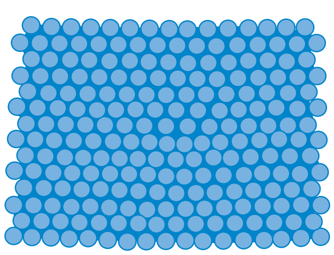
electrons in each atom become detached from the parent atom (leaving a positive ion). The detached electrons are free to move through the crystal and are not localized near the ion from which they originated. So we can picture a metallic crystal as an array of positive ions immersed in a sea of freed electrons whose attraction for the positive ions holds the crystal together (**Fig. 42.15**).

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.

Figure **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.

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'll discuss in Section 42.6, is the *substitutional impurity,* a foreign atom replacing a regular atom (for example, arsenic in a silicon crystal).

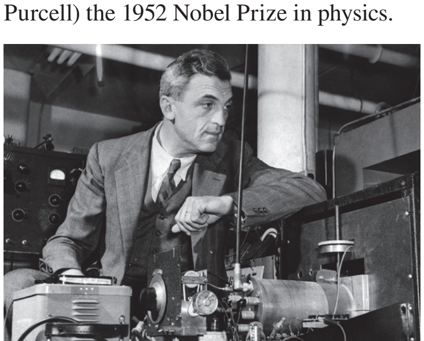
There are several basic types of extended defects called *dislocations*. One type is the *edge dislocation,* shown schematically in **Fig. 42.16**, in which one plane of atoms slips relative to another. The mechanical properties of metallic crystals are influenced strongly by the presence of dislocations. 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.

ANSWER

(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 an Na^+ ion is on the opposite corner of a cube of side a. The distance between these two ions is $\sqrt{a^2 + a^2 + a^2} = \sqrt{3a^2} = a\sqrt{3}$.

Figure **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.

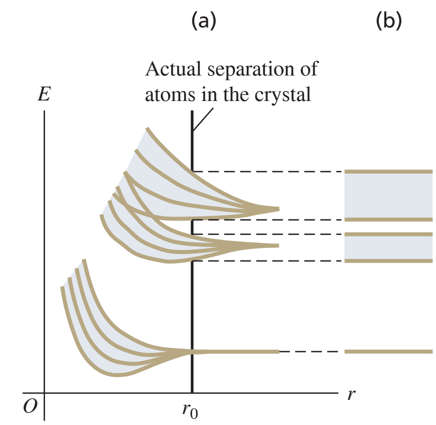


42.4 ENERGY BANDS

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.

Figure 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.



Insulators, Semiconductors, and Conductors

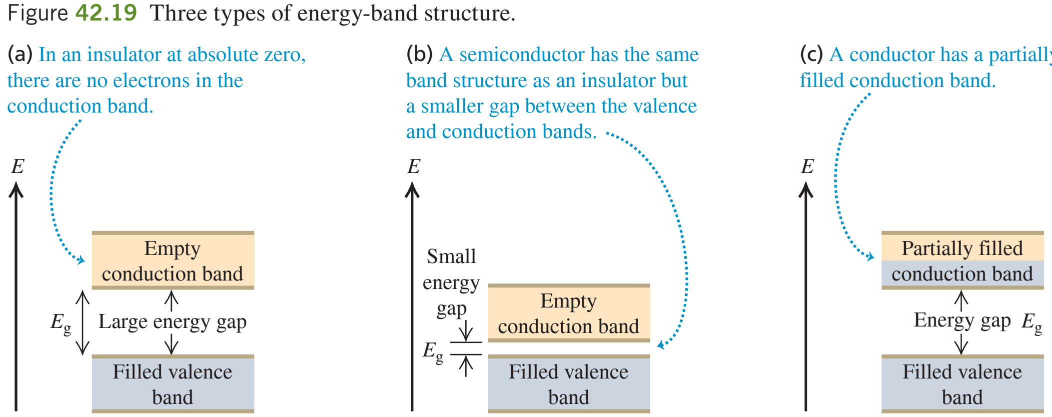
The nature of the energy bands determines whether the material is an electrical insulator, a semiconductor, or a conductor. In particular, what matters are the extent to which the states in each band are occupied and the spacing, called the *band gap* or *energy gap,* between adjacent bands.

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 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 the order of 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

Figure 42.19 Three types of energy-band structure.



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'll 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 $3p$ states) are about 2.1 eV above the two $3s$ ground states. In solid sodium, however, the atoms are so close together that the $3s$ and $3p$ *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 $3s$, six $3p$), 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.

EXAMPLE 42.5 Photoconductivity in germanium

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?

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 $E_{\text{min}} = 0.67$ eV. From $E = hc/\lambda$, the *maximum* wavelength λ_{max} (our target variable) corresponds to this *minimum* photon energy.

EXECUTE The wavelength of a photon with energy $E_{\text{min}} = 0.67$ eV is

$$\lambda_{\text{max}} = \frac{hc}{E_{\text{min}}} = \frac{(4.136 \times 10^{-15}\ \text{eV} \cdot \text{s})(3.00 \times 10^{8}\ \text{m/s})}{0.67\ \text{eV}}$$
$$= 1.9 \times 10^{-6}\ \text{m} = 1.9\ \mu\text{m} = 1900\ \text{nm}$$

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.

KEYCONCEPT The electron energy levels in a semiconductor form a valence band that is essentially filled, plus a conduction band that is empty and separated from the valence band by a small energy gap. The conductivity of a semiconductor can be increased by using photons to excite electrons from the valence band into the conduction band. To do this, the photon energy must be at least as great as the band gap; lower-energy photons will not be absorbed.

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.

ANSWER

(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 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.

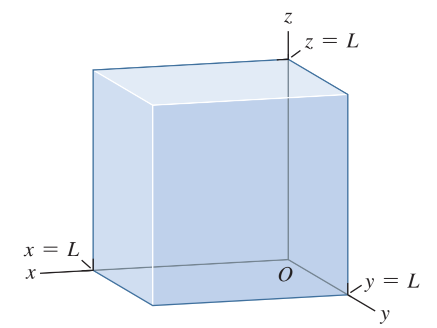
The **free-electron model** assumes that these electrons don't interact at all with the ions or with each other, but that there are infinite potential-energy barriers at the surfaces. The idea is that a typical electron moves so rapidly within the metal that it "sees" the effect of the ions and other 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

$$E_{n_X,n_Y,n_Z} = \frac{(n_X{}^2 + n_Y{}^2 + n_Z{}^2)\pi^2\hbar^2}{2mL^2} \qquad \begin{pmatrix} n_X = 1, 2, 3, \ldots; \\ n_Y = 1, 2, 3, \ldots; \\ n_Z = 1, 2, 3, \ldots \end{pmatrix} \tag{42.10}$$

Each state is labeled by the three positive-integer quantum numbers (n_X, n_Y, n_Z).

Figure **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).

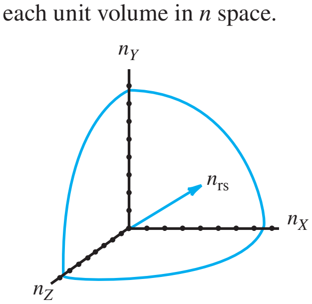


Density of States

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

$$n = \frac{\pi n_{rs}{}^3}{3} \tag{42.11}$$

Figure **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.



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):

$$dn = \frac{(2m)^{3/2}VE^{1/2}}{2\pi^2\hbar^3}\,dE \tag{42.14}$$

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

$$g(E) = \frac{(2m)^{3/2}V}{2\pi^2\hbar^3}E^{1/2} \tag{42.15}$$

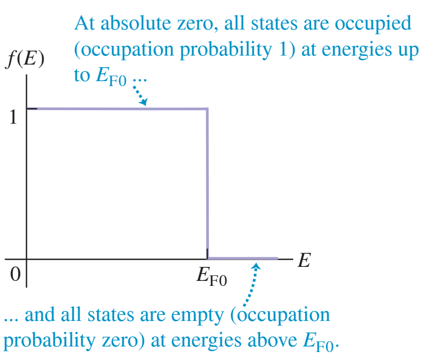
Volume

Electron energy

Planck's constant divided by 2π

Fermi–Dirac Distribution

Figure 42.22 The probability distribution for occupation of free-electron energy states at absolute zero.



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 $f(E) = 0$.

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

$$f(E) = \frac{1}{e^{(E - E_F)/kT} + 1} \tag{42.16}$$

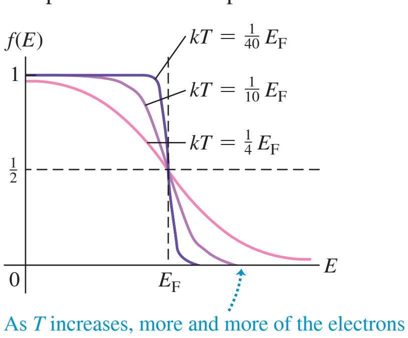
Absolute temperature

Energy of state

Fermi energy

Boltzmann constant

Figure 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.



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.

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

WITH VARIATION PROBLEMS

For free electrons in a solid at temperature T, at what energy is the probability that a particular state is occupied equal to (a) 0.01 and (b) 0.99?

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

$$E = E_{\mathrm{F}} + kT\ln\left(\frac{1}{f(E)} - 1\right)$$

(a) When $f(E) = 0.01$,

$$E = E_{\mathrm{F}} + kT\ln\left(\frac{1}{0.01} - 1\right) = E_{\mathrm{F}} + 4.6kT$$

The probability that a state $4.6kT$ above the Fermi level is occupied is only 0.01, or 1%.

(b) When $f(E) = 0.99$,

$$E = E_{\mathrm{F}} + kT\ln\left(\frac{1}{0.99} - 1\right) = E_{\mathrm{F}} - 4.6kT$$

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.

KEYCONCEPT In the free-electron model of a metal, the Fermi–Dirac distribution gives the probability that a state of a given energy E is occupied at a temperature T. The probability is greater than $\frac{1}{2}$ for energies less than the Fermi energy E_{F} and less than $\frac{1}{2}$ for energies greater than E_{F}.

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

$$dN = g(E)f(E)\,dE = \frac{(2m)^{3/2}VE^{1/2}}{2\pi^2\hbar^3}\frac{1}{e^{(E-E_{\mathrm{F}})/kT} + 1}\,dE \qquad (42.17)$$

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}:

$$N = \frac{(2m)^{3/2}VE_{\mathrm{F0}}{}^{3/2}}{3\pi^2\hbar^3} \qquad (42.18)$$

Solving Eq. (42.18) for E_{F0}, we get

$$E_{\mathrm{F0}} = \frac{3^{2/3}\pi^{4/3}\hbar^2}{2m}\left(\frac{N}{V}\right)^{2/3} \qquad (42.19)$$

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

$$E_{\mathrm{F0}} = \frac{3^{2/3}\pi^{4/3}\hbar^2 n^{2/3}}{2m} \qquad (42.20)$$

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.

EXAMPLE 42.7 The Fermi energy in copper

WITH VARIATION PROBLEMS

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.

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 $0.8\text{–}2.2 \times 10^6\ \text{m/s}$, respectively. Note that the calculated Fermi speed is far less than the speed of light $c = 3.00 \times 10^8\ \text{m/s}$, 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 $T = 20°\text{C} = 293\ \text{K}$, the quantity kT equals $(1.381 \times 10^{-23}\ \text{J/K})(293\ \text{K}) = 4.04 \times 10^{-21}\ \text{J} = 0.0254\ \text{eV}$.) 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 $E = 7.03\ \text{eV}$ 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.

KEYCONCEPT The Fermi energy in the free-electron model of a metal depends on the electron concentration (number of electrons per unit volume). In a typical metal at room temperature T, the Fermi energy is much greater than kT.

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

$$E_{av} = \frac{3}{2E_{F0}^{3/2}}\int_0^{E_{F0}} E^{3/2}\,dE = \tfrac{3}{5}E_{F0} \tag{42.21}$$

At absolute zero the average free-electron energy equals $\frac{3}{5}$ of the Fermi energy.

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.

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 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 $1.126 \times 10^{-18}\ \text{J} = 7.03\ \text{eV}$. According to Eq. (42.21), the average energy is $\frac{3}{5}$ of this, or $6.76 \times 10^{-19}\ \text{J} = 4.22\ \text{eV}$.

(b) In Example 42.7 we found that $kT = 4.04 \times 10^{-21}$ J = 0.0254 eV at room temperature $T = 20°\text{C} = 293$ 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

$$v = \sqrt{\frac{2E_{\text{av}}}{m}} = \sqrt{\frac{2(6.07 \times 10^{-21}\ \text{J})}{9.11 \times 10^{-31}\ \text{kg}}}$$
$$= 1.15 \times 10^5\ \text{m/s}$$

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).

KEYCONCEPT The exclusion principle, not the effects of temperature, is the dominant reason for the distribution of electron energies in a metal at room temperature.

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?

ANSWER

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 SEMICONDUCTORS

A **semiconductor** has an electrical resistivity that is intermediate between those of good conductors and those 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.

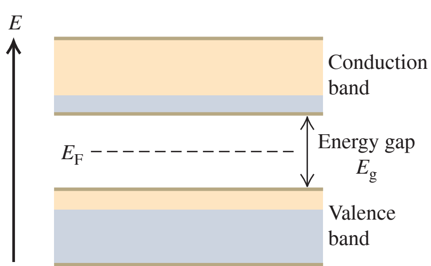
EXAMPLE 42.9 Jumping a band gap

WITH VARIATION PROBLEMS

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.

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.

Figure **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.



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.

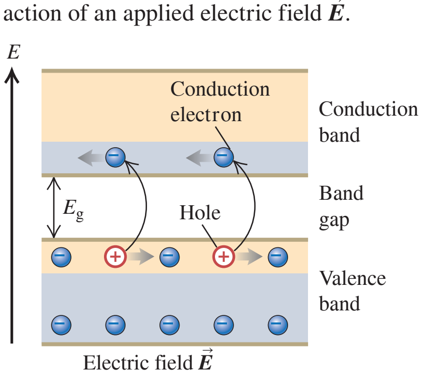
(c) For $E_g = 5.0$ eV, the exponents are 96.7 and 93.6; the values of $f(E)$ are 1.0×10^{-42} and 2.3×10^{-41}. The (extremely low) probability increases by a factor of 23 for a 10 K temperature rise.

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.

KEYCONCEPT The properties of the Fermi–Dirac distribution explain why the number of electrons in the conduction band of a semiconductor is sensitive to both the temperature and the size of the band gap.

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.

Figure **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 $\vec{E}$.



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 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 $(\mathrm{Ge}, Z = 32)$ a small amount of arsenic $(\mathrm{As}, Z = 33)$, the next element after germanium in the periodic table. We then allow the mixture to cool and crystallize. This deliberate addition of impurity elements is called *doping*. Arsenic is in Group V; it has *five* valence electrons compared to the four valence electrons of germanium. When one of these five valence electrons is removed from an arsenic atom, 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.

The fifth valence electron in arsenic 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.

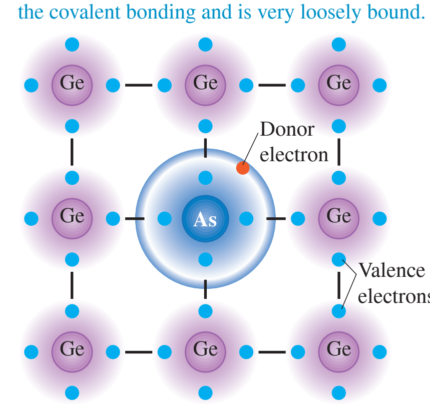
Example 42.9 shows that at ordinary temperatures and with a band gap of 1.0 eV, only a very small fraction (of the order of 10^{-9}) of the states at the bottom of the conduction band in a pure semiconductor contain electrons to participate in intrinsic conductivity. Thus we expect the conductivity of 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 $(\mathrm{Ga}, Z = 31)$; 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**, next page). The resulting atom has the same electron configuration as germanium 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

Figure **42.26** An n-type semiconductor: germanium (Ge) with an arsenic (As) impurity.

(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.

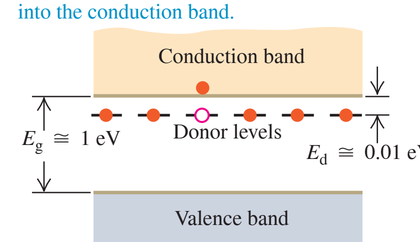
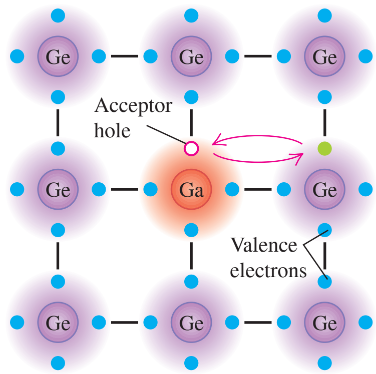


Figure **42.27** A p-type semiconductor: germanium (Ge) with a gallium (Ga) impurity.

(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.

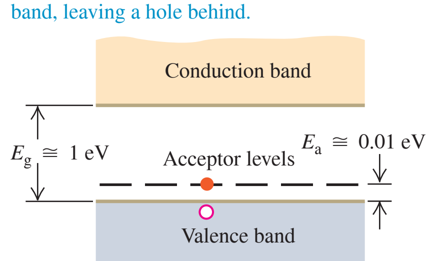
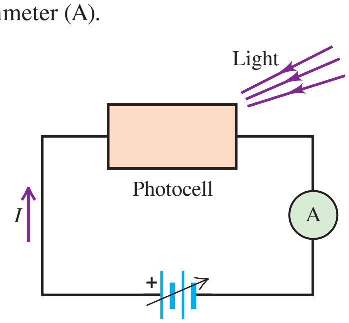


Figure **42.28** A semiconductor photocell in a circuit. The more intense the light falling on the photocell, the greater the conductivity of the photocell and the greater the current measured by the ammeter (A).



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" It's a common misconception that a p-type semiconductor has a net positive charge and an n-type semiconductor has a net negative charge. In fact, both types of semiconductors have a net *zero* charge because they are made up of neutral atoms. Instead, "p-type" means that the majority of the mobile carriers of charge within the semiconductor are positive (holes), and "n-type" means that the majority of mobile 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?

ANSWER

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 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.

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.)

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

$$I = I_S(e^{eV/kT} - 1) \tag{42.22}$$

Current through a *p*-*n* junction; Saturation current; Voltage; Absolute temperature; Boltzmann constant; Magnitude of electron charge; $e = 2.71828\ldots$

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.*

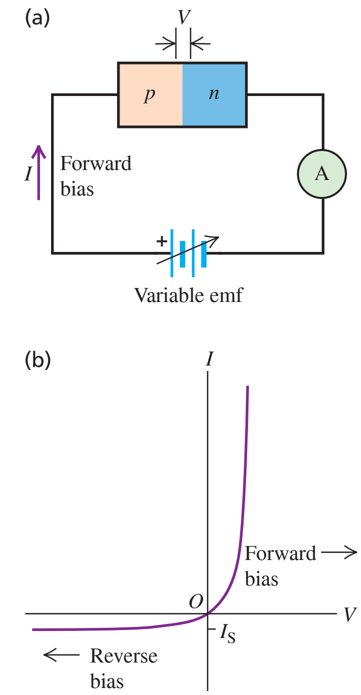


Figure 42.29 (a) A semiconductor *p*-*n* junction in a circuit. (b) Graph showing the asymmetric current–voltage relationship. The curve is described by Eq. (42.22).

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

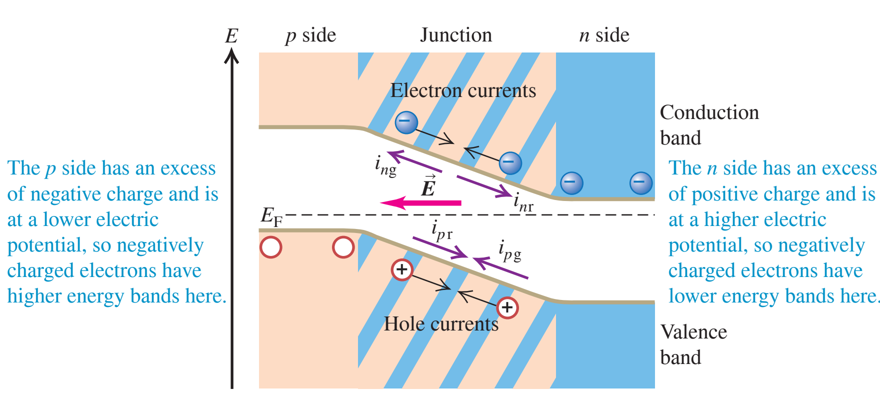


Figure 42.30 A *p*-*n* junction in equilibrium, with no externally applied field or potential difference. The generation currents (subscript g) and recombination currents (subscript r) exactly balance. The Fermi energy E_F is the same on both sides of the junction. The excess positive and negative charges on the n and p sides produce an electric field $\vec{E}$ in the direction shown.

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:

$$|i_{pg}| = |i_{pr}| \quad \text{and} \quad |i_{ng}| = |i_{nr}| \tag{42.23}$$

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

$$\begin{aligned} i_{p\text{tot}} &= i_{pr} - |i_{pg}| \\ &= |i_{pg}|e^{eV/kT} - |i_{pg}| \\ &= |i_{pg}|(e^{eV/kT} - 1) \end{aligned} \tag{42.24}$$

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.

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.)

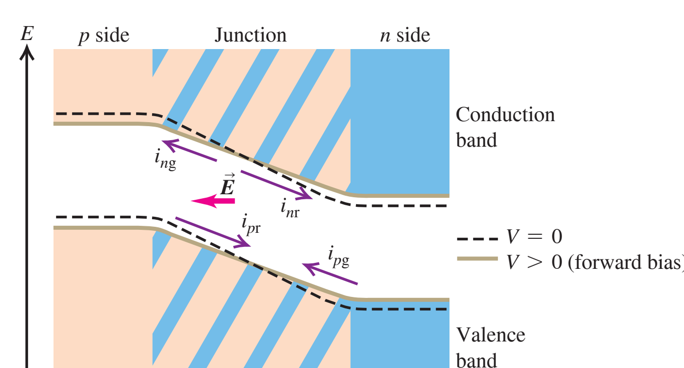
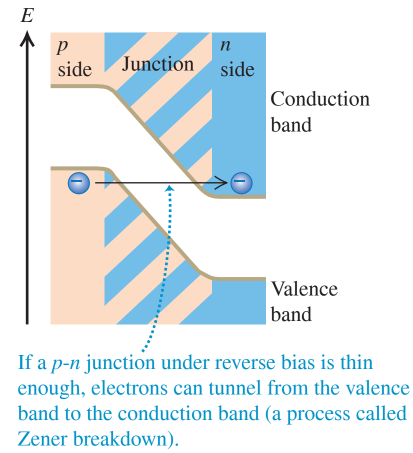


Figure 42.31 A p-n junction under forward-bias conditions. The potential difference between p and n regions is reduced, as is the electric field within the junction. The recombination currents increase but the generation currents are nearly constant, causing a net current from left to right. (Compare Fig. 42.30.)

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.

Figure 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.

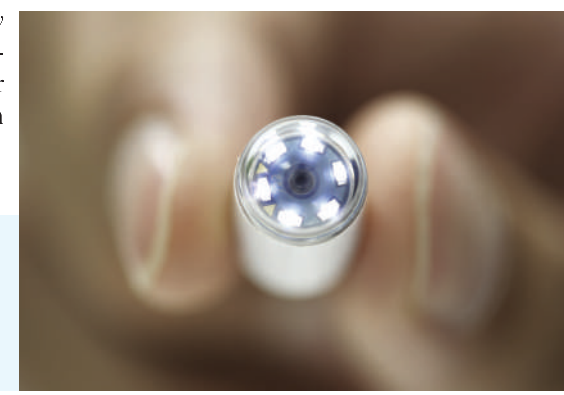


Semiconductor Devices and Light

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.

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.



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** (next page). 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.

Figure 42.33 Schematic diagram of a p-n-p transistor and circuit.

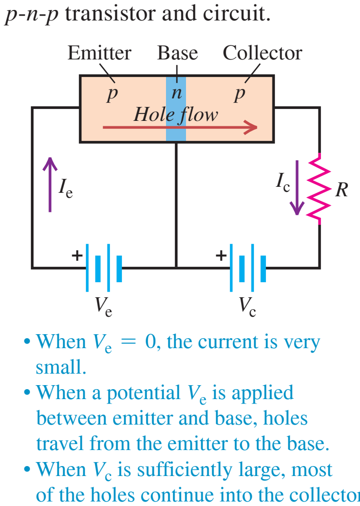
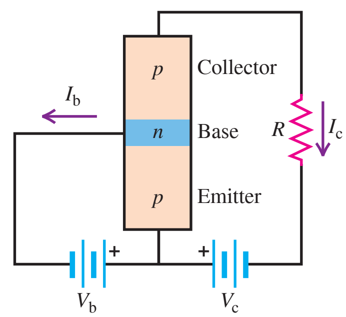
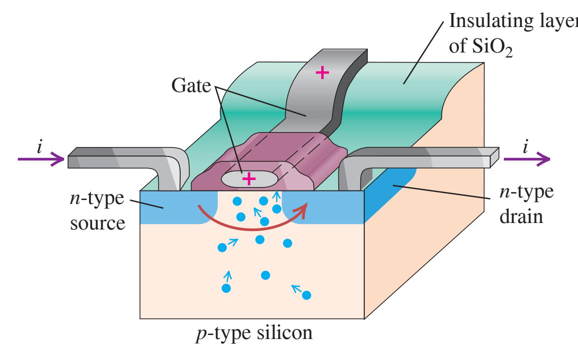


Figure **42.34** A common-emitter circuit.



- 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.

Figure **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.

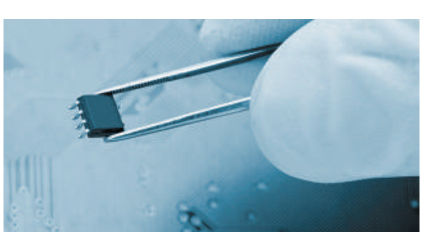


The *field-effect transistor* (**Fig. 42.35**) is an important type. In one variation a slab of *p*-type silicon is made with two *n*-type regions on the top, called the *source* and the *drain;* a metallic conductor is fastened to each. A third electrode called the *gate* is separated from the slab, 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

Figure **42.36** An integrated circuit chip smaller than your fingertip can contain millions of transistors.



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, mobile devices, and present-day computers (**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?

ANSWER

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.

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

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.)

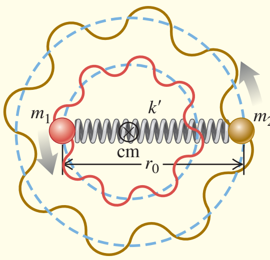
$$E_l = l(l+1)\frac{\hbar^2}{2I} \quad (l = 0, 1, 2, \dots) \qquad (42.3)$$

$$I = m_r r_0^2 \qquad (42.6)$$

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \qquad (42.4)$$

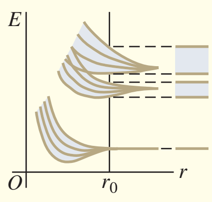
$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\omega = \left(n + \tfrac{1}{2}\right)\hbar\sqrt{\frac{k'}{m_r}} \qquad (42.7)$$

$$(n = 0, 1, 2, \dots)$$



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.)

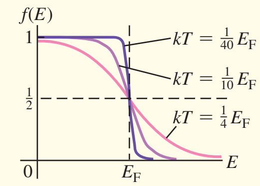


Continued

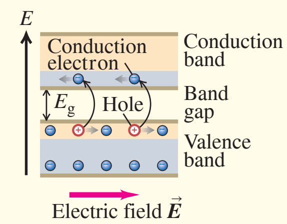
Free-electron model of metals: In the free-electron model of the behavior of conductors, the electrons are treated as completely free particles within the conductor. In this model the density of states is given by Eq. (42.15). The probability that an energy state of energy E is occupied is given by the 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.)

$$g(E) = \frac{(2m)^{3/2}V}{2\pi^2\hbar^3}E^{1/2} \qquad (42.15)$$

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1} \qquad (42.16)$$

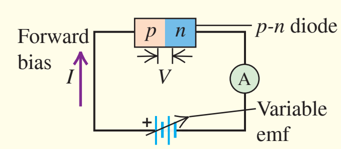


Semiconductors: A semiconductor has an energy gap of about 1 eV between its valence and conduction bands. Its electrical properties may be drastically changed by the addition of small concentrations of donor impurities, giving an n-type semiconductor, or acceptor impurities, giving a p-type semiconductor. (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).

$$I = I_S(e^{eV/kT} - 1) \qquad (42.22)$$



Chapter 42 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 42.2 and 42.3 (Section 42.2) before attempting these problems.

VP42.3.1 The two nuclei in the nitric oxide (NO) molecule are 0.1154 nm apart. The mass of the most common nitrogen atom is 2.326×10^{-26} kg, and the mass of the most common oxygen atom is 2.656×10^{-26} kg. Find (a) the reduced mass of the NO molecule, (b) the moment of inertia of the NO molecule, and (c) the energies, in meV, of the lowest three rotational energy levels of NO.

VP42.3.2 The $l = 2$ rotational level of the hydrogen chloride (HCl) molecule has energy 7.90 meV. The mass of the most common hydrogen atom is 1.674×10^{-27} kg, and the mass of the most common chlorine atom is 5.807×10^{-26} kg. Find (a) the moment of inertia of the HCl molecule, (b) the reduced mass of the HCl molecule, and (c) the distance between the H and Cl nuclei.

VP42.3.3 The mass of the most common silicon atom is 4.646×10^{-26} kg, and the mass of the most common oxygen atom is 2.656×10^{-26} kg. When a molecule of silicon monoxide (SiO) makes a transition between the $l = 1$ and $l = 0$ rotational levels, it emits a photon of wavelength 6.882 mm. Find (a) the moment of inertia of the SiO molecule, (b) the reduced mass of the SiO molecule, and (c) the distance between the Si and O nuclei.

VP42.3.4 A CO molecule is initially in the $n = 2$ vibrational level. If this molecule loses both vibrational and rotational energy and emits a photon, what are the photon wavelength and frequency if the initial angular momentum quantum number is (a) $l = 3$ and (b) $l = 2$?

Be sure to review EXAMPLES 42.6 and 42.7 (Section 42.5) before attempting these problems.

VP42.7.1 For free electrons in a solid with Fermi energy E_F and temperature T, find the energy for which the probability that a state at that energy is occupied is (a) 0.33 and (b) 0.90.

VP42.7.2 The Boltzmann constant is $k = 8.617 \times 10^{-5}$ eV/K. For a metallic solid at room temperature (293 K), what is the probability that an electron state is occupied if its energy is (a) 0.0250 eV below the Fermi level, (b) 0.0400 eV above the Fermi level, and (c) 0.100 eV above the Fermi level?

VP42.7.3 Silver contains 5.8×10^{28} free electrons per cubic meter. At absolute zero, what are (a) the Fermi energy (in J and eV) of silver, (b) the speed of an electron with this energy, and (c) the density of states (in states/J and states/eV) at the Fermi energy for a block of silver of volume 1.0 cm^3?

VP42.7.4 If kT is small compared to the Fermi energy at absolute zero, the Fermi energy at temperature T is essentially the same as at absolute zero. Use the results of the previous problem to find the electron energy, in eV, for which the probability that an electron state is occupied is (a) 0.92 and (b) 1.0×10^{-4} for silver at room temperature (293 K).

Be sure to review EXAMPLE 42.9 (Section 42.6) before attempting these problems.

VP42.9.1 For a certain semiconductor, the Fermi energy is in the middle of its band gap. If the temperature of the semiconductor is 285 K, find the probability that a state at the bottom of the conduction band is occupied if the band gap is (a) 0.500 eV and (b) 1.50 eV.

VP42.9.2 Find the ratio of the probability that a state at the bottom of the conduction band of a semiconductor is occupied at 315 K to the probability at 295 K if the band gap is (a) 0.400 eV and (b) 0.800 eV. Assume that the Fermi energy is in the middle of the band gap.
VP42.9.3 Consider a material whose Fermi energy is in the middle of its band gap. What band-gap width provides a probability of 0.00190 that a state at the bottom of the conduction band is occupied if the temperature of the material is (a) 305 K and (b) 325 K?
VP42.9.4 In a certain semiconductor, the Fermi energy lies above the top of the valence band by an amount equal to $\frac{3}{4}$ of the band gap. Find the probability that a state at the bottom of the conduction band is occupied if the band gap is 0.300 eV and the temperature is (a) 275 K and (b) 325 K.

BRIDGING PROBLEM Molecular Vibration and Semiconductor Band Gap

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

•, ••, •••: 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 1*s*, 2*s*, 2*p*, 3*s*, 3*p*, and 4*s* 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 • **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.
42.2 • 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.
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 $l = 1$.

Section 42.2 Molecular Spectra

42.4 •• The H_2 molecule has a moment of inertia of $4.6 \times 10^{-48}\ \text{kg} \cdot \text{m}^2$. 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?
42.5 • 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 • 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.
42.7 •• **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)?
42.8 • 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) $n = 0, l = 4 \rightarrow n = 1, l = 3$.
42.9 • 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?
42.10 • 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.
42.11 •• 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.

Section 42.3 Structure of Solids

42.12 • Potassium bromide (KBr) has a density of $2.75 \times 10^3\ \text{kg/m}^3$ 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 42.13)? Is the relationship between the two values qualitatively what you would expect? Explain.
42.13 • **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.

Section 42.4 Energy Bands

42.14 • 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.
42.15 • 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.

42.16 • 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

42.17 • Calculate the density of states $g(E)$ for the free-electron model of a metal if $E = 7.0$ eV and $V = 1.0\ \text{cm}^3$. Express your answer in units of states per electron volt.

42.18 • 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.)

42.19 • Consider the density of states in the free-electron model for an energy of 1.60 eV. At what energy will the density of states be (a) doubled and (b) halved?

42.20 • For a metal with kT equal to 25% of the Fermi energy, what is the probability that a state will be occupied by an electron if its energy is (a) 50% of the Fermi energy and (b) 50% greater than the Fermi energy? Are your results consistent with Fig. 42.23?

42.21 • **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, $C_V = 25.3\ \text{J/mol} \cdot \text{K}$. (c) Is the value of C_V due principally to the electrons? If not, to what is it due? (*Hint:* See Section 18.4.)

42.22 •• At the Fermi temperature T_F, $E_F = kT_F$ (see Exercise 42.18). When $T = T_F$, what is the probability that a state with energy $E = 2E_F$ is occupied?

42.23 •• 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

42.24 • 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.

42.25 • 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

42.26 •• (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?

42.27 • **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?

42.28 • For a certain *p-n* junction diode, the saturation current at room temperature $(20°\text{C})$ 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?

42.29 •• (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?

42.30 •• 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?

42.31 • The saturation current for a junction diode is 10 microamps. Calculate the resistance of the diode at forward and reverse potential differences of 0.20 V when $T = 17°\text{C}$.

PROBLEMS

42.32 • 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?

42.33 •• **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 $3.0 \times 10^{-29}\ \text{C} \cdot \text{m}$. 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 $1.5 \times 10^{-30}\ \text{C} \cdot \text{m}$. 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.

42.34 • 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.

42.35 •• **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

$$\frac{n_l}{n_0} = (2l + 1)e^{-[l(l+1)\hbar^2]/2IkT}$$

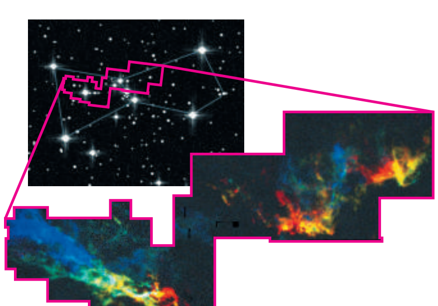
(*Hint:* For each value of l, how many states are there with different values of m_l?) (b) Determine the ratio n_l/n_0 for a gas of CO molecules at 300 K for (i) $l = 1$; (ii) $l = 2$; (iii) $l = 10$; (iv) $l = 20$; (v) $l = 50$. The moment of inertia of the CO molecule is given in Example 42.2 (Section 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.

42.36 ••• **CALC** Part (a) of Problem 42.35 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.)

42.37 •• Semiconductor devices can radiate heat up to a maximum power rating. If the device dissipates more power than that limit, it will heat up and explode, melt, or simply break down. A typical power rating for a diode is 500 mW. The power dissipated by any two-terminal device is the product of the current through it and the voltage across it. (a) Use Eq. (42.22) to estimate the power dissipated by a diode when it is forward biased at $V = 0.6$ V at a temperature of 300 K. The saturation current is 1.2×10^{-11} A. (b) Estimate to one significant figure the forward voltage at which the device will dissipate 500 mW. (c) Then what is the current?

42.38 •• 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. P42.38**. 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 42.35). 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 **P42.38**



42.39 • 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?

42.40 • 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.)

42.41 • 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) $n = 2, l = 2 \rightarrow n = 1, l = 3$.

42.42 • 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?

42.43 •• 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.

42.44 •• **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 an 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?

42.45 ••• 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.

42.46 •• **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?

42.47 •• 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.

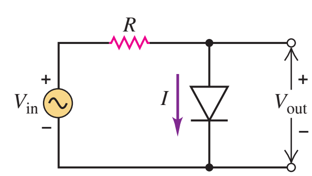
$f(E)$	0.064	0.173	0.390	0.661	0.856	0.950
E (eV)	3.0	2.5	2.0	1.5	1.0	0.5

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.

42.48 •• 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.]

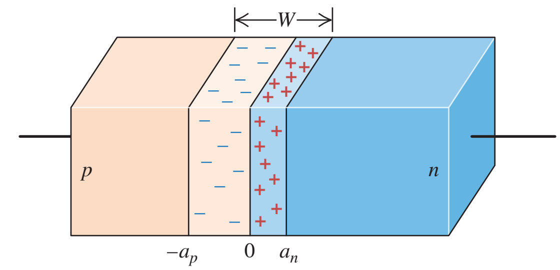
42.49 •• The circuit shown in **Fig. P42.49** is called a half-wave rectifier. The triangular symbol with the flat lower edge represents a diode, such that a forward bias drives the current downward. Use Eq. (42.22) to study the behavior of this circuit. Use a typical value for the saturation current: $I_S = 5 \times 10^{-13}$ A. Use a temperature of 300 K. (a) Any predicted current less than 1 μA can be considered a zero current. Determine the minimum voltage V_{out} for which the current I exceeds 1 μA. (b) The current increases dramatically for diode voltages that exceed that value. At a given diode voltage V_{out}, the diode has an effective resistance given by $R_D = V_{out}/I$. Use Eq. (42.22) to estimate the diode voltage $V_{out} = V_{max}$ for which $R_D = 100\ \Omega$. (c) What is the current I in that case? (d) If the resistance R is greater than 10 kΩ, then $R_D \ll R$ for larger currents, and there is a negligible additional voltage drop across the diode. This effectively pins $V_{out} = V_{max}$ for all input voltages $V_{in} > V_{max}$. With this perspective, estimate how this circuit would respond to a sinusoidal input voltage with amplitude 5 V.

Figure **P42.49**



42.50 ••• CP CALC A p-n junction includes p-type silicon, with donor atom density N_D, adjacent to n-type silicon, with acceptor atom density N_A. Near the junction, free electrons from the n side have diffused into the p side while free holes from the p side have diffused into the n side, leaving a "depletion region" of width W where there are no free charge carriers (**Fig. P42.50**). The width of the depletion region on the n side is a_n, and the width of the depletion region on the p side is a_p. Each depleted region has a constant charge density with magnitude equal to the corresponding donor atom density. Assume there is no net charge outside the depleted regions. Define an x-axis pointing from the p side toward the n side with the origin at the center of the junction. An electric field $\vec{E} = -E\hat{\imath}$ has developed in the depletion region, stabilizing the diffusion. (a) Use Gauss's law to determine the electric field on the p side of the junction, for $-a_p \le x \le 0$. (*Hint:* Use a cylindrical Gaussian surface parallel to the x-axis, with the left end outside the depletion region (where $\vec{E} = 0$) and the right side inside the region.) Note that the dielectric constant of silicon is $K = 11.7$. (b) Determine the electric field on the n side of the junction, for $0 \le x \le a_n$.

Figure **P42.50**



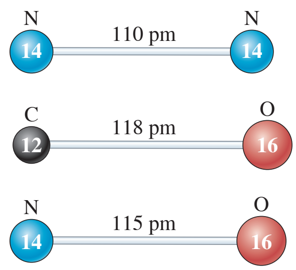
(c) Use continuity at $x = 0$ to determine a relationship between the depths a_p and a_n and in terms of N_A and N_D. (d) The electric potential $V(x)$ may be obtained by integrating the electric field, as shown in Eq. (23.18). Determine $V(x)$ in the region $-a_p \le x \le 0$ using the convention that $V(x) = 0$ for $x < -a_p$. (e) Similarly, determine $V(x)$ in the region $0 \le x \le a_n$. (f) What is the "barrier potential" $V_b = V(a_n) - V(-a_p)$? (g) The p side is doped with boron atoms with density $N_A = 1.00 \times 10^{16}\ \text{cm}^{-3}$. The n side is doped with arsenic atoms with density $N_D = 5.00 \times 10^{16}\ \text{cm}^{-3}$. The n side depletion depth is $a_n = 55.0$ nm. What is the p side depletion depth a_p? (h) What is the peak magnitude of the electric field, at $x = 0$? (i) What is the value of the barrier potential V_b?

42.51 ••• CP Consider the CO_2 molecule shown in Fig. 42.10c. The oxygen molecules have mass M_O and the carbon atom has mass M_C. Parameterize the positions of the left oxygen atom, the carbon atom, and the right oxygen atom using x_1, x_2, and x_3 as the respective rightward deviations from equilibrium. Treat the bonds as Hooke's-law springs with common spring constant k'. (a) Use Newton's second law to obtain expressions for $M_i\ddot{x}_i = M_i\, d^2x/dt^2$ in each case $i = 1, 2, 3$, where $M_{1,2,3} = (M_O, M_C, M_O)$. (*Note:* We represent time derivatives using dots.) Assume $X_C = 0$ at $t = 0$. (b) To ascertain the motion of the asymmetric stretching mode, set $x_1 = x_3 \equiv X_O$ and set $x_2 \equiv X_C$. Write the two independent equations that remain from your previous result. (c) Eliminate the sum $X_O + X_C$ from your equations. Use what remains to ascertain X_C in terms of X_O. (d) Substitute your expression for X_C into your equation for X_O to derive a harmonic oscillator equation $M_{eff}\ddot{X}_O = -kX_O$. What is M_{eff}? (e) This equation has the solution $X_O(t) = A\cos(\omega t)$. What is the angular frequency ω? (f) Using the experimentally determined spring constant $k' = 1860$ N/m and the atomic masses $M_C = 12$ u and $M_O = 16$ u, where $u = 1.6605 \times 10^{-27}$ kg, to determine the oscillation frequency $f = \omega/2\pi$.

CHALLENGE PROBLEMS

42.52 ••• A colorless, odorless gas in a tank needs to be identified. The gas is nitrogen (N_2), carbon monoxide (CO), or nitric oxide (NO). (See **Fig. P42.52**.) White light is passed through a sample to obtain an absorption spectrum. Among others, the following wavelengths present dark bands on the spectrum: $\lambda = 4.2680\ \mu$m, $4.2753\ \mu$m, $4.2826\ \mu$m, $4.2972\ \mu$m, and $4.3046\ \mu$m. These represent transitions from the lowest rotational states. (a) What photon energies correspond to these wavelengths? Specify your answers to five significant figures. Note that the speed of light is 2.9979×10^8 m/s and $h = 4.1357 \times 10^{-15}$ eV · s. (b) Which one of the absorbed energies corresponds to transitions *from* an $l = 0$ state of the gas? (*Hint:* Use Eq. (42.9) along with the transition rules as a guide.) (c) Which one of the absorbed energies corresponds to a transition *to* an $l = 0$ state? (d) Determine the value of $\hbar\sqrt{k'/m_r}$, where k' is the effective spring constant of the molecule and m_r is its reduced mass. (*Hint:* Use the previous two results and Eq. (42.9).) (e) Determine the molecule's moment of inertia I. (f) The atomic masses and the bond

Figure **P42.52**



lengths for the three molecule options are shown in Fig. P42.52, where u = 1.6605×10^{-27} kg is one atomic mass unit. Based on the data, what is the mystery gas? (g) What is the reduced mass m_r in terms of u? (h) What is the spring constant k'?

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 Challenge 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 Challenge 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?

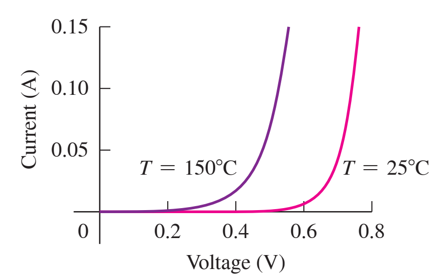
42.55 •• In the discussion of free electrons in Section 42.5, we assumed that we could ignore the effects of relativity. This is not a safe assumption if the Fermi energy is greater than about $\frac{1}{100}mc^2$ (that is, more than about 1% of the rest energy of an electron). (a) Assume that the Fermi 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 8×10^9 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 Challenge 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.

MCAT-STYLE 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.

Key Example VARIATION Problems

VP42.3.1 (a) 1.240×10^{-26} kg (b) 1.651×10^{-46} kg $\cdot$ m^2 (c) 0, 0.4203 meV, 1.261 meV
VP42.3.2 (a) 2.64×10^{-47} kg $\cdot$ m^2 (b) 1.627×10^{-27} kg (c) 0.127 nm
VP42.3.3 (a) 3.853×10^{-46} kg $\cdot$ m^2 (b) 1.690×10^{-26} kg (c) 0.1510 nm
VP42.3.4 (a) $f = 6.538 \times 10^{13}$ Hz, $\lambda = 4.585\ \mu$m (b) $f = 6.527 \times 10^{13}$ Hz, $\lambda = 4.593\ \mu$m
VP42.7.1 (a) $E_F + 0.71kT$ (b) $E_F - 2.2kT$
VP42.7.2 (a) 0.729 (b) 0.170 (c) 0.0187
VP42.7.3 (a) 8.8×10^{-19} J $= 5.5$ eV (b) 1.4×10^6 m/s (c) 9.9×10^{40} states/J $= 1.6 \times 10^{22}$ states/eV
VP42.7.4 (a) 5.4 eV (b) 5.7 eV
VP42.9.1 (a) 3.79×10^{-5} (b) 5.46×10^{-14}
VP42.9.2 (a) 1.65 (b) 2.72
VP42.9.3 (a) 0.329 eV (b) 0.351 eV
VP42.9.4 (a) 0.0405 (b) 0.0643

Bridging Problem

(a) 0.278 eV
(b) 1.74×10^{-25}

This sculpture of a woolly 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 OUTCOMES

In this chapter, you'll learn...

- **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.

You'll need to review...

- **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

$$R = R_0 A^{1/3} \qquad (43.1)$$

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 value is

$$1\ \text{u} = 1.660539040(20) \times 10^{-27}\ \text{kg}$$

In Section 43.2 we'll discuss the masses of nuclei in more detail.

CAUTION Nuclear masses are rest masses When we speak of the masses of the proton, the neutron, and nuclei, we always mean their *rest* masses (see Section 37.7). We'll use this same convention when we discuss other subatomic particles in Chapter 44.

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 approximately proportional to A. Since the proton and the neutron both have about the same mass, the *mass* of a nucleus is also approximately proportional to the total number A of protons and neutrons. When we divide the mass by the volume to calculate the density, the factors of A cancel. Thus *all nuclei have approximately the same density.* We'll see in Section 43.2 that this observation is important in understanding nuclear structure.

EXAMPLE 43.1 Calculating nuclear properties

The most common kind of iron nucleus has mass number $A = 56$. Find the approximate radius, mass, and density of the nucleus.

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 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

$$\begin{aligned} V &= \tfrac{4}{3}\pi R^3 = \tfrac{4}{3}\pi R_0^3 A = \tfrac{4}{3}\pi(4.6 \times 10^{-15}\ \text{m})^3 \\ &= 4.1 \times 10^{-43}\ \text{m}^3 \end{aligned}$$

$$\rho = \frac{m}{V} \approx \frac{9.3 \times 10^{-26}\ \text{kg}}{4.1 \times 10^{-43}\ \text{m}^3} = 2.3 \times 10^{17}\ \text{kg/m}^3$$

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!

KEYCONCEPT The radius of a nucleus is approximately proportional to the $\frac{1}{3}$ power of the number of nucleons A that the nucleus contains. The volume is approximately proportional to A.

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

$$\begin{aligned} &\text{Proton:} && m_p = 1.007276\ \text{u} = 1.672622 \times 10^{-27}\ \text{kg} \\ &\text{Neutron:} && m_n = 1.008665\ \text{u} = 1.674927 \times 10^{-27}\ \text{kg} \\ &\text{Electron:} && m_e = 0.000548580\ \text{u} = 9.10938 \times 10^{-31}\ \text{kg} \end{aligned}$$

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:

$$A = Z + N \qquad (43.2)$$

A single nuclear species having specific values of both Z and N is called a **nuclide. Table 43.1** (next page) 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

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).

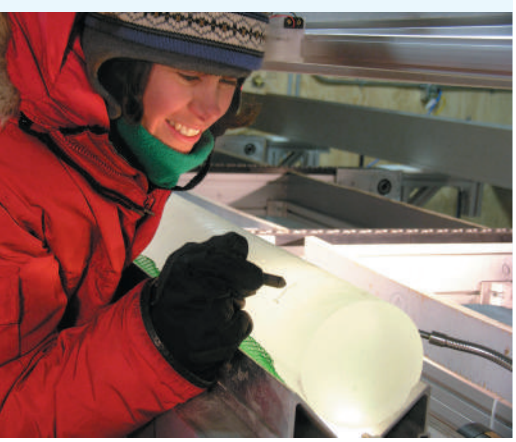


TABLE 43.1 Compositions of Some Common Nuclides

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

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.

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

$$S = \sqrt{\tfrac{1}{2}(\tfrac{1}{2} + 1)}\hbar = \sqrt{\tfrac{3}{4}}\hbar \qquad (43.3)$$

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 ($^{1}_{1}H$)	1	0	1.007825	1
Deuterium ($^{2}_{1}H$)	1	1	2.014102	2
Tritium ($^{3}_{1}H$)	1	2	3.016049	3
Helium ($^{3}_{2}He$)	2	1	3.016029	3
Helium ($^{4}_{2}He$)	2	2	4.002603	4
Lithium ($^{6}_{3}Li$)	3	3	6.015123	6
Lithium ($^{7}_{3}Li$)	3	4	7.016003	7
Beryllium ($^{9}_{4}Be$)	4	5	9.012183	9
Boron ($^{10}_{5}B$)	5	5	10.012937	10
Boron ($^{11}_{5}B$)	5	6	11.009305	11
Carbon ($^{12}_{6}C$)	6	6	12.000000	12
Carbon ($^{13}_{6}C$)	6	7	13.003355	13
Nitrogen ($^{14}_{7}N$)	7	7	14.003074	14
Nitrogen ($^{15}_{7}N$)	7	8	15.000109	15
Oxygen ($^{16}_{8}O$)	8	8	15.994915	16
Oxygen ($^{17}_{8}O$)	8	9	16.999132	17
Oxygen ($^{18}_{8}O$)	8	10	17.999160	18

Source: International Atomic Energy Agency.

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'll 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:

$$\mu_n = \frac{e\hbar}{2m_p} = 5.05078 \times 10^{-27}\ \text{J/T} = 3.15245 \times 10^{-8}\ \text{eV/T} \qquad \text{(nuclear magneton)} \tag{43.7}$$

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

$$|\mu_{sz}|_{\text{proton}} = 2.7928\mu_n \tag{43.8}$$

Even more surprising, the neutron, which has zero charge, has a spin magnetic moment; its z-component has magnitude

$$|\mu_{sz}|_{\text{neutron}} = 1.9130\mu_n \tag{43.9}$$

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

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.

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

$$\begin{aligned} U &= -|\mu_z|B = -(2.7928)(3.152 \times 10^{-8}\ \text{eV/T})(2.30\ \text{T}) \\ &= -2.025 \times 10^{-7}\ \text{eV} \end{aligned}$$

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.

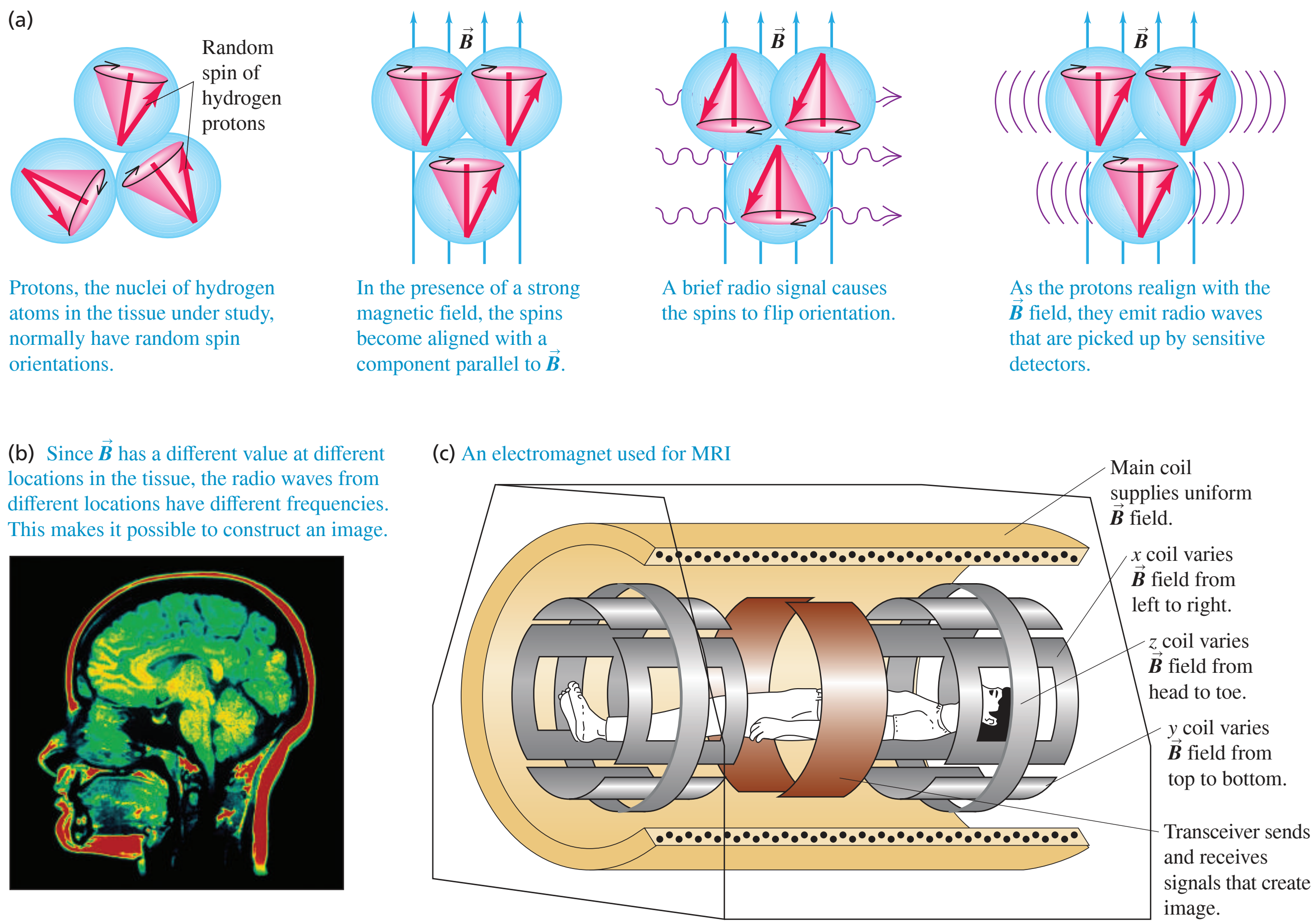
KEYCONCEPT Both the proton and the neutron have a magnetic moment $\vec{\boldsymbol{\mu}}$ associated with its spin. When placed in a magnetic field $\vec{\boldsymbol{B}}$, a proton or neutron has two states of different energy: one where the component of $\vec{\boldsymbol{\mu}}$ in the direction of $\vec{\boldsymbol{B}}$ is positive and one where it is negative.

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

Figure 43.1 Magnetic resonance imaging (MRI).



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.

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.

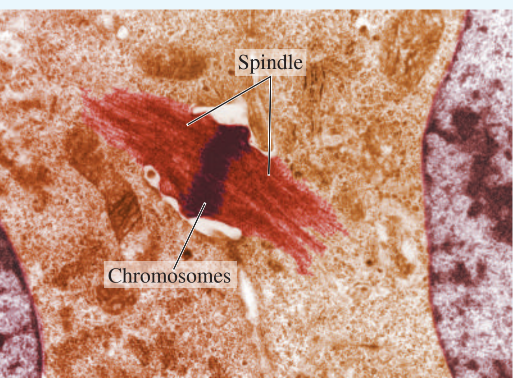
ANSWER

(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 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 minimum 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

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.



$$\underbrace{E_B}_{\substack{\text{Binding energy}\\ \text{of a nucleus}\\ \text{with } Z \text{ protons,}\\ N \text{ neutrons}}} = (\overset{\text{Atomic number}}{Z}\underset{\substack{\text{Mass of}\\ \text{hydrogen atom}}}{M_H} + \overset{\text{Neutron number}}{N}\underset{\substack{\text{Neutron}\\ \text{mass}}}{m_n} - \underset{\substack{\text{Mass of neutral atom}\\ \text{containing nucleus}}}{{}^{A}_{Z}M})\overset{\substack{(\text{Speed of light in vacuum})^2\\ = 931.5\text{ MeV/u}}}{c^2} \qquad (43.10)$$

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 ${}^2_1H$, 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 $(2.224\text{ MeV})/(2\text{ nucleons}) = 1.112\text{ MeV}$ per nucleon, ${}^2_1H$ has the lowest binding energy per nucleon of all nuclides. The nuclide with the highest value of E_B/A is ${}^{62}_{28}Ni$ (see Example 43.3).

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 $\Delta M = E_B/c^2 = (2.224\text{ MeV})/(931.5\text{ MeV/u}) = 0.002388\text{ u}$.

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 from this section and Section 43.1.

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.

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

WITH VARIATION PROBLEMS

Find the mass defect, total binding energy, and binding energy per nucleon of ${}^{62}_{28}Ni$, which has neutral atomic mass 61.928345 u.

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

$$E_B = (0.585365\text{ u})(931.5\text{ MeV/u}) = 545.3\text{ MeV}$$

The binding energy *per nucleon* is $E_B/A = (545.3\text{ MeV})/62$, or 8.795 MeV per nucleon. This value is the highest of any nuclide (**Fig. 43.2**).

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.

KEYCONCEPT The mass defect of a nucleus is the difference between the sum of the masses of its nucleons and the mass of the nucleus itself. We calculate this using the mass of a hydrogen atom rather than the mass of a proton, and using the mass of a neutral atom that contains the nucleus rather than the nuclear mass. To find the binding energy, multiply the mass defect by c^2; to then find the binding energy per nucleon, divide by the nucleon number A.

Figure **43.2** Approximate binding energy per nucleon as a function of mass number A (the total number of nucleons) for stable nuclides.

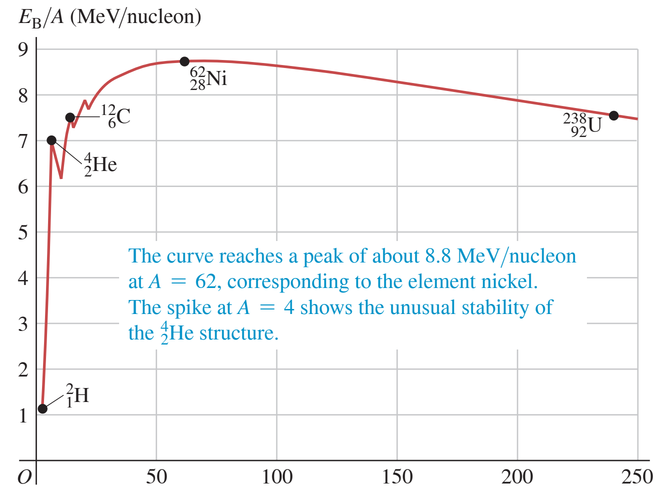


Figure 43.2 is a graph of binding energy per nucleon as a function of the mass number A. This graph shows that nearly all stable nuclides have binding energies per nucleon in the range 7–9 MeV. Note the spike at $A = 4$, showing the unusually large binding energy per nucleon of the $^{4}_{2}\mathrm{He}$ 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:

1. The nuclear force does not depend on charge; the binding is the same for both neutrons and protons. It also has a short range, of the order of nuclear dimensions—that is, 10^{-15} m. (If it had a longer range, a nucleus would grow by pulling in additional protons and neutrons.) Within its range, the attractive nuclear force is much stronger than the repulsive electric forces between protons; otherwise, the nucleus could never be stable. Unfortunately, there is no simple equation like Coulomb's law for how the nuclear force depends on the distance between nucleons.
2. The observation that nuclei all have about the same density and about the same binding energy per nucleon, E_B/A, shows that a particular nucleon cannot interact simultaneously through the nuclear force with *all* the other nucleons in a nucleus. (If it could, then in larger nuclei the nucleons would crowd closer together, making the nucleus denser and increasing the value of E_B/A.) Instead, a nucleon interacts with only the few other nucleons in its immediate vicinity. (This is different from electric forces; *every* proton in the nucleus exerts an electric repulsion on every other proton.) This limit on the number of interactions via the nuclear force is called *saturation*.
3. The nuclear force favors the 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, with each pair having opposite spins. This explains the spike in binding energy per nucleon shown in Fig. 43.2 for $^{4}_{2}\mathrm{He}$ (two protons and two neutrons). 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 a superconductor.)

Analyzing nuclear structure is a challenge because the nature of the nuclear force is so complex and because electric forces are also involved. 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 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. An individual nucleon interacts with only a few of its nearest neighbors (saturation). If every nucleon in a nucleus were surrounded by nearest neighbors, the binding energy per nucleon, E_B/A, would have the same constant value for all nucleons. If we call this constant C_1, then the binding energy of a nucleus with A nucleons would be C_1A.
2. The nucleons on the surface of the nucleus have no neighbors outside the surface and so are less tightly bound than those in the interior. 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:

$$E_B = C_1A - C_2A^{2/3} - C_3\frac{Z(Z-1)}{A^{1/3}} - C_4\frac{(A-2Z)^2}{A} \pm C_5A^{-4/3} \quad (43.11)$$

(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. The other terms in Eq. (43.11) are responsible for the difference. The negative terms with C_3 and C_4 become more important for larger nuclides that have more protons and have A greater than $2Z$; that explains why the binding energy per nucleon gradually decreases beyond $A = 62$, as Fig. 43.2 shows.

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.

EXAMPLE 43.4 Estimating binding energy and mass

WITH VARIATION PROBLEMS

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.

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):

$$ {}^{62}_{28}M = 28(1.007825\text{ u}) + 34(1.008665\text{ u}) - \frac{548.5\text{ MeV}}{931.5\text{ MeV/u}} $$
$$ = 61.925\text{ u} $$

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!

KEYCONCEPT You can use the liquid-drop model and its associated semiempirical mass formula to calculate the binding energy and mass of a nuclide.

The liquid-drop model and the mass formula derived from it are quite successful in correlating nuclear masses, and we'll 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

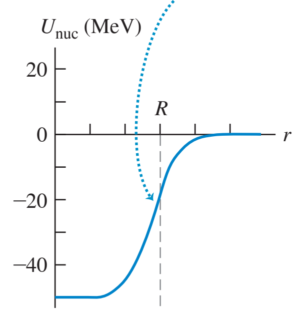
The **shell model** of nuclear structure is analogous to the central-field approximation in atomic physics (see Section 41.6). We picture each nucleon as moving in a potential that represents the averaged-out effect of all the other nucleons. Although this is a very simplified model, in several respects it works out quite well.

The potential-energy function for the nuclear force is the same for protons as for neutrons. **Figure 43.3a** (next page) 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 function is somewhat rounded because the nucleus doesn't have a sharply defined surface at radius R. For a proton there is an additional electric potential energy due to interactions with the other $Z - 1$ protons, which we collectively treat as a uniformly charged sphere of 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.

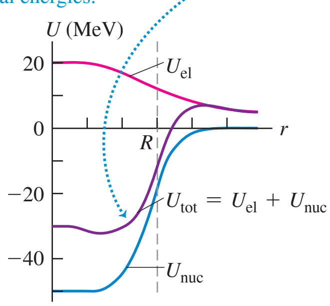
In principle, we could solve the Schrödinger equation for a proton or neutron moving in such a potential. For any spherically symmetric potential energy, the angular-momentum states are the same as for the electrons in the central-field approximation in atomic physics.

Figure 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.



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 multielectron 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 filled-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) $^{4}_{2}$He; (ii) $^{52}_{24}$Cr; (iii) $^{152}_{62}$Sm; (iv) $^{200}_{80}$Hg; (v) $^{252}_{92}$Cf.

ANSWER

(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 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 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 $^{4}_{2}$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 $^{8}_{4}$Be nucleus immediately splitting into two $^{4}_{2}$He nuclei.

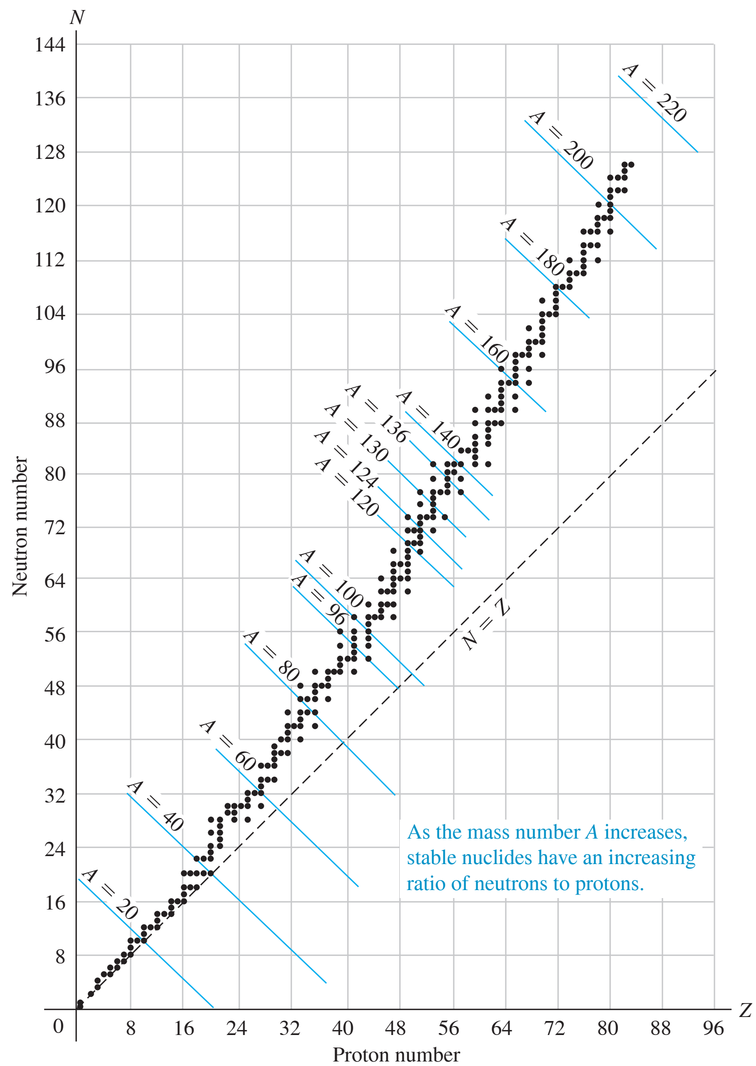
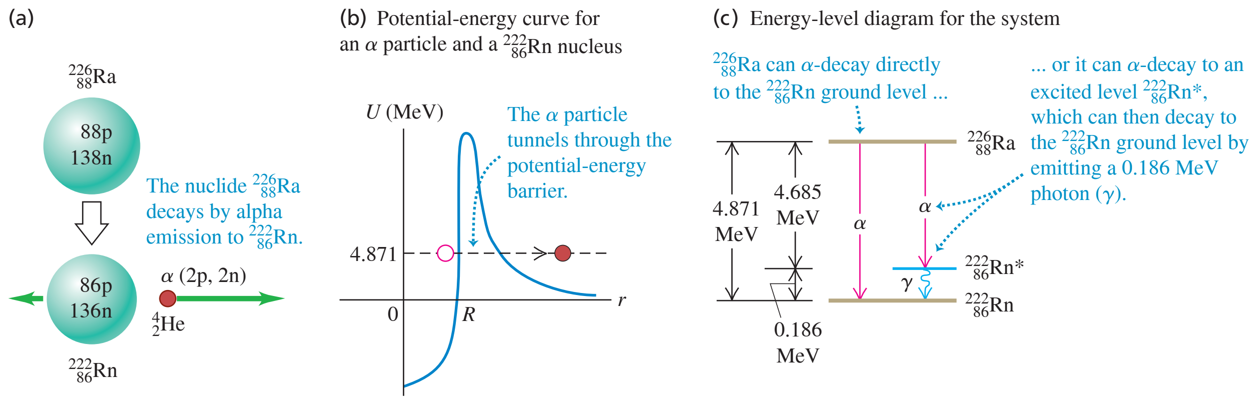


Figure **43.4** Segrè chart showing neutron number and proton number for stable nuclides.

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.

Figure 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 Decay

When unstable nuclides decay into different nuclides, they usually emit alpha (α) or beta (β) particles. An **alpha particle** is a $^{4}_{2}He$ 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. The nucleus gets rid of the excess nucleons in the form of a $^{4}_{2}He$ nucleus because this is a particularly stable grouping (see Fig. 43.2). 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 $^{4}_{2}He$ 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.)

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.

EXAMPLE 43.5 Alpha decay of radium — WITH VARIATION PROBLEMS

Show that the α-emission process $^{226}_{88}Ra \rightarrow {}^{222}_{86}Rn + {}^{4}_{2}He$ (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 $^{226}_{88}Ra$, 222.017578 u for $^{222}_{86}Rn$, and 4.002603 u for $^{4}_{2}He$.

IDENTIFY and SET UP Alpha emission is possible if the mass of the $^{226}_{88}Ra$ atom is greater than the sum of the atomic masses of $^{222}_{86}Rn$ and $^{4}_{2}He$. 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 $^{222}_{86}Rn$ atom are in motion after the decay; we'll 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

$$226.025410 \text{ u} - (222.017578 \text{ u} + 4.002603 \text{ u}) = +0.005229 \text{ u}$$

Since this is positive, α decay is energetically possible. The energy equivalent of this mass difference is

$$E = (0.005229 \text{ u})(931.5 \text{ MeV/u}) = 4.871 \text{ MeV}$$

In this process the $^{222}_{86}\text{Rn}$ 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 $^{226}_{88}\text{Ra}$ nucleus is at rest, the daughter $^{222}_{86}\text{Rn}$ 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 $^{226}_{88}\text{Ra}$ does emit α particles with a kinetic energy of 4.78 MeV. Check your results by verifying that the alpha particle and the $^{222}_{86}\text{Rn}$ 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 $^{222}_{86}\text{Rn}$ nucleus gets $4/(222 + 4)$ of the 4.871 MeV released]. You'll find that the alpha particle moves at $0.0506c = 1.52 \times 10^7 \text{ m/s}$; if momentum is conserved, you should find that the $^{222}_{86}\text{Rn}$ nucleus moves $\frac{4}{222}$ as fast. Does it?

KEYCONCEPT In alpha decay, a large, unstable nucleus with Z protons and N neutrons emits an alpha particle (a ^4_2He nucleus). The daughter nucleus has $Z - 2$ protons and $N - 2$ neutrons. The combined mass of the decay products is less than the mass of the parent; the lost rest energy goes into kinetic energy of the decay products.

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

$$\text{n} \rightarrow \text{p} + \beta^- + \bar{\nu}_e \tag{43.13}$$

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

WITH VARIATION PROBLEMS

The nuclide ${}^{60}_{27}\mathrm{Co}$, 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 ${}^{60}_{27}\mathrm{Co}$ and 59.930786 u for ${}^{60}_{28}\mathrm{Ni}$.

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 ${}^{60}_{27}\mathrm{Co}$ undergoes β^- decay and then compare its neutral atomic mass to that of ${}^{60}_{27}\mathrm{Co}$.

EXECUTE In the presumed β^- decay of ${}^{60}_{27}\mathrm{Co}$, Z increases by 1 from 27 to 28 and A remains at 60, so the final nuclide is ${}^{60}_{28}\mathrm{Ni}$. The neutral atomic mass of ${}^{60}_{27}\mathrm{Co}$ is greater than that of ${}^{60}_{28}\mathrm{Ni}$ by 0.003031 u, so β^- decay *can* occur.

EVALUATE With three decay products in β^- decay—the ${}^{60}_{28}\mathrm{Ni}$ 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 ${}^{60}_{27}\mathrm{Co}$ nucleus. By contrast, in alpha decay there are just two decay products, and their energies and momenta are determined uniquely (see Example 43.5).

KEYCONCEPT In beta-minus decay, an unstable nucleus with Z protons and N neutrons decays by converting a neutron into a proton and emitting an electron and an antineutrino. The daughter nucleus has $Z + 1$ protons and $N - 1$ neutrons. This is possible only if the neutral parent atom has a greater mass than the neutral daughter atom.

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

$$\mathrm{p} \rightarrow \mathrm{n} + \beta^+ + \nu_\mathrm{e} \tag{43.14}$$

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

$$\mathrm{p} + \beta^- \rightarrow \mathrm{n} + \nu_\mathrm{e} \tag{43.15}$$

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

WITH VARIATION PROBLEMS

The nuclide $^{57}_{27}\text{Co}$ 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 $^{57}_{27}\text{Co}$ and 56.935394 u for $^{57}_{26}\text{Fe}$.

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 $^{57}_{27}\text{Co}$ undergoes β^+ decay or electron capture and then find the corresponding mass difference.

EXECUTE The original nuclide is $^{57}_{27}\text{Co}$. 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 $^{57}_{26}\text{Fe}$. Its mass is less than that of $^{57}_{27}\text{Co}$ 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.

KEYCONCEPT In beta-plus decay, an unstable nucleus with Z protons and N neutrons decays by converting a proton into a neutron and emitting a positron and a neutrino. The daughter nucleus has $Z - 1$ protons and $N + 1$ neutrons. This is possible only if the neutral parent atom has a greater mass than the neutral daughter atom plus two electrons. If this is not the case, electron capture may be possible.

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}Ra have two possible kinetic energies, either 4.784 MeV or 4.602 MeV. Including the recoil energy of the resulting ^{222}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}Rn nucleus is left in an excited state. It then decays to its ground state by emitting a gamma-ray photon with energy $(4.871 - 4.685)\ \text{MeV} = 0.186\ \text{MeV}$.

CAUTION **Comparing α, β, and γ decays** 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. Remember that in each of these decays, the initial unstable nucleus does not simply disappear; it's just replaced by a different one (or, in the case of γ decay, a less-excited version of the initial nucleus).

Radioactive Decay Series

When a radioactive nucleus decays, the daughter nucleus may also be unstable. In this case a *series* of successive decays occurs until a stable configuration is reached. One of the most abundant radioactive nuclides 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** (next page). 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

$$^{238}\text{U} \rightarrow {}^{234}\text{Th} + \alpha \qquad \text{and} \qquad {}^{234}\text{Th} \rightarrow {}^{234}\text{Pa} + \beta^- + \overline{\nu}_e$$

Figure 43.6 This earthquake damage to a road was 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. Other radioactive nuclides whose decays heat our planet's interior include ^{235}U, ^{232}Th, and ^{40}K.

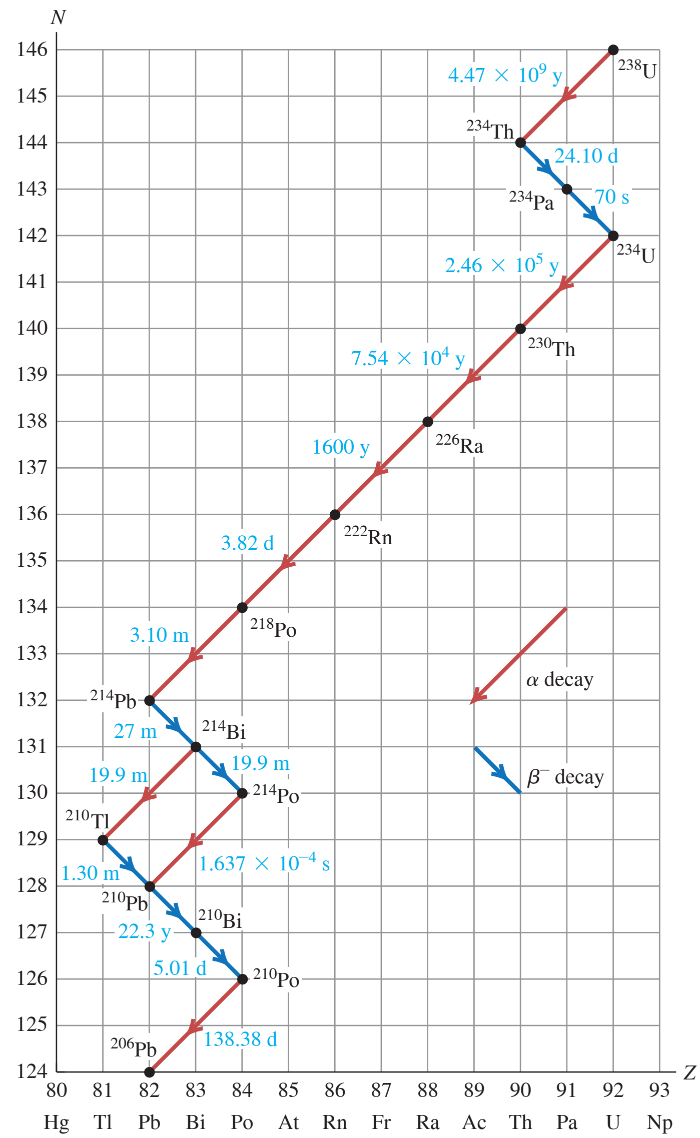


Figure **43.7** Segrè chart showing the uranium ^{238}U decay series, terminating with the stable nuclide ^{206}Pb. The times are half-lives (discussed in the next section), given in years (y), days (d), hours (h), minutes (m), or seconds (s).

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

$$^{234}\mathrm{Pa}^* \longrightarrow {}^{234}\mathrm{Pa} + \gamma \qquad \text{or} \qquad {}^{234}\mathrm{Pa}^* \xrightarrow{\gamma} {}^{234}\mathrm{Pa}$$

Note the branching of the ^{238}U decay series 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. Note also that the series includes unstable isotopes of several elements that also have stable isotopes, including thallium (Tl), lead (Pb), and bismuth (Bi). These unstable isotopes of these elements 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.

ANSWER

(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 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.

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 $t = 0$ — Time — Decay constant

$$N(t) = N_0 e^{-\lambda t} \qquad (43.17)$$

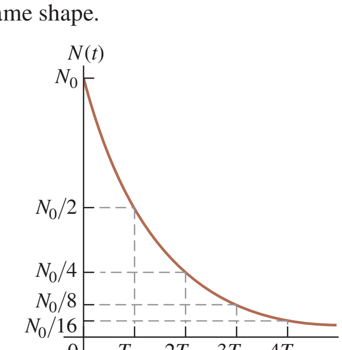
Figure 43.8 is a graph of this function.

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

$$\tfrac{1}{2} = e^{-\lambda T_{1/2}}$$

Figure 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.



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 the time for $N(t)$ to decrease by $1/e$. It is proportional to the half-life $T_{1/2}$:

Lifetime of unstable nucleus or particle; Half-life of nucleus or particle; Decay constant of nucleus or particle

$$\underbrace{T_{\text{mean}}}_{\text{Lifetime of unstable nucleus or particle}} = \frac{1}{\underbrace{\lambda}_{\text{Decay constant of nucleus or particle}}} = \frac{T_{1/2}}{\ln 2} = \frac{T_{1/2}}{0.693} \tag{43.19}$$

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

WITH VARIATION PROBLEMS

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?

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:

$$\begin{aligned} T_{1/2} &= (272 \text{ d})(86{,}400 \text{ s/d}) \\ &= 2.35 \times 10^7 \text{ s} \end{aligned}$$

From Eq. (43.19), the mean lifetime and the decay constant are

$$\begin{aligned} T_{\text{mean}} &= \frac{T_{1/2}}{\ln 2} = \frac{2.35 \times 10^7 \text{ s}}{0.693} \\ &= 3.39 \times 10^7 \text{ s} = 392 \text{ days} \end{aligned}$$

$$\lambda = \frac{1}{T_{\text{mean}}} = 2.95 \times 10^{-8} \text{ s}^{-1}$$

(b) The activity $-dN(t)/dt$ is given as 2.00 μCi, so

$$\begin{aligned} -\frac{dN(t)}{dt} &= 2.00\ \mu\text{Ci} = (2.00 \times 10^{-6})(3.70 \times 10^{10} \text{ s}^{-1}) \\ &= 7.40 \times 10^4 \text{ decays/s} \end{aligned}$$

From Eq. (43.16) this is equal to $\lambda N(t)$, so we find

$$\begin{aligned} N(t) = -\frac{dN(t)/dt}{\lambda} &= \frac{7.40 \times 10^4 \text{ s}^{-1}}{2.95 \times 10^{-8} \text{ s}^{-1}} \\ &= 2.51 \times 10^{12} \text{ nuclei} \end{aligned}$$

If you feel we're being too cavalier about the "units" decays and nuclei, you can use decays/(nucleus · s) as the unit for λ.

(c) From Eq. (43.17) the number $N(t)$ of nuclei remaining after one year (3.156×10^7 s) is

$$N(t) = N_0 e^{-\lambda t} = N_0 e^{-(2.95\times10^{-8}\,\mathrm{s}^{-1})(3.156\times10^{7}\,\mathrm{s})} = 0.394 N_0$$

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 $(0.394)(2.00\ \mu\mathrm{Ci}) = 0.788\ \mu\mathrm{Ci}$.

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 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 $N_0/4$.

KEYCONCEPT Because radioactive decay is a statistical process, the number of radioactive nuclei $N(t)$ present in a radioactive sample and the activity $-dN(t)/dt$ of the sample both decay exponentially with time. Both $N(t)$ and $-dN(t)/dt$ decrease by $\frac{1}{2}$ in one half-life and by $1/e$ in one mean lifetime.

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

WITH VARIATION PROBLEMS

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?

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):

$$T_{1/2} = 5730\ \mathrm{y} = (5730\ \mathrm{y})(3.156 \times 10^7\ \mathrm{s/y}) = 1.808 \times 10^{11}\ \mathrm{s}$$

$$\lambda = \frac{\ln 2}{T_{1/2}} = \frac{0.693}{1.808 \times 10^{11}\ \mathrm{s}} = 3.83 \times 10^{-12}\ \mathrm{s}^{-1}$$

Then, from Eq. (43.16),

$$N(t) = \frac{-dN/dt}{\lambda} = \frac{0.255\ \mathrm{s}^{-1}}{3.83 \times 10^{-12}\ \mathrm{s}^{-1}} = 6.65 \times 10^{10}\ \text{atoms}$$

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.

Continued

(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 $174\ \text{h}^{-1}$. 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.

KEYCONCEPT To find the age of a sample using radioactive dating, measure its present activity and determine what its activity would have been when it formed. The age is proportional to the logarithm of the ratio of these two activities.

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.

ANSWER

(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 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); $1\text{ Gy} = 1\text{ J/kg}$. 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.

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):

$$\text{Equivalent dose (Sv)} = \text{RBE} \times \text{Absorbed dose (Gy)} \tag{43.20}$$

A more common unit, corresponding to the rad, is the rem (an abbreviation of *röntgen equivalent for man*):

$$\text{Equivalent dose (rem)} = \text{RBE} \times \text{Absorbed dose (rad)} \tag{43.21}$$

Thus the unit of the RBE is 1 Sv/Gy or 1 rem/rad, and $1\text{ rem} = 0.01\text{ Sv}$.

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

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?

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 $1\text{ rem} = 0.01\text{ Sv}$ for equivalent dose. Table 43.3 gives the RBE for x rays; we use this

Continued

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.

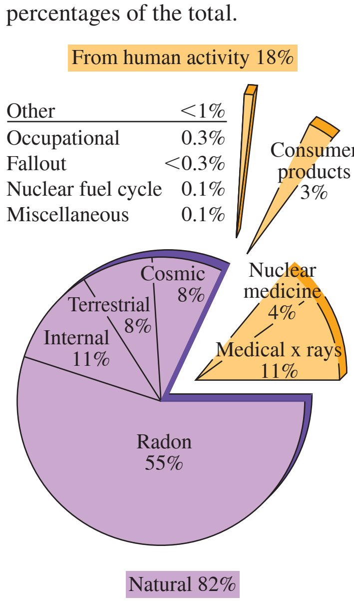
KEYCONCEPT The absorbed dose for a sample of tissue exposed to radiation is the amount of energy delivered per kilogram to the tissue. The equivalent dose equals the absorbed dose multiplied by the relative biological effectiveness (RBE), which depends on the type of radiation used.

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.

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.

Figure 43.9 Contribution of various sources to the total average radiation exposure in the U.S. population, expressed as percentages of the total.



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 $^{131}\mathrm{I}$ and the stable isotope $^{127}\mathrm{I}$. A minute quantity of $^{131}\mathrm{I}$ 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 $^{131}\mathrm{I}$.

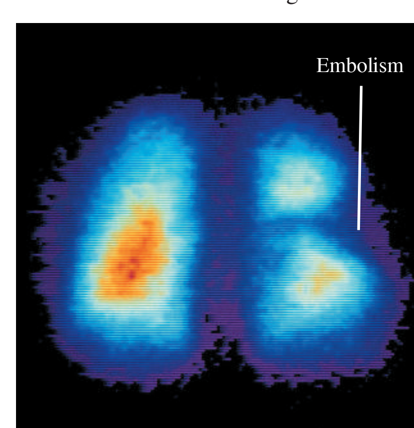
Another useful nuclide for nuclear medicine is technetium-99 $(^{99}\mathrm{Tc})$, which is formed in an excited state by the β^- decay of molybdenum $(^{99}\mathrm{Mo})$. The technetium then decays to its ground state by emitting a γ-ray photon with energy 143 keV. The half-life is 6.01 hours, unusually long for γ emission. (The ground state of $^{99}\mathrm{Tc}$ is also unstable, with a half-life of 2.11×10^5 y; it decays by β^- emission to the stable ruthenium nuclide $^{99}\mathrm{Ru}$.) 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 $^{99}\mathrm{Tc}$ 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 $(^3\mathrm{H})$, 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.

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.

Figure 43.10 This colored scintigram shows where a chemical containing radioactive $^{99}\mathrm{Tc}$ was taken up by a patient's lungs. The orange color in the lung on the left indicates strong γ-ray emission by the $^{99}\mathrm{Tc}$, 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.



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.

ANSWER

(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 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:

$$^4_2\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{17}_{8}\mathrm{O} + {}^1_1\mathrm{H} \tag{43.22}$$

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

$$Q = (M_A + M_B - M_C - M_D)c^2 \quad \text{(reaction energy)} \tag{43.23}$$

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 $^4_2\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{17}_{8}\mathrm{O} + {}^1_1\mathrm{H}$.

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: $^7_3\mathrm{Li}$	7.016003 u	*D:* $^4_2\mathrm{He}$	4.002603 u
	8.023828 u		8.005206 u

The mass decreases by 0.018622 u. From Eq. (43.23), the reaction energy is

$$Q = (0.018622\ \mathrm{u})(931.5\ \mathrm{MeV/u}) = +17.35\ \mathrm{MeV}$$

(b) The initial and final masses are

A: $^4_2\mathrm{He}$	4.002603 u	*C:* $^{17}_{8}\mathrm{O}$	16.999132 u
B: $^{14}_{7}\mathrm{N}$	14.003074 u	*D:* $^1_1\mathrm{H}$	1.007825 u
	18.005677 u		18.006957 u

The mass increases by 0.001280 u, and the corresponding reaction energy is

$$Q = (-0.001280\ \mathrm{u})(931.5\ \mathrm{MeV/u}) = -1.192\ \mathrm{MeV}$$

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.

KEYCONCEPT To find the energy released in a nuclear reaction (the reaction energy), calculate the difference between the total mass of the initial nuclei and the total mass of the final nuclei and then multiply the result by c^2.

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

$$K_{\mathrm{cm}} = \frac{M}{M + m}K \tag{43.24}$$

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, $M = 14.003074\ \mathrm{u}$ and $m = 4.002603\ \mathrm{u}$, so $M/(M + m) = (14.003074\ \mathrm{u})/(18.005677\ \mathrm{u}) = 0.7777$ and $K_{\mathrm{cm}} = 0.7777K$. Since K_{cm} must be at least 1.192 MeV, the α particle's kinetic energy K must be at least $(1.192\ \mathrm{MeV})/0.7777 = 1.533\ \mathrm{MeV}$.

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

$$\begin{aligned} U &= \frac{1}{4\pi\epsilon_0}\frac{(e)(3e)}{r} = (9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2})\frac{(3)(1.6 \times 10^{-19}\ \mathrm{C})^2}{3.5 \times 10^{-15}\ \mathrm{m}} \\ &= 2.0 \times 10^{-13}\ \mathrm{J} = 1.2\ \mathrm{MeV} \end{aligned}$$

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 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.

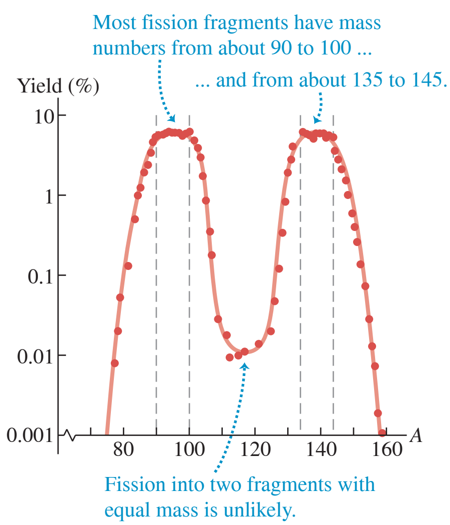
TEST YOUR UNDERSTANDING OF SECTION 43.6 The reaction described in part (a) of Example 43.11 is exoergic. Can it happen naturally when a sample of solid lithium is placed in a flask of hydrogen gas?

ANSWER

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 electric 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 NUCLEAR FISSION

Figure **43.11** Mass distribution of fission fragments from the fission of $^{236}\text{U}^*$ (an excited state of ^{236}U), which is produced when ^{235}U absorbs a neutron. The vertical scale is logarithmic.



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 Strassmann 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 ^3H.

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 $^{236}\text{U}^*$ is in a highly excited state and splits into two fragments almost instantaneously. Strictly speaking, it is $^{236}\text{U}^*$, 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.

Fission Reactions

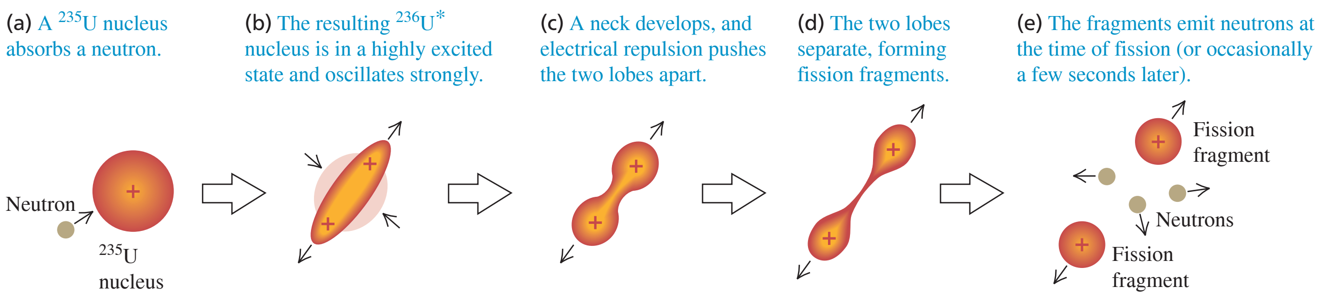
You should check the following two typical fission reactions for conservation of nucleon number and charge:

$$^{235}_{92}\text{U} + ^1_0\text{n} \rightarrow ^{236}_{92}\text{U}^* \rightarrow ^{144}_{56}\text{Ba} + ^{89}_{36}\text{Kr} + 3^1_0\text{n}$$
$$^{235}_{92}\text{U} + ^1_0\text{n} \rightarrow ^{236}_{92}\text{U}^* \rightarrow ^{140}_{54}\text{Xe} + ^{94}_{38}\text{Sr} + 2^1_0\text{n}$$

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.

Figure **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. Physicists have found evidence for at least seven 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.

Figure **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.

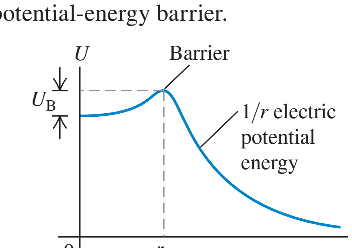
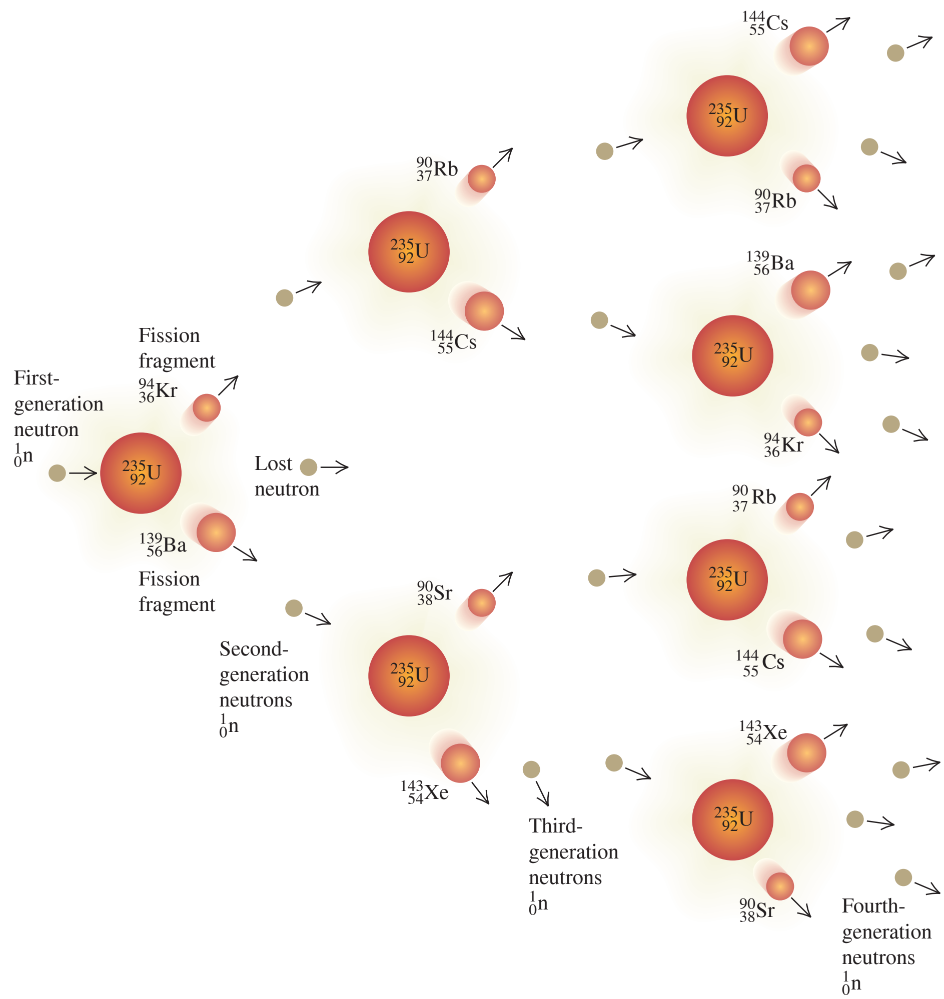


Figure **43.14** Schematic diagram of a nuclear fission chain reaction.



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

$$\mathrm{U} + \mathrm{O_2} \rightarrow \mathrm{UO_2}$$

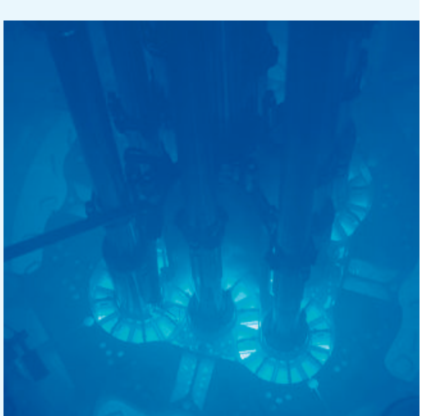
the heat of combustion is about 4500 J/g. Expressed as energy per atom, this is about 11 eV per atom. By contrast, fission liberates about 200 MeV per atom, nearly 20 million times as much energy.

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

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 0.75c). 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.



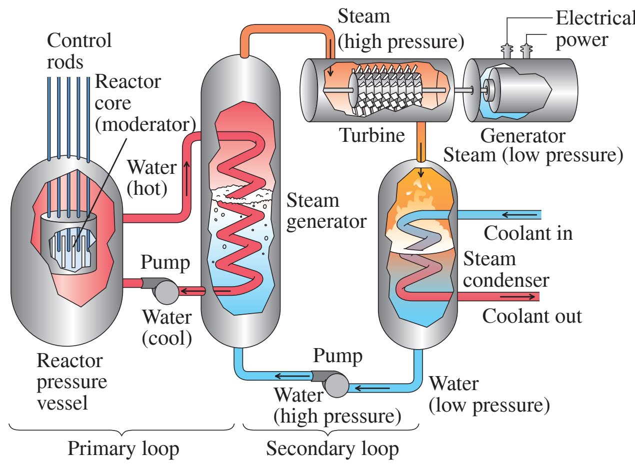


Figure **43.15** Schematic diagram of a nuclear power plant.

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

What mass of ^{235}U must undergo fission each day to provide 3000 MW of thermal power?

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!

KEYCONCEPT The thermal power output of a power plant equals the energy released per reaction multiplied by the number of reactions that take place per second. Since the energy released per nuclear fission reaction is very large, a substantial power output can be obtained from relatively few fission reactions per second.

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?

ANSWER

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 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.

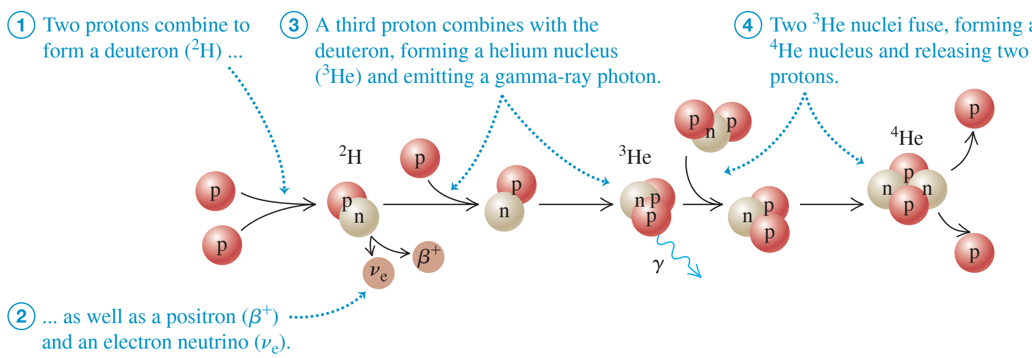
Here are three examples of energy-liberating fusion reactions, written in terms of the neutral atoms:

$$ {}^1_1\text{H} + {}^1_1\text{H} \rightarrow {}^2_1\text{H} + \beta^+ + \nu_e $$
$$ {}^2_1\text{H} + {}^1_1\text{H} \rightarrow {}^3_2\text{He} + \gamma $$
$$ {}^3_2\text{He} + {}^3_2\text{He} \rightarrow {}^4_2\text{He} + {}^1_1\text{H} + {}^1_1\text{H} $$

In the first reaction, two protons combine to form a deuteron $({}^2\text{H})$, 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, ^{3}He, with the emission of a gamma ray. Now double the first two reactions to provide the two ^{3}He nuclei that fuse in the third reaction to form an alpha particle $({}^4\text{He})$ and two protons. Together the reactions make up the process called the *proton-proton chain* (**Fig. 43.16**).

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

Figure **43.16** The proton-proton chain.



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 $4(0.511\text{ MeV}) = 2.044$ MeV of gamma radiation. Thus the total energy released is $(24.69 + 2.044)\text{ MeV} = 26.73$ 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'll 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 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.

Figure **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

Two deuterons fuse to form a *triton* (a nucleus of tritium, or ^{3}H) and a proton. How much energy is liberated?

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

$$Q = [2(2.014102\text{ u}) - 3.016049\text{ u} - 1.007825\text{ u}](931.5\text{ MeV/u})$$
$$= 4.03\text{ MeV}$$

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.

KEYCONCEPT When two light nuclei undergo fusion to form a more massive nucleus, the binding energy per nucleon increases. Hence energy is released in such fusion reactions.

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.

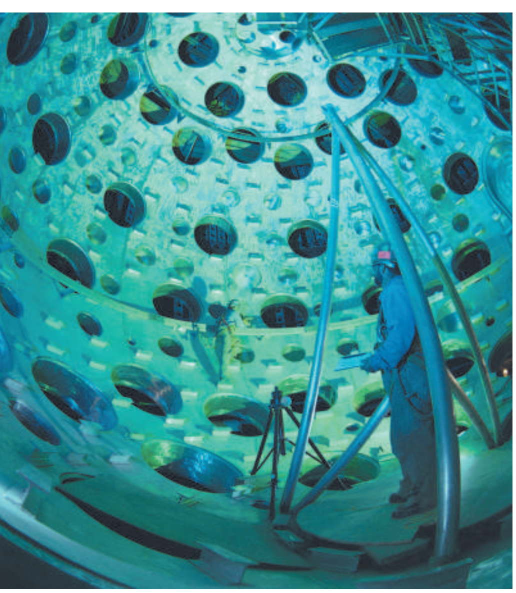
Atoms have this much energy only at extremely high temperatures. The discussion of Section 18.3 showed that the average translational kinetic energy of a gas molecule at temperature T is $\frac{3}{2}kT$, where k is Boltzmann's constant. The temperature at which this is equal to $E = 0.6 \times 10^{-13}$ J is determined by the relationship

$$E = \frac{3}{2}kT$$

$$T = \frac{2E}{3k} = \frac{2(0.6 \times 10^{-13}\ \text{J})}{3(1.38 \times 10^{-23}\ \text{J/K})} = 3 \times 10^{9}\ \text{K}$$

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.

Figure **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.



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

$$^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{1}\text{H} + {}^{1}_{1}\text{H} + 4.0\ \text{MeV} \tag{1}$$

$$^{3}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{0}\text{n} + 17.6\ \text{MeV} \tag{2}$$

$$^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{2}\text{He} + {}^{1}_{0}\text{n} + 3.3\ \text{MeV} \tag{3}$$

$$^{3}_{2}\text{He} + {}^{2}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{1}\text{H} + 18.3\ \text{MeV} \tag{4}$$

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 (^{2}H) and a triton (^{3}H) 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?

ANSWER

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.

CHAPTER 43 SUMMARY

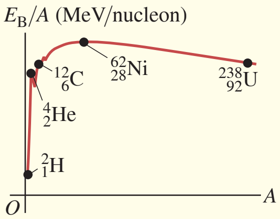
Nuclear properties: A nucleus is composed of A nucleons (Z protons and N neutrons). All nuclei have about the same density. The radius of a nucleus with mass number A is given approximately by Eq. (43.1). A single nuclear species of a given Z and N is called a nuclide. Isotopes are nuclides of the same 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.)

$$R = R_0 A^{1/3} \tag{43.1}$$

$$(R_0 = 1.2 \times 10^{-15}\ \text{m})$$

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.)

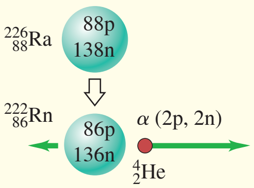
$$E_B = (ZM_H + Nm_n - {}^{A}_{Z}M)c^2 \tag{43.10}$$



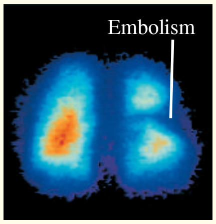
Radioactive decay: Unstable nuclides usually emit an alpha particle (a $^4_2$He nucleus) or a beta particle (an electron) in the process of changing to another nuclide, sometimes followed by a gamma-ray photon. The rate of decay of an unstable nucleus is described by the decay constant λ, the half-life $T_{1/2}$, 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.)

$$N(t) = N_0 e^{-\lambda t} \tag{43.17}$$

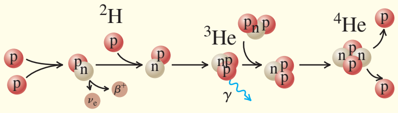
$$T_{\text{mean}} = \frac{1}{\lambda} = \frac{T_{1/2}}{\ln 2} = \frac{T_{1/2}}{0.693} \tag{43.19}$$



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.)



Chapter 43 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLES 43.3 and 43.4 (Section 43.2) before attempting these problems.

VP43.4.1 Find (a) the mass defect, (b) the total binding energy, and (c) the binding energy per nucleon of $^{16}_{8}\mathrm{O}$, which has a neutral atomic mass of 15.994915 u.

VP43.4.2 The mass defect of $^{63}_{29}\mathrm{Cu}$ is 0.5919378 u. Find (a) its binding energy, (b) its binding energy per nucleon, and (c) its neutral atomic mass.

VP43.4.3 Calculate the binding energy E_B/A for the nuclides (a) $^{75}_{33}\mathrm{As}$ (mass defect 0.7005604 u), (b) $^{150}_{62}\mathrm{Sm}$ (mass defect 1.330388 u), and (c) $^{225}_{88}\mathrm{Ra}$ (mass defect 1.852094 u). (d) Figure 43.2 shows that for nuclei with nucleon numbers $A > 62$, the binding energy per nucleon E_B/A decreases as A increases. Do your results agree with this?

VP43.4.4 Use the semiempirical mass formula to calculate the total estimated binding energy E_B and the binding energy per nucleon E_B/A for the nuclides (a) ruthenium-100 ($^{100}_{44}\mathrm{Ru}$) and (b) mercury-200 ($^{200}_{80}\mathrm{Hg}$). (c) Are your results consistent with Fig. 43.2?

Be sure to review EXAMPLES 43.5, 43.6, and 43.7 (Section 43.3) before attempting these problems.

VP43.7.1 Household smoke detectors contain a tiny amount of radioactive americium that decays by α emission to neptunium: $^{241}_{95}\mathrm{Am} \rightarrow {}^{237}_{93}\mathrm{Np} + {}^{4}_{2}\mathrm{He}$. The atomic masses are 241.056827 u for $^{241}_{95}\mathrm{Am}$, 237.048172 u for $^{237}_{93}\mathrm{Np}$, and 4.002603 u for $^{4}_{2}\mathrm{He}$. If the $^{241}_{95}\mathrm{Am}$ nucleus is originally at rest, find (a) the total energy released in the decay, (b) the kinetic energy of the emitted α particle, and (c) the speed of the emitted α particle.

VP43.7.2 The beryllium isotope $^{8}_{4}\mathrm{Be}$ is very unstable, and rapidly decays into two α particles: $^{8}_{4}\mathrm{Be} \rightarrow {}^{4}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He}$. The atomic masses are 8.0053051 u for $^{8}_{4}\mathrm{Be}$ and 4.0026033 u for $^{4}_{2}\mathrm{He}$. If the $^{8}_{4}\mathrm{Be}$ nucleus is originally at rest, find (a) the kinetic energy and (b) the speed of each emitted α particle.

VP43.7.3 State whether each of the following proposed β^- decays is possible or impossible. If it is possible, find the total energy released in the decay. The atomic mass is given for each nuclide. (a) $^{12}_{4}\mathrm{Be}$ (12.026922 u) to $^{12}_{5}\mathrm{B}$ (12.014353 u); (b) $^{33}_{17}\mathrm{Cl}$ (32.977452 u) to $^{33}_{18}\mathrm{Ar}$ (32.989926 u); (c) $^{82}_{35}\mathrm{Br}$ (81.9168018 u) to $^{82}_{36}\mathrm{Kr}$ (81.9134812 u).

VP43.7.4 State whether each of the following proposed β^+ decays is possible or impossible. If it is possible, find the total energy released in the decay; if it is impossible, state whether the decay can occur by electron capture. The atomic mass is given for each nuclide. The combined mass of two electrons is 0.001097 u. (a) $^{46}_{23}\mathrm{V}$ (45.960198 u) to $^{46}_{22}\mathrm{Ti}$ (45.952627 u); (b) $^{134}_{57}\mathrm{La}$ (133.908514 u) to $^{134}_{56}\mathrm{Ba}$ (133.904508 u); (c) $^{67}_{31}\mathrm{Ga}$ (66.9282024u) to $^{67}_{30}\mathrm{Zn}$ (66.9271275 u).

Be sure to review EXAMPLES 43.8 and 43.9 (Section 43.4) before attempting these problems.

VP43.9.1 The sodium isotope $^{24}_{11}\mathrm{Na}$ undergoes β^- decay to $^{24}_{12}\mathrm{Mg}$ (a stable isotope) with a half-life of 15.0 h. Initially a sample of $^{24}_{11}\mathrm{Na}$ has β^- activity of 1.50 μCi. Find (a) the mean lifetime, (b) the decay constant, (c) the initial number of $^{24}_{11}\mathrm{Na}$ nuclei in the sample, and (d) the activity of the sample after 24.0 h.

VP43.9.2 You are given a sample that was initially 100% $^{95}_{41}\mathrm{Nb}$. This isotope undergoes β^- decay to $^{95}_{42}\mathrm{Mo}$ (a stable isotope) with a half-life of 35.0 h. You find that the sample now contains 23.8% $^{95}_{41}\mathrm{Nb}$ and 76.2% $^{95}_{42}\mathrm{Mo}$ and has a β^- activity of 0.514 μCi. Find (a) the decay constant, (b) the present number of $^{95}_{41}\mathrm{Nb}$ nuclei in the sample, (c) the initial number of $^{95}_{41}\mathrm{Nb}$ nuclei in the sample, and (d) how much time has elapsed since the sample was 100% $^{95}_{41}\mathrm{Nb}$.

VP43.9.3 A certain isotope undergoes γ decay from an excited state to its ground state. Initially a sample of this isotope has an activity of 0.425 μCi; 45.0 s later, the activity has decreased to 0.121 μCi. Find (a) the decay constant and half-life of the excited state and (b) the number of nuclei in the excited state initially and 45.0 s later.

VP43.9.4 Before 1900 the activity per unit mass of atmospheric carbon due to the presence of $^{14}\mathrm{C}$ (half-life 5730 y) averaged about 0.255 Bq per gram of carbon. You detect 113 decays of $^{14}\mathrm{C}$ in 20.0 min from a sample of carbon of mass 0.510 g. Assuming that its activity per unit mass of carbon when it formed was the pre-1900 average value for the air, find (a) the age of the sample and (b) the current number of $^{14}\mathrm{C}$ nuclei in the sample.

BRIDGING PROBLEM Saturation of $^{128}\mathrm{I}$ Production

In an experiment, the iodine isotope $^{128}\mathrm{I}$ is created by irradiating a sample of $^{127}\mathrm{I}$ with a beam of neutrons, yielding 1.50×10^6 $^{128}\mathrm{I}$ nuclei per second. Initially no $^{128}\mathrm{I}$ nuclei are present. A $^{128}\mathrm{I}$ nucleus decays by β^- emission with a half-life of 25.0 min. (a) To what nuclide does $^{128}\mathrm{I}$ decay? (b) Could that nuclide decay back to $^{128}\mathrm{I}$ by β^+ emission? Why or why not? (c) After the sample has been irradiated for a long time, what is the maximum number of $^{128}\mathrm{I}$ 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 $^{128}\mathrm{I}$ 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 $^{128}\mathrm{I}$ atoms in the sample, taking account of both the creation of $^{128}\mathrm{I}$ by the neutron irradiation and the decay of any $^{128}\mathrm{I}$ 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 $^{128}\mathrm{I}$. What element is this?
5. Decide whether this nuclide can decay back to $^{128}\mathrm{I}$.
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

•, ••, •••: 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 ${}^{1}_{1}\mathrm{H}$ and ${}^{3}_{2}\mathrm{He}$. 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 ${}^{238}\mathrm{U}$ decay series shown in Fig. 43.7 stop at lead, ${}^{214}\mathrm{Pb}$?

Q43.10 In the ${}^{238}\mathrm{U}$ decay series shown in Fig. 43.7, some nuclides in the series are found much more abundantly in nature than others, even though every ${}^{238}\mathrm{U}$ nucleus goes through every step in the series before finally becoming ${}^{206}\mathrm{Pb}$. 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 ${}^{A}_{Z}\mathrm{El}_{\mathrm{i}}$ represents the initial nuclide, what is the decay process or processes if the final nuclide is (a) ${}_{Z+1}{}^{A}\mathrm{El}_{\mathrm{f}}$; (b) ${}^{A-4}_{Z-2}\mathrm{El}_{\mathrm{f}}$; (c) ${}_{Z-1}{}^{A}\mathrm{El}_{\mathrm{f}}$?

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, ${}^{226}\mathrm{Ra}$, 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, ${}^{28}_{14}\mathrm{Si}$; (b) rubidium, ${}^{85}_{37}\mathrm{Rb}$; (c) thallium, ${}^{205}_{81}\mathrm{Tl}$?

43.2 • 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

43.3 • The most common isotope of boron is ${}^{11}_{5}\mathrm{B}$. (a) Determine the total binding energy of ${}^{11}_{5}\mathrm{B}$ 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 ${}^{11}_{5}\mathrm{B}$ to its accuracy for ${}^{62}_{28}\mathrm{Ni}$ (see Example 43.4).

43.4 •• The nuclei ${}^{11}_{5}\mathrm{B}$ and ${}^{11}_{6}\mathrm{C}$ are called *mirror nuclei,* because the number of protons in ${}^{11}_{5}\mathrm{B}$ equals the number of neutrons in ${}^{11}_{6}\mathrm{C}$ and the number of neutrons in ${}^{11}_{5}\mathrm{B}$ equals the number of protons in ${}^{11}_{6}\mathrm{C}$. The atomic mass of ${}^{11}_{6}\mathrm{C}$ is 11.011434 u, and the atomic mass of ${}^{11}_{5}\mathrm{B}$ is given in Table 43.2. (a) Calculate the binding energies of ${}^{11}_{5}\mathrm{B}$ and ${}^{11}_{6}\mathrm{C}$. (b) Which of the two nuclei has the larger binding energy? Why is this so?

43.5 • Calculate (a) the total binding energy and (b) the binding energy per nucleon of ${}^{12}\mathrm{C}$. (c) What percent of the rest mass of this nucleus is its total binding energy?

43.6 • The most common isotope of uranium, ${}^{238}_{92}\mathrm{U}$, 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 • The binding energy per nucleon for ${}^{56}_{26}\mathrm{Fe}$ is 8.79 MeV/nucleon. What is the mass, in amu, of a neutral ${}^{56}_{26}\mathrm{Fe}$ atom? Give your answer to five significant figures.

43.8 •• An alpha particle is strongly bound. The ${}^{12}_{6}\mathrm{C}$ nucleus might be modeled as a composite of three alpha particles. Compare the binding energy of ${}^{12}_{6}\mathrm{C}$ 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, ${}^{14}_{7}\mathrm{N}$, and (b) the helium nucleus, ${}^{4}_{2}\mathrm{He}$. (c) How does the binding energy per nucleon compare for these two nuclei?

Section 43.3 Nuclear Stability and Radioactivity

43.11 • The carbon isotope $^{11}_{6}\text{C}$ undergoes β^+ (positron) decay. The atomic mass of $^{11}_{6}\text{C}$ is 11.011433 u. (a) How many protons and how many neutrons are in the daughter nucleus produced by this decay? (b) How much energy, in MeV, is released in the decay of one $^{11}_{6}\text{C}$ nucleus?

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 $^{239}_{94}\text{Pu}$; (b) β^- decay of $^{24}_{11}\text{Na}$; (c) β^+ decay of $^{15}_{8}\text{O}$.

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 radioactive isotope $^{12}_{7}\text{N}$ undergoes β^+ decay. The energy released in the decay of one nucleus is 16.316 MeV. What is the mass, in amu, of a neutral $^{12}_{7}\text{N}$ atom? Give your answer to eight significant figures.

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 • Tritium $(^{3}_{1}\text{H})$ 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 ($^{3}_{2}\text{He}$ 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

43.18 • A sample of radioactive nuclei has N_0 nuclei at time $t = 0$. The half-life of the decay is $T_{1/2}$. In terms of N_0, how many decays occur in the time period between $t = 0$ and $t = 0.500T_{1/2}$?

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 isotope $^{90}_{39}\text{Y}$ undergoes β^- decay with a half-life of 64.0 h. You measure an activity of 8.0×10^{16} Bq. (a) How many $^{90}_{39}\text{Y}$ nuclei are present in the sample at the time you make this measurement? (b) How many will be present after 12.0 days?

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 •• **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?

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.

Section 43.5 Biological Effects of Radiation

43.30 • **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)?

43.31 •• **BIO** A nuclear chemist receives an accidental radiation dose of 5.0 Gy from slow neutrons (RBE = 4.0). What does she receive in rad, rem, and J/kg?

43.32 •• **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?

43.33 • **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.

43.34 • **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?

43.35 •• The radioactive isotope $^{90}_{38}\text{Sr}$ is found in the fallout from atmospheric tests of nuclear weapons (see Problem 43.49). It replaces calcium in items such as bone and milk. It has a half-life of 29 years, and the net energy absorbed from each decay averages about 1.1 MeV. (a) Calculate the absorbed dose from 1.0 microgram of $^{90}_{38}\text{Sr}$ in one year if the energy is absorbed in 50 kg of body tissue. (b) The radiation from $^{90}_{38}\text{Sr}$ involves beta-minus particles and gamma rays. Calculate the equivalent dose in part (a) if you assume the RBE for the electrons is 1.0.

43.36 • **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°?

43.37 •• **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

43.38 •• Calculate the reaction energy Q for the reaction $p + {}^3_1\text{H} \rightarrow {}^2_1\text{H} + {}^2_1\text{H}$. Is this reaction exoergic or endoergic?

43.39 • Consider the nuclear reaction

$$^2_1\text{H} + {}^9_4\text{Be} \rightarrow \text{X} + {}^4_2\text{He}$$

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.

43.40 • **Energy from Nuclear Fusion.** Calculate the energy released in the fusion reaction

$$^3_2\text{He} + {}^2_1\text{H} \rightarrow {}^4_2\text{He} + {}^1_1\text{H}$$

43.41 •• 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.

43.42 •• Consider the nuclear reaction

$$^{28}_{14}\text{Si} + \gamma \rightarrow {}^{24}_{12}\text{Mg} + \text{X}$$

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 $^{28}_{14}\text{Si}$ atom is 27.976927 u, and the mass of a $^{24}_{12}\text{Mg}$ atom is 23.985042 u.

43.43 • Consider the nuclear reaction

$$^4_2\text{He} + {}^7_3\text{Li} \rightarrow \text{X} + {}^1_0\text{n}$$

where X is a nuclide. (a) What are Z and A for the nuclide X? (b) Is energy absorbed or liberated? How much?

PROBLEMS

43.44 •• **CP** The nucleus $^{10}_4\text{Be}$ undergoes β^- decay. The atomic mass of $^{10}_4\text{Be}$ is 10.013535 u. (a) How much energy, in MeV, is released in this decay? (b) If you assume that all the energy released in the decay appears as kinetic energy of the β^- particle (a very accurate assumption), what is the speed of the emitted β^-? Note that the energy released is nearly the same as the rest energy of the β^-.

43.45 • **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.)

43.46 •• (a) Calculate the minimum energy required to remove one proton from the nucleus $^{12}_6\text{C}$. This is called the proton-removal energy. (*Hint:* Find the difference between the mass of a $^{12}_6\text{C}$ nucleus and the mass of a proton plus the mass of the nucleus formed when a proton is removed from $^{12}_6\text{C}$.) (b) How does the proton-removal energy for $^{12}_6\text{C}$ compare to the binding energy per nucleon for $^{12}_6\text{C}$, calculated using Eq. (43.10)?

43.47 •• (a) Calculate the minimum energy required to remove one neutron from the nucleus $^{17}_8\text{O}$. This is called the neutron-removal energy. (See Problem 43.46.) (b) How does the neutron-removal energy for $^{17}_8\text{O}$ compare to the binding energy per nucleon for $^{17}_8\text{O}$, calculated using Eq. (43.10)?

43.48 •• Heat flows to the surface of the earth from its interior. Two-thirds of this heat is radiogenic in origin, which means it comes from the decay of radioactive elements in the earth's mantle, primarily from the isotopes ^{238}U and ^{232}Th. We can estimate the earth's internal energy supply due to the decay of uranium as follows: (a) There are 31 μg of ^{238}U in each kg of mantle. Estimate how many ^{238}U isotopes there are per kg of mantle. (b) The half-life of ^{238}U is 4.47 billion years. Calculate the decay constant in units of s^{-1}. (c) Multiply the results in parts (a) and (b) to obtain the activity in each kg of mantle. (d) Each decay chain $^{238}\text{U} \rightarrow {}^{206}\text{Pb}$ releases 52 MeV of energy. Use this value to convert the activity into watts per kg of mantle. (e) The mass of the earth is 6×10^{24} kg, and two-thirds of this is mantle mass. Use these values to estimate the power earth receives from uranium decays. (f) Uranium decays account for 39% of the radiogenic supply of heat. Estimate earth's total heat power related to radioactive decays. (g) Earth also releases heat that remains from its formation, with a power comparable with that due to radioactivity. Estimate the total heat power earth receives from its interior.

43.49 • **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?

43.50 •• **CP** Thorium $^{230}_{90}\text{Th}$ decays to radium $^{226}_{88}\text{Ra}$ by α emission. The masses of the neutral atoms are 230.033134 u for $^{230}_{90}\text{Th}$ and 226.025410 u for $^{226}_{88}\text{Ra}$. 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.)

43.51 •• 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?

43.52 •• The fusion reactions in the sun produce 26.73 MeV per alpha particle created. Each of these reactions produces two neutrinos. In total, the sun's power is about 3.8×10^{26} W. (a) Estimate how many solar fusion reactions take place every second. (b) Estimate the number of neutrinos that leave the sun each second. (c) The distance between the earth and the sun is 150 million km, and the radius of the earth is 6370 km. Using these figures, estimate the number of solar neutrinos that encounter the earth each second. (d) How many square centimeters does the earth present to the sun? (e) Divide your answer in part (c) by your answer in part (d) to obtain an estimate of the number of solar neutrinos that pass through every square centimeter on earth every second.

43.53 •• 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?

43.54 •• 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?

43.55 •• 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.

43.56 •• 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.

43.57 • **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."

43.58 • 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?

43.59 •• 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?

43.60 • 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?

43.61 •• 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?

43.62 ••• The proton-proton chain described in Section 43.8, known as the p-p I chain, accounts for 83.3% of the helium synthesis in the sun. This process uses six protons to produce one alpha particle, two protons, two neutrinos, and a least six gamma-ray photons. We can determine the energies of each of the products as follows: (a) The first reaction, $2\,^1_1H \rightarrow \,^2_1H + \beta^+ + v_e$, is followed immediately by the annihilation of the positron by an electron into two photons, each with energy equal to the rest energy of an electron. What is the remaining energy E_1 carried by the deuteron and the neutrino? (b) Work in the frame where the net momentum of the deuteron and neutrino is zero. Denote the (relativistic) energy of the neutrino by E_v and the (nonrelativistic) deuteron speed by v_d. Denote the mass of the deuteron by m_d. What are the conditions for energy and momentum conservation for these two particles, in terms of E_1? Treat the neutrino as massless. (c) Solve the equations from part (b) and use the deuteron mass in Table 43.2 to determine the percentage of the energy E_1 carried by the neutrino. (d) What is the total energy carried away by neutrinos in the p-p I chain? (*Note:* This chain creates two neutrinos by the above mechanism.) (e) What percentage of the energy generated by the p-p I chain is carried away by neutrinos? (*Note:* The total energy of the p-p I chain is computed in Section 43.8.)

43.63 ••• The sun produces a minority 16.68% of its helium-4 using the p-p II chain, in which four protons and an ambient alpha particle produce two alpha particles, two neutrinos, and at least four gamma-ray photons. The first two steps are the same as the first two steps of the p-p I chain (see Problem 43.62). However, the third step is $^3_2He + \,^4_2He \rightarrow \,^7_4Be + \gamma$, where ^7_4Be is an unstable beryllium nuclide with mass 7.016930 u. (a) What is the energy of the photon? (b) The beryllium nuclide then captures an electron, which changes one of its protons into a neutron, resulting in $^7_4Be + e^- \rightarrow \,^7_3Li + v_e$, where ^7_3Li is the common (and stable) nuclide of lithium, with mass 7.016003 u. The excess energy is carried away by the neutrino. What is the energy of that neutrino? (c) A proton then collides with the lithium nucleus, resulting in two alpha particles via $^1_1H + \,^7_3Li \rightarrow 2\,^4_2He$. How much energy is released in this step? (d) If each alpha particle has the same speed, then in the rest frame of the lithium disintegration, what is the speed of each alpha particle? (e) How much energy is released in total by the processes in this p-p II chain? (f) How does this compare with the energy released in the p-p I chain? (g) What percentage of the energy in the p-p II chain is carried away by neutrinos?

43.64 ••• In addition to the p-p I and p-p II chains (see Problems 43.62 and 43.63), the sun produces helium-4 in a relatively rare reaction called the p-p III chain. This produces only 0.1% of the sun's alpha particles but results in neutrinos with substantially higher energy, which have proved crucial to the study of neutrino physics. In the p-p III chain, an ambient ${}^3_2\text{He}$ fuses with ambient ${}^4_2\text{He}$ (alpha particles) to produce ${}^7_4\text{Be}$, precisely as in the p-p II chain. But in the p-p III chain, the ${}^7_4\text{Be}$ then absorbs a proton to produce ${}^8_5\text{B}$. The ${}^8_5\text{B}$ then decays to ${}^8_4\text{Be}$ along with a positron and a neutrino. (a) The total energy released in this process can be ascertained from its net input and output products $4\,{}^1_1\text{H} \rightarrow {}^4_2\text{He}$. How much energy is released? (b) Given the isotope masses 8.024607 u for ${}^8_5\text{B}$ and 8.0053051 u for ${}^8_4\text{Be}$, determine the energy of the neutrino emitted in the ${}^8_5\text{B}$ to ${}^8_4\text{Be}$ decay. (c) What percentage of the energy in the p-p III chain is carried away by neutrinos? (*Note:* This chain also produces a low-energy neutrino.)

43.65 • A bone fragment found in a cave believed to have been inhabited by early humans contains 0.29 times as much ${}^{14}\text{C}$ 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 ${}^{137}\text{Cs}$ present in the reactor was released. The isotope ${}^{137}\text{Cs}$ 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 ${}^{137}\text{Cs}$ is absorbed by plants, which are eaten by livestock and humans. How many ${}^{137}\text{Cs}$ 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 ${}^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_2\text{He} + {}^1_0\text{n}$. (a) Estimate the barrier energy by calculating the repulsive electrostatic potential energy of the two ${}^2_1\text{H}$ 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 ${}^{59}\text{Fe}$ 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?

MCAT-STYLE 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.)

Key Example √ARIATION Problems

VP43.4.1 (a) 0.137005 u (b) 127.6 MeV (c) 7.976 MeV

VP43.4.2 (a) 551.4 MeV (b) 8.752 MeV (c) 62.929597 u

VP43.4.3 (a) 8.701 MeV (b) 8.262 MeV (c) 7.668 MeV (d) yes

VP43.4.4 (a) $E_B = 868.1$ MeV, $E_B/A = 8.681$ MeV (b) $E_B = 1584$ MeV, $E_B/A = 7.922$ MeV (c) yes

VP43.7.1 (a) 5.637 MeV (b) 5.543 MeV (c) 1.63×10^7 m/s

VP43.7.2 (a) 0.0459 MeV = 45.9 keV (b) 1.49×10^6 m/s

VP43.7.3 (a) possible, 11.71 MeV (b) impossible (c) possible, 3.093 MeV

VP43.7.4 (a) possible, 6.030 MeV (b) possible, 2.710 MeV (c) impossible, yes

VP43.9.1 (a) 7.79×10^4 s (b) $1.28 \times 10^{-5}\ \text{s}^{-1}$ (c) 4.32×10^9 (d) 0.495 μCi

VP43.9.2 (a) $5.50 \times 10^{-6}\ \text{s}^{-1}$ (b) 3.46×10^9 (c) 1.45×10^{10} (d) 72.5 h

VP43.9.3 (a) $\lambda = 0.0279\ \text{s}^{-1}$, $T_{1/2} = 24.8$ s (b) 5.63×10^5 initially, 1.60×10^5 after 45.0 s

VP43.9.4 (a) 2.67×10^3 y (b) 2.46×10^{10}

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 Tarantula Nebula, a region some 170,000 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

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'll 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.

LEARNING OUTCOMES

In this chapter, you'll learn...

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.

You'll need to review...

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.

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 1910–1911 (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

$$ {}^4_2\mathrm{He} + {}^9_4\mathrm{Be} \longrightarrow {}^{12}_6\mathrm{C} + {}^1_0\mathrm{n} \tag{44.1} $$

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

$$ {}^1_0\mathrm{n} + {}^{10}_5\mathrm{B} \longrightarrow {}^7_3\mathrm{Li} + {}^4_2\mathrm{He} \tag{44.2} $$

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.*

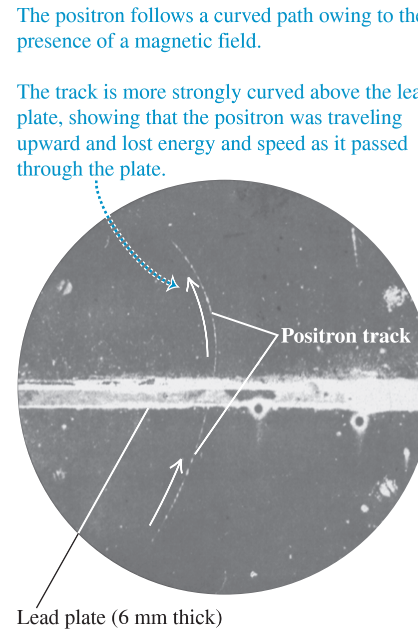
To theorists, the appearance of the positron was a welcome development. In 1928 the English physicist Paul Dirac had developed a relativistic generalization of the Schrödinger equation for the electron. In Section 41.5 we discussed how Dirac's ideas helped explain the spin magnetic moment of the electron.

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.)

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

Figure **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.



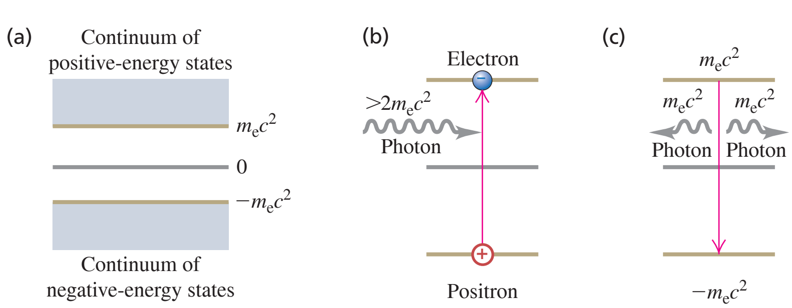
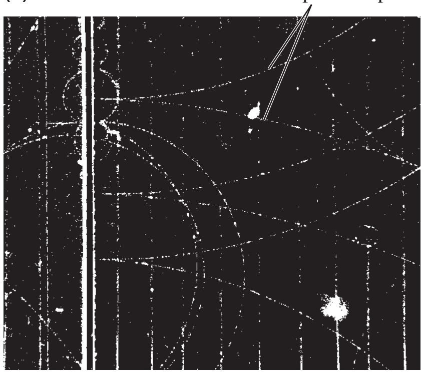


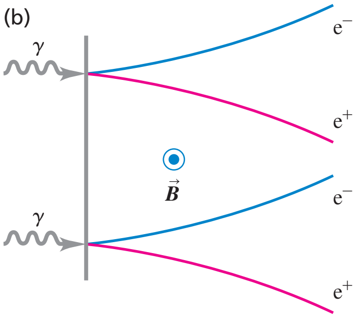
Figure **44.2** (a) Energy states for a free electron predicted by the Dirac equation. (b) Raising an electron from an $E < 0$ state to an $E > 0$ state corresponds to electron–positron pair production. (c) An electron dropping from an $E > 0$ state to a vacant $E < 0$ state corresponds to electron–positron pair annihilation.

Figure 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.

(a)



(b)



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:

$$E_{\min} = 2m_ec^2 = 2(9.109 \times 10^{-31}\ \text{kg})(2.998 \times 10^8\ \text{m/s})^2$$
$$= 1.637 \times 10^{-13}\ \text{J} = 1.022\ \text{MeV}$$

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

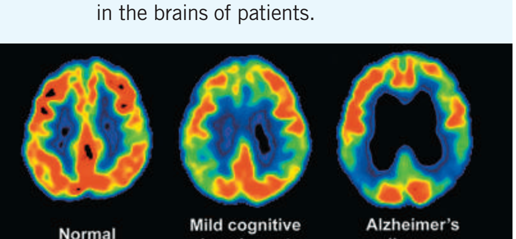
In classical physics we describe the interaction of charged particles in terms of electric and magnetic forces. In quantum mechanics we can describe this interaction in terms of emission and absorption of photons. Two electrons repel each other as one emits a photon and the other absorbs it, just as 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

$$\Delta E\,\Delta t \geq \frac{\hbar}{2} \tag{44.3}$$

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.

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.



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

$$U(r) = -f^2\left(\frac{e^{-r/r_0}}{r}\right) \quad \text{(nuclear potential energy)} \tag{44.4}$$

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:

$$U(r) = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r} \quad \text{(electric potential energy)} \tag{44.5}$$

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 $1.5 \times 10^{-15}\ \text{m} = 1.5\ \text{fm}$. 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:

$$\Delta m = \frac{\Delta E}{c^2} = \frac{2.1 \times 10^{-11}\ \text{J}}{(3.00 \times 10^8\ \text{m/s})^2} = 2.3 \times 10^{-28}\ \text{kg} = 130\ \text{MeV}/c^2$$

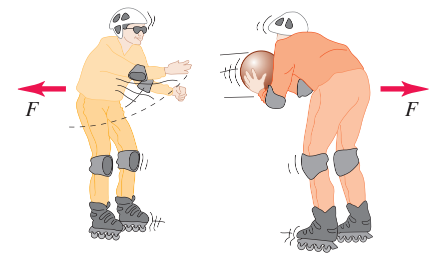
Yukawa postulated that an as yet undiscovered particle with this mass serves as the messenger for the nuclear force.

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 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.

Figure **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.

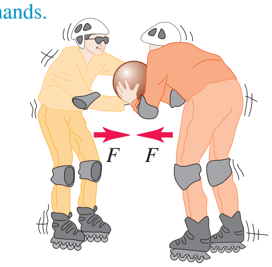
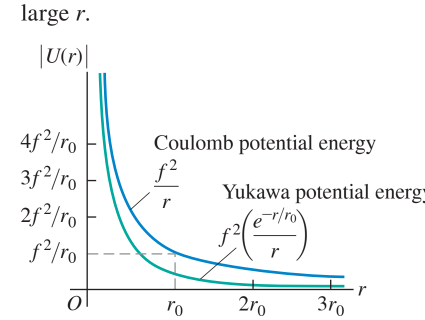


Figure **44.5** Graph of the magnitude of the Yukawa potential-energy function for nuclear forces, $|U(r)| = f^2e^{-r/r_0}/r$. The function $U(r) = f^2/r$, proportional to the potential energy for Coulomb's law, is also shown. The two functions are similar at small r, but the Yukawa potential energy drops off much more quickly at large r.



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$/c^2$.

ANSWER

(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 PARTICLE ACCELERATORS AND DETECTORS

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.

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.

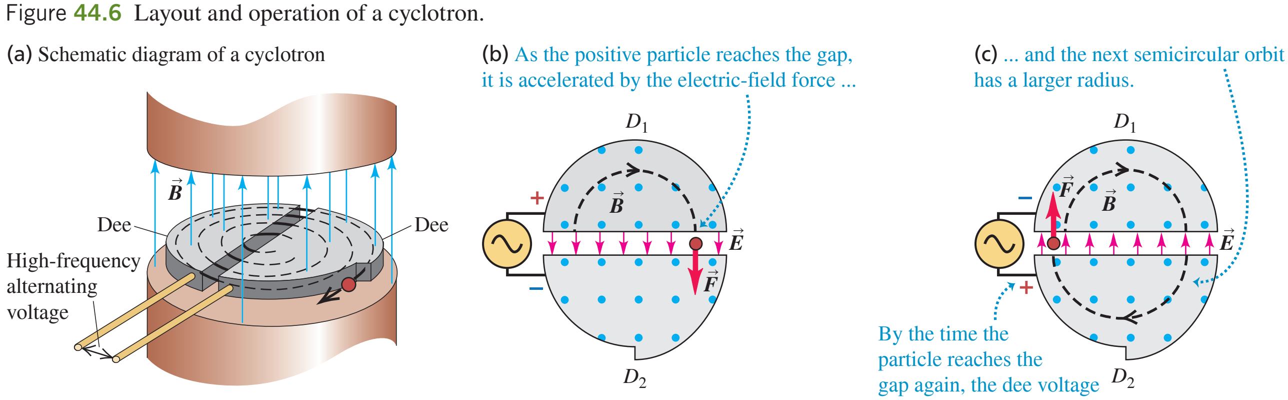
Linear Accelerators

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 charge q move inside a vacuum chamber in a uniform magnetic field $\vec{B}$ that is perpendicular to the plane of

Figure 44.6 Layout and operation of a cyclotron.



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 $v_{\text{max}} = |q|BR/m$. Assuming nonrelativistic speeds, we have

$$K_{\text{max}} = \tfrac{1}{2}mv_{\text{max}}{}^2 = \frac{q^2B^2R^2}{2m} \tag{44.8}$$

EXAMPLE 44.1 Frequency and energy in a proton cyclotron

WITH VARIATION PROBLEMS

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.

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

$$K_{\text{max}} = \frac{(1.60 \times 10^{-19}\ \text{C})^2(1.50\ \text{T})^2(0.50\ \text{m})^2}{2(1.67 \times 10^{-27}\ \text{kg})}$$
$$= 4.3 \times 10^{-12}\ \text{J} = 2.7 \times 10^{7}\ \text{eV} = 27\ \text{MeV}$$

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.

KEYCONCEPT In the type of particle accelerator called a cyclotron, particles follow a circular path in a plane perpendicular to a uniform magnetic field. Every one-half orbit the particles pass through a potential difference that increases their kinetic energy and the radius of their circular path. The maximum radius of the cyclotron path determines the maximum kinetic energy attainable.

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.

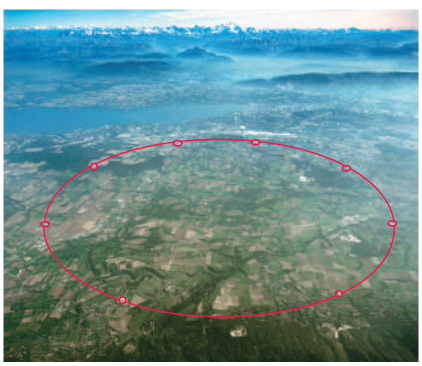
The Synchrotron

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.)

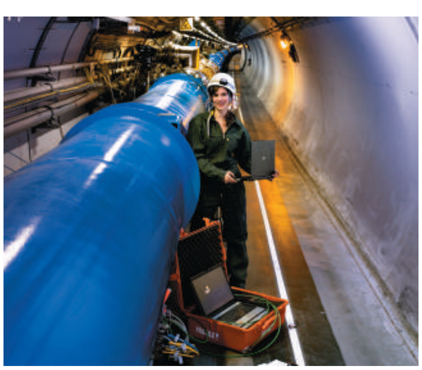
As we pointed out in Section 32.1, accelerated charges radiate electromagnetic energy. In an accelerator in which the particles move in curved paths, this radiation is often called *synchrotron radiation*. High-energy accelerators are typically constructed underground to provide protection from this 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.

Figure 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)



Available Energy

When a beam of high-energy particles collides with a stationary target, not all of the 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 available energy is greatest 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 energy (the sum of the rest energies and kinetic energies of the colliding particles) 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

$$E_a^2 = 2Mc^2E_m + (Mc^2)^2 + (mc^2)^2 \quad \text{(available energy)} \tag{44.9}$$

When the masses of the target and projectile particles are equal, we can simplify:

$$E_a^2 = 2mc^2(E_m + mc^2) \quad \text{(available energy, equal masses)} \tag{44.10}$$

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

$$E_a = \sqrt{2mc^2E_m} \quad \text{(available energy, equal masses, } E_m \gg mc^2\text{)} \tag{44.11}$$

The square root in Eq. (44.11) means that with a stationary target, 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.

CAUTION Available energy includes rest energies Note that the available energy given by Eqs. (44.9), (44.10), and (44.11) is the energy that goes into *all* of the particles that are present after the collision. If the initial particles survive the collision (see Example 44.2), some of the available energy goes back to these survivors; the rest goes to whatever additional particles are produced in the collision.

EXAMPLE 44.2 Threshold energy for pion production

WITH VARIATION PROBLEMS

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?

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:

$$4m^2c^4 + 4mm_\pi c^4 + m_\pi^2 c^4 = 2mc^2 E_m + 2(mc^2)^2$$

$$E_m = mc^2 + m_\pi c^2\left(2 + \frac{m_\pi c^2}{2mc^2}\right) = mc^2 + K$$

$$K = m_\pi c^2\left(2 + \frac{m_\pi c^2}{2mc^2}\right) = (135\text{ MeV})\left(2 + \frac{(135\text{ MeV})}{2(938\text{ MeV})}\right) = 280\text{ MeV}$$

EVALUATE The bombarding proton's kinetic energy must be more than twice the pion rest energy of 135 MeV. By contrast, in Example 37.11 (Section 37.8) 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.

KEYCONCEPT The available energy E_a in a particle collision is the energy available to produce the particles that are present after the collision. This equals the total energy of the colliding particles in the center-of-momentum system, a frame in which the total momentum of the particles is zero. Given E_a, you can calculate the required kinetic energy in a different frame in which one of the particles is initially at rest.

EXAMPLE 44.3 Increasing the available energy

WITH VARIATION PROBLEMS

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?

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 $mc^2 = 938\text{ MeV} = 0.938\text{ GeV}$, 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.

KEYCONCEPT In experiments in which a fast-moving particle strikes a second target particle that is initially stationary, a large increase in the energy of the moving particle produces only a relatively small increase in the available energy.

Colliding Beams

The limitation illustrated by Example 44.3 doesn't apply to *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: $E_a = 2 \times 7\text{ TeV} = 14\text{ TeV}$. 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).

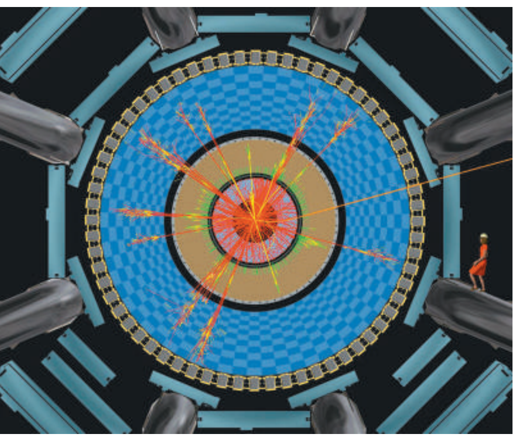


Figure **44.8** This computer-generated image shows the result of a simulated collision between two protons (not shown) in one of the interaction regions at the Large Hadron Collider. The view is along the beampipe. The different color tracks show different types of particles emerging from the collision. A variety of different detectors surround the collision region. (Note the woman in a red dress, drawn for scale.)

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.

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.

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?

ANSWER

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 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'll 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

Figure 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* $\left(s = 2, S = \sqrt{s(s+1)}\,\hbar = \sqrt{6}\,\hbar\right)$. 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 was confirmed in 1983 in experiments at CERN, for which Carlo Rubbia and Simon van der Meer were awarded the Nobel Prize in 1984.

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

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 GeV/c^2 for the W's and 91.2 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 $207m_e = 106$ 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 $273m_e = 140$ 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, $264m_e = 135$ 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* $\overline{p}$ had been suspected ever since the discovery of the positron. The $\overline{p}$ was found in 1955, when proton–antiproton $(p\overline{p})$ pairs were created by use of a beam of 6 GeV protons from the Bevatron at the University of California, Berkeley. The *antineutron* $\overline{n}$ 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'll 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 3×10^{-7} MeV/c^2 = 0.3 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 $3478m_e = 1777$ 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

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) $\pi^0 \rightarrow \mu^- + e^+ + \nu_e$

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: $-1 = 0 + 0 + (-1)$

(b) $\pi^- \rightarrow \mu^- + \overline{\nu}_\mu$

L_e: $0 = 0 + 0$

L_μ: $0 = 1 + (-1)$

Continued

(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.

KEYCONCEPT There are three types of lepton number: electron lepton number L_e, muon lepton number L_μ, and tau lepton number L_τ. Each of these is separately conserved in particle interactions; no violations of these conservation laws have ever been observed.

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 $\overline{p}$. 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'll 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 ($\overline{p}$, $\overline{n}$, $\overline{\Lambda}$, $\overline{\Sigma}$, 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	8.5×10^{-17}	$\gamma\gamma$	$u\overline{u}$, $d\overline{d}$
π^+	139.6	+1	0	0	0	2.60×10^{-8}	$\mu^+\nu_\mu$	$u\overline{d}$
π^-	139.6	−1	0	0	0	2.60×10^{-8}	$\mu^-\overline{\nu}_\mu$	$\overline{u}d$
K^+	493.7	+1	0	0	+1	1.24×10^{-8}	$\mu^+\nu_\mu$	$u\overline{s}$
K^-	493.7	−1	0	0	−1	1.24×10^{-8}	$\mu^-\overline{\nu}_\mu$	$\overline{u}s$
η^0	547.9	0	0	0	0	$\approx 10^{-18}$	$\gamma\gamma$	$u\overline{u}$, $d\overline{d}$, $s\overline{s}$
Baryons								
p	938.3	+1	$\frac{1}{2}$	1	0	Stable	—	uud
n	939.6	0	$\frac{1}{2}$	1	0	880	$pe^-\overline{\nu}_e$	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^-	1322	−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^+	2286	+1	$\frac{1}{2}$	1	0	2.0×10^{-13}	$pK^-\pi^+$	udc

EXAMPLE 44.5 Baryon number conservation

Check conservation of baryon number for these reactions:

(a) $n + p \rightarrow n + p + p + \bar{p}$

(b) $n + p \rightarrow n + p + \bar{n}$

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) $n + p \rightarrow n + p + p + \bar{p}$: $1 + 1 = 1 + 1 + 1 + (-1)$

(b) $n + p \rightarrow n + p + \bar{n}$: $1 + 1 \neq 1 + 1 + (-1)$

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.

KEYCONCEPT Baryon number, like the three lepton numbers, is always conserved in particle interactions.

EXAMPLE 44.6 Antiproton creation

What is the minimum proton energy required to produce an antiproton in a collision with a stationary proton?

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

$$p + p \rightarrow p + p + p + \bar{p}$$

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.

KEYCONCEPT To calculate the minimum energy required to produce a new particle, you must first consider the conservation laws associated with the particle. These laws will tell you which other particles may have to be produced along with the desired one. Once you have done this, you can determine how much available energy is required.

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

$$\Sigma^+ \rightarrow n + \pi^+$$
$$\Lambda^0 \rightarrow p + \pi^-$$
$$K^- \rightarrow \pi^+ + \pi^- + \pi^-$$

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.

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.

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?

ANSWER

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 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.

Figure 44.10 Quark content of four hadrons. The various color combinations that are needed for color neutrality are not shown.

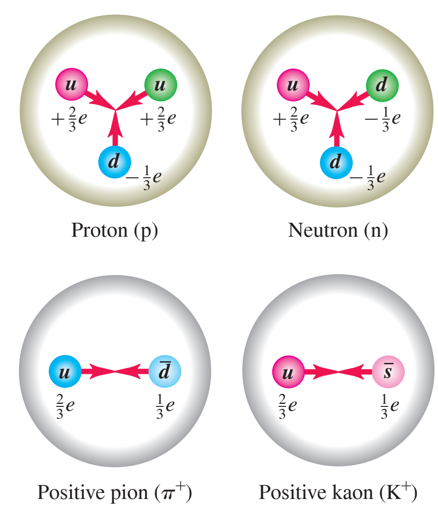


TABLE 44.4 Properties of the Three Original Quarks

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

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 $S = -1$.

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 $S = 1$.

KEYCONCEPT All baryons are composed of three quarks, and all antibaryons are composed of three antiquarks. For a baryon or antibaryon, the total electric charge equals the sum of the charges of its constituent quarks or antiquarks; the same is true of the total baryon number and total strangeness.

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 smaller 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 greater than 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 $\pm 200\ \mathrm{MeV}/c^2$ 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 Yuval Néeman. (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 pattern like pins in a bowling alley. When this pattern was first discovered, one of the particles was missing. But Gell-Mann gave

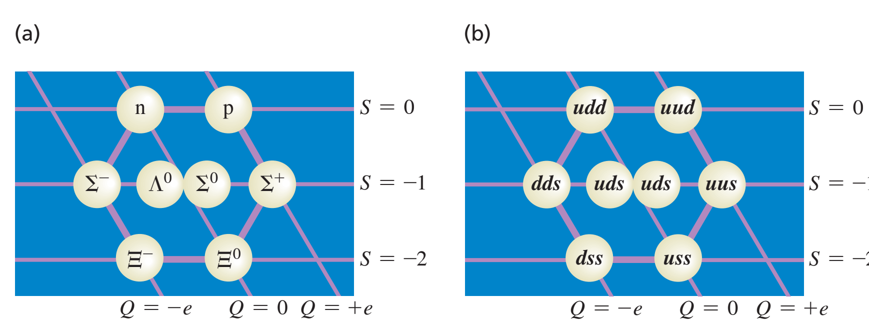


Figure 44.11 (a) Plot of S and Q values for spin-$\frac{1}{2}$ baryons, showing the symmetry pattern of the eightfold way. (b) Quark content of each spin-$\frac{1}{2}$ baryon. The quark contents of the Σ^0 and Λ^0 are the same; the Σ^0 is an excited state of the Λ^0 and can decay into it by photon emission.

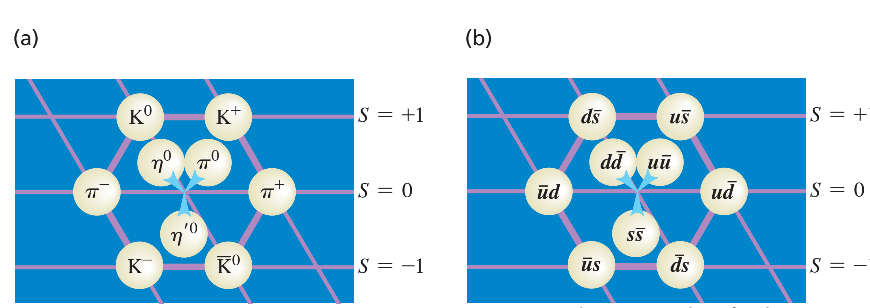


Figure 44.12 (a) Plot of S and Q values for nine spin-0 mesons, showing the symmetry pattern of the eightfold way. Each particle is on the opposite side of the hexagon from its antiparticle; each of the three particles in the center is its own antiparticle. (b) Quark content of each spin-0 meson. The particles in the center are different mixtures of the three quark–antiquark pairs shown.

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.

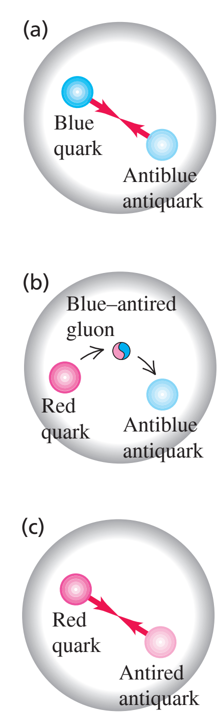
Color

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.

Figure 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 $\boldsymbol{c}$ (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 $3097\ \mathrm{MeV}/c^2$. 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 $\boldsymbol{c\bar{c}}$ system. Almost immediately after this, similar mesons of greater mass were observed and identified as excited states of the $\boldsymbol{c\bar{c}}$ system. A few years later, individual mesons with a nonzero net charm quantum number, D^0 ($\boldsymbol{c\bar{u}}$) and D^+ ($\boldsymbol{c\bar{d}}$), and a charmed baryon, Λ_c^+ ($\boldsymbol{udc}$), were also observed.

In 1977 a meson with mass $9460\ \mathrm{MeV}/c^2$, 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, $\boldsymbol{b}$ (the *bottom* quark), and its antiquark, $\boldsymbol{\bar{b}}$.

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	$+1$

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.

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.

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 $B = 0$.

TEST YOUR UNDERSTANDING OF SECTION 44.4 Is it possible to have a baryon with charge $Q = +e$ and strangeness $S = -2$?

ANSWER

no Only the s quark, with $S = -1$, has nonzero strangeness. For a baryon to have $S = -2$, it must have two s quarks and one quark of a different flavor. Since each 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 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.

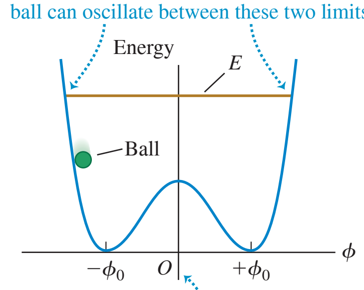
Between 1961 and 1967, Sheldon Glashow, Abdus Salam, and Steven Weinberg developed a theory that unifies the weak and electromagnetic forces. One outcome of their **electroweak theory** is a prediction of the weak-force mediator particles, the $W^{\pm}$ and Z^0

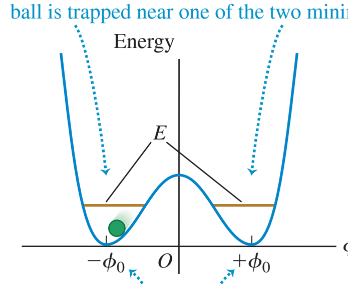
bosons, including their masses. The basic idea is that the mass difference between photons (zero mass) and the weak bosons ($\approx 100\ \mathrm{GeV}/c^2$) makes the electromagnetic and weak interactions behave quite differently at low energies. At sufficiently high energies (well above 100 GeV), however, 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 $\phi_{av} = 0$ (Fig. 44.14a). But at low energies, $\phi(\vec{r}, t)$ oscillates around either a positive average value $\phi_{av} = +\phi_0$ or a negative average value $\phi_{av} = -\phi_0$ (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 $|\phi_{av}|$, 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 $125\ \mathrm{GeV}/c^2$, even greater than the masses of the $W^{\pm}$ and Z^0 weak bosons.

Figure **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

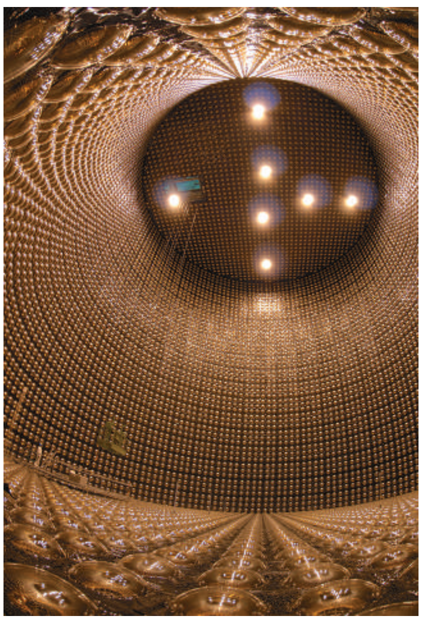
Perhaps at sufficiently high energies the strong interaction and the electroweak interaction have a convergence similar to that between the electromagnetic and weak interactions. If so, they can be unified to give a comprehensive theory of strong, weak, and electromagnetic interactions. Such schemes, 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.

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.

Figure **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.



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 spacetime 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 ***d*** quark can transform into a ***u*** quark, an electron, and an antineutrino by means of the weak interaction. If this happens to a ***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.

ANSWER

(i) If a ***d*** quark in a neutron (quark content ***udd***) undergoes the process $d \rightarrow u + e^- + \bar{\nu}_e$, the remaining baryon has quark content ***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 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.

Figure **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

WITH VARIATION PROBLEMS

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?

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 $z = 0.053$.

KEYCONCEPT The redshift of light from distant galaxies shows that they are moving away from us. The recession speed can be calculated from the ratio of the measured wavelength λ_0 of light from the galaxy to the wavelength λ_S measured in the rest frame of the galaxy.

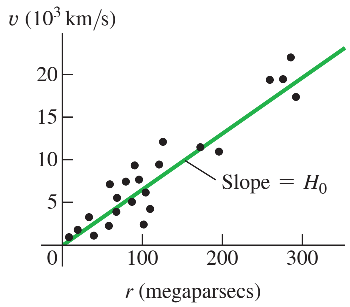


Figure 44.17 Graph of recession speed versus distance for several galaxies. The best-fit straight line illustrates Hubble's law. The slope of the line is the Hubble constant, H_0.

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 about $2.2 \times 10^{-18}\ \text{s}^{-1}$, with an uncertainty of about 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 = 9.46×10^{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:

$$H_0 = (2.2 \times 10^{-18}\ \text{s}^{-1})\left(\frac{9.46 \times 10^{12}\ \text{km}}{1\ \text{ly}}\right)\left(\frac{3.26\ \text{ly}}{1\ \text{pc}}\right)\left(\frac{10^6\ \text{pc}}{1\ \text{Mpc}}\right) = 68\ \frac{\text{km/s}}{\text{Mpc}}$$

EXAMPLE 44.9 Determining distance with the Hubble law

WITH VARIATION PROBLEMS

Use the Hubble law to find the distance from earth to the galaxy in Ursa Major described in Example 44.8.

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 $H_0 = 68\ (\text{km/s})/\text{Mpc} = 6.8 \times 10^4\ (\text{m/s})/\text{Mpc}$,

$$r = \frac{v}{H_0} = \frac{1.55 \times 10^7\ \text{m/s}}{6.8 \times 10^4\ (\text{m/s})/\text{Mpc}} = 230\ \text{Mpc}$$
$$= 2.3 \times 10^8\ \text{pc} = 7.5 \times 10^8\ \text{ly} = 7.1 \times 10^{24}\ \text{m}$$

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.

KEYCONCEPT The Hubble law is the observation that the speed at which a distant galaxy moves away from us is directly proportional to its distance from us. The speed of recession can be determined from the galaxy's redshift, and the distance can then be calculated from the Hubble law.

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

$$t = \frac{r}{v} = \frac{r}{H_0 r} = \frac{1}{H_0} = 4.5 \times 10^{17}\ \text{s} = 1.4 \times 10^{10}\ \text{y}$$

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.

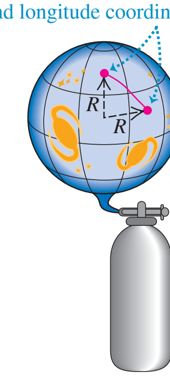
Any length that is measured in intergalactic space is proportional to R, so the wavelength of light traveling to us from a distant galaxy increases along with every other dimension as the universe expands. That is,

$$\frac{\lambda_0}{\lambda} = \frac{R_0}{R} \tag{44.16}$$

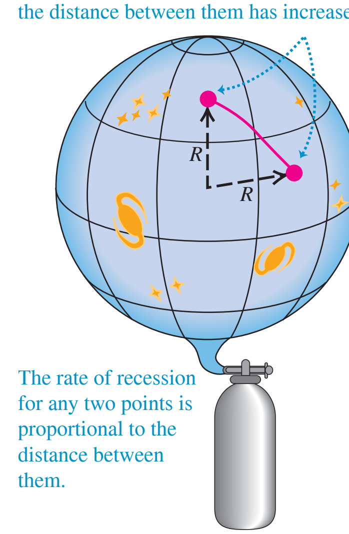
The zero subscripts refer to the values of the wavelength and scale factor *now,* just as H_0 is the current value of the Hubble constant. The quantities λ and R without subscripts are the values at *any* time—past, present, or future. For the galaxy described in Examples 44.8 and 44.9,

Figure 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.



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 12.1, meaning that the volume of space itself is about $(12.1)^3 \approx 1770$ times larger than it was when the light was emitted. Do *not* attempt to substitute $\lambda_0/\lambda_S = 12.1$ 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 $K = \tfrac{1}{2}mv^2$ 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.

The total energy E (kinetic plus potential) for our galaxy is

$$E = \tfrac{1}{2}mv^2 - \frac{GmM}{R} \tag{44.17}$$

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

$$\tfrac{1}{2}mv^2 = \frac{GmM}{R} \tag{44.18}$$

The total mass M inside the sphere is the volume $4\pi R^3/3$ times the density ρ_c:

$$M = \tfrac{4}{3}\pi R^3 \rho_c$$

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

$$\tfrac{1}{2}m(H_0 R)^2 = \frac{Gm}{R}\left(\tfrac{4}{3}\pi R^3 \rho_c\right) \quad \text{or}$$

$$\rho_c = \frac{3{H_0}^2}{8\pi G} \qquad \text{(critical density of the universe)} \tag{44.19}$$

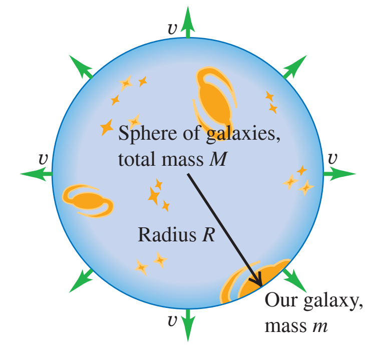
This is the *critical density.* If the average density is less than ρ_c, the universe should continue to expand indefinitely; if it is greater, the universe should eventually stop expanding and begin to contract.

Putting numbers into Eq. (44.19), we find

$$\rho_c = \frac{3(2.2 \times 10^{-18}\ \text{s}^{-1})^2}{8\pi(6.67 \times 10^{-11}\ \text{N}\cdot\text{m}^2/\text{kg}^2)} = 8.7 \times 10^{-27}\ \text{kg/m}^3$$

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.

Figure **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.)



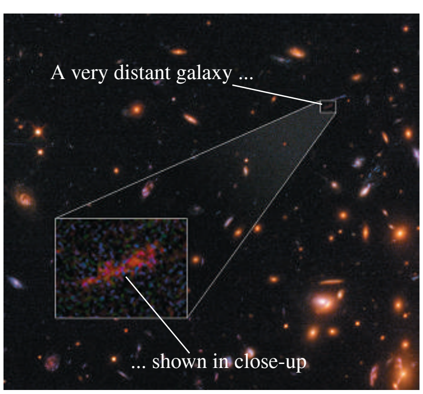
Dark Matter, Dark Energy, and the Accelerating Universe

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 31.0% 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 one-third* kilograms of dark matter.

Figure 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 13.3 billion years ago, when the universe was just 500 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.

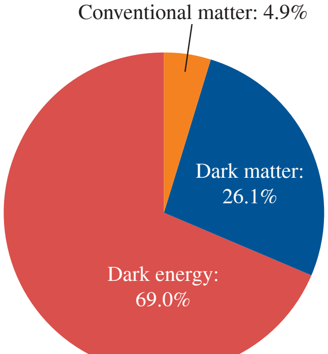


Since the average density of matter in the universe is less than the critical density, it might seem fair to conclude that the universe will continue to expand indefinitely, and that gravitational attraction between matter in different parts of the universe should slow the expansion down (albeit not 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.

Figure 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.



Observations show that the *energy* density of dark energy (measured in, say, joules per cubic meter) is 69.0% of the critical density times c^2; that is, it is equal to $0.690\rho_c c^2$. As described above, the average density of matter of all kinds is 31.0% of the critical density. From the Einstein relationship $E = mc^2$, the average *energy* density of matter in the universe is therefore $0.310\rho_c c^2$. 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 $0.690\rho_c c^2 + 0.310\rho_c c^2 = 1.00\rho_c c^2$. Of this, 69.0% is the mysterious dark energy, 26.1% 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.

TEST YOUR UNDERSTANDING OF SECTION 44.6 Is it accurate to say that your body is made of "ordinary" matter?

ANSWER

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.

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'll 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 ($T = 290$ K) and at the surface of the sun ($T = 5800$ 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?

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 $20(0.0375\ \text{eV}) = 0.75$ 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 ($E = 0.511$ MeV) and proton ($E = 938$ 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'll see, temperatures were this high in the very early universe.

KEYCONCEPT The average kinetic energy of particles in a gas is directly proportional to the absolute temperature T. At extremely high temperatures above 10^{10} K, this average kinetic energy can be comparable to the rest energies of subatomic particles.

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 $100\ \mathrm{GeV}/c^2$. 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:

$$l_{\mathrm{P}} = \sqrt{\frac{\hbar G}{c^3}} = 1.616 \times 10^{-35}\ \mathrm{m} \tag{44.21}$$

You should verify that this combination of constants has units of length. The *Planck time* $t_{\mathrm{P}} = l_{\mathrm{P}}/c$ is the time required for light to travel a distance l_{P}:

$$t_{\mathrm{P}} = \frac{l_{\mathrm{P}}}{c} = \sqrt{\frac{\hbar G}{c^5}} = 0.539 \times 10^{-43}\ \mathrm{s} \tag{44.22}$$

If we mentally go backward in time toward the Big Bang at $t = 0$, 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 on scales smaller 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 1514–1515 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

$$E \approx (10^{-13}\ \mathrm{GeV/K})(10^{32}\ \mathrm{K}) = 10^{19}\ \mathrm{GeV}$$

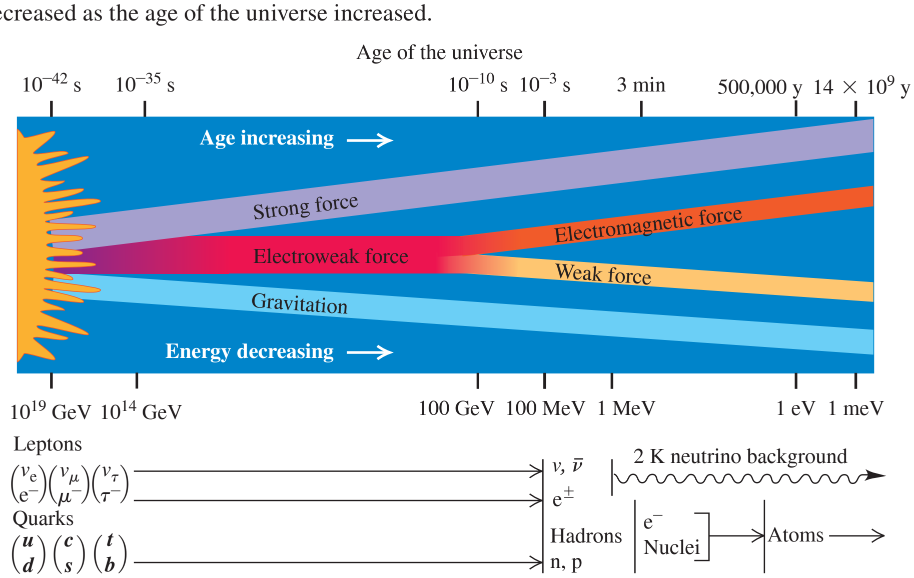
In a totally unified theory this is about the energy below which gravity begins to behave as a separate interaction. This time therefore marked the transition from any proposed TOE to the GUT period.

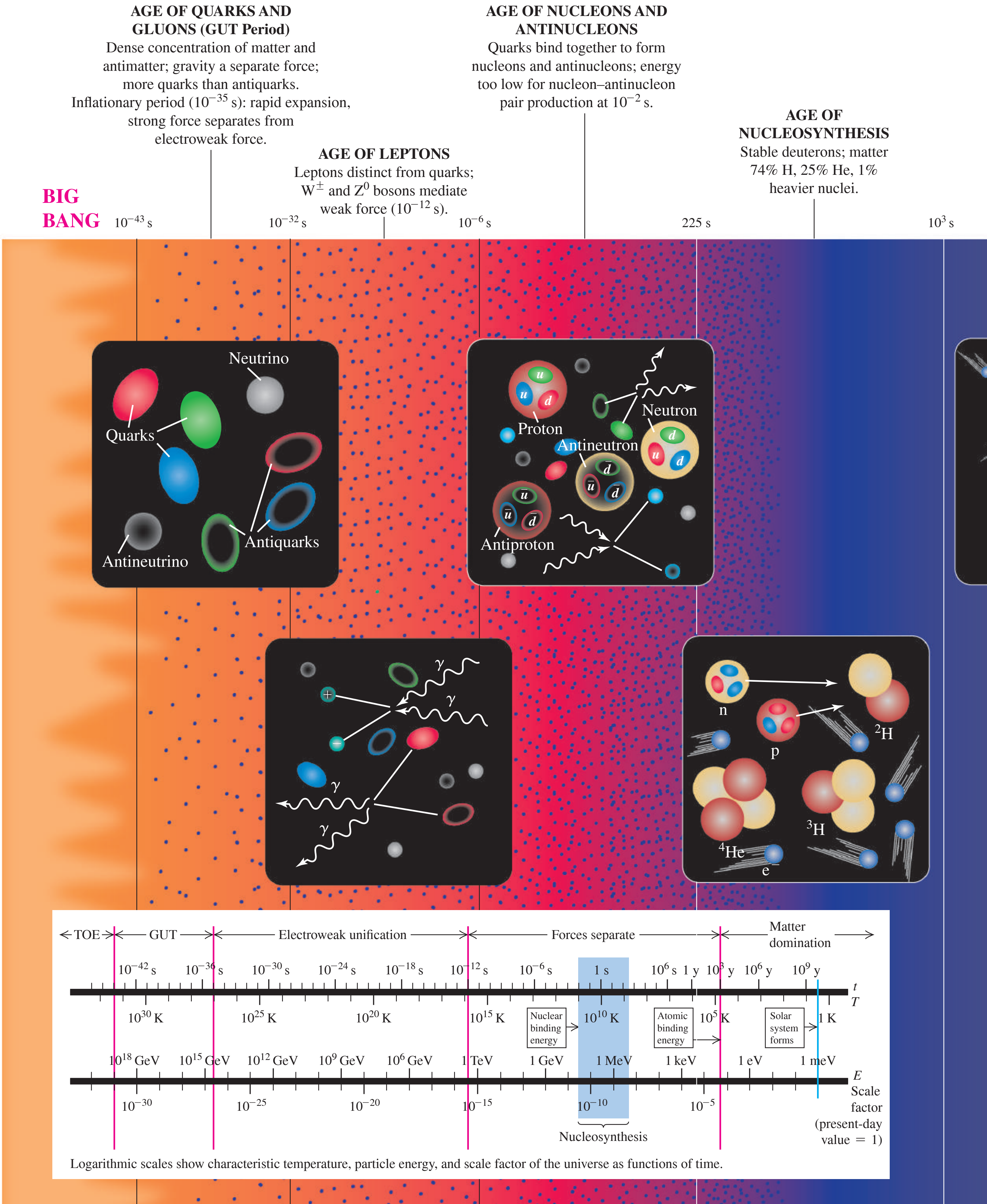
During the GUT period, roughly $t = 10^{-43}$ to 10^{-35} s, the strong and electroweak forces were still unified, and the universe consisted of a soup of quarks and leptons transforming into each other so freely that there was no distinction between the two families of particles. Other, much more massive 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 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.

Figure **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
Antineutrino
Antiquarks
Proton
Neutron
Antineutron
Antiproton
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.

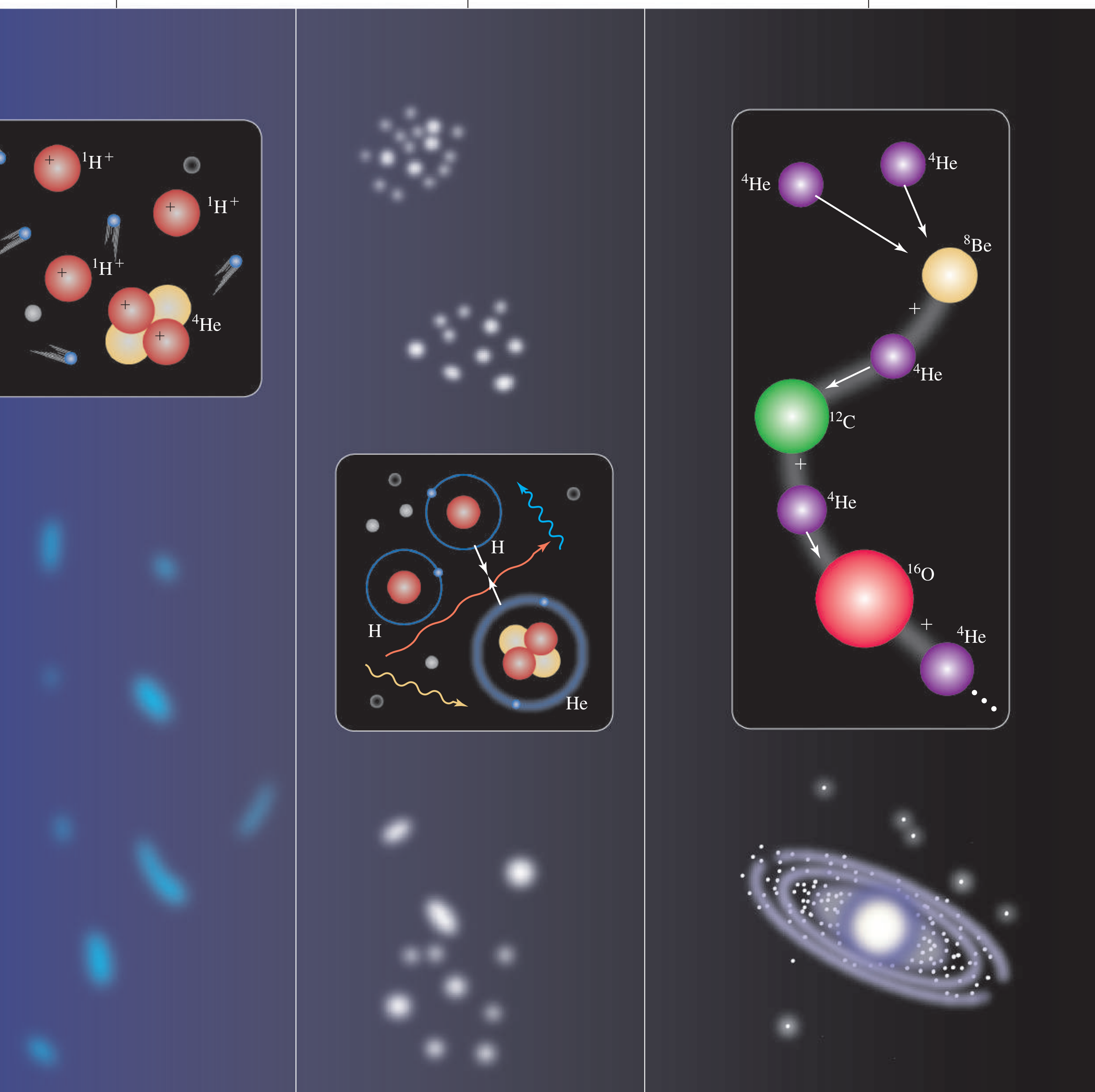
10^{13} s

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



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 average was critical because the binding energy of the *deuteron* (a neutron and a proton combined to form a ^2H nucleus) 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 ^7Li nuclei may also have formed by fusion of ^3H and ^4He nuclei. According to the theory, essentially all the ^1H and ^4He 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 ^8Be that is formed by fusion of two ^4He 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

Nearly all of the protons and neutrons in the seven-to-one ratio at $t = 225$ s either formed ^4He or remained as ^1H. After this time, what was the resulting relative abundance of ^1H and ^4He, by mass?

SOLUTION The ^4He 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 ^4He nucleus, leaving 12 protons (^1H nuclei). So there were eventually 12 ^1H nuclei for every ^4He nucleus. The masses of ^1H and ^4He are about 1 u and 4 u, respectively, so there were 12 u of ^1H for every 4 u of ^4He. Therefore the relative abundance, by mass, was 75% ^1H and 25% ^4He. 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.

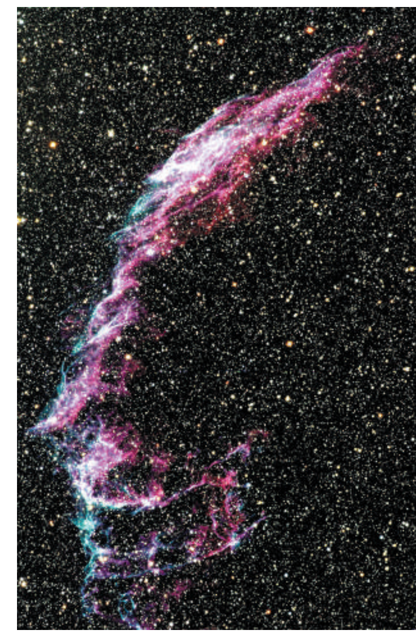
KEYCONCEPT The first atomic nuclei formed about three and a half minutes after the Big Bang. The present-day ratio of hydrogen nuclei to helium nuclei in the universe was determined by the ratio of protons to neutrons at that moment in the history of the universe.

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.

Figure **44.23** The Veil Nebula in the constellation Cygnus is a remnant of a supernova explosion that occurred about 8000 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.



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 a TV that's disconnected from cable and 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.

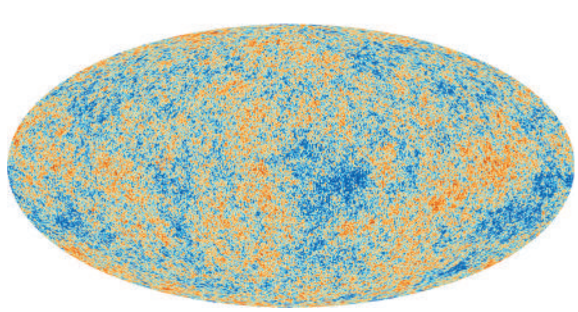


Figure **44.24** This false-color map shows microwave radiation from the entire sky mapped onto an oval. When this radiation was emitted 380,000 years after the Big Bang, the regions shown in blue were slightly cooler and denser than average. Within these cool, dense regions formed galaxies, including the Milky Way galaxy of which our solar system, our earth, and our selves are part.

EXAMPLE 44.12 Expansion of the universe

By approximately what factor has the universe expanded since $t = 380{,}000$ y?

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.

KEYCONCEPT The present-day temperature of the cosmic background radiation tells us how much the universe has expanded since the time 380,000 years after the Big Bang when the first atoms formed and the universe became transparent.

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?

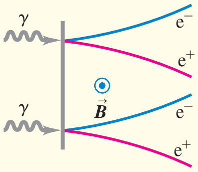
ANSWER

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.

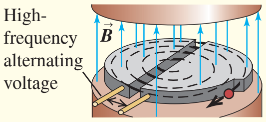
CHAPTER 44 SUMMARY

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.



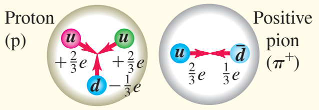
Particle accelerators and detectors: Cyclotrons, synchrotrons, and linear accelerators are used to accelerate charged particles to high energies for experiments with particle interactions. Only part of the beam energy is available to cause reactions with targets at rest. This problem is avoided in colliding-beam experiments. (See Examples 44.1–44.3.)



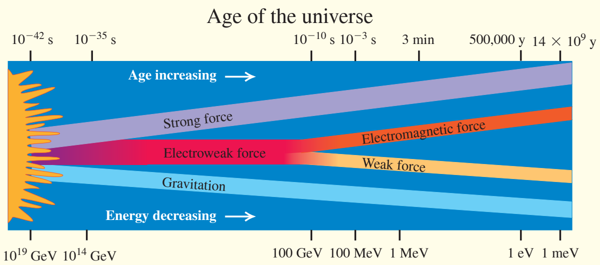
Particles and interactions: Four fundamental interactions are found in nature: the strong, electromagnetic, weak, and gravitational interactions. Particles can be described in terms of their interactions and of quantities that are conserved in all or some of the interactions.

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 charm, are conserved in some interactions. (See Examples 44.4–44.6.)

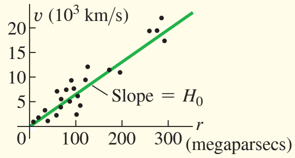
Quarks: Hadrons are composed of quarks. There are thought to be six types of quarks. The interaction between quarks is mediated by gluons. Quarks and gluons have an additional attribute called color. (See Example 44.7.)



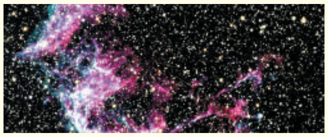
Symmetry and the unification of interactions: Symmetry considerations play a central role in all fundamental-particle theories. The electromagnetic and weak interactions become unified at high energies into the electroweak interaction. In grand unified theories the strong interaction is also unified 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 69.0% of the energy in the universe. Only 4.9% of the energy in the universe is in the form of conventional matter; the remaining 26.1% 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.)



Chapter 44 Media Assets

GUIDED PRACTICE

For assigned homework and other learning materials, go to Mastering Physics.

KEY EXAMPLE VARIATION PROBLEMS

Be sure to review EXAMPLE 44.1 (Section 44.2) before attempting these problems.

VP44.1.1 In a cyclotron the maximum radius for the path of protons is 0.800 m. The magnetic field is 0.600 T. Find (a) the frequency of the alternating voltage applied to the dees, (b) the maximum proton kinetic energy in MeV, and (c) the maximum proton speed in terms of c.

VP44.1.2 A cyclotron at the TRIUMF laboratory in Vancouver, Canada, has a magnetic field of 0.460 T. It accelerates negative hydrogen ions with charge $-e$ and mass 1.67×10^{-27} kg. Find (a) the radius of the path of the ions when they have speed 6.00×10^6 m/s and (b) the angular frequency of the ions' motion.

VP44.1.3 The first cyclotron (in 1930) had a diameter of 11.4 cm and was used to accelerate negative hydrogen ions (charge $-e$, mass 1.67×10^{-27} kg) to kinetic energies of 80.0 keV. Find (a) the magnetic field used and (b) the frequency of the alternating voltage applied to the dees.

VP44.1.4 A cyclotron is used to accelerate alpha particles (charge $+2e$, mass 6.64×10^{-27} kg). The frequency of the alternating voltage applied to the dees is 5.80 MHz. Find (a) the magnetic field used and (b) the kinetic energy, in MeV, of the alpha particles when the radius of their circular path is 0.650 m.

Be sure to review EXAMPLES 44.2 and 44.3 (Section 44.2) before attempting these problems.

VP44.3.1 A proton with kinetic energy K collides with a proton at rest. Both protons remain after the collision along with an eta-prime (η') particle that has rest energy 958 MeV. In order for this reaction to happen, what must be (a) the minimum available energy in the center-of-momentum system and (b) the minimum value of K?

VP44.3.2 A proton with kinetic energy K collides with a proton at rest. The particles that remain after the collision are one proton and a delta-plus (Δ^+) that has rest energy 1232 MeV. In order for this reaction to happen, what must be (a) the minimum available energy in the center-of-momentum system and (b) the minimum value of K?

VP44.3.3 A negative pion (π^-) that has rest energy 140 MeV and kinetic energy K collides with a proton at rest. What remains after the collision is a delta-zero (Δ^0) that has rest energy 1232 MeV. In order for this reaction to happen, what must be (a) the minimum available energy in the center-of-momentum system and (b) the minimum value of K?

VP44.3.4 The Large Hadron Collider gives each proton a total energy of 7 TeV $= 7.00 \times 10^3$ GeV. (a) If a proton that has this energy collides with a second proton that is stationary, what is the available energy? (b) After the collision, both protons remain as well as a neutral particle X. What is the maximum mass of X?

Be sure to review EXAMPLES 44.8 and 44.9 (Section 44.6) before attempting these problems.

VP44.9.1 The red spectral line of hydrogen has a wavelength of 656.3 nm when the hydrogen is at rest. This line is observed to have a wavelength of 725 nm in the spectrum of a galaxy. Find (a) the redshift z and (b) the recession speed of the galaxy in terms of c.

VP44.9.2 A prominent absorption line in the spectrum of the sun (see Fig. 39.9) and many galaxies is due to singly ionized calcium. In the rest frame, this line has an ultraviolet wavelength of 396.9 nm. If a galaxy is receding from us at $0.0711c$, find (a) the wavelength at which we'll observe this line and (b) the redshift z.

VP44.9.3 Find the distances to (a) a galaxy with a recession speed of $0.0992c$ and (b) a galaxy with a recession speed of $0.0711c$. Give your answers in pc.

VP44.9.4 The spectrum of a Type Ia supernova (an exploding star so luminous that it can be seen even in distant galaxies) has an emission line due to singly ionized silicon. If astronomers observe this emission line at 615.0 nm in the laboratory but 666 nm in the spectrum of a Type Ia supernova in a distant galaxy, find (a) the recession speed of the galaxy in m/s and (b) the distance to the galaxy in pc.

BRIDGING PROBLEM Hyperons, Pions, and the Expanding Universe

(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

•, ••, •••: 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 • 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?

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.

Section 44.2 Particle Accelerators and Detectors

44.7 • In a collision experiment, a proton at rest is struck by an antiproton. (a) What is the minimum kinetic energy of the antiproton if the available energy is 2.00 TeV? (b) If a colliding beam is used instead of a stationary target, what minimum kinetic energy for each beam is required for an available energy of 2.00 TeV?

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 • Bubble-chamber photographs taken with a 0.40 T magnetic field show at one point a positron moving perpendicular to the field in a 15 cm circle. What is the magnitude of the positron's momentum at that point?

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 • The first cyclotron used a 1.3 T magnetic field and had an 11 cm radius. (a) Find the maximum kinetic energy of protons accelerated by this cyclotron. Is it accurate to use nonrelativistic expressions in your calculation? (b) When the cyclotron was accelerating protons, at what frequency was it operated?

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 $\text{p} + \text{p} \rightarrow \text{p} + \text{p} + \eta^0$ reaction. The rest energy of the η^0 is 547.9 MeV (see Table 44.3).

44.15 •• 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 $\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \mathrm{p} + \mathrm{p} + \overline{\mathrm{p}}$ 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

44.16 • In about 8% of Ω^- decays, the decay products are a Ξ^- and a π^0. (a) What is the energy released in this decay? (b) In the decay, what is the change in baryon number, and what is the change in strangeness? Your results should show that this decay is allowed for the weak nuclear interaction but not for the strong interaction.

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 • 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?

44.21 • 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 • Which of the following reactions obey the conservation of baryon number? (a) $\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \mathrm{e}^+$; (b) $\mathrm{p} + \mathrm{n} \rightarrow 2\mathrm{e}^+ + \mathrm{e}^-$; (c) $\mathrm{p} \rightarrow \mathrm{n} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}$; (d) $\mathrm{p} + \overline{\mathrm{p}} \rightarrow 2\gamma$.

44.23 • In which of the following decays are the three lepton numbers conserved? In each case, explain your reasoning. (a) $\mu^- \rightarrow \mathrm{e}^- + \nu_\mathrm{e} + \overline{\nu}_\mu$; (b) $\tau^- \rightarrow \mathrm{e}^- + \overline{\nu}_\mathrm{e} + \nu_\tau$; (c) $\pi^+ \rightarrow \mathrm{e}^+ + \gamma$; (d) $\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}$.

44.24 • In which of the following reactions or decays is strangeness conserved? In each case, explain your reasoning. (a) $\mathrm{K}^+ \rightarrow \mu^+ + \nu_\mu$; (b) $\mathrm{n} + \mathrm{K}^+ \rightarrow \mathrm{p} + \pi^0$; (c) $\mathrm{K}^+ + \mathrm{K}^- \rightarrow \pi^0 + \pi^0$; (d) $\mathrm{p} + \mathrm{K}^- \rightarrow \Lambda^0 + \pi^0$.

Section 44.4 Quarks and Gluons

44.25 • 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.

44.26 •• 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 • 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.

44.28 • What is the total kinetic energy of the decay products when an upsilon particle at rest decays to $\tau^+ + \tau^-$?

Section 44.5 The Standard Model and Beyond

44.29 • Section 44.5 states that current experiments show that the mass of the Higgs boson is about 125 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

44.30 • 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.

44.31 • 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 $\lambda_0/\lambda_\mathrm{S}$ is expected for light from this galaxy?

44.32 • **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

44.33 • Calculate the reaction energy Q (in MeV) for the reaction $\mathrm{e}^- + \mathrm{p} \rightarrow \mathrm{n} + \nu_\mathrm{e}$. Is this reaction endoergic or exoergic?

44.34 • Calculate the energy (in MeV) released in the triple-alpha process $3\,{}^4\mathrm{He} \rightarrow {}^{12}\mathrm{C}$.

44.35 • Calculate the reaction energy Q (in MeV) for the nucleosynthesis reaction

$$ {}^{12}_{6}\mathrm{C} + {}^{4}_{2}\mathrm{He} \rightarrow {}^{16}_{8}\mathrm{O} $$

Is this reaction endoergic or exoergic?

PROBLEMS

44.36 •• The strong nuclear force can be crudely modeled as a Hooke's-law spring force, increasing linearly with the quark separation distance. The energy stored in this "spring" corresponds to the energy content of the gluon field. In this picture, as quarks are further separated, increasing energy is stored between them. At a critical separation distance, the energy converts to matter, and a new quark–antiquark pair is generated, guaranteeing that there can never be a free quark. (a) A proton has a diameter of about 1.5 fm. Estimate the repulsive Coulomb force between two up quarks separated by 0.5 fm. (b) Model the strong force as $F_\mathrm{s} = ks$, where s is the distance between two quarks. If this force balances the electrostatic repulsion between two up quarks when $s = 0.5$ fm, what is the effective spring constant k, in SI units? (c) Convert k into units of $\mathrm{MeV/fm^2}$. (d) How much energy is stored in the gluon field when $s = 0.5$ fm? The mass of an up quark is thought to be about 2.3 MeV/c^2. (e) How much energy is needed to produce an up quark and an antiup quark? (f) How far would two up quarks need to be separated so that the gluon energy $\frac{1}{2}ks^2$ matches the rest energy of an up–antiup quark pair?

44.37 •• 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.

44.38 •• 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.

44.39 •• The 7 TeV proton bunches that circulate in opposite directions around the 27 km ring of the Large Hadron Collider smash together at beam crossings every 25 ns. (a) How many bunches meet every second? (b) Each bunch has 115 billion protons. Of these, typically 20 collide during each crossing. Estimate the fraction of protons that collide per crossing. (c) Estimate how many collisions take place each second. (d) The bunches are 30 cm long and are squeezed to a diameter of 20 μm. Estimate the density of protons in a bunch, in units of protons/mm^3. (e) Estimate the density of hadrons in ordinary matter. (*Hint:* Divide your mass, which is mostly due to hadrons, by the mass of a proton to get the number of hadrons in your body. Then divide by the estimated volume.)

44.40 •• Calculate the threshold kinetic energy for the reaction $\pi^- + \text{p} \rightarrow \Sigma^0 + \text{K}^0$ if a π^- beam is incident on a stationary proton target. The K^0 has a mass of 497.7 MeV/c^2.

44.41 • 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) $\text{p} + \text{p} \rightarrow \text{p} + \Lambda^0 + ?$; (b) $\text{K}^- + \text{n} \rightarrow \Lambda^0 + ?$; (c) $\text{p} + \overline{\text{p}} \rightarrow \text{n} + ?$; (d) $\overline{\nu}_\mu + \text{p} \rightarrow \text{n} + ?$

44.42 •• 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.

44.43 • 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.

44.44 • 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?

44.45 •• **CP BIO** One proposed proton decay is $\text{p}^+ \rightarrow \text{e}^+ + \pi^0$, 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?

44.46 •• A ϕ meson (see Problem 44.43) at rest decays via $\phi \rightarrow \text{K}^+ + \text{K}^-$. 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 $\phi \rightarrow \text{K}^+ + \text{K}^- + \pi^0$ has not been observed. (c) Suggest reasons the decays $\phi \rightarrow \text{K}^+ + \pi^-$ and $\phi \rightarrow \text{K}^+ + \mu^-$ have not been observed.

44.47 ••• **CP** About 10,000 cosmic-ray protons, each with hundreds of MeV of energy, impinge on each square meter of our upper atmosphere each second. They collide with atmospheric nitrogen and oxygen to produce secondary showers of newly created particles, including many muons. Muons have a mass of 105.7 MeV/c^2 and an average lifetime of 2.197 μs. Consider a secondary cosmic muon produced at an altitude of 15.00 km aimed directly downward with an energy of 6.000 GeV. With such high energy, the muon can travel a great distance into the earth without slowing down. (a) Determine the speed of this muon. (b) How far would this muon travel in one lifetime if there were no relativistic effects? (c) In the frame of the muon, what distance separates its creation position from the earth's surface? (d) If one lifetime passes in the frame of the muon, how much time passes in the frame of the earth? (e) How far does a muon travel in this time as seen from the earth? (f) Does the muon survive its trip to the surface? How far will it penetrate the earth in its lifetime?

44.48 ••• **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.)

44.49 ••• **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.

44.50 ••• **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.

44.51 •• **DATA** While tuning up a medical cyclotron for use in isotope production, you obtain the data given in the table.

$\boldsymbol{B}$ **(T)**	0.10	0.20	0.30	0.40
$\boldsymbol{K_{\max}}$ **(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?

44.52 •• **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?

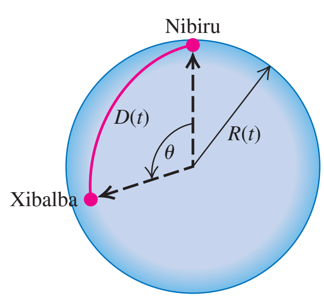
44.53 •• **DATA** You have entered a graduate program in particle physics and are learning about the use of symmetry. You begin by repeating the analysis that led to the prediction of the Ω^- particle. Nine of the spin-$\frac{3}{2}$ baryons are four Δ particles, each with mass 1232 MeV/c^2, strangeness 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?

44.54 ••• **CP** A positive kaon K^+ which is a bound state of an up quark and an antidown quark, can decay into an antimuon and a muon neutrino via a weak interaction process mediated by a W^+ boson. We can determine how much of the kaon's energy each of the daughter particles obtains using relativistic kinematics. (a) In the rest frame of the kaon, the speed of the antimuon is v and its Lorentz factor is $\gamma = (1 - v^2/c^2)^{-1/2}$. Using the notation E_ν for the energy of the neutrino, M_K for the mass of the kaon, and M_μ for the mass of the antimuon, write the relativistic expression for conservation of energy. Treat the neutrino as massless. (b) Write the relativistic equation for conservation of momentum. (c) Solve these two equations to determine v as a function of the mass ratio $\sigma = M_K/M_\mu$. (d) Using the masses M_K and M_μ found in Tables 44.3 and 44.2, respectively, determine the value of σ. (e) What is the energy of $E_\mu = \gamma M_\mu c^2$ of the antimuon? (f) What is the energy of the neutrino? (g) Do these energies add up to the rest energy of the kaon?

CHALLENGE PROBLEMS

44.55 ••• **CALC** Consider the following hypothetical universe: The Nibiruvians and the Xibalbans are minuscule "flat" societies located at an angular separation of $\theta = 60°$ on the surface of a two-dimensional spherical universe with radius R, as shown in **Fig. P44.55.** (a) What is the distance D between the two societies in terms of R and θ? (b) If R is increasing in time, what is the speed $V = dD/dt$ at which the civilizations are separating, as a function of R, dR/dt, and D? (c) The separation velocity and the distance D are related by $V = BD$. Determine the "Bubble parameter" B in terms of $R(t)$. Note that B is not constant. (Similarly, the Hubble "constant" H_0 is presumed to be changing with time.) (d) The radius of this universe was $R_0 = 500.0$ m upon its creation at time $t = 0$, and it has been increasing at a constant rate of 1.00 μm/s. What was the Bubble parameter exactly four years after creation? (e) What is the distance between Nibiru and Xibalba four years after creation? (f) At what speed are they separating at that time? (g) At that time, the Nibiruvians transmit isotropic ripple waves with their "light speed" of $c = 6.35$ μm/s and wavelength 1.00 nm. How long does it take these waves to reach Xibalba? (*Hint:* In time dt the waves travel an angular distance $d\theta = c\,dt/R(t)$. Integrate to obtain an expression for the total angular distance traveled as a function of time. Solve for the time needed to travel a given angular distance.) (h) At what wavelength will the Xibalbans observe these waves? (*Hint:* Use Eq. (44.16).)

Figure **P44.55**



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.

MCAT-STYLE 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 traveled 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, 26.1% is poorly understood dark matter and 69.0% is even more mysterious dark energy.

Key Example ✓ARIATION Problems

VP44.1.1 (a) 9.15 MHz (b) 11.0 MeV (c) 0.153c
VP44.1.2 (a) 0.136 m (b) 4.41×10^7 rad/s
VP44.1.3 (a) 0.717 T (b) 10.9 MHz
VP44.1.4 (a) 0.756 T (b) 11.6 MeV
VP44.3.1 (a) 2834 MeV (b) 2.41×10^3 MeV = 2.41 GeV
VP44.3.2 (a) 2170 MeV (b) 634 MeV
VP44.3.3 (a) 1232 MeV (b) 190 MeV
VP44.3.4 (a) 115 GeV (b) 113 GeV/c^2
VP44.9.1 (a) 0.105 (b) 0.0992c
VP44.9.2 (a) 426 nm (b) 0.0738
VP44.9.3 (a) 4.4×10^8 pc (b) 3.1×10^8 pc
VP44.9.4 (a) 2.38×10^7 m/s (b) 3.5×10^8 pc

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 defined by taking the value of Planck's constant h to be exactly $6.62607015 \times 10^{-34}\ \mathrm{kg \cdot m^2/s}$.

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 a current of one coloumb per second, where the coulomb is defined in terms of the elementary charge e.

kelvin (K) The *kelvin,* unit of thermodynamic temperature, is defined by taking the value of the Boltzmann constant k to be exactly 1.380649×10^{-23} J/K.

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 such that the elementary charge e is exactly $1.602176634 \times 10^{-19}$ C 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 SI unit of substance. One mole contains exactly $6.02214076 \times 10^{23}$ elementary entities.

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 $a = e^x$. $\quad \ln a + \ln b = \ln(ab) \quad \ln a - \ln b = \ln(a/b) \quad \ln(a^n) = n \ln a$

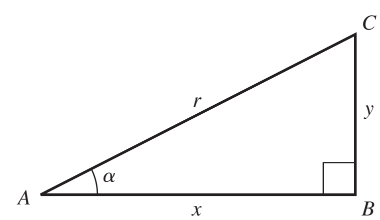
Quadratic formula: If $ax^2 + bx + c = 0$, $\quad x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

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:

$\sin^2\alpha + \cos^2\alpha = 1 \qquad \tan\alpha = \dfrac{\sin\alpha}{\cos\alpha}$

$\sin 2\alpha = 2\sin\alpha\cos\alpha \qquad \cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 2\cos^2\alpha - 1 = 1 - 2\sin^2\alpha$

$\sin\frac{1}{2}\alpha = \sqrt{\dfrac{1 - \cos\alpha}{2}} \qquad \cos\frac{1}{2}\alpha = \sqrt{\dfrac{1 + \cos\alpha}{2}}$

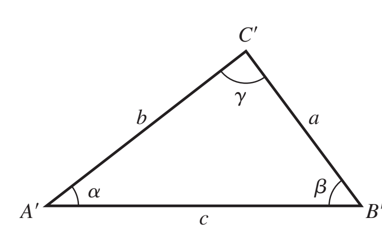
$\sin(-\alpha) = -\sin\alpha \qquad \sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$

$\cos(-\alpha) = \cos\alpha \qquad \cos(\alpha \pm \beta) = \cos\alpha\cos\beta \pm \sin\alpha\sin\beta$

$\sin(\alpha \pm \pi/2) = \pm\cos\alpha \qquad \sin\alpha + \sin\beta = 2\sin\frac{1}{2}(\alpha + \beta)\cos\frac{1}{2}(\alpha - \beta)$

$\cos(\alpha \pm \pi/2) = \mp\sin\alpha \qquad \cos\alpha + \cos\beta = 2\cos\frac{1}{2}(\alpha + \beta)\cos\frac{1}{2}(\alpha - \beta)$

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** (267)	105 **Db** (270)	106 **Sg** (269)	107 **Bh** (270)	108 **Hs** (269)	109 **Mt** (278)	110 **Ds** (281)	111 **Rg** (281)	112 **Cn** (285)	113 **Nh** (286)	114 **Fl** (289)	115 **Mc** (289)	116 **Lv** (293)	117 **Ts** (293)	118 **Og** (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 ($1\ u = 1.660539040(20) \times 10^{-27}$ kg), equivalent to grams per mole (g/mol).

APPENDIX E

UNIT CONVERSION FACTORS

Length

1 m = 100 cm = 1000 mm = 10^6 μm = 10^9 nm
1 km = 1000 m = 0.6214 mi
1 m = 3.281 ft = 39.37 in.
1 cm = 0.3937 in.
1 in. = 2.540 cm
1 ft = 30.48 cm
1 yd = 91.44 cm
1 mi = 5280 ft = 1.609 km
1 Å = 10^{-10} m = 10^{-8} cm = 10^{-1} nm
1 nautical mile = 6080 ft
1 light-year = 9.461×10^{15} m

Area

1 cm^2 = 0.155 $in.^2$
1 m^2 = 10^4 cm^2 = 10.76 ft^2
1 $in.^2$ = 6.452 cm^2
1 ft^2 = 144 $in.^2$ = 0.0929 m^2

Volume

1 liter = 1000 cm^3 = 10^{-3} m^3 = 0.03531 ft^3 = 61.02 $in.^3$
1 ft^3 = 0.02832 m^3 = 28.32 liters = 7.477 gallons
1 gallon = 3.788 liters

Time

1 min = 60 s
1 h = 3600 s
1 d = 86,400 s
1 y = 365.24 d = 3.156×10^7 s

Angle

1 rad = 57.30° = 180°/π
1° = 0.01745 rad = π/180 rad
1 revolution = 360° = 2π rad
1 rev/min (rpm) = 0.1047 rad/s

Speed

1 m/s = 3.281 ft/s
1 ft/s = 0.3048 m/s
1 mi/min = 60 mi/h = 88 ft/s
1 km/h = 0.2778 m/s = 0.6214 mi/h
1 mi/h = 1.466 ft/s = 0.4470 m/s = 1.609 km/h
1 furlong/fortnight = 1.662×10^{-4} m/s

Acceleration

1 m/s^2 = 100 cm/s^2 = 3.281 ft/s^2
1 cm/s^2 = 0.01 m/s^2 = 0.03281 ft/s^2
1 ft/s^2 = 0.3048 m/s^2 = 30.48 cm/s^2
1 mi/h · s = 1.467 ft/s^2

Mass

1 kg = 10^3 g = 0.0685 slug
1 g = 6.85×10^{-5} slug
1 slug = 14.59 kg
1 u = 1.661×10^{-27} kg
1 kg has a weight of 2.205 lb when g = 9.80 m/s^2

Force

1 N = 10^5 dyn = 0.2248 lb
1 lb = 4.448 N = 4.448×10^5 dyn

Pressure

1 Pa = 1 N/m^2 = 1.450×10^{-4} lb/$in.^2$ = 0.0209 lb/ft^2
1 bar = 10^5 Pa
1 lb/$in.^2$ = 6895 Pa
1 lb/ft^2 = 47.88 Pa
1 atm = 1.013×10^5 Pa = 1.013 bar
= 14.7 lb/$in.^2$ = 2117 lb/ft^2
1 mm Hg = 1 torr = 133.3 Pa

Energy

1 J = 10^7 ergs = 0.239 cal
1 cal = 4.186 J (based on 15° calorie)
1 ft · lb = 1.356 J
1 Btu = 1055 J = 252 cal = 778 ft · lb
1 eV = 1.602×10^{-19} J
1 kWh = 3.600×10^6 J

Mass–Energy Equivalence

1 kg $\leftrightarrow$ 8.988×10^{16} J
1 u $\leftrightarrow$ 931.5 MeV
1 eV $\leftrightarrow$ 1.074×10^{-9} u

Power

1 W = 1 J/s
1 hp = 746 W = 550 ft · lb/s
1 Btu/h = 0.293 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	$1.602176634 \times 10^{-19}$ C
Gravitational constant	G	$6.67408(31) \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$
Planck's constant	h	$6.62607015 \times 10^{-34}\ \mathrm{J \cdot s}$
Boltzmann constant	k	1.380649×10^{-23} J/K
Avogadro's number	N_A	$6.02214076 \times 10^{23}$ molecules/mol
Gas constant	$R = N_A K$	$8.314462618\ldots\ \mathrm{J/mol \cdot K}$
Mass of electron	m_e	$9.10938356(11) \times 10^{-31}$ kg
Mass of proton	m_p	$1.672621898(21) \times 10^{-27}$ kg
Mass of neutron	m_n	$1.674927471(21) \times 10^{-27}$ kg
Magnetic constant	μ_0	$1.25663706 \times 10^{-6}\ \mathrm{Wb/A \cdot m}$ (approximate) $\cong 4\pi \times 10^{-7}\ \mathrm{Wb/A \cdot m}$
Electric constant	$\epsilon_0 = 1/\mu_0 c^2$	$8.854187817 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2}$ (approximate)
	$1/4\pi\epsilon_0$	$8.987551787 \times 10^{9}\ \mathrm{N \cdot m^2/C^2}$ (approximate)

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	$1.6021766209(98) \times 10^{-19}$ J
Atomic mass unit	1 u	$1.660539040(20) \times 10^{-27}$ kg
Electron rest energy	$m_e c^2$	0.5109989461(31) MeV
Volume of ideal gas (0°C and 1 atm)		22.413962(13) 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 $\pm$ 0.0021. Values shown without uncertainties are exact. The exact values of the magnitude of the charge of the electron, Planck's constant, the Boltzmann constant, Avogadro's number, and the gas constant are from the redefinitions adopted in 2018. As consequences of these redefinitions, the values of the magnetic constant and electric constant now have fractional uncertainties of about 2×10^{-10}. As of this writing (2018) it was expected that updated values of the magnetic and electric constants, as well as of all other constants with uncertainties, were to be announced in May 2019. These updated values will be available at **http://physics.nist.gov/cuu**.

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‡	1.30×10^{22}	1.19×10^{6}	5.91×10^{12}	248.0 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 31.7 y
1.7 (a) 23.4 km/L (b) 1.4 tanks
1.9 9.0 cm
1.11 (a) 72 mm^2 (b) 0.50 (c) 36 mm (d) 6 mm (e) 2.0
1.13 0.45%
1.15 (a) no (b) no (c) no (d) no (e) no
1.17 $300 million
1.19 2×10^5
1.21 7.8 km, 38° north of east
1.23 144 m, 41° south of west.
1.25 $A_x = 0$, $A_y = -8.00$ 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
1.27 (a) −6.00 m (b) 11.3 m
1.29 (a) 9.01 m, 33.7° (b) 9.01 m, 33.7° (c) 22.3 m, 250° (d) 22.3 m, 70.3°
1.31 2.81 km, 38.5° north of west
1.33 (a) 2.48 cm, 18.4° (b) 4.09 cm, 83.7° (c) 4.09 cm, 264°
1.35 $\vec{A} = -(8.00\text{ m})\,\hat{\jmath}$, $\vec{B} = (7.50\text{ m})\,\hat{\imath} + (+13.0\text{ m})\,\hat{\jmath}$, $\vec{C} = (-10.9\text{ m})\hat{\imath} + (-5.07\text{ m})\hat{\jmath}$, $\vec{D} = (-7.99\text{ m})\hat{\imath} + (6.02\text{ m})\hat{\jmath}$
1.37 (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°
1.39 (a) $A = 5.38$, $B = 4.36$ (b) $-5.00\,\hat{\imath} + 2.00\,\hat{\jmath} + 7.00\,\hat{k}$ (c) 8.83, yes
1.41 (a) -104 m^2 (b) -148 m^2 (c) 40.6 m^2
1.43 (a) 165° (b) 28° (c) 90°
1.45 (a) $(-63.9\text{ m}^2)\hat{k}$ (b) $(63.9\text{ m}^2)\hat{k}$
1.47 −2.0 m, $-x$-direction
1.49 (a) 5.49 g/cm^3 (b) $1.1 \times 10^6\text{ g/cm}^3$ (c) $4.7 \times 10^{14}\text{ g/cm}^3$
1.51 (a) 1.64×10^4 km (b) $2.57r_E$
1.53 $\approx 6 \times 10^{27}$
1.55 (a) $0.60\,\hat{\imath} - 0.80\,\hat{k}$ (b) $-0.60\,\hat{\imath} + 0.80\,\hat{k}$ (c) $\vec{B}_+ = 0.80\,\hat{\imath} + 0.60\,\hat{k}$ and $\vec{B}_- = -0.80\,\hat{\imath} - 0.60\,\hat{k}$
1.57 144 m, 41° south of west
1.59 361 N/C, 184.73° from the $+x$-axis
1.61 7.55 N
1.63 60.9 km, 33.0° south of west
1.65 28.8 m, 11.4° north of east
1.67 71.9 m, 64.1° north of west
1.69 (a) 818 m, 15.8° west of south
1.71 18.6° east of south, 29.6 m
1.73 28.2 m
1.75 −2.00 J
1.77 124°
1.79 (a) $-bd$, $-bc\hat{\imath} - ad\hat{\jmath} - ac\hat{k}$ (b) $-bd$, ad in the $-y$-direction
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) 0.313 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) 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
2.15 (a) 0.500 m/s^2 (b) 0, 1.00 m/s^2
2.17 (a) 2.17 m, 9.60 m/s^2; 15.0 m, −38.4 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) $3.1 \times 10^6\text{ m/s}^2 = 3.2 \times 10^5\,g$ (b) 1.6 ms (c) no
2.29 (a) 0, 6.3 m/s^2, −11.2 m/s^2 (b) 100 m, 230 m, 320 m
2.31 (a) 2.94 m/s (b) 0.600 s
2.33 1.67 s
2.35 (a) 33.5 m (b) 15.8 m/s
2.37 (a) upward, downward, decreasing (b) downward, downward, increasing
2.39 (a) $t = \sqrt{2d/g}$ (b) 0.190 s
2.41 (a) 646 m (b) 16.4 s, 112 m/s
2.43 15.9 m
2.45 0.0868 m/s^2
2.47 37.6 m/s
2.49 (a) 467 m (b) 110 m/s
2.51 (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
2.53 (a) 10.0 m (b) (i) 8.33 m/s (ii) 9.09 m/s (iii) 9.52 m/s
2.55 250 km
2.57 (a) 197 m/s (b) 169 m/s
2.59 (a) 4.24 m/s (b) 7.54 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 3.2 m/s^2
2.69 (a) 2.0 s (b) 16 m/s in the $-x$-direction, −40 m/s^2 in the $-x$-direction
2.71 (a) −4.00 m/s (b) 12.0 m/s
2.73 38.2 m
2.75 (a) 6.69 m/s (b) 4.49 m (c) 1.42 s
2.77 (a) 3.3 s (b) $9H$
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, −2.33 m/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
3.25 (a) 0.034 m/s^2 = 0.0034g (b) 1.4 h
3.27 52.8 m/s^2
3.29 120 m/s, 270 mph
3.31 (a) 2.57 m/s^2 upward (b) 2.57 m/s^2 downward (c) 14.7 s
3.33 (a) 32.9 m/s (b) 27.7 m/s^2 (c) 35.5 rpm
3.35 (a) 14 s (b) 70 s
3.37 0.36 m/s, 52.5° south of west
3.39 (a) 24° west of south (b) 5.5 h
3.41 (a) 4.7 m/s, 25° south of east (b) 120 s (c) 240 m
3.43 (a) 14°, north of west (b) 310 km/h
3.45 $2b/3c$
3.47 (a) 128 m (b) 315 m
3.49 274 m
3.51 (a) 55.5 m south
3.53 33.7 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) $v = \sqrt{v_0^2 + 2gH}$ (b) $v = \sqrt{v_0^2 + 2gH}$ (c) $\sqrt{v_0^2 + 2gH}$ (d) stays the same
3.61 (a) 6.91 m (c) no
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 6.25 s
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, $B = 12.5\text{ N/s}^2$ (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 7.4×10^{-23} m/s^2
4.27 (b) yes
4.29 (a) yes (b) no
4.33 2.58 s
4.35 (a) 17 N, 90° clockwise from the $+x$-axis (b) 840 N
4.37 (a) 2.50 m/s^2 (b) 10.0 N (c) to the right, $F > T$ (d) 25.0 N
4.39 (a) 4.4 m (b) 300 m/s (c) (i) 2.7×10^4 N (ii) 9.0×10^3 N
4.41 (a) 0.603 m/s^2, upward (b) 1.26 m/s^2, downward
4.43 1630 N away from him
4.45 (a) 4.34 kg (b) 5.30 kg
4.47 309 N
4.49 7.78 m
4.51 (a) largest: Ferrari; smallest: Alpha Romeo and Honda Civic (b) largest: Ferrari; smallest: Volvo (c) 7.5 kN, smaller (d) zero
4.53 (b) 26 kg, 8.3 m/s^2
4.55 (a) 0.593 m, 1.33 s, 4.00 N (b) 2.00 s, 2.00 m/s
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) 3.4 m/s (c) $2.2w$
5.21 (a) 14.0 m (b) 18.0 m/s
5.23 50°
5.25 (a) 33 N (b) 3.1 m
5.27 (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 (upper cable)
5.53 (a) 1.5 rev/min (b) 0.92 rev/min
5.55 (a) $\sqrt{Lg}$ (b) $5mg$
5.57 2.42 m/s
5.59 (a) rope making 60° angle (b) 6400 N
5.61 $T_B = 4960$ N, $T_C = 1200$ N
5.63 3.18 m/s
5.65 (a) 101 N (b) 33.8 N (c) 62.3 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
5.71 245 N, equal to it
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 1.30 kg, 2.19 kg
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 (a) 7.75 N (b) 0.242
5.95 g/μ_s
5.97 (b) 0.452
5.99 (b) 0.28 (c) no
5.101 (a) 1.42×10^4 N (b) 817 m/s^2 (c) 27.1 m/s^2
5.103 (b) 8.8 N (c) 31.0 N (d) 1.54 m/s^2
5.105 $v = (2mg/k)\left[\frac{1}{2} + e^{-(k/m)t}\right]$
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 −6.7 J
6.11 (a) 0° (b) 4
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 $4W_1$
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) twice as fast (b) both decrease by a factor of $1/\sqrt{3}$
6.31 (a) $v_0^2/2\mu_k g$ (b) (i) $^1/_2$ (ii) 4 (iii) 2
6.33 (b) 13.1 cm (bottom), 14.1 cm (middle), 15.2 cm (top)
6.35 (a) 2.83 m/s (b) 3.46 m/s
6.37 8.5 cm
6.39 (a) 1.76 (b) 0.666 m/s
6.41 (a) 4.0 J (b) 0 (c) −1.0 J (d) 3.0 J (e) −1.0 J
6.43 (a) 2.83 m/s (b) 2.40 m/s
6.45 8.17 m/s
6.47 360,000 J; 100 m/s
6.53 (a) 84.6/min (b) 22.7/min
6.55 29.6 kW
6.57 0.20 W
6.59 (a) 608 J (b) −395 J (c) 0 (d) −189 J (e) 24 J (f) 1.5 m/s
6.61 (a) −35.5 J (b) decreases (c) −51.6 J
6.63 (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)
6.65 (a) 1.8 m/s = 4.0 mi/h (b) 180 m/s^2 ≈ $18g$, 900 N
6.67 (a) 5.11 m (b) 0.304 (c) 10.3 m
6.69 (a) $V/\sqrt{2}$, greater than (b) $W_D/4$, less than
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 (a) 47.0 J, −29.6 J, 2.41 m/s (b) 23.2 J, 29.6 J, −6.4 J (c) 40.6 J, 47.0 J, −6.4 J
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) 11.8 J, −15.7 J (b) (0.200 m)T, −(0.200 m)T (c) 0, −3.9 J, 0.75 m/s
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 10.8 J
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: $-1.5\text{ m} < x < -1.0\text{ m}$ and $0 < x < 1.0\text{ m}$
negative: $-1.0\text{ m} < x < 0$ and $1.0\text{ m} < x < 1.5\text{ m}$ (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 (a) 8.00×10^2 kg·m/s (b) 8.00×10^3 J (c) 160 N opposite to her velocity
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 0.31 m/s toward the south
8.33 (a) 0.846 m/s (b) 2.10 J
8.35 13.5 J
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 (a) B (b) 3 times greater (c) 3.0 m/s to the left
8.53 2520 km
8.55 0.488 m
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
8.65 (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 $v_A/2$ for each one
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 (a) 1 (b) $3 - 2\sqrt{2}, 3 + 2\sqrt{2}$
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 4.04 m/s
8.91 1.33 m
8.93 $H/9$
8.95 (a) 0.150 J (b) 15.4 J (c) 0.560 m
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
9.21 (a) 15.1 m/s^2 (b) 15.1 m/s^2
9.25 0.107 m, no
9.27 5.97 rad/s^2
9.29 0.180 m/s^2, 0.432 m/s^2, 0.468 m/s^2
9.31 (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
9.33 (a) 1.93 kg·m^2 (b) 6.53 kg·m^2 (c) 0 (d) 1.15 kg·m^2
9.35 0.193 kg·m^2
9.37 0.208 s
9.39 $\omega_1\sqrt{3/5}$
9.41 6.49 m/s
9.43 (a) wheel B (b) $\dfrac{3}{16}$
9.45 7.35×10^4 J
9.47 (a) 0.673 m (b) 45.5%
9.49 46.5 kg
9.51 an axis that is parallel to a diameter and is 0.516R from the center
9.53 $M(a^2 + b^2)/3$
9.55 $\frac{1}{2}MR^2$
9.57 (a) $\gamma L^2/2$ (b) $ML^2/2$ (c) $ML^2/6$
9.59 7.68 m
9.61 (a) 0.600 m/s^3 (b) $\alpha = (2.40\text{ rad/s}^3)t$ (c) 3.54 s (d) 17.7 rad
9.63 1.22 kg
9.65 8.44 kg
9.67 2.99 cm
9.69 68.4 J
9.71 4.65 kg·m^2
9.73 (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
9.75 1.46 m/s
9.77 $\sqrt{\dfrac{2gd(m_B - \mu_k m_A)}{m_A + m_B + I/R^2}}$
9.79 (a) 2.25×10^{-3} kg·m^2 (b) 3.40 m/s (c) 4.95 m/s
9.81 13.9 m
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 0.0131 kg·m^2
10.19 (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
10.21 (a) $\frac{1}{3}$ (b) $\frac{2}{7}$ (c) $\frac{2}{5}$ (d) $\frac{5}{13}$
10.23 (a) 0.613 (b) no (c) no slipping
10.25 14.0 m
10.29 (a) 3.76 m (b) 8.58 m/s
10.31 (a) 67.9 rad/s (b) 8.35 J
10.33 (a) 0.309 rad/s (b) 100 J (c) 6.67 W
10.35 (a) 0.704 N·m (b) 157 rad (c) 111 J (d) 111 J
10.37 (a) 115 kg·m^2/s into the page (b) 125 kg·m^2/s^2 out of the page
10.39 4.71×10^{-6} kg·m^2/s
10.41 (a) A: rad/s^2; B: rad/s^4 (b) (i) 59.0 kg·m^2/s (ii) 56.1 N·m
10.43 4600 rad/s
10.45 1.14 rev/s
10.47 (a) 1.38 rad/s (b) 1080 J, 495 J
10.49 (a) 0.120 rad/s (b) 3.20×10^{-4} J (c) work done by the bug
10.51 4.82 m/s^2
10.53 $\dfrac{m_{\text{rock}}v}{\left(m + \dfrac{M}{2}\right)R}$
10.55 2.4×10^{-12} N·m
10.59 0.483
10.61 (a) 16.3 rad/s^2 (b) no, decrease (c) 5.70 rad/s
10.63 0.921 m/s^2, 7.68 rad/s^2, 35.5 N (at A), 21.4 N (at B)
10.65 $\dfrac{7v_{\text{cm}}^2}{10g}, \dfrac{5v_{\text{cm}}^2}{6g}$, hollow
10.67 68.0 N
10.69 270 N
10.71 $a = \dfrac{2g}{2 + (R/b)^2}$, $\alpha = \dfrac{2g}{2b + R^2/b}$,
$T = \dfrac{2mg}{2(b/R)^2 + 1}$
10.73 (a) $3H_0/5$
10.75 29.0 m/s
10.77 (a) 26.0 m/s (b) no change
10.79 $g/3$

10.81 (a) $\frac{6}{19}v/L$ (b) $\frac{3}{19}$
10.83 3200 J
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) $mv_1^2 r_1^2/r^3$ (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 9.70 cm
11.7 6.6 kN
11.9 (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
11.11 (a) 550 N (b) 0.614 m from A
11.13 (a) 1920 N (b) 1140 N
11.15 (a) $T = 2.60w$; $3.28w$, 37.6° (b) $T = 4.10w$; $5.39w$, 48.8°
11.17 (a) 3410 N (b) 3410 N, 7600 N
11.19 (b) 530 N (c) 600 N, 270 N; downward
11.21 220 N (left), 255 N (right), 42°
11.23 (a) 0.800 m (b) clockwise (c) 0.800 m, clockwise
11.25 6.12 kg
11.27 1.9 mm
11.29 2.0×10^{11} Pa
11.31 2.0×10^3 atm
11.33 (a) 150 atm (b) 1.5 km, no
11.35 4.8×10^9 Pa, 2.1×10^{-10} Pa^{-1}
11.37 (a) 2.4×10^{-2} (b) 2.4×10^{-3} m
11.39 $1.4T_{\text{steel}}$
11.41 $0.19\Delta p_1$
11.43 3.41×10^7 Pa
11.47 20.0 kg
11.49 (a) 525 N (b) 222 N, 328 N (c) 1.48
11.51 (a) 140 N (b) 6 cm to the right
11.53 (a) 409 N (b) 161 N
11.55 49.9 cm
11.57 (a) $\arctan(mg/2T)$ (b) $\sqrt{\dfrac{3g\sin\theta}{L}}$
11.59 (a) $\arctan\left[\left(\dfrac{1}{\alpha} - 1\right)\tan\theta\right]$ (b) 1/2 (c) 0
11.61 5500 N
11.63 (b) 2000 N = $2.72mg$ (c) 4.4 mm
11.65 (a) 1.39 m (b) 1.8×10^{-3} m
11.67 (a) 175 N at each hand, 200 N at each foot (b) 91 N at each hand and at each foot
11.69 (a) 1150 N (b) 1940 N (c) 918 N (d) 0.473
11.71 590 N (person above), 1370 N (person below); person above
11.73 (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
11.75 (a) 375 N (b) 325 N (c) 512 N
11.77 (a) 0.424 N (A), 1.47 N (B), 0.424 N (C) (b) 0.848 N
11.79 (a) 27° to tip, 31° to slip, tips first (b) 27° to tip, 22° to slip, slips first
11.81 (a) 80 N (A), 870 N (B) (b) 1.92 m
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 (a) 1.86×10^6 Pa (b) 184 m
12.9 0.581 m
12.11 828 kg/m^3
12.13 2.8 m
12.15 6.0×10^4 Pa
12.17 2.27×10^5 N
12.19 (a) 636 Pa (b) (i) 1170 Pa (ii) 1170 Pa
12.21 10.9
12.23 10.3 m
12.25 0.0898 m^3, 1020 kg/m^3
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
12.37 501 N
12.39 9.6 m/s
12.41 (a) 17.0 m/s (b) 0.317 m
12.43 (a) 0.536 mm (b) 30.8 s, yes
12.45 28.4 m/s
12.47 1.47×10^5 Pa
12.49 2.03×10^4 Pa
12.51 2.25×10^5 Pa
12.53 $1.19D$
12.55 (a) 5.9×10^5 N (b) 1.8×10^5 N
12.57 2.61×10^4 N·m
12.59 0.800 kg
12.61 0.964 cm, rise
12.63 (a) 1470 Pa (b) 13.9 cm
12.65 (a) 0.0500 m^3 (b) 10.0 kg
12.67 9.8×10^6 kg, yes
12.71 (a) 16.5 cm (b) 1.75 m
12.73 (a) 5.07 m/s, 1.28 (b) 32.4 min, 2.08
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 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) $F_{\text{moon}}/F_{\text{earth}} = 3.5 \times 10^{-6}$
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 (a) 12.3 m/s^2 (b) 80.0% the density of earth
13.17 (a) 5020 m/s (b) 60,600 m/s
13.19 9.03 m/s^2
13.21 2.40×10^2 N
13.23 -2.00×10^6 J, -4.00×10^6 J
13.25 (a) 7410 m/s (b) 1.71 h
13.27 7330 m/s
13.29 (a) 4.1 m/s = 9.1 mph, yes (b) 2.6 h
13.31 (a) 82,700 m/s (b) 14.5 days
13.33 (a) 7.84×10^9 s = 248 y (b) 4.44×10^{12} m, 7.38×10^{12} m
13.35 (a) 0 (b) 1.07×10^{-12} N (c) 1.25×10^{-12} N
13.37 (a) (i) 5.31×10^{-9} N (ii) 2.67×10^{-9} N
13.39 (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) $U = -GmM/a$, $F_x = 0$
13.41 (a) 33.7 N (b) 32.8 N
13.43 (a) 4.3×10^{37} kg = $(2.1 \times 10^7)M_S$ (b) no (c) 6.32×10^{10} m, yes
13.45 (a) 9.67×10^{-12} N, at 45° above $+x$-axis (b) 3.02×10^{-5} m/s
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 $4\pi GM$
13.61 $v = \sqrt{\dfrac{2Gm_E h}{R_E(R_E + h)}}$
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) 7910 s (b) 1.53 (c) 8430 m/s (perigee), 5510 m/s (apogee) (d) 2410 m/s; 3250 m/s; perigee
13.69 5.36×10^9 J
13.71 9.36 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) 7.90×10^3 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) 90 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) $T/6$ (b) $T/12$ (c) no
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) $\omega^3 A \sin \omega t$ (b) 0 (c) 0 (d) $\pm A$ (e) 1.3 s
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 $\frac{\sqrt{3}}{2}A$, larger
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 (a) 0.25 s (b) 0.25 s
14.45 0.407 swings per second
14.47 10.7 m/s^2
14.49 (a) $g \sin\theta$ (b) $g\sqrt{\cos^2(\theta/2) + 4[\cos(\theta/2) - \cos\theta]^2}$ (c) $2g(1 - \cos\theta)$, $\sqrt{2gL(1 - \cos\theta)}$
14.51 A: $2\pi\sqrt{\frac{L}{g}}$, B: $\frac{2\sqrt{2}}{3}\left(2\pi\sqrt{\frac{L}{g}}\right)$; pendulum A
14.53 A: $2\pi\sqrt{\frac{L}{g}}$, B: $\sqrt{\frac{11}{10}}\left(2\pi\sqrt{\frac{L}{g}}\right)$, pendulum B
14.55 $\frac{\sqrt{3}}{2}\omega$
14.57 (a) 0.393 Hz (b) 1.73 kg/s
14.59 (a) $A_1/3$ (b) $2A_1$
14.61 (a) 12 (b) 72
14.63 0.353 m
14.65 (a) $16k_1$ (b) double it
14.67 599 N/m
14.69 (a) 24.4 cm (b) 0.221 s (c) 1.19 m/s
14.71 2.00 m
14.73 $0.921\left(\frac{1}{2\pi}\sqrt{\frac{g}{L}}\right)$
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 $2\pi\sqrt{\frac{M}{3k}}$
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) $y(x, t) = (0.0700 \text{ m}) \cos[(19.6 \text{ m}^{-1})x + (157 \text{ rad/s})t]$ (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
15.23 4.10 mm
15.25 (a) 95 km (b) 0.25 μW/m^2 (c) 110 kW
15.27 (a) 0.050 W/m^2 (b) 22 kJ
15.35 (a) $(1.33 \text{ m})n$, $n = 0, 1, 2, \ldots$ (b) $(1.33 \text{ m})(n + \frac{1}{2})$, $n = 0, 1, 2, \ldots$
15.37 (a) 96.0 m/s (b) 461 N (c) 1.13 m/s, 4.26 m/s^2
15.39 (b) 2.80 cm (c) 277 cm (d) 185 cm, 7.96 Hz, 0.126 s, 1470 cm/s (e) 280 cm/s (f) $y(x, t) = (5.60 \text{ cm}) \times \sin[(0.0906 \text{ rad/cm})x] \sin[(133 \text{ rad/s})t]$
15.41 (a) 0.240 m (b) 0.0600 m
15.43 (a) 45.0 cm (b) no
15.45 1.80×10^3 m/s, 1.06×10^{-4} m
15.47 (a) 20.0 Hz, 126 rad/s, 3.49 rad/m (b) $y(x, t) = (2.50 \times 10^{-3} \text{ m}) \cos[(3.49 \text{ rad/m})x - (126 \text{ rad/s})t]$ (c) $y(0, t) = (2.50 \times 10^{-3} \text{ m}) \cos[(126 \text{ rad/s})t]$ (d) $y(1.35\ m, t) = (2.50 \times 10^{-3} \text{ m}) \times \cos[(126 \text{ rad/s})t - 3\pi/2 \text{ rad}]$ (e) 0.315 m/s (f) -2.50×10^{-3} m, 0
15.49 6.00×10^{-2} kg
15.51 (a) 62.1 m
15.53 13.7 Hz, 25.0 m
15.55 1.83 m
15.57 360 Hz (copper), 488 Hz (aluminum)
15.59 (a) 18.8 cm (b) 0.0169 kg
15.61 (a) 7.07 cm (b) 0.400 kW
15.63 (a) 0.0673 kg/m (b) 14.3 N (c) 14.6 m/s (d) 19.5 Hz (e) 1, 26, 51, 76, 101 (f) 0.0615 m (g) $(n-1)(15.0 \text{ mm})$ (h) 3.64 m/s
15.65 (a) 2.22 g (b) 2.24×10^4 m/s^2
15.67 233 N
15.69 1780 kg/m^3
15.71 (a) 148 N (b) 26%
15.73 (c) 47.5 Hz (d) 138 g
15.75 (a) because $P_{av} = 2\pi^2\sqrt{\mu F}A^2 f^2$ (b) 280 m/s (c) 1100 rad/s
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 0.156 s
16.7 90.8 m
16.9 81.4°C
16.11 (a) 5.5×10^{-15} J (b) 0.074 mm/s
16.13 15.0 cm
16.15 (a) 4.14 Pa (b) 0.0208 W/m^2 (c) 103 dB
16.17 (a) 4.4×10^{-12} W/m^2 (b) 6.4 dB (c) 5.8×10^{-11} m
16.19 14.0 dB
16.21 (a) 2.0×10^{-7} W/m^2 (b) 6.0 m (c) 290 m (d) yes, no
16.23 (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
16.25 506 Hz, 1517 Hz, 2529 Hz
16.27 (a) 35.2 Hz (b) 17.6 Hz
16.29 (a) closed (b) 5th harmonic (c) 0.453 m (d) 190 Hz
16.31 (a) 614 Hz (b) 1230 Hz
16.33 (a) 172 Hz (b) 86 Hz
16.35 0.125 m
16.37 (a) $(820 \text{ Hz})n$, $n = 1, 2, 3, \ldots$ (b) $(410 \text{ Hz})(2n + 1)$, $n = 0, 1, 2, \ldots$
16.39 (a) 433 Hz (b) loosen
16.41 780 m/s
16.43 (a) 375 Hz (b) 371 Hz (c) 4 Hz
16.45 (a) 0.25 m/s (b) 0.91 m
16.47 19.8 m/s
16.49 (a) 355 m/s (b) 0.710 m (c) 500 Hz
16.51 (a) 7.02 m/s, toward (b) 1404 Hz
16.53 (a) 36.0° (b) 2.94 s
16.55 (b) 344 m/s (c) no
16.57 (a) 0.879 s (b) 27.3 Hz
16.59 (a) stopped (b) 7th and 9th (c) 0.439 m
16.61 (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
16.63 (a) 0.0823 m (b) 120 Hz
16.67 (b) 2.0 m/s
16.69 (a) 38 Hz (b) no
16.71 (a) 375 m/s (b) 1.39 (c) 0.8 cm
16.73 (a) 441 Hz (b) 0.430 m
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 $\frac{L_{\text{Invar}}}{L_{\text{Al}}} = \frac{1}{0.225} = 4$
17.25 554 kJ
17.27 23 min
17.29 240 J/kg·K
17.31 0.526 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 54.5 kJ, 13.0 kcal, 51.7 Btu
17.45 357 m/s
17.47 3.45 L
17.49 5.05×10^{15} kg
17.51 0.0674 kg
17.53 2370 J/kg·K
17.55 157 W/m·K
17.57 (a) −0.86°C (b) 24 W/m^2
17.61 105.5°C
17.63 (a) 21 kW (b) 6.4 kW
17.65 2.1 cm^2
17.67 (a) 1.61×10^{11} m (b) 5.43×10^6 m
17.69 35.0°C
17.71 $c = \frac{Q\beta}{\rho\Delta V}$
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 2.00×10^3 J/kg·K

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 bottles/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) 1.04 kW (b) 87.1 W (c) 1.13 kW (d) 28 g (e) 1.1 bottles
17.105 (c) 170 h (d) 1.5×10^{10} s ≈ 500 y, no
17.107 5.82 g
17.109 (a) 3.00×10^4 J/kg (b) 1.00×10^3 J/kg · K (liquid), 1.33×10^3 J/kg · K (solid)
17.111 *A:* 216 W/m · K, *B:* 130 W/m · K
17.113 $H = \dfrac{k\pi R_1 R_2}{L}(T_H - T_C)$
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) 5.7×10^6 Pa (b) 3.79×10^6 Pa, −25.2%
18.7 503°C
18.9 19.7 kPa
18.11 0.159 L
18.13 $0.0508V$
18.15 0.421 mol
18.17 (a) 6.95×10^{-16} kg (b) 2.32×10^{-13} kg/m^3
18.19 55.6 mol, 3.35×10^{25} molecules
18.21 (a) 2.20×10^6 molecules (b) 2.44×10^{19} molecules
18.23 (a) 3.45×10^{-9} m (b) about 10 times the diameter of a typical molecule (c) 10 times the spacing of atoms in solids
18.25 (a) 5.83×10^7 J (b) 242 m/s
18.27 6.46×10^{23}
18.29 (a) 1.93×10^6 m/s, no (b) 7.3×10^{10} K
18.31 (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
18.33 3800°C
18.35 1100 m/s
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) 1.34 (b) 1.23
18.45 (a) 610 Pa (b) 22.12 MPa
18.49 (a) 11.8 kPa (b) 0.566 L
18.51 272°C
18.53 0.195 kg
18.55 (a) $a = \left(\dfrac{T_{in}}{T_{out}} - 1\right)g$ (b) $T_{in} = T_{out}\left(\dfrac{a}{g} + 1\right)$
18.57 (a) 0.0959 m^3 (b) 17.8 mol/s (c) 42.6 m
18.59 (a) 30.7 cylinders (b) 8420 N (c) 7800 N
18.61 (a) 26.2 m/s (b) 16.1 m/s, 5.44 m/s (c) 1.74 m
18.63 $\approx 5 \times 10^{27}$ atoms
18.65 (a) 4.65×10^{-26} kg (b) 6.11×10^{-21} J (c) 2.04×10^{24} molecules (d) 12.5 kJ
18.67 (b) r_2 (c) $r_1 = \dfrac{R_0}{2^{1/6}}$, $r_2 = R_0$, $2^{-1/6}$ (d) U_0
18.69 (a) $2R$ = 16.6 J/mol · K (b) less than
18.71 (b) 1.40×10^5 K (N_2), 1.01×10^4 K (H_2) (c) 6370 K (N_2), 459 K (H_2)
18.73 $3kT/m$, same
18.75 (b) $0.0421N$ (c) $(2.94 \times 10^{-21})N$ (d) $0.0297N$, $(2.08 \times 10^{-21})N$ (e) $0.0595N$, $(4.15 \times 10^{-21})N$
18.77 (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
18.79 (a) 42.6% (b) 3 km (c) 1 km
18.81 (a) 4.1×10^{-46} kg · m^2 (c) $v_f = 0$, $\omega = v_i/r$ (d) 5.5×10^{12} rad/s
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) Q = +100 J, W = −300 J (b) Q = −100 J, W = −300 J (c) Q = +100 J, W = +300 J (d) −100 J, W = +300 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 $4\sqrt{2}$
19.35 (a) increase (b) 4.8 kJ
19.37 (a) 0.681 mol (b) 0.0333 m^3 (c) 2.23 Kj d) 0
19.39 (a) 45.0 J (b) liberate, 65.0 J (c) 23.0 J, 22.0 J
19.41 (a) the same (b) 4.0 kJ, absorb (c) 8.0 kJ
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 (b) −300 J, out of gas
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) 0.169 m (b) 198°C (c) 71.3 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 (a) 53.5 J + $v^{0.40}$[(21.4 J)ln v − 53.5 J] (b) 7 (c) 7 (d) 7 (e) 63.1 J
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) 12 (b) 7.4 kJ, 20 kJ
20.9 (a) 58% (b) 1.4%
20.11 1.2 h
20.13 576 K, 504 K
20.15 (a) 215 J (b) 378 K (c) 39.0%
20.17 (a) 38 kJ (b) 590°C
20.19 (a) 492 J (b) 212 W (c) 5.4
20.21 161 K
20.23 (a) 429 J/K (b) −393 J/K (c) 36 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 10.0 J/K
20.31 (a) 121 J (b) 3800 cycles
20.33 (a) 90.2 J (b) 320 J (c) 45°C (d) 0 (e) 263 g
20.35 −5.8 J/K, decrease
20.37 (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%
20.39 (a) 21.0 kJ (enters), 16.6 kJ (leaves) (b) 4.4 kJ, 21% (c) $e = 0.31e_{max}$
20.41 (a) 7.0% (b) 3.0 MW; 2.8 MW (c) 6×10^5 kg/h = 6×10^5 L/h
20.43 (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 → 2: 1422 J, 405 J; 2 → 3: −1355 J, 0; 3 → 4: −274 J, −274 J; 4 → 1: 339 J, 0
(c) 131 J (d) 7.44%; $e = 0.168e_C$
20.45 $1 - T_C/T_H$, same
20.47 (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%
20.49 6.23
20.53 (a) A: 28.9%, B: 38.3%, C: 53.8%, D: 24.4% (b) C (c) B > D > A
20.55 (a) 4.83% (b) 4.83% (c) 6.25%
(d) $e = \dfrac{0.80T_d - 200}{12T_d - 2700}$, 6.67%
20.57 (a) 151 K (b) 305 Pa (c) 5.39 kJ (d) 4.39 kJ
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 2.58 μN, in the $-y$-direction
21.19 Attractive
21.21 (a) 4.40×10^{-16} N (b) 2.63×10^{11} m/s^2
(c) 2.63×10^5 m/s
21.23 (a) 3.30×10^6 N/C, to the left (b) 14.2 ns
(c) 1.80×10^3 N/C, to the right
21.25 (a) -21.9 μC (b) 1.02×10^{-7} N/C
21.27 (a) 364 N/C
(b) no; 2.73 μm, downward
21.29 (a) $-\hat{\jmath}$ (b) $\frac{\sqrt{2}}{2}\hat{\imath} + \frac{\sqrt{2}}{2}\hat{\jmath}$ (c) $-0.39\,\hat{\imath} + 0.92\,\hat{\jmath}$
21.31 3.09×10^{-3} N in the $+y$-direction
21.33 20.0 cm
21.35 (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
21.37 1.04×10^7 N/C, toward the -2.00-μC charge
21.39 (a) 8740 N/C, to the right (b) 6540 N/C, to the right (c) 1.40×10^{-15} N, to the right
21.41 1.73×10^{-8} N, toward the point midway between the electrons
21.43 (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
21.45 113 N/C, $-x$ direction
21.47 (a) 1340 N/C, $+x$-direction
(b) 1340 N/C, $-x$-direction
(c) 8260 N/C, $+x$-direction
(d) 8260 N/C, $-x$-direction
21.49 (a) (7.0 N/C)$\hat{\imath}$ (b) (1.75×10^{-5} N)$\hat{\imath}$
21.51 (a) 1.4×10^{-11} C·m, from q_1 toward q_2
(b) 860 N/C
21.53 (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
21.55 (a) 1680 N, from the $+5.00$-μC charge toward the -5.00-μC charge
(b) 22.3 N·m, clockwise
21.57 (b) $\frac{Q^2}{8\pi\epsilon_0 L^2}(1 + 2\sqrt{2})$, away from the center of the square
21.59 (a) 8.63×10^{-5} N, -5.52×10^{-5} N
(b) 1.02×10^{-4} N, 32.6° below the $+x$-axis
21.61 (b) 2.80 μC (c) 39.5°
21.63 3.41×10^4 N/C, to the left
21.65 (a) 1 nC (b) 6×10^{10}
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) $\frac{6q^2}{4\pi\epsilon_0 L^2}$, away from the vacant corner
(b) $\frac{3q^2}{4\pi\epsilon_0 L^2}\left(\sqrt{2} + \frac{1}{2}\right)$, 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) $\frac{mv_0^2\sin^2\alpha}{2eE}$ (b) $\frac{mv_0^2\sin^2\alpha}{eE}$
(d) h_{max}: 0.418 m, d: 2.89 m
21.79 (a) $E_x = \frac{Q}{4\pi\epsilon_0 a}\left(\frac{1}{x-a} - \frac{1}{x}\right)$, $E_y = 0$
(b) $\frac{qQ}{4\pi\epsilon_0 a}\left(\frac{1}{r} - \frac{1}{r+a}\right)\hat{\imath}$
21.81 (a) 1.56 N/C, $+x$-direction (c) smaller
(d) 4.7%
21.83 $E_x = E_y = \frac{Q}{2\pi^2\epsilon_0 a^2}$
21.85 (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
21.87 (a) $\pi(R_2^2 - R_1^2)\sigma$
(b) $\frac{\sigma}{2\epsilon_0}\left(\frac{1}{\sqrt{(R_1/x)^2+1}} - \frac{1}{\sqrt{(R_2/x)^2+1}}\right)\frac{|x|}{x}\hat{\imath}$
(c) $\frac{\sigma}{2\epsilon_0}\left(\frac{1}{R_1} - \frac{1}{R_2}\right)x\hat{\imath}$; $x \ll R_1$
(d) $\frac{1}{2\pi}\sqrt{\frac{q\sigma}{2\epsilon_0 m}\left(\frac{1}{R_1} - \frac{1}{R_2}\right)}$
21.89 (a) $q_1 = 8.00$ μC, $q_2 = 3.00$ μC
(b) 7.50 N, in the $-x$-direction (c) $x = 0.248$ m
21.91 (a) $\frac{\sigma}{2\epsilon_0}\left[\frac{1}{\sqrt{1+(R/\Delta)^2}}\right]$ downward (b) $q\sigma/2g\epsilon_0$
(c) $-\frac{q\sigma}{2\epsilon_0}\left(\sqrt{\Delta^2+R^2} - R\right)$ (d) $mg\Delta$
(e) $\frac{2R\alpha}{1-\alpha^2}$ (f) $\sqrt{2g\left(\Delta - \frac{\sqrt{\Delta^2+R^2}-R}{\alpha}\right)}$
(g) 57.7 g, 10.7 cm
21.93 (b) $-\frac{\lambda}{4\pi\epsilon_0 a}\left[\frac{y_2 - L}{\sqrt{(y_2-L)^2+a^2}} - \frac{y_2}{\sqrt{y_2^2+a^2}}\right]\lambda dy_2$
(c) $\frac{Q^2}{2\pi\epsilon_0 L^2}\left(\sqrt{(L/a)^2+1} - 1\right)\hat{\imath}$
(e) $\frac{Q^2}{2\pi\epsilon_0 L^2}\left[L\ln\left(\frac{L+\sqrt{a^2+L^2}}{a}\right) + a - \sqrt{a^2+L^2}\right]$
(f) 21.5 m/s
21.95 (b) $q_1 < 0$, $q_2 > 0$ (c) 0.843μC (d) 56.2 N
21.97 (a) $\frac{Q}{2\pi\epsilon_0 L}\left(\frac{1}{2x+a} - \frac{1}{2L+2x+a}\right)$
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 7.70×10^3 N/C
22.7 14.8 cm^2
22.9 (a) 0 (b) 1.22×10^8 N/C, radially inward
(c) 3.64×10^7 N/C, radially inward
22.11 0.977 N·m^2/C, inward
22.13 0.0810 N
22.15 (a) 350 N/C (b) 700 N/C (c) 1400 N/C
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 (a) $2\pi R\sigma$ (b) $\frac{\sigma R}{\epsilon_0 r}$ (c) $\frac{\lambda}{2\pi\epsilon_0 r}$
22.29 1.16 km/s
22.31 10.2°
22.33 (a) 750 N·m^2/C
(b) 0 (c) 577 N/C, $+x$-direction
(d) within and outside
22.35 (a) $\frac{Q}{4\pi\epsilon_0 r^2}$, toward the center of the shell (b) 0
22.37 (a) $\frac{\lambda}{2\pi\epsilon_0 r}$, radially outward
(b) $\frac{\lambda}{2\pi\epsilon_0 r}$, radially outward
(d) $-\lambda$ (inner), $+\lambda$ (outer)
22.39 (a) $\frac{\rho r}{2\epsilon_0}$ (b) $\frac{\lambda}{2\pi\epsilon_0 r}$ (c) They are equal.
22.41 (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
22.43 (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) $+6q$
22.45 (a) $\frac{\alpha}{2\epsilon_0}\left(1 - \frac{a^2}{r^2}\right)$ (b) $q = +2\pi\alpha a^2$, $E = \frac{\alpha}{2\epsilon_0}$
22.47 (c) $\frac{Qr}{4\pi\epsilon_0 R^3}\left(4 - \frac{3r}{R}\right)$ (e) $2R/3$, $\frac{Q}{3\pi\epsilon_0 R^2}$
22.49 $\frac{\alpha r}{\epsilon_0}\left(\frac{1}{2} - \frac{r}{3R}\right)$, $\frac{\alpha R^2}{6\epsilon_0 r}$, yes
22.51 $|x| > d$ (outside the slab): $\frac{\rho_0 d}{3\epsilon_0}\frac{x}{|x|}\hat{\imath}$;
$|x| < d$ (inside the slab): $\frac{\rho_0 x^3}{3\epsilon_0 d^2}\hat{\imath}$
22.53 (b) $\frac{\rho\vec{b}}{3\epsilon 0}$
22.55 (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
22.57 (i) 377 N/C (ii) 653 N/C (iii) 274 N/C
(iv) 0
22.59 (a) $\frac{\pi\rho_0 R^2}{3}$ (b) $T = 2\pi\left(\frac{R_{\text{orbit}}}{R_{\text{cylinder}}}\right)\sqrt{\frac{6\epsilon_0 M}{Q\rho_0}}$
22.61 (a) $\frac{QL}{2\epsilon_0\sqrt{(L/2)^2+R^2}}$
(b) $\frac{QL}{4\epsilon_0}\left(\frac{2}{L} - \frac{1}{\sqrt{R^2+(L/2)^2}}\right)$
(c) $\frac{QL}{4\epsilon_0}\left(\frac{2}{L} - \frac{1}{\sqrt{R^2+(L/2)^2}}\right)$ (d) Q/ϵ_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 89.3 V, point a
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 (a) negative (b) 171 V/m (c) -17.1 V
23.23 (a) b (b) 800 V/m (c) -48.0 μJ
23.25 (a) 0.0539 J (b) 3.00 m/s
23.27 (a) (i) 180 V (ii) -270 V (iii) -450 V
(b) 719 V, inner shell
23.29 (a) oscillatory (b) 1.67×10^7 m/s
23.31 150 m/s
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) $V = \dfrac{q}{4\pi\epsilon_0}\left(\dfrac{1}{r_a} - \dfrac{1}{r_b}\right)$
(ii) $V = \dfrac{q}{4\pi\epsilon_0}\left(\dfrac{1}{r} - \dfrac{1}{r_b}\right)$ (iii) $V = 0$
(d) 0 (e) $E = \dfrac{q - Q}{4\pi\epsilon_0 r^2}$
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) $-0.360\ \mu$J (b) $x = 0.074$ m
23.49 $+1.90\ 10^5$ V, point A
23.51 4.2×10^6 V
23.53 (a) $-21.5\ \mu$J (b) -2.83 kV (c) 35.4 kV/m
23.55 (a) $7.85 \times 10^4\ \text{V/m}^{4/3}$
(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) $-\dfrac{1.46q^2}{\pi\epsilon_0 d}$
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) $1.93 \times 10^{14}\ \text{m/s}^2$, 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 $\frac{3}{5}\left(\dfrac{Q^2}{4\pi\epsilon_0 R}\right)$
23.69 360 kV
23.71 (a) $\dfrac{Q}{4\pi\epsilon_0 a}\ln\left(\dfrac{x + a}{x}\right)$
(b) $\dfrac{Q}{4\pi\epsilon_0 a}\ln\left(\dfrac{a + \sqrt{a^2 + y^2}}{y}\right)$
(c) (a): $\dfrac{Q}{4\pi\epsilon_0 x}$, (b): $\dfrac{Q}{4\pi\epsilon_0 y}$
23.73 (a) $U = \dfrac{1}{2}kd^2 + \dfrac{1}{4\pi\epsilon_0}\dfrac{e^2}{d}$
(b) $d_0 = \left(\dfrac{e^2}{4\pi\epsilon_0 k}\right)^{1/3}$ (c) 2.30×10^{17} N/m
(d) 2.16 MeV (e) 1.44×10^7 m/s
23.75 2.48×10^{-14} m
23.77 (a) $A = -6.0\ \text{V/m}^2$, $B = -4.0\ \text{V/m}^3$, $C = -2.0\ \text{V/m}^6$, $D = 10$ 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 23.79
23.79 (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 23.81
23.81 (a) 4.75×10^{-15} C (b) 94.9 kV/m
(c) 5.81×10^{-15} C (d) 36,300
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 $\mu\text{C/m}^2$
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 5.57 μF
24.15 (a) series (b) 5000
24.17 (a) $Q_1 = Q_2 = 22.4\ \mu$C, $Q_3 = 44.8\ \mu$C, $Q_4 = 67.2\ \mu$C
(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) $Q_1 = 156\ \mu$C, $Q_2 = 260\ \mu$C
(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) $Q_{35} = 7.7\ \mu$C, $Q_{75} = 16.5\ \mu$C
(c) 2.66 mJ (d) $U_{35} = 0.85$ mJ, $U_{75} = 1.81$ mJ
(e) 220 V
24.31 (a) 3.60 mJ (before), 13.5 mJ (after)
(b) 9.9 mJ, increase
24.33 (a) 0.620 $\mu\text{C/m}^2$ (b) 1.28
24.35 $\dfrac{2KQ_0}{1 + K}$, increases
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) $\dfrac{Q}{\epsilon_0 AK}$ (b) $\dfrac{Qd}{\epsilon_0 AK}$ (c) $K\dfrac{\epsilon_0 A}{d} = KC_0$
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) $Q_1 = 550\ \mu$C, $Q_2 = 370\ \mu$C,
$Q_3 = Q_4 = 180\ \mu$C, $Q_5 = 550\ \mu$C;
$V_1 = 65$ V, $V_2 = 87$ V,
$V_3 = V_4 = 43$ V, $V_5 = 65$ V
24.55 $C_2 = 6.00\ \mu$F, $C_3 = 4.50\ \mu$F
24.57 (a) 76 μC (b) 1.4 mJ (c) 11 V (d) 1.3 mJ
24.59 (a) 2.3 μF (b) $Q_1 = 970\ \mu$C, $Q_2 = 640\ \mu$C
(c) 47 V
24.61 1.67 μF
24.63 (a) $\dfrac{Q^2}{2A\epsilon_0 K}$ (b) $\dfrac{d_0}{2A^2\epsilon_0 YK}$ (c) $7.53 \times 10^7\ \text{m/C}^2$
(d) 95.7 kV (e) 133 kV
24.65 (a) $\dfrac{\pi\epsilon_0 r^2 V^2}{2z^2}$ (b) $\dfrac{\pi\epsilon_0 r^2 V^2}{4d}$ (c) $\dfrac{\pi\epsilon_0 r^2 V^2}{2d}$
(d) $\dfrac{\pi\epsilon_0 r^2 V^2}{4d}$ (e) no
24.67 (a) $C_1 = 6.00\ \mu$F, $C_2 = 3.00\ \mu$F
(b) same charge; C_2 stores more energy
(c) C_1 stores more charge and energy
24.69 (a) first (connected) (b) 144 cm^2
(c) disconnected
24.71 (a) $\dfrac{\epsilon_0 A(K - 1)}{d\ln K}$
24.73 Choice (c)
24.75 Choice (a)

Chapter 25

25.1 1.0 C
25.3 (a) 3.12×10^{19} (b) $1.51 \times 10^6\ \text{A/m}^2$
(c) 0.111 mm/s (d) both (b) and
(c) would increase
25.5 (a) 0.26 mm/s (b) 0.0120 N/C
25.7 (a) 330 C (b) 41 A
25.9 9.0 μA
25.11 $4.02 \times 10^{-8}\ \Omega \cdot \text{m}$
25.13 (a) $1.06 \times 10^{-5}\ \Omega \cdot \text{m}$ (b) 0.00105 $(\text{C}°)^{-1}$
25.15 (a) 0.206 mV (b) 0.176 mV
25.17 0.457
25.19 0.125 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 21 W, 21 W
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) $3.65 \times 10^{-8}\ \Omega \cdot \text{m}$ (b) 172 A
(c) 2.58 mm/s
25.53 0.060 Ω
25.55 (a) 2.5 mA (b) 21.4 μV/m (c) 85.5 μV/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) 0.20 Ω (b) 8.7 V
25.63 (a) 1.0 kΩ (b) 100 V (c) 10 W
25.65 (a) \$78.90 (b) \$140.27
25.67 (a) 171 $\mu\Omega$ (b) 176 μV/m (c) left: 54.7 $\mu\Omega$; right: 116 $\mu\Omega$
25.69 (a) 4.25 ppt (b) 510 pC
(c) 1.33 W (d) 1.60 ppt
25.71 6.67 V
25.73 (b) no (c) yes (d) 9.40 W (e) 4.12 W
25.75 (a) inward (b) $\dfrac{V}{\ln(r_{\text{outer}}/r_{\text{inner}})}$
(c) $\dfrac{\rho}{2\pi L}\ln(r_{\text{outer}}/r_{\text{inner}})$
(d) 616 $\Omega \cdot \text{m}$
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) $I = \dfrac{\mathcal{E}}{R}$ (b) $I = \dfrac{\mathcal{E}}{6R}$ (c) parallel
26.11 (a) $I_1 = 8.00$ A, $I_3 = 12.0$ A
(b) 84.0 V
26.13 5.00 Ω; $I_{3.00} = 8.00$ 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) 2.00 A (b) 5.00 Ω (c) 42.0 V (d) 3.50 A

26.25 (a) 8.00 A (b) $\mathcal{E}_1 = 36.0$ V, $\mathcal{E}_2 = 54.0$ V (c) 9.00 Ω

26.27 (a) 1.60 A (top), 1.40 A (middle), 0.20 A (bottom) (b) 10.4 V

26.29 (a) 36.4 V (b) 0.500 A

26.31 (a) 2.14 V, a (b) 0.050 A, 0; down

26.33 (a) 0.641 Ω (b) 975 Ω

26.35 (a) 17.9 V (b) 22.7 V (c) 21.4%

26.37 (a) 0.849 μF (b) 2.89 s

26.39 (a) 0 (b) 245 V (c) 0 (d) 32.7 mA (e) (a): 245 V; (b): 0; (c): 1.13 mC; (d): 0

26.41 (a) 4.21 ms (b) 0.125 A

26.43 192 μC

26.45 13.6 A

26.47 (a) 0.937 A (b) 0.606 A

26.49 (a) 165 μC (b) 463 Ω (c) 12.6 ms

26.51 900 W

26.53 (a) $U_C = \frac{1}{2}C\mathcal{E}^2$ (b) $U_\mathcal{E} = C\mathcal{E}^2$ (c) $U_R = \frac{1}{2}C\mathcal{E}^2$ (d) 1/2, 1/2

26.55 (a) 2.29 A (b) 3.71 A (c) 1.43 A

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

26.63 (a) 109 V; no (b) 13.5 s

26.67 (a) −12.0 V (b) 1.71 A (c) 4.21 Ω

26.69 (a) $\epsilon_0\rho K$ (b) 1.28 nC (c) 165 s (d) 2.68 pA

26.71 (a) 11.8 pC (b) 0.917 A (c) 4.92 pJ (d) 703 pJ (e) 1300 pJ (f) 595 pJ

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 μC

26.77 1.7 MΩ, 3.1 μF

26.79 (a) −1.23 ms (slope), 79.5 μC (y-intercept) (b) 247 Ω, 15.9 V (c) 1.22 ms (d) 11.9 V

26.81 (a) $V_{\text{out}} = 21\text{V} - 10V_{\text{in}}$ (b) 1.35 V (c) −10

26.85 (b) 4 (c) 3.2 MΩ, 4.0×10^{-3} (d) 3.4×10^{-4} (e) 0.88

26.87 Choice (d)

Chapter 27

27.1 (a) $(-6.68 \times 10^{-4}\ \text{N})\hat{\boldsymbol{k}}$ (b) $(6.68 \times 10^{-4}\ \text{N})\hat{\imath} + (7.27 \times 10^{-4}\ \text{N})\hat{\jmath}$

27.3 (a) positive (b) 0.0505 N

27.5 172 m/s, $-x$-direction

27.7 (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

27.9 (a) 3.05 mWb (b) 1.83 mWb (c) 0

27.11 −0.78 mWb

27.13 (a) 0 (b) −0.0115 Wb (c) +0.0115 Wb (d) 0

27.15 (a) 0.160 mT, into the page (b) 0.111 μs

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 0.838 mT

27.23 (a) 7900 N/C, $\hat{\imath}$ (b) 7900 N/C, $\hat{\imath}$

27.25 0.0445 T, out of the page

27.27 (a) 4.92 km/s (b) 9.96×10^{-26} kg

27.29 0.724 N, 63.4° below the current direction in the upper wire segment

27.31 (a) 817 V (b) 113 m/s^2

27.33 (a) a (b) 3.21 kg

27.35 (a) 55 turns (b) counterclockwise (c) 8.73 N·m, upward

27.37 (b) $F_{cd} = 1.20$ N (c) 0.420 N·m

27.39 (a) A_2 (b) 290 rad/s^2

27.41 (a) $-NIAB\hat{\imath}$, 0 (b) 0, $-NIAB$ (c) $+NIAB\hat{\imath}$, 0 (d) 0, $+NIAB$

27.43 (a) 1.13 A (b) 3.69 A (c) 98.2 V (d) 362 W

27.45 (a) 4.7 mm/s (b) $+4.5 \times 10^{-3}$ V/m, $+z$-direction (c) 53 μV

27.47 (a) $x = 0.565$ m, $y = 0.825$ m, $z = 0.0900$ m (b) 31.3 m/s

27.49 (a) 8.3×10^6 m/s (b) 0.14 T

27.51 4.81×10^7 C/kg

27.53 1.6 mm

27.55 $\dfrac{Mg\tan\theta}{LB}$, right to left

27.57 (a) 8.46 mT (b) 27.2 cm (c) 2.2 cm; yes

27.59 (a) ILB, to the right (b) $\dfrac{v^2m}{2ILB}$ (c) 1960 km

27.61 (a) 1.13 N, toward the origin (b) 0.800 N, $+z$-direction

27.63 (a) 48,000 A (b) not feasible (c) 19 G (d) 0.83 N

27.65 0.024 T, $+y$-direction

27.67 (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

27.69 $-(0.444\ \text{N})\hat{\jmath}$

27.71 (b) left: $(B_0LI/2)\hat{\imath}$; top: $-IB_0L\hat{\jmath}$; right: $-(B_0LI/2)\hat{\imath}$; bottom: 0 (c) $-IB_0L\hat{\jmath}$

27.73 (a) $\dfrac{Mg}{2WNB}$ (b) $R(1 + \pi/2)$ (c) $Mg(\sigma - 1)\sqrt{h(2R - h)}$ (d) $\left[\sigma\cos\left(\dfrac{h}{R} - 1\right) - 1\right]MgR$ (e) $(1 - \sigma)Mgh$ (f) $\left\{\dfrac{h}{R} - \sigma\left[\sin\left(\dfrac{h}{R} - 1\right) + 1\right]\right\}(MgR)$ (g) $R[1 + \arccos(1/\sigma)]$ (h) $\sigma < \dfrac{1}{2}\left(1 + \dfrac{\pi}{2}\right)$

27.75 (a) $-IA\hat{\boldsymbol{k}}$ (b) $B_x = \dfrac{3D}{IA}, B_y = \dfrac{4D}{IA}, B_z = -\dfrac{12D}{IA}$

27.77 (b) 1.85×10^{-28} kg (c) 1.20 kV (d) 8.32×10^5 m/s

27.79 (a) $r = R\sin\theta$ (b) $dI = \sigma\omega R^2\sin\theta d\theta$ (c) $d\mu = \pi R^4\sigma\omega\sin^3\theta d\theta$ (d) $\dfrac{Q\omega R^2}{3}\hat{\boldsymbol{k}}$ (e) $\dfrac{Q\omega BR^2}{3}(\sin\alpha\hat{\jmath} - \cos\alpha\hat{\imath})B$

27.81 (a) 5.14 m (b) 1.72 μs (c) 6.08 mm (d) 3.05 cm

27.83 Choice (c)

27.85 Choice (a)

Chapter 28

28.1 (a) $-(19.2\ \mu\text{T})\hat{\boldsymbol{k}}$ (b) 0 (c) $(19.2\ \mu\text{T})\hat{\imath}$ (d) $(6.79\ \mu\text{T})\hat{\imath}$

28.3 (a) 60.0 nT, out of the page at A and B (b) 0.120 μT, out of the page (c) 0

28.5 (a) 0 (b) $-(1.31\ \mu\text{T})\hat{\boldsymbol{k}}$ (c) $-(0.462\ \mu\text{T})\hat{\boldsymbol{k}}$ (d) $(1.31\ \mu\text{T})\hat{\jmath}$

28.7 (a) left (b) 0 (c) 1.21 pT, out of the page (d) 1.21 pT, into the page

28.9 (a) 0.440 μT, out of the page (b) 16.7 nT, out of the page (c) 0

28.11 (a) $(50.0\ \text{pT})\hat{\jmath}$ (b) $-(50.0\ \text{pT})\hat{\imath}$ (c) $-(17.7\ \text{pT})(\hat{\imath} - \hat{\jmath})$ (d) 0

28.13 17.6 μT, into the page

28.15 (a) 0.8 mT (b) 40 μT (20 times larger)

28.17 25 μA

28.19 (a) $-(0.10\ \mu\text{T})\hat{\imath}$ (b) 2.19 μT, at 46.8° from the $+x$-axis to the $+z$-axis (c) $(7.9\ \mu\text{T})\hat{\imath}$

28.21 (a) 0 (b) 6.67 μT, toward the top of the page (c) 7.54 μT, to the left

28.23 (a) 0 (b) 0 (c) 0.40 mT, to the left

28.25 (a) P: 41 μT, into the page; Q: 25 μT, out of the page (b) P: 9.0 μT, out of the page; Q: 9.0 μT, into the page

28.27 $\dfrac{\mu_0 I^2}{2\pi g\lambda}$

28.29 (a) 6.00 μN; repulsive (b) 24.0 μN

28.31 0.38 μA

28.33 $\dfrac{\mu_0|I_1 - I_2|}{4R}$; 0

28.35 18.0 A, counterclockwise

28.37 (a) 305 A (b) -3.83×10^{-4} T·m

28.39 (a) $\mu_0 I/2\pi r$ (b) 0

28.41 (a) $\dfrac{\mu_0 I_1}{2\pi r}$ (b) $\dfrac{\mu_0(I_1 + I_2)}{2\pi r}$

28.43 (a) 1790 turns per meter (b) 63.0 m

28.45 (a) 3.72 MA (b) 124 kA (c) 237 A

28.47 (a) (i) 1.13 mT (ii) 4.68 MA/m (iii) 5.88 T

28.49 (a) 1.00 μT, into the page (b) $(74.9\ \text{nN})\hat{\jmath}$

28.51 0.200 m

28.53 (a) $\mu_0 I^2R/d$ (b) $\dfrac{4B^2}{\mu_0 d}\left(\dfrac{A}{\pi}\right)^{3/2}$ (c) $\dfrac{1}{2}\sqrt{\mu_0 Fd}\left(\dfrac{\pi}{A}\right)^{3/4}$

28.55 (a) 5.7×10^{12} m/s^2, away from the wire (b) 32.5 N/C, away from the wire (c) no

28.59 (a) 2.00 A, out of the page (b) 2.13 μT, to the right (c) 2.06 μT

28.61 23.2 A

28.63 (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

28.65 (a) $\dfrac{3I}{2\pi R^3}$ (b) (i) $B = \dfrac{\mu_0 Ir^2}{2\pi R^3}$ (ii) $B = \dfrac{\mu_0 I}{2\pi r}$

28.67 (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) $B = \dfrac{\mu_0 I_0 r}{2\pi a^2}\left(2 - \dfrac{r^2}{a^2}\right)$

28.69 (a) $B = \mu_0 In/2$, $+x$-direction (b) $B = \mu_0 In/2$, $-x$-direction

28.71 (a) $\dfrac{Q_1\omega_1}{2\pi}$ (b) $\dfrac{\mu_0\omega_1 Q_1}{2\pi H}\hat{\boldsymbol{k}}$ (c) $\dfrac{\mu_0 Q_1Q_2\omega_1\omega_2R_2^2}{8\pi H}\sin\theta$ (d) $\dfrac{1}{2}MR_2^2\omega_2$ (e) $\dfrac{\mu_0}{4\pi H}\dfrac{Q_1Q_2\omega_1\sin\theta}{M}$

28.73 (a) no (c) 64.9 A, 1.19 cm

28.75 (a) nq (b) $n^{2/3}q$ (c) $\frac{1}{2}\mu_0 nqvR\,\hat{\boldsymbol{\phi}}$. (d) $n^{2/3}qvRd\phi$ (e) $\dfrac{\mu_0}{2}n^{5/3}q^2v^2R^2Ld\phi$ (f) $\pi\mu_0 n^{5/3}q^2v^2R^2L$, $\dfrac{\mu_0}{2}n^{5/3}q^2v^2R$ (g) 1.9×10^{-5} Pa

28.77 (a) $\dfrac{Q\omega}{4\pi}\sin\theta d\theta$ (b) $\dfrac{1}{4}Q\omega R^2\sin^3\theta\, d\theta$, $+z$-direction (c) $\dfrac{Q\omega R^2}{3}$, $+z$-direction (d) $\dfrac{2}{3}MR^2\omega$ (e) $\dfrac{Q}{2M}$ (f) 1

28.79 (b) $\dfrac{1}{2g}\left(\dfrac{\mu_0 Q_0^2}{4\pi\lambda RCd}\right)^2$

28.81 Choice (b)

28.83 Choice (c)

Chapter 29

29.1 (a) 17.1 mV (b) 28.5 mA
29.3 (a) Wb/s, Wb/s^3 (b) $\alpha = (0.750\ \text{s}^2)\beta$ (c) -1.20 V
29.5 (a) 34 V (b) counterclockwise
29.7 (a) $\mu_0 i/2\pi r$, into the page (b) $\frac{\mu_0 i}{2\pi r}L dr$ (c) $\frac{\mu_0 iL}{2\pi}\ln(b/a)$ (d) $\frac{\mu_0 L}{2\pi}\ln(b/a)\frac{di}{dt}$ (e) 0.506 μV
29.9 (a) 5.44 mV (b) clockwise
29.11 (a) 12.6 mA (b) 626 μA
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 (a) 0.72 V (b) 0
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 (a) 3.00 V (b) b to a (c) 0.800 N, to the right (d) 6.00 W for each
29.29 (a) counterclockwise (b) 42.4 mW
29.31 35.0 m/s, to the right
29.33 (a) 0.225 A, clockwise (b) 0 (c) 0.225 A, counterclockwise
29.35 9.21 A/s
29.37 0.950 mV
29.39 (a) 0.599 nC (b) 6.00 mA (c) 6.00 mA
29.41 (a) 2.2 nA (b) 0.75 s
29.43 (a) 2.5×10^5 C (b) 65 A (c) 11 T
29.45 (a) 3.7 A (b) 1.33 mA (c) counterclockwise
29.47 16.2 μV
29.49 (a) $\frac{\mu_0 Iabv}{2\pi r(r+a)}$ (b) clockwise
29.51 (a) 17.9 mV (b) a to b
29.53 $\mu_0 IW/4\pi$
29.55 (a) $\frac{\mu_0 Iv}{2\pi}\ln(1 + L/d)$ (b) a (c) 0
29.57 (a) 0.165 V (b) 0.165 V (c) 0; 0.0412 V
29.59 (a) B^2L^2v/R
29.61 a: $\frac{qr}{2}\frac{dB}{dt}$, to the left; b: $\frac{qr}{2}\frac{dB}{dt}$, toward the top of the page; c: 0
29.63 5.0 s
29.65 (a) $0.3071\ \text{s}^{-1}$ (b) 3.69 T (c) a (d) 2.26 s
29.67 (a) $\vec{\boldsymbol{v}} - (1.26\ \text{m/s})\sin\theta\,\hat{\boldsymbol{\imath}}$ (b) $-1.26\sin\theta\ \text{T}\cdot\text{m/s}\,\hat{\boldsymbol{\jmath}}$ (c) $(10.0\ \text{cm})(\cos\theta\,\hat{\boldsymbol{\jmath}} - \sin\theta\,\hat{\boldsymbol{k}})d\theta$ (d) 4.73 A (e) upward
29.69 (a) a to b (b) $\frac{Rmg\tan\phi}{L^2B^2\cos\phi}$ (c) $\frac{mg\tan\phi}{LB}$ (d) $\frac{Rm^2g^2\tan^2\phi}{L^2B^2}$ (e) $\frac{Rm^2g^2\tan^2\phi}{L^2B^2}$; same
29.71 Choice (c)
29.73 Choice (a)

Chapter 30

30.1 (a) 0.270 V; yes (b) 0.270 V
30.3 (a) 1.96 H (b) 7.11 mWb
30.5 (a) 1940 (b) 800 A/s
30.7 (a) 0.250 H (b) 0.450 mWb
30.9 (a) 4.68 mV (b) a
30.11 (b) 0.111 μH
30.13 0.0300 T
30.15 2850
30.17 (a) 0.161 T (b) 10.3 kJ/m^3 (c) 0.129 J (d) 40.2 μH
30.19 91.7 J
30.21 (a) 2.40 A/s (b) 0.800 A/s (c) 0.413 A (d) 0.750 A
30.23 (a) 17.3 μs (b) 30.7 μs
30.25 15.4 V
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) 0 (b) 0 (c) $\mathcal{E}^2/4R$
30.33 (a) 443 nC (b) 358 nC
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
30.43 (a) 59 NH (b) 26.8 μV
30.45 20 km/s; about 30 times smaller
30.47 (a) 5.00 H (b) 31.7 m; no
30.49 222 μF, 9.31 μH
30.51 (a) 111 N (b) 37.1 ms (c) $4.00 \times 10^3\ \text{rad/s}^2$
30.53 (a) 15.0 mC (b) 93.8 μS (c) 2.67 kHz (d) 11.3 J (e) 1.30 ms
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 (b) 5.0 Ω, 8.5 H (c) 1.7 kJ; 2.0 kW
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) $v_L = \frac{\mathcal{E}}{R + R_L}(R_L + Re^{-(R+R_L)t/L})$ (b) 50.0 V (c) 30.0 V; 3.00 A (d) 6.67 Ω (e) 40.0 mH
30.69 (a) $\frac{\mu_0 I(N_1 - N_2)\pi a^2}{\lambda}$ (b) $\frac{\mu_0 I\pi}{\lambda}(N_1 b^2 - N_2 a^2)$ (c) $\frac{\mu_0\pi}{\lambda}(N_1^2 b^2 - N_2^2 a^2)$ (d) $\frac{\mu_0\pi}{\lambda}(N_1^2 b^2 + 2N_1N_2a^2 + N_2^2a^2)$ (e) 10.3 mH (f) 16.0 mH
30.71 (a) $i_1 = \frac{\mathcal{E}}{R_1}(1 - e^{-R_1 t/L})$, $i_2 = \frac{\mathcal{E}}{R_2}e^{-t/R_2C}$, $q_2 = \mathcal{E}C(1 - e^{-t/R_2C})$ (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) 90°; lead (b) 193 Hz
31.5 (a) 1510 Ω (b) 0.239 H (c) 497 Ω (d) 16.6 μF
31.7 (a) 1700 Ω (b) 106 Ω (c) 150 Hz
31.9 (a) $(12.5\ \text{V})\cos[(480\ \text{rad/s})t]$ (b) 7.17 V
31.11 (a) $i = (0.0253\ \text{A})\cos[(720\ \text{rad/s})t]$ (b) 180 Ω (c) $v_L = -(4.56\ \text{V})\sin[(720\ \text{rad/s})t]$
31.13 (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
31.15 50.0 V
31.17 32.8°
31.19 (a) 40.0 W (b) 0.167 A (c) 720 Ω
31.21 (b) 32.4 Ω
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 $\frac{V_{out}}{V_s} = \sqrt{\frac{R^2 + \omega^2L^2}{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}$
31.51 (a) 102 Ω (b) 0.882 A (c) 270 V
31.53 (a) $C = 5\epsilon_0 a^2\theta/g$ (b) 131° (c) 7.41 V (d) 1230 kHz (e) 7.74 V
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) $\frac{V}{R}$ (b) $\frac{V}{R}\sqrt{\frac{L}{C}}$ (c) $\frac{V}{R}\sqrt{\frac{L}{C}}$ (d) $\frac{1}{2}L\frac{V^2}{R^2}$ (e) $\frac{1}{2}L\frac{V^2}{R^2}$
31.63 20.0 Ω, 0.180 H
31.65 (a) $\frac{V_{in}}{\sqrt{R^2 + (\omega L)^2}}$ (b) $\arctan\left(\frac{\omega L}{R}\right)$ (c) $\frac{1}{\sqrt{1 + (R/\omega L)^2}}$ (d) $\frac{R}{2.00\pi L}$ (e) 1.6 mH
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) (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
32.9 (a) $+y$-direction (b) 0.149 mm (c) $\vec{\boldsymbol{B}} = (1.03\ \text{mT})\cos[(4.22 \times 10^4\ \text{rad/m})y - (1.265 \times 10^{13}\ \text{rad/s})t]\hat{\boldsymbol{\imath}}$
32.11 (a) 361 m (b) 0.0174 rad/m (c) 5.22×10^6 rad/s (d) 0.0144 V/m
32.13 (a) 0.381 μm (b) 0.526 μm (c) 1.38 (d) 1.90
32.15 (a) 330 W/m^2 (b) 500 V/m, 1.7 μT
32.17 2.5×10^{25} W
32.19 93.8 W
32.21 12.0 V/m, 40.0 nT
32.23 (a) 0.18 mW (b) 274 V/m, 0.913 μT (c) 0.18 mJ/s (d) 0.010 W/cm^2

32.25 (a) 637 W/m^2 (b) 693 V/m, 2.31 μT
(c) 2.12 $\mu J/m^3$
32.27 (a) 30.5 cm (b) 2.46 GHz (c) 2.11 GHz
32.29 7.50 cm
32.31 (a) 0.375 mJ (b) 4.08 mPa
(c) 604 nm, 3.70×10^{14} Hz
(d) 30.3 kV/m, 101 μT
32.33 (a) 6.02×10^{-3} W/m^2 (b) 2.13 V/m, 7.10 nT
(c) 1.20 pN; no
32.35 (a) at $r = R$: 64 MW/m^2, 0.21 Pa; at $r = R/2$: 260 MW/m^2, 0.85 Pa (b) no
32.37 3.89×10^{-13} rad/s^2
32.39 (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) $\dfrac{\rho l I^2}{\pi a^2} = I^2 R$
32.41 (a) 1.364 m (b) 10.91 m
32.43 (a) 9.75×10^{-15} W/m^2
(b) 2.71 $\mu V/m$, 9.03×10^{-15} T, 67.3 ms
(c) 3.25×10^{-23} Pa (d) 0.190 m
32.45 (a) $\dfrac{4\rho G\pi M R^3}{3r^2}$ (b) $\dfrac{LR^2}{4cr^2}$ (c) 0.19 μm; no
32.47 (b) 3.00×10^8 m/s
32.49 (a) $\dfrac{I_0}{c}(2 - e)$ (b) 3.3×10^{-16} N (c) 0.55
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 2.11×10^8 m/s, 417 nm, 5.07×10^{14} Hz
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 38.5°
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 (a) red: 1.36, violet: 1.40 (b) red: 2.21×10^8 m/s, violet: 2.14×10^8 m/s
33.27 $0.375 I_0$
33.29 (a) A: $I_0/2$, B: $0.125 I_0$, C: $0.0938 I_0$ (b) 0
33.31 (a) 1.40 (b) 35.5°
33.33 6.38 W/m^2
33.35 (a) $0.374 I$ (b) $2.35 I$
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 23.3°
33.55 (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
33.57 (a) 35° (b) $I_0 = 10.1$ W/m^2, $I_p = 19.9$ W/m^2
33.59 (a) $\theta = \arcsin\left(\dfrac{2R - d}{2R + d}\right)$ (b) $R = \dfrac{d(n_1 + n_2)}{2(n_1 - n_2)}$
(c) 2.07 cm (d) ≈50 cm (e) 1.49 μs
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
34.19 2.67 cm
34.21 3.30 m
34.23 (a) at the center of the bowl, 1.33 (b) no
34.25 39.5 cm
34.27 (a) 30 cm, 9.60 mm, inverted (b) 60.0 cm, 9.60 mm, upright
34.29 (a) 107 cm to the right of the lens, 17.8 mm; real; inverted (b) the same
34.31 71.2 cm to the right of the lens; −2.97
34.33 3.69 cm; 2.82 cm to the left of the lens
34.35 1.67
34.37 12.0 mm
34.39 −12 cm
34.41 26.3 cm from the lens, 12.4 mm; erect; same side
34.43 (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
34.45 (a) 53.0 cm (b) real (c) 2.50 mm; inverted
34.47 (a) $s' = -2f$, virtual, left (b) $3h$, upright (c) yes
34.49 8.69 cm; no
34.51 (a) 12 cm to the right of the converging lens
(b) −8 cm (c) 24 cm to the right of the diverging lens (d) 12 cm to the right of the converging lens, –4 cm, 6 cm to the right of the diverging lens
34.53 (a) 80.0 cm (b) 76.9 cm
34.55 49.4 cm, 2.02 diopters
34.57 (a) far point: 202 cm, near point: 44.6 cm
(b) 56.1 cm
34.59 (a) 6.06 cm (b) 4.12 mm
34.61 (a) −6.33 (b) 1.90 cm (c) 0.127 rad
34.63 (a) 0.661 m (b) 59.1
34.65 16.0 mm
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 (a) 6.49 cm (b) 6.76 cm $\le s \le$ 7.05 cm
(c) 66.2 cm (d) 1.36 m/s, downward (e) 24.1 cm
34.87 (a) 30.9 cm (b) 29.2 cm
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 (d) 36.0 cm, 21.6 cm; $d = 1.2$ cm
34.101 (a) −16.6 cm (b) 20.0 cm to the right
34.103 (b) 10.0 cm (c) 90 cm in front of the glass, virtual
34.105 (a) $4f$
34.107 (b) 1.74 cm
34.109 Choice (d)
34.111 Choice (b)

Chapter 35

35.1 0.75 m, 2.00 m, 3.25 m, 4.50 m, 5.75 m, 7.00 m, 8.25 m
35.3 (a) 2.0 m (b) constructively
(c) 1.0 m, destructively
35.5 (a) 11 (b) 2.50 m
35.7 1.14 mm
35.9 0.83 mm
35.11 (a) 39 (b) ±73.3°
35.13 12.6 cm
35.15 1200 nm
35.17 (a) $0.750 I_0$ (b) 80 nm
35.19 1670 rad
35.21 (a) 4.52 rad (b) $0.404 I_0$
35.23 114 nm
35.25 (a) 55.6 nm (b) (i) 2180 nm
(ii) 11.0 wavelengths
35.27 (a) 514 nm; green (b) 603 nm; orange
35.29 0.11 μm
35.31 0.570 mm
35.33 1.57
35.35 (a) 96.0 nm (b) no, no
35.37 (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
35.39 1.730
35.41 761 m, 219 m, 90.1 m, 20.0 m
35.43 $6.8 \times 10^{-5} (C°)^{-1}$
35.45 1.33 μm
35.47 600 nm, 467 nm; no
35.49 (a) 1.54 (b) ±15.0°
35.51 (a) 50 MHz (b) 237.0 m
35.53 (a) $\lambda/2$ (b) 6
(c) ±9.64 cm, ±31.7 cm, ±84.1 cm
(d) 2, ±11.3 cm from the center
35.55 14.0
35.57 Choice (d)
35.59 Choice (c)

Chapter 36

36.1 506 nm
36.3 (a) 226 (b) ±83.0°
36.5 9.07 m
36.7 (a) 63.8 cm
(b) ±22.1°, ±34.3°, ±48.8°, ±70.1°
36.9 ±16.0°, ±33.4°, ±55.6°
36.11 (a) 10.9 mm (b) 5.4 mm
36.13 (a) 580 nm (b) 0.128

36.15 (a) $\pm 13.0°, \pm 26.7°, \pm 42.4°, \pm 64.1°$
(b) 2.08 W/m^2
36.17 (a) $d = 2a$ (b) 0.101 W/m^2
36.19 (a) 3 (b) 2
36.21 (a) 0.0627°, 0.125° (b) $0.249 I_0$, $0.0256 I_0$
36.23 (a) 3.00 mm (b) 6.00 mm
36.25 (a) 4830 lines/cm (b) 4; $\pm 37.7°, \pm 66.5°$
36.27 (a) 2.40 μm (b) 417 slits/mm (c) 1.80 μm
36.29 (a) 2.69×10^4 (b) $m = 3$
36.31 (a) 467 nm (b) 27.8°
36.33 (a) 1820 slits (b) 41.0297°, 0.0137°
36.35 0.232 nm
36.37 92 cm
36.39 1.88 m
36.41 220 m
36.43 (a) 77 m (Hubble), 1100 km (Arecibo)
(b) 1500 km
36.45 30.2 μm
36.47 (a) 78 (b) $\pm 80.8°$ (c) 555 $\mu W/m^2$
36.49 1.68
36.51 (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°
36.53 −0.033 mm; decrease
36.55 360 nm
36.57 second
36.59 (a) 65° (b) 0.16 nm (c) 4.4×10^6 m/s
(d) 2.6×10^{16} Hz (e) 3.3×10^{-34} J·s
(f) 6.6×10^{-34} J·s (g) yes
36.61 (a) 1.03 mm (b) 0.148 mm
36.63 (a) 12.1 μm (b) 10.4 cm, 15.2 cm
36.65 (a) 10.6 μm (b) 532 nm/s (c) 1.60 μm
(d) 1.11 s (e) 92.7% (f) 2.88°
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 (a) 12.0 ms measured by the first officer
(b) $0.997c$
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 (a) 2.30 (b) 4.11
37.27 $3.06 p_0$
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) 0.867 nJ (b) 0.270 nJ (c) 0.452
37.37 (a) 5.34 pJ (nonrel), 5.65 pJ (rel), 1.06
(b) 67.8 pJ (nonrel), 331 pJ (rel), 4.88
37.39 (a) 2.06 MV (b) 0.330 pJ = 2.06 MeV
37.41 $0.700c$
37.43 42.5 y
37.45 (a) 1.0 kg (b) A: 1.7 kg, B: 1.8 kg
37.47 (a) $\Delta = 9 \times 10^{-9}$ (b) $7000m$
37.49 5.01 ns, clock on plane
37.51 0.168 MeV
37.53 (a) 1.08×10^{14} J (b) 2.70×10^{19} W
(c) 1.10×10^{10} kg
37.55 (a) $IHL\hat{\boldsymbol{j}}$
(b) $IHLB\hat{\boldsymbol{\imath}}$
(c) $\vec{\mu}\,' = \vec{\mu}/\gamma$
(d) $\vec{B}\,' = B\gamma\hat{\boldsymbol{k}}$ (e) $\vec{B}\,'_\perp = \vec{B}_\perp\gamma$
37.57 $0.357c$; receding
37.59 154 km/h
37.61 2.04×10^{-13} N
37.63 (a) 2.6×10^{-8} s (b) 0.97
37.65 (a) 2.0×10^{-18} kg (b) 4.0×10^4 m/s^2
37.67 (a) (0, 0), $(-L\gamma, Lv\gamma/c^2)$, $(-vT\gamma, T\gamma)$,
$((-L + vT)\gamma, (T + Lv/c^2)\gamma)$
(c) cLT (d) cLT (e) yes
37.69 (a) 2494 MeV (b) 2.526 times
(c) 987.4 MeV, twice as much
37.71 Choice (c)
37.73 Choice (b)

CREDITS

Cover John Woodworth/Alamy

Front Matter John Woodworth/Alamy; iii: Hugh D. Young; iii: Roger A. Freedman; xiii: Brian A Jackson/Shutterstock; xiv: Blanscape/Shutterstock; xv: George Burba/Shutterstock; xvi: 06photo/Shutterstock; xvii: Tomatito/Shutterstock

Chapter 1 p. 1: Minerva Studio/Shutterstock; p. 2: Michele Perbellini/Shutterstock; p. 2: NASA; p. 5: Science History Images/Alamy Stock Photo; p. 6: Courtesy of Bruker Corporation; p. 6: DGLimages/Alamy Stock Photo; p. 6: ESB Basic/Shutterstock; p. 6: JPL/NASA; p. 6: NASA; p. 6: Sebastian Kaulitzki/Shutterstock; p. 8: Chronicle/Alamy Stock Photo; p. 10: Tyler Olson/Shutterstock

Chapter 2 p. 34: Vibe Images/Fotolia; p. 37: moodboard/Brand X Pictures/Getty Images; p. 37: Stefan Holm/Shutterstock; p. 46: NASA; p. 50: Richard Megna/Fundamental Photographs; p. 53: RUSS ROHDE/Getty Images; p. 53: Thomas_EyeDesign/E+/Getty Images

Chapter 3 p. 67: Feng Li/Getty Images Sport/Getty Images; p. 71: Luca Lozzi / Getty Images News/Getty Images; p. 72: Dominique Douieb/PhotoAlto Agency RF Collections/Getty Images; p. 75: Richard Megna /Fundamental Photographs; p. 77: Richard Megna /Fundamental Photographs; p. 84: hubblesite/STScI/NASA; p. 85: David Wall/Alamy; p. 86: Chris Putnam/123rf; p. 89: CERN

Chapter 4 p. 100: ilbusca/Getty Images; p. 105: FotoPrisma/Fotolia; p. 107: Wayne Eastep/ The Image Bank/Getty Images; p. 110: FLUKY FLUKY/ Shutterstock; p. 112: Cheryl A. Meyer/Shutterstock; p. 116: Kadmy/Fotolia; p. 119: Maksym Gorpenyuk/Shutterstock; p. 120: John W. McDonough/Sports Illustrated/ Getty Images; p. 120: PCN Photography/Alamy Stock Photo; p. 120: Roy Pedersen/Fotolia; p. 127: www.autosnout.com

Chapter 5 p. 130: Brian A Jackson/Shutterstock; p. 139: NASA; p. 142: Efired/Shutterstock; p. 144: Alex Kosev/Shutterstock; p. 148: Eye of Science/Science Source; p. 149: dzphotovideo/Getty Images; p. 155: T-Photo/Shutterstock; p. 156: BrunoWeltmann/Shutterstock; p. 156: Comaniciu Dan/Shutterstock; p. 156: Jason Stitt/Shutterstock; p. 156: NASA/JPL/Space Science Institute; p. 166: Data from Rothschild, M. et al. "The Flying Leap of a Flea," Scientific American, Vol. 229, No. 5 (November 1973), pp. 92–101.; p. 169: From www.engineeringtoolbox.com

Chapter 6 p. 172: Brocreative/Shutterstock; p. 173: mariiya/Fotolia; p. 173: Steve Gschmeissner/Science Source; p. 181: Mikadun/Shutterstock; p. 186: Steve Gschmeissner/Science Source; p. 190: Anthony Hall/Fotolia; p. 190: Fox Photos/Stringer/Getty Images; p. 190: Gayvoronskaya_Yana/Shutterstock; p. 190: Keystone/Stringer/Shutterstock; p. 191: fandu/Fotolia; p. 200: www.sciencebuddies.org

Chapter 7 p. 203: Nagel Photography/Shutterstock; p. 204: 0tvalo/Shutterstock; p. 205: Erni/Shutterstock; p. 205: Robert F. Bukaty/Apimages; p. 207: hinnamsaisuy/Shutterstock; p. 212: threerocksimages/Shutterstock; p. 214: G. Ronald Austing/Science Source; p. 214: JNP/Shutterstock; p. 220: 135pixels/Shutterstock; p. 220: Kletr/Shutterstock; p. 223: LaiQuocAnh/Shutterstock; p. 224: Thorsten Eckert/Alamy; p. 235: www.engineershandbook.com

Chapter 8 p. 237: Colorful High Speed Photographs/Getty Images; p. 238: Alex Emanuel Koch/Shutterstock; p. 239: Willie Linn/Shutterstock; p. 240: Jim Cummins/Getty Images; p. 240: Vereshchagin Dmitry/Shutterstock; p. 242: Ted Kinsman/Science Source; p. 248: FPG/Getty Images; p. 251: David Leah/Getty Images; p. 256: Richard Megna /Fundamental Photographs; p. 258: Elliotte Rusty Harold/Shutterstock; p. 259: NASA; p. 270: From www.chuckhawks.com

Chapter 9 p. 273: Blanscape/Shutterstock; p. 274: connel/Shutterstock; p. 278: Hybrid Medical Animation/Science Source; p. 285: Ammit Jack/Shutterstock; p. 285: Dan Rodney/Shutterstock; p. 288: David J. Phillip/Associated Press; p. 288: DenisNata/Shutterstock; p. 290: NASA; p. 300: NASA

Chapter 10 p. 303: Dan Rodney/Shutterstock; p. 307: 68/Ocean/Corbis; p. 311: Bruce MacQueen/Shutterstock; p. 311: David Lentink/Science Source; p. 311: Robert Young/Fotolia; p. 313: gorillaimages/Shutterstock; p. 315: Chris DeRidder/Shutterstock; p. 316: Bjorn Heller/Shutterstock; p. 320: NHPA/Photoshot

Chapter 11 p. 339: nito/Shutterstock; p. 342: haveseen/Shuttterstock; p. 344: Maridav/Shutterstock; p. 347: Andrew Bret Wallis/Getty Images; p. 347: Benjamin Marin Rubio/Shutterstock; p. 347: Djomas/Shutterstock; p. 347: Richard Carey/Fotolia; p. 350: Dante Fenolio/Science Source

Chapter 12 p. 366: Isabelle Kuehn/Shutterstock; p. 366: orlandin/Shutterstock; p. 367: travelview/Shutterstock; p. 370: Pascal, Blaise. 1653. Treatise on the Equilibrium of Liquids.; p. 370: SPL/Science Source; p. 371: Lisa F. Young/Shutterstock; p. 372: Andrey Jitkov/Shutterstock; p. 373: Archimedes; p. 375: Pali A/fotolia; p. 376: Pearson; p. 376: Popov Nikolay/Shutterstock; p. 377: Fribus Mara/Shutterstock; p. 379: Sunychka Sol/Shutterstock; p. 382: TechWizard/Shutterstock; p. 383: Edward Lara/Shutterstock; p. 383: Jurik Peter/Shutterstock; p. 383: VitaminCo/Shutterstock; p. 384: Harold Edgerton/MIT courtesy of Palm Press, INc.

Chapter 13 p. 395: NASA/JPL-Caltech/SSI/Cornell; p. 395: Newton published the law of gravitation in 1687. Philosophiae naturalis principia mathematica (MacLehose, 1726) [1871] Reprinted for Sir William Thomson and Hugh Blackburn.; p. 397: JPL/Cornell University/NASA; p. 397: NASA/JPL/University of Arizona; p. 399: Robert Nemiroff & Jerry Bonnell/NASA; p. 400: NASA; p. 400: swisshippo/Fotolia; p. 404: NASA; p. 405: Based on Newton, Isaac. 1728. A Treatise on the System of the World.; p. 405: NASA/Dembinsky Photo Associates/Alamy Stock Photo; p. 406: M.Showalter/NASA; p. 407: Kepler, Johannes. 1597. Mysterium Cosmographicum.; p. 409: NASA; p. 410: NASA; p. 416: 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.; p. 417: NASA; p. 418: NASA; p. 419: These images/animations were created by Prof. Andrea Ghez and her research team at UCLA and are from data sets obtained with the W. M. Keck Telescopes.; p. 426: ssd.jpl.nasa.gov; p. 427: nssdc.gsfc.nasa.gov/planetary/factsheet; p. 428: Based on S. Seager et al., (2007) "Mass–Radius Relationships for Solid Exoplanets" The Astrophysical Journal, vol. 669, no.2, p. 1297.; arXiv:0707.2895 [astro-ph].

Chapter 14 p. 429: blurAZ/Shutterstock; p. 430: Steve Byland/Shutterstock; p. 434: American Diagnostic Corporation; p. 446: Zurijeta/Shutterstock; p. 449: Christopher Griffin/Alamy Stock Photo; p. 452: symbiot/Shutterstock; p. 453: sonya etchison/Shutterstock

Chapter 15 p. 464: Walter D. Mooney/U.S. Geological Survey; p. 465: MARCO POLO COLLECTION/Balan Madhavan / Alamy; p. 466: imageBROKER/Alamy; p. 467: EpicStockMedia/Shutterstock; p. 476: Kodda/Shutterstock; p. 478: Christian Delbert/shutterstock; p. 483: Richard Megna /Fundamental Photographs; p. 487: Shmeliova Natalia/Shutterstock; p. 488: NSO/AURA/NSF

Chapter 16 p. 501: Eduard Kyslynskyy/Shutterstock; p. 505: Anton Havelaar/123rf; p. 505: 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.; p. 505: Ollyy/Shutterstock; p. 506: deusexlupus/123rf; p. 506: Steve Thorne/Redferns/Getty Images; p. 508: Kretztechnik/Science Source; p. 511: Pius Lee/Shutterstock; p. 516: Digoarpi/Shutterstock; p. 519: B-D-S Piotr Marcinski/Shutterstock; p. 519: Martin Bough/Fundamental Photographs; p. 522: Roger A. Freedman; p. 523: Gino Santa Maria/Shutterstock; p. 526: cleanfotos/Shutterstock; p. 529: NASA; p. 30: NASA

Chapter 17 p. 541: huyangshu/Shutterstock; p. 545: David Thyberg/Shutterstock; p. 545: hofhauser/Shutterstock; p. 545: luckyraccoon/123rf.com; p. 545: Roger A. Freedman; p. 548: Evannovostro/Shutterstock; p. 550: Eric Schrader/Pearson Education, Inc.; p. 552: Hugh D. Young; p. 554: Roman Sigaev/Shutterstock; p. 556: Adam Hart-Davis/Science Source; p. 557: Richard Megna/Fundamental Photographs; p. 558: NASA; p. 558: Wavebreak Media Ltd/123rf; p. 562: DmitryND/Getty Images; p. 562: Paul Nicklen/ Getty Images; p. 564: Hugh D. Young; p. 565: Ted Kinsman/Science Source; p. 567: From https://data.giss.nasa.

gov/gistemp/graphs/; p. 567: 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; p. 575: Nature Production/Nature Picture Library

Chapter 18 p. 579: Stockyimages/Shutterstock; p. 581: Topham/The Image Works; p. 586: Courtesy of Bruker Corporation; p. 590: Rich Iwasaki/Getty Images; p. 591: ARCTIC IMAGES/Alamy Stock Photo; p. 593: David Grossman/The Image Works; p. 599: Ray Coleman/Science Source; p. 600: Melissa King/Shutterstock; p. 606: NOAO/AURA/NSF; p. 611: T.A. Rector/University of Alaska Anchorage, H. Schweiker/WIYN and NOAO/AURA/NSF

Chapter 19 p. 613: Ken Paul/Getty Images; p. 614: John P. Surey; p. 614: Phumphao Sumrankong/Shutterstock; p. 614: Stock/Getty Images; p. 619: Diego Cervo/Shutterstock; p. 620: Liv friis-larsen/Shutterstock; p. 624: Rob Byron/Shutterstock; p. 624: Tom Branch/Science Source; p. 630: dean bertoncelj/Shutterstock

Chapter 20 p. 642: George Burba/Shutterstock; p. 644: James Morgan/Shutterstock; p. 645: Tyler Olson/Shutterstock; p. 653: Bill Bachman/Science Source; p. 658: Staff Sgt. Robert Zoellner; p. 659: cowardlion/Shutterstock; p. 664: DenisNata/Shutterstock; p. 665: Eric Schrader/Pearson Education; p. 670: Eric Schrader/Pearson Education; p. 677: Based on the Atlantic Ocean Atlas, part of the Woods Hole Oceanographic Institution Atlas Series, Vol. 1. (Woods Hole: Woods Hole Oceanographic Institution, 1960). From https://commons.wikimedia.org/wiki/File:Thermocline.jpg

Chapter 21 p. 678: Mikael Damkier/Shutterstock; p. 682: Clayton Stalter/The Image Works; p. 684: George Resch/Fundamental Photographs; p. 685: Knape/Getty Images; p. 690: Matt9122/Shutterstock; 701: otnaydur/Shutterstock; p. 701: Richard Megna / Fundamental Photographs; p. 704: Mark Smith/Science Source; p. 716: Data from Dominic Clarke et al., "Detection and Learning of Floral Electric Fields by Bumblebees," Science, April 5, 2013: Vol. 340, no. 6128, pp. 66–69.

Chapter 22 p. 718: U. S. Department of Energy; p. 723: Mark Webster/Getty Images; p. 725: Akg-images; p. 734: Dr. David Furness, Keele University/Science Source; p. 736: CTK/Alamy Stock Photo; p. 736: yevgeniy11/Shutterstock

Chapter 23 p. 747: Ted Kurihara Photography/Getty Images; p. 752: NASA; p. 754: Vladyslav Danilin/Fotolia; p. 755: StockLite/Shutterstock; p. 757: BSIP/Newscom; p. 762: Keystone/Getty Images; p. 765: Map from the USGS, U.S. Geological Survey Library; p. 768: 3D4Medical/Science Source

Chapter 24 p. 783: Stock image/Shutterstock; p. 784: Andrew Lambert Photography/Science Source; p. 786: Eric Schrader/Pearson Education; p. 787: otnaydur/Shutterstock; p. 791: National Technology and Engineering Solutions of Sandia, LLC.; p. 795: Cabeca de Marmore/Shutterstock; p. 798: Based on H Heller, M Schaefer, & K Schulten, "Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases", J. Phys. Chem. p. 97:8343-60, 1993/Pearson Science; p. 800: Luis Molinero Martínez/123rf; iii: Caiaimage/Chris Ryan/GettyImages

Chapter 25 p. 812: JGI/Tom Grill/Getty Images; p. 815: Eric Schrader/Pearson Education; p. 817: Stockbyte/Getty Images; p. 818: Dr. David Furness, Keele University/Science Source; p. 820: TFoxFoto/Shutterstock; p. 822: Laurence Gough/Shutterstock; p. 824: Andrew J. Martinez/Science Source; p. 824: Richard Megna /Fundamental Photographs; p. 843: 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).

Chapter 26 p. 844: 06photo/Shutterstock; p. 846: Claro Cortes/Reuters; p. 849: John P. Surey; p. 854: Yury Kosourov/Shutterstock; p. 855: Gary Ombler/Getty Images; p. 857: Richard Megna/Fundamental Photographs; p. 860: Mehau Kulyk/Science Source; p. 864: Evgeny Tomeev/Shutterstock; p. 864: Vladimir Arndt/Shutterstock

Chapter 27 p. 878: Michael Flippo/Fotolia; p. 879: harukaze-k/Fotolia; p. 880: Robert Smith/Getty Images; p. 882: LStockStudio/Shutterstock; p. 885: Richard Megna/Fundamental Photographs; p. 885: Richard Megna/Fundamentalphotographs; p. 888: Richard Megna/Fundamental Photographs; p. 889: NASA; p. 889: Photo courtesy of Berkeley Lab; p. 889: Roman Krochuk/Shuttestock; p. 898: Simon Fraser/Science Source; p. 903: heptagon.circle/Shutterstock; p. 905: Microgen/Shutterstock

Chapter 28 p. 918: Image courtesy of CERN; p. 922: Tristan3D/Shutterstock; p. 925: TaweePhoto/Shutterstock; p. 927: Jeremy Walker/Science Source; p. 929: James Steidl/Fotolia; p. 936: Eric Schrader; p. 941: Kenneth Scarberry

Chapter 29 p. 953: jabejon/Getty Images; p. 954: Simon Fraser/University of Durham/Science Source; p. 960: CTK/Alamy Stock Photo; p. 968: Mino Surkala/Shutterstock; p. 968: Susan & Allan Parker/Alamy Stock Photo; p. 970: NASA; p. 976: Richard Megna/Fundamental Photographs

Chapter 30 p. 988: EschCollection/GettyImages; p. 989: Petr Malyshev/Shutterstock; p. 992: Ron Chapple/Shutterstock; p. 994: Turleyt/Fotolia; p. 997: Ensuper/Shutterstock; p. 997: NASA

Chapter 31 p. 1018: Roger A. Freedman; p. 1021: Rubberball/Mike Kemp/Getty Images; p. 1028: Lawrence Migdale/Science Source; p. 1028: Mauro Fermariello/Science Source; p. 1036: Katharina Wittfeld/Shutterstock; p. 1037: Roger A. Freedman; p. 1037: Thomas Barrat/Shutterstock; p. 1038: egd/Shutterstock

Chapter 32 p. 1048: JLImages/Alamy Stock Photo; p. 1049: Albert Einstein in James Clerk Maxwell, A Commemoration Volume (New York: The Macmillan Company, 1931), p. 71.; p. 1049: Bettmann/Getty Images; p. 1050: Jirapol Konroo/Shutterstock; p. 1050: Pablo Hidalgo/123rf; p. 1052: Ted Kinsman/Science Source; p. 1058: John A Davis/Shutterstock; p. 1060: Dudarev Mikhail/Shutterstock; p. 1062: Craig Holmes Premium/Alamy Stock Photo; p. 1063: Alexander Tsiaras/Science Source; p. 1065: NASA; p. 1068: grzymkiewicz/Getty Images; p. 1075: "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.

Chapter 33 p. 1077: Rozaliya/Getty Images; p. 1078: bradleym/Getty Images; p. 1078: NIH/Science Source; p. 1082: Mead Treadwell/NASA; p. 1082: Passakorn Umpornmaha/Shutterstock; p. 1082: Peter Herring/Image Quest Marine; p. 1082: Roger A. Freedman; p. 1085: Ken Kay /Fundamental Photographs; p. 1087: VEM/Science Source; p. 1089: mtilghma/Fotolia; p. 1091: Dougal Waters/Getty Images; p. 1091: Pixtal/Agefotostock; p. 1093: Diane Hirsch /Fundamental Photographs; p. 1096: Nestor Rizhniak/Shutterstock; p. 1097: Cefo design/Shutterstock; p. 1097: Linda Hirst; p. 1097: Peter Aprahamian/Sharples Stress Engineers Ltd./Science Source; p. 1098: lauriek/Getty Images; p. 1098: Roger A. Freedman; p. 1101: Loskutnikov/Shutterstock

Chapter 34 p. 1110: Corbis/VCG/Getty Images; p. 1113: Eric Schrader; p. 1115: NASA; p. 1117: OlegD/Shutterstock; p. 1124: Vaclav Volrab/Shutterstock; p. 1126: Richard Megna/Fundamnetal Photographs, NYC; p. 1130: Hoby Finn/Getty Images; p. 1130: John Hedgecoe/TopFoto/The Image Works; p. 1136: John Hedgecoe/TopFoto/The Image Works; p. 1137: kak2s/Shutterstock; p. 1137: THK Photo Products; p. 1138: Omikron/Science Source; p. 1139: ethylalkohol/Shutterstock; p. 1142: Jan Zoetekouw/Shutterstock; p. 1144: Jan Hinsch/Science Source; p. 1147: NOAO

Chapter 35 p. 1159: Ilya Andriyanov/Shutterstock; p. 1162: Peter Dazeley/Getty Images; p. 1163: Roger A. Freedman; p. 1164: Pearson Education Inc.; p. 1170: RBFried/Getty Images; p. 1173: Bausch & Lomb Incorporated; p. 1174: Bausch & Lomb Incorporated; p. 1175: Vladimirkarp/Shutterstock; p. 1175: Vukusic, Pete; p. 1176: Chaiwut Siriphithakwong/Shutterstock

Chapter 36 p. 1185: Tomatito/Shutterstock; p. 1186: Hugh D. Young; p. 1186: Ken Kay/Fundamental Photograph; p. 1189: Pearson Education Inc.; p. 1191: Richard Megna/Fundamental Photographs; p. 1193: Roger A. Freedman; p. 1197: Dr.OGA/Shutterstock; p. 1198: National Optical Astronomy Observatory (NOAO); p. 1198: SPL/Science Source; p. 1199: Tek Image/Science Source; p. 1200: Science History Images/Alamy Stock Photo; p. 1202: Omikron/Science Source; p. 1203: Image Courtesy of National Radio Astronomy Observatory.; p. 1204: GIPhotoStock/Science Source; p. 1204: Pearson Education Inc.; p. 1204: Pearson Education, Inc; p. 1205: Michal Ninger/Shutterstock; p. 1207: Paul Silverman/Fundamental Photographs; p. 1208: GIPhotoStock/Science Source; p. 1213: NASA

Chapter 37 p. 1217: Courtesy of Brookhaven National Laboratory; p. 1220: Adimas/Fotolia; p. 1224: John F. Kennedy Space Center/NASA; p. 1227: Eva-Katalin/Getty Images; p. 1228: Jack Fields/Science Source; p. 1233: sirtravelalot/Shutterstock; p. 1236: John F. Kennedy Space Center/NASA; p. 1236: sirtravelalot/Shutterstock; p. 1241: Geng yuhe/Associated Press; p. 1244: Matt Cooper/Shutterstock

Rear Matter A-1: Adapted from the National Institute of Standards and Technology (http://physics.nist.gov/cuu); A-7: National Institute of Standards and Technology (http://physics.nist.gov>cuu).; A-8: NASA (http://solarsystem.nasa.gov/planets/).

INDEX

D

E

N

O

P

Q

R

S

T

U

V

W

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[‡]	1.30×10^{22}	1.19×10^{6}	5.91×10^{12}	248.0 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 $= 1\ \mu$K $= 10^{-6}$ 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

Fundamental Physical Constants*

Name	Symbol	Value
Speed of light in vacuum	c	2.99792458×10^{8} m/s
Magnitude of charge of electron	e	$1.602176634 \times 10^{-19}$ C
Gravitational constant	G	$6.6.67408(31) \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$
Planck's constant	h	$6.62607015 \times 10^{-34}\ \mathrm{J \cdot s}$
Boltzmann constant	k	1.380649×10^{-23} J/k
Avogadro's number	N_A	$6.02214076 \times 10^{23}$ molecules/mol
Gas constant	R	$8.314462618\ldots\ \mathrm{J/mol \cdot K}$
Mass of electron	m_e	$9.10938356(11) \times 10^{-31}$ kg
Mass of proton	m_p	$1.672621898(21) \times 10^{-27}$ kg
Mass of neutron	m_n	$1.674927471(21) \times 10^{-27}$ kg
Magnetic constant	μ_0	$1.25663706 \times 10^{-6}\ \mathrm{Wb/A \cdot m}$ (approximate) $\cong 4\pi \times 10^{-7}\ \mathrm{Wb/A \cdot m}$
Electric constant	$\epsilon_0 = 1/\mu_0 c^2$	$8.854187817 \times 10^{-12}\ \mathrm{C^2/N \cdot m^2}$ (approximate)
	$1/4\pi\epsilon_0$	$8.987551787 \times 10^{9}\ \mathrm{N \cdot m^2/C^2}$ (approximate)

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	$1.6021766209(98) \times 10^{-19}$ J
Atomic mass unit	1 u	$1.660539040(20) \times 10^{-27}$ kg
Electron rest energy	$m_e c^2$	0.5109989461(31) MeV
Volume of ideal gas (0°C and 1 atm)		22.413962(13) 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. The exact values of the magnitude of the charge of the electron, Planck's constant, the Boltzmann constant, Avogadro's number, and the gas constant are from the redefinitions adopted in 2018. As consequences of these redefinitions, the values of the magnetic constant and electric constant now have fractional uncertainties of about 2×10^{-10}. As of this writing (2018) it was expected that updated values of the magnetic and electric constants, as well as of all other constants with uncertainties, were to be announced in May 2019. These updated values will be available at **http://physics.nist.gov/cuu**.