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.

Example 32.3 Energy in a nonsinusoidal wave

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.

SOLUTION

IDENTIFY and SET UP: In this wave $\vec{E}$ and $\vec{B}$ are uniform behind the wave front (and zero ahead of it). Hence the target variables B, u, and S must also be uniform behind the wave front. Given the magnitude E, we use Eq. (32.4) to find B, Eq. (32.25) to find u, and Eq. (32.27) to find S. (We cannot use Eq. (32.29), which applies to sinusoidal waves only.)

EXECUTE: From Eq. (32.4),

$$B = \frac{E}{c} = \frac{100\ \text{V/m}}{3.00 \times 10^8\ \text{m/s}} = 3.33 \times 10^{-7}\ \text{T}$$

From Eq. (32.25),

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

The magnitude of the Poynting vector is

$$\begin{aligned} S &= \frac{EB}{\mu_0} = \frac{(100\ \text{V/m})(3.33 \times 10^{-7}\ \text{T})}{4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A}} \\ &= 26.5\ \text{V}\cdot\text{A/m}^2 = 26.5\ \text{W/m}^2 \end{aligned}$$

EVALUATE: We can check our result for S by using Eq. (32.26):

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

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.

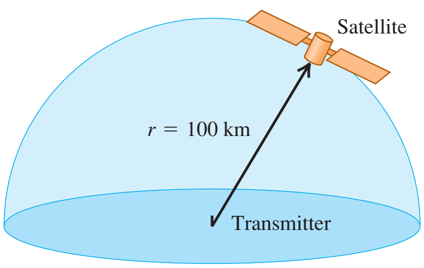
Example 32.4 Energy in a sinusoidal wave

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

SOLUTION

IDENTIFY and SET UP: We are given the transmitter's average total power P. The intensity I is just the average power per unit area, so to find I at 100 km from the transmitter we divide P by the surface area of the hemisphere shown 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 Eqs. (32.29) to find E_{max}; Eq. (32.4) then yields B_{max}.

32.19 A radio station radiates waves into the hemisphere shown.



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

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

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 Eqs. (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).

Electromagnetic Momentum Flow and Radiation Pressure

By using the observation that energy is required to establish electric and magnetic fields, we have shown that electromagnetic waves transport energy. It can also be shown that electromagnetic waves carry *momentum* p, with a corresponding momentum density (momentum dp per volume dV) of magnitude

$$\frac{dp}{dV} = \frac{EB}{\mu_0 c^2} = \frac{S}{c^2} \tag{32.30}$$

This momentum is a property of the field; it is not associated with the mass of a moving particle in the usual sense.

There is also a corresponding momentum flow rate. The volume dV occupied by an electromagnetic wave (speed c) that passes through an area A in time dt is $dV = Ac\,dt$. When we substitute this into Eq. (32.30) and rearrange, we find that the momentum flow rate per unit area is

$$\frac{1}{A}\frac{dp}{dt} = \frac{S}{c} = \frac{EB}{\mu_0 c} \quad \text{(flow rate of electromagnetic momentum)} \tag{32.31}$$

This is the momentum transferred per unit surface area per unit time. We obtain the *average* rate of momentum transfer per unit area by replacing S in Eq. (32.31) by $S_{av} = I$.

This momentum is responsible for **radiation pressure.** When an electromagnetic wave is completely absorbed by a surface, the wave's momentum is also transferred to the surface. For simplicity we'll consider a surface perpendicular to the propagation direction. Using the ideas developed in Section 8.1, we see that the rate dp/dt at which momentum is transferred to the absorbing surface equals the *force* on the surface. The average force per unit area due to the wave, or *radiation pressure* p_{rad}, is the average value of dp/dt divided by the absorbing area A. (We use the subscript "rad" to distinguish pressure from momentum, for which the symbol p is also used.) From Eq. (32.31) the radiation pressure is

$$p_{rad} = \frac{S_{av}}{c} = \frac{I}{c} \quad \text{(radiation pressure, wave totally absorbed)} \tag{32.32}$$

If the wave is totally reflected, the momentum change is twice as great, and the pressure is

$$p_{rad} = \frac{2S_{av}}{c} = \frac{2I}{c} \quad \text{(radiation pressure, wave totally reflected)} \tag{32.33}$$

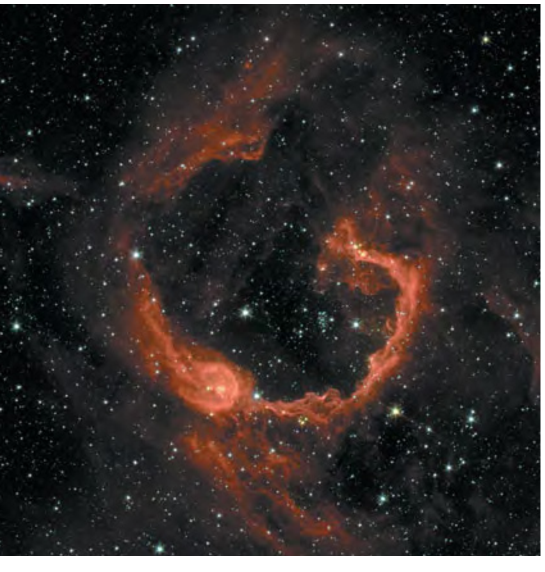
For example, the value of I (or S_{av}) for direct sunlight, before it passes through the earth's atmosphere, is approximately $1.4\ \text{kW/m}^2$. From Eq. (32.32) the corresponding average pressure on a completely absorbing surface is

$$p_{rad} = \frac{I}{c} = \frac{1.4 \times 10^3\ \text{W/m}^2}{3.0 \times 10^8\ \text{m/s}} = 4.7 \times 10^{-6}\ \text{Pa}$$

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 32.45). Inside stars that are much more massive and luminous than the sun, radiation pressure is so great that it substantially augments the gas pressure within the star and so helps to prevent the star from collapsing under its own gravity. In some cases the radiation pressure of stars can have dramatic effects on the material surrounding them (Fig. 32.20).

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

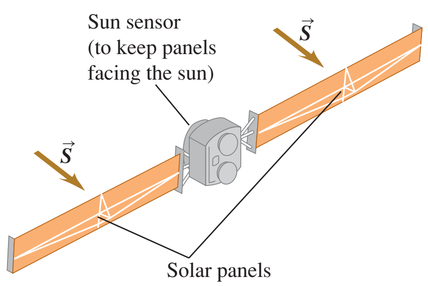


Example 32.5 Power and pressure from sunlight

An earth-orbiting satellite has solar energy–collecting panels with a total area of 4.0 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.

SOLUTION

IDENTIFY and SET UP: This problem uses the relationships among intensity, power, radiation pressure, and force. In the above discussion we calculated the intensity I (power per unit area) of sunlight as well as the radiation pressure p_{rad} (force per unit area) of sunlight on an absorbing surface. (We calculated these values 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.



32.21 Solar panels on a satellite.

EXECUTE: The intensity I (power per unit area) is 1.4×10^3 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:

$$P = IA = (1.4 \times 10^3 \text{ W/m}^2)(4.0 \text{ m}^2)$$
$$= 5.6 \times 10^3 \text{ W} = 5.6 \text{ kW}$$

The radiation pressure of sunlight on an absorbing surface is $p_{rad} = 4.7 \times 10^{-6}$ Pa $= 4.7 \times 10^{-6}$ N/m^2. The total force F is the pressure p_{rad} times the area A:

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

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 earth) of a single grain of salt. Over time, however, this small force can have a noticeable effect on the orbit of a satellite like that in Fig. 32.21, and so radiation pressure must be taken into account.

Test Your Understanding of Section 32.4 Figure 32.13 shows one wavelength of a sinusoidal electromagnetic wave at time $t = 0$. For which of the following four values of x is (a) the energy density a maximum; (b) the energy density a minimum; (c) the magnitude of the instantaneous (not average) Poynting vector a maximum; (d) the magnitude of the instantaneous (not average) Poynting vector a minimum? (i) $x = 0$; (ii) $x = \lambda/4$; (iii) $x = \lambda/2$; (iv) $x = 3\lambda/4$.

MP

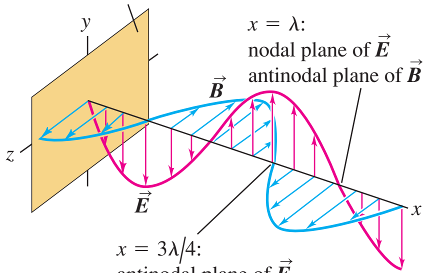
32.5 Standing Electromagnetic Waves

Electromagnetic waves can be *reflected;* the surface of a conductor (like a polished sheet of metal) or of a dielectric (such as a sheet of glass) can serve as a reflector. The superposition principle holds for electromagnetic waves just as for electric and magnetic fields. 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; you should review that discussion.

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

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.



MasteringPHYSICS
PhET: Microwaves

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, using the identities

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

The results are

$$E_y(x, t) = -2E_{\max} \sin kx \sin \omega t \quad (32.34)$$
$$B_z(x, t) = -2B_{\max} \cos kx \cos \omega t \quad (32.35)$$

Equation (32.34) is analogous to Eq. (15.28) for a stretched string. We see that at $x = 0$ the electric field $E_y(x = 0, t)$ is *always* zero; this is required by the nature of the ideal conductor, which plays the same role as a fixed point at the end of a string. Furthermore, $E_y(x, t)$ is zero at *all* times at points in those planes perpendicular to the x-axis for which $\sin kx = 0$—that is, $kx = 0, \pi, 2\pi, \ldots$. Since $k = 2\pi/\lambda$, the positions of these planes are

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

These planes are called the **nodal planes** of the $\vec{E}$ field; they are the equivalent of the nodes, or nodal points, of a standing wave on a string. Midway between any two adjacent nodal planes is a plane on which $\sin kx = \pm 1$; on each such plane, the magnitude of $E(x, t)$ equals the maximum possible value of $2E_{\max}$ twice per oscillation cycle. These are the **antinodal planes** of $\vec{E}$, corresponding to the antinodes of waves on a string.

The total magnetic field is zero at all times at points in planes on which $\cos kx = 0$. This occurs where

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

These are the nodal planes of the $\vec{B}$ field; there is an antinodal plane of $\vec{B}$ midway between any two adjacent nodal planes.

Figure 32.22 shows a standing-wave pattern at one instant of time. The magnetic field is *not* zero at the conducting surface $(x = 0)$. The surface currents that must be present to make $\vec{E}$ exactly zero at the surface cause magnetic fields at the surface. The nodal planes of each field are separated by one half-wavelength. The nodal planes of one field are midway between those of the other; hence the nodes of $\vec{E}$ coincide with the antinodes of $\vec{B}$, and conversely. Compare this situation to the distinction between pressure nodes and displacement nodes in Section 16.4.

The total electric field is a *sine* function of t, and the total magnetic field is a *cosine* function of t. The sinusoidal variations of the two fields are therefore 90° out of phase at each point. At times when $\sin \omega t = 0$, the electric field is zero *everywhere,* and the magnetic field is maximum. When $\cos \omega t = 0$, the magnetic field is zero everywhere, and the electric field is maximum. This is in contrast to a wave traveling in one direction, as described by Eqs. (32.16) or (32.19) separately, in which the sinusoidal variations of $\vec{E}$ and $\vec{B}$ at any particular point are *in phase.* You can show that Eqs. (32.34) and (32.35) satisfy the wave equation, Eq. (32.15). You can also show that they satisfy Eqs. (32.12) and (32.14), the equivalents of Faraday's and Ampere's laws (see Exercise 32.36).

Standing Waves in a Cavity

Let's now insert a second conducting plane, parallel to the first and a distance L from it, along the $+x$-axis. The cavity between the two planes is analogous to a stretched string held at the points $x = 0$ and $x = L$. Both conducting planes must be nodal planes for $\vec{E}$; a standing wave can exist only when the second plane is placed at one of the positions where $E(x, t) = 0$, so L must be an integer multiple of $\lambda/2$. The wavelengths that satisfy this condition are

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

The corresponding frequencies are

$$f_n = \frac{c}{\lambda_n} = n\frac{c}{2L} \qquad (n = 1, 2, 3, \dots) \tag{32.39}$$

Thus there is a set of *normal modes,* each with a characteristic frequency, wave shape, and node pattern (Fig. 32.23). By measuring the node positions, we can measure the wavelength. If the frequency is known, the wave speed can be determined. This technique was first used by Hertz in the 1880s in his pioneering investigations of electromagnetic waves.

Conducting surfaces are not the only reflectors of electromagnetic waves. Reflections also occur at an interface between two insulating materials with different dielectric or magnetic properties. The mechanical analog is a junction of two strings with equal tension but different linear mass density. In general, a wave incident on such a boundary surface is partly transmitted into the second material and partly reflected back into the first. For example, light is transmitted through a glass window, but its surfaces also reflect light.

32.23 A typical microwave oven sets up a standing electromagnetic wave with $\lambda = 12.2$ cm, a wavelength that is strongly absorbed by the water in food. Because the wave has nodes spaced $\lambda/2 = 6.1$ cm apart, the food must be rotated while cooking. Otherwise, the portion that lies at a node—where the electric-field amplitude is zero—will remain cold.

Example 32.6 Intensity in a standing wave

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

SOLUTION

IDENTIFY and SET UP: The intensity I of the wave is the time-averaged value S_{av} of the magnitude of the Poynting vector $\vec{S}$. To find S_{av}, we first use Eq. (32.28) to find the instantaneous value of $\vec{S}$ and then average it over a whole number of cycles of the wave.

EXECUTE: Using the wave functions of Eqs. (32.34) and (32.35) in Eq. (32.28) for the Poynting vector $\vec{S}$, we find

$$\begin{aligned}\vec{S}(x, t) &= \frac{1}{\mu_0}\vec{E}(x, t) \times \vec{B}(x, t) \\ &= \frac{1}{\mu_0}[-2\hat{\jmath}E_{max}\sin kx\cos\omega t] \times [-2\hat{k}B_{max}\cos kx\sin\omega t] \\ &= \hat{\imath}\frac{E_{max}B_{max}}{\mu_0}(2\sin kx\cos kx)(2\sin\omega t\cos\omega t) \\ &= \hat{\imath}S_x(x, t)\end{aligned}$$

Using the identity $\sin 2A = 2\sin A\cos A$, we can rewrite $S_x(x, t)$ as

$$S_x(x, t) = \frac{E_{max}B_{max}\sin 2kx\sin 2\omega t}{\mu_0}$$

The average value of a sine function over any whole number of cycles is zero. Thus *the time average of $\vec{S}$ at any point is zero;* $I = S_{av} = 0$.

EVALUATE: This result is what we should expect. The standing wave is a superposition of two waves with the same frequency and amplitude, traveling in opposite directions. All the energy transferred by one wave is cancelled by an equal amount transferred in the opposite direction by the other wave. When we use electromagnetic waves to transmit power, it is important to avoid reflections that give rise to standing waves.

Example 32.7 Standing waves in a cavity

Electromagnetic standing waves are set up in a cavity with two parallel, highly conducting walls 1.50 cm apart. (a) Calculate the longest wavelength λ and lowest frequency f of these standing waves. (b) For a standing wave of this wavelength, where in the cavity does $\vec{E}$ have maximum magnitude? Where is $\vec{E}$ zero? Where does $\vec{B}$ have maximum magnitude? Where is $\vec{B}$ zero?

Continued

SOLUTION

IDENTIFY and SET UP: Only certain normal modes are possible for electromagnetic waves in a cavity, just as only certain normal modes are possible for standing waves on a string. The longest possible wavelength and lowest possible frequency correspond to the $n = 1$ mode in Eqs. (32.38) and (32.39); we use these to find λ and f. Equations (32.36) and (32.37) then give the locations of the nodal planes of $\vec{E}$ and $\vec{B}$. The antinodal planes of each field are midway between adjacent nodal planes.

EXECUTE: (a) From Eqs. (32.38) and (32.39), the $n = 1$ wavelength and frequency are

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

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

(b) With $n = 1$ there is a single half-wavelength between the walls. The electric field has nodal planes ($\vec{E} = \mathbf{0}$) at the walls and an antinodal plane (where $\vec{E}$ has its maximum magnitude) midway between them. The magnetic field has *antinodal* planes at the walls and a nodal plane midway between them.

EVALUATE: One application of such standing waves is to produce an oscillating $\vec{E}$ field of definite frequency, which is used to probe the behavior of a small sample of material placed in the cavity. To subject the sample to the strongest possible field, it should be placed near the center of the cavity, at the antinode of $\vec{E}$.

Test Your Understanding of Section 32.5 In the standing wave described in Example 32.7, is there any point in the cavity where the energy density is zero at all times? If so, where? If not, why not?

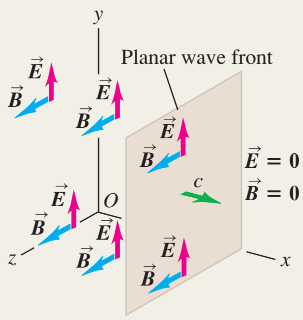
CHAPTER 32 SUMMARY

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 only 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 both give relationships between the magnitudes of $\vec{E}$ and $\vec{B}$; requiring both of these relationships to be 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 $\vec{E} \times \vec{B}$.

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

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

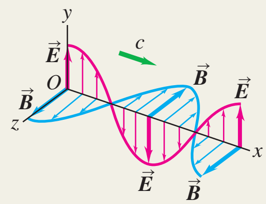
$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \qquad (32.9)$$



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.)

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

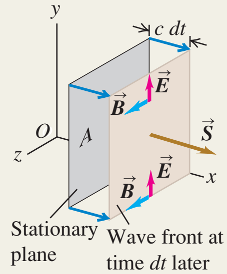
Energy and momentum in electromagnetic waves: The energy flow rate (power per unit area) in an electromagnetic wave in vacuum is given by the Poynting vector $\vec{S}$. The magnitude of the time-averaged value of the Poynting vector is called the intensity I of the wave. Electromagnetic waves also carry 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.)

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

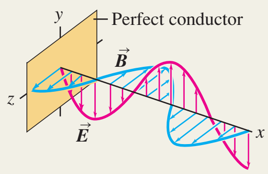
$$I = S_{av} = \frac{E_{max}B_{max}}{2\mu_0} = \frac{E_{max}{}^2}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}} E_{max}{}^2 = \frac{1}{2}\epsilon_0 c E_{max}{}^2 \qquad (32.29)$$

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

(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.)



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? Assume that the plane of the antenna loop is perpendicular to the direction of the radiation's magnetic field and that the source radiates uniformly in all directions.

SOLUTION GUIDE

See MasteringPhysics® study area for a Video Tutor solution.

IDENTIFY and SET UP:

1. The electromagnetic wave has an oscillating magnetic field. This causes a magnetic flux through the loop antenna that varies sinusoidally with time. By Faraday's law, this produces an emf equal in magnitude to the rate of change of the flux. The target variable is the magnitude of this emf.
2. Select the equations that you will need to find (i) the intensity of the wave at the position of the loop, a distance $r = 2.50$ km from the source of power $P = 55.0$ kW; (ii) the amplitude of the sinusoidally varying magnetic field at that position; (iii) the magnetic flux through the loop as a function of time; and (iv) the emf produced by the flux.

EXECUTE

3. Find the wave intensity at the position of the loop.
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 be able to pick up signals from the source.)

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **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 According to Ampere's law, is it possible to have both a conduction current and a displacement current at the same time? Is it possible for the effects of the two kinds of current to cancel each other exactly so that *no* magnetic field is produced? Explain.

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?

Q32.14 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?

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. 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 not harmful to the skin and is necessary for the production of vitamin D. UVB, with a wavelength between 280 nm and 320 nm, is much more dangerous because it causes skin cancer. (a) Find the frequency ranges of UVA and UVB. (b) What are the ranges of the wave numbers for UVA and UVB?

32.7 • A sinusoidal electromagnetic wave having a magnetic field of amplitude 1.25 μT and a wavelength of 432 nm is traveling in the $+x$-direction through empty space. (a) What is the frequency of this wave? (b) What is the amplitude of the associated electric field? (c) Write the equations for the electric and magnetic fields as functions of x and t in the form of Eqs. (32.17).

32.8 • An electromagnetic wave of wavelength 435 nm is traveling in vacuum in the $-z$-direction. The electric field has amplitude 2.70×10^{-3} V/m and is parallel to the x-axis. What are (a) the frequency and (b) the magnetic-field amplitude? (c) Write the vector equations for $\vec{E}(z, t)$ and $\vec{B}(z, t)$.

32.9 • Consider electromagnetic waves propagating in air. (a) Determine the frequency of a wave with a wavelength of (i) 5.0 km, (ii) 5.0 μm, (iii) 5.0 nm. (b) What is the wavelength (in meters and nanometers) of (i) gamma rays of frequency 6.50×10^{21} Hz and (ii) an AM station radio wave of frequency 590 kHz?

32.10 • The electric field of a sinusoidal electromagnetic wave obeys the equation $E = (375 \text{ V/m})\cos[(1.99 \times 10^7 \text{ rad/m})x + (5.97 \times 10^{15} \text{ rad/s})t]$. (a) What are the amplitudes of the electric and magnetic fields of this wave? (b) What are the frequency, wavelength, and period of the wave? Is this light visible to humans? (c) What is the speed of the wave?

32.11 • An electromagnetic wave has an electric field given by $\vec{E}(y, t) = (3.10 \times 10^5 \text{ V/m})\hat{k}\cos[ky - (12.65 \times 10^{12} \text{ rad/s})t]$. (a) In which direction is the wave traveling? (b) What is the wavelength of the wave? (c) Write the vector equation for $\vec{B}(y, t)$.

32.12 • An electromagnetic wave has a magnetic field given by $\vec{B}(x, t) = -(8.25 \times 10^{-9} \text{ T})\hat{\jmath}\cos[(1.38 \times 10^4 \text{ rad/m})x + \omega t]$. (a) In which direction is the wave traveling? (b) What is the frequency f of the wave? (c) Write the vector equation for $\vec{E}(x, t)$.

32.13 • Radio station WCCO in Minneapolis broadcasts at a frequency of 830 kHz. At a point some distance from the transmitter, the magnetic-field amplitude of the electromagnetic wave from WCCO is 4.82×10^{-11} T. Calculate (a) the wavelength; (b) the wave number; (c) the angular frequency; (d) the electric-field amplitude.

32.14 • An electromagnetic wave with frequency 65.0 Hz travels in an insulating magnetic material that has dielectric constant 3.64 and relative permeability 5.18 at this frequency. The electric field has amplitude 7.20×10^{-3} V/m. (a) What is the speed of propagation of the wave? (b) What is the wavelength of the wave? (c) What is the amplitude of the magnetic field?

32.15 • An electromagnetic wave with frequency 5.70×10^{14} Hz propagates with a speed of 2.17×10^8 m/s in a certain piece of glass. Find (a) the wavelength of the wave in the glass; (b) the wavelength of a wave of the same frequency propagating in air; (c) the index of refraction n of the glass for an electromagnetic wave with this frequency; (d) the dielectric constant for glass at this frequency, assuming that the relative permeability is unity.

Section 32.4 Energy and Momentum in Electromagnetic Waves

32.16 • BIO **High-Energy Cancer Treatment.** Scientists are working on a new technique to kill cancer cells by zapping them with ultrahigh-energy (in the range of 10^{12} W) pulses of light that last for an extremely short time (a few nanoseconds). These short pulses scramble the interior of a cell without causing it to explode, as long pulses would do. We can model a typical such cell as a disk 5.0 μm in diameter, with the pulse lasting for 4.0 ns with an average power of 2.0×10^{12} W. We shall assume that the energy is spread uniformly over the faces of 100 cells for each pulse. (a) How much energy is given to the cell during this pulse? (b) What is the intensity (in W/m^2) delivered to the cell? (c) What are the maximum values of the electric and magnetic fields in the pulse?

32.17 •• **Fields from a Light Bulb.** We can reasonably model a 75-W incandescent light bulb as a sphere 6.0 cm in diameter. Typically, only about 5% of the energy goes to visible light; the rest goes largely to nonvisible infrared radiation. (a) What is the visible-light intensity (in W/m^2) at the surface of the bulb? (b) What are the amplitudes of the electric and magnetic fields at this surface, for a sinusoidal wave with this intensity?

32.18 •• A sinusoidal electromagnetic wave from a radio station passes perpendicularly through an open window that has area 0.500 m^2. At the window, the electric field of the wave has rms value 0.0200 V/m. How much energy does this wave carry through the window during a 30.0-s commercial?

32.19 •• **Testing a Space Radio Transmitter.** You are a NASA mission specialist on your first flight aboard the space shuttle. Thanks to your extensive training in physics, you have been assigned to evaluate the performance of a new radio transmitter on board the International Space Station (ISS). Perched on the shuttle's movable arm, you aim a sensitive detector at the ISS, which is 2.5 km away. You find that the electric-field amplitude of the radio waves coming from the ISS transmitter is 0.090 V/m and that the frequency of the waves is 244 MHz. Find the following: (a) the intensity of the radio wave at your location; (b) the magnetic-field amplitude of the wave at your location; (c) the total power output of the ISS radio transmitter. (d) What assumptions, if any, did you make in your calculations?

32.20 • The intensity of a cylindrical laser beam is $0.800\ \text{W/m}^2$. The cross-sectional area of the beam is $3.0 \times 10^{-4}\ \text{m}^2$ and the intensity is uniform across the cross section of the beam. (a) What is the average power output of the laser? (b) What is the rms value of the electric field in the beam?

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

32.22 • A sinusoidal electromagnetic wave emitted by a cellular phone has a wavelength of 35.4 cm and an electric-field amplitude of 5.40×10^{-2} V/m at a distance of 250 m from the phone. Calculate (a) the frequency of the wave; (b) the magnetic-field amplitude; (c) the intensity of the wave.

32.23 • A monochromatic light source with power output 60.0 W radiates light of wavelength 700 nm uniformly in all directions. Calculate $E_{\max}$ and $B_{\max}$ for the 700-nm light at a distance of 5.00 m from the source.

32.24 • For the electromagnetic wave represented by Eqs. (32.19), show that the Poynting vector (a) is in the same direction as the propagation of the wave and (b) has average magnitude given by Eqs. (32.29).

32.25 •• An intense light source radiates uniformly in all directions. At a distance of 5.0 m from the source, the radiation pressure on a perfectly absorbing surface is 9.0×10^{-6} Pa. What is the total average power output of the source?

32.26 • **Television Broadcasting.** Public television station KQED in San Francisco broadcasts a sinusoidal radio signal at a power of 316 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?

32.27 •• BIO **Laser Safety.** If the eye receives an average intensity greater than $1.0 \times 10^2\ \text{W/m}^2$, 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.

32.28 • In the 25-ft Space Simulator facility at NASA's Jet Propulsion Laboratory, a bank of overhead arc lamps can produce light of intensity $2500\ \text{W/m}^2$ 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.

32.29 • **Laboratory Lasers.** He–Ne lasers are often used in physics demonstrations. They produce light of wavelength 633 nm and a power of 0.500 mW spread over a cylindrical beam 1.00 mm in diameter (although these quantities can vary). (a) What is the intensity of this laser beam? (b) What are the maximum values of the electric and magnetic fields? (c) What is the average energy density in the laser beam?

32.30 •• **Solar Sail 1.** During 2004, Japanese scientists successfully tested two solar sails. One had a somewhat complicated shape that we shall model as a disk 9.0 m in diameter and 7.5 μm thick. The intensity of solar energy at that location was about 1400 W/m^2. (a) What force did the sun's light exert on this sail, assuming that it struck perpendicular to the sail and that the sail was perfectly reflecting? (b) If the sail was made of magnesium, of density $1.74\ \text{g/cm}^3$, what acceleration would the sun's radiation give to the sail? (c) Does the acceleration seem large enough to be feasible for space flight? In what ways could the sail be modified to increase its acceleration?

Section 32.5 Standing Electromagnetic Waves

32.31 • **Microwave Oven.** The microwaves in a certain microwave oven have a wavelength of 12.2 cm. (a) How wide must this oven be so that it will contain five antinodal planes of the electric field along its width in the standing-wave pattern? (b) What is the frequency of these microwaves? (c) Suppose a manufacturing error occurred and the oven was made 5.0 cm longer than specified in part (a). In this case, what would have to be the frequency of the microwaves for there still to be five antinodal planes of the electric field along the width of the oven?

32.32 • An electromagnetic standing wave in air of frequency 750 MHz is set up between two conducting planes 80.0 cm apart. At which positions between the planes could a point charge be placed at rest so that it would *remain* at rest? Explain.

32.33 • A standing electromagnetic wave in a certain material has frequency 2.20×10^{10} Hz. The nodal planes of $\vec{B}$ are 3.55 mm apart. Find (a) the wavelength of the wave in this material; (b) the distance between adjacent nodal planes of the $\vec{E}$ field; (c) the speed of propagation of the wave.

32.34 • 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 $\vec{B}$?

32.35 • An electromagnetic standing wave in a certain material has frequency 1.20×10^{10} Hz and speed of propagation 2.10×10^8 m/s. (a) What is the distance between a nodal plane of $\vec{B}$ and the closest antinodal plane of $\vec{B}$? (b) What is the distance between an antinodal plane of $\vec{E}$ and the closest antinodal plane of $\vec{B}$? (c) What is the distance between a nodal plane of $\vec{E}$ and the closest nodal plane of $\vec{B}$?

32.36 • CALC Show that the electric and magnetic fields for standing waves given by Eqs. (32.34) and (32.35) (a) satisfy the wave equation, Eq. (32.15), and (b) satisfy Eqs. (32.12) and (32.14).

PROBLEMS

32.37 •• 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 does the pulse of the laser beam exert on the retina as it is fully absorbed by the circular spot? (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?

32.38 •• 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.)

32.39 •• You want to support a sheet of fireproof paper horizontally, using only a vertical upward beam of light spread uniformly over the sheet. There is no other light on this paper. The sheet measures 22.0 cm by 28.0 cm and has a mass of 1.50 g. (a) If the paper is black and hence absorbs all the light that hits it, what must be the intensity of the light beam? (b) For the light in part (a), what are the amplitudes of its electric and magnetic fields? (c) If the paper is white and hence reflects all the light that hits it, what intensity of light beam is needed to support it? (d) To see if it is physically reasonable to expect to support a sheet of paper this way, calculate the intensity in a typical 0.500-mW laser beam that is 1.00 mm in diameter, and compare this value with your answer in part (a).

32.40 •• For a sinusoidal electromagnetic wave in vacuum, such as that described by Eq. (32.16), show that the *average* energy density in the electric field is the same as that in the magnetic field.

32.41 • 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?

32.42 • A plane sinusoidal electromagnetic wave in air has a wavelength of 3.84 cm and an $\vec{E}$-field amplitude of 1.35 V/m. (a) What is the frequency? (b) What is the $\vec{B}$-field amplitude? (c) What is the intensity? (d) What average force does this radiation exert on a totally absorbing surface with area 0.240 m^2 perpendicular to the direction of propagation?

32.43 • A small helium–neon laser emits red visible light with a power of 4.60 mW in a beam that has a diameter of 2.50 mm. (a) What are the amplitudes of the electric and magnetic fields of the 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?

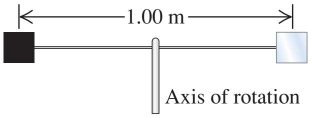
32.44 •• The electric-field component of a sinusoidal electromagnetic wave traveling through a plastic cylinder is given by the equation $E = (5.35\ \text{V/m})\cos[(1.39 \times 10^7\ \text{rad/m})x - (3.02 \times 10^{15}\ \text{rad/s})t]$. (a) Find the frequency, wavelength, and speed of this wave in the plastic. (b) What is the index of refraction of the plastic? (c) Assuming that the amplitude of the electric field does not change, write a comparable equation for the electric field if the light is traveling in air instead of in plastic.

32.45 • 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?

32.46 •• A source of sinusoidal electromagnetic waves radiates uniformly in all directions. At 10.0 m from this source, the amplitude of the electric field is measured to be 1.50 N/C. What is the electric-field amplitude at a distance of 20.0 cm from the source?

32.47 •• **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. P32.47). 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 **P32.47**



32.48 •• **CP** A circular loop of wire has radius 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 0.0195 W/m^2, and the wavelength of the wave is 6.90 m. What is the maximum emf induced in the loop?

32.49 • **CALC** **CP** A cylindrical conductor with a circular cross section has a radius a and a resistivity ρ and carries a constant current I. (a) What are the magnitude and direction of the electric-field vector $\vec{E}$ at a point just inside the wire at a distance a from the axis? (b) What are the magnitude and direction of the magnetic-field vector $\vec{B}$ at the same point? (c) What are the magnitude and direction of the Poynting vector $\vec{S}$ at the same point? (The direction of $\vec{S}$ is the direction in which electromagnetic energy flows into or out of the conductor.) (d) Use the result in part (c) to find the rate of flow of energy into the volume occupied by a length l of the conductor. (*Hint:* Integrate $\vec{S}$ over the surface of this volume.) Compare your result to the rate of generation of thermal energy in the same volume. Discuss why the energy dissipated in a current-carrying conductor, due to its resistance, can be thought of as entering through the cylindrical sides of the conductor.

32.50 • 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?

32.51 •• **CP** **Flashlight to the Rescue.** You are the sole crew member of the interplanetary spaceship *T:1339 Vorga,* which makes regular cargo runs between the earth and the mining colonies in the asteroid belt. You are working outside the ship one day while at a distance of 2.0 AU from the sun. [1 AU (astronomical unit) is the average distance from the earth to the sun, 149,600,000 km.] Unfortunately, you lose contact with the ship's hull and begin to drift away into space. You use your spacesuit's rockets to try to push yourself back toward the ship, but they run out of fuel and stop working before you can return to the ship. You find yourself in an awkward position, floating 16.0 m from the spaceship with zero velocity relative to it. Fortunately, you are carrying a 200-W flashlight. You turn on the flashlight and use its beam as a "light rocket" to push yourself back toward the ship. (a) If you, your spacesuit, and the flashlight have a combined mass of 150 kg, how long will it take you to get back to the ship? (b) Is there another way you could use the flashlight to accomplish the same job of returning you to the ship?

32.52 • The 19th-century inventor Nikola Tesla proposed to transmit electric power via sinusoidal electromagnetic waves. Suppose power is to be transmitted in a beam of cross-sectional area 100 m^2. What electric- and magnetic-field amplitudes are required to transmit an amount of power comparable to that handled by modern transmission lines (that carry voltages and currents of the order of 500 kV and 1000 A)?

32.53 •• **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?

32.54 •• **CP** **Solar Sail 2.** 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.

32.55 •• **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).]

CHALLENGE PROBLEMS

32.56 ••• **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 $\vec{E}(x, t) = E_y(x, t)\hat{j}$ propagating in the $+x$-direction within a conductor is

$$\frac{\partial^2 E_y\,(x, t)}{\partial x^2} = \frac{\mu}{\rho}\frac{\partial E_y(x, t)}{\partial t}$$

where μ is the permeability of the conductor and ρ is its resistivity. (a) A solution to this wave equation is

$$E_y(x, t) = E_{\max} e^{-k_C x}\cos(k_C x - \omega t)$$

where $k_C = \sqrt{\omega\mu/2\rho}$. Verify this by substituting $E_y(x, t)$ into the above wave equation. (b) The exponential term shows that the electric field decreases in amplitude as it propagates. Explain why this happens. (*Hint:* The field does work to move charges within the conductor. The current of these moving charges causes i^2R heating within the conductor, raising its temperature. Where does the energy to do this come from?) (c) Show that the electric-field amplitude decreases by a factor of $1/e$ in a distance $1/k_C = \sqrt{2\rho/\omega\mu}$, and calculate this distance for a radio wave with frequency f =1.0 MHz in copper (resistivity $1.72 \times 10^{-8}\ \Omega \cdot$ m; permeability $\mu = \mu_0$). Since this distance is so short, electromagnetic waves of this frequency can hardly propagate at all into copper. Instead, they are reflected at the surface of the metal. This is why radio waves cannot penetrate through copper or other metals, and why radio reception is poor inside a metal structure.

32.57 ••• **CP** Electromagnetic radiation is emitted by accelerating charges. The rate at which energy is emitted from an accelerating charge that has charge q and acceleration a is given by

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

where c is the speed of light. (a) Verify that this equation is dimensionally correct. (b) If a proton with a kinetic energy of 6.0 MeV is traveling in a particle accelerator in a circular orbit of radius 0.750 m, what fraction of its energy does it radiate per second? (c) Consider an electron orbiting with the same speed and radius. What fraction of its energy does it radiate per second?

32.58 ••• **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 32.57)? What does this tell you about the use of classical physics in describing the atom?

Answers

Chapter Opening Question ?

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, this reflected wave is just as intense as the incident wave. Tarnished metals are less shiny because their surface is oxidized and less conductive; polishing the metal removes the oxide and exposes the conducting metal.

Test Your Understanding Questions

32.1 Answers: (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. (29.20), 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. (29.21).

32.2 Answers: (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 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.

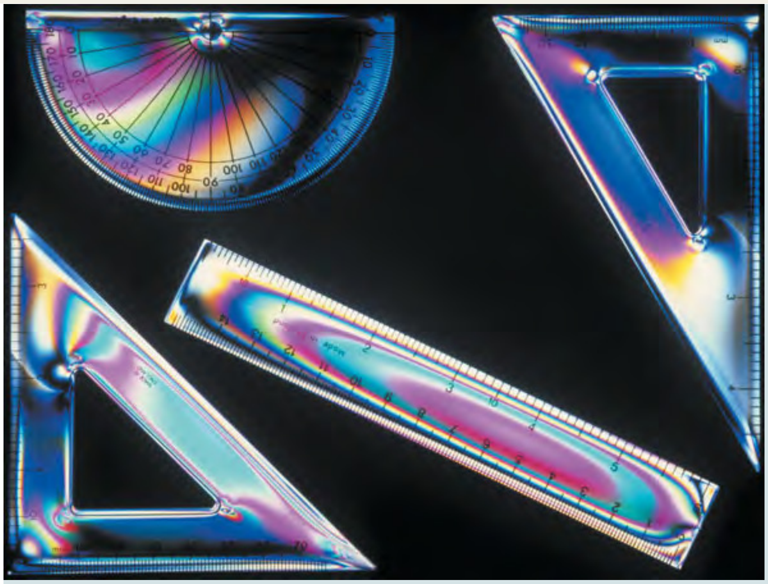
32.4 Answers: (a) (i) and (iii), (b) (ii) and (iv), (c) (i) and (iii), (d) (ii) and (iv) Both the energy density u and the Poynting vector magnitude S are maximum where the $\vec{E}$ and $\vec{B}$ fields have their maximum magnitudes. (The directions of the fields doesn't matter.) From Fig. 32.13, this 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 Answer: no There are places where $\vec{E} = \mathbf{0}$ at all times (at the walls) and the electric energy density $\frac{1}{2}\epsilon_0 E^2$ is always zero. There are also places where $\vec{B} = \mathbf{0}$ at all times (on the plane midway between the walls) and the magnetic energy density $B^2/2\mu_0$ is always zero. However, there are *no* locations where both $\vec{E}$ and $\vec{B}$ are always zero. Hence the energy density at any point in the standing wave is always nonzero.

Bridging Problem

Answer: 0.0368 V

33 THE NATURE AND PROPAGATION OF LIGHT



? These drafting tools are made of clear plastic, but a rainbow of colors appears when they are placed between two special filters called polarizers. How does this cause the colors?

LEARNING GOALS

By studying this chapter, you will learn:

- What light rays are, and how they are related to wave fronts.
- The laws that govern the reflection and refraction of light.
- The circumstances under which light is totally reflected at an interface.
- How to make polarized light out of ordinary light.
- How Huygens's principle helps us analyze reflection and refraction.

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, optical computers, and new techniques in medical imaging.

The importance of optics to physics, and to science and engineering in general, is so great that we will devote the next four chapters to its study. In this chapter we begin with a study of the laws of reflection and refraction and the concepts of dispersion, polarization, and scattering of light. Along the way we compare the various possible descriptions of light in terms of particles, rays, or waves, and we introduce Huygens's principle, an important link that connects the ray and wave viewpoints. In Chapter 34 we'll use the ray description of light to understand how mirrors and lenses work, and we'll see how mirrors and lenses are used in optical instruments such as cameras, microscopes, and telescopes. We'll explore the wave characteristics of light further in Chapters 35 and 36.

33.1 The Nature of Light

Until the time of Isaac Newton (1642–1727), most scientists thought that light consisted of streams of particles (called *corpuscles*) emitted by light sources. Galileo and others tried (unsuccessfully) to measure the speed of light. Around 1665, evidence of *wave* properties of light began to be discovered. By the early 19th century, evidence that light is a wave had grown very persuasive.

In 1873, James Clerk Maxwell predicted the existence of electromagnetic waves and calculated their speed of propagation, as we learned in Chapter 32. This development, along with the experimental work of Heinrich Hertz starting in 1887, showed conclusively that light is indeed an electromagnetic wave.

The Two Personalities of Light

The wave picture of light is not the whole story, however. Several effects associated with emission and absorption of light reveal a particle aspect, in that the energy carried by light waves is packaged in discrete bundles called *photons* or *quanta.* These apparently contradictory wave and particle properties have been reconciled since 1930 with the development of quantum electrodynamics, a comprehensive theory that includes *both* wave and particle properties. The *propagation* of light is best described by a wave model, but understanding emission and absorption requires a particle approach.

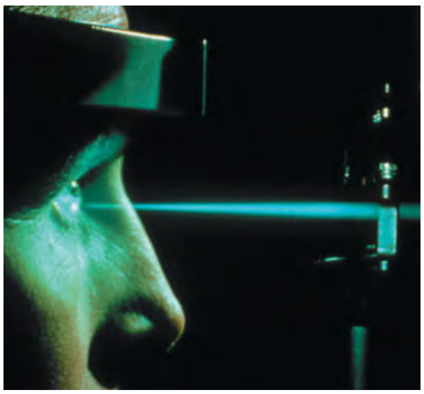
The fundamental sources of all electromagnetic radiation are electric charges in accelerated motion. All bodies emit electromagnetic radiation as a result of thermal motion of their molecules; this radiation, called *thermal radiation,* is a mixture of different wavelengths. At sufficiently high temperatures, all matter emits enough visible light to be self-luminous; a very hot body appears "red-hot" (Fig. 33.1) or "white-hot." Thus hot matter in any form is a light source. Familiar examples are a candle flame, hot coals in a campfire, the coils in an electric room heater, and an incandescent lamp filament (which usually operates at a temperature of about 3000°C).

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

Light is also produced during electrical discharges through ionized gases. The bluish light of mercury-arc lamps, the orange-yellow of sodium-vapor lamps, and the various colors of "neon" signs are familiar. A variation of the mercury-arc lamp is the *fluorescent* lamp (see Fig. 30.7). This light source uses a material called a *phosphor* to convert the ultraviolet radiation from a mercury arc into visible light. This conversion makes fluorescent lamps more efficient than incandescent lamps in transforming electrical energy into light.

In most light sources, light is emitted independently by different atoms within the source; in a *laser,* by contrast, atoms are induced to emit light in a cooperative, coherent fashion. The result is a very narrow beam of radiation that can be enormously intense and that is much more nearly *monochromatic,* or single-frequency, than light from any other source. Lasers are used by physicians for microsurgery, in a DVD or Blu-ray player to scan the information recorded on a video disc, in industry to cut through steel and to fuse high-melting-point materials, and in many other applications (Fig. 33.2).

33.2 Ophthalmic surgeons use lasers for repairing detached retinas and for cauterizing blood vessels in retinopathy. Pulses of blue-green light from an argon laser are ideal for this purpose, since they pass harmlessly through the transparent part of the eye but are absorbed by red pigments in the retina.



No matter what its source, electromagnetic radiation travels in vacuum at the same speed. As we saw in Sections 1.3 and 32.1, the speed of light in vacuum is defined to be

$$c = 2.99792458 \times 10^8 \text{ m/s}$$

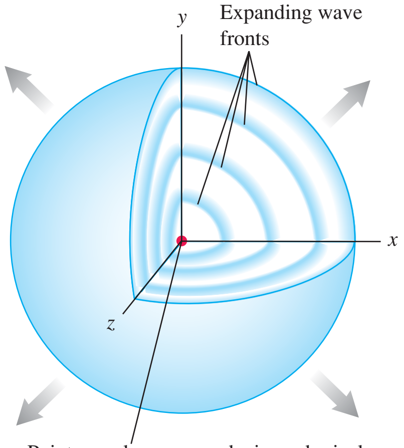
or 3.00×10^8 m/s to three significant figures. The duration of one second is defined by the cesium clock (see Section 1.3), so one meter is defined to be the distance that light travels in 1/299,792,458 s.

Waves, Wave Fronts, and Rays

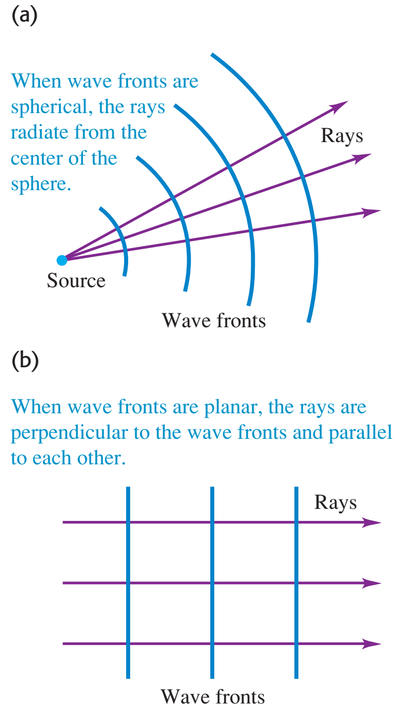
We often use the concept of a **wave front** to describe wave propagation. We introduced this concept in Section 32.2 to describe the leading edge of a wave. More generally, we define a wave front as *the locus of all adjacent points at which the phase of vibration of a physical quantity associated with 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

33.3 Spherical wave fronts of sound spread out uniformly in all directions from a point source in a motionless medium, such as still air, that has the same properties in all regions and in all directions. Electromagnetic waves in vacuum also spread out as shown here.



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



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 will often use diagrams that show the shapes of the wave fronts or their cross sections in some reference plane. For example, when electromagnetic waves are radiated by a small light source, we can represent the wave fronts as spherical surfaces concentric with the source or, as in Fig. 33.4a, by the circular intersections of these surfaces with the plane of the diagram. Far away from the source, where the radii of the spheres have become very large, a section of a spherical surface can be considered as a plane, and we have a *plane* wave like those discussed in Sections 32.2 and 32.3 (Fig, 33.4b).

To describe the directions in which light propagates, it's often convenient to represent a light wave by **rays** rather than by wave fronts. Rays were used to describe light long before its wave nature was firmly established. In a particle theory of light, rays are the paths of the particles. From the wave viewpoint *a ray is an imaginary line along the direction of travel of the wave.* In Fig. 33.4a the rays are the radii of the spherical wave fronts, and in Fig. 33.4b they are straight lines perpendicular to the wave fronts. When waves travel in a homogeneous isotropic material (a material with the same properties in all regions and in all directions), the rays are always straight lines normal to the wave fronts. At a boundary surface between two materials, such as the surface of a glass plate in air, the wave speed and the direction of a ray may change, but the ray segments in the air and in the glass are straight lines.

The next several chapters will give you many opportunities to see the interplay of the ray, wave, and particle descriptions of light. The branch of optics for which the ray description is adequate is called **geometric optics;** the branch dealing specifically with wave behavior is called **physical optics.** This chapter and the following one are concerned mostly with geometric optics. In Chapters 35 and 36 we will study wave phenomena and physical optics.

Test Your Understanding of Section 33.1 Some crystals are *not* isotropic: Light travels through the crystal at a higher speed in some directions than in others. In a crystal in which light travels at the same speed in the x- and z-directions but at a faster speed in the y-direction, what would be the shape of the wave fronts produced by a light source at the origin? (i) spherical, like those shown in Fig. 33.3; (ii) ellipsoidal, flattened along the y-axis; (iii) ellipsoidal, stretched out along the y-axis.

33.2 Reflection and Refraction

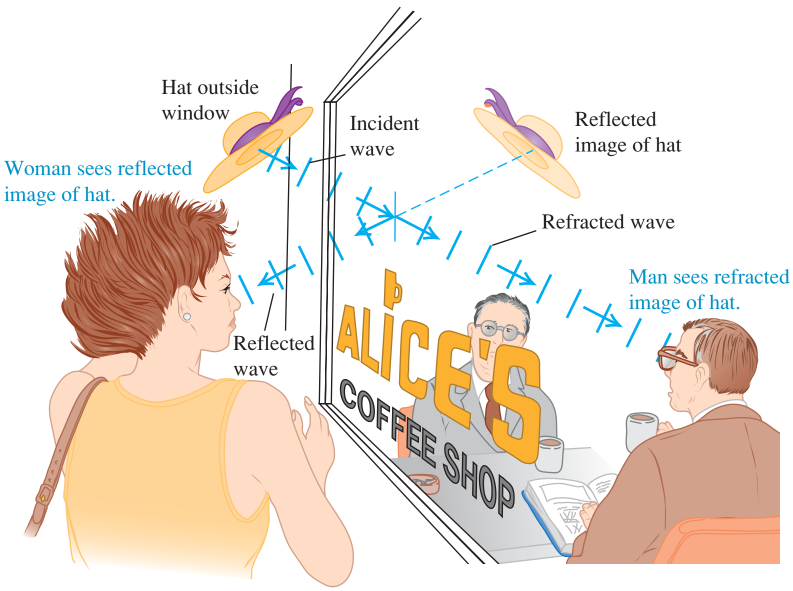
In this section we'll use the *ray* model of light to explore two of the most important aspects of light propagation: **reflection** and **refraction.** When a light wave strikes a smooth interface separating two transparent materials (such as air and glass or water and glass), the wave is in general partly *reflected* and partly *refracted* (transmitted) into the second material, as shown in Fig. 33.5a. For example, when you look into a restaurant window from the street, you see a reflection of the street scene, but a person inside the restaurant can look out through the window at the same scene as light reaches him by refraction.

MasteringPHYSICS

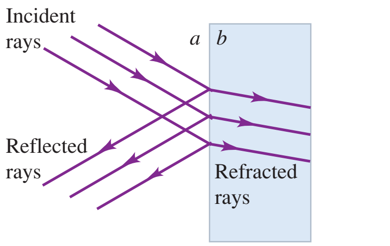
ActivPhysics 15.1: Reflection and Refraction
ActivPhysics 15.3: Refraction Applications

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.

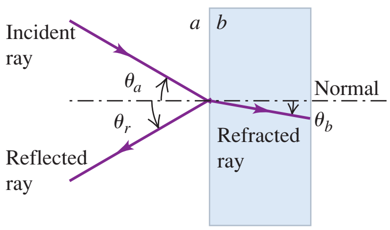
(a) Plane waves reflected and refracted from a window



(b) The waves in the outside air and glass represented by rays



(c) The representation simplified to show just one set of rays

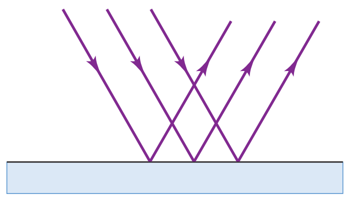


The segments of plane waves shown in Fig. 33.5a can be represented by bundles of rays forming *beams* of light (Fig. 33.5b). For simplicity we often draw only one ray in each beam (Fig. 33.5c). Representing these waves in terms of rays is the basis of geometric optics. We begin our study with the behavior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

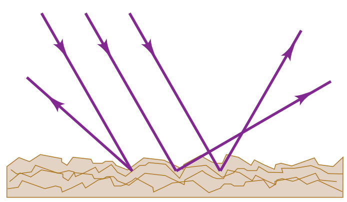
We describe the directions of the incident, reflected, and refracted (transmitted) rays at a smooth interface between two optical materials in terms of the angles they make with the *normal* (perpendicular) to the surface at the point of incidence, as shown in Fig. 33.5c. If the interface is rough, both 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.** This distinction is shown in Fig. 33.6. Both kinds of reflection can occur with either transparent materials or *opaque* materials that do not transmit light. The vast majority of objects in your environment (including plants, other people, and this book) are visible to you because they reflect light in a diffuse manner from their surfaces. Our primary concern, however, will be with specular reflection from a very smooth surface such as highly polished glass or metal. Unless stated otherwise, when referring to "reflection" we will always mean *specular* reflection.

33.6 Two types of reflection.

(a) Specular reflection



(b) Diffuse reflection



The **index of refraction** of an optical material (also called the **refractive index**), denoted by n, plays a central role in geometric optics. It is the ratio of the speed of light c in vacuum to the speed v in the material:

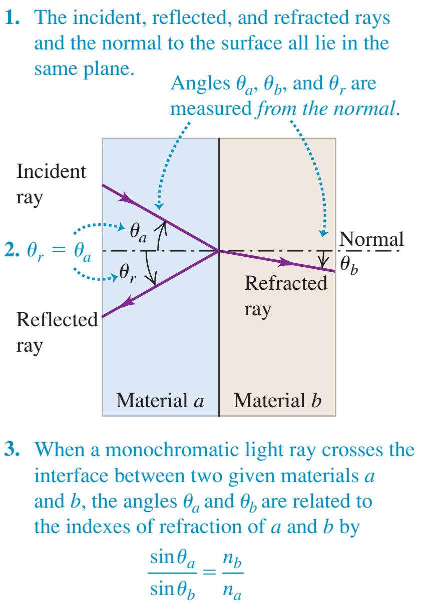
$$n = \frac{c}{v} \qquad \text{(index of refraction)} \qquad (33.1)$$

Light always travels *more slowly* in a material than in vacuum, so the value of n in anything other than vacuum is always greater than unity. For vacuum, $n = 1$.

Since n is a ratio of two speeds, it is a pure number without units. (The relationship of the value of n to the electric and magnetic properties of a material is described in Section 32.3.)

CAUTION **Wave speed and index of refraction** Keep in mind that the wave speed v is *inversely* proportional to the index of refraction n. The greater the index of refraction in a material, the *slower* the wave speed in that material. Failure to remember this point can lead to serious confusion!

33.7 The laws of reflection and refraction.



The Laws of Reflection and Refraction

Experimental studies of the directions of the incident, reflected, and refracted rays 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.** The plane of the three rays and the normal, 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,

$$\theta_r = \theta_a \qquad \text{(law of reflection)} \tag{33.2}$$

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

$$n_a \sin\theta_a = n_b \sin\theta_b \qquad \text{(law of refraction)} \tag{33.4}$$

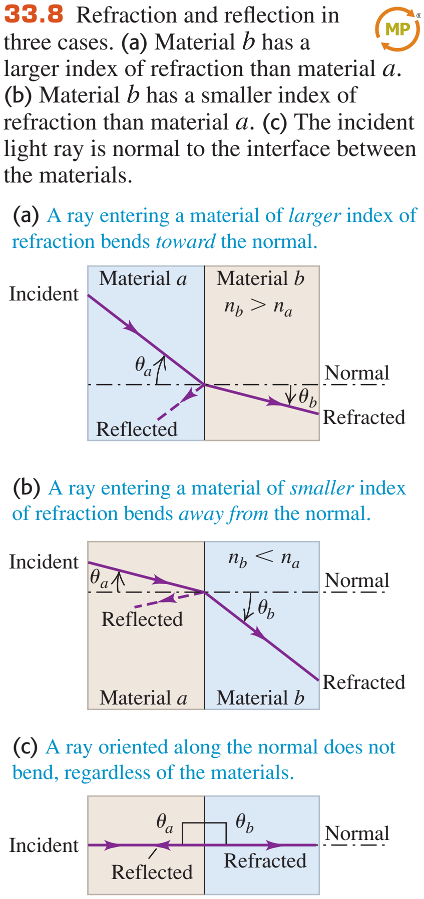
This experimental result, together with the observation that the incident and refracted rays and the normal all lie in the same plane, is called the **law of refraction** or **Snell's law,** after the Dutch scientist Willebrord Snell (1591–1626). There is some doubt that Snell actually discovered it. 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

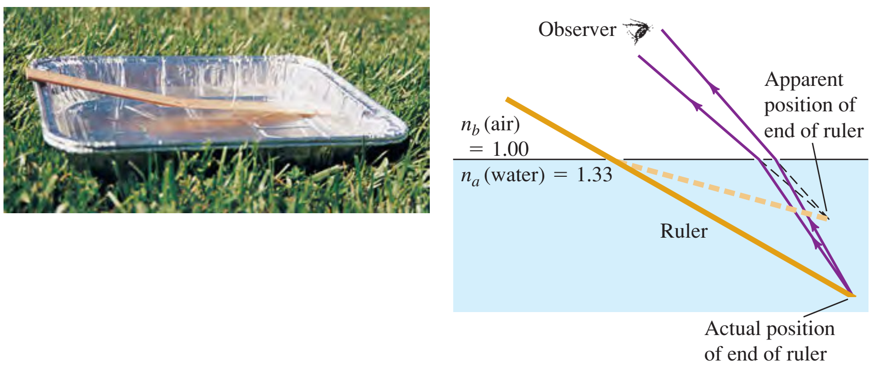
33.8 Refraction and reflection in three cases. (a) Material b has a larger index of refraction than material a. (b) Material b has a smaller index of refraction than material a. (c) The incident light ray is normal to the interface between the materials.



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



$\theta_a = 0$ and $\sin\theta_a = 0$, so from Eq. (33.4) θ_b is also equal to zero, so the transmitted ray is also normal to the interface. Equation (33.2) shows that θ_r, too, is equal to zero, so the reflected ray travels back along the same path as the incident ray.

The law of refraction explains why a partially submerged ruler or drinking straw appears bent; light rays coming from below the surface change in direction at the air–water interface, so the rays appear to be coming from a position above 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; these two rays, the incident ray, and the normal to the surface again lie in the same plane. 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 using Eq. (33.4).] Since reflected and incident rays make the same angle with the normal, the path of a reflected ray is also reversible. That's why when you see someone's eyes in a mirror, they can also see you.

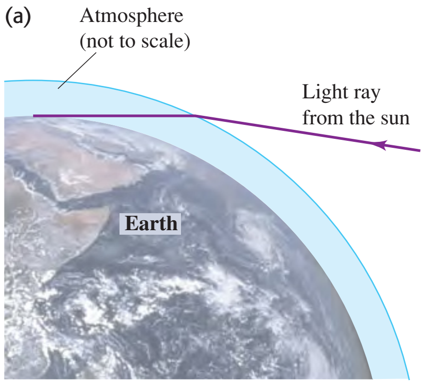
The *intensities* of the reflected and refracted rays depend on the angle of incidence, the two indexes of refraction, and the polarization (that is, the direction of the electric-field vector) of the incident ray. The fraction reflected is smallest at normal incidence ($\theta_a = 0°$), where it is about 4% for an air–glass interface. This fraction increases with increasing angle of incidence to 100% at grazing incidence, when $\theta_a = 90°$.

It's possible to use Maxwell's equations to predict the amplitude, intensity, phase, and polarization states of the reflected and refracted waves. Such an analysis is beyond our scope, however.

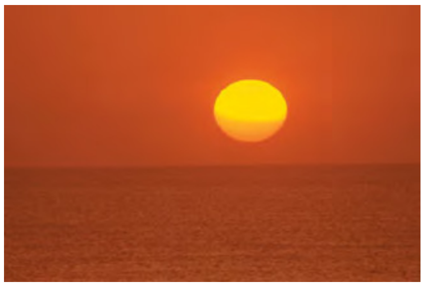
The index of refraction depends not only on the substance but also on the wavelength of the light. The dependence on wavelength is called *dispersion;* we will consider it in Section 33.4. Indexes of refraction for several solids and liquids are given in Table 33.1 for a particular wavelength of yellow light.

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 33.55.)

(a)



(b)



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.

Table 33.1 Index of Refraction for Yellow Sodium Light, $\lambda_0 = 589$ nm

Substance	Index of Refraction, n
Solids	
Ice (H_2O)	1.309
Fluorite (CaF_2)	1.434
Polystyrene	1.49
Rock salt (NaCl)	1.544
Quartz (SiO_2)	1.544
Zircon ($ZrO_2 \cdot SiO_2$)	1.923
Diamond (C)	2.417
Fabulite ($SrTiO_3$)	2.409
Rutile (TiO_2)	2.62
Glasses (typical values)	
Crown	1.52
Light flint	1.58
Medium flint	1.62
Dense flint	1.66
Lanthanum flint	1.80
Liquids at 20°C	
Methanol (CH_3OH)	1.329
Water (H_2O)	1.333
Ethanol (C_2H_5OH)	1.36
Carbon tetrachloride (CCl_4)	1.460
Turpentine	1.472
Glycerine	1.473
Benzene	1.501
Carbon disulfide (CS_2)	1.628

The index of refraction of air at standard temperature and pressure is about 1.0003, and we will usually take it to be exactly unity. The index of refraction of a gas increases as its density increases. Most glasses used in optical instruments have indexes of refraction between about 1.5 and 2.0. A few substances have larger indexes; one example is diamond, with 2.417.

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. It's also important to see what happens to the *wave* characteristics of the light 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

$$\lambda = \frac{\lambda_0}{n} \quad \text{(wavelength of light in a material)} \qquad (33.5)$$

When a wave passes from one material into a second material with larger index of refraction, so that $n_b > n_a$, 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 that $n_b < n_a$, 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

MP

IDENTIFY *the relevant concepts:* Use geometric optics, discussed in this section, whenever light (or electromagnetic radiation of *any* frequency and wavelength) encounters a boundary between materials. In general, part of the light is reflected back into the first material and part is refracted into the second material.

SET UP *the problem* using the following steps:

1. In problems involving rays and angles, start by drawing a large, neat diagram. Label all known angles and indexes of refraction.
2. Identify the target variables.

EXECUTE *the solution* as follows:

1. Apply the laws of reflection, Eq. (33.2), and refraction, Eq. (33.4). Measure angles of incidence, reflection, and refraction with respect to the *normal* to the surface, *never* from the surface itself.
2. Apply geometry or trigonometry in working out angular relationships. Remember that the sum of the acute angles of a right triangle is 90° (they are *complementary*) and the sum of the interior angles in any triangle is 180°.
3. The frequency of the electromagnetic radiation does not change when it moves from one material to another; the wavelength changes in accordance with Eq. (33.5), $\lambda = \lambda_0/n$.

EVALUATE *your answer:* In problems that involve refraction, check that your results are consistent with Snell's law ($n_a \sin\theta_a = n_b \sin\theta_b$). If the second material has a higher index of refraction than the first, the angle of refraction must be *smaller* than the angle of incidence: The refracted ray bends toward the normal. If the first material has the higher index of refraction, the refracted angle must be larger than the incident angle: The refracted ray bends away from the normal.

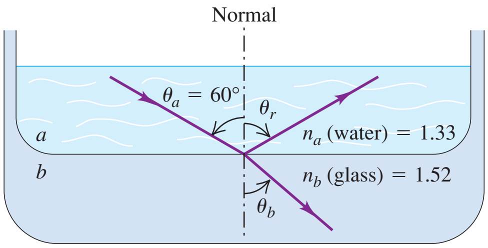
Example 33.1 Reflection and refraction

In Fig. 33.11, material a is water and material b is glass with index of refraction 1.52. The incident ray makes an angle of 60.0° with the normal; find the directions of the reflected and refracted rays.

SOLUTION

IDENTIFY and SET UP: This is a problem in geometric optics. We are given the angle of incidence $\theta_a = 60.0°$ and the indexes of refraction $n_a = 1.33$ and $n_b = 1.52$. We must find the angles of reflection and refraction θ_r and θ_b; to do this we use Eqs. (33.2) and (33.4), respectively. Figure 33.11 shows the rays and angles; n_b is slightly greater than n_a, so by Snell's law [Eq. (33.4)] θ_b is slightly smaller than θ_a, as the figure shows.

33.11 Reflection and refraction of light passing from water to glass.



EXECUTE: According to Eq. (33.2), the angle the reflected ray makes with the normal is the same as that of the incident ray, so $\theta_r = \theta_a = 60.0°$.

To find the direction of the refracted ray we use Snell's law, Eq. (33.4):

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

$$\sin\theta_b = \frac{n_a}{n_b}\sin\theta_a = \frac{1.33}{1.52}\sin 60.0° = 0.758$$

$$\theta_b = \arcsin(0.758) = 49.3°$$

EVALUATE: The second material has a larger refractive index than the first, as in Fig. 33.8a. Hence the refracted ray is bent toward the normal and $\theta_b < \theta_a$.

Example 33.2 Index of refraction in the eye

The wavelength of the red light from a helium-neon laser is 633 nm in air but 474 nm in the aqueous humor inside your eyeball. Calculate the index of refraction of the aqueous humor and the speed and frequency of the light in it.

SOLUTION

IDENTIFY and SET UP: The key ideas here are (i) the definition of index of refraction n in terms of the wave speed v in a medium and the speed c in vacuum, and (ii) the relationship between wavelength λ_0 in vacuum and wavelength λ in a medium of index n. We use Eq. (33.1), $n = c/v$; Eq. (33.5), $\lambda = \lambda_0/n$; and $v = \lambda f$.

EXECUTE: The index of refraction of air is very close to unity, so we assume that the wavelength λ_0 in vacuum is the same as that in air, 633 nm. Then from Eq. (33.5),

$$\lambda = \frac{\lambda_0}{n} \qquad n = \frac{\lambda_0}{\lambda} = \frac{633\text{ nm}}{474\text{ nm}} = 1.34$$

This is about the same index of refraction as for water. Then, using $n = c/v$ and $v = \lambda f$, we find

$$v = \frac{c}{n} = \frac{3.00 \times 10^8\text{ m/s}}{1.34} = 2.25 \times 10^8\text{ m/s}$$

$$f = \frac{v}{\lambda} = \frac{2.25 \times 10^8\text{ m/s}}{474 \times 10^{-9}\text{ m}} = 4.74 \times 10^{14}\text{ Hz}$$

EVALUATE: Note that while the speed and wavelength have different values in air and in the aqueous humor, the *frequency* in air, f_0, is the same as the frequency f in the aqueous humor:

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

When a light wave passes from one material into another, the wave speed and wavelength both change but the wave frequency is unchanged.

Example 33.3 A twice-reflected ray

Two mirrors are perpendicular to each other. A ray traveling in a plane perpendicular to both mirrors is reflected from one mirror, then the other, as shown in Fig. 33.12. What is the ray's final direction relative to its original direction?

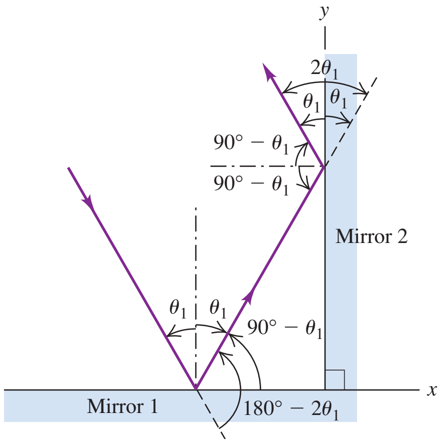
SOLUTION

IDENTIFY and SET UP: This problem involves the law of reflection, which we must apply twice (once for each mirror).

EXECUTE: For mirror 1 the angle of incidence is θ_1, and this equals the angle of reflection. The sum of interior angles in the triangle shown in the figure is 180°, so we see that the angles of incidence and reflection for mirror 2 are both $90° - \theta_1$. The total change in direction of the ray after both reflections is therefore $2(90° - \theta_1) + 2\theta_1 = 180°$. That is, the ray's final direction is opposite to its original direction.

Continued

33.12 A ray moving in the *xy*-plane. The first reflection changes the sign of the *y*-component of its velocity, and the second reflection changes the sign of the *x*-component. For a different ray with a *z*-component of velocity, a third mirror (perpendicular to the two shown) could be used to change the sign of that component.



EVALUATE: An alternative viewpoint is that reflection reverses the sign of the component of light velocity perpendicular to the surface but leaves the other components unchanged. We invite you to verify this in detail. You should also be able to use this result to show that when a ray of light is successively reflected by three mirrors forming a corner of a cube (a "corner reflector"), its final direction is again opposite to its original direction. This principle is widely used in tail-light lenses and bicycle reflectors to improve their night-time visibility. Apollo astronauts placed arrays of corner reflectors on the moon. By use of laser beams reflected from these arrays, the earth–moon distance has been measured to within 0.15 m.

Test Your Understanding of Section 33.2 You are standing on the shore of a lake. You spot a tasty fish swimming some distance below the lake surface. (a) If you want to spear the fish, should you aim the spear (i) above, (ii) below, or (iii) directly at the apparent position of the fish? (b) If instead you use a high-power laser to simultaneously kill and cook the fish, should you aim the laser (i) above, (ii) below, or (iii) directly at the apparent position of the fish?

MasteringPHYSICS

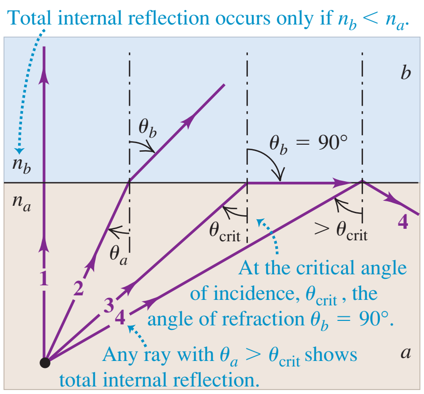
ActivPhysics 15.2: Total Internal Reflection

33.3 Total Internal Reflection

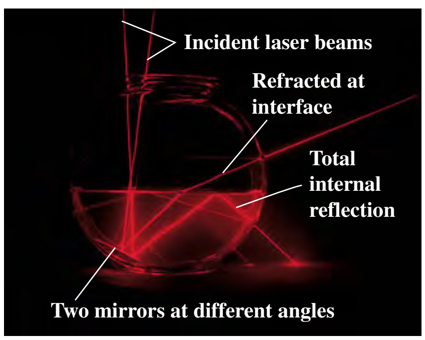
We have described how light is partially reflected and partially transmitted at an interface between two materials with different indexes of refraction. Under certain circumstances, however, *all* of the light can be reflected back from the interface, with none of it being transmitted, even though the 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

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

(a) Total internal reflection



(b) Total internal reflection demonstrated with a laser, mirrors, and water in a fishbowl



strike the surface of a second material b with index n_b, where $n_a > n_b$. (Materials a and b could be water and air, respectively.) From Snell's law of refraction,

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

Because n_a/n_b is greater than unity, $\sin\theta_b$ is larger than $\sin\theta_a$; the ray is bent *away from* the normal. Thus there must be some value of θ_a *less than* 90° for which $\sin\theta_b = 1$ and $\theta_b = 90°$. This is shown by ray 3 in the diagram, which emerges just grazing the surface at an angle of refraction of 90°. Compare Fig. 33.13a to the photograph of light rays in Fig. 33.13b.

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, the sine of the angle of refraction, as computed by Snell's law, 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

$$\sin\theta_{\text{crit}} = \frac{n_b}{n_a} \qquad \text{(critical angle for total internal reflection)} \qquad (33.6)$$

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

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

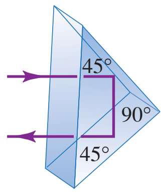
The light will be *totally reflected* if it strikes the glass–air surface at an angle of 41.1° or larger. Because the critical angle is slightly smaller than 45°, it is possible to use a prism with angles of 45°–45°–90° as a totally reflecting surface. As reflectors, totally reflecting prisms have some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

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 the rod even if the rod is curved, provided that the curvature is not too great. Such a

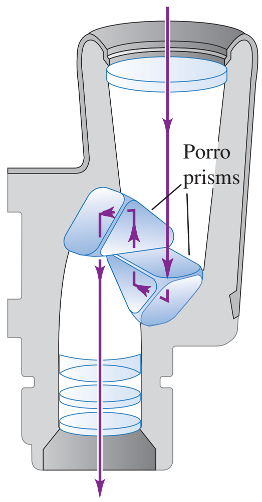
33.14 (a) Total internal reflection in a Porro prism. (b) A combination of two Porro prisms in binoculars.

(a) Total internal reflection in a Porro prism

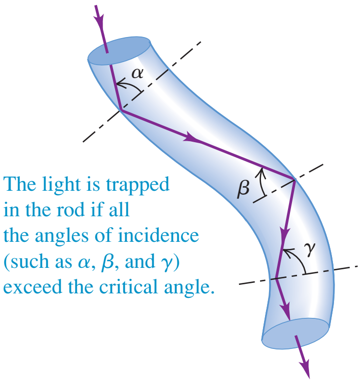


If the incident beam is oriented as shown, total internal reflection occurs on the 45° faces (because, for a glass–air interface, $\theta_{\text{crit}} = 41.1$).

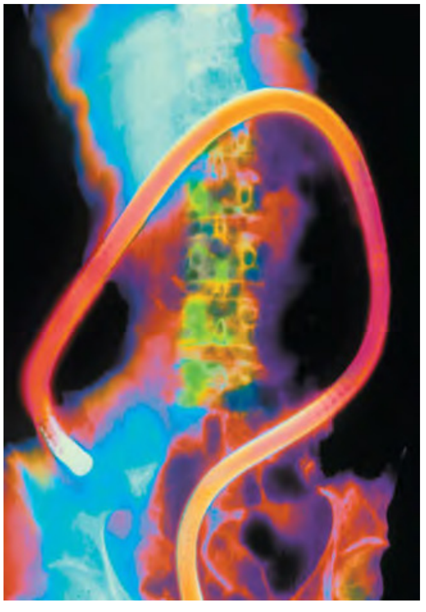
(b) Binoculars use Porro prisms to reflect the light to each eyepiece.



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



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



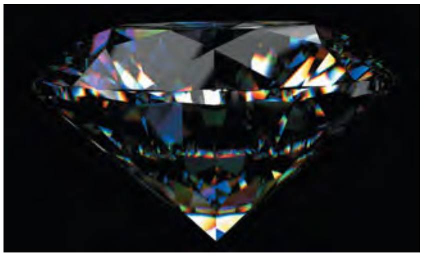
rod is sometimes called a *light pipe.* A bundle of fine glass or plastic fibers behaves in the same way and has the advantage of being flexible. A bundle may consist of thousands of individual fibers, each of the order of 0.002 to 0.01 mm in diameter. If the fibers are assembled in the bundle so that 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, in which they are used to transmit a modulated laser beam. 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.

33.17 To maximize their brilliance, diamonds are cut so that there is total internal reflection on their back surfaces.



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 (Fig. 33.17). "Imitation diamond" gems, such as cubic zirconia, are made from less expensive crystalline materials with comparable indexes of refraction.

Conceptual Example 33.4 A leaky periscope

A submarine periscope uses two totally reflecting $45°-45°-90°$ prisms with total internal reflection on the sides adjacent to the $45°$ angles. Explain why the periscope will no longer work if it springs a leak and the bottom prism is covered with water.

SOLUTION

The critical angle for water $(n_b = 1.33)$ on glass $(n_a = 1.52)$ is

$$\theta_{\text{crit}} = \arcsin\frac{1.33}{1.52} = 61.0°$$

The $45°$ angle of incidence for a totally reflecting prism is *smaller* than this new $61°$ critical angle, so total internal reflection does not occur at the glass–water interface. Most of the light is transmitted into the water, and very little is reflected back into the prism.

Test Your Understanding of Section 33.3 In which of the following situations is there total internal reflection? (i) Light propagating in water $(n = 1.33)$ strikes a water–air interface at an incident angle of 70°; (ii) light propagating in glass $(n = 1.52)$ strikes a glass–water interface at an incident angle of 70°; (iii) light propagating in water strikes a water–glass interface at an incident angle of 70°.

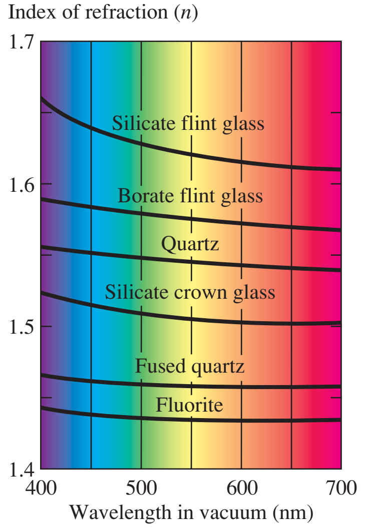
33.4 Dispersion

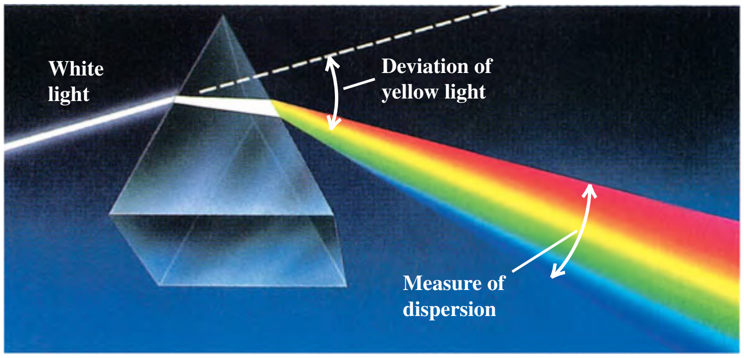
Ordinary white light is a superposition of waves with wavelengths extending throughout the visible spectrum. The speed of light *in vacuum* is the same for all wavelengths, but the speed in a material substance is different for different wavelengths. Therefore the index of refraction of a material depends on wavelength. The dependence of wave speed and index of refraction on wavelength is called **dispersion.**

Figure 33.18 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 33.19 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. Violet light is deviated most, and red is deviated least; other colors are in intermediate positions. 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. 33.18 we can see that for a substance such as 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.

33.18 Variation of index of refraction n with wavelength for different transparent materials. The horizontal axis shows the wavelength λ_0 of the light *in vacuum;* the wavelength in the material is equal to $\lambda = \lambda_0/n$.





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

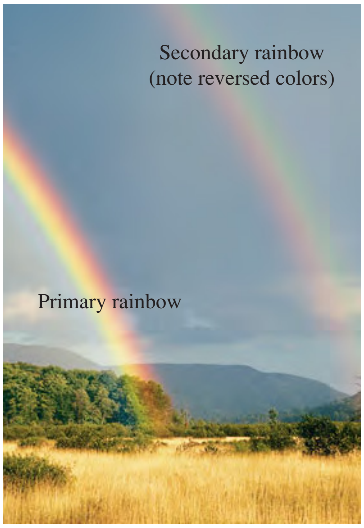
As we mentioned in Section 33.3, the brilliance of diamond is due in part to its unusually large refractive index; another important factor is its large dispersion, which causes white light entering a diamond to emerge as a multicolored spectrum. Crystals of rutile and of strontium titanate, which can be produced synthetically, have about eight times the dispersion of diamond.

Rainbows

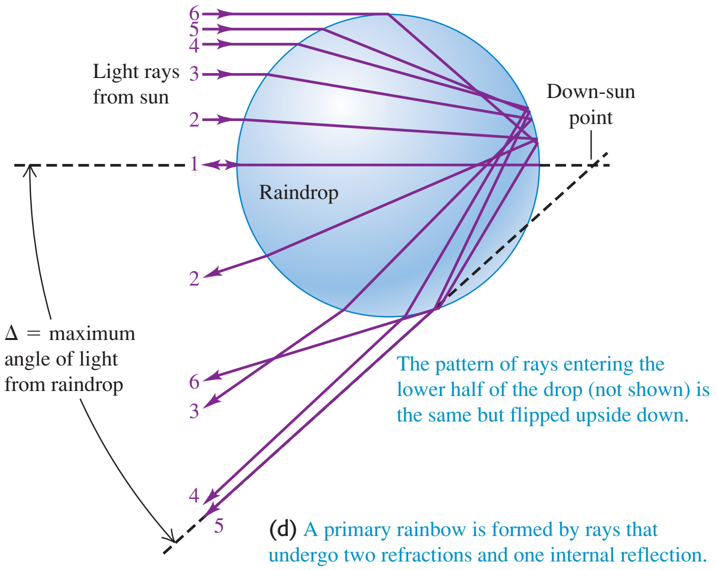
When you experience the beauty of a rainbow, as in Fig. 33.20a, you are seeing the combined effects of dispersion, refraction, and reflection. Sunlight comes from behind you, enters a water droplet, is (partially) reflected from the back surface of the droplet, and is refracted again upon exiting the droplet (Fig. 33.20b). A light ray that enters the middle of the raindrop is reflected straight back. All other rays

33.20 How rainbows form.

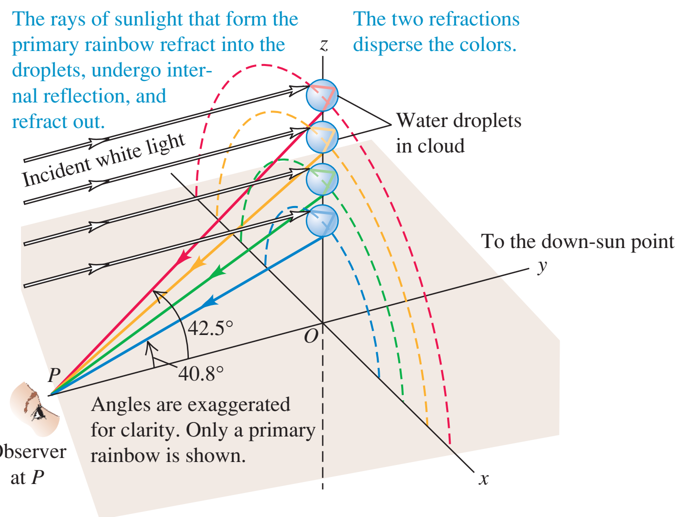
(a) A double rainbow



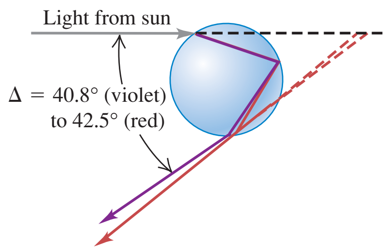
(b) The paths of light rays entering the upper half of a raindrop



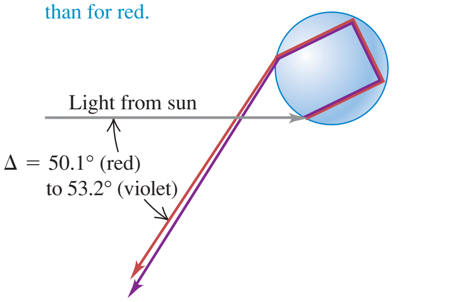
(c) Forming a rainbow. The sun in this illustration is directly behind the observer at *P*.



(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.



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. 33.20c). Because no light reaches your eye from angles larger than Δ, the sky looks dark outside the rainbow (see Fig. 33.20a). 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. 33.20d). The bright disk of red light is slightly larger than that for orange light, which in turn is slightly larger than that for yellow light, and so on. As a result, you see the rainbow as a band of colors.

In many cases you can see a second, larger rainbow. It is the result of dispersion, refraction, and *two* reflections from the back surface of the droplet (Fig. 33.20e). Each time a light ray hits the back surface, part of the light is refracted out of the drop (not shown in Fig. 33.20); 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. 33.20a.

33.5 Polarization

ActivPhysics 16.9: Physical Optics: 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. 33.21a. In this case the string always lies in the xy-plane. But the displacements might instead be along the z-axis, as in Fig. 33.21b; 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. 33.21c 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),

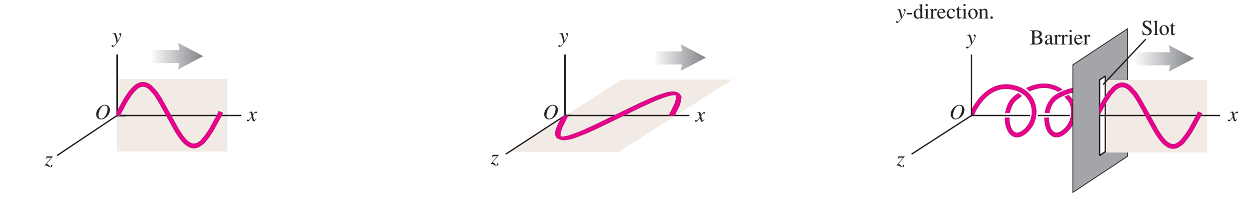
$$\vec{E}(x, t) = \hat{\jmath}E_{\max}\cos(kx - \omega t)$$
$$\vec{B}(x, t) = \hat{k}B_{\max}\cos(kx - \omega t)$$

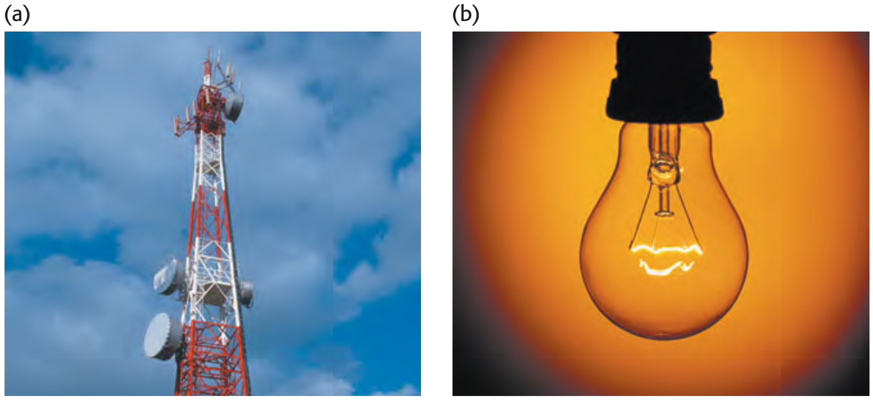
33.21 (a), (b) Polarized waves on a string. (c) Making a polarized wave on a string from an unpolarized one using a polarizing filter.

(a) Transverse wave linearly polarized in the y-direction

(b) Transverse wave linearly polarized in the z-direction

(c) The slot functions as a polarizing filter, passing only components polarized in the y-direction.





33.22 (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.

is said to be polarized in the y-direction because the electric field has only a y-component.

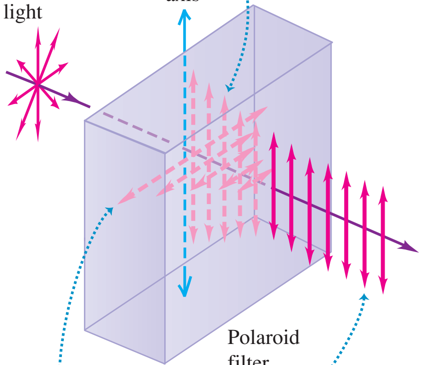
CAUTION **The meaning of "polarization"** It's unfortunate that the same word "polarization" that is used to describe the direction of $\vec{E}$ in an electromagnetic wave is also used to describe the shifting of electric charge within a body, such as in response to a nearby charged body; we described this latter kind of polarization in Section 21.2 (see Fig. 21.7). You should remember that while these two concepts have the same name, they do not describe the same phenomenon.

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. 33.22a).

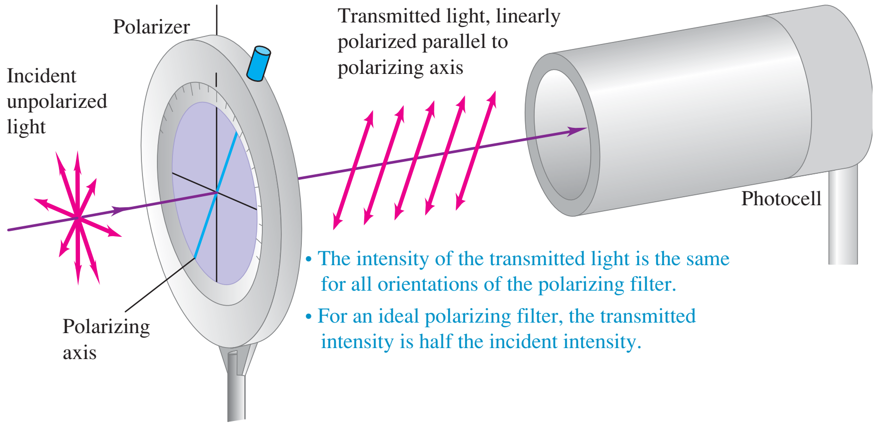
The situation is different for visible light. Light from incandescent light bulbs and fluorescent light fixtures is *not* polarized (Fig. 33.22b). The "antennas" that radiate light waves are the molecules that make up the sources. The waves emitted by any one molecule may be linearly polarized, like 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. 33.21c.

33.23 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).



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. Developed originally by the American scientist Edwin H. Land, 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. 33.23). A Polaroid filter transmits 80% or more of the intensity of a wave that is polarized parallel to a certain axis in the material, called the **polarizing**



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

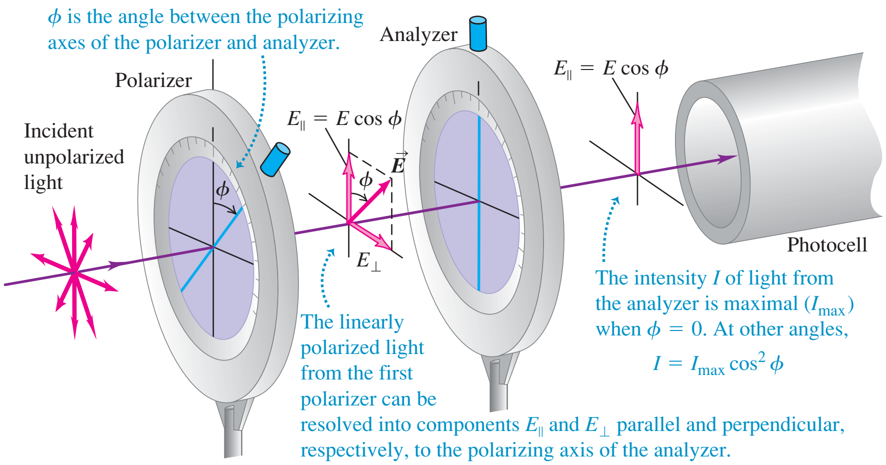
axis, but only 1% or less for waves that are polarized perpendicular to this axis. In one type of Polaroid filter, long-chain molecules within the filter are oriented with their axis perpendicular to the polarizing axis; these molecules preferentially absorb light that is polarized along their length, much like the conducting wires in a polarizing filter for microwaves.

Using Polarizing Filters

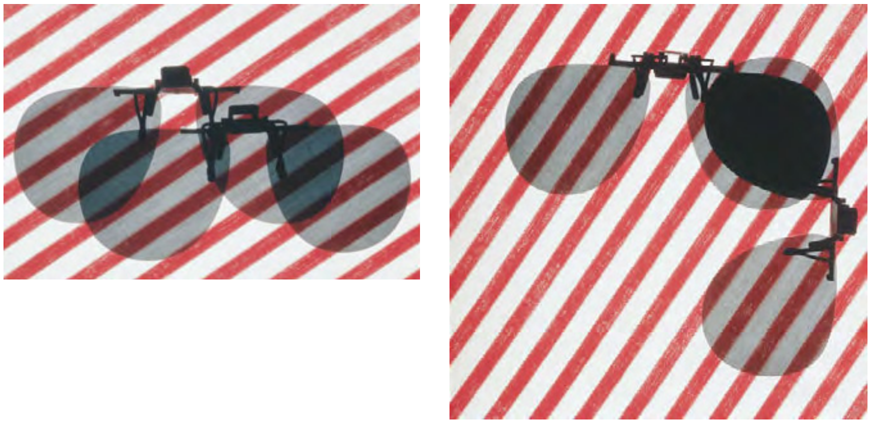
An *ideal* polarizing filter (polarizer) passes 100% of the incident light that is polarized parallel to the filter's polarizing axis but completely blocks all light that is polarized perpendicular to this axis. Such a device is an unattainable idealization, but the concept is useful in clarifying the basic ideas. In the following discussion we will assume that all polarizing filters are ideal. In Fig. 33.24 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. 33.24, 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. 33.25? Suppose the polarizing axis of the analyzer makes an angle ϕ with the polarizing axis of the



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



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

first polarizer. We can resolve the linearly polarized light that is transmitted by the first polarizer into two components, as shown in Fig. 33.25, 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. 33.26). To determine the direction of polarization of the light transmitted by the first polarizer, rotate the analyzer until the photocell in Fig. 33.25 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 our energy discussion in Section 32.4 that the intensity of an electromagnetic wave is proportional to the *square* of the amplitude of the wave [see Eq. (32.29)]. The ratio of transmitted to incident *amplitude* is $\cos\phi$, so the ratio of transmitted to incident *intensity* is $\cos^2\phi$. Thus the intensity of the light transmitted through the analyzer is

$$I = I_{max}\cos^2\phi \qquad \text{(Malus's law, polarized light passing through an analyzer)} \tag{33.7}$$

where I_{max} is the maximum intensity of light transmitted (at $\phi = 0$) and I is the amount transmitted at angle ϕ. This relationship, discovered experimentally by Etienne 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

MP

IDENTIFY *the relevant concepts:* In all electromagnetic waves, including light waves, the direction of polarization is the direction of the $\vec{E}$ field and is perpendicular to the propagation direction. Problems about polarizers are therefore about the components of $\vec{E}$ parallel and perpendicular to the polarizing axis.

SET UP *the problem* using the following steps:

1. Start by drawing a large, neat diagram. Label all known angles, including the angles of all polarizing axes.
2. Identify the target variables.

EXECUTE *the solution* as follows:

1. Remember that a polarizer lets pass only electric-field components parallel to its polarizing axis.
2. If the incident light is linearly polarized and has amplitude E and intensity I_{max}, the light that passes through an ideal polarizer has amplitude $E\cos\phi$ and intensity $I_{max}\cos^2\phi$, where ϕ is the angle between the incident polarization direction and the filter's polarizing axis.
3. Unpolarized light is a random mixture of all possible polarization states, so on the average it has equal components in any two perpendicular directions. When passed through an ideal polarizer, unpolarized light becomes linearly polarized light with half the incident intensity. Partially linearly polarized light is a superposition of linearly polarized and unpolarized light.
4. The intensity (average power per unit area) of a wave is proportional to the *square* of its amplitude. If you find that two waves differ in amplitude by a certain factor, their intensities differ by the square of that factor.

EVALUATE *your answer:* Check your answer for any obvious errors. If your results say that light emerging from a polarizer has greater intensity than the incident light, something's wrong: A polarizer can't add energy to a light wave.

Example 33.5 Two polarizers in combination

In Fig. 33.25 the incident unpolarized light has intensity I_0. Find the intensities transmitted by the first and second polarizers if the angle between the axes of the two filters is 30°.

SOLUTION

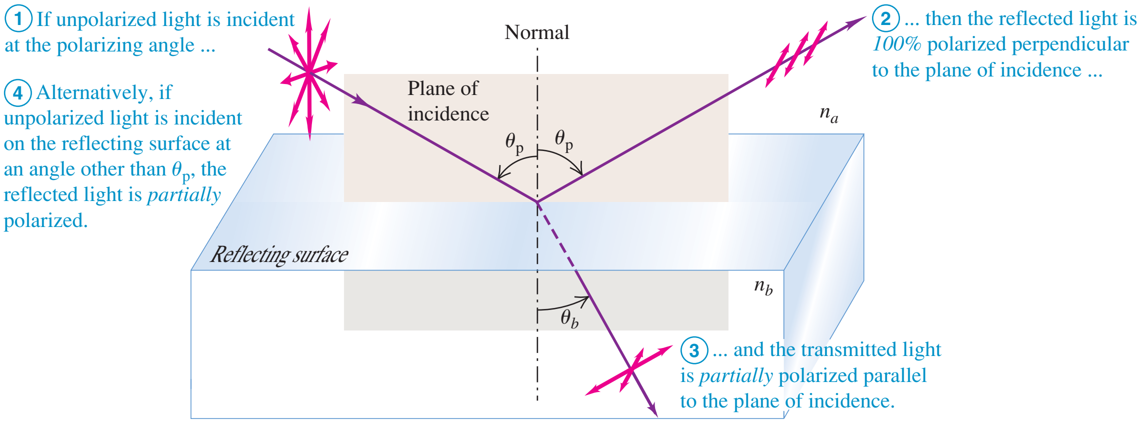
IDENTIFY and SET UP: This problem involves a polarizer (a polarizing filter on which unpolarized light shines, producing polarized light) and an analyzer (a second polarizing filter on which the polarized light shines). We are given the intensity I_0 of the incident light and the angle $\phi = 30°$ between the axes of the polarizers. We use Malus's law, Eq. (33.7), to solve for the intensities of the light emerging from each polarizer.

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

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

EVALUATE: Note that the intensity decreases after each passage through a polarizer. The only situation in which the transmitted intensity does *not* decrease is if the polarizer is ideal (so it absorbs none of the light that passes through it) and if the incident light is linearly polarized along the polarizing axis, so $\phi = 0$.

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



Polarization by Reflection

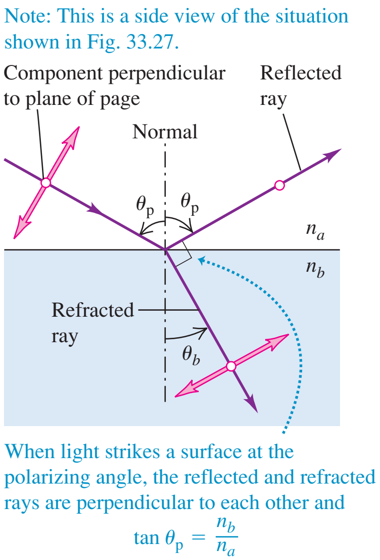
Unpolarized light can be polarized, either partially or totally, by *reflection.* In Fig. 33.27, unpolarized natural light is incident on a reflecting surface between two transparent optical materials. For most angles of incidence, waves for which the electric-field vector $\vec{E}$ is perpendicular 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. 33.27. 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. 33.28). In this case the angle of refraction θ_b equals $90° - \theta_p$. From the law of refraction,

$$n_a \sin\theta_p = n_b \sin\theta_b$$

33.28 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.



so we find

$$n_a \sin\theta_p = n_b \sin(90° - \theta_p) = n_b \cos\theta_p$$

$$\tan\theta_p = \frac{n_b}{n_a} \quad \text{(Brewster's law for the polarizing angle)} \tag{33.8}$$

This relationship is known as **Brewster's law.** Although discovered experimentally, it can also be *derived* from a wave model using Maxwell's equations.

Polarization by reflection is the reason polarizing filters are widely used in sunglasses (Fig. 33.26). 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 or the surface of a lake, it causes unwanted glare. Vision can be improved by eliminating this glare. The manufacturer makes the polarizing axis of the lens material 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.

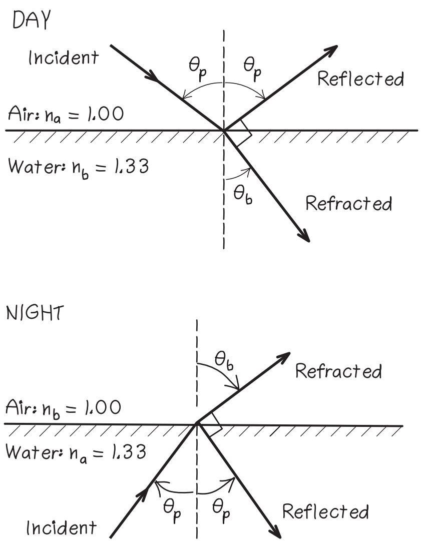
Example 33.6 Reflection from a swimming pool's surface

Sunlight reflects off the smooth surface of a swimming pool. (a) For what angle of reflection is the reflected light completely polarized? (b) What is the corresponding angle of refraction? (c) At night, an underwater floodlight is turned on in the pool. Repeat parts (a) and (b) for rays from the floodlight that strike the surface from below.

SOLUTION

IDENTIFY and SET UP: This problem involves polarization by reflection at an air–water interface in parts (a) and (b) and at a water–air interface in part (c). Figure 33.29 shows our sketches.

33.29 Our sketches for this problem.



For both cases our first target variable is the polarizing angle θ_p, which we find using Brewster's law, Eq. (33.8). For this angle of reflection, the angle of refraction θ_b is the complement of θ_p (that is, $\theta_b = 90° - \theta_p$).

EXECUTE: (a) During the day (shown in the upper part of Fig. 33.29) 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. 33.29) the light moves in water toward air, so now $n_a = 1.33$ and $n_b = 1.00$. Again using Eq. (33.8), we have

$$\theta_p = \arctan\frac{1.00}{1.33} = 36.9°$$

$$\theta_b = 90° - 36.9° = 53.1°$$

EVALUATE: We check our answer in part (b) by using Snell's law, $n_a \sin\theta_a = n_b \sin\theta_b$, to solve for θ_b:

$$\sin\theta_b = \frac{n_a \sin\theta_p}{n_b} = \frac{1.00 \sin 53.1°}{1.33} = 0.600$$

$$\theta_b = \arcsin(0.600) = 36.9°$$

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

Circular and Elliptical Polarization

Light and other electromagnetic radiation can also have *circular* or *elliptical* polarization. To introduce these concepts, let's return once more to mechanical waves on a stretched string. In Fig. 33.21, 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.

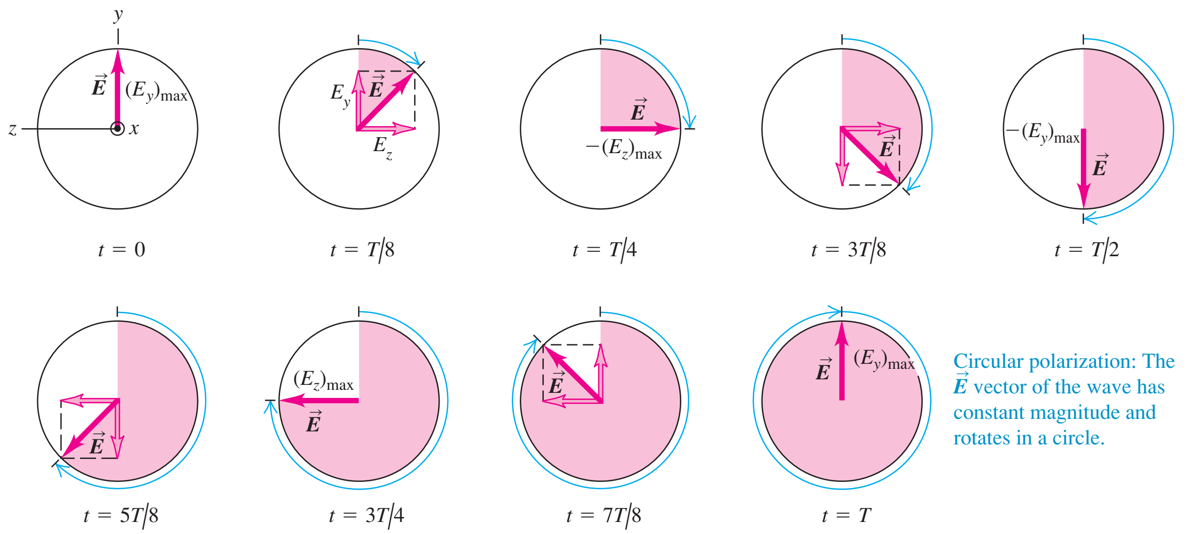
But now suppose the two equal-amplitude waves differ in phase by a quarter-cycle. Then the resultant motion of each point corresponds to a superposition of two simple harmonic motions at right angles, with a quarter-cycle phase difference. The y-displacement at a point is greatest at times when the z-displacement is zero, and vice versa. The motion of the string as a whole then no longer takes place 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. This is shown to the left of the polarizing filter in Fig. 33.21c. This particular superposition of two linearly polarized waves is called **circular polarization.**

Figure 33.30 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. 33.30 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

33.30 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.



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

33.31 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 metal artificial hip joint used in medicine.

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 33.31 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. 33.31 is that for which the transmitted light is most nearly polarized along the analyzer's polarization axis.

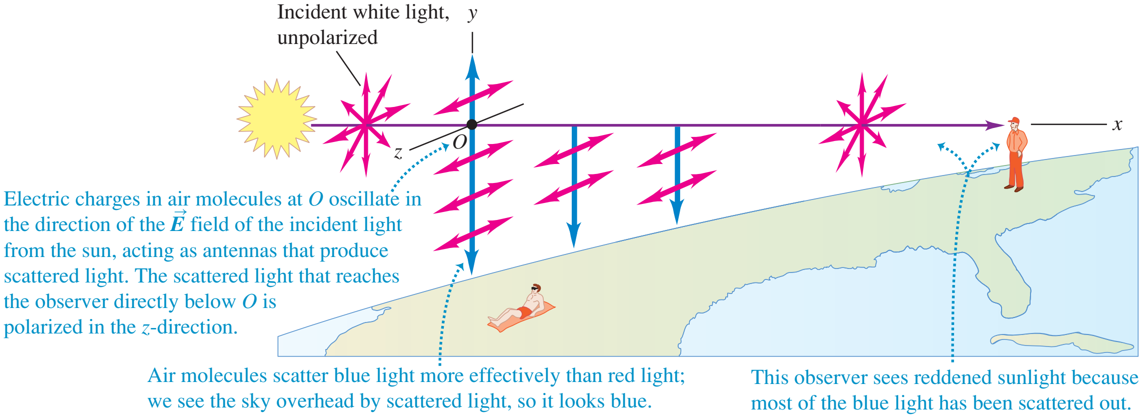
Test Your Understanding of Section 33.5 You are taking a photograph of a sunlit high-rise office building. In order to minimize the reflections from the building's windows, you place a polarizing filter on the camera lens. How should you orient the filter? (i) with the polarizing axis vertical; (ii) with the polarizing axis horizontal; (iii) either orientation will minimize the reflections just as well; (iv) neither orientation will have any effect.

33.6 Scattering of Light

The sky is blue. Sunsets are red. Skylight is partially polarized; that's why the sky looks darker from some angles than from others when it is viewed through Polaroid sunglasses. As we will see, a single phenomenon is responsible for all of these effects.

When you look at the daytime sky, the light that you see is sunlight that has been absorbed and then re-radiated in a variety of directions. This process is called **scattering.** (If the earth had no atmosphere, the sky would appear as black in the daytime as it does at night, just as it does to an astronaut in space or on the moon.) Figure 33.32 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

33.32 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.



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. 33.32 show the direction of polarization of the light reaching the observer.

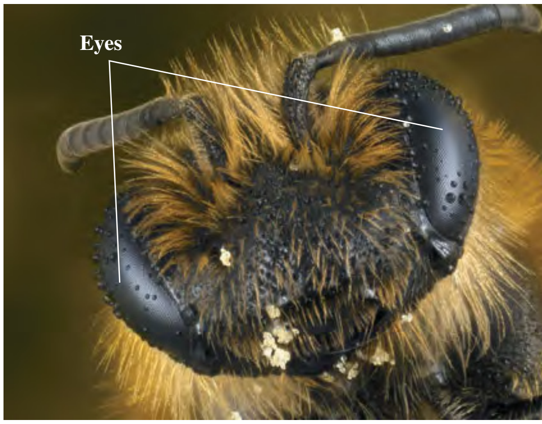
As the original beam of sunlight passes though the atmosphere, its intensity decreases as its energy goes into the scattered light. Detailed analysis of the scattering process shows that the intensity of the light scattered from air molecules increases in proportion to the fourth power of the frequency (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 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. 33.33). Milk looks white for the same reason; the scattering is due to fat globules 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. 33.32).

Application Bee Vision and Polarized Light from the Sky

The eyes of a bee can detect the polarization of light. Bees use this to help them navigate between the hive and food sources. As Fig. 33.32 shows, a bee sees unpolarized light if it looks directly toward the sun and sees completely polarized light if it looks 90° away from the sun. These polarizations are unaffected by the presence of clouds, so a bee can navigate relative to the sun even on an overcast day.



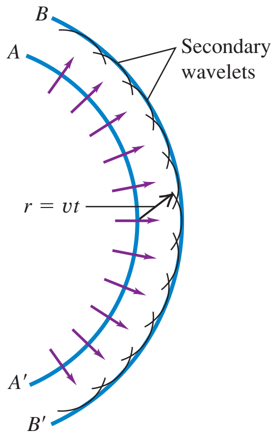
33.33 Clouds are white because they efficiently scatter sunlight of all wavelengths.

33.7 Huygens's Principle

The laws of reflection and refraction of light rays that we 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. Thus it is not an independent principle, but it is often very convenient for calculations with wave phenomena.

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



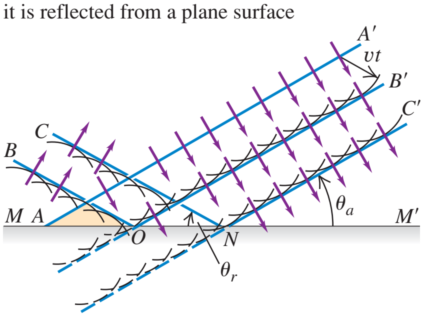
Huygens's principle is shown in Fig. 33.34. The original wave front AA' is traveling outward from a source, as indicated by the arrows. We want to find the shape of the wave front after a time interval t. We assume that v, the speed of propagation of the wave, is the same at all points. Then in time t the wave front travels a distance vt. We construct several circles (traces of spherical wavelets) with radius $r = vt$, centered at points along AA'. The trace of the envelope of these wavelets, which is the new wave front, is the curve BB'.

Reflection and Huygens's Principle

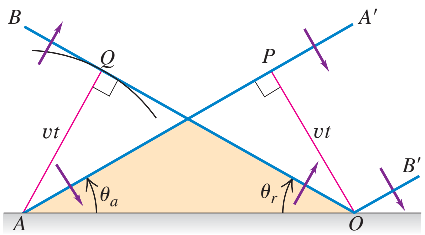
To derive the law of reflection from Huygens's principle, we consider a plane wave approaching a plane reflecting surface. In Fig. 33.35a the lines AA', OB', and NC' represent successive positions of a wave front approaching the surface MM'. Point A on the wave front AA' has just arrived at the reflecting 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.

33.35 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)



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. 33.35b. From O we draw $OP = vt$, perpendicular to AA'. Now OB, by construction, is tangent to a circle of radius vt with center at A. If we draw AQ from A to the point of tangency, the triangles APO and OQA are congruent because they are right triangles with the side AO in common and with $AQ = OP = vt$. The angle θ_a therefore equals the angle θ_r, and we have the law of reflection.

Refraction and Huygens's Principle

We can derive the law of *refraction* by a similar procedure. In Fig. 33.36a 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 in the figure; they proceed as in Fig. 33.35.) 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. 33.36b. We draw $OQ = v_a t$, perpendicular to AQ, and we draw $AB = v_b t$, perpendicular to BO. From the right triangle AOQ,

$$\sin\theta_a = \frac{v_a t}{AO}$$

and from the right triangle AOB,

$$\sin\theta_b = \frac{v_b t}{AO}$$

Combining these, we find

$$\frac{\sin\theta_a}{\sin\theta_b} = \frac{v_a}{v_b} \qquad (33.9)$$

We have defined the index of refraction n of a material as the ratio of the speed of light c in vacuum to its speed v in the material: $n_a = c/v_a$ and $n_b = c/v_b$. Thus

$$\frac{n_b}{n_a} = \frac{c/v_b}{c/v_a} = \frac{v_a}{v_b}$$

and we can rewrite Eq. (33.9) as

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

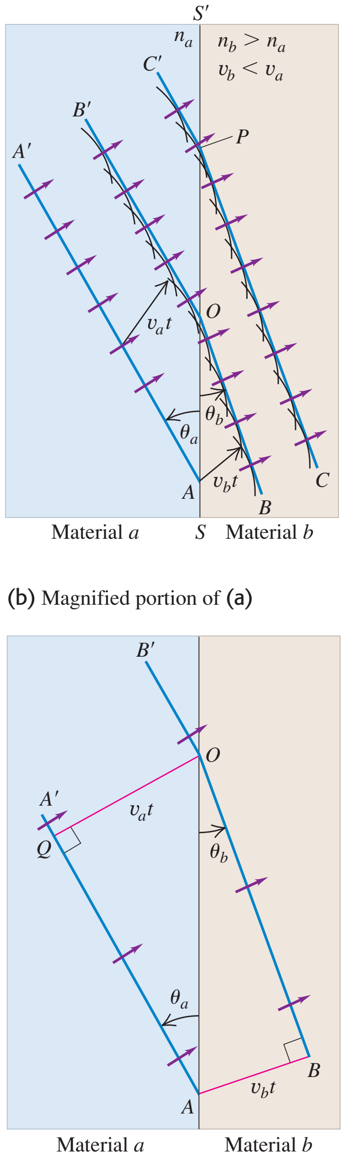
$$n_a \sin\theta_a = n_b \sin\theta_b$$

which we recognize as Snell's law, Eq. (33.4). So we have derived Snell's law from a wave theory. Alternatively, we may choose to regard Snell's law as an experimental result that defines the index of refraction of a material; in that case this analysis helps to confirm the relationship $v = c/n$ for the speed of light in a material.

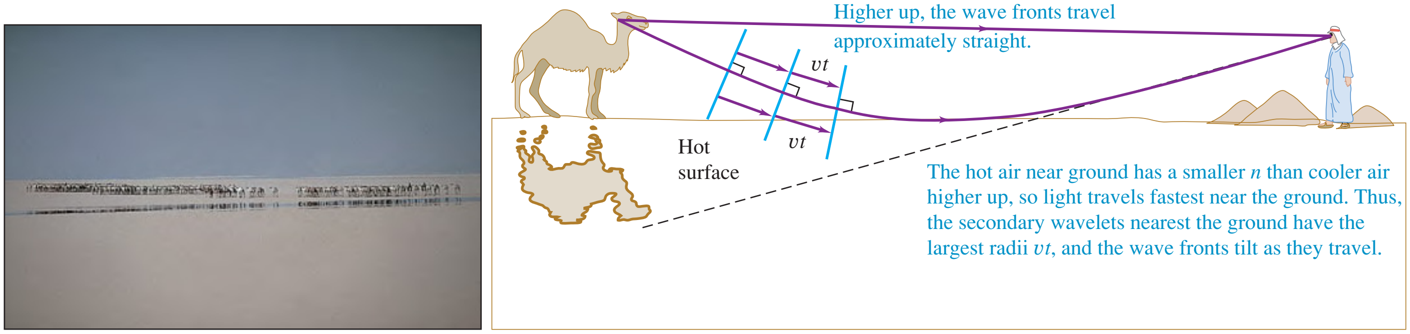
Mirages offer an interesting example of Huygens's principle in action. When the surface of pavement or desert sand is heated intensely by the sun, a hot, less dense, smaller-n layer of air forms near the surface. The speed of light is slightly greater in the hotter air near the ground, the Huygens wavelets 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. 33.37. 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. The mind of a thirsty traveler can interpret the apparent reflecting surface as a sheet of water.

33.36 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



33.37 How mirages are formed.



It is important to keep in mind that Maxwell's equations are the fundamental relationships for electromagnetic wave propagation. But Huygens's principle provides a convenient way to visualize this propagation.

Test Your Understanding of Section 33.7 Sound travels faster in warm air than in cold air. Imagine a weather front that runs north-south, with warm air to the west of the front and cold air to the east. A sound wave traveling in a northeast direction in the warm air encounters this front. How will the direction of this sound wave change when it passes into the cold air? (i) The wave direction will deflect toward the north; (ii) the wave direction will deflect toward the east; (iii) the wave direction will be unchanged. ❙

CHAPTER 33 SUMMARY

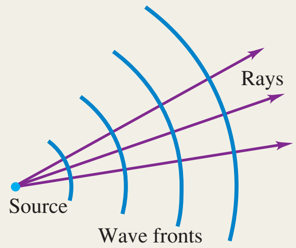
Light and its properties: Light is an electromagnetic wave. When emitted or absorbed, it also shows particle properties. It is emitted by accelerated electric charges.

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.)

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

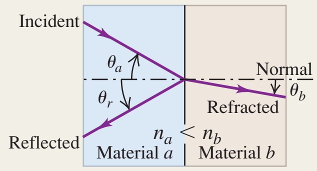
$$\lambda = \frac{\lambda_0}{n} \qquad (33.5)$$



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.)

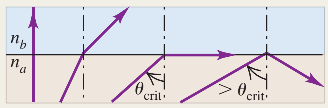
$$\theta_r = \theta_a \qquad (33.2)$$
(law of reflection)

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



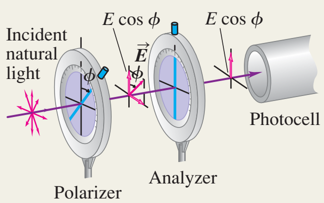
Total internal reflection: When a ray travels in a material of greater index of refraction n_a toward a material of smaller index n_b, total internal reflection occurs at the interface when the angle of incidence exceeds a critical angle θ_{crit}. (See Example 33.4.)

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



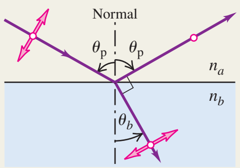
Polarization of light: The direction of polarization of a linearly polarized electromagnetic wave is the direction of the $\vec{E}$ field. A polarizing filter passes waves that are linearly polarized along its polarizing axis and blocks waves polarized perpendicularly to that axis. When polarized light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

$$I = I_{max}\cos^2\phi \qquad (33.7)$$
(Malus's law)

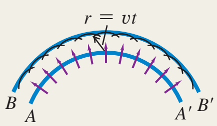


Polarization by reflection: When unpolarized light strikes an interface between two materials, Brewster's law states that the reflected light is completely polarized perpendicular to the plane of incidence (parallel to the interface) if the angle of incidence equals the polarizing angle θ_p. (See Example 33.6.)

$$\tan\theta_p = \frac{n_b}{n_a} \qquad (33.8)$$
(Brewster's law)



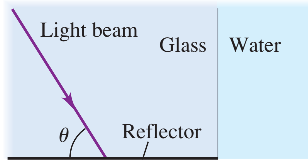
Huygens's principle: Huygens's principle states that if the position of a wave front at one instant is known, then the position of the front at a later time can be constructed by imagining the front as a source of secondary wavelets. Huygens's principle can be used to derive the laws of reflection and refraction.



BRIDGING PROBLEM Reflection and Refraction

Figure 33.38 shows a rectangular glass block that has a metal reflector on one face and water on an adjoining face. A light beam strikes the reflector as shown. You gradually increase the angle θ of the light beam. If $\theta \geq 59.2°$, no light enters the water. What is the speed of light in this glass?

33.38



SOLUTION GUIDE

See MasteringPhysics® study area for a Video Tutor solution.

IDENTIFY and SET UP

1. Specular reflection occurs where the light ray in the glass strikes the reflector. If no light is to enter the water, we require that there be reflection only and no refraction where this ray strikes the glass–water interface—that is, there must be total internal reflection.
2. The target variable is the speed of light v in the glass, which you can determine from the index of refraction n of the glass. (Table 33.1 gives the index of refraction of water.) Write down the equations you will use to find n and v.

EXECUTE

3. Use the figure to find the angle of incidence of the ray at the glass–water interface.
4. Use the result of step 3 to find n.
5. Use the result of step 4 to find v.

EVALUATE

6. How does the speed of light in the glass compare to the speed in water? Does this make sense?

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **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 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. 33.32, 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. 33.37 for light.)

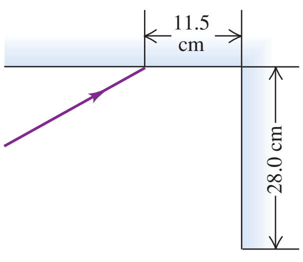
Q33.25 Can water waves be reflected and refracted? Give examples. Does Huygens's principle apply to water waves? Explain.

EXERCISES

Section 33.2 Reflection and Refraction

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

Figure **E33.1**



33.2 • **BIO** **Light Inside the Eye.** The vitreous humor, a transparent, gelatinous fluid that fills most of the eyeball, has an index of refraction of 1.34. Visible light ranges in wavelength from 380 nm (violet) to 750 nm (red), as measured in air. This light travels through the vitreous humor and strikes the rods and cones at the surface of the retina. What are the ranges of (a) the wavelength, (b) the frequency, and (c) the speed of the light just as it approaches the retina within the vitreous humor?

33.3 • A beam of light has a wavelength of 650 nm in vacuum. (a) What is the speed of this light in a liquid whose index of refraction at this wavelength is 1.47? (b) What is the wavelength of these waves in the liquid?

33.4 • Light with a frequency of 5.80×10^{14} Hz travels in a block of glass that has an index of refraction of 1.52. What is the wavelength of the light (a) in vacuum and (b) in the glass?

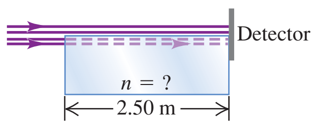
33.5 • A light beam travels at 1.94×10^8 m/s in quartz. The wavelength of the light in quartz is 355 nm. (a) What is the index of refraction of quartz at this wavelength? (b) If this same light travels through air, what is its wavelength there?

33.6 •• Light of a certain frequency has a wavelength of 438 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**

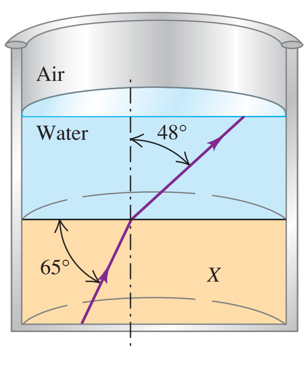


33.9 • Light traveling in air is incident on the surface of a block of plastic at an angle of 62.7° to the normal and is bent so that it makes a 48.1° angle with the normal in the plastic. Find the speed of light in the plastic.

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 •• In a material having an index of refraction n, a light ray has frequency f, wavelength λ, and speed v. What are the frequency, wavelength, and speed of this light (a) in vacuum and

(b) in a material having refractive index n'? In each case, express your answers in terms of *only* f, λ, v, n, and n'.

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 42.0° with respect to the normal to the interface, what angle does the refracted ray in the glass make with respect to the normal?

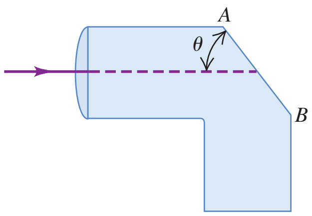
33.15 • A ray of light is incident on a plane surface separating two sheets of glass with refractive indexes 1.70 and 1.58. The angle of incidence is 62.0°, and the ray originates in the glass with $n = 1.70$. Compute the angle of refraction.

Section 33.3 Total Internal Reflection

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 • **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. E33.17). You want to cut the face AB so that all the light will reflect back into the pipe after it first strikes that face. (a) What is the largest that θ can be if the pipe is in air? (b) If the pipe is immersed in water of refractive index 1.33, what is the largest that θ can be?

Figure **E33.17**



33.18 • A beam of light is traveling inside a solid glass cube having index of refraction 1.53. 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?

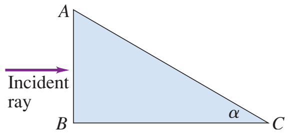
33.19 •• The critical angle for total internal reflection at a liquid–air interface is 42.5°. (a) If a ray of light traveling in the liquid has an angle of incidence at the interface of 35.0°, what angle does the refracted ray in the air make with the normal? (b) If a ray of light traveling in air has an angle of incidence at the interface of 35.0°, what angle does the refracted ray in the liquid make with the normal?

33.20 • 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 • 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?

33.22 • Light is incident along the normal on face AB of a glass prism of refractive index 1.52, as shown in Fig. E33.22. 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 **E33.22**



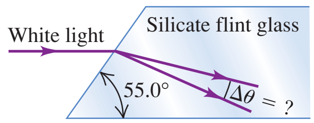
33.23 • A piece of glass with a flat surface is at the bottom of a tank of water. If a ray of light traveling in the glass is incident on the interface with the water at an angle with respect to the normal that is greater than 62.0°, no light is refracted into the water. For smaller angles of incidence, part of the ray is refracted into the water. If the light has wavelength 408 nm in the glass, what is the wavelength of the light in the water?

33.24 •• 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. The speed of a sound wave is 344 m/s in air and 1320 m/s in water. (a) Which medium has the higher index of refraction for sound? (b) What is the critical angle for a sound wave incident on the surface between air and water? (c) For total internal reflection to occur, must the sound wave be traveling in the air or in the water? (d) Use your results to explain why it is possible to hear people on the opposite shore of a river or small lake extremely clearly.

Section 33.4 Dispersion

33.25 •• 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. E33.25. 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. 33.18.)

Figure **E33.25**



33.26 • 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?

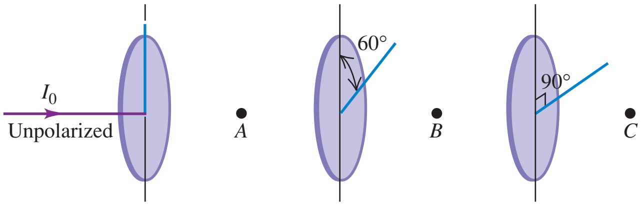
Section 33.5 Polarization

33.27 • 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) At what angle above the horizontal is the sun if sunlight reflected from the surface of a calm lake is completely polarized? (b) What is the plane of the electric-field vector in the reflected light?

33.29 •• A beam of unpolarized light of intensity I_0 passes through a series of ideal polarizing filters with their polarizing directions 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 traveling in water strikes a glass plate at an angle of incidence of 53.0°; part of the beam is reflected and part is refracted. If the reflected and refracted portions make an angle of 90.0° with each other, what is the index of refraction of the glass?

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 •• Light of original intensity I_0 passes through two ideal polarizing filters having their polarizing axes oriented as shown in Fig. E33.32. 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 **E33.32**

33.33 •• A beam of polarized light passes through a polarizing filter. When the angle between the polarizing axis of the filter and the direction of polarization of the light is θ, the intensity of the emerging beam is I. If you now want the intensity to be $I/2$, what should be the angle (in terms of θ) between the polarizing angle of the filter and the original direction of polarization of the light?

33.34 • 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?

33.35 •• 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?

33.36 ••• Three polarizing filters are stacked, with the polarizing axis of the second and third filters at 23.0° and 62.0°, respectively, to that of the first. If unpolarized light is incident on the stack, the light has intensity 75.0 W/cm^2 after it passes through the stack. If the incident intensity is kept constant, what is the intensity of the light after it has passed through the stack if the second polarizer is removed?

33.37 • **Three Polarizing Filters.** Three polarizing filters are stacked with the polarizing axes of the second and third at 45.0° and 90.0°, respectively, with that of the first. (a) If unpolarized light of intensity I_0 is incident on the stack, find the intensity and state of polarization of light emerging from each filter. (b) If the second filter is removed, what is the intensity of the light emerging from each remaining filter?

Section 33.6 Scattering of Light

33.38 • 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

33.39 •• **The Corner Reflector.** An inside corner of a cube is lined with mirrors to make a corner reflector (see Example 33.3 in Section 33.2). A ray of light is reflected successively from each of three mutually perpendicular mirrors; show that its final direction is always exactly opposite to its initial direction.

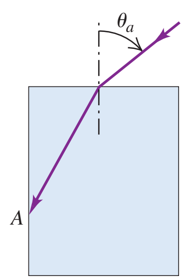
33.40 • A light beam is directed parallel to the axis of a hollow cylindrical tube. When the tube contains only air, it takes the light 8.72 ns to travel the length of the tube, but when the tube is filled with a transparent jelly, it takes the light 2.04 ns longer to travel its length. What is the refractive index of this jelly?

33.41 •• 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?

33.42 ••• 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?

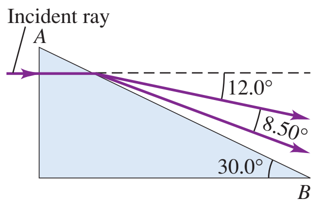
33.43 •• A ray of light is incident in air on a block of a transparent solid whose index of refraction is n. If $n = 1.38$, what is the *largest* angle of incidence θ_a for which total internal reflection will occur at the vertical face (point A shown in Fig. P33.43)?

Figure **P33.43**



33.44 • A light ray in air strikes the right-angle prism shown in Fig. P33.44. 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 **P33.44**



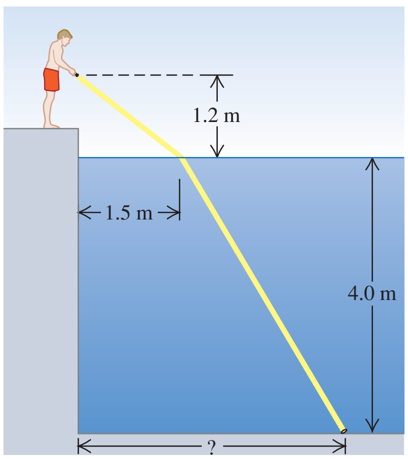
33.45 •• A ray of light traveling *in* a block of glass ($n = 1.52$) is incident on the top surface at an angle of 57.2° with respect to the normal in the glass. If a layer of oil is placed on the top surface of the glass, the ray is totally reflected. What is the maximum possible index of refraction of the oil?

33.46 ••• 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?

33.47 •• Old photographic plates were made of glass with a light-sensitive emulsion on the front surface. This emulsion was somewhat transparent. When a bright point source is focused on the front of the plate, the developed photograph will show a halo around the image of the spot. If the glass plate is 3.10 mm thick and the halos have an inner radius of 5.34 mm, what is the index of refraction of the glass? (*Hint:* Light from the spot on the front surface is scattered in all directions by the emulsion. Some of it is then totally reflected at the back surface of the plate and returns to the front surface.)

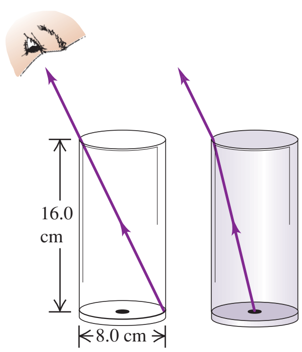
33.48 • After a long day of driving you take a late-night swim in a motel swimming pool. When you go to your room, you realize that you have lost your room key in the pool. You borrow a powerful flashlight and walk around the pool, shining the light into it. The light shines on the key, which is lying 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. P33.48). If the water here is 4.0 m deep, how far is the key from the edge of the pool?

Figure **P33.48**



33.49 • 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. P33.49a). 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. P33.49b). What is the index of refraction of the liquid?

Figure **P33.49**

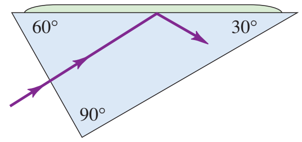


33.50 •• A 45°–45°–90° prism is immersed in water. A ray of light is incident normally on one of its shorter faces. What is the minimum index of refraction that the prism must have if this ray is to be totally reflected within the glass at the long face of the prism?

33.51 • 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?

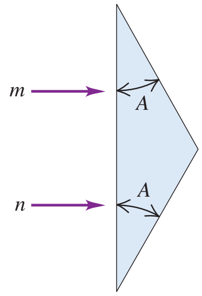
33.52 •• Light is incident normally on the short face of a 30°–60°–90° prism (Fig. P33.52). A drop of liquid is placed on the hypotenuse of the prism. If the index of refraction of the prism is 1.62, find the maximum index that the liquid may have if the light is to be totally reflected.

Figure **P33.52**



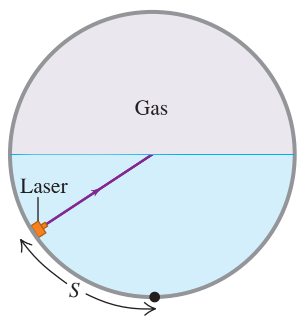
33.53 •• The prism shown in Fig. P33.53 has a refractive index of 1.66, and the angles A are 25.0°. Two light rays m and n are parallel as they enter the prism. What is the angle between them after they emerge?

Figure **P33.53**



33.54 •• A horizontal cylindrical tank 2.20 m in diameter is half full of water. The space above the water is filled with a pressurized gas of unknown refractive index. A small laser can move along the curved bottom of the water and aims a light beam toward the center of the water surface (Fig. P33.54). You observe that when the laser has moved a distance $S = 1.09$ m or more (measured along the curved surface) from the lowest point in the water, no light enters the gas. (a) What is the index of refraction of the gas? (b) What minimum time does it take the light beam to travel from the laser to the rim of the tank when (i) $S > 1.09$ m and (ii) $S < 1.09$ m?

Figure **P33.54**



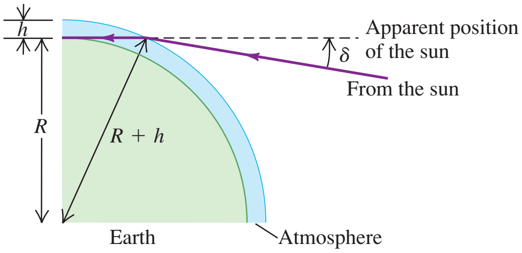
33.55 •• When the sun is either rising or setting and appears to be just on the horizon, it is in fact *below* the horizon. The explanation for this seeming paradox is that light from the sun bends slightly when entering the earth's atmosphere, as shown in Fig. P33.55. Since our perception is based on 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 **P33.55**

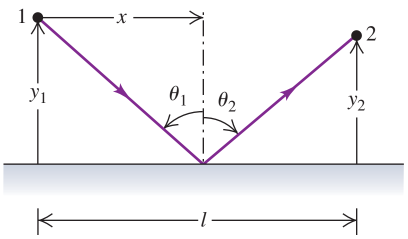


33.56 •• CALC Fermat's Principle of Least Time. A ray of light traveling with speed c leaves point 1 shown in Fig. P33.56 and is reflected to point 2. The ray strikes the reflecting surface a horizontal distance x from point 1. (a) Show that the time t required for the light to travel from 1 to 2 is

$$t = \frac{\sqrt{y_1^2 + x^2} + \sqrt{y_2^2 + (l - x)^2}}{c}$$

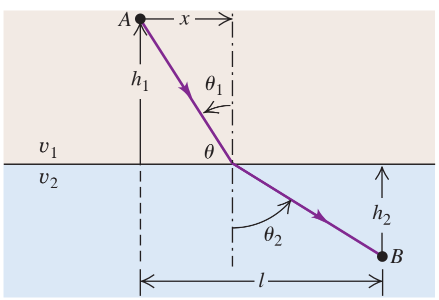
(b) Take the derivative of t with respect to x. Set the derivative equal to zero to show that this time reaches its *minimum* value when $\theta_1 = \theta_2$, which is the law of reflection and corresponds to the actual path taken by the light. This is an example of Fermat's *principle of least time,* which states that among all possible paths between two points, the one actually taken by a ray of light is that for which the time of travel is a *minimum.* (In fact, there are some cases in which the time is a maximum rather than a minimum.)

Figure **P33.56**



33.57 •• CALC A ray of light goes from point A in a medium in which the speed of light is v_1 to point B in a medium in which the speed is v_2 (Fig. P33.57). The ray strikes the interface a horizontal distance x to the right of point A. (a) Show that the time required for the light to go from A to B is

Figure **P33.57**



$$t = \frac{\sqrt{h_1^2 + x^2}}{v_1} + \frac{\sqrt{h_2^2 + (l - x)^2}}{v_2}$$

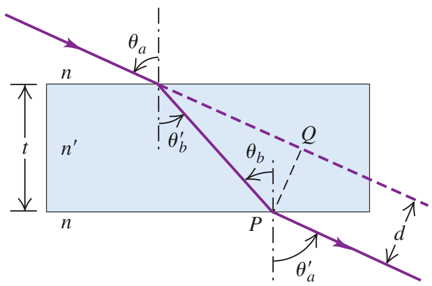
(b) Take the derivative of t with respect to x. Set this derivative equal to zero to show that this time reaches its *minimum* value when $n_1 \sin\theta_1 = n_2 \sin\theta_2$. This is Snell's law and corresponds to the actual path taken by the light. This is another example of Fermat's principle of least time (see Problem 33.56).

33.58 •• Light is incident in air at an angle θ_a (Fig. P33.58) 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

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

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 **P33.58**

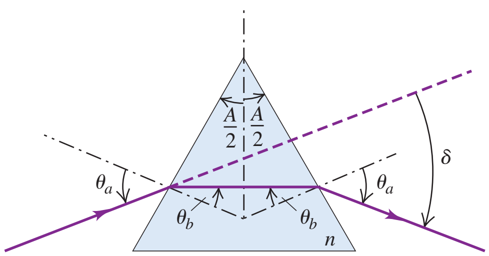


33.59 •• Angle of Deviation. The incident angle θ_a shown in Fig. P33.59 is chosen so that the light passes symmetrically through the prism, which has refractive index n and apex angle A. (a) Show that the angle of deviation δ (the angle between the initial and final directions of the ray) is given by

$$\sin\frac{A + \delta}{2} = n\sin\frac{A}{2}$$

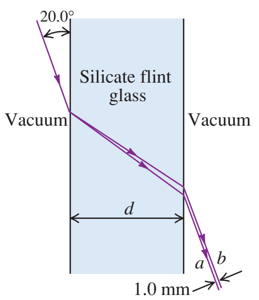
(When the light passes through symmetrically, as shown, the angle of deviation is a minimum.) (b) Use the result of part (a) to find the angle of deviation for a ray of light passing symmetrically through a prism having three equal angles ($A = 60.0°$) and $n = 1.52$. (c) A certain glass has a refractive index of 1.61 for red light (700 nm) and 1.66 for violet light (400 nm). If both colors pass through symmetrically, as described in part (a), and if $A = 60.0°$, find the difference between the angles of deviation for the two colors.

Figure **P33.59**



33.60 • A thin beam of white light is directed at a flat sheet of silicate flint glass at an angle of 20.0° to the surface of the sheet. Due to dispersion in the glass, the beam is spread out in a spectrum as shown in Fig. P33.60. The refractive index of silicate flint glass versus wavelength is graphed in Fig. 33.18. (a) The rays a and b shown in Fig. P33.60 correspond to the extreme wavelengths shown in Fig. 33.18. Which corresponds to red and which to violet? Explain your reasoning. (b) For what thickness d of the glass sheet will the spectrum be 1.0 mm wide, as shown (see Problem 33.58)?

Figure **P33.60**



33.61 • 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 of θ with respect to the vertical. The data in the table give the intensity measured through a polarizer with an orientation of ϕ with respect to the vertical. (a) What is the orientation of the polarized component? (That is, what is the angle θ?) (b) What are the values of I_0 and I_p?

ϕ (°)	I_{total} (W/m²)	ϕ (°)	I_{total} (W/m²)
0	18.4	100	8.6
10	21.4	110	6.3
20	23.7	120	5.2
30	24.8	130	5.2
40	24.8	140	6.3
50	23.7	150	8.6
60	21.4	160	11.6
70	18.4	170	15.0
80	15.0	180	18.4
90	11.6		

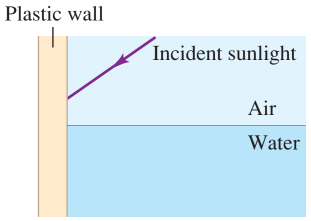
33.62 •• BIO **Optical Activity of Biological Molecules.** Many biologically important molecules are optically active. When linearly polarized light traverses a solution of compounds containing these molecules, its plane of polarization is rotated. Some compounds rotate the polarization clockwise; others rotate the polarization counterclockwise. The amount of rotation depends on the amount of material in the path of the light. The following data give the amount of rotation through two amino acids over a path length of 100 cm:

Rotation (°)		
***l*-leucine**	***d*-glutamic acid**	**Concentration (g/100 mL)**
−0.11	0.124	1.0
−0.22	0.248	2.0
−0.55	0.620	5.0
−1.10	1.24	10.0
−2.20	2.48	20.0
−5.50	6.20	50.0
−11.0	12.4	100.0

From these data, find the relationship between the concentration C (in grams per 100 mL) and the rotation of the polarization (in degrees) of each amino acid. (*Hint:* Graph the concentration as a function of the rotation angle for each amino acid.)

33.63 •• 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. P33.63). 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 **P33.63**



33.64 •• A certain birefringent material has indexes of refraction n_1 and n_2 for the two perpendicular components of linearly polarized light passing through it. The corresponding wavelengths are $\lambda_1 = \lambda_0/n_1$ and λ_0/n_2, where λ_0 is the wavelength in vacuum. (a) If the crystal is to function as a quarter-wave plate, the number of wavelengths of each component within the material must differ by $\frac{1}{4}$. Show that the minimum thickness for a quarter-wave plate is

$$d = \frac{\lambda_0}{4(n_1 - n_2)}$$

(b) Find the minimum thickness of a quarter-wave plate made of siderite $(FeO \cdot CO_2)$ if the indexes of refraction are $n_1 = 1.875$ and $n_2 = 1.635$ and the wavelength in vacuum is $\lambda_0 = 589$ nm.

CHALLENGE PROBLEMS

33.65 ••• Consider two vibrations of equal amplitude and frequency but differing in phase, one along the x-axis,

$$x = a\sin(\omega t - \alpha)$$

and the other along the y-axis,

$$y = a\sin(\omega t - \beta)$$

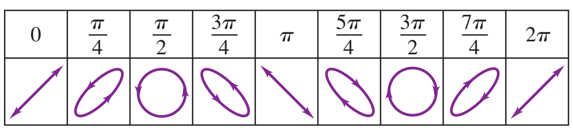
These can be written as follows:

$$\frac{x}{a} = \sin\omega t\cos\alpha - \cos\omega t\sin\alpha \qquad (1)$$

$$\frac{y}{a} = \sin\omega t\cos\beta - \cos\omega t\sin\beta \qquad (2)$$

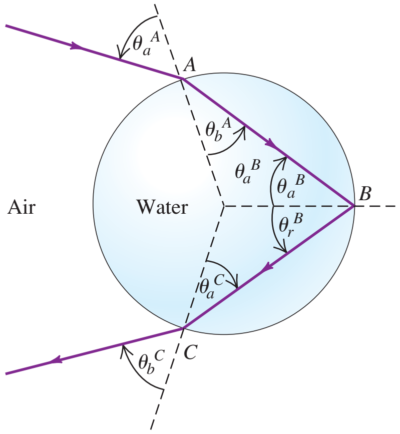
(a) Multiply Eq. (1) by $\sin\beta$ and Eq. (2) by $\sin\alpha$, and then subtract the resulting equations. (b) Multiply Eq. (1) by $\cos\beta$ and Eq. (2) by $\cos\alpha$, and then subtract the resulting equations. (c) Square and add the results of parts (a) and (b). (d) Derive the equation $x^2 + y^2 - 2xy\cos\delta = a^2\sin^2\delta$, where $\delta = \alpha - \beta$. (e) Use the above result to justify each of the diagrams in Fig. P33.65. In the figure, the angle given is the phase difference between two simple harmonic motions of the same frequency and amplitude, one horizontal (along the x-axis) and the other vertical (along the y-axis). The figure thus shows the resultant motion from the superposition of the two perpendicular harmonic motions.

Figure **P33.65**



33.66 ••• CALC A rainbow is produced by the reflection of sunlight by spherical drops of water in the air. Figure P33.66 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 $\theta_a{}^A$ 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 $\theta_a{}^A$—that is, when $d\Delta/d\theta_a{}^A = 0$. If this condition is satisfied, all the rays with incident angles close to $\theta_a{}^A$ will be sent back in the same direction, producing a bright zone in the sky. Let θ_1 be the value of $\theta_a{}^A$ 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. 33.20d? When you view the rainbow, which color, red or violet, is higher above the horizon?

Figure **P33.66**



33.67 ••• CALC A *secondary rainbow* is formed when the incident light undergoes two internal reflections in a spherical drop of water as shown in Fig. 33.20e. (See Challenge Problem 33.66.) (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 33.66. 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. 33.20e? When you view a secondary rainbow, is red or violet higher above the horizon? Explain.

Answers

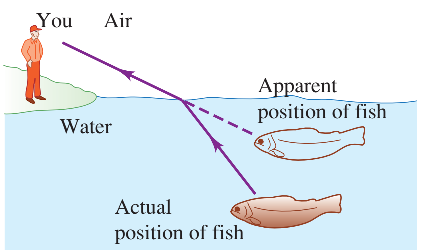
Chapter Opening Question ?

This is the same effect as shown in Fig. 33.31. The drafting tools are placed between two polarizing filters whose polarizing axes are perpendicular. In places where the clear plastic is under stress, the plastic becomes birefringent; that is, light travels through it at a speed that depends on its polarization. The result is that the light that emerges from the plastic has a different polarization than the light that enters. A spot on the plastic appears bright if the emerging light has the same polarization as the second polarizing filter. The amount of birefringence depends on the wavelength of the light as well as the amount of stress on the plastic, so different colors are seen at different locations on the plastic.

Test Your Understanding Questions

33.1 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 Answers: (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 Answers: (i), (ii) Total internal reflection can occur only if two conditions are met: n_b must be less than n_a, and the critical angle θ_{crit} (where $\sin\theta_{crit} = n_b/n_a$) must be smaller than the angle of incidence θ_a. In the first two cases both conditions are met: The critical angles are (i) $\theta_{crit} = \sin^{-1}(1/1.33) = 48.8°$ and (ii) $\theta_{crit} = \sin^{-1}(1.33/1.52) = 61.0°$, both of which are smaller than $\theta_a = 70°$. In the third case $n_b = 1.52$ is greater than $n_a = 1.33$, so total internal reflection cannot occur for any incident angle.

33.5 Answer: (ii) The sunlight reflected from the windows of the high-rise building is partially polarized in the vertical direction, since each window lies in a vertical plane. The Polaroid filter in front of the lens is oriented with its polarizing axis perpendicular to the dominant direction of polarization of the reflected light.

33.7 Answer: (ii) Huygens's principle applies to waves of all kinds, including sound waves. Hence this situation is exactly like that shown in Fig. 33.36, 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. 33.36 shows that the rays (which indicate the direction of propagation) deflect toward the east.

Bridging Problem

Answer: 1.93×10^8 m/s

34 GEOMETRIC OPTICS

? How do magnifying lenses work? At what distance from the object being examined do they provide the sharpest view?

LEARNING GOALS

By studying this chapter, you will learn:

- How a plane mirror forms an image.
- Why concave and convex mirrors form different kinds of image.
- How images can be formed by a curved interface between two transparent materials.
- What aspects of a lens determine the type of image that it produces.
- What determines the field of view of a camera lens.
- What causes various defects in human vision, and how they can be corrected.
- The principle of the simple magnifier.
- How microscopes and telescopes work.

Your reflection in the bathroom mirror, the view of the moon through a telescope, the patterns seen in a kaleidoscope—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 objects seen in a kaleidoscope seem to be in many places at the same time. 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, and some simple geometry and trigonometry. The key role played by geometry in our analysis is the reason for the name *geometric optics* that is given 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.

MasteringPHYSICS®

ActivPhysics 15.4: Geometric Optics: Plane Mirrors

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*. For an observer to see this object directly, there must be no obstruction between the object and the observer's eyes. Note that light rays from the object reach the observer's left and right eyes at different angles; these differences are processed by the observer's brain to infer the *distance* from the observer to the object.

The object in Fig. 34.1 is a **point object** that has no physical extent. Real objects with length, width, and height are called **extended objects.** To start with, we'll consider only an idealized point object, since we can always think of an extended object as being made up of a very large number of point objects.

Suppose some of the rays from the object strike a smooth, plane reflecting surface (Fig. 34.2). This could be the surface of a material with a different index of refraction, which reflects part of the incident light, or a polished metal surface that reflects almost 100% of the light that strikes it. We will always draw the reflecting surface as a black line with a shaded area behind it, as in Fig. 34.2. Bathroom mirrors have a thin sheet of glass that lies in front of and protects the reflecting surface; we'll ignore the effects of this thin sheet.

According to the law of reflection, all rays striking the surface are reflected at an angle from the normal equal to the angle of incidence. Since the surface is plane, the normal is in the same direction at all points on the surface, and we have *specular* reflection. After the rays are reflected, their directions are the same as though they had come from point P'. We call point P an *object point* and point P' the corresponding *image point,* and we say that the reflecting surface forms an **image** of point P. An observer who can see only the rays reflected from the surface, and who doesn't know that he's seeing a reflection, *thinks* that the rays originate from the image point P'. The image point is therefore a convenient way to describe the directions of the various reflected rays, just as the object point P describes the directions of the rays arriving at the surface *before* reflection.

If the surface in Fig. 34.2 were *not* smooth, the reflection would be *diffuse,* and rays reflected from different parts of the surface would go in uncorrelated directions (see Fig. 33.6b). In this case there would not be a definite image point P' from which all reflected rays seem to emanate. You can't see your reflection in the surface of 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.

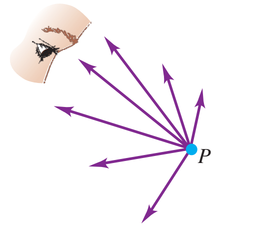
An image is also formed by a plane *refracting* surface, as shown in Fig. 34.3. Rays coming from point P are refracted at the interface between two optical materials. When the angles of incidence are small, the final directions of the rays after refraction are the same as though they had come from point P', as shown, and again we call P' an *image point.* In Section 33.2 we described how this effect makes underwater objects appear closer to the surface than they really are (see Fig. 33.9).

In both Figs. 34.2 and 34.3 the rays do not actually pass through the image point P'. Indeed, if the mirror in Fig. 34.2 is opaque, there is no light at all on its right side. If the outgoing rays don't actually pass through the image point, we call the image a **virtual image.** Later we will see cases in which the outgoing rays really *do* pass through an image point, and we will call the resulting image a **real image.** The images that are formed on a projection screen, on the photographic film in a camera, and on the retina of your eye are real images.

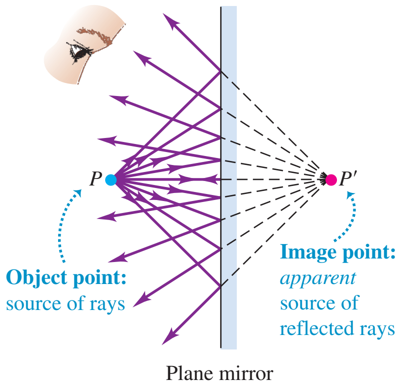
Image Formation by a Plane Mirror

Let's concentrate for now on images produced by *reflection;* we'll return to refraction later in the chapter. To find the precise location of the virtual image P' that a plane mirror forms of an object at P, we use the construction shown in Fig. 34.4. The figure shows two rays diverging from an object point P at a

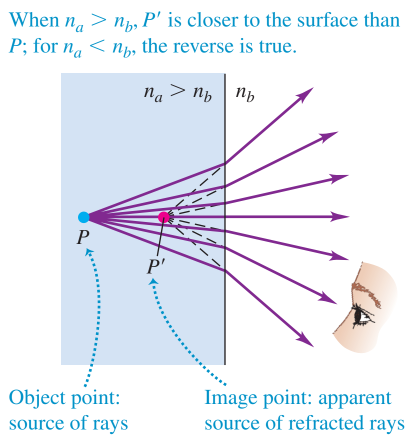
34.1 Light rays radiate from a point object P in all directions.



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



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

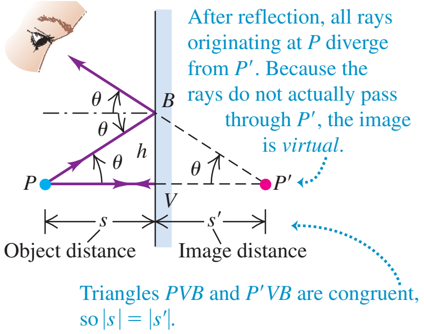


distance s to the left of a plane mirror. We call s the **object distance.** The ray PV is incident normally on the mirror (that is, it is perpendicular to the mirror surface), and it returns along its original path.

The ray PB makes an angle θ with PV. It strikes the mirror at an angle of incidence θ and is reflected at an equal angle with the normal. When we extend the two reflected rays backward, they intersect at point P', at a distance s' behind the mirror. We call s' the **image distance.** The line between P and P' is perpendicular to the mirror. The two triangles PVB and $P'VB$ are congruent, so P and P' are at equal distances from the mirror, and s and s' have equal magnitudes. The image point P' is located exactly opposite the object point P as far *behind* the mirror as the object point is from the front of the mirror.

We can repeat the construction of Fig. 34.4 for each ray diverging from P. The directions of *all* the outgoing reflected rays are the same as though they had originated at point P', confirming that P' is the *image* of P. No matter where the observer is located, she will always see the image at the point P'.

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



Sign Rules

Before we go further, let's introduce some general sign rules. These may seem unnecessarily complicated for the simple case of an image formed by a plane mirror, but we want to state the rules in a form that will be applicable to *all* the situations we will encounter later. These will include image formation by a plane or spherical reflecting or refracting surface, or by a pair of refracting surfaces forming a lens. Here are the rules:

1. **Sign rule for the object distance:** When the object is on the same side of the reflecting or refracting surface as the incoming light, the 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, the image distance s' is positive; otherwise, it is negative.
3. **Sign rule for the radius of curvature of a spherical surface:** When the center of curvature C is on the same side as the outgoing light, the radius of curvature is positive; otherwise, it is negative.

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

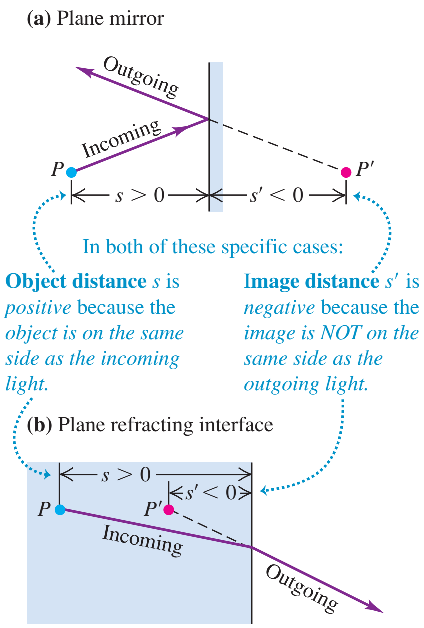


Figure 34.5 illustrates rules 1 and 2 for two different situations. For a mirror the incoming and outgoing sides are always the same; for example, in Figs. 34.2, 34.4, and 34.5a they are both on the left side. For the refracting surfaces in Figs. 34.3 and 34.5b the incoming and outgoing sides are on the left and right sides, respectively, of the interface between the two materials. (Note that other textbooks may use different rules.)

In Figs. 34.4 and 34.5a the object distance s is *positive* because the object point P is on the incoming side (the left side) of the reflecting surface. The image distance s' is *negative* because the image point P' is *not* on the outgoing side (the left side) of the surface. The object and image distances s and s' are related simply by

$$s = -s' \qquad \text{(plane mirror)} \tag{34.1}$$

For a plane reflecting or refracting surface, the radius of curvature is infinite and not a particularly interesting or useful quantity; in these cases we really don't need sign rule 3. But this rule will be of great importance when we study image formation by *curved* reflecting and refracting surfaces later in the chapter.

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

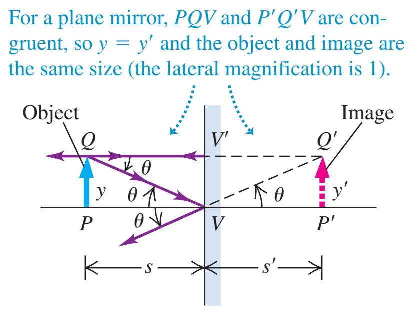


Image of an Extended Object: Plane Mirror

Next we consider an *extended* object with finite size. For simplicity we often consider an object that has only one dimension, like a slender arrow, oriented parallel to the reflecting surface; an example is the arrow PQ in Fig. 34.6. The distance from the head to the tail of an arrow oriented in this way is called its *height;* in Fig. 34.6 the height is y. The image formed by such an extended object is an extended image; to each point on the object, there corresponds a point on the image. Two of the rays from Q are shown; *all* the rays from Q appear to diverge from its image point Q' after reflection. The image of the arrow is the line $P'Q'$, with height y'. Other points of the object PQ have image points between P' and Q'. The triangles PQV and $P'Q'V$ are congruent, so the object PQ and image $P'Q'$ have the same size and orientation, and $y = y'$.

The ratio of image height to object height, y'/y, in *any* image-forming situation is called the **lateral magnification** m; that is,

$$m = \frac{y'}{y} \qquad \text{(lateral magnification)} \tag{34.2}$$

Thus for a plane mirror the lateral magnification m is unity. When you look at yourself in a plane mirror, your image is the same size as the real you.

In Fig. 34.6 the image arrow points in the *same* direction as the object arrow; we say that the image is **erect.** In this case, y and y' have the same sign, and the lateral magnification m is positive. The image formed by a plane mirror is always erect, so y and y' have both the same magnitude and the same sign; from Eq. (34.2) the lateral magnification of a plane mirror is always $m = +1$. Later we will encounter situations in which the image is **inverted;** that is, the image arrow points in the direction *opposite* to that of the object arrow. For an inverted image, y and y' have *opposite* signs, and the lateral magnification m is *negative.*

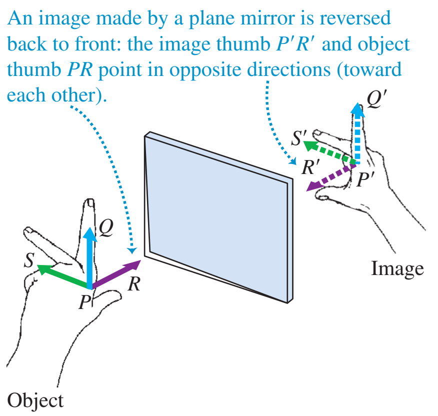
The object in Fig. 34.6 has only one dimension. Figure 34.7 shows a *three*-dimensional object and its three-dimensional virtual image formed by a plane mirror. The object and image are related in the same way as a left hand and a right hand.

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.* To verify this object-image relationship, point your thumbs along PR and $P'R'$, your forefingers along PQ and $P'Q'$, and your middle fingers along PS and $P'S'$. 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. Figure 34.8 illustrates this point.

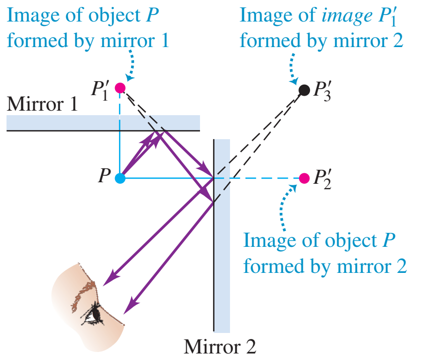
An important property of all images formed by reflecting or refracting surfaces is that an *image* formed by one surface or optical device can serve as the *object* for a second surface or device. Figure 34.9 shows a simple example. Mirror 1 forms an image P'_1 of the object point P, and mirror 2 forms another image P'_2, each in the way we have just discussed. But in addition, the image P'_1 formed by mirror 1 serves 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 in geometric optics. We will use it later in this chapter to locate the image formed by two successive curved-surface refractions in a lens. This idea will help us to understand image formation by combinations of lenses, as in a microscope or a refracting telescope.

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



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 H and A reversed?

34.9 Images P'_1 and P'_2 are formed by a single reflection of each ray from the object at P. Image P'_3, located by treating either of the other images as an object, is formed by a double reflection of each ray.



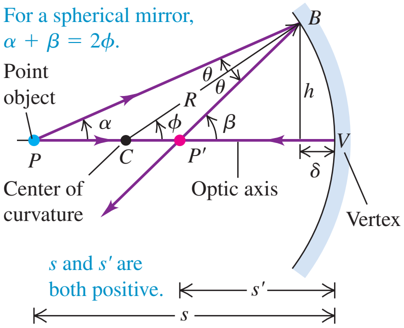
Test Your Understanding of Section 34.1 If you walk directly toward a plane mirror at a speed v, at what speed does your image approach you? (i) slower than v; (ii) v; (iii) faster than v but slower than $2v$; (iv) $2v$; (v) faster than $2v$.

34.2 Reflection at a Spherical Surface

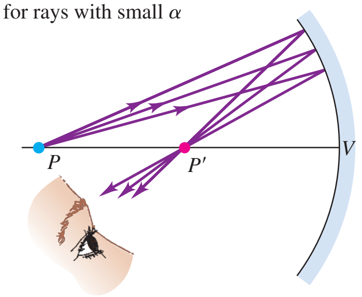
A plane mirror produces an image that is the same size as the object. But there are many applications for mirrors in which the image and object must be of different sizes. A magnifying mirror used when applying makeup gives an image that is *larger* than the object, and surveillance mirrors (used in stores to help spot shoplifters) give an image that is *smaller* than the object. There are also applications of mirrors in which a *real* image is desired, so light rays do indeed pass through the image point P'. A plane mirror by itself cannot perform any of these tasks. Instead, *curved* mirrors are used.

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

(a) Construction for finding the position P' of an image formed by a concave spherical mirror



(b) The paraxial approximation, which holds for rays with small α



All rays from P that have a small angle α pass through P', forming a real image.

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

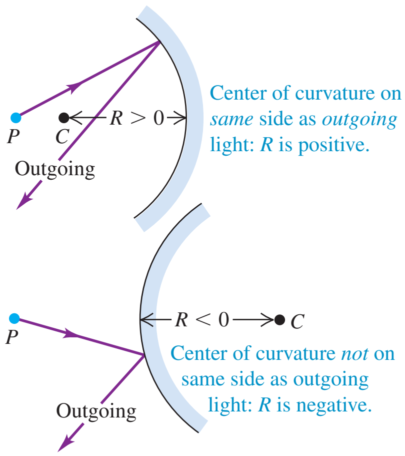


Image of a Point Object: Spherical Mirror

We'll consider the special (and easily analyzed) case of image formation by a *spherical* mirror. Figure 34.10a shows a spherical mirror with radius of curvature R, with its concave side facing the incident light. The **center of curvature** of the surface (the center of the sphere of which the surface is a part) is at C, and the **vertex** of the mirror (the center of the mirror surface) is at V. The line CV is called the **optic axis.** Point P is an object point that lies on the optic axis; for the moment, we assume that the distance from P to V is greater than R.

Ray PV, passing through C, strikes the mirror normally and is reflected back on itself. Ray PB, at an angle α with the axis, strikes the mirror at B, where the angles of incidence and reflection are θ. The reflected ray intersects the axis at point P'. We will show shortly that *all* rays from P intersect the axis at the *same* point P', as in Fig. 34.10b, provided that the angle α is small. Point P' is therefore the *image* of object point P. Unlike the reflected rays in Fig. 34.1, the reflected rays in Fig. 34.10b actually do intersect at point P', then diverge from P' *as if* they had originated at this point. Thus P' is a *real* image.

To see the usefulness of having a real image, suppose that the mirror is in a darkened room in which the only source of light is a self-luminous object at P. If you place a small piece of photographic film at P', all the rays of light coming from point P that reflect off the mirror will strike the same point P' on the film; when developed, the film will show a single bright spot, representing a sharply focused image of the object at point P. This principle is at the heart of most astronomical telescopes, which use large concave mirrors to make photographs of celestial objects. With a *plane* mirror like that in Fig. 34.2, placing a piece of film at the image point P' would be a waste of time; 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 find the location of the real image point P' in Fig. 34.10a and prove the assertion 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 neglect 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:

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} \qquad \text{(object–image relationship, spherical mirror)} \tag{34.4}$$

This equation does not contain the angle α. Hence *all* rays from P that make sufficiently small angles with the axis intersect at P' after they are reflected; this verifies our earlier assertion. Such rays, nearly parallel to the axis and close to it, are called **paraxial rays.** (The term **paraxial approximation** is often used for the approximations we have just described.) Since all such reflected light rays converge on the image point, a concave mirror is also called a *converging mirror.*

Be sure you understand that Eq. (34.4), as well as many similar relationships that we will derive later in this chapter and the next, is only *approximately* correct. It results from a calculation containing approximations, and it is valid only for paraxial rays. If we increase the angle α that a ray makes with the optic axis, the point P' where the ray intersects the optic axis moves somewhat closer to the vertex than for a paraxial ray. As a result, a spherical mirror, unlike a plane mirror, does not form a precise point image of a point object; the image is "smeared out." This property of a spherical mirror is called **spherical aberration.** When the primary mirror of the Hubble Space Telescope (Fig. 34.12a) was manufactured, tiny errors were made in its shape that led to an unacceptable amount of spherical aberration (Fig. 34.12b). The performance of the telescope improved dramatically after the installation of corrective optics (Fig. 34.12c).

If the radius of curvature becomes infinite ($R = \infty$), the mirror becomes *plane,* and Eq. (34.4) reduces to Eq. (34.1) for a plane reflecting surface.

Focal Point and Focal Length

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

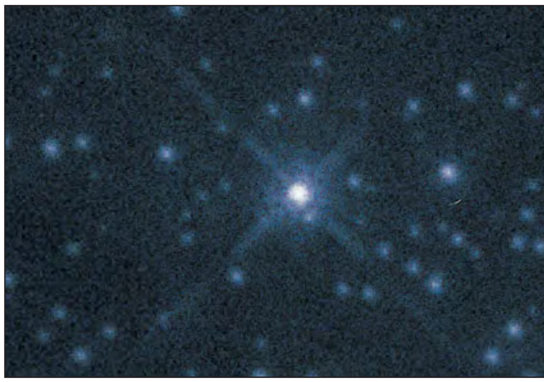
$$\frac{1}{\infty} + \frac{1}{s'} = \frac{2}{R} \qquad s' = \frac{R}{2}$$

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

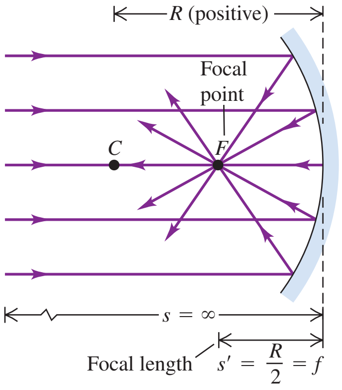
(b) A star seen with the original mirror

(c) The same star with corrective optics

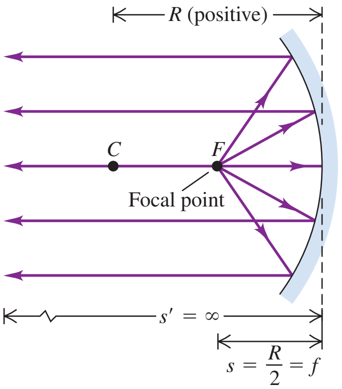


34.13 The focal point and focal length of a concave mirror.

(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.



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

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

The opposite situation is shown in Fig. 34.13b. 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; this is why parabolic mirrors are preferred for astronomical telescopes. 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; light from the sun is collected by the mirrors and directed to the focal point, where a steam boiler is placed. (The concepts of focal point and focal length also apply to lenses, as we'll see in Section 34.4.)

We will usually express the relationship between object and image distances for a mirror, Eq. (34.4), in terms of the focal length f:

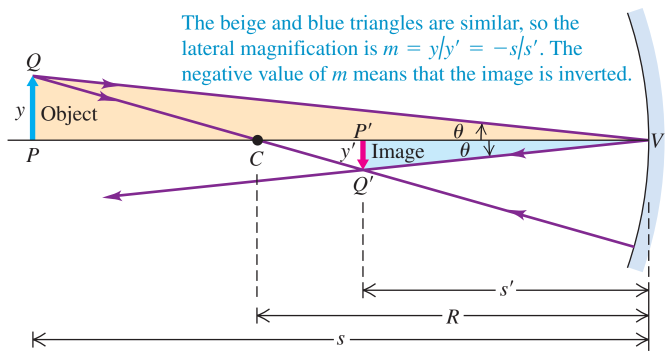
$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \quad \text{(object–image relationship, spherical mirror)} \tag{34.6}$$

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. In Eq. (34.2) we defined the *lateral magnification* m as the ratio of image size y' to object size y:

$$m = \frac{y'}{y}$$

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



Because triangles PVQ and $P'VQ'$ in Fig. 34.14 are *similar,* we also have the relationship $y/s = -y'/s'$. The negative sign is needed because object and image are on opposite sides of the optic axis; if y is positive, y' is negative. Therefore

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

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 placed *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 applying 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 an enlarged, erect image is seen. You can prove these statements about concave mirrors by applying Eqs. (34.6) and (34.7) (see Exercise 34.11). 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.

MasteringPHYSICS

ActivPhysics 15.5: Spherical Mirrors: Ray Diagrams
ActivPhysics 15.6: Spherical Mirrors: The Mirror Equation
ActivPhysics 15.7: Spherical Mirrors: Linear Magnification
ActivPhysics 15.8: Spherical Mirrors: Problems

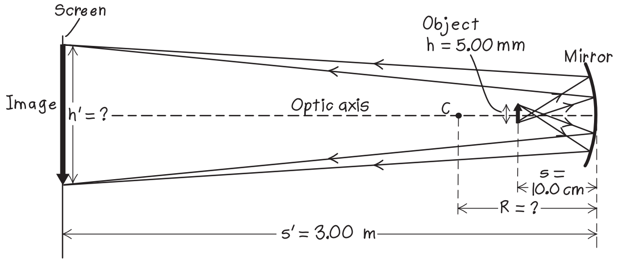
Example 34.1 Image formation by a concave mirror I

A concave mirror forms an image, on a wall 3.00 m in front of the mirror, of a headlamp filament 10.0 cm in front of the mirror. (a) What are the radius of curvature and focal length of the mirror? (b) What is the lateral magnification? What is the image height if the object height is 5.00 mm?

SOLUTION

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

34.15 Our sketch for this problem.



EXECUTE: (a) Both the object and the image are on the concave (reflective) side of the mirror, so both s and s' are positive; we have $s = 10.0$ cm and $s' = 300$ cm. We solve Eq. (34.4) for R:

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

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

The focal length of the mirror is $f = R/2 = 9.7$ cm.

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

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

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

Continued

EVALUATE: Our sketch indicates that the image is inverted; our calculations agree. Note that the object (at $s = 10.0$ cm) is just outside the focal point ($f = 9.7$ cm). This is very similar to what is done in automobile headlights. With the filament close to the focal point, the concave mirror produces a beam of nearly parallel rays.

Conceptual Example 34.2 Image formation by a concave mirror II

In Example 34.1, suppose that the lower half of the mirror's reflecting surface is covered with nonreflective soot. What effect will this have on the image of the filament?

SOLUTION

It would be natural to guess that the image would now show only half of the filament. But in fact the image will still show the *entire* filament. You can see why by examining Fig. 34.10b. Light rays coming from any object point P are reflected from *all* parts of the mirror and converge on the corresponding image point P'. If part of the mirror surface is made nonreflective (or is removed altogether), rays from the remaining reflective surface still form an image of every part of the object.

Reducing the reflecting area reduces the light energy reaching the image point, however: The image becomes *dimmer*. If the area is reduced by one-half, the image will be one-half as bright. Conversely, *increasing* the reflective area makes the image brighter. To make reasonably bright images of faint stars, astronomical telescopes use mirrors that are up to several meters in diameter (see Fig. 34.12a).

Convex Mirrors

In Fig. 34.16a the *convex* side of a spherical mirror faces the incident light. The center of curvature is on the side opposite to the outgoing rays; according to sign rule 3 in Section 34.1, R is negative (see Fig. 34.11). Ray PB is reflected, with the angles of incidence and reflection both equal to θ. The reflected ray, projected backward, intersects the axis at P'. As with a concave mirror, *all* rays from P that are reflected by the mirror diverge from the same point P', provided that the angle α is small. Therefore P' is the image of P. The object distance s is positive, the image distance s' is negative, and the radius of curvature R is *negative* for a *convex* mirror.

Figure 34.16b shows two rays diverging from the head of the arrow PQ and the virtual image $P'Q'$ of this arrow. The same procedure that we used for a concave mirror can be used to show that for a convex mirror,

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

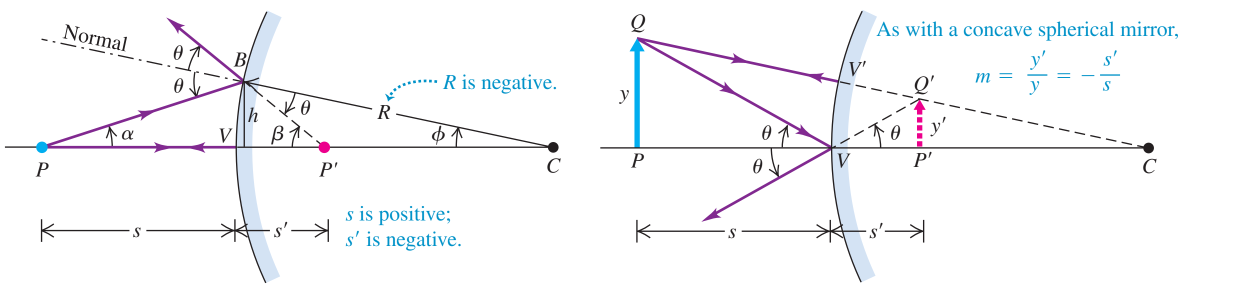
and the lateral magnification is

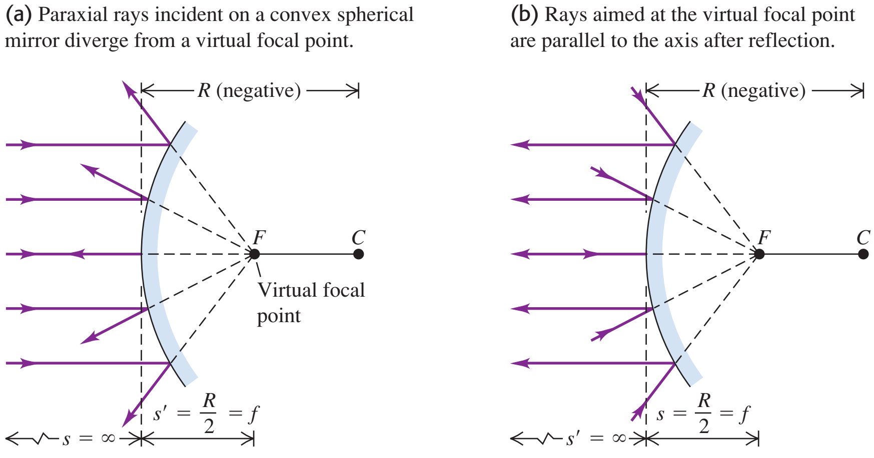
$$m = \frac{y'}{y} = -\frac{s'}{s}$$

34.16 Image formation by a convex mirror.

(a) Construction for finding the position of an image formed by a convex mirror

(b) Construction for finding the magnification of an image formed by a convex mirror





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

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.

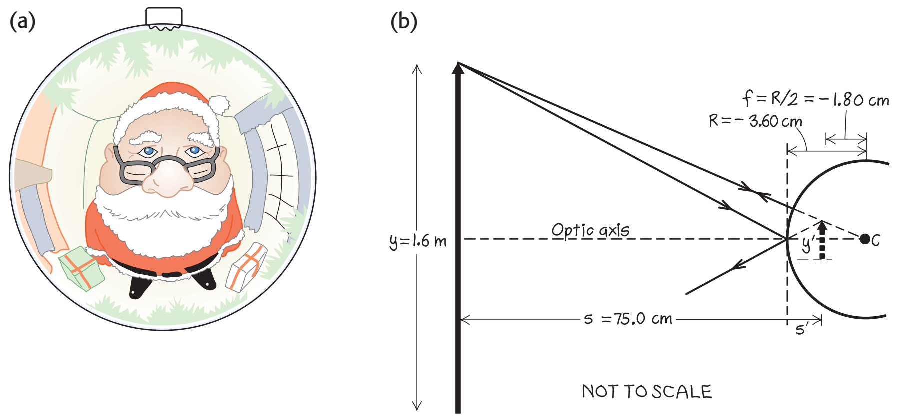
Example 34.3 Santa's image problem

Santa checks himself for soot, using his reflection in a silvered Christmas tree ornament 0.750 m away (Fig. 34.18a). The diameter of the ornament is 7.20 cm. Standard reference texts state that he is a "right jolly old elf," so we estimate his height to be 1.6 m. Where and how tall is the image of Santa formed by the ornament? Is it erect or inverted?

SOLUTION

IDENTIFY and SET UP: Figure 34.18b shows the situation. Santa is the object, and the surface of the ornament closest to him acts as a convex mirror. The relationships among object distance, image distance, focal length, and magnification are the same as for

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

concave mirrors, provided we use the sign rules consistently. The radius of curvature and the focal length of a convex mirror are *negative*. The object distance is $s = 0.750\text{ m} = 75.0\text{ cm}$, and Santa's height is $y = 1.6\text{ m}$. We solve Eq. (34.6) to find the image distance s', and then use Eq. (34.7) to find the lateral magnification m and the image height y'. The sign of m tells us whether the image is erect or inverted.

EXECUTE: The radius of the mirror (half the diameter) is $R = -(7.20\text{ cm})/2 = -3.60\text{ cm}$, and the focal length is $f = R/2 = -1.80\text{ cm}$. From Eq. (34.6),

$$\frac{1}{s'} = \frac{1}{f} - \frac{1}{s} = \frac{1}{-1.80\text{ cm}} - \frac{1}{75.0\text{ cm}}$$

$$s' = -1.76\text{ cm}$$

Because s' is negative, the image is behind the mirror—that is, on the side opposite to the outgoing light (Fig. 34.18b)—and it is virtual. The image is about halfway between the front surface of the ornament and its center.

From Eq. (34.7), the lateral magnification and the image height are

$$m = \frac{y'}{y} = -\frac{s'}{s} = -\frac{-1.76\text{ cm}}{75.0\text{ cm}} = 0.0234$$

$$y' = my = (0.0234)(1.6\text{ m}) = 3.8 \times 10^{-2}\text{ m} = 3.8\text{ cm}$$

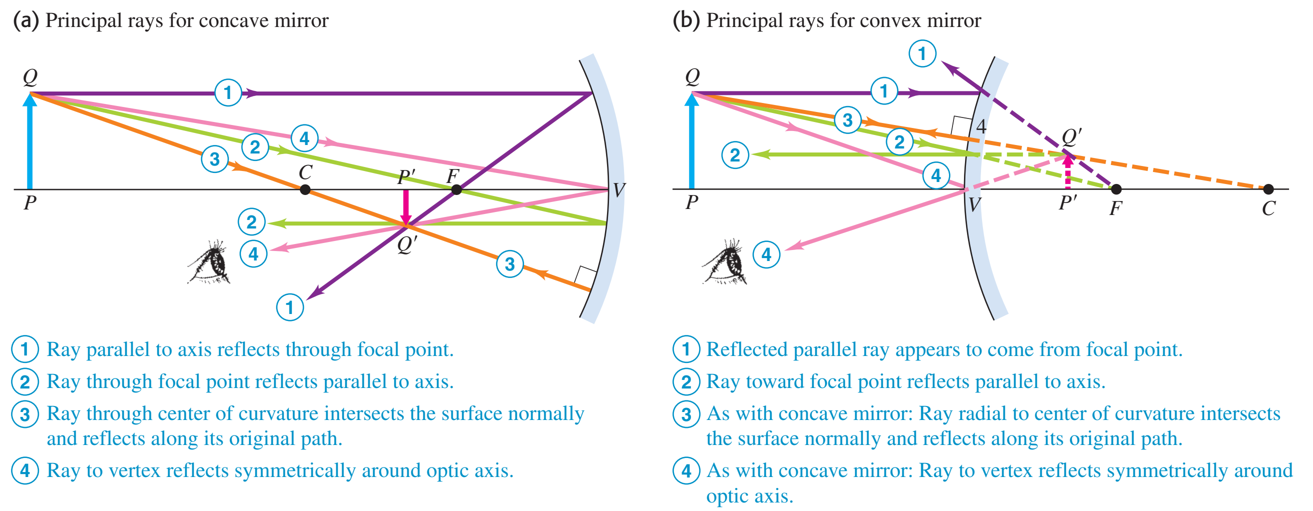
EVALUATE: Our sketch indicates that the image is erect so both m and y' are positive; our calculations agree. When the object distance s is positive, a convex mirror *always* forms an erect, virtual, reduced, reversed image. For this reason, convex mirrors are used at blind intersections, for surveillance in stores, and as wide-angle rear-view mirrors for cars and trucks. (Many such mirrors read "Objects in mirror are closer than they appear.")

Graphical Methods for Mirrors

In Examples 34.1 and 34.3, we used Eqs. (34.6) and (34.7) to find the position and size of the image formed by a mirror. We can also determine the properties of the image by a simple *graphical* method. This method consists of finding the point of intersection of a few particular rays that diverge from a point of the object (such as point Q in Fig. 34.19) and are reflected by the mirror. Then (neglecting aberrations) *all* rays from this object point that strike the mirror will intersect at the same point. For this construction we always choose an object point that is *not* on the optic axis. Four rays that we can usually draw easily are shown in Fig. 34.19. These are called **principal rays.**

1. *A ray parallel to the axis,* after reflection, passes through the focal point F of a concave mirror or appears to come from the (virtual) focal point of a convex mirror.
2. *A ray through (or proceeding toward) the focal point F* is reflected parallel to the axis.
3. *A ray along the radius* through or away from the center of curvature C intersects the surface normally and is reflected back along its original path.
4. *A ray to the vertex V* is reflected forming equal angles with the optic axis.

34.19 The graphical method of locating an image formed by a spherical mirror. The colors of the rays are for identification only; they do not refer to specific colors of light.



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 only need to draw the principal rays, because these are all you need 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 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

A concave mirror has a radius of curvature with absolute value 20 cm. Find graphically the image of an object in the form of an arrow perpendicular to the axis of the mirror at object distances of (a) 30 cm, (b) 20 cm, (c) 10 cm, and (d) 5 cm. Check the construction by *computing* the size and lateral magnification of each image.

SOLUTION

IDENTIFY and SET UP: We must use graphical methods *and* calculations to analyze the image made by a mirror. The mirror is concave, so its radius of curvature is $R = +20$ cm and its focal length is $f = R/2 = +10$ cm. Our target variables are the image distances s' and lateral magnifications m corresponding to four cases with successively smaller object distances s. In each case we solve Eq. (34.6) for s' and use $m = -s'/s$ to find m.

EXECUTE: Figure 34.20 shows the principal-ray diagrams for the four cases. Study each of these diagrams carefully and confirm that each numbered ray is drawn in accordance with the rules given earlier (under "Graphical Methods for Mirrors"). Several points are worth noting. First, in case (b) the object and image distances are equal. Ray 3 cannot be drawn in this case because a ray from Q through the center of curvature C does not strike the mirror. In case (c), ray 2 cannot be drawn because a ray from Q through F does not strike the mirror. In this case the outgoing rays are parallel, corresponding to an infinite image distance. In case (d), the outgoing rays diverge; they have been extended backward to the *virtual image point* Q', from which they appear to diverge. Case (d) illustrates the general observation that an object placed inside the focal point of a concave mirror produces a virtual image.

Measurements of the figures, with appropriate scaling, give the following approximate image distances: (a) 15 cm; (b) 20 cm; (c) ∞ or $-\infty$ (because the outgoing rays are parallel and do not converge at any finite distance); (d) -10 cm. To *compute* these distances, we solve Eq. (34.6) for s' and insert $f = 10$ cm:

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

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

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

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

The signs of s' tell us that the image is real in cases (a) and (b) and virtual in case (d).

Continued

The lateral magnifications measured from the figures are approximately (a) $-\frac{1}{2}$; (b) -1; (c) ∞ or $-\infty$; (d) $+2$. From Eq. (34.7),

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

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

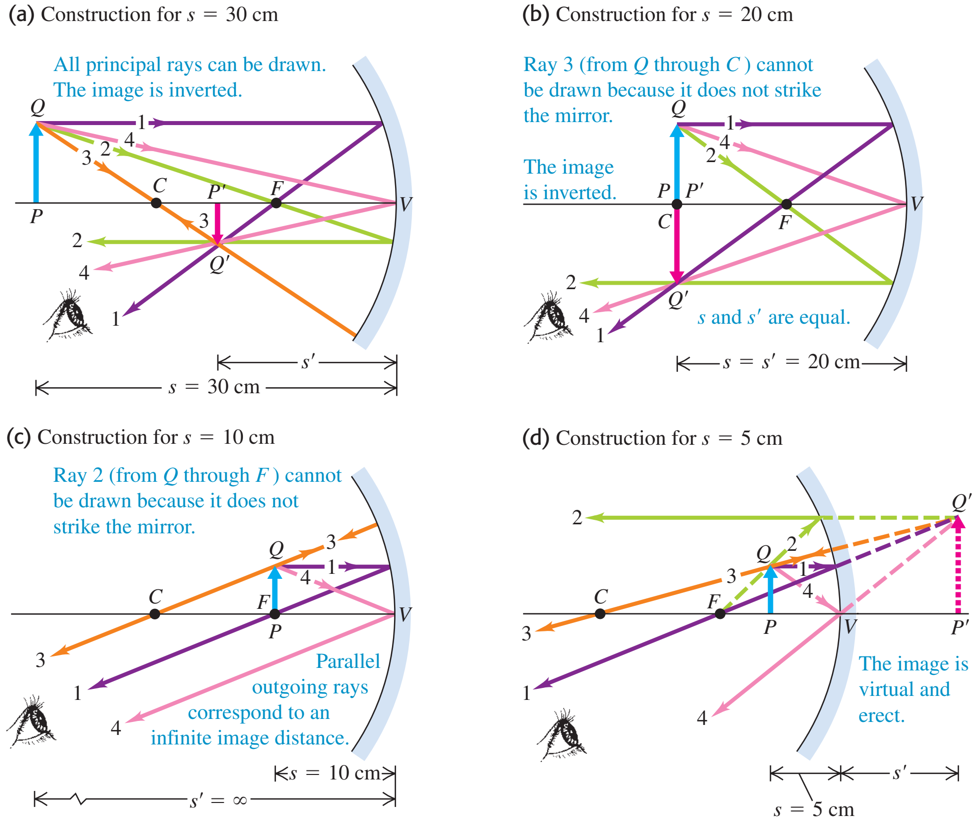
$$\text{(c) } m = -\frac{\infty \text{ cm}}{10 \text{ cm}} = -\infty \text{ (or } +\infty\text{)}$$

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

The signs of m tell us that the image is inverted in cases (a) and (b) and erect in case (d).

EVALUATE: Notice the trend of the results in the four cases. When the object is far from the mirror, as in Fig. 34.20a, the image is smaller than the object, inverted, and real. As the object distance s decreases, the image moves farther from the mirror and gets larger (Fig. 34.20b). When the object is at the focal point, the image is at infinity (Fig. 34.20c). When the object is inside the focal point, the image becomes larger than the object, erect, and virtual (Fig. 34.20d). You can confirm these conclusions by looking at objects reflected in the concave bowl of a shiny metal spoon.

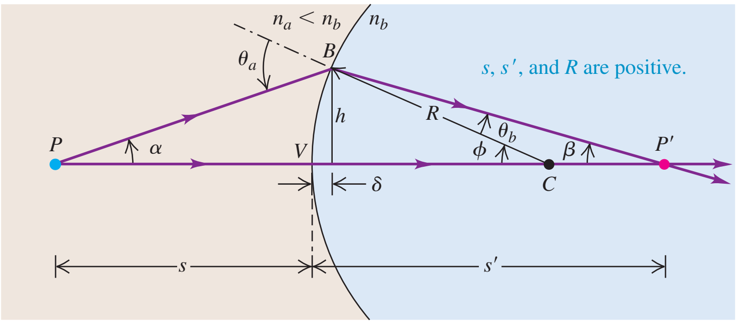
34.20 Using principal-ray diagrams to locate the image $P'Q'$ made by a concave mirror.



Test Your Understanding of Section 34.2 A cosmetics mirror is designed so that your reflection appears right-side up and enlarged. (a) Is the mirror concave or convex? (b) To see an enlarged image, what should be the distance from the mirror (of focal length f) to your face? (i) $|f|$; (ii) less than $|f|$; (iii) greater than $|f|$.

34.3 Refraction at a Spherical Surface

As we mentioned in Section 34.1, images can be formed by refraction as well as by reflection. To begin with, let's consider refraction at a spherical surface—that is, at a spherical interface between two optical materials with different indexes of refraction. This analysis is directly applicable to some real optical systems, such as the human eye. It also provides a stepping-stone for the analysis of lenses, which usually have *two* spherical (or nearly spherical) surfaces.



34.21 Construction for finding the position of the image point P' of a point object P formed by refraction at a spherical surface. The materials to the left and right of the interface have refractive indexes n_a and n_b, respectively. In the case shown here, $n_a < n_b$.

Image of a Point Object: Spherical Refracting Surface

In Fig. 34.21 a spherical surface with radius R forms an interface between two materials with different indexes of refraction n_a and n_b. The surface forms an image P' of an object point P; we want to find how the object and image distances (s and s') are related. We will use the same sign rules that we used for spherical mirrors. The center of curvature C is on the outgoing side of the surface, so R is positive. Ray PV strikes the vertex V and is perpendicular to the surface (that is, to the plane that is tangent to the surface at the point of incidence V). It passes into the second material without deviation. Ray PB, making an angle α with the axis, is incident at an angle θ_a with the normal and is refracted at an angle θ_b. These rays intersect at P', a distance s' to the right of the vertex. The figure is drawn for the case $n_a < n_b$. The object and image distances are both 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

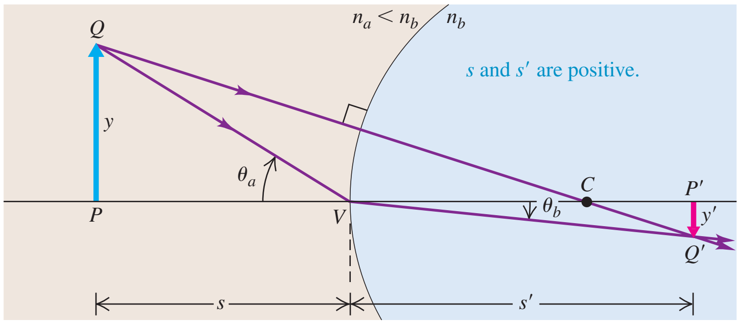
$$\theta_b = \frac{n_a}{n_b}(\alpha + \phi)$$

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 neglect the small distance δ; those equations then become

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



34.22 Construction for determining the height of an image formed by refraction at a spherical surface. In the case shown here, $n_a < n_b$.

Finally, we substitute these into Eq. (34.10) and divide out the common factor h. We obtain

$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R} \quad \text{(object–image relationship, spherical refracting surface)} \tag{34.11}$$

This equation does not contain the angle α, so the image distance is the same for *all* paraxial rays emanating from P; this proves our assertion 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

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

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

Equations (34.11) and (34.12) can be applied 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, you should construct diagrams like Figs. 34.21 and 34.22 for the following three cases: (i) $R > 0$ and $n_a > n_b$, (ii) $R < 0$ and $n_a < n_b$, and (iii) $R < 0$ and $n_a > n_b$. Then in each case, use your diagram to again derive Eqs. (34.11) and (34.12).

Here's a final note on the sign rule for the radius of curvature R of a surface. For the convex reflecting surface in Fig. 34.16, we considered R negative, but the convex *refracting* surface in Fig. 34.21 has a *positive* value of R. This may seem inconsistent, but it isn't. The rule is that R is positive if the center of curvature C is on the outgoing side of the surface and negative if C is on the other side. For the convex reflecting surface in Fig. 34.16, R is negative because point C is to the right of the surface but outgoing rays are to the left. For the convex refracting surface in Fig. 34.21, R is positive because both C and the outgoing rays are to the right of the surface.

Refraction at a curved surface is one reason gardeners avoid watering plants at midday. As sunlight enters a water drop resting on a leaf (Fig. 34.23), the light rays are refracted toward each other as in Figs. 34.21 and 34.22. The sunlight that strikes the leaf is therefore more concentrated and able to cause damage.

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 object appears to have a sharp bend at the water surface because 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.)

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

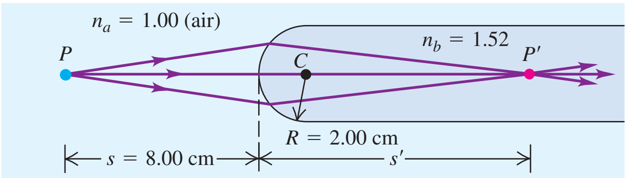
Example 34.5 Image formation by refraction I

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.

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of refraction at a curved surface. Our target variables are the image distance s' and the lateral magnification m. Here material a is air $(n_a = 1.00)$ and material b is the glass of which the rod is made $(n_b = 1.52)$. We are given $s = 8.00\text{ cm}$. The center of curvature of the spherical surface is on the outgoing side of the surface, so the radius is positive: $R = +2.00\text{ cm}$. We solve Eq. (34.11) for s', and we use Eq. (34.12) to find m.

34.24 The glass rod in air forms a real image.



EXECUTE: (a) From Eq. (34.11),

$$\frac{1.00}{8.00\text{ cm}} + \frac{1.52}{s'} = \frac{1.52 - 1.00}{+2.00\text{ cm}}$$
$$s' = +11.3\text{ cm}$$

(b) From Eq. (34.12),

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

EVALUATE: Because the image distance s' is positive, the image is formed 11.3 cm to the *right* of the vertex (on the outgoing side), as Fig. 34.24 shows. The value of m tells us that the image is somewhat smaller than the object and that it is inverted. If the object is an arrow 1.000 mm high, pointing upward, the image is an arrow 0.929 mm high, pointing downward.

Example 34.6 Image formation by refraction II

The glass rod of Example 34.5 is immersed in water, which has index of refraction $n = 1.33$ (Fig. 34.25). The object distance is again 8.00 cm. Find the image distance and lateral magnification.

SOLUTION

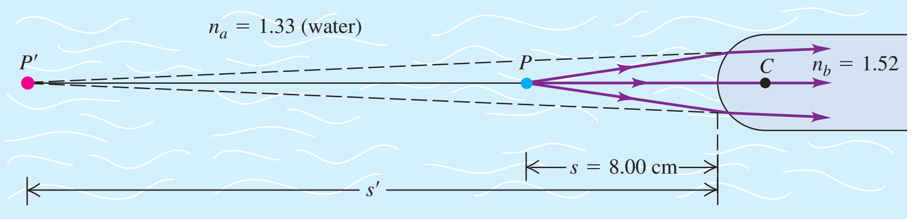
IDENTIFY and SET UP: The situation is the same as in Example 34.5 except that now $n_a = 1.33$. We again use Eqs. (34.11) and (34.12) to determine s' and m, respectively.

EXECUTE: Our solution of Eq. (34.11) in Example 34.5 yields

$$\frac{1.33}{8.00\text{ cm}} + \frac{1.52}{s'} = \frac{1.52 - 1.33}{+2.00\text{ cm}}$$
$$s' = -21.3\text{ cm}$$

Continued

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



The lateral magnification in this case is

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

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.

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?

SOLUTION

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

EXECUTE: The object distance is the actual depth of the pool, $s = 2.00$ m. Material a is water $(n_a = 1.33)$ and material b is air $(n_b = 1.00)$. From Eq. (34.13),

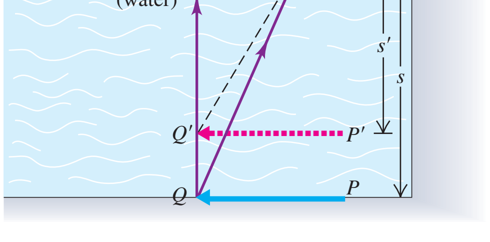
$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{1.33}{2.00\text{ m}} + \frac{1.00}{s'} = 0$$

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

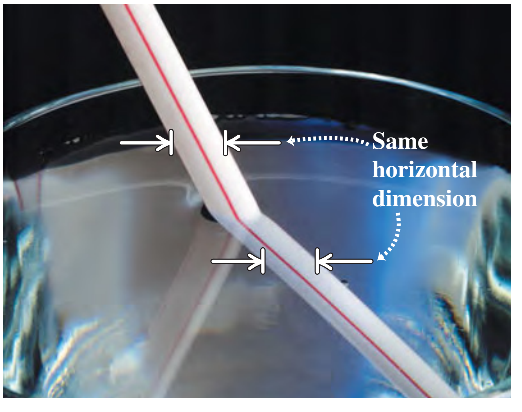
The image distance is negative. By the sign rules in Section 34.1, this means that the image is virtual and on the incoming side of the refracting surface—that is, on the same side as the object, namely underwater. The pool's apparent depth is 1.50 m, or just 75% of its true depth.

EVALUATE: Recall that the lateral magnification for a plane refracting surface is $m = 1$. Hence the image $P'Q'$ of the arrow has the same *horizontal length* as the actual arrow PQ (Fig. 34.27). Only its depth is different from that of PQ.

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



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



Test Your Understanding of Section 34.3 The water droplets in Fig. 34.23 have radius of curvature R and index of refraction $n = 1.33$. Can they form an image of the sun on the leaf?

34.4 Thin Lenses

MasteringPHYSICS

PhET: Geometric Optics
ActivPhysics 15.9: Thin Lens Ray Diagram
ActivPhysics 15.10: Converging Thin Lenses
ActivPhysics 15.11: Diverging 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 neglect 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. We can analyze thin lenses in detail using the results of Section 34.3 for refraction by a single spherical surface. However, we postpone this analysis until later in the section so that we can first discuss the properties of thin lenses.

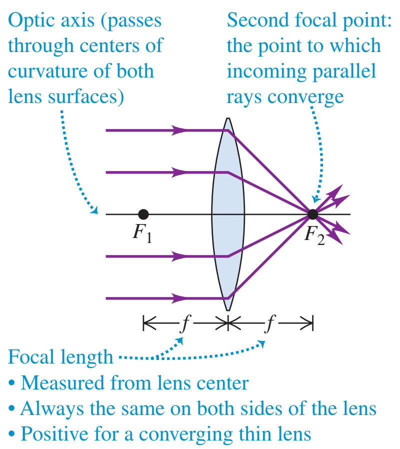
Properties of a Lens

A lens of the shape shown in Fig. 34.28 has the property that when a beam of rays parallel to the axis passes through the lens, the rays converge to a point F_2 (Fig. 34.28a) and form a real image at that point. Such a lens is called a **converging lens.** Similarly, rays passing through point F_1 emerge from the lens as a beam of parallel rays (Fig. 34.28b). The points F_1 and F_2 are called the first and second *focal points,* and the distance f (measured from the center of the lens) is called the *focal length.* Note the similarities between the two focal points of a converging lens and the single focal point of a concave mirror (see Fig. 34.13). As for a concave mirror, the focal length of a converging lens is defined to be a *positive* quantity, and such a lens is also called a *positive lens.*

The central horizontal line in Fig. 34.28 is called the *optic axis,* as with spherical mirrors. The centers of curvature of the two spherical surfaces lie on and define the optic axis. The two focal lengths in Fig. 34.28, both labeled f, *are always equal* for a thin lens, even when the two sides have different curvatures. We will derive this somewhat surprising 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.

34.28 F_1 and F_2 are the first and second focal points of a converging thin lens. The numerical value of f is positive.

(a)



(b)

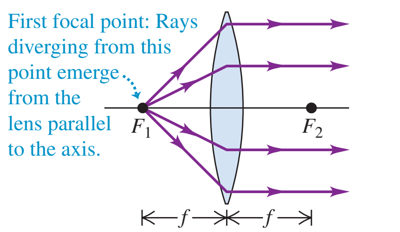
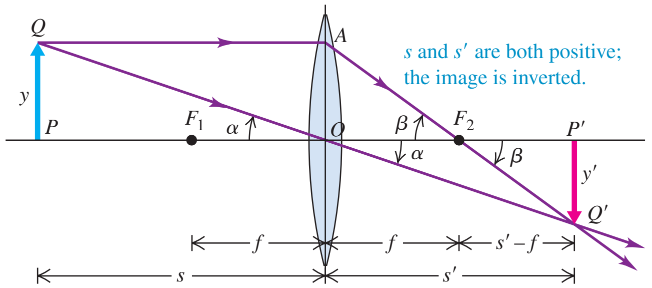


Image of an Extended Object: Converging Lens

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.



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 two angles labeled α in Fig. 34.29 are equal. Therefore the two right triangles PQO and $P'Q'O$ are *similar,* and ratios of corresponding sides are equal. Thus

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

(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

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad \text{(object–image relationship, thin lens)} \tag{34.16}$$

This analysis also gives the lateral magnification $m = y'/y$ for the lens; from Eq. (34.14),

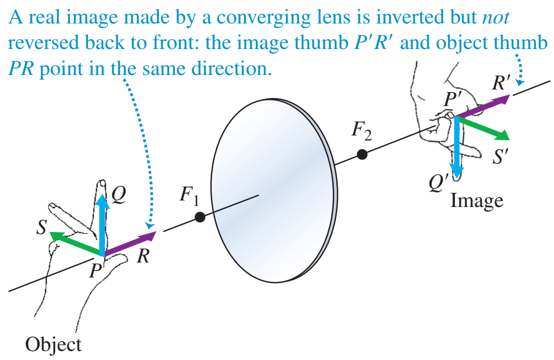
$$m = -\frac{s'}{s} \qquad \text{(lateral magnification, thin lens)} \tag{34.17}$$

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 will see, the same sign rules that we used for spherical mirrors are also applicable to lenses. In particular, consider a lens with a positive focal length (a converging lens). When an object is outside the first focal point F_1 of this lens (that is, when $s > f$), the image distance s' is positive (that is, the image is on the same side as the outgoing rays); this image is real and inverted, as in Fig. 34.29. An object placed inside the first focal point of a converging lens, so that $s < f$, produces an image with a negative value of s'; this image is located on the same side of the lens as the object and is virtual, erect, and larger than the object. You can verify these statements algebraically 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 $P'R'$

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



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, 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.

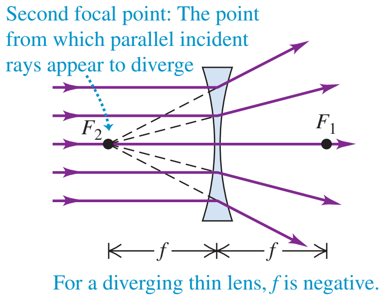
Diverging Lenses

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.

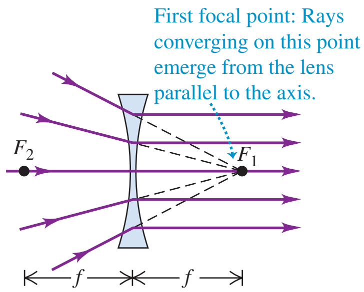
Equations (34.16) and (34.17) apply to *both* positive and negative lenses. Figure 34.32 shows various types of lenses, both converging and diverging. Here's an important observation: *Any lens that is thicker at its center than at its edges is a converging lens with positive f; and any lens that is 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 using the *lensmaker's equation,* which it is our next task to derive.

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.

(a)



(b)



The Lensmaker's Equation

We'll now derive Eq. (34.16) in more detail and at the same time derive the *lensmaker's equation,* which is a relationship among the focal length f, the index of refraction n of the lens, and the radii of curvature R_1 and R_2 of the lens surfaces. We use the principle that an image formed by one reflecting or refracting surface can serve as the object for a second reflecting or refracting surface.

We begin with the somewhat more general problem of two spherical interfaces separating three materials with indexes of refraction n_a, n_b, and n_c, as shown in Fig. 34.33. The object and image distances for the first surface are s_1 and s'_1, and those for the second surface are s_2 and s'_2. We assume that the lens is thin, so that the distance t between the two surfaces is small in comparison with the object and image distances and can therefore be neglected. 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

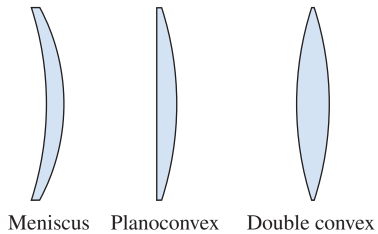
$$\frac{n_a}{s_1} + \frac{n_b}{s'_1} = \frac{n_b - n_a}{R_1}$$

$$\frac{n_b}{s_2} + \frac{n_c}{s'_2} = \frac{n_c - n_b}{R_2}$$

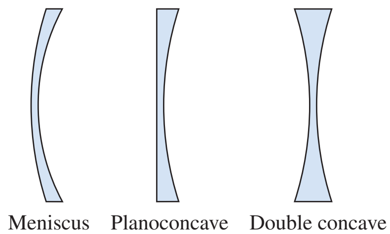
Ordinarily, the first and third materials are air or vacuum, so we set $n_a = n_c = 1$. The second index n_b is that of the lens, which we can call simply n. Substituting these values and the relationship $s_2 = -s'_1$, we get

34.32 Various types of lenses.

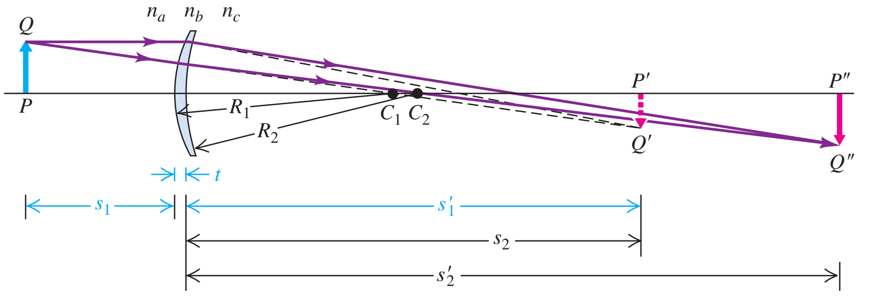
(a) **Converging lenses**



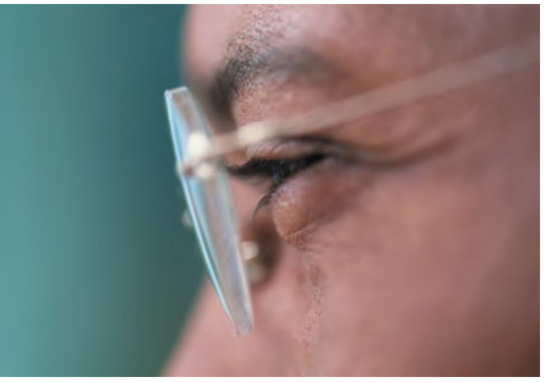
(b) **Diverging lenses**



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.



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



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

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

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, and we obtain

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

Finally, thinking of the lens as a single unit, we call the object distance simply s instead of s_1, and we call the final image distance s' instead of s'_2. Making these substitutions, we have

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

Now we compare this with the other thin-lens equation, Eq. (34.16). We see that the object and image distances s and s' appear in exactly the same places in both equations and that the focal length f is given by

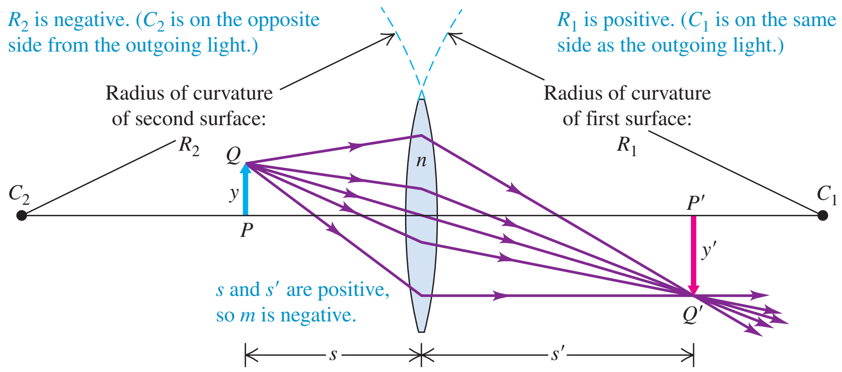
$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad \text{(lensmaker's equation for a thin lens)} \tag{34.19}$$

This is the **lensmaker's equation.** In the process of rederiving the relationship among object distance, image distance, and focal length for a thin lens, we have also derived 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 positive focal lengths and that all the lenses in Fig. 34.32b are diverging lenses with negative focal lengths.

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.



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

Example 34.8 Determining the focal length of a lens

(a) Suppose the absolute values of the radii of curvature of the lens surfaces in Fig. 34.35 are both equal to 10 cm and the index of refraction of the glass is $n = 1.52$. What is the focal length f of the lens? (b) Suppose the lens in Fig. 34.31 also has $n = 1.52$ and the absolute values of the radii of curvature of its lens surfaces are also both equal to 10 cm. What is the focal length of this lens?

SOLUTION

IDENTIFY and SET UP: We are asked to find the focal length f of (a) a lens that is convex on both sides (Fig. 34.35) and (b) a lens that is concave on both sides (Fig. 34.31). In both cases we solve the lensmaker's equation, Eq. (34.19), to determine f. We apply the sign rules given in Section 34.1 to the radii of curvature R_1 and R_2 to take account of whether the surfaces are convex or concave.

EXECUTE: (a) The lens in Fig. 34.35 is *double convex:* The center of curvature of the first surface (C_1) is on the outgoing side of the lens, so R_1 is positive, and the center of curvature of the second surface (C_2) is on the *incoming* side, so R_2 is negative. Hence $R_1 = +10$ cm and $R_2 = -10$ cm. Then from Eq. (34.19),

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

(b) The lens in Fig. 34.31 is *double concave:* The center of curvature of the first surface is on the *incoming* side, so R_1 is negative, and the center of curvature of the second surface is on the outgoing side, so R_2 is positive. Hence in this case $R_1 = -10$ cm and $R_2 = +10$ cm. Again using Eq. (34.19),

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

EVALUATE: In part (a) the focal length is *positive,* so this is a converging lens; this makes sense, since the lens is thicker at its center than at its edges. In part (b) the focal length is *negative,* so this is a diverging lens; this also makes sense, since the lens is thicker at its edges than at its center.

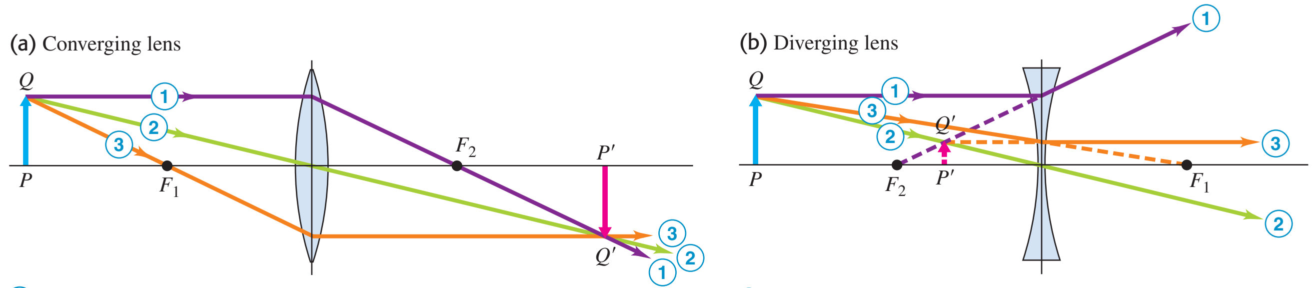
Graphical Methods for Lenses

We can determine the position and size of an image formed by a thin lens by using a graphical method very similar to the one we used in Section 34.2 for spherical mirrors. Again we draw a few special rays called *principal rays* that diverge from a point of the object that is *not* on the optic axis. The intersection of these rays, after they pass through the lens, determines the position and size of the image. In using this graphical method, we will consider the entire deviation of a ray as occurring at the midplane of the lens, as shown in Fig. 34.36. This is consistent with the assumption that the distance between the lens surfaces is negligible.

The three principal rays whose paths are usually easy to trace for lenses are shown in Fig. 34.36:

1. *A ray parallel to the axis* emerges from the lens in a direction that passes through the second focal point F_2 of a converging lens, or appears to come from the second focal point of a diverging lens.
2. *A ray through the center of the lens* is not appreciably deviated; at the center of the lens the two surfaces are parallel, so this ray emerges at essentially the same angle at which it enters and along essentially the same line.
3. *A ray through (or proceeding toward) the first focal point* F_1 emerges parallel to the axis.

34.36 The graphical method of locating an image formed by a thin lens. The colors of the rays are for identification only; they do not refer to specific colors of light. (Compare Fig. 34.19 for spherical mirrors.)



(1) Parallel incident ray refracts to pass through second focal point F_2.
(2) Ray through center of lens does not deviate appreciably.
(3) Ray through the first focal point F_1 emerges parallel to the axis.

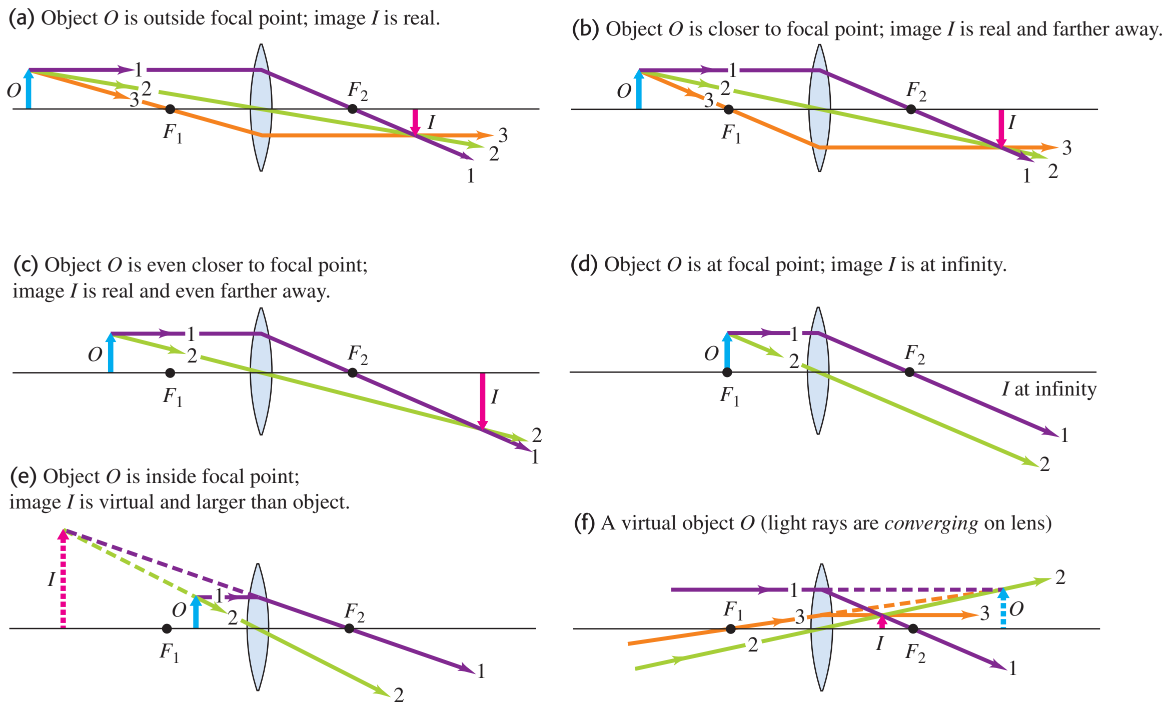
(1) Parallel incident ray appears after refraction to have come from the second focal point F_2.
(2) Ray through center of lens does not deviate appreciably.
(3) Ray aimed at the first focal point F_1 emerges parallel to the axis.

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.

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.)



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 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 film is moved farther behind the lens (i.e., the lens is moved farther in front of the film). In Fig. 34.37d the object is at the focal point; ray 3 can't be drawn because it doesn't pass through the lens. In Fig. 34.37e the object distance is less than the focal length. The outgoing rays are divergent, and the image is *virtual;* its position is located by extending the outgoing rays backward, so the image distance s' is negative. Note also that the image is erect and larger than the object. (We'll see the usefulness of this in Section 34.6.) Figure 34.37f corresponds to a *virtual object.* The incoming rays do not diverge from a real object, but are *converging* as though they would meet at the tip of the virtual object O on the right side; the object distance s is negative in this case. The image is real and is located between the lens and the second focal point. This situation can arise if the rays that strike the lens in Fig. 34.37f emerge from another converging lens (not shown) to the left of the figure.

Problem-Solving Strategy 34.2 Image Formation by Thin Lenses

IDENTIFY *the relevant concepts:* Review Problem-Solving Strategy 34.1 (Section 34.2) for mirrors, which is equally applicable here. As for mirrors, you should solve problems involving image formation by lenses using *both* principal-ray diagrams and equations.

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. Make sure that you 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

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 using Eqs. (34.16) and (34.17), respectively.

SOLUTION

IDENTIFY and SET UP: We are given the focal length $f = 20$ cm and four object distances s. Our target variables are the corresponding image distances s' and lateral magnifications m. We solve Eq. (34.16) for s', and find m from Eq. (34.17), $m = -s'/s$.

EXECUTE: Figures 34.37a, d, e, and f, respectively, show the appropriate principal-ray diagrams. You should be able to reproduce these without referring to the figures. Measuring these diagrams yields the approximate results: $s' = 35$ cm, $-\infty$, -40 cm, and 15 cm, and $m = -\frac{2}{3}$, $+\infty$, $+3$, and $+\frac{1}{3}$, respectively.

Calculating the image distances from Eq. (34.16), we find

$$\text{(a)}\ \frac{1}{50\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = 33.3\text{ cm}$$

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

$$\text{(c)}\ \frac{1}{15\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = -60\text{ cm}$$

$$\text{(d)}\ \frac{1}{-40\text{ cm}} + \frac{1}{s'} = \frac{1}{20\text{ cm}} \qquad s' = 13.3\text{ cm}$$

The graphical results are fairly close to these except for part (c); the accuracy of the diagram in Fig. 34.37e is limited because the rays extended backward have nearly the same direction.

From Eq. (34.17),

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

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

Continued

EVALUATE: Note that the image distance s' is positive in parts (a) and (d) but negative in part (c). This makes sense: The image is real in parts (a) and (d) but virtual in part (c). The light rays that emerge from the lens in part (b) are parallel and never converge, so the image can be regarded as being at either $+\infty$ or $-\infty$.

The values of magnification m tell us that the image is inverted in part (a) and erect in parts (c) and (d), in agreement with the principal-ray diagrams. The infinite value of magnification in part (b) is another way of saying that the image is formed infinitely far away.

Example 34.10 Image formation by a diverging lens

A beam of parallel rays spreads out after passing through a thin diverging lens, as if the rays all came from a point 20.0 cm from the center of the lens. You want to use this lens to form an erect, virtual image that is $\frac{1}{3}$ the height of the object. (a) Where should the object be placed? Where will the image be? (b) Draw a principal-ray diagram.

SOLUTION

IDENTIFY and SET UP: The result with parallel rays shows that the focal length is $f = -20$ cm. We want the lateral magnification to be $m = +\frac{1}{3}$ (positive because the image is to be erect). Our target variables are the object distance s and the image distance s'. In part (a), we solve the magnification equation, Eq. (34.17), for s' in terms of s; we then use the object–image relationship, Eq. (34.16), to find s and s' individually.

EXECUTE: (a) From Eq. (34.17), $m = +\frac{1}{3} = -s'/s$, so $s' = -s/3$. We insert this result into Eq. (34.16) and solve for the object distance s:

$$\frac{1}{s} + \frac{1}{-s/3} = \frac{1}{s} - \frac{3}{s} = -\frac{2}{s} = \frac{1}{f}$$

$$s = -2f = -2(-20.0\text{ cm}) = 40.0\text{ cm}$$

The object should be 40.0 cm from the lens. The image distance will be

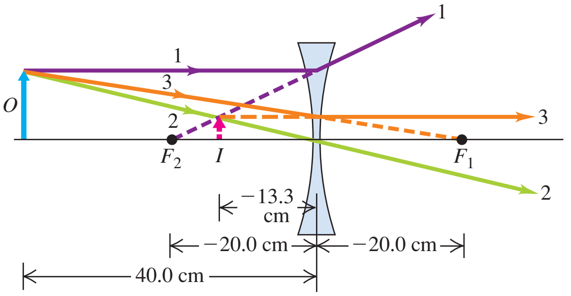
$$s' = -\frac{s}{3} = -\frac{40.0\text{ cm}}{3} = -13.3\text{ cm}$$

The image distance is negative, so the object and image are on the same side of the lens.

(b) Figure 34.38 is a principal-ray diagram for this problem, with the rays numbered as in Fig. 34.36b.

EVALUATE: You should be able to draw a principal-ray diagram like Fig. 34.38 without referring to the figure. From your diagram, you can confirm our results in part (a) for the object and image distances. You can also check our results for s and s' by substituting them back into Eq. (34.16).

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



Example 34.11 An image of an image

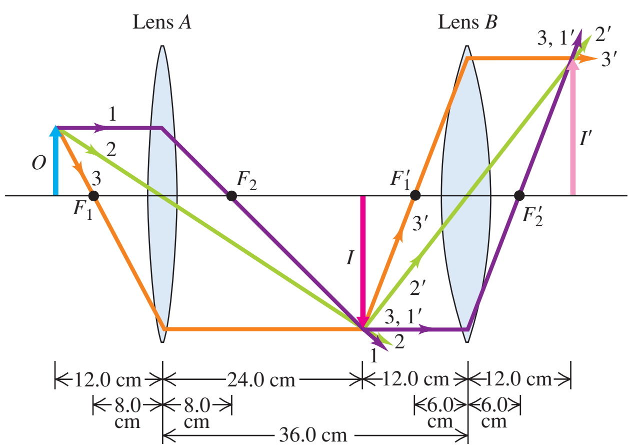
Converging lenses A and B, of focal lengths 8.0 cm and 6.0 cm, respectively, are placed 36.0 cm apart. Both lenses have the same optic axis. An object 8.0 cm high is placed 12.0 cm to the left of lens A. Find the position, size, and orientation of the image produced by the lenses in combination. (Such combinations are used in telescopes and microscopes, to be discussed in Section 34.7.)

SOLUTION

IDENTIFY and SET UP: Figure 34.39 shows the situation. The object O lies outside the first focal point F_1 of lens A, which therefore produces a real image I. The light rays that strike lens B diverge from this real image just as if I was a material object; image I therefore acts as an *object* for lens B. Our goal is to determine the properties of the image I' made by lens B. We use both ray-diagram and computational methods to do this.

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.

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).



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.

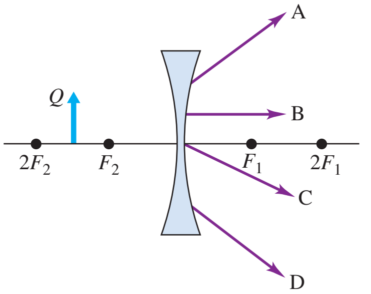
Image I is $36.0\text{ cm} - 24.0\text{ cm} = 12.0$ cm to the left of lens B, so the object distance for lens B is $+12.0$ cm. Applying Eq. (34.16) to lens B then gives

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

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

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

Test Your Understanding of Section 34.4 A diverging lens and an object are positioned as shown in the figure at right. Which of the rays A, B, C, and D could emanate from point Q at the top of the object?

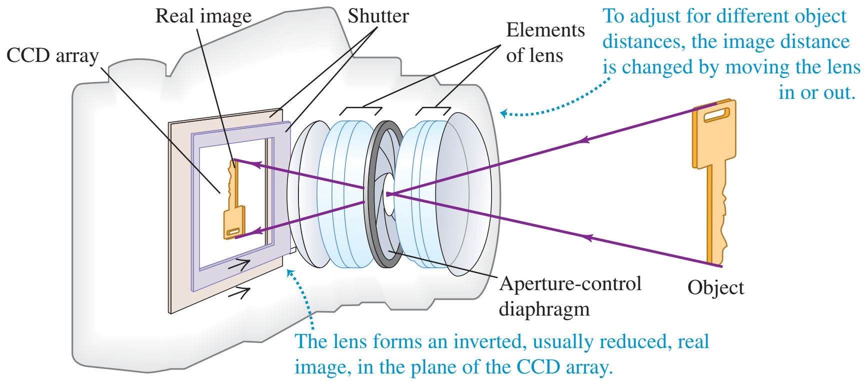


34.5 Cameras

The concept of *image,* which is so central to understanding simple mirror and lens systems, plays an equally important role in the analysis of optical instruments (also called *optical devices*). Among the most common optical devices are cameras, which make an image of an object and record it either electronically or on film.

The basic elements of a **camera** are a light-tight box ("camera" is a Latin word meaning "a room or enclosure"), a converging lens, a shutter to open the lens for a prescribed length of time, and a light-sensitive recording medium (Fig. 34.40). In a digital camera this is an electronic detector called a charge-coupled device (CCD) array; in an older camera, this is photographic film. The lens forms an inverted real image on the recording medium of the object being photographed. High-quality camera lenses have several elements, permitting partial correction of various *aberrations,* including the dependence of index of refraction on wavelength and the limitations imposed by the paraxial approximation.

When the camera is in proper *focus,* the position of the recording medium coincides with the position of the real image formed by the lens. The resulting photograph will then be as sharp as possible. With a converging lens, the image distance increases as the object distance decreases (see Figs. 34.41a, 34.41b, and 34.41c, and the discussion in Section 34.4). Hence in "focusing" the camera, we move the lens closer to the film for a distant object and farther from the film for a nearby object.



34.40 Key elements of a digital camera.

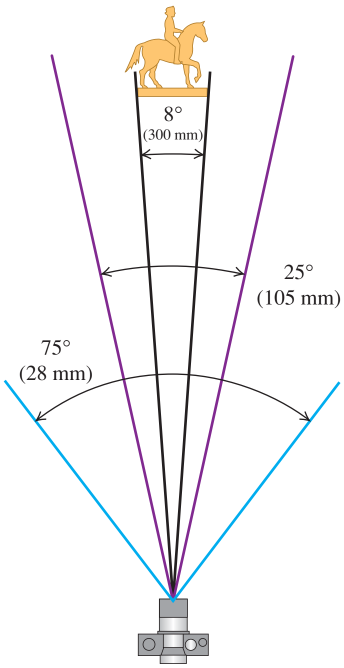
34.41 (a), (b), (c) Three photographs taken with the same camera from the same position in the Boston Public Garden using lenses with focal lengths $f = 28$ mm, 105 mm, and 300 mm. Increasing the focal length increases the image size proportionately. (d) The larger the value of f, the smaller the angle of view. The angles shown here are for a camera with image area 24 mm $\times$ 36 mm (corresponding to 35-mm film) and refer to the angle of view along the diagonal dimension of the film.

(a) $f = 28$ mm

(b) $f = 105$ mm

(c) $f = 300$ mm

(d) The angles of view for the photos in (a)–(c)



Camera Lenses: Focal Length

The choice of the focal length f for a camera lens depends on the film size 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 small 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 was 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 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 film. Hence the longer the focal length, the narrower the angle of view (Fig. 34.41d).

Camera Lenses: f-Number

For the film to record the image properly, the total light energy per unit area reaching the 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 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 film with a particular lens is proportional to D^2/f^2. The light-gathering capability of a lens is commonly expressed by photographers in terms of the ratio f/D, called the ***f*-number** of the lens:

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

For example, a lens with a focal length $f = 50$ mm and an aperture diameter $D = 25$ mm is said to have an f-number of 2, or "an aperture of $f/2$." The light intensity reaching the 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 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 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.

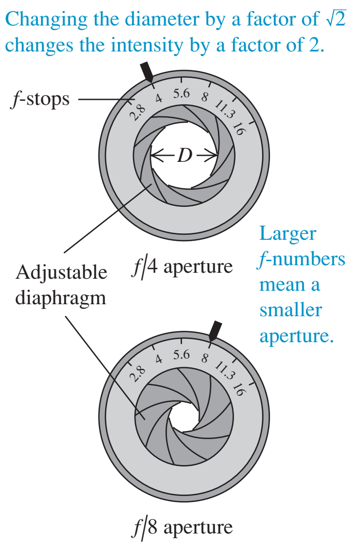
Zoom Lenses and Projectors

Many photographers use a *zoom lens,* which is not a single lens but a complex collection of several lens elements that give a continuously variable focal length, often over a range as great as 10 to 1. Figures 34.43a and 34.43b show a simple system with variable focal length, and Fig. 34.43c shows a typical zoom lens for a 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.

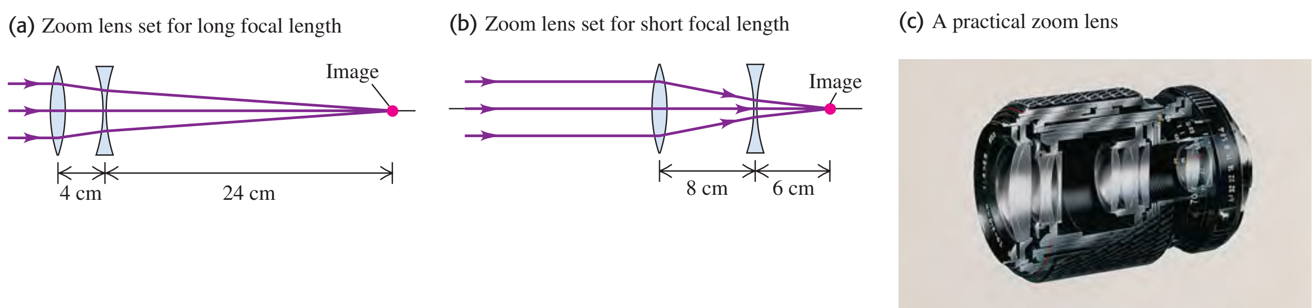
A *projector* for viewing slides, digital images, or motion pictures operates very much like a camera in reverse. In a movie projector, a lamp shines on the film, which acts as an object for the projection lens. The lens forms a real, enlarged, inverted image of the film on the projection screen. Because the image is inverted, the film goes through the projector upside down so that the image on the screen appears right-side up.

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.

34.42 A camera lens with an adjustable diaphragm.



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) A typical zoom lens for a single-lens reflex camera, containing twelve elements arranged in four groups.



Example 34.12 Photographic exposures

A common telephoto lens for a 35-mm camera has a focal length of 200 mm; its f-stops range from $f/2.8$ to $f/22$. (a) What is the corresponding range of aperture diameters? (b) What is the corresponding range of image intensities on the film?

SOLUTION

IDENTIFY and SET UP: Part (a) of this problem uses the relationship among lens focal length f, aperture diameter D, and f-number. Part (b) uses the relationship between intensity and aperture diameter. We use Eq. (34.20) to relate D (the target variable) to the f-number and the focal length $f = 200\text{ mm}$. The intensity of the light reaching the film is proportional to D^2/f^2; since f is the same in each case, we conclude that the intensity in this case is proportional to D^2, the square of the aperture diameter.

EXECUTE: (a) From Eq. (34.20), the diameter ranges from

$$D = \frac{f}{f\text{-number}} = \frac{200\text{ mm}}{2.8} = 71\text{ mm}$$

to

$$D = \frac{200\text{ mm}}{22} = 9.1\text{ mm}$$

(b) Because the intensity is proportional to D^2, the ratio of the intensity at $f/2.8$ to the intensity at $f/22$ is

$$\left(\frac{71\text{ mm}}{9.1\text{ mm}}\right)^2 = \left(\frac{22}{2.8}\right)^2 = 62 \qquad (\text{about } 2^6)$$

EVALUATE: If the correct exposure time at $f/2.8$ is $\frac{1}{1000}$ s, then the exposure at $f/22$ is $(62)\left(\frac{1}{1000}\text{ s}\right) = \frac{1}{16}$ s to compensate for the lower intensity. In general, the smaller the aperture and the larger the f-number, the longer the required exposure. Nevertheless, many photographers prefer to use small apertures so that only the central part of the lens is used to make the image. This minimizes aberrations that occur near the edges of the lens and gives the sharpest possible image.

Test Your Understanding of Section 34.5 When used with 35-mm film (image area 24 mm × 36 mm), a lens with $f = 50\text{ mm}$ gives a 45° angle of view and is called a "normal lens." When used with a CCD array that measures 5 mm × 5 mm, this same lens is (i) a wide-angle lens; (ii) a normal lens; (iii) a telephoto lens.

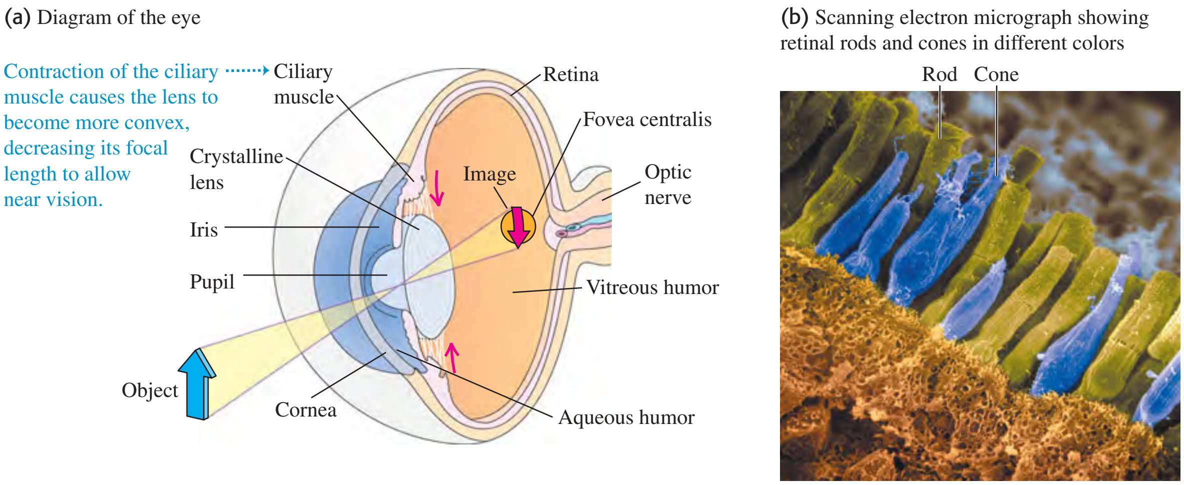
MasteringPHYSICS

PhET: Color Vision

34.6 The Eye

The optical behavior of the eye is similar to that of a camera. The essential parts of the human eye, considered as an optical system, are shown in Fig. 34.44a. The eye is nearly spherical and about 2.5 cm in diameter. The front portion is somewhat more sharply curved and is covered by a tough, transparent membrane called the *cornea.* The region behind the cornea contains a liquid called the *aqueous humor.* Next comes the *crystalline lens,* a capsule containing a fibrous jelly, hard at the center and progressively softer at the outer portions. The crystalline lens is

34.44 (a) The eye. (b) There are two types of light-sensitive cells on the retina. The rods are more sensitive to light than the 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.



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 film in a 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-film distance is changed.) For the normal eye, an object at infinity is sharply focused when the ciliary muscle is relaxed. To permit sharp imaging on the retina of closer objects, 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 the focal length. This process is called *accommodation.*

The extremes of the range over which distinct vision is possible are known as the *far point* and the *near point* of the eye. The far point of a normal eye is at infinity. The position of the near point depends on the amount by which the ciliary muscle can increase the curvature of the crystalline lens. The range of accommodation gradually diminishes with age because the crystalline lens grows throughout a person's life (it is about 50% larger at age 60 than at age 20) and the ciliary muscles are less able to distort a larger lens. For this reason, the near point gradually recedes as one grows older. This recession of the near point is called *presbyopia.* Table 34.1 shows the approximate position of the near point for an average person at various ages. For example, an average person 50 years of age cannot focus on an object that is closer than about 40 cm.

Table 34.1 Receding of Near Point with Age

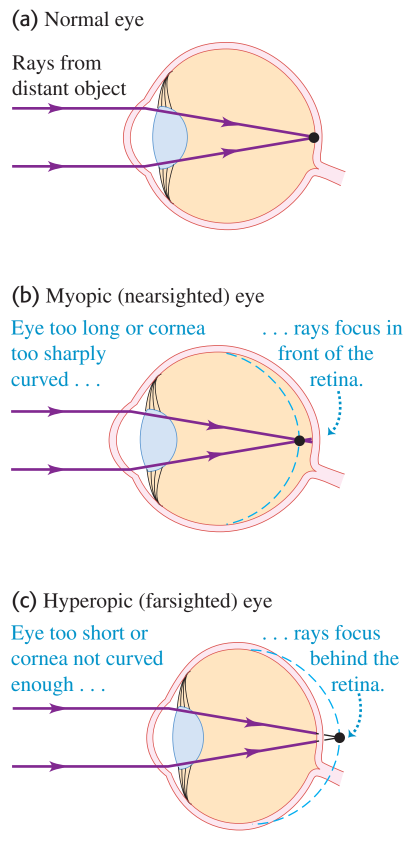
Age (years)	Near Point (cm)
10	7
20	10
30	14
40	22
50	40
60	200

Defects of Vision

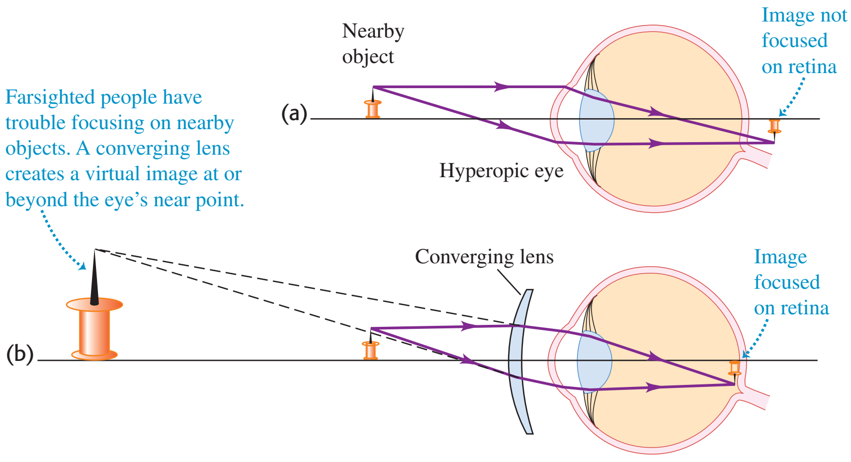
Several common defects of vision result from incorrect distance relationships in the eye. A normal eye forms an image on the retina of an object at infinity when the eye is relaxed (Fig. 34.45a). In the *myopic* (nearsighted) eye, the eyeball is too long from front to back in comparison with the radius 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

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. The dashed blue curve indicates the required position of the retina.



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.



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 refractive index 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.

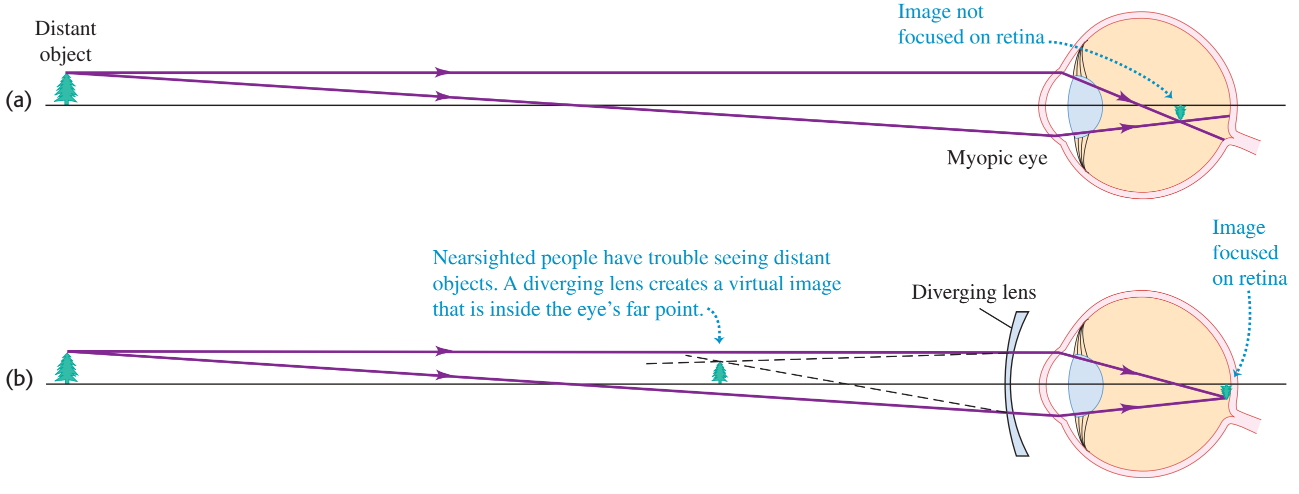
converging (positive) lens, as shown in Fig. 34.46. In effect the lens moves the object farther away from the eye to a point where a sharp retinal image can be formed. Similarly, correcting the myopic eye involves using a diverging (negative) lens to move the image closer to the eye than the actual object, as shown in Fig. 34.47.

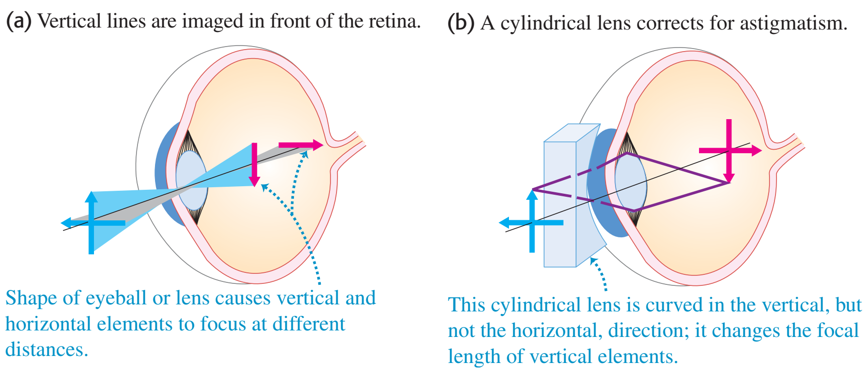
Astigmatism is a different type of defect in which the surface of the cornea is not spherical but rather more sharply curved in one plane than in another. As a result, horizontal lines may be imaged in a different plane from vertical lines (Fig. 34.48a). Astigmatism may make it impossible, for example, to focus clearly on both the horizontal and vertical bars of a window at the same time.

Astigmatism can be corrected by use of a lens with a *cylindrical* surface. For example, suppose the curvature of the cornea in a horizontal plane is correct to 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

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.





34.48 One type of astigmatism and how it is corrected.

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.

SOLUTION

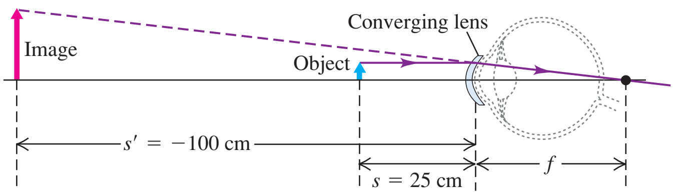
IDENTIFY and SET UP: Figure 34.49 shows the situation. We want the lens to form a virtual image of the object at the near point of the eye, 100 cm from it. The contact lens (which we treat as having negligible thickness) is at the surface of the cornea, so the object distance is $s = 25$ cm. The virtual image is on the incoming side of the contact lens, so the image distance is $s' = -100$ cm. We use Eq. (34.16) to determine the required focal length f of the contact lens; the corresponding power is $1/f$.

EXECUTE: From Eq. (34.16),

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

$$f = +33\text{ cm}$$

34.49 Using a contact lens to correct for farsightedness. For clarity, the eye and contact lens are shown much larger than the scale of the figure; the 2.5-cm diameter of the eye is actually much smaller than the focal length f of the contact lens.



We need a converging lens with focal length $f = 33$ cm and power $1/(0.33\text{ m}) = +3.0$ diopters.

EVALUATE: In this example we used a contact lens to correct hyperopia. Had we used eyeglasses, we would have had to account for the separation between the eye and the eyeglass lens, and a somewhat different power would have been required (see Example 34.14).

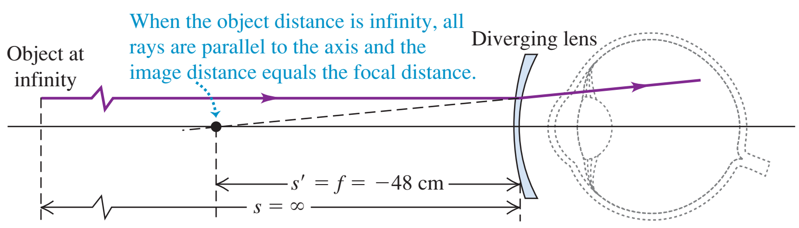
Example 34.14 Correcting for nearsightedness

The far point of a certain myopic eye is 50 cm in front of the eye. Find the focal length and power of the eyeglass lens that will permit the wearer to see clearly an object at infinity. Assume that the lens is worn 2 cm in front of the eye.

SOLUTION

IDENTIFY and SET UP: Figure 34.50 shows the situation. The far point of a myopic eye is nearer than infinity. To see clearly objects

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

beyond the far point, we need a lens that forms a virtual image of such objects no farther from the eye than the far point. Assume that the virtual image of the object at infinity is formed at the far point, 50 cm in front of the eye (48 cm in front of the eyeglass lens). Then when the object distance is $s = \infty$, we want the image distance to be $s' = -48$ cm. As in Example 34.13, we use the values of s and s' to calculate the required focal length.

EXECUTE: From Eq. (34.16),

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} = \frac{1}{\infty} + \frac{1}{-48\text{ cm}}$$
$$f = -48\text{ cm}$$

We need a *diverging* lens with focal length $f = -48$ cm and power $1/(-0.48\text{ m}) = -2.1$ diopters.

EVALUATE: If a *contact* lens were used to correct this myopia, we would need $f = -50$ cm and a power of -2.0 diopters. Can you see why?

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.

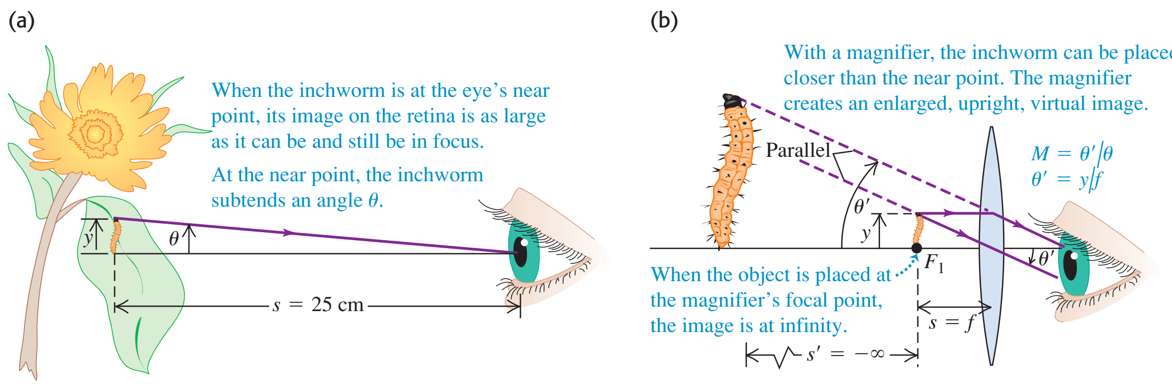
34.7 The Magnifier

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 the angular size of an object is greatest (that is, it subtends the largest possible viewing angle) when it is placed at the near point. In the following discussion we will assume an average viewer for whom the near point is 25 cm from the eye.

A converging lens can be used to form a virtual image that is larger and farther from the eye than the object itself, as shown in Fig. 34.51b. Then the object can be moved closer to the eye, and the angular size of the image may be substantially larger than the angular size of the object at 25 cm without the lens. A lens used in this way is called a **magnifier,** otherwise known as a *magnifying glass* or a *simple magnifier.* The virtual image is most comfortable to view when it is placed at infinity, so that the ciliary muscle of the eye is relaxed; this means that the object is placed at the focal point F_1 of the magnifier. In the following discussion we assume that this is done.

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.



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 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 the *angular* magnification M with the *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! (That's why we didn't attempt to draw an infinitely large inchworm in Fig. 34.51b.) When dealing with a magnifier, M is useful but m is not.

To find the value of M, we first assume that the angles are small enough that each angle (in radians) is equal to its sine and its tangent. Using Fig. 34.451a 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

$$M = \frac{\theta'}{\theta} = \frac{y/f}{y/25\text{ cm}} = \frac{25\text{ cm}}{f} \qquad \begin{array}{l}\text{(angular magnification}\\ \text{for a simple magnifier)}\end{array} \tag{34.22}$$

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×. When greater magnification than this is needed, we usually use a compound microscope, discussed in the next section.

Test Your Understanding of Section 34.7 You are examining a gem using a magnifier. If you change to a different magnifier with twice the focal length of the first one, (i) you will have to hold the object at twice the distance and the angular magnification will be twice as great; (ii) you will have to hold the object at twice the distance and the angular magnification will be $\frac{1}{2}$ as great; (iii) you will have to hold the object at $\frac{1}{2}$ the distance and the angular magnification will be twice as great; (iv) you will have to hold the object at $\frac{1}{2}$ the distance and the angular magnification will be $\frac{1}{2}$ as great.

34.8 Microscopes and Telescopes

MasteringPHYSICS

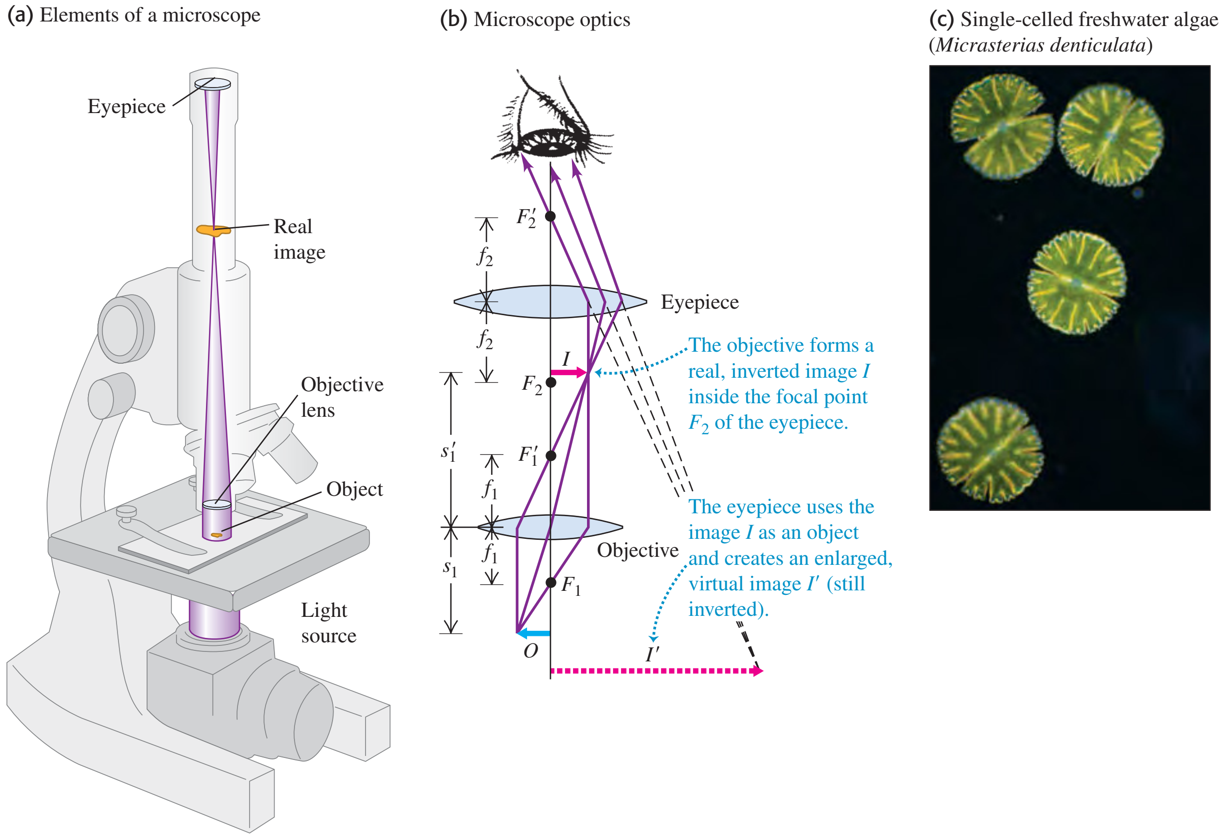
ActivPhysics 15.12: Two-Lens Optical Systems

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

When we need greater magnification than we can get with a simple magnifier, the instrument that we usually use is the **microscope,** sometimes called a *compound*

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.



microscope. The essential elements of a microscope are shown in Fig. 34.52a. 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 element. We used this principle in Section 34.4 when we derived the thin-lens 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_1' of a second converging lens called the **eyepiece** or *ocular*. (The reason the image should lie just *inside* F_1' is left for you to discover; see Problem 34.108.) The eyepiece acts as a simple magnifier, as discussed in Section 34.7, and forms a final virtual image I' of I. The position of I' may be anywhere between the near and far points of the eye. Both the objective and the eyepiece of an actual microscope are highly corrected compound lenses with several optical elements, but for simplicity we show them here as simple thin lenses.

As for a simple magnifier, what matters when viewing through a microscope is the *angular* magnification M. The overall angular magnification of the compound microscope is the product of two factors. The first factor is the *lateral* magnification m_1 of the objective, which determines the linear size of the real image I; the second factor is the *angular* magnification M_2 of the eyepiece, which relates the angular size of the virtual image seen through the eyepiece to the angular size that the real image I would have if you viewed it *without* the eyepiece. The first of these factors is given by

$$m_1 = -\frac{s_1'}{s_1} \tag{34.23}$$

where s_1 and s_1' are the object and image distances, respectively, for the objective lens. Ordinarily, the object is very close to the focal point, and the resulting image distance s_1' is very great in comparison to the focal length f_1 of the objective lens. Thus s_1 is approximately equal to f_1, and we can write $m_1 = -s_1'/f_1$.

The real image I is close to the focal point F_1' 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:

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

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 for microscope components 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 take a photograph using a microscope (called a *photomicrograph* or *micrograph*), the eyepiece is removed and a camera placed so that the real image I falls on the camera's CCD array 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).

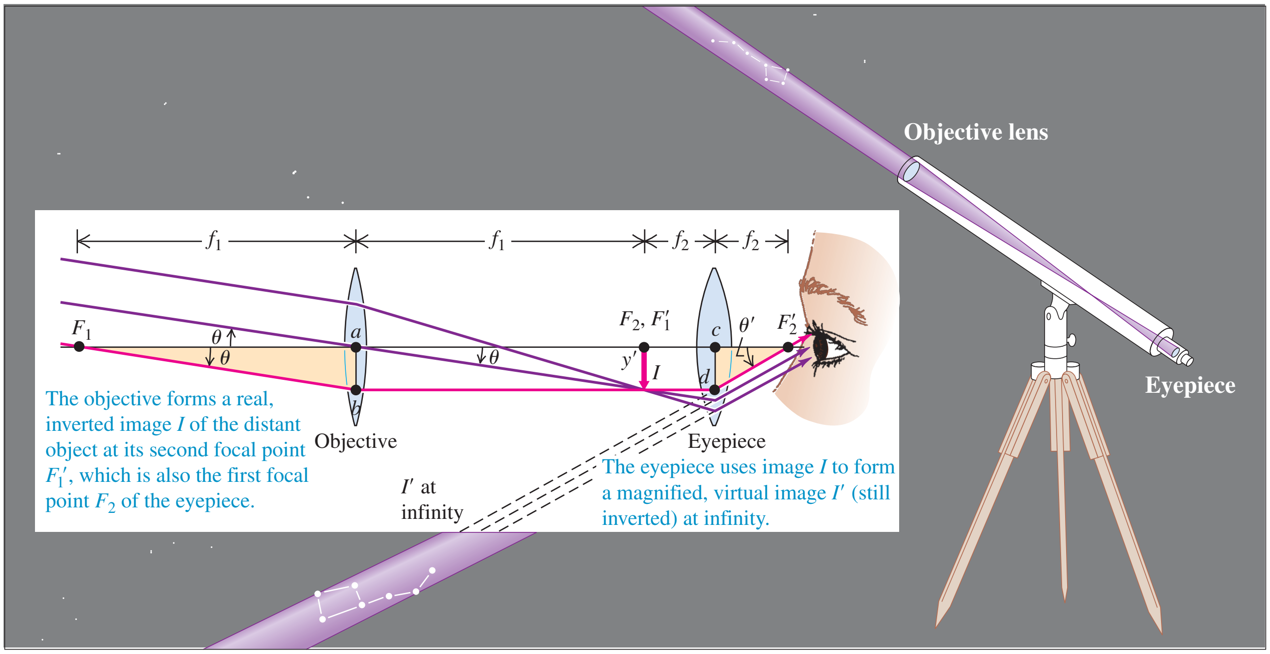
Telescopes

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 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

34.53 Optical system of an astronomical refracting telescope.



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 $F_2'cd$,

$$\theta = \frac{-y'}{f_1} \qquad \theta' = \frac{y'}{f_2}$$

and the angular magnification M is

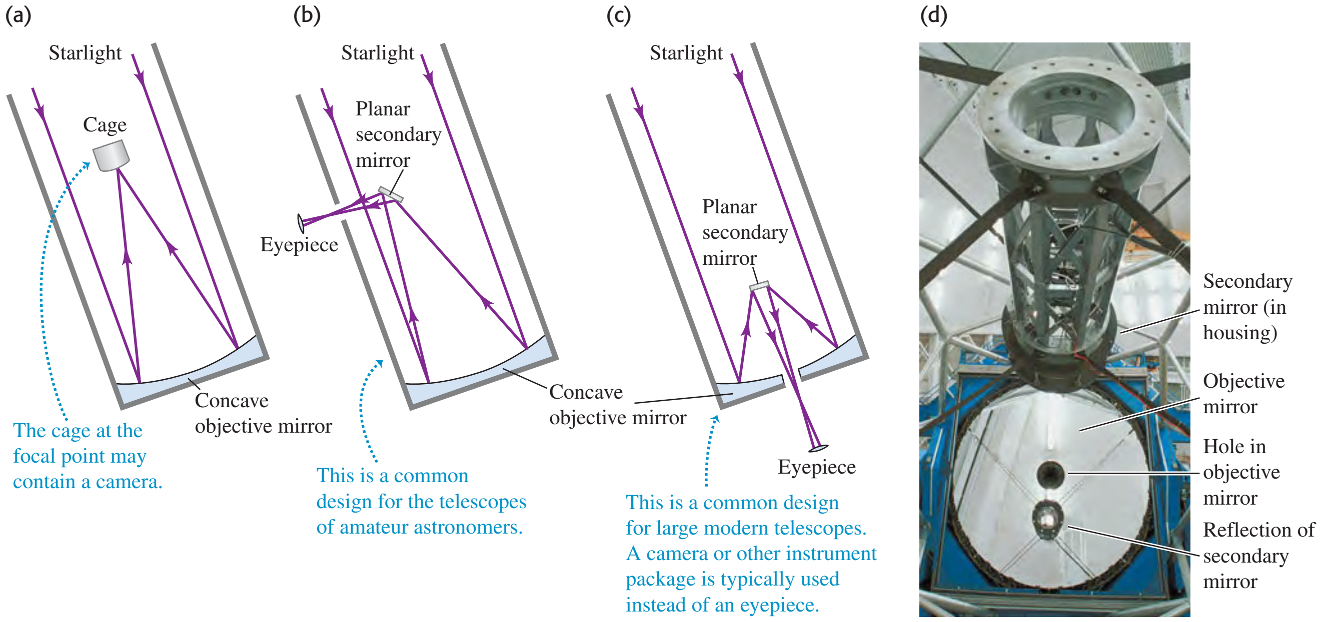
$$M = \frac{\theta'}{\theta} = -\frac{y'/f_2}{y'/f_1} = -\frac{f_1}{f_2} \qquad \text{(angular magnification for a telescope)} \tag{34.25}$$

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.

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

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.



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 parabolic rather than spherical. 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 telescopes in the world, the Keck telescopes atop Mauna Kea in Hawaii, each have an objective mirror of overall diameter 10 m made up of 36 separate hexagonal reflecting elements.

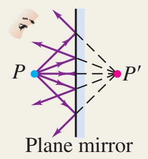
One challenge in designing reflecting telescopes is that the image is formed in front of the objective mirror, in a region traversed by incoming rays. Isaac Newton devised one solution to this problem. A flat secondary mirror oriented at 45° to the optic axis causes the image to be formed in a hole on 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 a CCD array or photographic film 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?

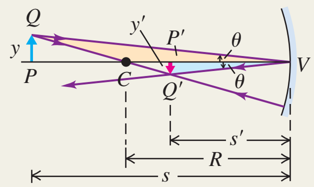
CHAPTER 34 SUMMARY

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.

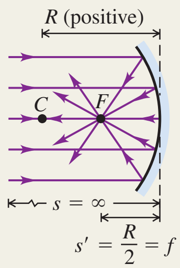


Lateral magnification: The lateral magnification m in any reflecting or refracting situation is defined as the ratio of image height y' to object height y. When m is positive, the image is erect; when m is negative, the image is inverted.

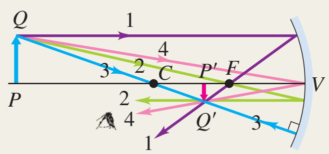
$$m = \frac{y'}{y} \tag{34.2}$$



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.)



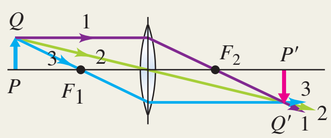
	Plane Mirror	Spherical Mirror	Plane Refracting Surface	Spherical Refracting Surface
Object and image distances	$\frac{1}{s} + \frac{1}{s'} = 0$	$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} = \frac{1}{f}$	$\frac{n_a}{s} + \frac{n_b}{s'} = 0$	$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R}$
Lateral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

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

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

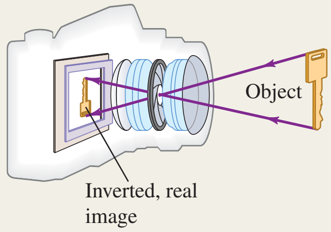


Sign rules: The following sign rules are used with all plane and spherical reflecting and refracting surfaces.

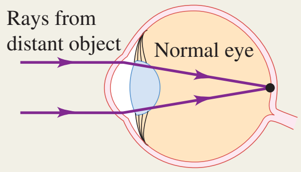
- $s > 0$ when the object is on the incoming side of the surface (a real object); $s < 0$ otherwise.
- $s' > 0$ when the image is on the outgoing side of the surface (a real image); $s' < 0$ otherwise.
- $R > 0$ when the center of curvature is on the outgoing side of the surface; $R < 0$ otherwise.
-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

Cameras: A camera forms a real, inverted, reduced image of the object being photographed on a light-sensitive surface. The amount of light striking this surface is controlled by the shutter speed and the aperture. The intensity of this light is inversely proportional to the square of the f-number of the lens. (See Example 34.12.)

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

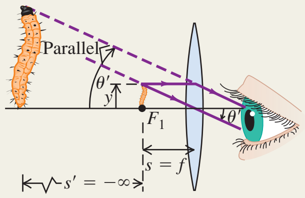


The eye: In the eye, refraction at the surface of the cornea forms a real image on the retina. Adjustment for various object distances is made by squeezing the lens, thereby making it bulge and decreasing its focal length. A nearsighted eye is too long for its lens; a farsighted eye is too short. The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

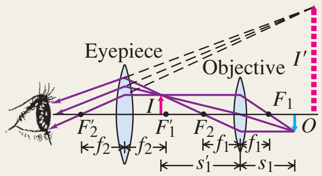


The simple magnifier: The simple magnifier creates a virtual image whose angular size θ' is larger than the angular size θ of the object itself at a distance of 25 cm, the nominal closest distance for comfortable viewing. The angular magnification M of a simple magnifier is the ratio of the angular size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

$$M = \frac{\theta'}{\theta} = \frac{25\text{ cm}}{f} \tag{34.22}$$



Microscopes and telescopes: In a compound microscope, the objective lens forms a first image in the barrel of the instrument, and the eyepiece forms a final virtual image, often at infinity, of the first image. The telescope operates on the same principle, but the object is far away. In a reflecting telescope, the objective lens is replaced by a concave mirror, which eliminates chromatic aberrations.



BRIDGING PROBLEM Image Formation by a Wine Goblet

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

See MasteringPhysics® study area for a Video Tutor solution.

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.
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

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q34.1 A spherical mirror is cut in half horizontally. Will an image be formed by the bottom half of the mirror? If so, where will the image be formed?

Q34.2 For the situation shown in Fig. 34.3, is the image distance s' positive or negative? Is the image real or virtual? Explain your answers.

Q34.3 The laws of optics also apply to electromagnetic waves invisible to the eye. A satellite TV dish is used to detect radio waves coming from orbiting satellites. Why is a curved reflecting surface (a "dish") used? The dish is always concave, never convex; why? The actual radio receiver is placed on an arm and suspended in front of the dish. How far in front of the dish should it be placed?

Q34.4 Explain why the focal length of a *plane* mirror is infinite, and explain what it means for the focal point to be at infinity.

Q34.5 If a spherical mirror is immersed in water, does its focal length change? Explain.

Q34.6 For what range of object positions does a concave spherical mirror form a real image? What about a convex spherical mirror?

Q34.7 When a room has mirrors on two opposite walls, an infinite series of reflections can be seen. Discuss this phenomenon in terms of images. Why do the distant images appear fainter?

Q34.8 For a spherical mirror, if $s = f$, then $s' = \infty$, and the lateral magnification m is infinite. Does this make sense? If so, what does it mean?

Q34.9 You may have noticed a small convex mirror next to your bank's ATM. Why is this mirror convex, as opposed to flat or concave? What considerations determine its radius of curvature?

Q34.10 A student claims that she can start a fire on a sunny day using just the sun's rays and a concave mirror. How is this done? Is the concept of image relevant? Can she do the same thing with a convex mirror? Explain.

Q34.11 A person looks at his reflection in the concave side of a shiny spoon. Is it right side up or inverted? Does it matter how far his face is from the spoon? What if he looks in the convex side? (Try this yourself!)

Q34.12 In Example 34.4 (Section 34.2), there appears to be an ambiguity for the case $s = 10$ cm as to whether s' is $+\infty$ or $-\infty$ and whether the image is erect or inverted. How is this resolved? Or is it?

Q34.13 Suppose that in the situation of Example 34.7 of Section 34.3 (see Fig. 34.26) a vertical arrow 2.00 m tall is painted on the side of the pool beneath the water line. According to the calculations in the example, this arrow would appear to the person shown in Fig. 34.26 to be 1.50 m long. But the discussion following Eq. (34.13) states that the magnification for a plane refracting surface is $m = 1$, which suggests that the arrow would appear to the person to be 2.00 m long. How can you resolve this apparent contradiction?

Q34.14 The bottom of the passenger-side mirror on your car notes, "Objects in mirror are closer than they appear." Is this true? Why?

Q34.15 How could you very quickly make an approximate measurement of the focal length of a converging lens? Could the same method be applied if you wished to use a diverging lens? Explain.

Q34.16 The focal length of a simple lens depends on the color (wavelength) of light passing through it. Why? Is it possible for a lens to have a positive focal length for some colors and negative for others? Explain.

Q34.17 When a converging lens is immersed in water, does its focal length increase or decrease in comparison with the value in air? Explain.

Q34.18 A spherical air bubble in water can function as a lens. Is it a converging or diverging lens? How is its focal length related to its radius?

Q34.19 Can an image formed by one reflecting or refracting surface serve as an object for a second reflection or refraction? Does it matter whether the first image is real or virtual? Explain.

Q34.20 If a piece of photographic film is placed at the location of a real image, the film will record the image. Can this be done with a virtual image? How might one record a virtual image?

Q34.21 According to the discussion in Section 34.2, light rays are reversible. Are the formulas in the table in this chapter's Summary still valid if object and image are interchanged? What does reversibility imply with respect to the *forms* of the various formulas?

Q34.22 You've entered a survival contest that will include building a crude telescope. You are given a large box of lenses. Which two lenses do you pick? How do you quickly identify them?

Q34.23 **BIO** You can't see clearly underwater with the naked eye, but you *can* if you wear a face mask or goggles (with air between your eyes and the mask or goggles). Why is there a difference? Could you instead wear eyeglasses (with water between your eyes and the eyeglasses) in order to see underwater? If so, should the lenses be converging or diverging? Explain.

Q34.24 You take a lens and mask it so that light can pass through only the bottom half of the lens. How does the image formed by the masked lens compare to the image formed before masking?

EXERCISES

Section 34.1 Reflection and Refraction at a Plane Surface

34.1 • A candle 4.85 cm tall is 39.2 cm to the left of a plane mirror. Where is the image formed by the mirror, and what is the height of this image?

34.2 • The image of a tree just covers the length of a plane mirror 4.00 cm tall when the mirror is held 35.0 cm from the eye. The tree is 28.0 m from the mirror. What is its height?

34.3 • A pencil that is 9.0 cm long is held perpendicular to the surface of a plane mirror with the tip of the pencil lead 12.0 cm from the mirror surface and the end of the eraser 21.0 cm from the mirror surface. What is the length of the image of the pencil that is formed by the mirror? Which end of the image is closer to the mirror surface: the tip of the lead or the end of the eraser?

Section 34.2 Reflection at a Spherical Surface

34.4 • A concave mirror has a radius of curvature of 34.0 cm. (a) What is its focal length? (b) If the mirror is immersed in water (refractive index 1.33), what is its focal length?

34.5 • An object 0.600 cm tall is placed 16.5 cm to the left of the vertex of a concave spherical mirror having a radius of curvature of 22.0 cm. (a) Draw a principal-ray diagram showing the formation of the image. (b) Determine the position, size, orientation, and nature (real or virtual) of the image.

34.6 • Repeat Exercise 34.5 for the case in which the mirror is convex.

34.7 •• The diameter of Mars is 6794 km, and its minimum distance from the earth is 5.58×10^7 km. When Mars is at this distance, find the diameter of the image of Mars formed by a spherical, concave telescope mirror with a focal length of 1.75 m.

34.8 •• An object is 24.0 cm from the center of a silvered spherical glass Christmas tree ornament 6.00 cm in diameter. What are the position and magnification of its image?

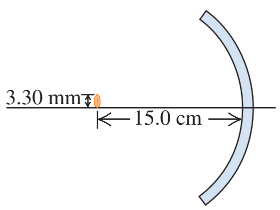
34.9 • A coin is placed next to the convex side of a thin spherical glass shell having a radius of curvature of 18.0 cm. Reflection from the surface of the shell forms an image of the 1.5-cm-tall coin that is 6.00 cm behind the glass shell. Where is the coin located? Determine the size, orientation, and nature (real or virtual) of the image.

34.10 • You hold a spherical salad bowl 90 cm in front of your face with the bottom of the bowl facing you. The salad bowl is made of polished metal with a 35-cm radius of curvature. (a) Where is the image of your 2.0-cm-tall nose located? (b) What are the image's size, orientation, and nature (real or virtual)?

34.11 • (a) Show that Eq. (34.6) can be written as $s' = sf/(s - f)$ and hence the lateral magnification given by Eq. (34.7) can be expressed as $m = f/(f - s)$. (b) A concave spherical mirror has focal length $f = +14.0$ cm. What is the nonzero distance of the object from the mirror vertex if the image has the same height as the object? In this case, is the image erect or inverted? (c) A convex spherical mirror has $f = -8.00$ cm. What is the nonzero distance of the object from the mirror vertex if the height of the image is one-half the height of the object?

34.12 • The thin glass shell shown in Fig. E34.12 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 **E34.12**



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 • 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.

Section 34.3 Refraction at a Spherical Surface

34.15 • 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?

34.16 •• A tank whose bottom is a mirror is filled with water to a depth of 20.0 cm. A small fish floats motionless 7.0 cm under the surface of the water. (a) What is the apparent depth of the fish when viewed at normal incidence? (b) What is the apparent depth of the image of the fish when viewed at normal incidence?

34.17 • A 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?

34.18 • A person is lying on a diving board 3.00 m above the surface of the water in a swimming pool. The person looks at a penny that is on the bottom of the pool directly below her. The penny appears to the person to be a distance of 8.00 m from her. What is the depth of the water at this point?

34.19 •• **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?

34.20 • 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.

34.21 •• The glass rod of Exercise 34.20 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?

34.22 •• 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?

34.23 •• Repeat Exercise 34.22 for the case in which the end of the rod is ground to a *concave* hemispherical surface with radius 4.00 cm.

34.24 •• The glass rod of Exercise 34.23 is immersed in a liquid. An object 14.0 cm from the vertex of the left end of the rod and on its axis is imaged at a point 9.00 cm from the vertex inside the liquid. What is the index of refraction of the liquid?

Section 34.4 Thin Lenses

34.25 • 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.

34.26 • 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.

34.27 • 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?

34.28 • A converging lens with a focal length of 90.0 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?

34.29 •• 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?

34.30 • 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?

34.31 •• A double-convex thin lens has surfaces with equal radii of curvature of magnitude 2.50 cm. Looking through 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?

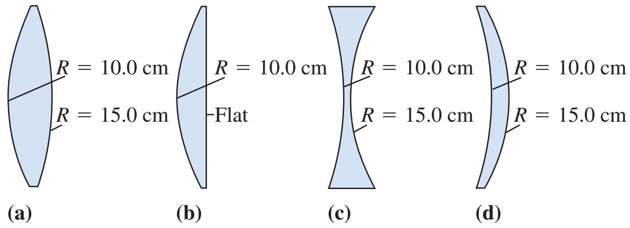
34.32 • BIO **The Lens of the Eye.** The crystalline lens of the human eye is a double-convex lens made of material having an index of refraction of 1.44 (although this varies). Its focal length in air is about 8.0 mm, which also varies. We shall assume that the radii of curvature of its two surfaces have the same magnitude. (a) Find the radii of curvature of this lens. (b) If an object 16 cm tall is placed 30.0 cm from the eye lens, where would the lens focus it and how tall would the image be? Is this image real or virtual? Is it erect or inverted? (*Note:* The results obtained here are not strictly accurate because the lens is embedded in fluids having refractive indexes different from that of air.)

34.33 • BIO **The Cornea As a Simple Lens.** The cornea behaves as a thin lens of focal length approximately 1.8 cm, although this varies a bit. The material of which it is made has an index of refraction of 1.38, and its front surface is convex, with a radius of curvature of 5.0 mm. (a) If this focal length is in air, what is the radius of curvature of the back side of the cornea? (b) The closest distance at which a typical person can focus on an object (called the near point) is about 25 cm, although this varies considerably with age. Where would the cornea focus the image of an 8.0-mm-tall object at the near point? (c) What is the height of the image in part (b)? Is this image real or virtual? Is it erect or inverted? (*Note:* The results obtained here are not strictly accurate because, on one side, the cornea has a fluid with a refractive index different from that of air.)

34.34 • A converging lens with a focal length of 7.00 cm forms an image of a 4.00-mm-tall real object that is to the left of the lens. The image is 1.30 cm tall and erect. Where are the object and image located? Is the image real or virtual?

34.35 • For each thin lens shown in Fig. E34.35, calculate the location of the image of an object that is 18.0 cm to the left of the lens. The lens material has a refractive index of 1.50, and the radii of curvature shown are only the magnitudes.

Figure **E34.35**



34.36 • 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.

34.37 • Repeat Exercise 34.36 for the case in which the lens is diverging, with a focal length of −48.0 cm.

34.38 • 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.

34.39 •• **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.

34.40 •• **Combination of Lenses II.** Repeat Problem 34.39 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 Problem 34.39.

34.41 •• **Combination of Lenses III.** Two thin lenses with a focal length of magnitude 12.0 cm, the first diverging and the second converging, are located 9.00 cm apart. An object 2.50 mm tall is placed 20.0 cm to the left of the first (diverging) lens. (a) How far from this first lens is the final image formed? (b) Is the final image real or virtual? (c) What is the height of the final image? Is it erect or inverted? (*Hint:* See the preceding two problems.)

Section 34.5 Cameras

34.42 • 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?

34.43 •• A camera lens has a focal length of 200 mm. How far from the lens should the subject for the photo be if the lens is 20.4 cm from the film?

34.44 • When a camera is focused, the lens is moved away from or toward the film. If you take a picture of your friend, who is standing 3.90 m from the lens, using a camera with a lens with a 85-mm focal length, how far from the film is the lens? Will the whole image of your friend, who is 175 cm tall, fit on film that is 24 × 36 mm?

34.45 • Figure 34.41 shows photographs of the same scene taken with the same camera with lenses of different focal length. If the object is 200 m from the lens, what is the magnitude of the lateral magnification for a lens of focal length (a) 28 mm; (b) 105 mm; (c) 300 mm?

34.46 • A photographer takes a photograph of a Boeing 747 airliner (length 70.7 m) when it is flying directly overhead at an altitude of 9.50 km. The lens has a focal length of 5.00 m. How long is the image of the airliner on the film?

34.47 • **Choosing a Camera Lens.** The picture size on ordinary 35-mm camera film is 24 mm $\times$ 36 mm. Focal lengths of lenses available for 35-mm cameras typically include 28, 35, 50 (the "normal" lens), 85, 100, 135, 200, and 300 mm, among others. Which of these lenses should be used to photograph the following objects, assuming that the object is to fill most of the picture area? (a) a building 240 m tall and 160 m wide at a distance of 600 m, and (b) a mobile home 9.6 m in length at a distance of 40.0 m.

34.48 • **Zoom Lens.** Consider the simple model of the zoom lens shown in Fig. 34.43a. The converging lens has focal length $f_1 = 12$ cm, and the diverging lens has focal length $f_2 = -12$ cm. The lenses are separated by 4 cm as shown in Fig. 34.43a. (a) For a distant object, where is the image of the converging lens? (b) The image of the converging lens serves as the object for the diverging lens. What is the object distance for the diverging lens? (c) Where is the final image? Compare your answer to Fig. 34.43a. (d) Repeat parts (a), (b), and (c) for the situation shown in Fig. 34.43b, in which the lenses are separated by 8 cm.

34.49 •• A camera lens has a focal length of 180.0 mm and an aperture diameter of 16.36 mm. (a) What is the f-number of the lens? (b) If the correct exposure of a certain scene is $\frac{1}{30}$ s at $f/11$, what is the correct exposure at $f/2.8$?

34.50 •• Recall that the intensity of light reaching film in a camera is proportional to the effective area of the lens. Camera A has a lens with an aperture diameter of 8.00 mm. It photographs an object using the correct exposure time of $\frac{1}{30}$ s. What exposure time should be used with camera B in photographing the same object with the same film if this camera has a lens with an aperture diameter of 23.1 mm?

34.51 • **Photography.** A 35-mm camera has a standard lens with focal length 50 mm and can focus on objects between 45 cm and infinity. (a) Is the lens for such a camera a concave or a convex lens? (b) The camera is focused by rotating the lens, which moves it on the camera body and changes its distance from the film. In what range of distances between the lens and the film plane must the lens move to focus properly over the 45 cm to infinity range?

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 refraction 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 Problem 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 •• BIO If the person in Problem 34.56 chooses ordinary glasses over contact lenses, what power lens (in diopters) does she need to correct her vision if the lenses are 2.0 cm in front of the eye?

Section 34.7 The Magnifier

34.58 •• 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 an insect 2.00 mm in length through a magnifier. 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 0.025 radian?

Section 34.8 Microscopes and Telescopes

34.61 • A certain microscope is provided with objectives that have focal lengths of 16 mm, 4 mm, and 1.9 mm and with eyepieces that have angular magnifications of $5\times$ and $10\times$. Each objective forms an image 120 mm beyond its second focal point. Determine (a) the largest overall angular magnification obtainable and (b) the smallest overall angular magnification obtainable.

34.62 •• **Resolution of a Microscope.** The image formed by a microscope objective with a focal length of 5.00 mm is 160 mm from its second focal point. The eyepiece has a focal length of 26.0 mm. (a) What is the angular magnification of the microscope? (b) The unaided eye can distinguish two points at its near point as separate if they are about 0.10 mm apart. What is the minimum separation between two points that can be observed (or resolved) through this microscope?

34.63 •• The focal length of the eyepiece of a certain microscope is 18.0 mm. The focal length of the objective is 8.00 mm. The distance between objective and eyepiece is 19.7 cm. The final image formed by the eyepiece is at infinity. Treat all lenses as thin. (a) What is the distance from the objective to the object being viewed? (b) What is the magnitude of the linear magnification produced by the objective? (c) What is the overall angular magnification of the microscope?

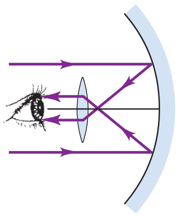
34.64 •• The eyepiece of a refracting telescope (see Fig. 34.53) has a focal length of 9.00 cm. The distance between objective and eyepiece is 1.80 m, and the final image is at infinity. What is the angular magnification of the telescope?

34.65 •• 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?

34.66 •• Saturn is viewed through the Lick Observatory refracting telescope (objective focal length 18 m). If the diameter of the image of Saturn produced by the objective is 1.7 mm, what angle does Saturn subtend from when viewed from earth?

34.67 •• A reflecting telescope (Fig. E34.67) 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 **E34.67**



PROBLEMS

34.68 • 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?

34.69 • If you run away from a plane mirror at 3.60 m/s, at what speed does your image move away from you?

34.70 • An object is placed between two plane mirrors arranged at right angles to each other at a distance d_1 from the surface of one mirror and a distance d_2 from the other. (a) How many images are formed? Show the location of the images in a diagram. (b) Draw the paths of rays from the object to the eye of an observer.

34.71 •• What is the size of the smallest vertical plane mirror in which a woman of height h can see her full-length image?

34.72 • 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 2.25 times the size of the object. How far should the mirror be from the wall? What should its radius of curvature be?

34.73 •• 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?

34.74 •• **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?

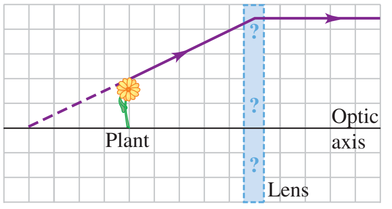
34.75 • Suppose the lamp filament shown in Example 34.1 (Section 34.2) is moved to a position 8.0 cm in front of the mirror. (a) Where is the image located now? Is it real or virtual? (b) What is the height of the image? Is it erect or inverted? (c) In Example 34.1, the filament is 10.0 cm in front of the mirror, and an image of the filament is formed on a wall 3.00 m from the mirror. If the filament is 8.0 cm from the mirror, can a wall be placed so that an image is formed on it? If so, where should the wall be placed? If not, why not?

34.76 •• A layer of benzene $(n = 1.50)$ 4.20 cm deep floats on water $(n = 1.33)$ that is 6.50 cm deep. What is the apparent distance from the upper benzene surface to the bottom of the water layer when it is viewed at normal incidence?

34.77 •• **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?

34.78 •• Figure P34.78 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 **P34.78**



34.79 •• **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 $\times$ 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 magnification of this camera?

34.80 ••• A microscope is focused on the upper surface of a glass plate. A second plate is then placed over the first. To focus on the bottom surface of the second plate, the microscope must be raised 0.780 mm. To focus on the upper surface, it must be raised another 2.50 mm. Find the index of refraction of the second plate.

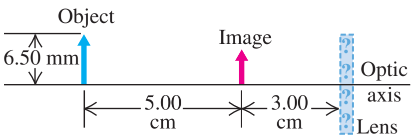
34.81 •• What should be the index of refraction of a transparent sphere in order for paraxial rays from an infinitely distant object to be brought to a focus at the vertex of the surface opposite the point of incidence?

34.82 •• **A Glass Rod.** Both ends of a glass rod with index of refraction 1.60 are ground and polished to convex hemispherical surfaces. The radius of curvature at the left end is 6.00 cm, and the radius of curvature at the right end is 12.0 cm. The length of the rod between vertices is 40.0 cm. The object for the surface at the left end is an arrow that lies 23.0 cm to the left of the vertex of this surface. The arrow is 1.50 mm tall and at right angles to the axis. (a) What constitutes the object for the surface at the right end of the rod? (b) What is the object distance for this surface? (c) Is the object for this surface real or virtual? (d) What is the position of the final image? (e) Is the final image real or virtual? Is it erect or inverted with respect to the original object? (f) What is the height of the final image?

34.83 •• The rod in Problem 34.82 is shortened to a distance of 25.0 cm between its vertices; the curvatures of its ends remain the same. As in Problem 34.82, the object for the surface at the left end is an arrow that lies 23.0 cm to the left of the vertex of this surface. The arrow is 1.50 mm tall and at right angles to the axis. (a) What is the object distance for the surface at the right end of the rod? (b) Is the object for this surface real or virtual? (c) What is the position of the final image? (d) Is the final image real or virtual? Is it erect or inverted with respect to the original object? (e) What is the height of the final image?

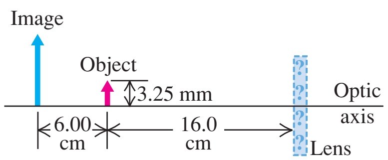
34.84 • Figure P34.84 shows an object and its image formed by a thin lens. (a) What is the focal length of the lens, and what type of lens (converging or diverging) is it? (b) What is the height of the image? Is it real or virtual?

Figure **P34.84**



34.85 • Figure P34.85 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 **P34.85**



34.86 ••• 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 viewed from the flat end of the rod, the apparent depth of the object is 9.50 cm from the flat end. What is its apparent depth when viewed from the curved end?

34.87 • 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?

34.88 ••• A transparent rod 50.0 cm long and with a refractive index of 1.60 is cut flat at the right end and rounded to a hemispherical surface with a 15.0-cm radius at the left end. An object is placed on the axis of the rod 12.0 cm to the left of the vertex of the hemispherical end. (a) What is the position of the final image? (b) What is its magnification?

34.89 ••• A glass rod with a refractive index of 1.55 is ground and polished at both ends to hemispherical surfaces with radii of 6.00 cm. When an object is placed on the axis of the rod, 25.0 cm to the left of the left-hand end, the final image is formed 65.0 cm to the right of the right-hand end. What is the length of the rod measured between the vertices of the two hemispherical surfaces?

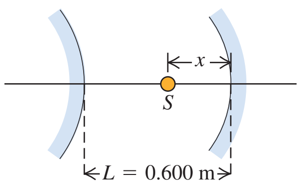
34.90 • The radii of curvature of the surfaces of a thin converging meniscus lens are $R_1 = +12.0$ cm and $R_2 = +28.0$ cm. The index of refraction is 1.60. (a) Compute the position and size of the image of an object in the form of an arrow 5.00 mm tall, perpendicular to the lens axis, 45.0 cm to the 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.

34.91 • An object to the left of a lens is imaged by the lens on a screen 30.0 cm to the right of the lens. When the lens is moved 4.00 cm to the right, the screen must be moved 4.00 cm to the left to refocus the image. Determine the focal length of the lens.

34.92 • An object is placed 18.0 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?

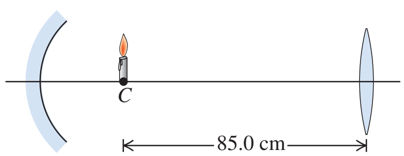
34.93 •• A convex mirror and a concave mirror are placed on the same optic axis, separated by a distance $L = 0.600$ m. The radius of curvature of each mirror has a magnitude of 0.360 m. A light source is located a distance x from the concave mirror, as shown in Fig. P34.93. (a) What distance x will result in the rays from the source returning to the source after reflecting first from the convex mirror and then from the concave mirror? (b) Repeat part (a), but now let the rays reflect first from the concave mirror and then from the convex one.

Figure **P34.93**



34.94 •• As shown in Fig. P34.94 the candle is at the center of curvature of the concave mirror, whose focal length is 10.0 cm. The converging lens has a focal length of 32.0 cm and is 85.0 cm to the right of the candle. The candle is viewed looking through the lens from the right. The lens forms two images of the candle. The first is formed by light passing directly through the lens. The second image is formed from the light that goes from the candle to the mirror, is reflected, and then passes through the lens. (a) For each of these two images, draw a principal-ray diagram that locates the image. (b) For each image, answer the following questions: (i) Where is the image? (ii) Is the image real or virtual? (iii) Is the image erect or inverted with respect to the original object?

Figure **P34.94**



34.95 ••• One end of a long glass rod is ground to a convex hemispherical shape. This glass has an index of refraction of 1.55. When a small leaf is placed 20.0 cm in front of the center of the hemisphere along the optic axis, an image is formed inside the glass 9.12 cm from the spherical surface. Where would the image be formed if the glass were now immersed in water (refractive index 1.33) but nothing else were changed?

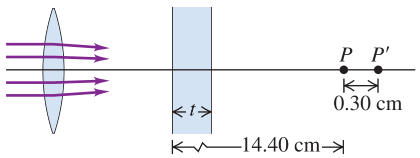
34.96 •• Two Lenses in Contact. (a) Prove that when two thin lenses with focal lengths f_1 and f_2 are placed *in contact,* the focal length f of the combination is given by the relationship

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$$

(b) A converging meniscus lens (see Fig. 34.32a) has an index of refraction of 1.55 and radii of curvature for its surfaces of magnitudes 4.50 cm and 9.00 cm. The concave surface is placed upward and filled with carbon tetrachloride (CCl_4), which has $n = 1.46$. What is the focal length of the CCl_4–glass combination?

34.97 ••• Rays from a lens are converging toward a point image P located to the right of the lens. What thickness t of glass with index of refraction 1.60 must be interposed between the lens and P for the image to be formed at P', located 0.30 cm to the right of P? The locations of the piece of glass and of points P and P' are shown in Fig. P34.97.

Figure **P34.97**



34.98 •• A Lens in a Liquid. A lens obeys Snell's law, bending light rays at each surface an amount determined by the index of refraction of the lens and the index of the medium in which the lens is located. (a) Equation (34.19) assumes that the lens is surrounded by air. Consider instead a thin lens immersed in a liquid with refractive index n_{liq}. Prove that the focal length f' is then given by Eq. (34.19) with n replaced by n/n_{liq}. (b) A thin lens with index n has focal length f in vacuum. Use the result of part (a) to show that when this lens is immersed in a liquid of index n_{liq}, it will have a new focal length given by

$$f' = \left[\frac{n_{\text{liq}}(n-1)}{n - n_{\text{liq}}}\right] f$$

34.99 ••• 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?

34.100 ••• 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?

34.101 ••• A glass plate 3.50 cm thick, with an index of refraction of 1.55 and plane parallel faces, is held with its faces horizontal and its lower face 6.00 cm above a printed page. Find the position of the image of the page formed by rays making a small angle with the normal to the plate.

34.102 •• A symmetric, double-convex, thin lens made of glass with index of refraction 1.52 has a focal length in air of 40.0 cm. The lens is sealed into an opening in the left-hand end of a tank filled with water. At the right-hand end of the tank, opposite the lens, is a plane mirror 90.0 cm from the lens. The index of refraction of the water is $\frac{4}{3}$. (a) Find the position of the image formed by the lens–water–mirror system of a small object outside the tank on the lens axis and 70.0 cm to the left of the lens. (b) Is the image real or virtual? (c) Is it erect or inverted? (d) If the object has a height of 4.00 mm, what is the height of the image?

34.103 • You have a camera with a 35.0-mm-focal-length lens and 36.0-mm-wide film. You wish to take a picture of a 12.0-m-long sailboat but find that the image of the boat fills only $\frac{1}{4}$ of the width of the film. (a) How far are you from the boat? (b) How much closer must the boat be to you for its image to fill the width of the film?

34.104 •• 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.)

34.105 • 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.

34.106 •• 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 film be changed? To refocus on the more distant object, is the lens moved toward or away from the film?

34.107 •• The derivation of the expression for angular magnification, Eq. (34.22), assumed a near point of 25 cm. In fact, the near point changes with age as shown in Table 34.1. In order to achieve an angular magnification of 2.0×, what focal length should be used by a person of (a) age 10; (b) age 30; (c) age 60? (d) If the lens that gives $M = 2.0$ for a 10-year-old is used by a 60-year-old, what angular magnification will the older viewer obtain? (e) Does your answer in part (d) mean that older viewers are able to see more highly magnified images than younger viewers? Explain.

34.108 •• Angular Magnification. In deriving Eq. (34.22) for the angular magnification of a magnifier, we assumed that the object is placed at the focal point of the magnifier so that the virtual image is formed at infinity. Suppose instead that the object is placed so that the virtual image appears at an average viewer's near point of 25 cm, the closest point at which the viewer can bring an object into focus. (a) Where should the object be placed to achieve this? Give your answer in terms of the magnifier focal length f. (b) What angle θ' will an object of height y subtend at the position found in part (a)? (c) Find the angular magnification M with the object at the position found in part (a). The angle θ is the same as in Fig. 34.51a, since it refers to viewing the object *without* the magnifier. (d) For a convex lens with $f = +10.0$ cm, what is the value of M with the object at the position found in part (a)? How many times greater is M in this case than in the case where the image is formed at infinity? (e) In the description of a compound microscope in Section 34.8, it is stated that in a properly designed instrument, the real image formed by the objective lies *just inside* the first focal point F_1' of the eyepiece. What advantages are gained by having the image formed by the objective be just inside F_1', as opposed to precisely at F_1'? What happens if the image formed by the objective is *just outside* F_1'?

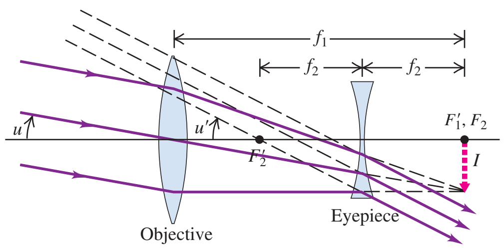
34.109 •• BIO In one form of cataract surgery the person's natural lens, which has become cloudy, is replaced by an artificial lens. The refracting properties of the replacement lens can be chosen so that the person's eye focuses on distant objects. But there is no accommodation, and glasses or contact 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?

34.110 •• 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?

34.111 •• BIO A person with a near point of 85 cm, but excellent distant vision, normally wears corrective glasses. But he loses them while traveling. Fortunately, he has his old pair as a spare. (a) If the lenses of the old pair have a power of +2.25 diopters, what is his near point (measured from his eye) when he is wearing the old glasses if they rest 2.0 cm in front of his eye? (b) What would his near point be if his old glasses were contact lenses instead?

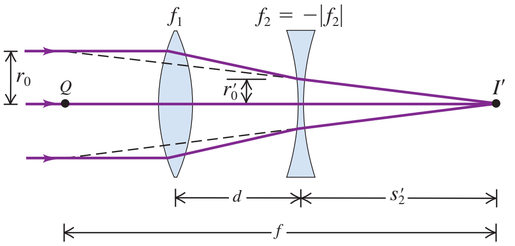
34.112 •• The Galilean Telescope. Figure P34.112 is a diagram of a *Galilean telescope,* or *opera glass,* with both the object and its final image at infinity. The image I serves as a virtual object for the eyepiece. The final image is virtual and erect. (a) Prove that the angular magnification is $M = -f_1/f_2$. (b) A Galilean telescope is to be constructed with the same objective lens as in Exercise 34.65. What focal length should the eyepiece have if this telescope is to have the same magnitude of angular magnification as the one in Exercise 34.65? (c) Compare the lengths of the telescopes.

Figure **P34.112**



34.113 ••• Focal Length of a Zoom Lens. Figure P34.113 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 **P34.113**



34.114 • A certain reflecting telescope, constructed as shown in Fig. E34.67, has a spherical mirror with a radius of curvature of 96.0 cm and an eyepiece with a focal length of 1.20 cm. If the angular magnification has a magnitude of 36 and the object is at infinity, find the position of the eyepiece and the position and nature (real or virtual) of the final image. (*Note:* $|M|$ is *not* equal to $|f_1|/|f_2|$, so the image formed by the eyepiece is *not* at infinity.)

34.115 • A microscope with an objective of focal length 8.00 mm and an eyepiece of focal length 7.50 cm is used to project an image on a screen 2.00 m from the eyepiece. Let the image distance of the objective be 18.0 cm. (a) What is the lateral magnification of the image? (b) What is the distance between the objective and the eyepiece?

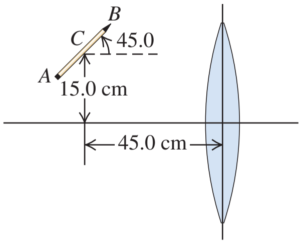
CHALLENGE PROBLEMS

34.116 ••• *Spherical aberration* is a blurring of the image formed by a spherical mirror. It occurs because parallel rays striking the mirror far from the optic axis are focused at a different point than are rays near the axis. This problem is usually minimized by using only the center of a spherical mirror. (a) Show that for a spherical concave mirror, the focus moves toward the mirror as the parallel rays move toward the outer edge of the mirror. (*Hint:* Derive an analytic expression for the distance from the vertex to the focus of the ray for a particular parallel ray. This expression should be in terms of (i) the radius of curvature R of the mirror and (ii) the angle θ between the incident ray and a line connecting the center of curvature of the mirror with the point where the ray strikes the mirror.) (b) What value of θ produces a 2% change in the location of the focus, compared to the location for θ very close to zero?

34.117 ••• 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)?

34.118 ••• 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. P34.118. (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 **P34.118**



34.119 ••• 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 34.98. 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.)

Answers

Chapter Opening Question ?

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 doing microsurgery would not appreciate having to strain his 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.

Test Your Understanding Questions

34.1 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 $2v$.

34.2 Answers: (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.

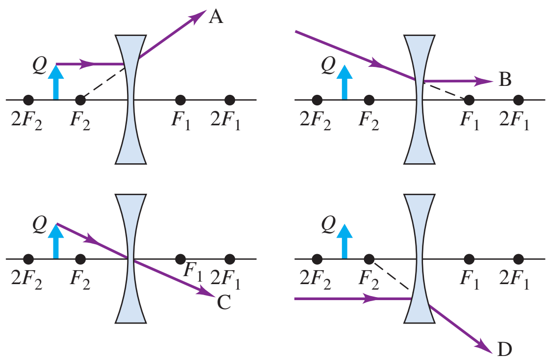
34.3 Answer: no The sun is very far away, so the object distance is essentially infinite: $s = \infty$ and $1/s = 0$. Material a is air ($n_a = 1.00$) and material b is water ($n_b = 1.33$), so the image position s' is given by

$$\frac{n_a}{s} + \frac{n_b}{s'} = \frac{n_b - n_a}{R} \quad \text{or} \quad 0 + \frac{1.33}{s'} = \frac{1.33 - 1.00}{R}$$

$$s' = \frac{1.33}{0.33}R = 4.0R$$

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. While the rays are not focused to a point, they are nonetheless concentrated and can cause damage to the leaf.

34.4 Answers: 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 Answer: (iii) The smaller image area of the CCD array 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 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 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 (24.22) 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 Answer: (i) The objective lens of a microscope is designed to make 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 is designed to make *reduced* images. For example, the moon is thousands of kilometers in diameter, but its image may fit on a CCD array a few centimeters across. Thus $|m|$ is much less than 1 for a refracting telescope. (In both cases m is negative because the objective makes an inverted image, which is why the question asks about the absolute value of m.)

Bridging Problem

Answers: (a) 29.9 cm to the left of the goblet
(b) 3.73 cm to the right of the goblet

35 INTERFERENCE

? Soapy water is colorless, but when blown into bubbles it shows vibrant colors. How does the thickness of the bubble walls determine the particular colors that appear?

LEARNING GOALS

By studying this chapter, you will learn:

- What happens when two waves combine, or interfere, in space.
- How to understand the interference pattern formed by the interference of two coherent light waves.
- How to calculate the intensity at various points in an interference pattern.
- How interference occurs when light reflects from the two surfaces of a thin film.
- How interference makes it possible to measure extremely small distances.

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 many aspects of the behavior of light *can't* be understood on the basis of rays. We have already learned that 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. The colors seen in oil films and soap bubbles are a result of interference between light reflected from the front and back surfaces of a thin film of oil or soap solution. Effects that occur when *many* sources of waves are present are called *diffraction* phenomena; we'll study these in Chapter 36. In that chapter we'll see that diffraction effects occur whenever a wave passes through an aperture or around an obstacle. They are important in practical applications of physical optics such as diffraction gratings, x-ray diffraction, and holography.

While our primary concern is with light, interference and diffraction can occur with waves of *any* kind. As we go along, we'll point out applications to other types of waves such as sound and water waves.

35.1 Interference and Coherent Sources

As we discussed in Chapter 15, the term **interference** refers to any situation in which two or more waves overlap in space. When this occurs, the total wave at any point at any instant of time is governed by the **principle of superposition,** which we introduced in Section 15.6 in the context of waves on a string. This principle also applies to electromagnetic waves and is the most important principle in all of physical optics. The principle of superposition states:

When two or more waves overlap, the resultant displacement at any point and at any instant is found by adding the instantaneous displacements that would be produced at the point by the individual waves if each were present alone.

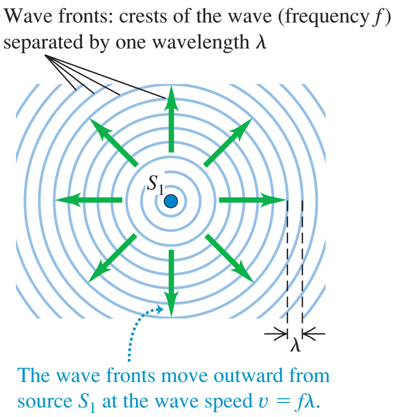
(In some special situations, such as electromagnetic waves propagating in a crystal, this principle may not apply. A discussion of these is beyond our scope.)

We use the term "displacement" in a general sense. With waves on the surface of a liquid, we mean the actual displacement of the surface above or below its normal level. With sound waves, the term refers to the excess or deficiency of pressure. For electromagnetic waves, we usually mean a specific component of electric or magnetic field.

Interference in Two or Three Dimensions

We have already discussed one important case of interference, in which two identical waves propagating in opposite directions combine to produce a *standing wave.* We saw this in Chapters 15 and 16 for transverse waves on a string and 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.

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 source that is available at present 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. As we analyze interference and diffraction effects in this chapter and the next, we will assume that we are working with monochromatic waves (unless we explicitly state otherwise).

Constructive and Destructive Interference

Two identical sources of monochromatic waves, S_1 and S_2, are shown in Fig. 35.2a. The two sources produce waves of the same amplitude and the same wavelength λ.

In addition, the two sources are permanently *in phase;* they vibrate in unison. They might be two loudspeakers driven by the same amplifier, two radio antennas powered by the same transmitter, or two small slits in an opaque screen, illuminated by the same monochromatic light source. We will see that if there were not a constant phase relationship between the two sources, the phenomena we are about to discuss would not occur. Two monochromatic sources of the same frequency and with a constant phase relationship (not necessarily in phase) are said to be **coherent.** We also use the term *coherent waves* (or, for light waves, *coherent light*) to refer to the waves emitted by two such sources.

If the waves emitted by the two coherent sources are *transverse,* like electromagnetic waves, then we will also assume that the wave disturbances produced by both sources have the same *polarization* (that is, they lie along the same line). For example, the sources S_1 and S_2 in Fig. 35.2a could be two radio antennas in the form of long rods oriented parallel to the z-axis (perpendicular to the plane of the figure); at any point in the xy-plane the waves produced by both antennas have $\vec{E}$ fields with only a z-component. Then we need only a single scalar function to describe each wave; this makes the analysis much easier.

We position the two sources of equal amplitude, equal wavelength, and (if the waves are transverse) the same polarization along the y-axis in Fig. 35.2a, equidistant from the origin. Consider a point a on the x-axis. From symmetry the two distances from S_1 to a and from S_2 to a are *equal;* waves from the two sources thus require equal times to travel to a. Hence waves that leave S_1 and S_2 in phase arrive at a in phase. The two waves add, and the total amplitude at a is *twice* the amplitude of each individual wave. This is true for *any* point on the x-axis.

Similarly, the distance from S_2 to point b is exactly two wavelengths *greater* than the distance from S_1 to b. A wave crest from S_1 arrives at b exactly two cycles earlier than a crest emitted at the same time from S_2, and again the two waves arrive in phase. As at point a, the total amplitude is the sum of the amplitudes of the waves from S_1 and S_2.

In general, when waves from two or more sources arrive at a point *in phase,* they reinforce each other: The amplitude of the resultant wave is the *sum* of the amplitudes of the individual waves. This is called **constructive interference** (Fig. 35.2b). Let the distance from S_1 to any point P be r_1, and let the distance from S_2 to P be r_2. For constructive interference to occur at P, the path difference $r_2 - r_1$ for the two sources must be an integral multiple of the wavelength λ:

$$r_2 - r_1 = m\lambda \qquad (m = 0, \pm 1, \pm 2, \pm 3, \dots) \qquad \text{(constructive interference, sources in phase)} \tag{35.1}$$

In Fig. 35.2a, points a and b satisfy Eq. (35.1) with $m = 0$ and $m = +2$, respectively.

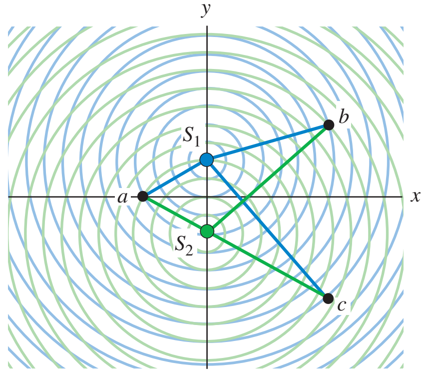
Something different occurs at point c in Fig. 35.2a. At this point, the path difference $r_2 - r_1 = -2.50\lambda$, which is a *half-integral* number of wavelengths. Waves from the two sources arrive at point c exactly a half-cycle out of phase. A crest of one wave arrives at the same time as a crest in the opposite direction (a "trough") of the other wave (Fig. 35.2c). The resultant amplitude is the *difference* between the two individual amplitudes. If the individual amplitudes are equal, then the total amplitude is *zero*! This cancellation or partial cancellation of the individual waves is called **destructive interference.** The condition for destructive interference in the situation shown in Fig. 35.2a is

$$r_2 - r_1 = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, \pm 1, \pm 2, \pm 3, \dots) \quad \text{(destructive interference, sources in phase)} \tag{35.2}$$

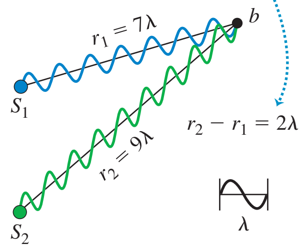
The path difference at point c in Fig. 35.2a satisfies Eq. (35.2) with $m = -3$.

35.2 (a) A "snapshot" of sinusoidal waves spreading out from two coherent sources S_1 and S_2. Constructive interference occurs at point a (equidistant from the two sources) and (b) at point b. (c) Destructive interference occurs at point c.

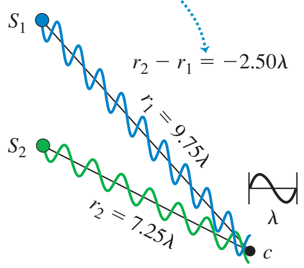
(a) Two coherent wave sources separated by a distance 4λ



(b) Conditions for constructive interference: Waves interfere constructively if their path lengths differ by an integral number of wavelengths: $r_2 - r_1 = m\lambda$.



(c) Conditions for destructive interference: Waves interfere destructively if their path lengths differ by a half-integral number of wavelengths: $r_2 - r_1 = (m + \frac{1}{2})\lambda$.



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.

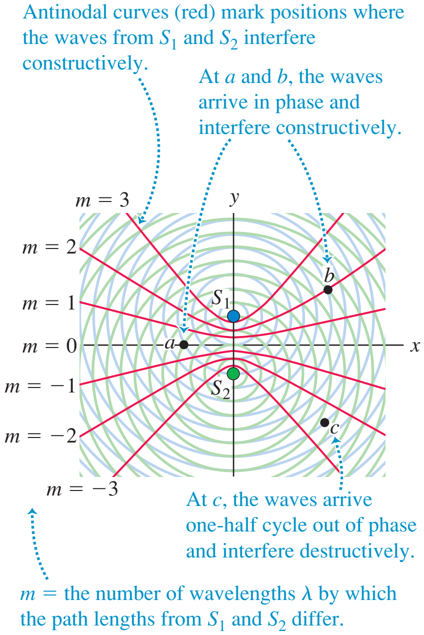


Figure 35.3 shows the same situation as in Fig. 35.2a, but with red curves that show all positions where *constructive* interference occurs. On each curve, the path difference $r_2 - r_1$ is equal to an integer m times the wavelength, as in Eq. (35.1). These curves are called **antinodal curves.** They 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** The interference patterns in Figs. 35.2a and 35.3 are *not* standing waves, though they have some similarities to the standing-wave patterns described in Chapters 15 and 16 and Section 32.5. In a standing wave, the interference is between two waves propagating in opposite directions; there is *no* net energy flow in either direction (the energy in the wave is left "standing"). In the situations shown in Figs. 35.2a and 35.3, there is likewise a stationary pattern of antinodal and nodal curves, but there is a net flow of energy *outward* from the two sources. All that interference does is to "channel" the energy flow so that it is greatest along the antinodal curves and least along the nodal curves.

For Eqs. (35.1) and (35.2) to hold, the two sources must have the same wavelength and must *always* be in phase. These conditions are rather easy to satisfy for sound waves. But with *light* waves there is no practical way to achieve a constant phase relationship (coherence) with two independent sources. This is because of the way light is emitted. In ordinary light sources, atoms gain excess energy by thermal agitation or by impact with accelerated electrons. Such an "excited" atom begins to radiate energy and continues until it has lost all the energy it can, typically in a time of the order of 10^{-8} s. The many atoms in a source ordinarily radiate in an unsynchronized and random phase relationship, and the light that is emitted from *two* such sources has no definite phase relationship.

However, the light from a single source can be split so that parts of it emerge from two or more regions of space, forming two or more *secondary sources.* Then any random phase change in the source affects these secondary sources equally and does not change their *relative* phase.

The distinguishing feature of light from a *laser* is that the emission of light from many atoms is synchronized in frequency and phase. As a result, the random phase changes mentioned above occur much less frequently. Definite phase relationships are preserved over correspondingly much greater lengths in the beam, and laser light is much more coherent than ordinary light.

Test Your Understanding of Section 35.1 Consider a point in Fig. 35.3 on the positive y-axis above S_1. Does this point lie on (i) an antinodal curve; (ii) a nodal curve; or (iii) neither? (*Hint:* The distance between S_1 and S_2 is 4λ.)

35.2 Two-Source Interference of Light

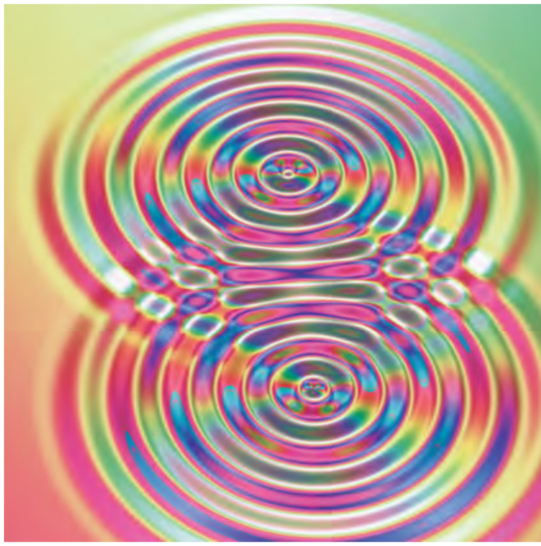
The interference pattern produced by two coherent sources of *water* waves of the same wavelength can be readily seen in a ripple tank with a shallow layer of water (Fig. 35.4). This pattern is not directly visible when the interference is between *light* waves, since light traveling in a uniform medium cannot be seen. (A shaft of afternoon sunlight in a room is made visible by scattering from airborne dust particles.)

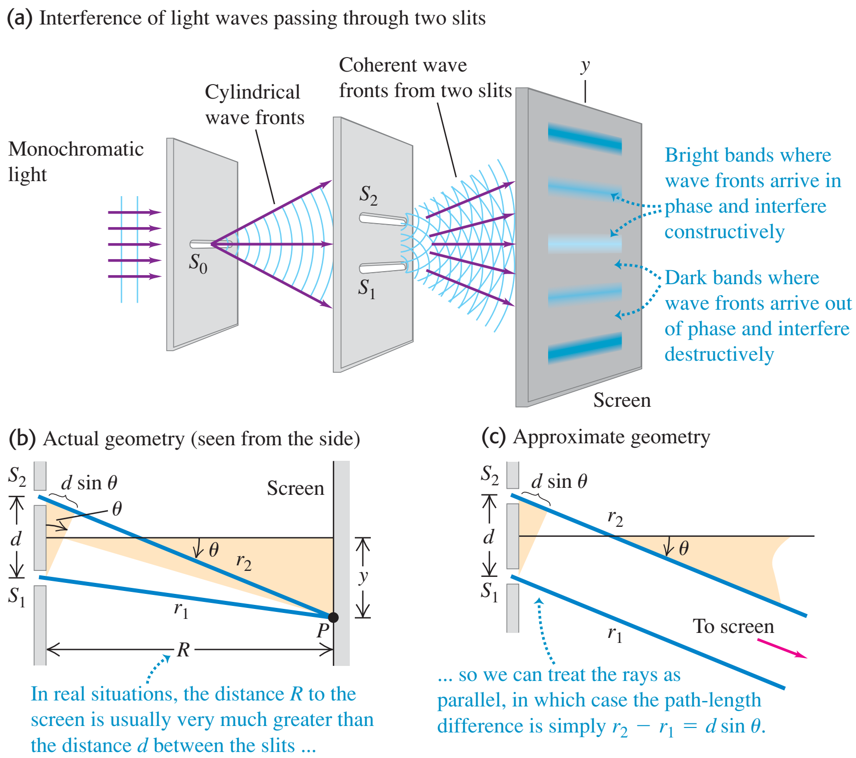
One of the earliest quantitative experiments to reveal the interference of light from two sources was performed in 1800 by the English scientist Thomas Young. We will refer back to this experiment several times in this and later chapters, so it's important to understand it in detail. Young's apparatus is shown in perspective in Fig. 35.5a. A light source (not shown) emits monochromatic light; however, this light is not suitable for use in an interference experiment because emissions from different parts of an ordinary source are not synchronized. To remedy this, the light is directed at a screen with a narrow slit S_0, 1 μm or so wide. The light emerging from the slit originated from only a small region of the light source; thus slit S_0 behaves more nearly like the idealized source shown in Fig. 35.1. (In modern versions of the experiment, a laser is used as a source of coherent light, and the slit S_0 isn't needed.) The light from slit S_0 falls on a screen with two other narrow slits S_1 and S_2, each 1 μm or so wide and a few tens or hundreds of micrometers apart. Cylindrical wave fronts spread out from slit S_0 and reach slits S_1 and S_2 *in phase* because they travel equal distances from S_0. The waves *emerging* from slits S_1 and S_2 are therefore also always in phase, so S_1 and S_2 are *coherent* sources. The interference of waves from S_1 and S_2 produces a pattern in space like that to the right of the sources in Figs. 35.2a and 35.3.

To visualize the interference pattern, a screen is placed so that the light from S_1 and S_2 falls on it (Fig. 35.5b). The screen will be most brightly illuminated at points P, where the light waves from the slits interfere constructively, and will be darkest at points where the interference is destructive.

To simplify the analysis of Young's experiment, we assume that the distance R from the slits to the screen is so large in comparison to the distance d between the slits that the lines from S_1 and S_2 to P are very nearly parallel, as in Fig. 35.5c. 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

35.4 The concepts of constructive interference and destructive interference apply to these water waves as well as to light waves and sound waves.





35.5 **(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.

MasteringPHYSICS

PhET: Wave Interference
ActivPhysics 16.1: Two-Source Interference: Introduction
ActivPhysics 16.2: Two-Source Interference: Qualitative Questions
ActivPhysics 16.3: Two-Source Interference: Problems

$$r_2 - r_1 = d\sin\theta \qquad (35.3)$$

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

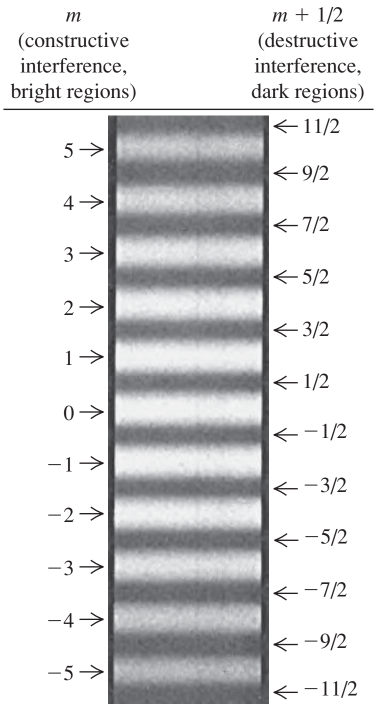
We found in Section 35.1 that constructive interference (reinforcement) occurs at points where the path difference is an integral number of wavelengths, $m\lambda$, where $m = 0, \pm1, \pm2, \pm3, \ldots$. So the bright regions on the screen in Fig. 35.5a occur at angles θ for which

$$d\sin\theta = m\lambda \qquad (m = 0, \pm1, \pm2, \ldots) \qquad \text{(constructive interference, two slits)} \qquad (35.4)$$

Similarly, destructive interference (cancellation) occurs, forming dark regions on the screen, at points for which the path difference is a half-integral number of wavelengths, $\left(m + \frac{1}{2}\right)\lambda$:

$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \qquad (m = 0, \pm1, \pm2, \ldots) \qquad \text{(destructive interference, two slits)} \qquad (35.5)$$

35.6 Photograph of interference fringes produced on a screen in Young's double-slit experiment.



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. 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 experiments such as this, the distances y_m are often much smaller than the distance R from the slits to the screen. Hence θ_m is very small, $\tan\theta_m$ is very nearly equal to $\sin\theta_m$, and

$$y_m = R\sin\theta_m$$

Combining this with Eq. (35.4), we find that *for small angles only,*

$$y_m = R\frac{m\lambda}{d} \qquad \text{(constructive interference in Young's experiment)} \qquad (35.6)$$

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.

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 only for *small* angles. 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 fringe.

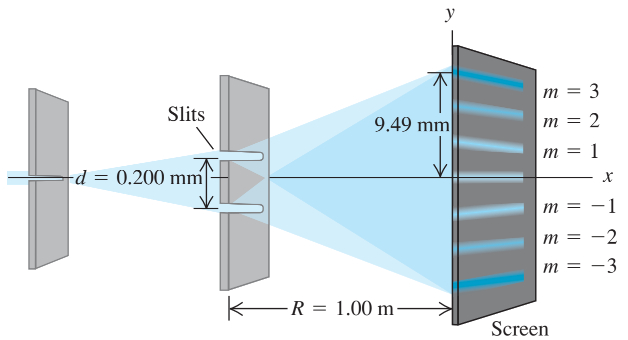
The distance between adjacent bright bands in the pattern is *inversely* proportional to the distance d between the slits. The closer together the slits are, the more the pattern spreads out. When the slits are far apart, the bands in the pattern are closer together.

While we have described the experiment that Young performed with visible light, the results given in Eqs. (35.4) and (35.5) are valid for *any* type of wave, provided that the resultant wave from two coherent sources is detected at a point that is far away in comparison to the separation d.

Example 35.1 Two-slit interference

Figure 35.7 shows a two-slit interference experiment in which the slits are 0.200 mm apart and the screen is 1.00 m from the slits. The $m = 3$ bright fringe in the figure is 9.49 mm from the central fringe. Find the wavelength of the light.

35.7 Using a two-slit interference experiment to measure the wavelength of light.



SOLUTION

IDENTIFY and SET UP: Our target variable in this two-slit interference problem is the wavelength λ. We are given the slit separation $d = 0.200$ mm, the distance from slits to screen $R = 1.00$ m, and the distance $y_3 = 9.49$ mm on the screen from the center of the interference pattern to the $m = 3$ bright fringe. We may use Eq. (35.6) to find λ, since the value of R is so much greater than the values of d or y_3.

EXECUTE: We solve Eq. (35.6) for λ for the case $m = 3$:

$$\lambda = \frac{y_m d}{mR} = \frac{(9.49 \times 10^{-3}\ \text{m})(0.200 \times 10^{-3}\ \text{m})}{(3)(1.00\ \text{m})}$$
$$= 633 \times 10^{-9}\ \text{m} = 633\ \text{nm}$$

EVALUATE: This bright fringe could also correspond to $m = -3$. Can you show that this gives the same result for λ?

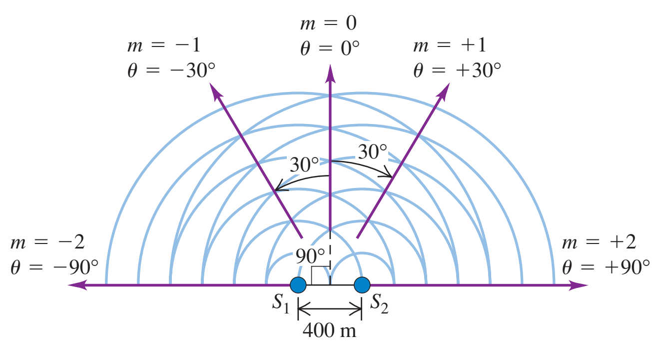
Example 35.2 Broadcast pattern of a radio station

It is often desirable to radiate most of the energy from a radio transmitter in particular directions rather than uniformly in all directions. Pairs or rows of antennas are often used to produce the desired radiation pattern. As an example, consider two identical vertical antennas 400 m apart, operating at 1500 kHz $= 1.5 \times 10^6$ Hz (near the top end of the AM broadcast band) and oscillating in phase. At distances much greater than 400 m, in what directions is the intensity from the two antennas greatest?

SOLUTION

IDENTIFY and SET UP: The antennas, shown in Fig. 35.8, correspond to sources S_1 and S_2 in Fig. 35.5. Hence we can apply the ideas of two-slit interference to this problem. Since the resultant wave is detected at distances much greater than $d = 400$ m, we may use Eq. (35.4) to give the directions of the intensity maxima, the values of θ for which the path difference is zero or a whole number of wavelengths.

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,

$$\sin\theta = \frac{(m + \frac{1}{2})\lambda}{d} = \frac{m + \frac{1}{2}}{2} \qquad \theta = \pm 14.5°, \pm 48.6°$$

These angles fall between the angles for intensity maxima, as they should. The angles are not small, so the angles for the minima are *not* exactly halfway between the angles for the maxima.

Test Your Understanding of Section 35.2 You shine a tunable laser (whose wavelength can be adjusted by turning a knob) on a pair of closely spaced slits. The light emerging from the two slits produces an interference pattern on a screen like that shown in Fig. 35.6. If you adjust the wavelength so that the laser light changes from red to blue, how will the spacing between bright fringes change? (i) The spacing increases; (ii) the spacing decreases; (iii) the spacing is unchanged; (iv) not enough information is given to decide.

35.3 Intensity in Interference Patterns

In Section 35.2 we found the positions of maximum and minimum intensity in a two-source interference pattern. Let's now see how to find the intensity at *any* point in the pattern. To do this, we have to combine the two sinusoidally varying fields (from the two sources) at a point P in the radiation pattern, taking proper account of the phase difference of the two waves at point P, which results from the path difference. The intensity is then proportional to the square of the resultant electric-field amplitude, as we learned in Chapter 32.

To calculate the intensity, we will assume that the two sinusoidal functions (corresponding to waves from the two sources) have equal amplitude E and that the $\vec{E}$ fields lie along the same line (have the same polarization). This assumes that the sources are identical and neglects 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 proportional to the difference in their path lengths, $(r_2 - r_1)$. If the phase angle between these arriving waves is ϕ, then we can use the following expressions for the two electric fields superposed at P:

$$E_1(t) = E\cos(\omega t + \phi)$$
$$E_2(t) = E\cos\omega t$$

The superposition of the two fields at P is a sinusoidal function with some amplitude E_P that depends on E and the phase difference ϕ. First we'll work on finding the amplitude E_P if E and ϕ are known. Then we'll find the intensity I of the resultant wave, which is proportional to $E_P{}^2$. Finally, we'll relate the phase difference ϕ to the path difference, which is determined by the geometry of the situation.

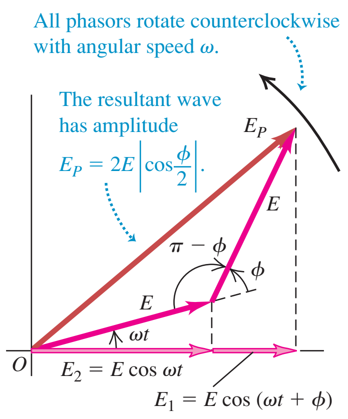
Amplitude in Two-Source Interference

To add the two sinusoidal functions with a phase difference, we use the same *phasor* representation that we used for simple harmonic motion (see Section 14.2) and for voltages and currents in ac circuits (see Section 31.1). We suggest that you review these sections now. Each sinusoidal function is represented by a rotating vector (phasor) whose projection on the horizontal axis at any instant represents the instantaneous value of the sinusoidal function.

In Fig. 35.9, E_1 is the horizontal component of the phasor representing the wave from source S_1, and E_2 is the horizontal component of the phasor for the wave from S_2. As shown in the diagram, both phasors have the same magnitude E, but E_1 is *ahead* of E_2 in phase by an angle ϕ. Both phasors rotate counterclockwise with constant angular speed ω, and the sum of the projections on the horizontal axis at any time gives the instantaneous value of the total E field at point P. Thus the amplitude E_P of the resultant sinusoidal wave at P is the magnitude of the dark red phasor in the diagram (labeled E_P); this is the *vector sum* of the other two phasors. To find E_P, we use the law of cosines and the trigonometric identity $\cos(\pi - \phi) = -\cos\phi$:

$$E_P{}^2 = E^2 + E^2 - 2E^2\cos(\pi - \phi)$$
$$= E^2 + E^2 + 2E^2\cos\phi$$

35.9 Phasor diagram for the superposition at a point P of two waves of equal amplitude E with a phase difference ϕ.



Then, using the identity $1 + \cos\phi = 2\cos^2(\phi/2)$, we obtain

$$E_P{}^2 = 2E^2(1 + \cos\phi) = 4E^2\cos^2\left(\frac{\phi}{2}\right)$$

$$E_P = 2E\left|\cos\frac{\phi}{2}\right| \qquad \text{(amplitude in two-source interference)} \tag{35.7}$$

You can also obtain this result without using phasors (see Problem 35.50).

When the two waves are in phase, $\phi = 0$ and $E_P = 2E$. When they are exactly a half-cycle out of phase, $\phi = \pi$ rad $= 180°$, $\cos(\phi/2) = \cos(\pi/2) = 0$, and $E_P = 0$. Thus the superposition of two sinusoidal waves with the same frequency and amplitude but with a phase difference yields a sinusoidal wave with the same frequency and an amplitude between zero and twice the individual amplitudes, depending on the phase difference.

Intensity in Two-Source Interference

To obtain the intensity I at point P, we recall from Section 32.4 that I is equal to the average magnitude of the Poynting vector, S_{av}. For a sinusoidal wave with electric-field amplitude E_P, this is given by Eq. (32.29) with E_{max} replaced by E_P. Thus we can express the intensity in several equivalent forms:

$$I = S_{av} = \frac{E_P{}^2}{2\mu_0 c} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}E_P{}^2 = \frac{1}{2}\epsilon_0 c E_P{}^2 \tag{35.8}$$

The essential content of these expressions is that I is proportional to $E_P{}^2$. When we substitute Eq. (35.7) into the last expression in Eq. (35.8), we get

$$I = \tfrac{1}{2}\epsilon_0 c E_P{}^2 = 2\epsilon_0 c E^2 \cos^2\frac{\phi}{2} \tag{35.9}$$

In particular, the *maximum* intensity I_0, which occurs at points where the phase difference is zero ($\phi = 0$), is

$$I_0 = 2\epsilon_0 c E^2$$

Note that the maximum intensity I_0 is *four times* (not twice) as great as the intensity $\frac{1}{2}\epsilon_0 c E^2$ from each individual source.

Substituting the expression for I_0 into Eq. (35.9), we can express the intensity I at any point very simply in terms of the maximum intensity I_0:

$$I = I_0\cos^2\frac{\phi}{2} \qquad \text{(intensity in two-source interference)} \tag{35.10}$$

For some phase angles ϕ the intensity is I_0, four times as great as for an individual wave source, but for other phase angles the intensity is 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 (as we mentioned in Section 35.1).

Phase Difference and Path Difference

Our next task is to find the phase difference ϕ between the two fields at any point P. We know that ϕ is proportional to the difference in path length from the two sources to point P. When the path difference is one wavelength, the phase difference is one cycle, and $\phi = 2\pi$ rad $= 360°$. When the path difference is

$\lambda/2$, $\phi = \pi$ rad $= 180°$, and so on. That is, the ratio of the phase difference ϕ to 2π is equal to the ratio of the path difference $r_2 - r_1$ to λ:

$$\frac{\phi}{2\pi} = \frac{r_2 - r_1}{\lambda}$$

Thus a path difference $(r_2 - r_1)$ causes a phase difference given by

$$\phi = \frac{2\pi}{\lambda}(r_2 - r_1) = k(r_2 - r_1) \qquad \text{(phase difference related to path difference)} \tag{35.11}$$

where $k = 2\pi/\lambda$ is the *wave number* introduced 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 the material has index of refraction n, then

$$\lambda = \frac{\lambda_0}{n} \qquad \text{and} \qquad k = nk_0 \tag{35.12}$$

where λ_0 and k_0 are the wavelength and the wave number, respectively, in vacuum.

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

$$I = I_0\cos^2\left(\tfrac{1}{2}kd\sin\theta\right) = I_0\cos^2\left(\frac{\pi d}{\lambda}\sin\theta\right) \qquad \text{(intensity far from two sources)} \tag{35.14}$$

The directions of *maximum* intensity occur 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

$$d\sin\theta = m\lambda$$

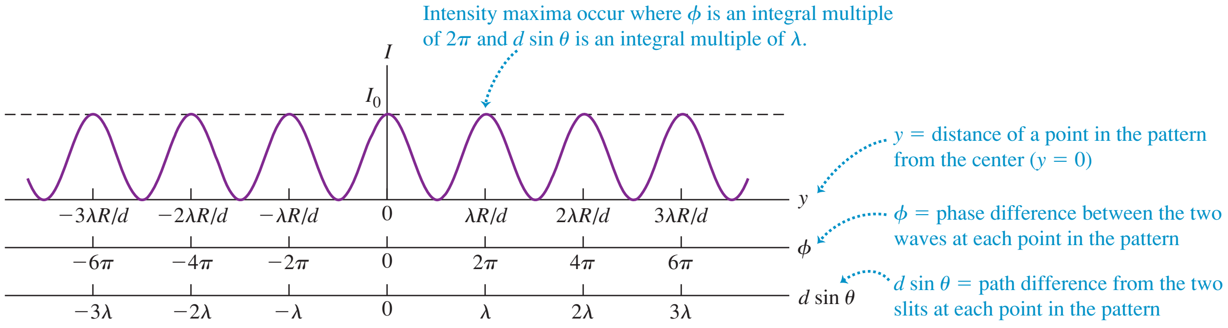
in agreement with Eq. (35.4). You can also derive Eq. (35.5) for the zero-intensity directions from Eq. (35.14).

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 shows a graph of Eq. (35.15); we can compare this with the photographically recorded pattern of Fig. 35.6. The peaks in Fig. 35.10 all 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.

35.10 Intensity distribution in the interference pattern from two identical slits.



Example 35.3 A directional transmitting antenna array

Suppose the two identical radio antennas of Fig. 35.8 are moved to be only 10.0 m apart and the broadcast frequency is increased to $f = 60.0$ MHz. At a distance of 700 m from the point midway between the antennas and in the direction $\theta = 0$ (see Fig. 35.8), the intensity is $I_0 = 0.020\ \text{W/m}^2$. At this same distance, find (a) the intensity in the direction $\theta = 4.0°$; (b) the direction near $\theta = 0$ for which the intensity is $I_0/2$; and (c) the directions in which the intensity is zero.

SOLUTION

IDENTIFY and SET UP: This problem involves the intensity distribution as a function of angle. Because the 700-m distance from the antennas to the point at which the intensity is measured is much greater than the distance $d = 10.0$ m between the antennas, the amplitudes of the waves from the two antennas are very nearly equal. Hence we can use Eq. (35.14) to relate intensity I and angle θ.

EXECUTE: The wavelength is $\lambda = c/f = 5.00$ m. The spacing $d = 10.0$ m between the antennas is just twice the wavelength (as was the case in Example 35.2), so $d/\lambda = 2.00$ and Eq. (35.14) becomes

$$I = I_0 \cos^2\left(\frac{\pi d}{\lambda}\sin\theta\right) = I_0 \cos^2[(2.00\pi \text{ rad})\sin\theta]$$

(a) When $\theta = 4.0°$,

$$I = I_0 \cos^2[(2.00\pi \text{ rad})\sin 4.0°] = 0.82 I_0$$
$$= (0.82)(0.020\ \text{W/m}^2) = 0.016\ \text{W/m}^2$$

(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 that $\sin\theta = \pm(1/8.00) = \pm 0.125$ and $\theta = \pm 7.2°$.

(c) The intensity is zero when $\cos[(2.00\pi \text{ rad})\sin\theta] = 0$. This occurs for $2.00\pi \sin\theta = \pm\pi/2, \pm 3\pi/2, \pm 5\pi/2, \ldots$, or $\sin\theta = \pm 0.250, \pm 0.750, \pm 1.25, \ldots$. Values of $\sin\theta$ greater than 1 have no meaning, so the answers are

$$\theta = \pm 14.5°, \pm 48.6°$$

EVALUATE: The condition in part (b) that $I = I_0/2$, so that $(2.00\pi \text{ rad})\sin\theta = \pm\pi/4$ rad, is also satisfied when $\sin\theta = \pm 0.375, \pm 0.625$, or ± 0.875 so that $\theta = \pm 22.0°, \pm 38.7°$, or $\pm 61.0°$. (Can you verify this?) It would be incorrect to include these angles in the solution, however, because the problem asked for the angle *near* $\theta = 0$ at which $I = I_0/2$. These additional values of θ aren't the ones we're looking for.

Test Your Understanding of Section 35.3 A two-slit interference experiment uses coherent light of wavelength 5.00×10^{-7} m. Rank the following points in the interference pattern according to the intensity at each point, from highest to lowest. (i) a point that is closer to one slit than the other by 4.00×10^{-7} m; (ii) a point where the light waves received from the two slits are out of phase by 4.00 rad; (iii) a point that is closer to one slit than the other by 7.50×10^{-7} m; (iv) a point where the light waves received by the two slits are out of phase by 2.00 rad.

35.4 Interference in Thin Films

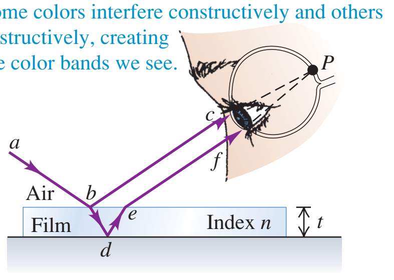
You often see bright bands of color when light reflects from a thin layer of oil floating on water or from a soap bubble (see the photograph that opens this chapter). These are the results of interference. Light waves are reflected from the front and back surfaces of such thin films, and constructive interference between the two reflected waves (with different path lengths) occurs in different

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) The rainbow fringes of an oil slick on water

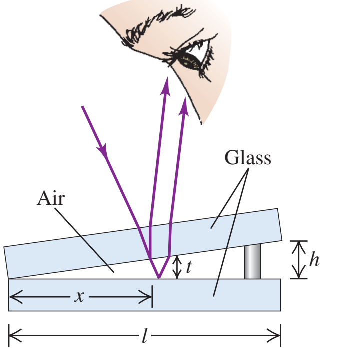
places for different wavelengths. Figure 35.11a shows the situation. Light shining on the upper surface of a thin film with thickness t is partly reflected at the upper surface (path *abc*). Light *transmitted* through the upper surface is partly reflected at the lower surface (path *abdef*). The two 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 rings or fringes in Fig. 35.11b (which shows a thin film of oil floating on water) and in the photograph that opens this chapter (which shows thin films of soap solution that make up the bubble walls). The complex shapes of the colored rings in each photograph result from variations in the thickness of the film.

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, as shown. (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. Along the line 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.

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.



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 proportional to the amplitude E_i of the incident wave and is given by

$$E_r = \frac{n_a - n_b}{n_a + n_b} E_i \quad \text{(normal incidence)} \tag{35.16}$$

This result shows that the incident and reflected amplitudes have the same sign when n_a is larger than n_b and opposite sign when n_b is larger than n_a. We can distinguish three cases, as shown in Fig. 35.13:

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 or a ring that can move vertically without friction.

Figure 35.13b: When $n_a = n_b$, the amplitude E_r of the reflected wave is zero. The incident light wave can't "see" the interface, and there is *no* reflected wave.

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.

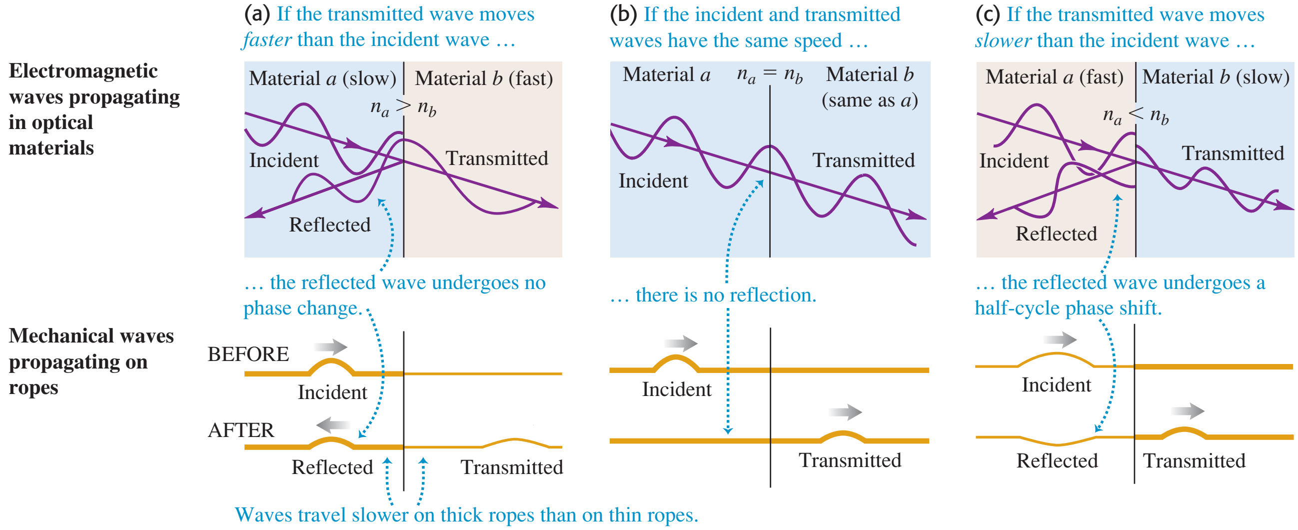


Figure 35.13c: When $n_a < n_b$, light travels more slowly in the second material than in the first. In this case, E_r and E_i have opposite signs, and the phase shift of the reflected wave relative to the incident wave is π rad (180° or a half-cycle). This is analogous to reflection (with inversion) of a transverse mechanical wave on a light rope at a point where it is tied to a heavier rope or a rigid support.

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

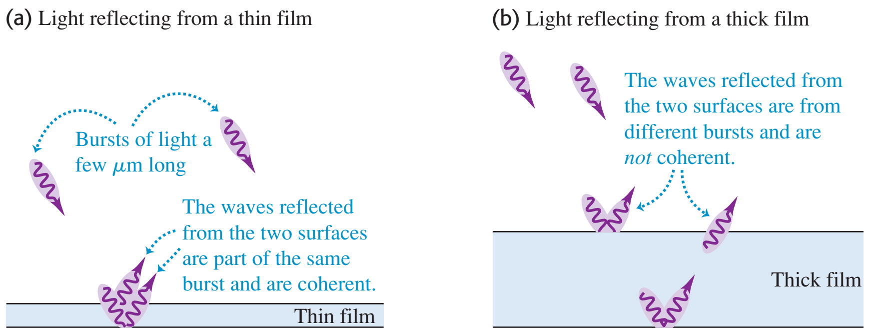
$$2t = m\lambda \qquad (m = 0, 1, 2, \dots) \qquad \text{(constructive reflection from thin film, no relative phase shift)} \tag{35.17a}$$

$$2t = \left(m + \tfrac{1}{2}\right)\lambda \qquad (m = 0, 1, 2, \dots) \qquad \text{(destructive reflection from thin film, no relative phase shift)} \tag{35.17b}$$

If *one* of the two waves has a half-cycle reflection phase shift, the conditions for constructive and destructive interference are reversed:

$$2t = \left(m + \tfrac{1}{2}\right)\lambda \qquad (m = 0, 1, 2, \dots) \qquad \text{(constructive reflection from thin film, half-cycle relative phase shift)} \tag{35.18a}$$

$$2t = m\lambda \qquad (m = 0, 1, 2, \dots) \qquad \text{(destructive reflection from thin film, half-cycle relative phase shift)} \tag{35.18b}$$



35.14 (a) Light reflecting from a thin film produces a steady interference pattern, but (b) light reflecting from a thick film does not.

Thin and Thick Films

We have emphasized *thin* films in our discussion because of a principle we introduced in Section 35.1: In order for two waves to cause a steady interference pattern, the waves must be *coherent,* with a definite and constant phase relationship. The sun and light bulbs emit light in a stream of short bursts, each of which is only a few micrometers long ($1 \text{ micrometer} = 1\ \mu m = 10^{-6}$ m). If light reflects from the two surfaces of a thin film, the two reflected waves are part of the same burst (Fig. 35.14a). Hence these waves are coherent and interference occurs as we have described. If the film is too thick, however, the two reflected waves will belong to different bursts (Fig. 35.14b). There is no definite phase relationship between different light bursts, so the two waves are incoherent and there is no fixed interference pattern. That's why you see interference colors in light reflected from an oil slick 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

MP

IDENTIFY *the relevant concepts:* Problems with thin films involve interference of two waves, one reflected from the film's front surface and one reflected from the back surface. Typically you will be asked to relate the wavelength, the film thickness, and the index of refraction of the film.

SET UP *the problem* using the following steps:

1. Make a drawing showing the geometry of the film. Identify the materials that adjoin the film; their properties determine whether one or both of the reflected waves have a half-cycle phase shift.
2. Identify the target variable.

EXECUTE *the solution* as follows:

1. Apply the rule for phase changes to each reflected wave: There is a half-cycle phase shift when $n_b > n_a$ and none when $n_b < n_a$.
2. If *neither* reflected wave undergoes a phase shift, or if *both* do, use Eqs. (35.17). If only one reflected wave undergoes a phase shift, use Eqs. (35.18).
3. Solve the resulting equation for the target variable. Use the wavelength $\lambda = \lambda_0/n$ of light *in the film* in your calculations, where n is the index of refraction of the film. (For air, $n = 1.000$ to four-figure precision.)
4. If you are asked about the wave that is transmitted through the film, remember that minimum intensity in the reflected wave corresponds to maximum *transmitted* intensity, and vice versa.

EVALUATE *your answer:* Interpret your results by examining what would happen if the wavelength were changed or if the film had a different thickness.

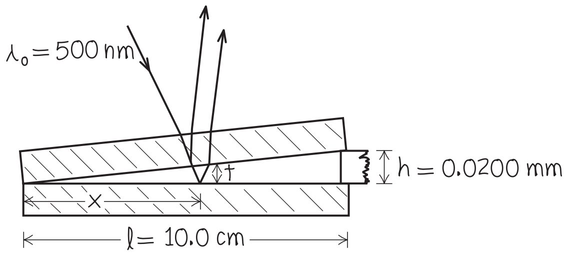
Example 35.4 Thin-film interference I

Suppose the two glass plates in Fig. 35.12 are two microscope slides 10.0 cm long. At one end they are in contact; at the other end they are separated by a piece of paper 0.0200 mm thick. What is the spacing of the interference fringes seen by reflection? Is the fringe at the line of contact bright or dark? Assume monochromatic light with a wavelength in air of $\lambda = \lambda_0 = 500$ nm.

SOLUTION

IDENTIFY and SET UP: Figure 35.15 depicts the situation. We'll consider only interference between the light reflected from the upper and lower surfaces of the air wedge between the microscope slides. [The top slide has a relatively great thickness, about 1 mm,

35.15 Our sketch for this problem.



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):

$$2t = m\lambda_0 \qquad (m = 0, 1, 2, \dots)$$

From similar triangles in Fig. 35.15 the thickness t of the air wedge at each point is proportional to the distance x from the line of contact:

$$\frac{t}{x} = \frac{h}{l}$$

Combining this with Eq. (35.18b), we find

$$\frac{2xh}{l} = m\lambda_0$$

$$x = m\frac{l\lambda_0}{2h} = m\frac{(0.100\text{ m})(500 \times 10^{-9}\text{ m})}{(2)(0.0200 \times 10^{-3}\text{ m})} = m(1.25\text{ mm})$$

Successive dark fringes, corresponding to $m = 1, 2, 3, \dots,$ 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 an oil film on water is illuminated by white light, as in Fig. 35.11b).

Example 35.5 Thin-film interference II

Suppose the glass plates of Example 35.4 have $n = 1.52$ and the space between plates contains water $(n = 1.33)$ instead of air. What happens now?

SOLUTION

IDENTIFY and SET UP: The index of refraction of the water wedge is still less than that of the glass on either side of it, so the phase shifts are the same as in Example 35.4. Once again we use Eq. (35.18b) to find the positions of the dark fringes; the only difference is that the wavelength λ in this equation is now the wavelength in water instead of in air.

EXECUTE: In the film of water $(n = 1.33)$, the wavelength is $\lambda = \lambda_0/n = (500\text{ nm})/(1.33) = 376$ nm. When we replace λ_0 by λ in the expression from Example 35.4 for the position x of the mth dark fringe, we find that the fringe spacing is reduced by the same factor of 1.33 and is equal to 0.940 mm. There is still a dark fringe at the line of contact.

EVALUATE: Can you see that to obtain the same fringe spacing as in Example 35.4, the dimension h in Fig. 35.15 would have to be reduced to $(0.0200\text{ mm})/1.33 = 0.0150$ mm? This shows that what matters in thin-film interference is the *ratio* t/λ between film thickness and wavelength. [You can see this by considering Eqs. (35.17) and (35.18).]

Example 35.6 Thin-film interference III

Suppose the upper of the two plates of Example 35.4 is a plastic with $n = 1.40$, the wedge is filled with a silicone grease with $n = 1.50$, and the bottom plate is a dense flint glass with $n = 1.60$. What happens now?

SOLUTION

IDENTIFY and SET UP: The geometry is again the same as shown in Fig. 35.15, but now half-cycle phase shifts occur at *both* surfaces of the grease wedge (see Fig. 35.13c). Hence there is no *relative* phase shift and we must use Eq. (35.17b) to find the positions of the dark fringes.

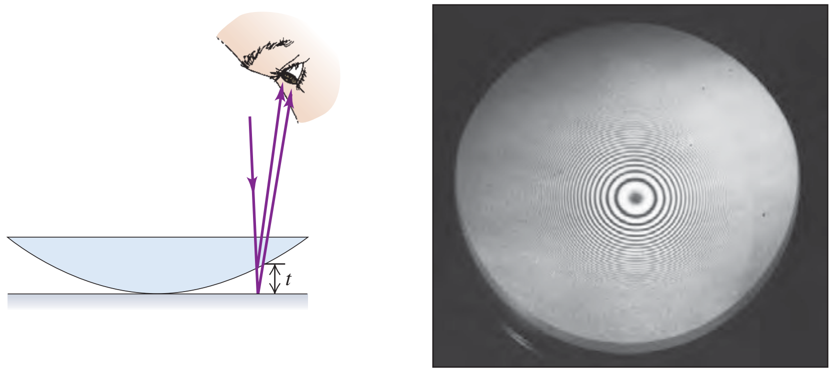
EXECUTE: The value of λ to use in Eq. (35.17b) is the wavelength in the silicone grease, $\lambda = \lambda_0/n = (500\text{ nm})/1.50 = 333$ nm. You can readily show that the fringe spacing is 0.833 mm. Note that the two reflected waves from the line of contact are in phase (they both undergo the same phase shift), so the line of contact is at a *bright* fringe.

EVALUATE: What would happen if you carefully removed the upper microscope slide so that the grease wedge retained its shape? There would still be half-cycle phase changes at the upper and lower surfaces of the wedge, so the pattern of fringes would be the same as with the upper slide present.

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

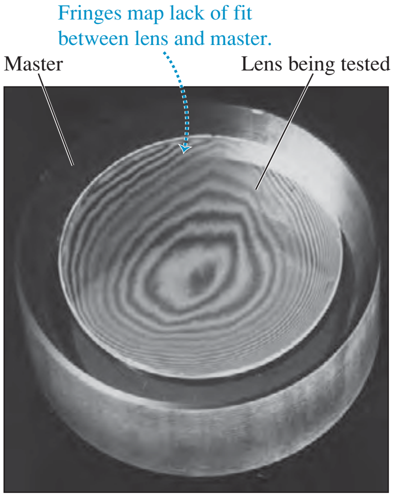


Newton's Rings

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.**

We can use interference fringes to compare the surfaces of two optical parts by placing the two in contact and observing the interference fringes. Figure 35.17 is a photograph made during the grinding of a telescope objective lens. The lower, larger-diameter, thicker disk is the correctly shaped master, and the smaller, upper disk is the lens under test. The "contour lines" are Newton's interference fringes; each one indicates an additional distance between the specimen and the master of one half-wavelength. At 10 lines from the center spot the distance between the two surfaces is five wavelengths, or about 0.003 mm. This isn't very good; high-quality lenses are routinely ground with a precision of less than one wavelength. The surface of the primary mirror of the Hubble Space Telescope was ground to a precision of better than $\frac{1}{50}$ wavelength. Unfortunately, it was ground to incorrect specifications, creating one of the most precise errors in the history of optical technology (see Section 34.2).

35.17 The surface of a telescope objective lens under inspection during manufacture.

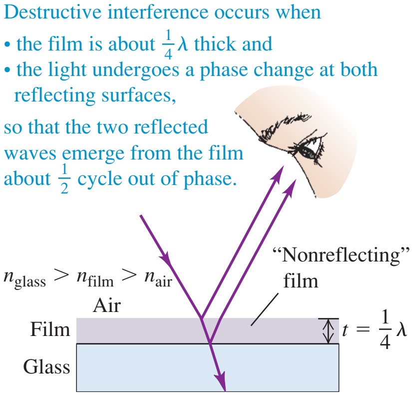


Nonreflective and Reflective Coatings

Nonreflective coatings for lens surfaces make use of thin-film interference. A thin layer or film of hard transparent material with an index of refraction smaller than that of the glass is deposited on the lens surface, as in Fig. 35.18. Light is reflected from both surfaces of the layer. In both reflections the light is reflected from a medium of greater index than that in which it is traveling, so the same phase change occurs in both reflections. If the film thickness is a quarter (one-fourth) of the wavelength *in the film* (assuming normal incidence), the total path difference is a half-wavelength. Light reflected from the first surface is then a half-cycle out of phase with light reflected from the second, and there is destructive interference.

The thickness of the nonreflective coating can be a quarter-wavelength for only one particular wavelength. This is usually chosen in the central yellow-green portion of the spectrum ($\lambda = 550$ nm), where the eye is most sensitive. Then there is somewhat more reflection at both longer (red) and shorter (blue) wavelengths, and the reflected light has a purple hue. The overall reflection from a lens or prism surface can be reduced in this way from 4–5% to less than 1%. This also increases the net amount of light that is *transmitted* through the lens, since light that is not reflected will be transmitted. The same principle is used to minimize reflection from silicon photovoltaic solar cells ($n = 3.5$) by use of a thin surface layer of silicon monoxide (SiO, $n = 1.45$); this helps to increase the amount of light that actually reaches the solar cells.

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% transmission or reflection for particular wavelengths. Some practical applications of these coatings are for color separation in television cameras and for infrared "heat reflectors" in motion-picture projectors, solar cells, and astronauts' visors.

Example 35.7 A nonreflective coating

A common lens coating material is magnesium fluoride (MgF_2), with $n = 1.38$. What thickness should a nonreflective coating have for 550-nm light if it is applied to glass with $n = 1.52$?

SOLUTION

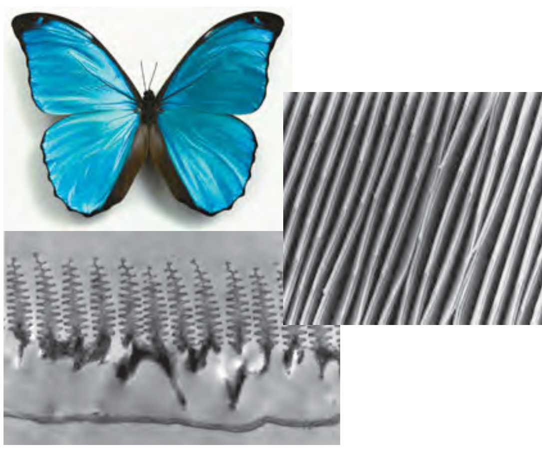
IDENTIFY and SET UP: This coating is of the sort shown in Fig. 35.18. The thickness must be one-quarter of the wavelength of this light *in the coating*.

EXECUTE: The wavelength in air is $\lambda_0 = 550$ nm, so its wavelength in the MgF_2 coating is $\lambda = \lambda_0/n = (550\text{ nm})/1.38 = 400$ nm. The coating thickness should be one-quarter of this, or $\lambda/4 = 100$ nm.

EVALUATE: This is a very thin film, no more than a few hundred molecules thick. Note that this coating is *reflective* for light whose wavelength is *twice* the coating thickness; light of that wavelength reflected from the coating's lower surface travels one wavelength farther than light reflected from the upper surface, so the two waves are in phase and interfere constructively. This occurs for light with a wavelength in MgF_2 of 200 nm and a wavelength in air of $(200\text{ nm})(1.38) = 276$ nm. This is an ultraviolet wavelength (see Section 32.1), so designers of optical lenses with nonreflective coatings need not worry about such enhanced reflection.

Test Your Understanding of Section 35.4 A thin layer of benzene ($n = 1.501$) ties 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.

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 mirror-like brilliance. (The undersides of the wings do not have these structures and are a dull brown.)

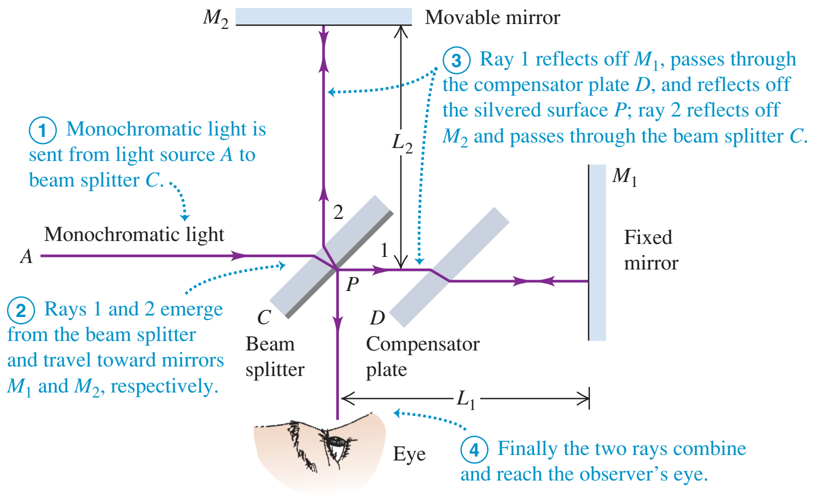


35.5 The Michelson Interferometer

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.

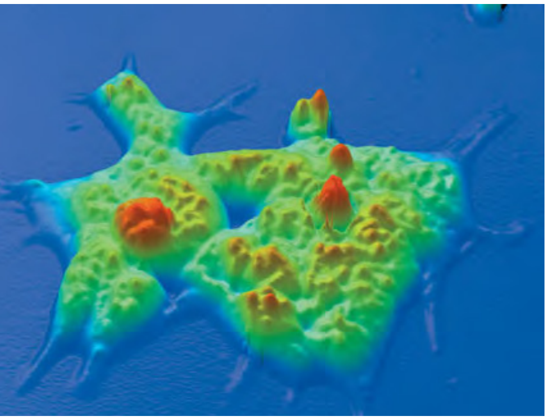
How a Michelson Interferometer Works

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.



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.

Application Imaging Cells with a Michelson Interferometer
This false-color image of a human colon cancer cell was made using a microscope that was mated to a Michelson interferometer. The cell is in one arm of the interferometer, and light passing through the cell undergoes a phase shift that depends on the cell thickness and the organelles within the cell. The fringe pattern can then be used to construct a three-dimensional view of the cell. Scientists have used this technique to observe how different types of cells behave when prodded by microscopic probes. Cancer cells turn out to be "softer" than normal cells, a distinction that may make cancer stem cells easier to identify.



The 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.

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

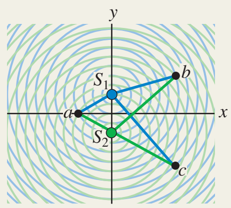
The original application of the Michelson interferometer was to the historic **Michelson-Morley experiment.** Before the electromagnetic theory of light became established, most physicists thought that the propagation of light waves occurred in a medium called the **ether,** which was believed to permeate 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 will study in detail in Chapter 37). Einstein postulated that the speed of a light wave in vacuum has the same magnitude c relative to *all* inertial reference frames, no matter what their velocity may be relative to each other. The presumed ether then plays no role, and the concept of an ether has been abandoned.

Test Your Understanding of Section 35.5 You are observing the pattern of fringes in a Michelson interferometer like that shown in Fig. 35.19. If you change the index of refraction (but not the thickness) of the compensator plate, will the pattern change?

CHAPTER 35 SUMMARY

Interference and coherent sources: Monochromatic light is light with a single frequency. Coherence is a definite, unchanging phase relationship between two waves. The overlap of waves from two coherent sources of monochromatic light forms an interference pattern. The principle of superposition states that the total wave disturbance at any point is the sum of the disturbances from the separate waves.

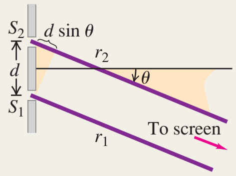


Two-source interference of light: When two sources are in phase, constructive interference occurs where the difference in path length from the two sources is zero or an integer number of wavelengths; destructive interference occurs where the path difference is a half-integer number of wavelengths. If two sources separated by a distance d are both very far from a point P, and the line from the sources to P makes an angle θ with the line perpendicular to the line of the sources, then the condition for constructive interference at P is Eq. (35.4). The condition for destructive interference is Eq. (35.5). When θ is very small, the position y_m of the mth bright fringe on a screen located a distance R from the sources is given by Eq. (35.6). (See Examples 35.1 and 35.2.)

$$d \sin\theta = m\lambda \quad (m = 0, \pm 1, \pm 2, \dots) \tag{35.4}$$
(constructive interference)

$$d \sin\theta = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, \pm 1, \pm 2, \dots) \tag{35.5}$$
(destructive interference)

$$y_m = R\frac{m\lambda}{d} \tag{35.6}$$
(bright fringes)

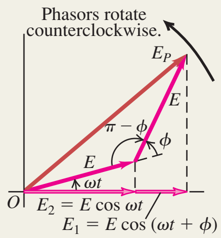


Intensity in interference patterns: When two sinusoidal waves with equal amplitude E and phase difference ϕ are superimposed, the resultant amplitude E_P and intensity I are given by Eqs. (35.7) and (35.10), respectively. If the two sources emit in phase, the phase difference ϕ at a point P (located a distance r_1 from source 1 and a distance r_2 from source 2) is directly proportional to the difference in path length $r_2 - r_1$. (See Example 35.3.)

$$E_P = 2E\left|\cos\frac{\phi}{2}\right| \tag{35.7}$$

$$I = I_0 \cos^2\frac{\phi}{2} \tag{35.10}$$

$$\phi = \frac{2\pi}{\lambda}(r_2 - r_1) = k(r_2 - r_1) \tag{35.11}$$



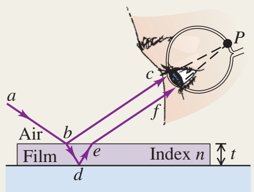
Interference in thin films: When light is reflected from both sides of a thin film of thickness t and no phase shift occurs at either surface, constructive interference between the reflected waves occurs when $2t$ is equal to an integral number of wavelengths. If a half-cycle phase shift occurs at one surface, this is the condition for destructive interference. A half-cycle phase shift occurs during reflection whenever the index of refraction in the second material is greater than that in the first. (See Examples 35.4–35.7.)

$$2t = m\lambda \quad (m = 0, 1, 2, \dots) \tag{35.17a}$$
(constructive reflection from thin film, no relative phase shift)

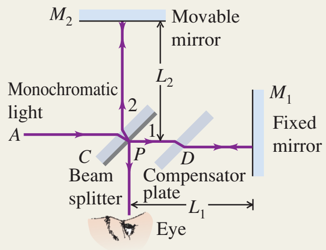
$$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \dots) \tag{35.17b}$$
(destructive reflection from thin film, no relative phase shift)

$$2t = \left(m + \tfrac{1}{2}\right)\lambda \quad (m = 0, 1, 2, \dots) \tag{35.18a}$$
(constructive reflection from thin film, half-cycle relative phase shift)

$$2t = m\lambda \quad (m = 0, 1, 2, \dots) \tag{35.18b}$$
(destructive reflection from thin film, half-cycle relative phase shift)



Michelson interferometer: The Michelson interferometer uses a monochromatic light source and can be used for high-precision measurements of wavelengths. Its original purpose was to detect motion of the earth relative to a hypothetical ether, the supposed medium for electromagnetic waves. The ether has never been detected, and the concept has been abandoned; the speed of light is the same relative to all observers. This is part of the foundation of the special theory of relativity.



BRIDGING PROBLEM Modifying a Two-Slit Experiment

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

See MasteringPhysics® study area for a Video Tutor solution.

IDENTIFY and SET UP

1. The oil layer acts as a thin film, so we must consider interference between light that reflects from the top and bottom surfaces of the oil. If a wavelength is prominent in the *transmitted* light, there is destructive interference for that wavelength in the *reflected* light.
2. Choose the appropriate interference equations that relate the thickness of the oil film and the wavelength of light. Take account of the indexes of refraction of the air, oil, and water.

EXECUTE

3. For part (a), find the wavelengths for which there is constructive interference as seen from above the oil film. Which of these are in the visible spectrum?
4. For part (b), find the visible wavelength for which there is destructive interference as seen from above the film. (This will ensure that there is substantial transmitted light at the wavelength.)

EVALUATE

5. If a diver below the water's surface shines a light up at the bottom of the oil film, at what wavelengths would there be constructive interference in the light that reflects back downward?

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **BIO**: Biosciences problems.

DISCUSSION QUESTIONS

Q35.1 A two-slit interference experiment is set up, and the fringes are displayed on a screen. Then the whole apparatus is immersed in the nearest swimming pool. How does the fringe pattern change?

Q35.2 Could an experiment similar to Young's two-slit experiment be performed with sound? How might this be carried out? Does it matter that sound waves are longitudinal and electromagnetic waves are transverse? Explain.

Q35.3 Monochromatic coherent light passing through two thin slits is viewed on a distant screen. Are the bright fringes equally spaced on the screen? If so, why? If not, which ones are closest to being equally spaced?

Q35.4 In a two-slit interference pattern on a distant screen, are the bright fringes midway between the dark fringes? Is this ever a good approximation?

Q35.5 Would the headlights of a distant car form a two-source interference pattern? If so, how might it be observed? If not, why not?

Q35.6 The two sources S_1 and S_2 shown in Fig. 35.3 emit waves of the same wavelength λ and are in phase with each other. Suppose S_1 is a weaker source, so that the waves emitted by S_1 have half the amplitude of the waves emitted by S_2. How would this affect the positions of the antinodal lines and nodal lines? Would there be total reinforcement at points on the antinodal curves? Would there be total cancellation at points on the nodal curves? Explain your answers.

Q35.7 Could the Young two-slit interference experiment be performed with gamma rays? If not, why not? If so, discuss differences in the experimental design compared to the experiment with visible light.

Q35.8 Coherent red light illuminates two narrow slits that are 25 cm apart. Will a two-slit interference pattern be observed when the light from the slits falls on a screen? Explain.

Q35.9 Coherent light with wavelength λ falls on two narrow slits separated by a distance d. If d is less than some minimum value, no dark fringes are observed. Explain. In terms of λ, what is this minimum value of d?

Q35.10 A fellow student, who values memorizing equations above understanding them, combines Eqs. (35.4) and (35.13) to "prove" that ϕ can *only* equal $2\pi m$. How would you explain to this student that ϕ can have values other than $2\pi m$?

Q35.11 If the monochromatic light shown in Fig. 35.5a were replaced by white light, would a two-slit interference pattern be seen on the screen? Explain.

Q35.12 In using the superposition principle to calculate intensities in interference patterns, could you add the intensities of the waves instead of their amplitudes? Explain.

Q35.13 A glass windowpane with a thin film of water on it reflects less than when it is perfectly dry. Why?

Q35.14 A *very* thin soap film ($n = 1.33$), whose thickness is much less than a wavelength of visible light, looks black; it appears to reflect no light at all. Why? By contrast, an equally thin layer of soapy water ($n = 1.33$) on glass ($n = 1.50$) appears quite shiny. Why is there a difference?

Q35.15 Interference can occur in thin films. Why is it important that the films be *thin*? Why don't you get these effects with a relatively *thick* film? Where should you put the dividing line between "thin" and "thick"? Explain your reasoning.

Q35.16 If we shine white light on an air wedge like that shown in Fig. 35.12, the colors that are weak in the light *reflected* from any point along the wedge are strong in the light *transmitted* through the wedge. Explain why this should be so.

Q35.17 Monochromatic light is directed at normal incidence on a thin film. There is destructive interference for the reflected light, so the intensity of the reflected light is very low. What happened to the energy of the incident light?

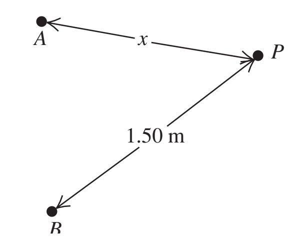
Q35.18 When a thin oil film spreads out on a puddle of water, the thinnest part of the film looks dark in the resulting interference pattern. What does this tell you about the relative magnitudes of the refractive indexes of oil and water?

EXERCISES

Section 35.1 Interference and Coherent Sources

35.1 • Two small stereo speakers A and B that are 1.40 m apart are sending out sound of wavelength 34 cm in all directions and all in phase. A person at point P starts out equidistant from both speakers and walks so that he is always 1.50 m from speaker B (Fig. E35.1). For what values of x will the sound this person hears be (a) maximally reinforced, (b) cancelled? Limit your solution to the cases where $x \leq 1.50$ m.

Figure **E35.1**



35.2 •• Two speakers that are 15.0 m apart produce in-phase sound waves of frequency 250.0 Hz in a room where the speed of sound is 340.0 m/s. A woman starts out at the midpoint between the two speakers. The room's walls and ceiling are covered with absorbers to eliminate reflections, and she listens with only one ear for best precision. (a) What does she hear: constructive or destructive interference? Why? (b) She now walks slowly toward one of the speakers. How far from the center must she walk before she first hears the sound reach a minimum intensity? (c) How far from the center must she walk before she first hears the sound maximally enhanced?

35.3 •• Two identical audio speakers connected to the same amplifier produce in-phase sound waves with a single frequency that can be varied between 300 and 600 Hz. The speed of sound is 340 m/s. You find that where you are standing, you hear minimum-intensity sound. (a) Explain why you hear minimum-intensity sound. (b) If one of the speakers is moved 39.8 cm toward you, the sound you hear has maximum intensity. What is the frequency of the sound? (c) How much closer to you from the position in part (b) must the speaker be moved to the next position where you hear maximum intensity?

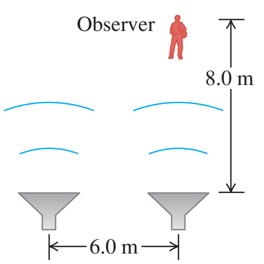
35.4 • **Radio Interference.** Two radio antennas A and B radiate in phase. Antenna B is 120 m to the right of antenna A. Consider point Q along the extension of the line connecting the antennas, a horizontal distance of 40 m to the right of antenna B. The frequency, and hence the wavelength, of the emitted waves can be varied. (a) What is the longest wavelength for which there will be destructive interference at point Q? (b) What is the longest wavelength for which there will be constructive interference at point Q?

35.5 •• 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?

35.6 • Two light sources can be adjusted to emit monochromatic light of any visible wavelength. The two sources are coherent, 2.04 μm apart, and in line with an observer, so that one source is 2.04 μm farther from the observer than the other. (a) For what visible wavelengths (380 to 750 nm) will the observer see the brightest light, owing to constructive interference? (b) How would your answers to part (a) be affected if the two sources were not in line with the observer, but were still arranged so that one source is 2.04 μm farther away from the observer than the other? (c) For what visible wavelengths will there be *destructive* interference at the location of the observer?

35.7 • 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. E35.7. (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 **E35.7**



35.8 •• Figure 35.3 shows the wave pattern produced by two identical, coherent sources emitting waves with wavelength λ and separated by a distance $d = 4\lambda$. (a) Explain why the positive y-axis above S_1 constitutes an antinodal curve with $m = +4$ and why the negative y-axis below S_2 constitutes an antinodal curve with $m = -4$. (b) Draw the wave pattern produced when the separation between the sources is reduced to 3λ. In your drawing, sketch all antinodal curves—that is, the curves on which $r_2 - r_1 = m\lambda$. Label each curve by its value of m. (c) In general, what determines the maximum (most positive) and minimum (most negative) values of the integer m that labels the antinodal lines? (d) Suppose the separation between the sources is increased to $7\frac{1}{2}\lambda$. How many antinodal curves will there be? To what values of m do they correspond? Explain your reasoning. (You should not have to make a drawing to answer these questions.)

Section 35.2 Two-Source Interference of Light

35.9 • Young's experiment is performed with light from excited helium atoms ($\lambda = 502$ nm). Fringes are measured carefully on a screen 1.20 m away from the double slit, and the center of the 20th fringe (not counting the central bright fringe) is found to be 10.6 mm from the center of the central bright fringe. What is the separation of the two slits?

35.10 •• Coherent light with wavelength 450 nm falls on a double slit. On a screen 1.80 m away, the distance between dark fringes is 4.20 mm. What is the separation of the slits?

35.11 •• 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?

35.12 •• If the entire apparatus of Exercise 35.11 (slits, screen, and space in between) is immersed in water, what then is the distance between the second and third dark lines?

35.13 •• 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?

35.14 • 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?

35.15 •• 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.)

35.16 •• Coherent light that contains two wavelengths, 660 nm (red) and 470 nm (blue), passes through two narrow slits separated by 0.300 mm, and the interference pattern is observed on a screen 5.00 m from the slits. What is the distance on the screen between the first-order bright fringes for the two wavelengths?

35.17 •• 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?

35.18 •• 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

35.19 •• 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 angle is 60.0°?

35.20 • Coherent sources A and B emit electromagnetic waves with wavelength 2.00 cm. Point P is 4.86 m from A and 5.24 m from B. What is the phase difference at P between these two waves?

35.21 • 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?

35.22 • Two slits spaced 0.260 mm apart are placed 0.700 m from a screen and illuminated by coherent light with a wavelength of 660 nm. 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 $I_0/2$?

35.23 •• Points A and B are 56.0 m apart along an east-west line. At each of these points, a radio transmitter is emitting a 12.5-MHz signal horizontally. These transmitters are in phase with each other and emit their beams uniformly in a horizontal plane. A receiver is taken 0.500 km north of the AB line and initially placed at point C, directly opposite the midpoint of AB. The receiver can be moved only along an east-west direction but, due to its limited sensitivity, it must always remain within a range so that the intensity of the signal it receives from the transmitter is no less than $\frac{1}{4}$ of its maximum value. How far from point C (along an east-west line) can the receiver be moved and always be able to pick up the signal?

35.24 • Consider two antennas separated by 9.00 m that radiate in phase at 120 MHz, as described in Exercise 35.5. 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

35.25 • 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?

35.26 •• **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.

35.27 •• Two rectangular pieces of plane glass are laid one upon the other on a table. A thin strip of paper is placed between them at one edge so that a very thin wedge of air is formed. The plates are illuminated at normal incidence by 546-nm light from a mercury-vapor lamp. Interference fringes are formed, with 15.0 fringes per centimeter. Find the angle of the wedge.

35.28 •• A plate of glass 9.00 cm long is placed in contact with a second plate and is held at a small angle with it by a metal strip

0.0800 mm thick placed under one end. The space between the plates is filled with air. The glass is illuminated from above with light having a wavelength in air of 656 nm. How many interference fringes are observed per centimeter in the reflected light?

35.29 •• 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.

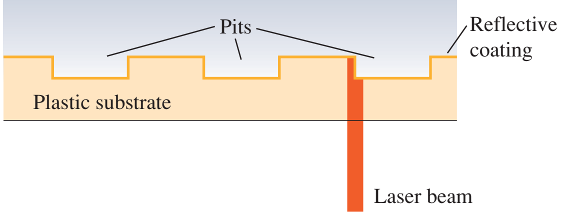
35.30 • A plastic film with index of refraction 1.85 is put on the surface of a car window to increase the reflectivity and thus to keep the interior of the car cooler. The window glass has index of refraction 1.52. (a) What minimum thickness is required if light with wavelength 550 nm in air reflected from the two sides of the film is to interfere constructively? (b) It is found to be difficult to manufacture and install coatings as thin as calculated in part (a). What is the next greatest thickness for which there will also be constructive interference?

35.31 • 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.

35.32 •• Light with wavelength 648 nm in air is incident perpendicularly from air on a film 8.76 μm thick and with refractive index 1.35. Part of the light is reflected from the first surface of the film, and part enters the film and is reflected back at the second surface, where the film is again in contact with air. (a) How many waves are contained along the path of this second part of the light in its round trip through the film? (b) What is the phase difference between these two parts of the light as they leave the film?

35.33 •• **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. E35.33). 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 **E35.33**



35.34 • What is the thinnest soap film (excluding the case of zero thickness) that appears black when illuminated with light with wavelength 480 nm? The index of refraction of the film is 1.33, and there is air on both sides of the film.

Section 35.5 The Michelson Interferometer

35.35 • 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 ($\lambda = 633$ nm) move across a line in the field of view?

35.36 • Jan first uses a Michelson interferometer with the 606-nm light from a krypton-86 lamp. He displaces the movable mirror away from him, counting 818 fringes moving across a line in his field of view. Then Linda replaces the krypton lamp with filtered 502-nm light from a helium lamp and displaces the movable mirror toward her. She also counts 818 fringes, but they move across the line in her field of view opposite to the direction they moved for Jan. Assume that both Jan and Linda counted to 818 correctly. (a) What distance did each person move the mirror? (b) What is the resultant displacement of the mirror?

PROBLEMS

35.37 ••• The radius of curvature of the convex surface of a planoconvex lens is 68.4 cm. The lens is placed convex side down on a perfectly flat glass plate that is illuminated from above with red light having a wavelength of 580 nm. Find the diameter of the second bright ring in the interference pattern.

35.38 •• Newton's rings can be seen 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 0.720 mm. If water ($n = 1.33$) now fills the space between the lens and the plate, what is the new diameter of this ring?

35.39 • 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?

35.40 •• 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?

35.41 •• 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 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?

35.42 •• In a setup similar to that of Problem 35.41, the glass has an index of refraction of 1.53, the plates are each 8.00 cm long, and the metal foil is 0.015 mm thick. The space between the plates is filled with a jelly whose refractive index is not known precisely, but is known to be greater than that of the glass. When you illuminate these plates from above with light of wavelength 525 nm, you observe a series of equally spaced dark fringes in the reflected light. You measure the spacing of these fringes and find that there are 10 of them every 6.33 mm. What is the index of refraction of the jelly?

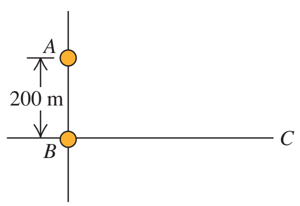
35.43 ••• Suppose you illuminate two thin slits by monochromatic coherent light in air and find that they produce their first interference *minima* at $\pm 35.20°$ on either side of the central bright spot. You then immerse these slits in a transparent liquid and illuminate them with the same light. Now you find that the first minima occur at $\pm 19.46°$ instead. What is the index of refraction of this liquid?

35.44 •• **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 $(20.0°\text{C})$, the first interference dark fringes occur at $\pm 32.5°$ from the original direction of the laser beam when viewed from some distance. If this sheet is now slowly heated up to 135°C, by how many degrees do these dark fringes change position? Do they move closer together or get farther apart? See Table 17.1 for pertinent information, and ignore any effects that might occur due to change in the thickness of the slits. (*Hint:* Since thermal expansion normally produces very small changes in length, you can use differentials to find the change in the angle.)

35.45 •• Two speakers, 2.50 m apart, are driven by the same audio oscillator so that each one produces a sound consisting of *two* distinct frequencies, 0.900 kHz and 1.20 kHz. The speed of sound in the room is 344 m/s. Find all the angles relative to the usual centerline in front of (and far from) the speakers at which *both* frequencies interfere constructively.

35.46 •• Two radio antennas radiating in phase are located at points A and B, 200 m apart (Fig. P35.46). 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. P35.46). 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 **P35.46**

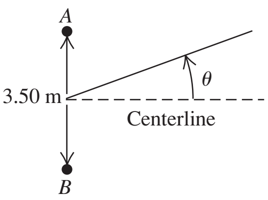


35.47 •• One round face of a 3.25-m, solid, cylindrical plastic pipe is covered with a thin black coating that completely blocks light. The opposite face is covered with a fluorescent coating that glows when it is struck by light. Two straight, thin, parallel scratches, 0.225 mm apart, are made in the 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?

35.48 • A uniform thin film of material of refractive index 1.40 coats a glass plate of refractive index 1.55. This film has the proper thickness to cancel normally incident light of wavelength 525 nm that strikes the film surface from air, but it is somewhat greater than the minimum thickness to achieve this cancellation. As time goes by, the film wears away at a steady rate of 4.20 nm per year. What is the minimum number of years before the reflected light of this wavelength is now enhanced instead of cancelled?

35.49 •• 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. P35.49) 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 **P35.49**



35.50 ••• **CP** The electric fields received at point P from two identical, coherent wave sources are $E_1(t) = E\cos(\omega t + \phi)$ and $E_2(t) = E\cos\omega t$. (a) Use the trigonometric identities in Appendix B to show that the resultant wave is $E_P(t) = 2E\cos(\phi/2)\cos(\omega t + \phi/2)$. (b) Show that the amplitude of this resultant wave is given by Eq. (35.7). (c) Use the result of part (a) to show that at an interference maximum, the amplitude of the resultant wave is in phase with the original waves $E_1(t)$ and $E_2(t)$. (d) Use the result of part (a) to show that near an interference minimum, the resultant wave is approximately $\frac{1}{4}$ cycle out of phase with either of the original waves. (e) Show that the *instantaneous* Poynting vector at point P has magnitude $S = 4\epsilon_0 cE^2\cos^2(\phi/2)\cos^2(\omega t + \phi/2)$ and that the *time-averaged* Poynting vector is given by Eq. (35.9).

35.51 •• **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 $(20.0°\text{C})$, 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.)

35.52 ••• **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 $2.00\ \text{W/m}^2$. (a) At how many other angles in the range $0° < \theta < 90°$ is the intensity also $2.00\ \text{W/m}^2$? (b) Find the four smallest angles in the range $0° < \theta < 90°$ for which the intensity is $2.00\ \text{W/m}^2$. (c) What is the intensity at a point on the circle at an angle of 4.65° from the centerline?

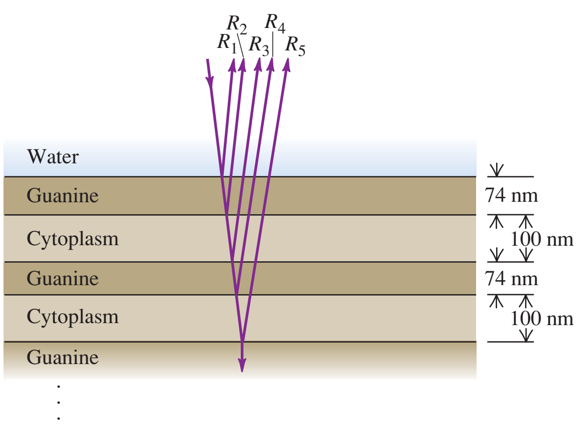
35.53 •• Consider a two-slit interference pattern, for which the intensity distribution is given by Eq. (35.14). Let θ_m be the angular position of the mth bright fringe, where the intensity is I_0. Assume that θ_m is small, so that $\sin\theta_m \cong \theta_m$. Let $\theta_m{}^+$ and $\theta_m{}^-$ be the two angles on either side of θ_m for which $I = \frac{1}{2}I_0$. The quantity $\Delta\theta_m = |\theta_m{}^+ - \theta_m{}^-|$ is the half-width of the mth fringe. Calculate $\Delta\theta_m$. How does $\Delta\theta_m$ depend on m?

35.54 •• 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?

35.55 ••• A source S of monochromatic light and a detector D are both located in air a distance h above a horizontal plane sheet of glass and are separated by a horizontal distance x. Waves reaching D directly from S interfere with waves that reflect off the glass. The distance x is small compared to h so that the reflection is at close to normal incidence. (a) Show that the condition for constructive interference is $\sqrt{x^2 + 4h^2} - x = \left(m + \frac{1}{2}\right)\lambda$, and the condition for destructive interference is $\sqrt{x^2 + 4h^2} - x = m\lambda$. (*Hint:* Take into account the phase change on reflection.) (b) Let $h = 24$ cm and $x = 14$ cm. What is the longest wavelength for which there will be constructive interference?

35.56 •• **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. P35.56). 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. P35.56, 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 **P35.56**



35.57 • Two thin parallel slits are made in an opaque sheet of film. When a monochromatic beam of light is shone through them at normal incidence, the first bright fringes in the transmitted light occur in air at $\pm 18.0°$ with the original direction of the light beam on a distant screen when the apparatus is in air. When the apparatus is immersed in a liquid, the same bright fringes now occur at $\pm 12.6°$. Find the index of refraction of the liquid.

35.58 •• 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?

35.59 ••• In a Young's two-slit experiment a piece of glass with an index of refraction n and a thickness L is placed in front of the upper slit. (a) Describe qualitatively what happens to the interference pattern. (b) Derive an expression for the intensity I of the light at points on a screen as a function of n, L, and θ. Here θ is the usual angle measured from the center of the two slits. That is, determine the equation analogous to Eq. (35.14) but that also involves L and n for the glass plate. (c) From your result in part (b) derive an expression for the values of θ that locate the maxima in the interference pattern [that is, derive an equation analogous to Eq. (35.4)].

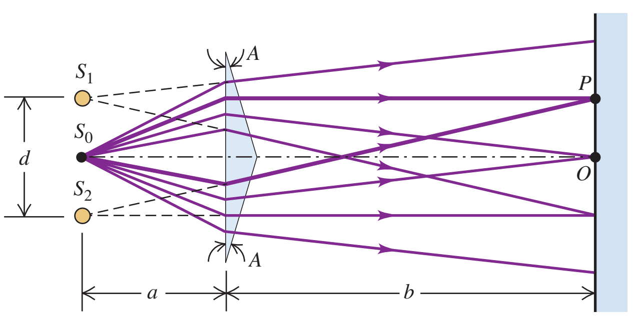
35.60 •• After a laser beam passes through two thin parallel slits, the first completely dark fringes occur at $\pm 19.0°$ with the original direction of the beam, as viewed on a screen far from the slits. (a) What is the ratio of the distance between the slits to the wavelength of the light illuminating 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?

CHALLENGE PROBLEMS

35.61 ••• **CP** The index of refraction of a glass rod is 1.48 at $T = 20.0°\text{C}$ and varies linearly with temperature, with a coefficient of $2.50 \times 10^{-5}/\text{C}°$. The coefficient of linear expansion of the glass is $5.00 \times 10^{-6}/\text{C}°$. 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 $5.00\ \text{C}°/\text{min}$. The light source has wavelength $\lambda = 589$ nm, and the rod initially is at $T = 20.0°\text{C}$. How many fringes cross the field of view each minute?

35.62 ••• **CP** Figure P35.62 shows an interferometer known as *Fresnel's biprism.* The magnitude of the prism angle A is

Figure **P35.62**



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 $n = 1.50$.

Answers

Chapter Opening Question ?

The colors appear due to constructive interference between light waves reflected from the outer and inner surfaces of the soap bubble. The thickness of the bubble walls at each point determines the wavelength of light for which the most constructive interference occurs and hence the color that appears the brightest at that point (see Example 35.4 in Section 35.4).

Test Your Understanding Questions

35.1 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.

35.2 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 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 Answers: (i) and (iii) Benzene has a larger index of refraction than air, so light that reflects off the upper surface of the benzene undergoes a half-cycle phase shift. Fluorite has a *smaller* index of refraction than benzene, so light that reflects off the benzene–fluorite interface does not undergo a phase shift. Hence the equation for constructive reflection is Eq. (35.18a), $2t = \left(m + \frac{1}{2}\right)\lambda$, which we can rewrite as $t = \left(m + \frac{1}{2}\right)\lambda/2 = \left(m + \frac{1}{2}\right)(400\text{ mm})/2 = 100$ nm, 300 nm, 500 nm, ...

35.5 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.

Bridging Problem

Answers: (a) 441 nm **(b)** 551 nm

36 DIFFRACTION

LEARNING GOALS

By studying this chapter, you will learn:

- What happens when coherent light shines on an object with an edge or aperture.
- How to understand the diffraction pattern formed when coherent light passes through a narrow slit.
- How to calculate the intensity at various points in a single-slit diffraction pattern.
- What happens when coherent light shines on an array of narrow, closely spaced slits.
- How scientists use diffraction gratings for precise measurements of wavelength.
- How x-ray diffraction reveals the arrangement of atoms in a crystal.
- How diffraction sets limits on the smallest details that can be seen with a telescope.

? The laser used to read a DVD has a wavelength of 650 nm, while the laser used to read a Blu-ray disc has a shorter 405-nm wavelength. How does this make it possible for a Blu-ray disc to hold more information than a DVD?

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 around a corner or the speech of a person whose back is turned to you. What may surprise you (and certainly surprised many scientists of the early 19th century) is that *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 behave precisely according to 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.* We'll find that 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 the resulting pattern of light and dark is a result of interference among the waves emanating from these sources.

Light emerging from arrays of apertures also forms patterns whose character 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 recorded on photographic film and reproduced. When properly illuminated, it forms a three-dimensional image of the original object.

36.1 Fresnel and Fraunhofer Diffraction

According to geometric optics, when an opaque object is placed between a point light source and a screen, as in Fig. 36.1, the shadow of the object forms a perfectly sharp line. No light at all strikes the screen at points within the shadow, and the area outside the shadow is illuminated nearly uniformly. But as we saw in Chapter 35, the *wave* nature of light causes effects that can't be understood with geometric optics. An important class of such effects occurs when light strikes a barrier that has an aperture or an edge. The interference patterns formed in such a situation are grouped under the heading **diffraction.**

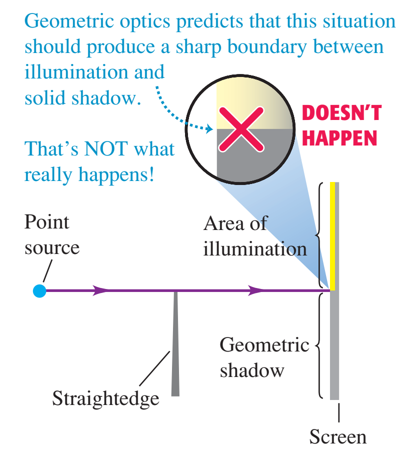
Figure 36.2 shows an example of diffraction. The photograph in Fig. 36.2a was made by placing a razor blade halfway between a pinhole, illuminated by monochromatic light, and a photographic film. The film recorded the shadow cast by the blade. Figure 36.2b is an enlargement of a region near the shadow of the right edge of the blade. The position of the *geometric* shadow line is indicated by arrows. The area outside the geometric shadow is bordered by alternating bright and dark bands. There is some light in the shadow region, although this is not very visible in the photograph. The first bright band in Fig. 36.2b, just to the right of the geometric shadow, is considerably brighter than in the region of uniform illumination to the extreme right. This simple experiment gives us some idea of the richness and complexity of what might seem to be a simple idea, the casting of a shadow by an opaque object.

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.

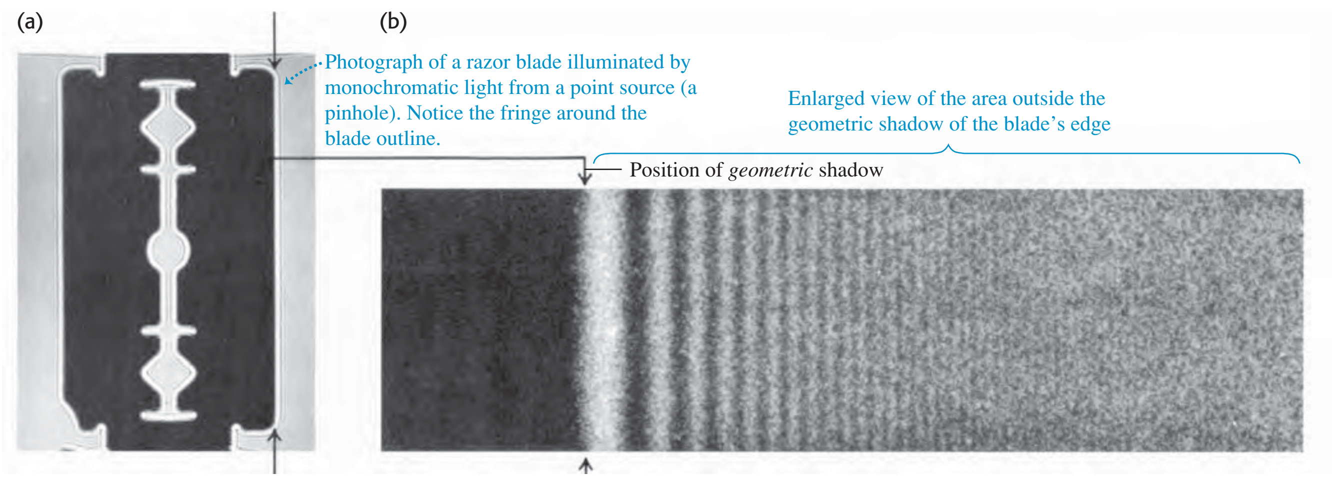
Diffraction and Huygens's Principle

We can analyze diffraction patterns using Huygens's principle (see Section 33.7). 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 combine all the individual displacements produced by these secondary waves, using the superposition principle and taking into account their amplitudes and relative phases.

36.1 A point source of light illuminates a straightedge.



36.2 An example of diffraction.



In Fig. 36.1, both the point source and the screen are relatively close to the obstacle forming the diffraction pattern. This situation is described as *near-field diffraction* or **Fresnel diffraction,** pronounced "Freh-nell" (after the French scientist Augustin Jean Fresnel, 1788–1827). By contrast, we use the term **Fraunhofer diffraction** (after the German physicist Joseph von Fraunhofer, 1787–1826) for situations in which the source, obstacle, and screen are far enough apart that we can consider all lines from the source to the obstacle to be parallel, and can likewise consider all lines from the obstacle to a given point on the screen to be parallel. We will restrict the following discussion to Fraunhofer diffraction, which is usually simpler to analyze in detail than Fresnel diffraction.

Diffraction is sometimes described as "the bending of light around an obstacle." But the process that causes diffraction is present in the propagation of *every* wave. When part of the wave is cut off by some obstacle, we observe diffraction effects that result from interference of the remaining parts of the wave fronts. Optical instruments typically use only a limited portion of a wave; for example, a telescope uses only the part of a wave that is admitted by its objective lens or mirror. Thus diffraction plays a role in nearly all optical phenomena.

Finally, we emphasize that there is no fundamental distinction between *interference* and *diffraction.* In Chapter 35 we used the term *interference* for effects involving waves from a small number of sources, usually two. *Diffraction* usually involves a *continuous* distribution of Huygens's wavelets across the area of an aperture, or a very large number of sources or apertures. But both interference and diffraction are consequences of superposition and Huygens's principle.

Test Your Understanding of Section 36.1 Can *sound* waves undergo diffraction around an edge?

36.2 Diffraction from a Single Slit

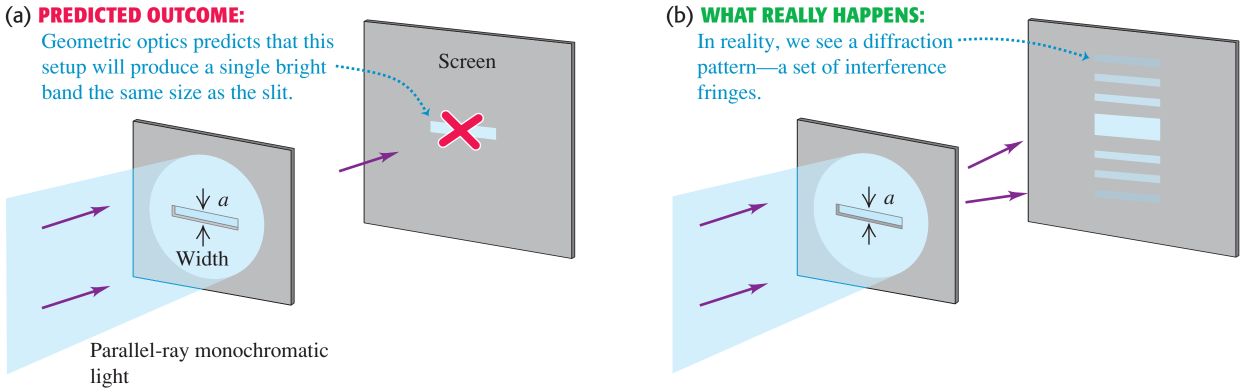
MasteringPHYSICS

PhET: Wave Interference
ActivPhysics 16.6: Single-Slit Diffraction

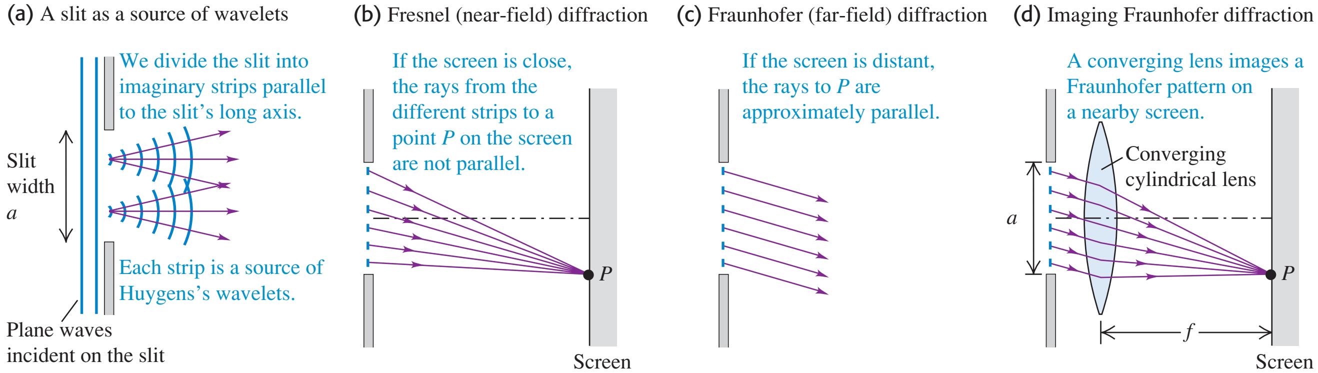
In this section we'll discuss the diffraction pattern formed by plane-wave (parallel-ray) monochromatic light when it emerges from a long, narrow slit, as shown in Fig. 36.3. We call the narrow dimension the *width,* even though in this figure it is a vertical dimension.

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

36.3 (a) The "shadow" of a horizontal slit as incorrectly predicted by geometric optics. (b) A horizontal slit actually produces a diffraction pattern. The slit width has been greatly exaggerated.



36.4 Diffraction by a single rectangular slit. The long sides of the slit are perpendicular to the figure.



transmitted beam is in the central bright band, whose width is *inversely* proportional to the width of the slit. In general, the smaller the width of 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 corresponds to the screen.

Single-Slit Diffraction: Locating the Dark Fringes

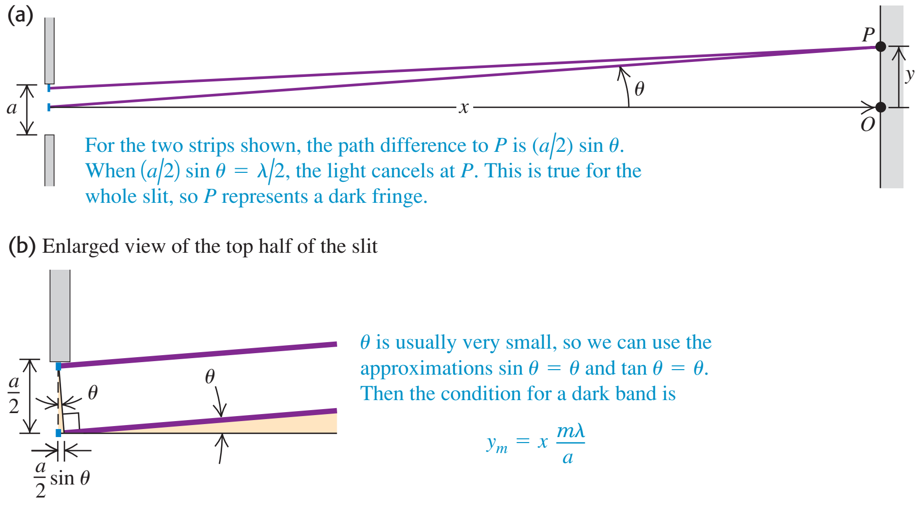
Figure 36.4 shows a side view of the same setup; the long sides of the slit are perpendicular to the figure, and plane waves are incident on the slit from the left. According to Huygens's principle, each element of area of the slit opening can be considered as a source of secondary waves. In particular, imagine dividing the slit into several narrow strips of equal width, parallel to the long edges and perpendicular to the page. Figure 36.4a shows two such strips. Cylindrical secondary wavelets, shown in cross section, spread out from each strip.

In Fig. 36.4b a screen is placed to the right of the slit. We can calculate the resultant intensity at a point P on the screen by adding the contributions from the individual wavelets, taking proper account of their various phases and amplitudes. It's easiest to do this calculation if we assume that the screen is far enough away that all the rays from various parts of the slit to a particular point P on the screen are parallel, as in Fig. 36.4c. An equivalent situation is Fig. 36.4d, in which the rays to the lens are parallel and the lens forms a reduced image of the same pattern that would be formed on an infinitely distant screen without the lens. We might expect that the various light paths through the lens would introduce additional phase shifts, but in fact it can be shown that all the paths have *equal* phase shifts, so this is not a problem.

The situation of Fig. 36.4b is Fresnel diffraction; those in Figs. 36.4c and 36.4d, where the outgoing rays are considered parallel, are Fraunhofer diffraction. We can derive quite simply the most important characteristics of the Fraunhofer diffraction pattern from a single slit. First consider two narrow strips, one just below the top edge of the drawing of the slit and one at its center, shown in end view in Fig. 36.5. The difference in path length to point P is $(a/2)\sin\theta$, where a is the slit width and θ is the angle between the perpendicular to the slit 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.

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.



Hence the combined light from the entire slit completely cancels at P, giving a dark fringe in the interference pattern. A dark fringe occurs whenever

$$\frac{a}{2}\sin\theta = \pm\frac{\lambda}{2} \quad \text{or} \quad \sin\theta = \pm\frac{\lambda}{a} \tag{36.1}$$

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 screen 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

$$\sin\theta = \frac{m\lambda}{a} \qquad (m = \pm 1, \pm 2, \pm 3, \dots) \qquad \text{(dark fringes in single-slit diffraction)} \tag{36.2}$$

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. We also 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). The central bright fringe is wider than the other bright fringes, as Fig. 36.3b shows. In the small-angle approximation that we will use below, it is exactly *twice* as wide.

With light, the wavelength λ is of the order of 500 nm $= 5 \times 10^{-7}$ m. This is often much smaller than the slit width a; a typical slit width is 10^{-2} cm $= 10^{-4}$ m. Therefore the values of θ in Eq. (36.2) are often so small that the approximation $\sin\theta \approx \theta$ (where θ is in radians) is a very good one. In that case we can rewrite this equation as

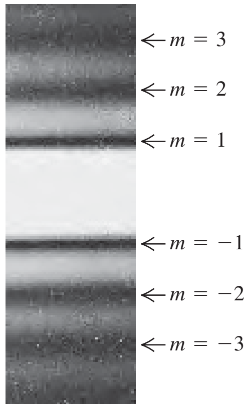
$$\theta = \frac{m\lambda}{a} \qquad (m = \pm 1, \pm 2, \pm 3, \dots) \qquad \text{(for small angles } \theta\text{)}$$

where θ is 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), and 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 $m = \pm1,\ \pm2$, and ±3 minima labeled.

36.6 Photograph of the Fraunhofer diffraction pattern of a single horizontal slit.



CAUTION **Single-slit diffraction vs. two-slit interference** Equation (36.3) has the same form as the equation for the two-slit pattern, Eq. (35.6), except that in Eq. (36.3) we use x rather than R for the distance to the screen. But Eq. (36.3) gives the positions of the *dark* fringes in a *single-slit* pattern rather than the *bright* fringes in a *double-slit* pattern. Also, $m = 0$ in Eq. (36.2) is *not* a dark fringe. Be careful!

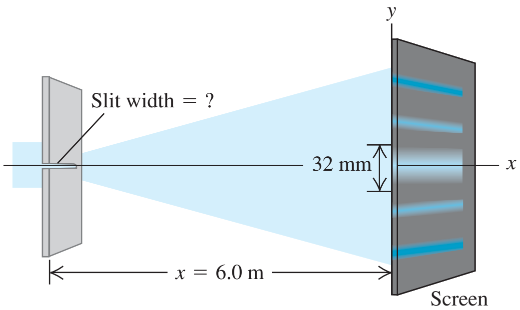
Example 36.1 Single-slit diffraction

You pass 633-nm laser light through a narrow slit and observe the diffraction pattern on a screen 6.0 m away. The distance on the screen between the centers of the first minima on either side of the central bright fringe is 32 mm (Fig. 36.7). How wide is the slit?

SOLUTION

IDENTIFY and SET UP: This problem involves the relationship between the positions of dark fringes in a single-slit diffraction pattern and the slit width a (our target variable). The distances between fringes on the screen are much smaller than the slit-to-screen distance, so the angle θ shown in Fig. 36.5a is very small and we can use Eq. (36.3) to solve for a.

36.7 A single-slit diffraction experiment.



EXECUTE: The first minimum corresponds to $m = 1$ in Eq. (36.3). The distance y_1 from the central maximum to the first minimum on either side is half the distance between the two first minima, so $y_1 = (32\text{ mm})/2 = 16$ mm. Solving Eq. (36.3) for a, we find

$$a = \frac{x\lambda}{y_1} = \frac{(6.0\text{ m})(633 \times 10^{-9}\text{ m})}{16 \times 10^{-3}\text{ m}} = 2.4 \times 10^{-4}\text{ m} = 0.24\text{ mm}$$

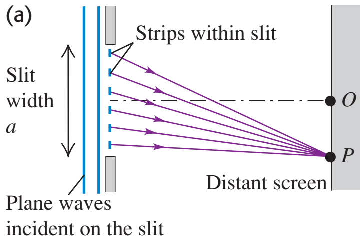
EVALUATE: The angle θ is small only if the wavelength is small compared to the slit width. Since $\lambda = 633\text{ nm} = 6.33 \times 10^{-7}$ m and we have found $a = 0.24\text{ mm} = 2.4 \times 10^{-4}$ m, our result is consistent with this: The wavelength is $(6.33 \times 10^{-7}\text{ m})/(2.4 \times 10^{-4}\text{ m}) = 0.0026$ as large as the slit width. Can you show that the distance between the *second* minima on either side is $2(32\text{ mm}) = 64$ mm, and so on?

Test Your Understanding of Section 36.2 Rank the following single-slit diffraction experiments in order of the size of the angle from the center of the diffraction pattern to the first dark fringe, from largest to smallest: (i) wavelength 400 nm, slit width 0.20 mm; (ii) wavelength 600 nm, slit width 0.20 mm; (iii) wavelength 400 nm, slit width 0.30 mm; (iv) wavelength 600 nm, slit width 0.30 mm.

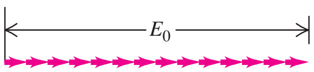
36.3 Intensity in the Single-Slit Pattern

We can derive an expression for the intensity distribution for the single-slit diffraction pattern by the same phasor-addition method that we used in Section 35.3 to obtain Eqs. (35.10) and (35.14) 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

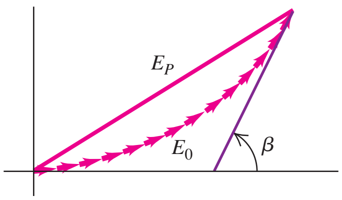
36.8 Using phasor diagrams to find the amplitude of the $\vec{E}$ field in single-slit diffraction. Each phasor represents the $\vec{E}$ field from a single strip within the slit.



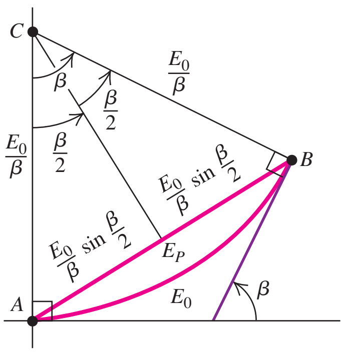
(b) At the center of the diffraction pattern (point O), the phasors from all strips within the slit are in phase.



(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



each individual 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 from the top strip of Fig. 36.8a and the wave from the bottom strip; that is, β is the phase of the wave received at P from the top strip with respect to the wave received at P from the bottom strip.

We may imagine dividing the slit into narrower and narrower strips. In the limit that there is an infinite number of infinitesimally narrow strips, the curved trail of phasors becomes an *arc of a circle* (Fig. 36.8d), with arc length equal to the length E_0 in Fig. 36.8b. The center C of this arc is found by constructing perpendiculars at A and B. From the relationship among arc length, radius, and angle, the radius of the arc is E_0/β; the amplitude E_P of the resultant electric field at P is equal to the chord AB, which is $2(E_0/\beta)\sin(\beta/2)$. (Note that β *must* be in radians!) We then have

$$E_P = E_0\frac{\sin(\beta/2)}{\beta/2} \qquad \text{(amplitude in single-slit diffraction)} \qquad (36.4)$$

The intensity at each point on the screen is proportional to the square of the amplitude given by Eq. (36.4). If I_0 is the intensity in the straight-ahead direction where $\theta = 0$ and $\beta = 0$, then the intensity I at any point is

$$I = I_0\left[\frac{\sin(\beta/2)}{\beta/2}\right]^2 \qquad \text{(intensity in single-slit diffraction)} \qquad (36.5)$$

We can express the phase difference β in terms of geometric quantities, as we did for the two-slit pattern. From Eq. (35.11) the phase difference is $2\pi/\lambda$ times the path difference. Figure 36.5 shows that the path difference between the ray from the top of the slit and the ray from the middle of the slit is $(a/2)\sin\theta$. The path difference between the rays from the top of the slit and the bottom of the slit is twice this, so

$$\beta = \frac{2\pi}{\lambda}a\sin\theta \qquad (36.6)$$

and Eq. (36.5) becomes

$$I = I_0\left\{\frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda}\right\}^2 \qquad \text{(intensity in single-slit diffraction)} \qquad (36.7)$$

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, using 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

$$\frac{a\sin\theta}{\lambda} = m \qquad (m = \pm1, \pm2, \ldots)$$

$$\sin\theta = \frac{m\lambda}{a} \qquad (m = \pm1, \pm2, \ldots) \tag{36.8}$$

This agrees with our previous result, Eq. (36.2). Note again that $\beta = 0$ (corresponding to $\theta = 0$) is *not* a minimum. Equation (36.5) is indeterminate at $\beta = 0$, but we can evaluate the limit as $\beta \rightarrow 0$ 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 $\beta = \pm\pi, \pm3\pi, \pm5\pi$, or in general,

$$\beta \approx \pm(2m + 1)\pi \qquad (m = 0, 1, 2, \ldots) \tag{36.9}$$

This is *approximately* correct, but because of the factor $(\beta/2)^2$ in the denominator of Eq. (36.5), the maxima don't occur precisely at these points. When we take the derivative of Eq. (36.5) with respect to β and set it equal to zero to try to find the maxima and minima, we get a transcendental equation that has to be solved numerically. In fact there is *no* maximum near $\beta = \pm\pi$. The first maxima on either side of the central maximum, near $\beta = \pm3\pi$, actually occur at $\pm2.860\pi$. The second side maxima, near $\beta = \pm5\pi$, are actually at $\pm4.918\pi$, and so on. The error in Eq. (36.9) vanishes in the limit of large m—that is, for intensity maxima far from the center of the pattern.

To find the intensities at the side maxima, we substitute these values of β back into Eq. (36.5). Using the approximate expression in Eq. (36.9), we get

$$I_m \approx \frac{I_0}{\left(m + \frac{1}{2}\right)^2\pi^2} \tag{36.10}$$

where I_m is the intensity of the mth side maximum and I_0 is the intensity of the central maximum. Equation (36.10) gives the series of intensities

$$0.0450I_0 \qquad 0.0162I_0 \qquad 0.0083I_0$$

and so on. As we have pointed out, this equation is only approximately correct. The actual intensities of the side maxima turn out to be

$$0.0472I_0 \qquad 0.0165I_0 \qquad 0.0083I_0 \qquad \ldots$$

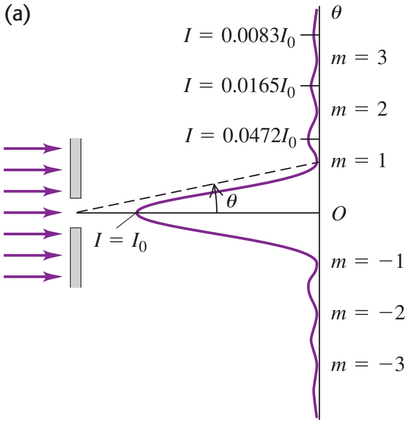
Note that the intensities of the side maxima decrease very rapidly, as Fig. 36.9a also shows. Even the first side maxima have less than 5% of the intensity of the central maximum.

Width of the Single-Slit Pattern

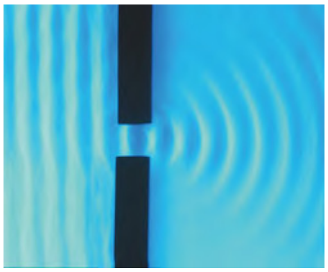
For small angles the angular spread of the diffraction pattern is inversely proportional to the slit width a or, more precisely, to the ratio of a to the wavelength λ. Figure 36.10 shows graphs of intensity I as a function of the angle θ for three values of the ratio a/λ.

36.9 (a) Intensity versus angle in single-slit diffraction. The values of m label intensity minima given by Eq. (36.8). Most of the wave power goes into the central intensity peak (between the $m = 1$ and $m = -1$ intensity minima). (b) These water waves passing through a small aperture behave exactly like light waves in single-slit diffraction. Only the diffracted waves within the central intensity peak are visible; the waves at larger angles are too faint to see.

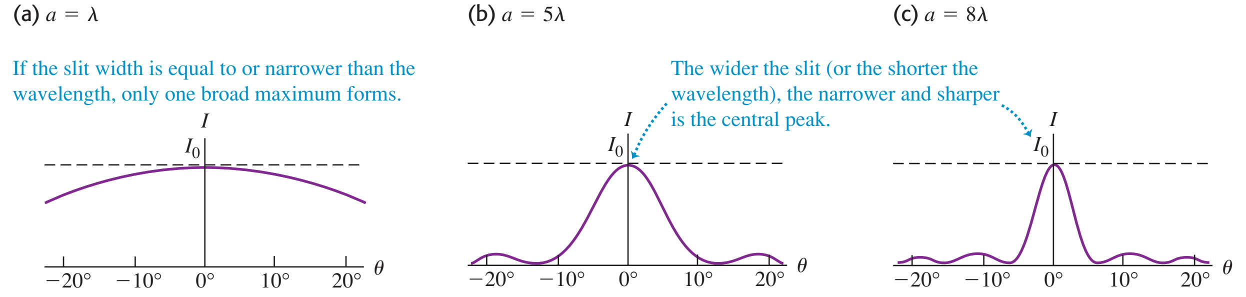
(a)



(b)



36.10 The single-slit diffraction pattern depends on the ratio of the slit width a to the wavelength 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 minimum beside the central maximum, corresponding to $\beta/2 = \pi$, is, from Eq. (36.7),

$$\theta_1 = \frac{\lambda}{a} \tag{36.11}$$

This characterizes the width (angular spread) of the central maximum, and we see that it is *inversely* proportional to the slit width a. When the small-angle approximation is valid, the central maximum is exactly twice as wide as each side maximum. When a is of the order of a centimeter or more, θ_1 is so small that we can consider practically all the light to be concentrated at the geometrical focus. But when a is less than λ, the central maximum spreads over 180°, and the fringe pattern is not seen at all.

It's important to keep in mind that diffraction occurs for *all* kinds of waves, not just light. Sound waves undergo diffraction when they pass through a slit or aperture such as an ordinary doorway. The sound waves used in speech have wavelengths of about a meter or greater, and a typical doorway is less than 1 m wide; in this situation, a is less than λ, and the central intensity maximum extends over 180°. This is why the sounds coming through an open doorway can easily be heard by an eavesdropper hiding out of sight around the corner. In the same way, sound waves can bend around the head of an instructor who faces the blackboard while lecturing (Fig. 36.11). By contrast, there is essentially no diffraction of visible light through a doorway because the width a is very much greater than the wavelength λ (of order 5×10^{-7} m). You can *hear* around corners because typical sound waves have relatively long wavelengths; you cannot *see* around corners because the wavelength of visible light is very short.

36.11 The sound waves used in speech have a long wavelength (about 1 m) and can easily bend around this instructor's head. By contrast, light waves have very short wavelengths and undergo very little diffraction. Hence you can't *see* around his head!

Example 36.2 Single-slit diffraction: Intensity I

(a) The intensity at the center of a single-slit diffraction pattern is I_0. What is the intensity at a point in the pattern where there is a 66-radian phase difference between wavelets from the two edges of the slit? (b) If this point is 7.0° away from the central maximum, how many wavelengths wide is the slit?

SOLUTION

IDENTIFY and SET UP: In our analysis of Fig. 36.8 we used the symbol β for the phase difference between wavelets from the two edges of the slit. In part (a) we use Eq. (36.5) to find the intensity I at the point in the pattern where $\beta = 66$ rad. In part (b) we need to find the slit width a as a multiple of the wavelength λ so our target variable is a/λ. We are given the angular position θ of the point where $\beta = 66$ rad, so we can use Eq. (36.6) to solve for a/λ.

EXECUTE: (a) We have $\beta/2 = 33$ rad, so from Eq. (36.5),

$$I = I_0\left[\frac{\sin(33\text{ rad})}{33\text{ rad}}\right]^2 = (9.2 \times 10^{-4})I_0$$

(b) From Eq. (36.6),

$$\frac{a}{\lambda} = \frac{\beta}{2\pi\sin\theta} = \frac{66\text{ rad}}{(2\pi\text{ rad})\sin 7.0°} = 86$$

For example, for 550-nm light the slit width is $a = (86)(550\text{ nm}) = 4.7 \times 10^{-5}\text{ m} = 0.047$ mm, or roughly $\frac{1}{20}$ mm.

EVALUATE: To what point in the diffraction pattern does this value of β correspond? To find out, note that $\beta = 66$ rad is approximately equal to 21π. This is an odd multiple of π, corresponding to the form $(2m + 1)\pi$ found in Eq. (36.9) for the intensity *maxima*. Hence $\beta = 66$ rad corresponds to a point near the tenth $(m = 10)$ maximum. This is well beyond the range shown in Fig. 36.9a, which shows only maxima out to $m = \pm 3$.

Example 36.3 Single-slit diffraction: Intensity II

In the experiment described in Example 36.1 (Section 36.2), the intensity at the center of the pattern is I_0. What is the intensity at a point on the screen 3.0 mm from the center of the pattern?

SOLUTION

IDENTIFY and SET UP: This is similar to Example 36.2, except that we are not given the value of the phase difference β at the point in question. We use geometry to determine the angle θ for our point and then use Eq. (36.7) to find the intensity I (the target variable).

EXECUTE: Referring to Fig. 36.5a, we have $y = 3.0$ mm and $x = 6.0$ m, so $\tan\theta = y/x = (3.0 \times 10^{-3}\text{ m})/(6.0\text{ m}) = 5.0 \times 10^{-4}$. This is so small that the values of $\tan\theta$, $\sin\theta$, and θ (in radians) are all nearly the same. Then, using Eq. (36.7),

$$\frac{\pi a \sin\theta}{\lambda} = \frac{\pi(2.4 \times 10^{-4}\text{ m})(5.0 \times 10^{-4})}{6.33 \times 10^{-7}\text{ m}} = 0.60$$

$$I = I_0\left(\frac{\sin 0.60}{0.60}\right)^2 = 0.89 I_0$$

EVALUATE: Figure 36.9a shows that an intensity this high can occur only within the central intensity maximum. This checks out; from Example 36.1, the first intensity minimum ($m = 1$ in Fig. 36.9a) is $(32\text{ mm})/2 = 16$ mm from the center of the pattern, so the point in question here at $y = 3$ mm does, indeed, lie within the central maximum.

Test Your Understanding of Section 36.3 Coherent electromagnetic radiation is sent through a slit of width 0.0100 mm. For which of the following wavelengths will there be *no* points in the diffraction pattern where the intensity is zero? (i) blue light of wavelength 500 nm; (ii) infrared light of wavelength 10.6 μm; (iii) microwaves of wavelength 1.00 mm; (iv) ultraviolet light of wavelength 50.0 nm.

36.4 Multiple Slits

In Sections 35.2 and 35.3 we analyzed interference from two point sources or 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.

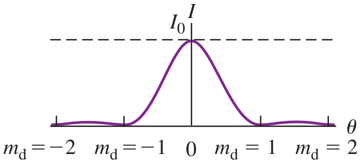
Two Slits of Finite Width

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.

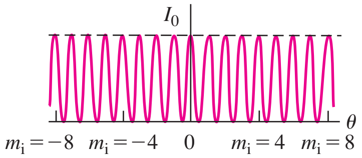
Figure 36.12a shows the intensity in a single-slit diffraction pattern with slit width a. The *diffraction minima* are labeled by the integer $m_{\text{d}} = \pm 1, \pm 2, \ldots$ ("d" for "diffraction"). Figure 36.12b shows the pattern formed by two very narrow slits with distance d between slits, where 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 $m_{\text{i}} = 0, \pm 1, \pm 2, \ldots$ ("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

36.12 Finding the intensity pattern for two slits of finite width.

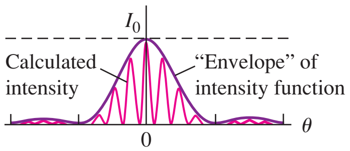
(a) Single-slit diffraction pattern for a slit width a



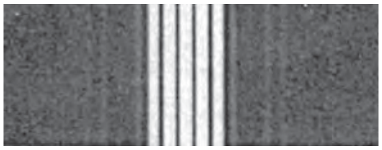
(b) Two-slit interference pattern for narrow slits whose separation d is four times the width of the slit in (a)



(c) Calculated intensity pattern for two slits of width a and separation $d = 4a$, including both interference and diffraction effects

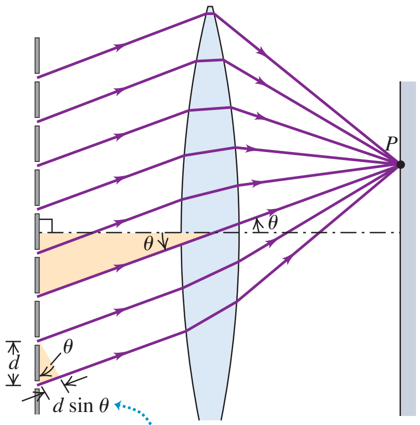


(d) Actual photograph of the pattern calculated in (c)



For $d = 4a$, every fourth interference maximum at the sides ($m_i = \pm 4, \pm 8, \ldots$) is missing.

36.13 Multiple-slit diffraction. Here a lens is used to give a Fraunhofer pattern on a nearby screen, as in Fig. 36.4d.



Maxima occur where the path difference for adjacent slits is a whole number of wavelengths: $d \sin\theta = m\lambda$.

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):

$$I = I_0 \cos^2 \frac{\phi}{2} \left[\frac{\sin(\beta/2)}{\beta/2} \right]^2 \qquad \text{(two slits of finite width)} \tag{36.12}$$

where, as before,

$$\phi = \frac{2\pi d}{\lambda} \sin\theta \qquad \beta = \frac{2\pi a}{\lambda} \sin\theta$$

Note that in Fig. 36.12c, every fourth interference maximum at the sides is *missing* because these interference maxima ($m_i = \pm 4, \pm 8, \ldots$) coincide with diffraction minima ($m_d = \pm 1, \pm 2, \ldots$). This can also be seen in Fig. 36.12d, which is a photograph of an actual pattern with $d = 4a$. You should be able to convince yourself that there will be "missing" maxima whenever d is an integer multiple of a.

Figures 36.12c and 36.12d show that as you move away from the central bright maximum of the two-slit pattern, the intensity of the maxima decreases. This is a result of the single-slit modulating pattern shown in Fig. 36.12a; mathematically, the decrease in intensity arises from the factor $(\beta/2)^2$ in the denominator of Eq. (36.12). This decrease in intensity can also be seen 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

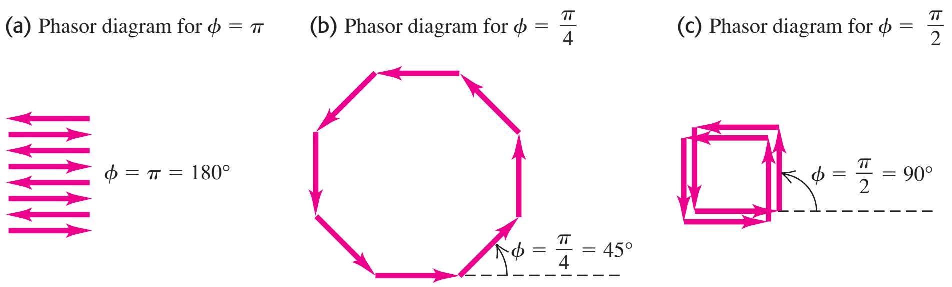
Next let's consider patterns produced by *several* very narrow slits. As we will see, systems of narrow slits are of tremendous practical importance in *spectroscopy*, the determination of the particular wavelengths of light coming from a source. Assume that each slit is narrow in comparison to the wavelength, so its diffraction pattern spreads out nearly uniformly. Figure 36.13 shows an array of eight narrow slits, with distance d between adjacent slits. Constructive interference occurs for rays at angle θ to the normal that arrive at point P with a path difference between adjacent slits equal to an integer number of wavelengths:

$$d \sin\theta = m\lambda \qquad (m = 0, \pm 1, \pm 2, \ldots)$$

This means that reinforcement occurs when the phase difference ϕ at P for light from adjacent slits is an integer multiple of 2π. That is, the maxima in the pattern occur at the *same* positions as for *two* slits with the same spacing. To this extent the pattern resembles the two-slit pattern.

But what happens *between* the maxima? 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

36.14 Phasor diagrams for light passing through eight narrow slits. Intensity maxima occur when the phase difference $\phi = 0, 2\pi, 4\pi, \ldots$. Between the maxima at $\phi = 0$ and $\phi = 2\pi$ are seven minima, corresponding to $\phi = \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2$, and $7\pi/4$. Can you draw phasor diagrams for the other minima?



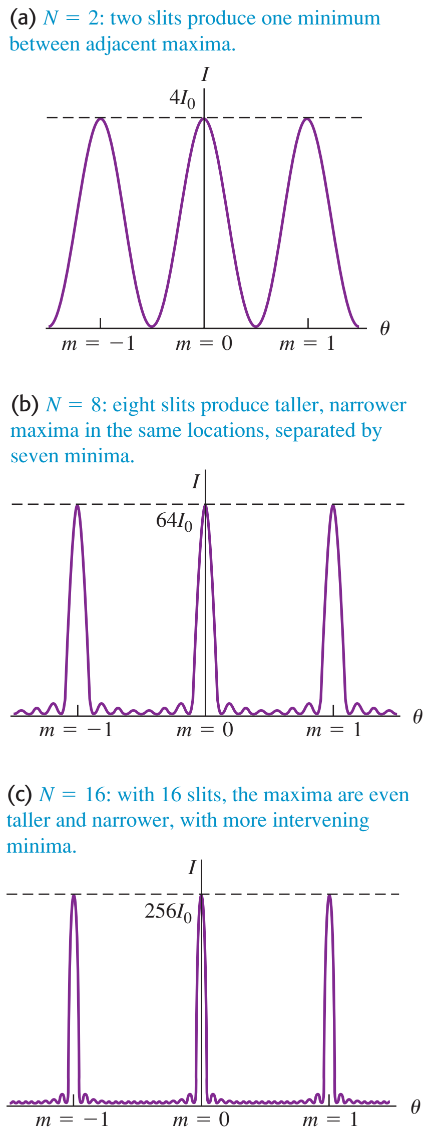
generally, the intensity with eight slits is zero whenever ϕ is an integer multiple of $\pi/4$, *except* when ϕ is a multiple of 2π. Thus there are seven minima for every maximum.

Figure 36.15b shows the result of a detailed calculation of the eight-slit pattern. The large maxima, called *principal maxima,* are in the same positions as for the two-slit pattern of Fig. 36.15a but are much narrower. If the phase difference ϕ between adjacent slits is slightly different from a multiple of 2π, the waves from slits 1 and 2 will be only a little out of phase; however, the phase difference between slits 1 and 3 will be greater, that between slits 1 and 4 will be greater still, and so on. This leads to a partial cancellation for angles that are only slightly different from the angle for a maximum, giving the narrow maxima in Fig. 36.15b. The maxima are even narrower with 16 slits (Fig. 36.15c).

You should show that when there are N slits, there are $(N - 1)$ minima between each pair of principal maxima and a minimum occurs whenever ϕ is an integral multiple of $2\pi/N$ (except when ϕ is an integral multiple of 2π, which gives a principal maximum). There are small *secondary* intensity maxima between the minima; these become smaller in comparison to the principal maxima as N increases. The greater the value of N, the narrower the principal maxima become. From an energy standpoint the total power in the entire pattern is proportional to N. The height of each principal maximum is proportional to N^2, so from energy conservation the width of each principal maximum must be proportional to $1/N$. As we will see in the next section, the narrowness of the principal maxima in a multiple-slit pattern is of great practical importance.

Test Your Understanding of Section 36.4 Suppose two slits, each of width a, are separated by a distance $d = 2.5a$. Are there any missing maxima in the interference pattern produced by these slits? If so, which are missing? If not, why not? ❙

36.15 Interference patterns for N equally spaced, very narrow slits. (a) Two slits. (b) Eight slits. (c) Sixteen slits. The vertical scales are different for each graph; I_0 is the maximum intensity for a single slit, and the maximum intensity for N slits is $N^2 I_0$. The width of each peak is proportional to $1/N$.

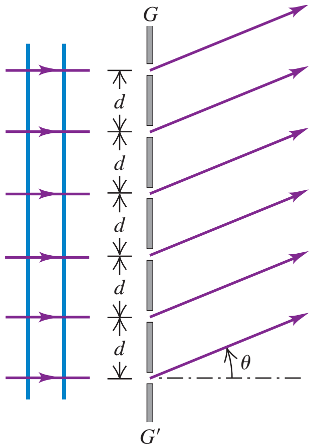


36.5 The Diffraction Grating

We have just seen that increasing the number of slits in an interference experiment (while keeping the spacing of adjacent slits constant) gives interference patterns in which the maxima are in the same positions, but progressively narrower, than with two slits. Because these maxima are so narrow, their angular position, and hence the wavelength, can be measured to very high precision. As we will see, this effect has many important applications.

An array of a large number of parallel slits, all with the same width a and spaced equal distances d between centers, is called a **diffraction grating.** The first one was constructed by Fraunhofer using fine wires. Gratings can be made by using a diamond point to scratch many equally spaced grooves on a glass or metal surface,

36.16 A portion of a transmission diffraction grating. The separation between the centers of adjacent slits is d.



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 given by

$$d\sin\theta = m\lambda \qquad (m = 0, \pm1, \pm2, \pm3, \dots) \qquad \text{(intensity maxima, multiple slits)} \tag{36.13}$$

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.

MasteringPHYSICS®

ActivPhysics 16.4: The Grating: Introduction and Questions
ActivPhysics 16.5: The Grating: Problems

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 $m = \pm1$ lines are called the *first-order lines,* the $m = \pm2$ lines the *second-order lines,* and so on. If the grating is illuminated by white light with a continuous distribution of wavelengths, each value of m corresponds to a continuous spectrum in the pattern. The angle for each wavelength is determined by Eq. (36.13); for a given value of m, long wavelengths (the red end of the spectrum) lie at larger angles (that is, are deviated more from the straight-ahead direction) than do the shorter wavelengths at the violet end of the spectrum.

As Eq. (36.13) shows, the sines of the deviation angles of the maxima are proportional to the ratio λ/d. For substantial deviation to occur, the grating spacing d should be of the same order of magnitude as the wavelength λ. Gratings for use with visible light (λ from 400 to 700 nm) usually have about 1000 slits per millimeter; the value of d is the *reciprocal* of the number of slits per unit length, so d is of the order of $\frac{1}{1000}$ mm $= 1000$ nm.

36.17 Microscopic pits on the surface of this DVD act as a reflection grating, splitting white light into its component colors.

In a *reflection grating,* the array of equally spaced slits shown in Fig. 36.16 is replaced by an array of equally spaced ridges or grooves on a reflective screen. The reflected light has maximum intensity at angles where the phase difference between light waves reflected from adjacent ridges or grooves is an integral multiple of 2π. If light of wavelength λ is incident normally on a reflection grating with a spacing d between adjacent ridges or grooves, the *reflected* angles at which intensity maxima occur are given by Eq. (36.13).

The rainbow-colored reflections from the surface of a DVD are a reflection-grating effect (Fig. 36.17). The "grooves" are tiny pits 0.12 μm deep in the surface of the disc, with a uniform radial spacing of 0.74 μm $=$ 740 nm. Information is coded on the DVD by varying the *length* of the pits. The reflection-grating aspect of the disc is merely an aesthetic side benefit.

Example 36.4 Width of a grating spectrum

The wavelengths of the visible spectrum are approximately 380 nm (violet) to 750 nm (red). (a) Find the angular limits of the first-order visible spectrum produced by a plane grating with 600 slits per millimeter when white light falls normally on the grating. (b) Do the first-order and second-order spectra overlap? What about the second-order and third-order spectra? Do your answers depend on the grating spacing?

SOLUTION

IDENTIFY and SET UP: We must find the angles spanned by the visible spectrum in the first-, second-, and third-order spectra. These correspond to $m = 1$, 2, and 3 in Eq. (36.13).

EXECUTE: (a) The grating spacing is

$$d = \frac{1}{600\ \text{slits/mm}} = 1.67 \times 10^{-6}\ \text{m}$$

We solve Eq. (36.13) for θ:

$$\theta = \arcsin\frac{m\lambda}{d}$$

Then for $m = 1$, the angular deviations θ_{v1} and θ_{r1} for violet and red light, respectively, are

$$\theta_{v1} = \arcsin\left(\frac{380 \times 10^{-9}\ \text{m}}{1.67 \times 10^{-6}\ \text{m}}\right) = 13.2°$$

$$\theta_{r1} = \arcsin\left(\frac{750 \times 10^{-9}\ \text{m}}{1.67 \times 10^{-6}\ \text{m}}\right) = 26.7°$$

That is, the first-order visible spectrum appears with deflection angles from $\theta_{v1} = 13.2°$ (violet) to $\theta_{r1} = 26.7°$ (red).

(b) With $m = 2$ and $m = 3$, our equation $\theta = \arcsin(m\lambda/d)$ for 380-mm violet light yields

$$\theta_{v2} = \arcsin\left(\frac{2(380 \times 10^{-9}\ \text{m})}{1.67 \times 10^{-6}\ \text{m}}\right) = 27.1°$$

$$\theta_{v3} = \arcsin\left(\frac{3(380 \times 10^{-9}\ \text{m})}{1.67 \times 10^{-6}\ \text{m}}\right) = 43.0°$$

For 750-nm red light, this same equation gives

$$\theta_{r2} = \arcsin\left(\frac{2(750 \times 10^{-9}\ \text{m})}{1.67 \times 10^{-6}\ \text{m}}\right) = 63.9°$$

$$\theta_{r3} = \arcsin\left(\frac{3(750 \times 10^{-9}\ \text{m})}{1.67 \times 10^{-6}\ \text{m}}\right) = \arcsin(1.35) = \text{undefined}$$

Hence the second-order spectrum extends from 27.1° to 63.9° and the third-order spectrum extends from 43.0° to 90° (the largest possible value of θ). The undefined value of θ_{r3} means that the third-order spectrum reaches $\theta = 90° = \arcsin(1)$ at a wavelength shorter than 750 nm; you should be able to show that this happens for $\lambda = 557$ nm. Hence the first-order spectrum (from 13.2° to 26.7°) does not overlap with the second-order spectrum, but the second- and third-order spectra do overlap. You can convince yourself that this is true for any value of the grating spacing d.

EVALUATE: The fundamental reason the first-order and second-order visible spectra don't overlap is that the human eye is sensitive to only a narrow range of wavelengths. Can you show that if the eye could detect wavelengths from 380 nm to 900 nm (in the near-infrared range), the first and second orders *would* overlap?

Grating Spectrographs

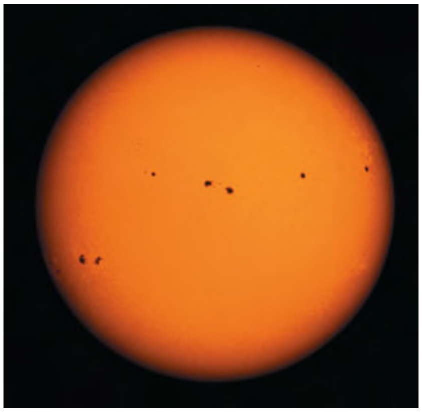
Diffraction gratings are widely used to measure the spectrum of light emitted by a source, a process called *spectroscopy* or *spectrometry*. Light incident on a grating of known spacing is dispersed into a spectrum. The angles of deviation of the maxima are then measured, and Eq. (36.13) is used to compute the wavelength. With a grating that has many slits, very sharp maxima are produced, and the angle of deviation (and hence the wavelength) can be measured very precisely.

An important application of this technique is to astronomy. As light generated within the sun passes through the sun's atmosphere, certain wavelengths are selectively absorbed. The result is that the spectrum of sunlight produced by a diffraction grating has dark *absorption lines* (Fig. 36.18). Experiments in the laboratory show that different types of atoms and ions absorb light of different wavelengths. By comparing these laboratory results with the wavelengths of absorption lines in the spectrum of sunlight, astronomers can deduce the chemical composition of the sun's atmosphere. The same technique is used to make chemical assays of galaxies that are millions of light-years away.

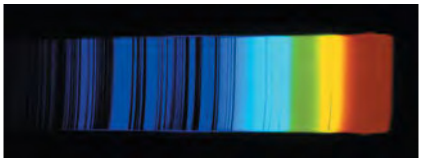
Figure 36.19 shows one design for a *grating spectrograph* used in astronomy. A transmission grating is used in the figure; in other setups, a reflection grating is used. In older designs a prism was used rather than a grating, and a spectrum was formed by dispersion (see Section 33.4) rather than diffraction. However, there is no simple relationship between wavelength and angle of deviation for a prism, prisms absorb some of the light that passes through them, and they are less effective for many nonvisible wavelengths that are important in astronomy. For these and other reasons, gratings are preferred in precision applications.

36.18 (a) A visible-light photograph of the sun. (b) Sunlight is dispersed into a spectrum by a diffraction grating. Specific wavelengths are absorbed as sunlight passes through the sun's atmosphere, leaving dark lines in the spectrum.

(a)

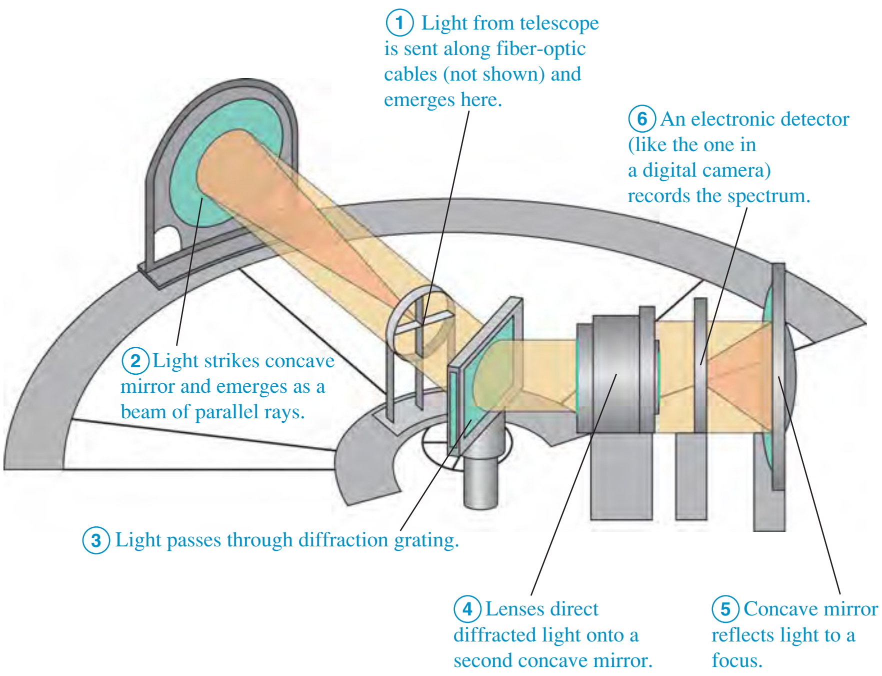


(b)



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



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). (See Problem 36.64.)

$$R = \frac{\lambda}{\Delta\lambda} \quad \text{(chromatic resolving power)} \tag{36.14}$$

As an example, when sodium atoms are heated, they emit strongly at the yellow wavelengths 589.00 nm and 589.59 nm. A spectrograph that can barely distinguish these two lines in the spectrum (called the *sodium doublet*) has a chromatic resolving power $R = (589.00 \text{ nm})/(0.59 \text{ nm}) = 1000$. (You can see these wavelengths when boiling water on a gas range. If the water boils over onto the flame, dissolved sodium from table salt emits a burst of yellow light.)

We can derive an expression for the resolving power of a diffraction grating used in a spectrograph. Two different wavelengths give diffraction maxima at slightly different angles. As a reasonable (though arbitrary) criterion, let's assume that we can distinguish them as two separate peaks if the maximum of one coincides with the first minimum of the other.

From our discussion in Section 36.4 the mth-order maximum occurs when the phase difference ϕ for adjacent slits is $\phi = 2\pi m$. The first minimum beside that maximum occurs when $\phi = 2\pi m + 2\pi/N$, where N is the number of slits. The phase difference is also given by $\phi = (2\pi d \sin\theta)/\lambda$, so the angular interval $d\theta$ corresponding to a small increment $d\phi$ in the phase shift can be obtained from the differential of this equation:

$$d\phi = \frac{2\pi d \cos\theta \, d\theta}{\lambda}$$

When $d\phi = 2\pi/N$, this corresponds to the angular interval $d\theta$ between a maximum and the first adjacent minimum. Thus $d\theta$ is given by

$$\frac{2\pi}{N} = \frac{2\pi d \cos\theta \, d\theta}{\lambda} \quad \text{or} \quad d\cos\theta \, d\theta = \frac{\lambda}{N}$$

CAUTION **Watch out for different uses of the symbol d** Don't confuse the spacing d with the differential "d" in the angular interval $d\theta$ or in the phase shift increment $d\phi$!

Now we need to find the angular spacing $d\theta$ between maxima for two slightly different wavelengths. We have $d\sin\theta = m\lambda$, so the differential of this equation gives

$$d\cos\theta \, d\theta = m \, d\lambda$$

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.

According to our criterion, the limit or resolution is reached when these two angular spacings are equal. Equating the two expressions for the quantity $(d\cos\theta\, d\theta)$, we find

$$\frac{\lambda}{N} = m\, d\lambda \qquad \text{and} \qquad \frac{\lambda}{d\lambda} = Nm$$

If $\Delta\lambda$ is small, we can replace $d\lambda$ by $\Delta\lambda$, and the resolving power R is given by

$$R = \frac{\lambda}{\Delta\lambda} = Nm \tag{36.15}$$

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.

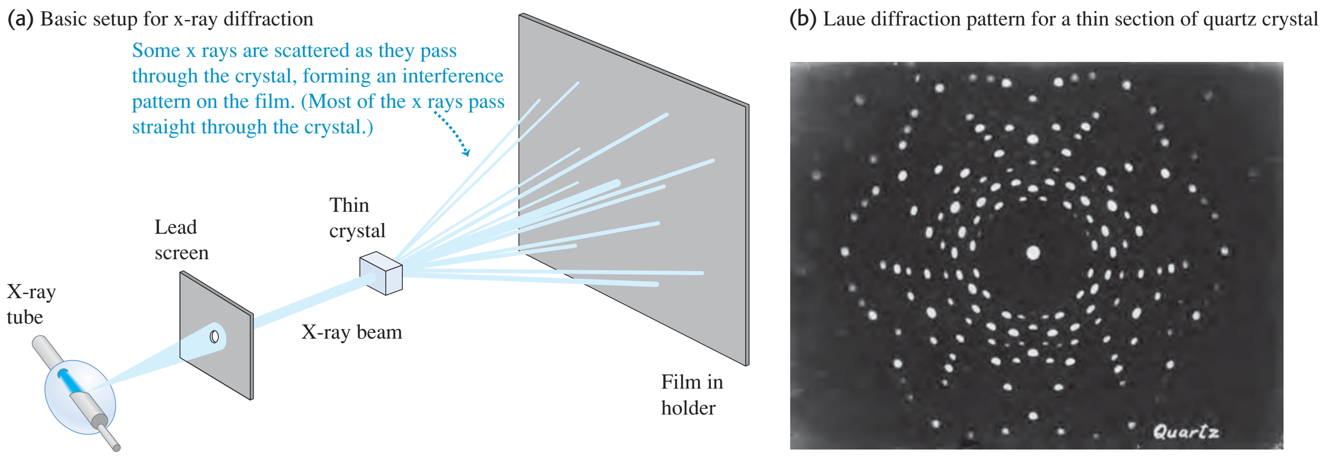
Test Your Understanding of Section 36.5 What minimum number of slits would be required in a grating to resolve the sodium doublet in the fourth order? (i) 250; (ii) 400; (iii) 1000; (iv) 4000.

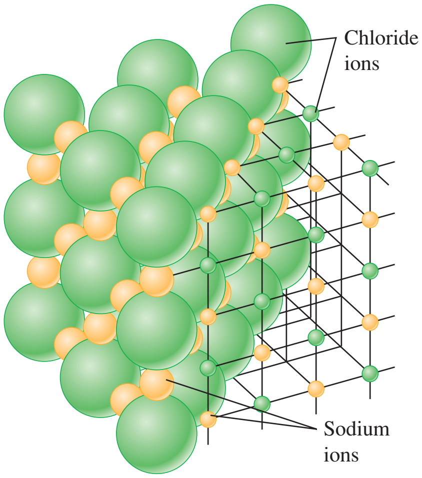
36.6 X-Ray Diffraction

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 Friederich, 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.

36.20 (a) An x-ray diffraction experiment. (b) Diffraction pattern (or *Laue pattern*) formed by directing a beam of x rays at a thin section of quartz crystal.





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.)

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, 3-cm microwaves striking an array of small conducting spheres, 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.

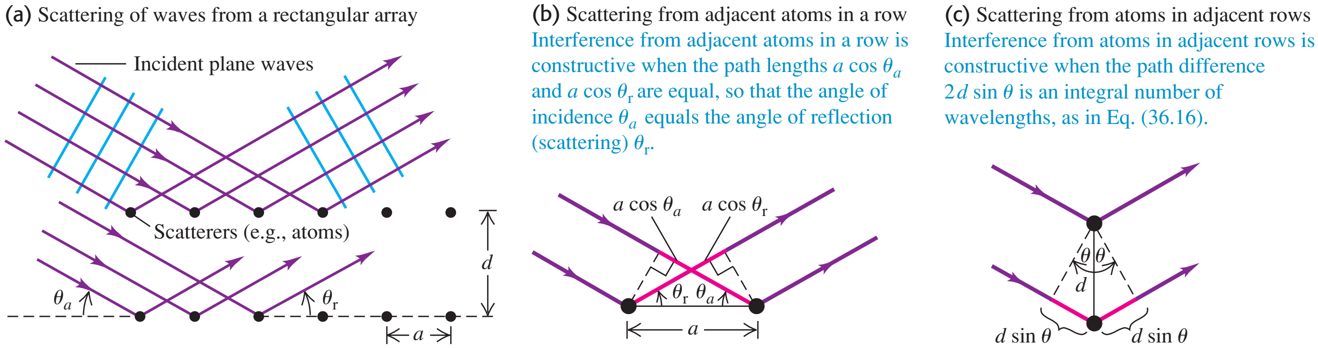
As Fig. 36.22b shows, the path length from source to observer is the same for all the scatterers in a single row if the two angles θ_a and θ_r are equal. Scattered radiation from *adjacent* rows is *also* in phase if the path difference for adjacent rows is an integer number of wavelengths. Figure 36.22c 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 as

$$2d\sin\theta = m\lambda \qquad (m = 1, 2, 3, \dots) \qquad \text{(Bragg condition for constructive interference from an array)} \tag{36.16}$$

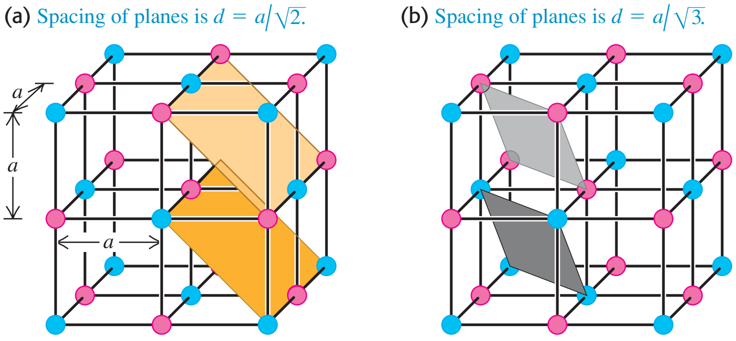
CAUTION **Scattering from an array** In Eq. (36.16) the angle θ is measured with respect to the *surface* of the crystal, rather than with respect to the *normal* to the plane of an array of slits or a grating. Also, note that the path difference in Eq. (36.16) is $2d\sin\theta$, not $d\sin\theta$ as in Eq. (36.13) for a diffraction grating.

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 $m\lambda$

36.22 A two-dimensional model of scattering from a rectangular array. Note that the angles in **(b)** are measured from the *surface* of the array, not from its normal.



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.



must be less than $2d$ and so λ must be less than $2d/m$. For example, the value of d in an NaCl crystal (see Fig. 36.21) is only 0.282 nm. Hence to have the mth-order maximum present in the diffraction pattern, λ must be less than $2(0.282\text{ nm})/m$; that is, $\lambda < 0.564$ nm for $m = 1$, $\lambda < 0.282$ nm for $m = 2$, $\lambda < 0.188$ nm for $m = 3$, and so on. These are all x-ray wavelengths (see Fig. 32.4), which is why x rays are used for studying crystal structure.

We can extend this discussion to a three-dimensional array by considering *planes* of scatterers instead of *rows.* Figure 36.23 shows two different sets of parallel planes that pass through all the scatterers. Waves from all the scatterers in a given plane interfere constructively if the angles of incidence and scattering are equal. There is also constructive interference between planes when Eq. (36.16) is satisfied, where d is now the distance between adjacent planes. Because there are many different sets of parallel planes, there are also many values of d and many sets of angles that give constructive interference for the whole crystal lattice. This phenomenon is called **Bragg reflection,** and Eq. (36.16) is called the **Bragg condition,** in honor of Sir William Bragg and his son Laurence Bragg, two pioneers in x-ray analysis.

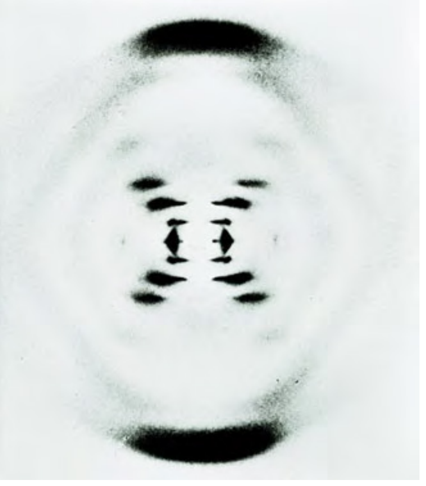
CAUTION **Bragg *reflection* is really Bragg *interference*** While we are using the term *reflection,* remember that we are dealing with an *interference* effect. In fact, the reflections from various planes are closely analogous to interference effects in thin films (see Section 35.4).

As Fig. 36.20b shows, in x-ray diffraction there is nearly complete cancellation in all but certain very specific directions in which constructive interference occurs and forms bright spots. Such a pattern is usually called an x-ray *diffraction* pattern, although *interference* pattern might be more appropriate.

We can determine the wavelength of x rays by examining the diffraction pattern for a crystal of known structure and known spacing between atoms, just as we determined wavelengths of visible light by measuring patterns from slits or gratings. (The spacing between atoms in simple crystals of known structure, such as sodium chloride, can be found from the density of the crystal and Avogadro's number.) Then, once we know the x-ray wavelength, we can use x-ray diffraction to explore the structure and determine the spacing between atoms in crystals with unknown structure.

X-ray diffraction is by far the most important experimental tool in the investigation of crystal structure of solids. X-ray diffraction also plays an important role in studies of the structures of liquids and of organic molecules. It has been one of the chief experimental techniques in working out the double-helix structure of DNA (Fig. 36.24) and subsequent advances in molecular genetics.

36.24 The British scientist Rosalind Franklin made this groundbreaking x-ray diffraction image of DNA in 1953. The dark bands arranged in a cross provided the first evidence of the helical structure of the DNA molecule.



Example 36.5 X-ray diffraction

You direct a beam of 0.154-nm x rays at certain planes of a silicon crystal. As you increase the angle of incidence of the beam from zero, the first strong interference maximum occurs when the beam makes an angle of 34.5° with the planes. (a) How far apart are the planes? (b) Will you find other interference maxima from these planes at greater angles of incidence?

SOLUTION

IDENTIFY and SET UP: This problem involves Bragg reflection of x rays from the planes of a crystal. In part (a) we use the Bragg condition, Eq. (36.16), to find the distance d between adjacent planes from the known wavelength $\lambda = 0.154$ nm and angle of incidence $\theta = 34.5°$ for the $m = 1$ interference maximum. Given the value of d, we use the Bragg condition again in part (b) to find the values of θ for interference maxima corresponding to other values of m.

EXECUTE: (a) We solve Eq. (36.16) for d and set $m = 1$:

$$d = \frac{m\lambda}{2\sin\theta} = \frac{(1)(0.154\text{ nm})}{2\sin 34.5°} = 0.136\text{ nm}$$

This is the distance between adjacent planes.

(b) To calculate other angles, we solve Eq. (36.16) for $\sin\theta$:

$$\sin\theta = \frac{m\lambda}{2d} = m\frac{0.154\text{ nm}}{2(0.136\text{ nm})} = m(0.566)$$

Values of m of 2 or greater give values of $\sin\theta$ greater than unity, which is impossible. Hence there are *no* other angles for interference maxima for this particular set of crystal planes.

EVALUATE: Our result in part (b) shows that there *would* be a second interference maximum if the quantity $2\lambda/2d = \lambda/d$ were less than 1. This would be the case if the wavelength of the x rays were less than $d = 0.136$ nm. How short would the wavelength need to be to have *three* interference maxima?

Test Your Understanding of Section 36.6 You are doing an x-ray diffraction experiment with a crystal in which the atomic planes are 0.200 nm apart. You are using x rays of wavelength 0.100 nm. Will the fifth-order maximum be present in the diffraction pattern?

MasteringPHYSICS

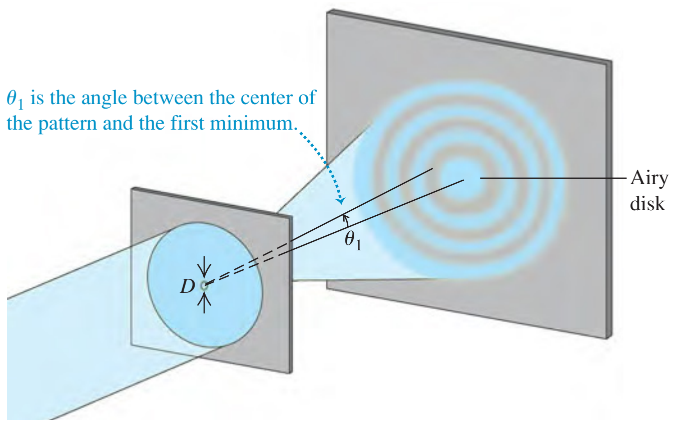
ActivPhysics 16.7: Circular Hole Diffraction
ActivPhysics 16.8: Resolving Power

36.7 Circular Apertures and Resolving Power

We have studied in detail the diffraction patterns formed by long, thin slits or arrays of slits. But an aperture of *any* shape forms a diffraction pattern. The diffraction pattern formed by a *circular* aperture is of special interest because of its role in limiting how well an optical instrument can resolve fine details. In principle, we could compute the intensity at any point P in the diffraction pattern by dividing the area of the aperture into small elements, finding the resulting wave amplitude and phase at P, and then integrating over the aperture area to find the resultant amplitude and intensity at P. In practice, the integration cannot be carried out in terms of elementary functions. We will simply *describe* the pattern and quote a few relevant numbers.

The diffraction pattern formed by a circular aperture consists of a central bright spot surrounded by a series of bright and dark rings, as Fig. 36.25 shows.

36.25 Diffraction pattern formed by a circular aperture of diameter D. The pattern consists of a central bright spot and alternating dark and bright rings. The angular radius θ_1 of the first dark ring is shown. (This diagram is not drawn to scale.)



We can describe the pattern in terms of the angle θ, representing the angular radius of each ring. If the aperture diameter is D and the wavelength is λ, the angular radius θ_1 of the first *dark* ring is given by

$$\sin\theta_1 = 1.22\frac{\lambda}{D} \qquad \text{(diffraction by a circular aperture)} \tag{36.17}$$

The angular radii of the next two dark rings are given by

$$\sin\theta_2 = 2.23\frac{\lambda}{D} \qquad \sin\theta_3 = 3.24\frac{\lambda}{D} \tag{36.18}$$

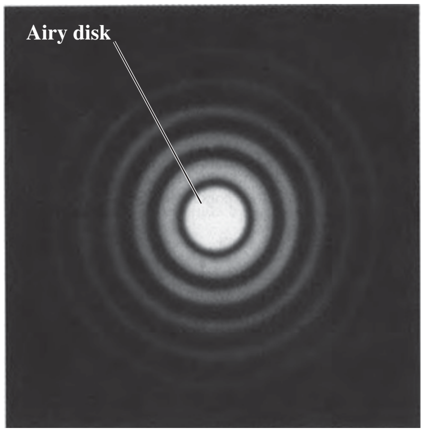
Between these are bright rings with angular radii given by

$$\sin\theta = 1.63\frac{\lambda}{D}, \qquad 2.68\frac{\lambda}{D}, \qquad 3.70\frac{\lambda}{D} \tag{36.19}$$

and so on. The central bright spot is called the **Airy disk,** in honor of Sir George Airy (1801–1892), Astronomer Royal of England, 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 intensities in the bright rings drop off very quickly with increasing angle. When D is much larger than the wavelength λ, the usual case for optical instruments, the peak intensity in the first ring is only 1.7% of the value at the center of the Airy disk, and the peak intensity of the second ring is only 0.4%. Most (85%) of the light energy falls within the Airy disk. Figure 36.26 shows a diffraction pattern from a circular aperture.

36.26 Photograph of the diffraction pattern formed by a circular aperture.



Diffraction and Image Formation

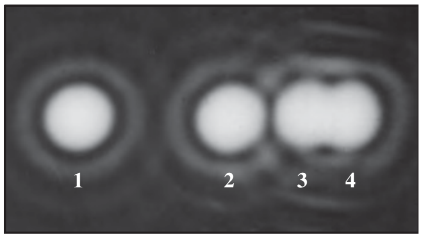
Diffraction has far-reaching implications for image formation by lenses and mirrors. In our study of optical instruments in Chapter 34 we assumed that a lens with focal length f focuses a parallel beam (plane wave) to a *point* at a distance f from the lens. This assumption ignored diffraction effects. 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 the left-hand source is well separated from the others, but the images of the middle and right-hand sources have merged. In Fig. 36.27b, with a larger aperture diameter and hence smaller Airy disks, the middle and right-hand images are better resolved. In Fig. 36.27c, with a still larger aperture, they are well resolved.

A widely used criterion for resolution of two point objects, proposed by the English physicist Lord Rayleigh (1842–1919) and called **Rayleigh's criterion,** is that the objects are just barely resolved (that is, distinguishable) if the center of one diffraction pattern coincides with the first minimum of the other. In that case the angular separation of the image centers is given by Eq. (36.17). The angular separation of the *objects* is the same as that of the *images* made by a telescope, microscope, or other optical device. So two point objects are barely resolved, according to Rayleigh's criterion, when their angular separation is given by Eq. (36.17).

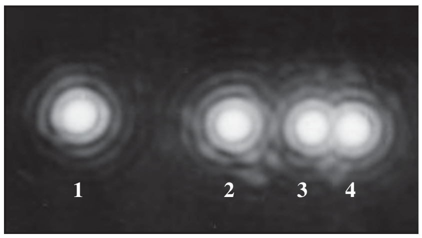
The minimum separation of two objects that can just be resolved by an optical instrument is called the **limit of resolution** of the instrument. The smaller the limit of resolution, the greater the *resolution,* or **resolving power,** of the instrument. Diffraction sets the ultimate limits on resolution 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

36.27 Diffraction patterns of four very small ("point") sources of light. The photographs were made with a circular aperture in front of the lens. (a) The aperture is so small that the patterns of sources 3 and 4 overlap and are barely resolved by Rayleigh's criterion. Increasing the size of the aperture decreases the size of the diffraction patterns, as shown in (b) and (c).

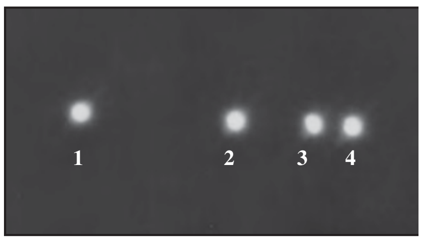
(a) Small aperture



(b) Medium aperture



(c) Large aperture



Application Bigger Telescope, Better Resolution

One reason for building very large telescopes is to increase the aperture diameter and thus minimize diffraction effects. The effective diameter of a telescope can be increased by using arrays of smaller telescopes. The Very Large Array (VLA) in New Mexico is a collection of 27 radio telescopes, each 25 m in diameter, that can be spread out in a Y-shaped arrangement 36 km across. Hence the effective aperture diameter is 36 km, giving the VLA a limit of resolution of 5×10^{-8} rad at a radio wavelength of 1.5 cm. If your eye had this angular resolution, you could read the "20/20" line on an eye chart more than 30 km away!

gain in detail. The images in Fig. 36.27 would not become sharper with further enlargement.

CAUTION **Resolving power vs. chromatic resolving power** Be careful not to confuse the resolving power of an optical instrument with the *chromatic* resolving power of a grating (described in Section 36.5). Resolving power refers to the ability to distinguish the images of objects that appear close to each other, when looking either through an optical instrument or at a photograph made with the instrument. Chromatic resolving power describes how well different wavelengths can be distinguished in a spectrum formed by a diffraction grating.

Rayleigh's criterion combined with Eq. (36.17) shows that resolution (resolving power) improves with larger diameter; it also improves with shorter wavelengths. Ultraviolet microscopes have higher resolution than visible-light microscopes. In electron microscopes the resolution is limited by the wavelengths associated with the electrons, which have wavelike aspects (to be discussed further in Chapter 39). These wavelengths can be made 100,000 times smaller than wavelengths of visible light, with a corresponding gain in resolution. Resolving power also explains the difference in storage capacity between DVDs (introduced in 1995) and Blu-ray discs (introduced in 2003). Information is stored in both of these in a series of tiny pits. In order not to lose information in the scanning process, the scanning optics must be able to resolve two adjacent pits so that they do not seem to blend into a single pit (see sources 3 and 4 in Fig. 36.27). The blue scanning laser used in a Blu-ray player has a shorter wavelength (405 nm) and hence better resolving power than the 650-nm red laser in a DVD player. Hence pits can be spaced closer together in a Blu-ray disc than in a DVD, and more information can be stored on a disc of the same size (50 gigabytes on a Blu-ray disc versus 4.7 gigabytes on a DVD).

Example 36.6 Resolving power of a camera lens

A camera lens with focal length $f = 50$ mm and maximum aperture $f/2$ forms an image of an object 9.0 m away. (a) If the resolution is limited by diffraction, what is the minimum distance between two points on the object that are barely resolved? What is the corresponding distance between image points? (b) How does the situation change if the lens is "stopped down" to $f/16$? Use $\lambda = 500$ nm in both cases.

SOLUTION

IDENTIFY and SET UP: This example uses the ideas about resolving power, image formation by a lens (Section 34.4), and f-number (Section 34.5). From Eq. (34.20), the f-number of a lens is its focal length f divided by the aperture diameter D. We use this equation to determine D and then use Eq. (36.17) (the Rayleigh criterion) to find the angular separation θ between two barely resolved points on the object. We then use the geometry of image formation by a lens to determine the distance y between those points and the distance y' between the corresponding image points.

EXECUTE: (a) The aperture diameter is $D = f/(f\text{-number}) = (50\text{ mm})/2 = 25\text{ mm} = 25 \times 10^{-3}$ m. From Eq. (36.17) the angular separation θ of two object points that are barely resolved is

$$\theta \approx \sin\theta = 1.22\frac{\lambda}{D} = 1.22\frac{500 \times 10^{-9}\text{ m}}{25 \times 10^{-3}\text{ m}} = 2.4 \times 10^{-5}\text{ rad}$$

We know from our thin-lens analysis in Section 34.4 that, apart from sign, $y/s = y'/s'$ [see Eq. (34.14)]. Thus the angular separations of the object points and the corresponding image points are both equal to θ. Because the object distance s is much greater than the focal length $f = 50$ mm, the image distance s' is approximately equal to f. Thus

$$\frac{y}{9.0\text{ m}} = 2.4 \times 10^{-5} \qquad y = 2.2 \times 10^{-4}\text{ m} = 0.22\text{ mm}$$

$$\frac{y'}{50\text{ mm}} = 2.4 \times 10^{-5} \qquad y' = 1.2 \times 10^{-3}\text{ mm}$$

$$= 0.0012\text{ mm} \approx \tfrac{1}{800}\text{ mm}$$

(b) The aperture diameter is now $(50\text{ mm})/16$, or one-eighth as large as before. The angular separation between barely resolved points is eight times as great, and the values of y and y' are also eight times as great as before:

$$y = 1.8\text{ mm} \qquad y' = 0.0096\text{ mm} = \tfrac{1}{100}\text{ mm}$$

Only the best camera lenses can approach this resolving power.

EVALUATE: Many photographers use the smallest possible aperture for maximum sharpness, since lens aberrations cause light rays that are far from the optic axis to converge to a different image point than do rays near the axis. But as this example shows, diffraction effects become more significant at small apertures. One cause of fuzzy images has to be balanced against another.

Test Your Understanding of Section 36.7 You have been asked to compare four different proposals for telescopes to be placed in orbit, above the blurring effects of the earth's atmosphere. Rank the proposed telescopes in order of their ability to resolve small details, from best to worst. (i) a radio telescope 100 m in diameter observing at a wavelength of 21 cm; (ii) an optical telescope 2.0 m in diameter observing at a wavelength of 500 nm; (iii) an ultraviolet telescope 1.0 m in diameter observing at a wavelength of 100 nm; (iv) an infrared telescope 2.0 m in diameter observing at a wavelength of 10 μm.

36.8 Holography

Holography is a technique for recording and reproducing an image of an object through the use of interference effects. Unlike the two-dimensional images recorded by an ordinary photograph or television system, a holographic image is truly three-dimensional. Such an image can be viewed from different directions to reveal different sides and from various distances to reveal changing perspective. If you had never seen a hologram, you wouldn't believe it was possible!

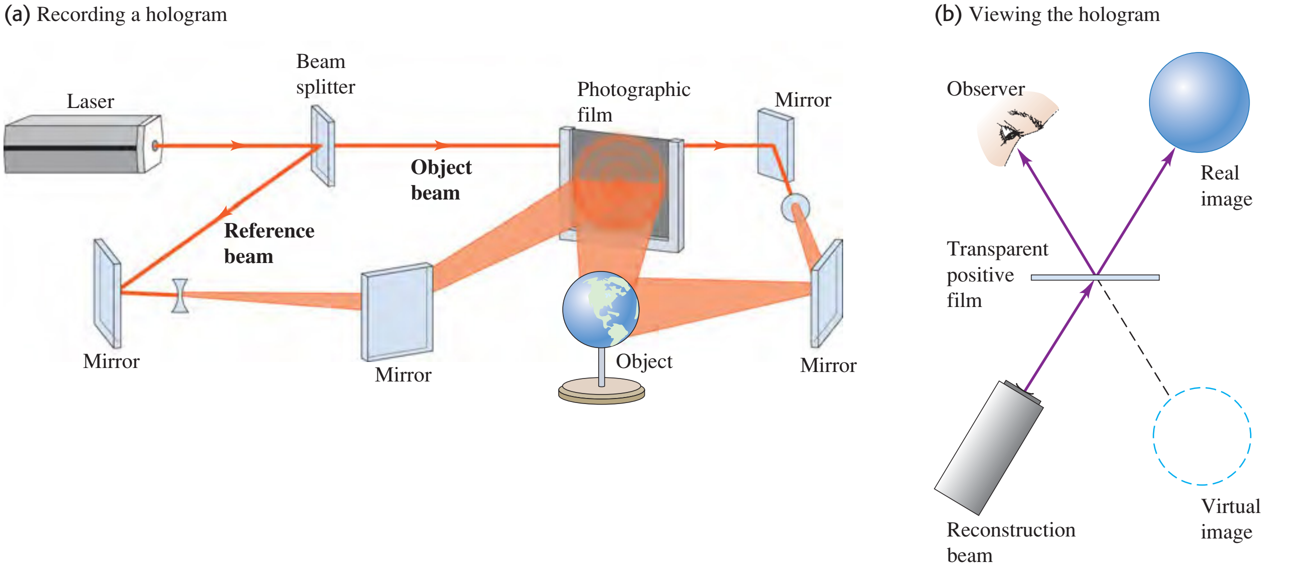
Figure 36.28a shows the basic procedure for making a hologram. We illuminate the object to be holographed with monochromatic light, and we place a photographic film so that it is struck by scattered light from the object and also by direct light from the source. In practice, the light source must be a laser, for reasons we will discuss later. Interference between the direct and scattered light leads to the formation and recording of a complex interference pattern 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.

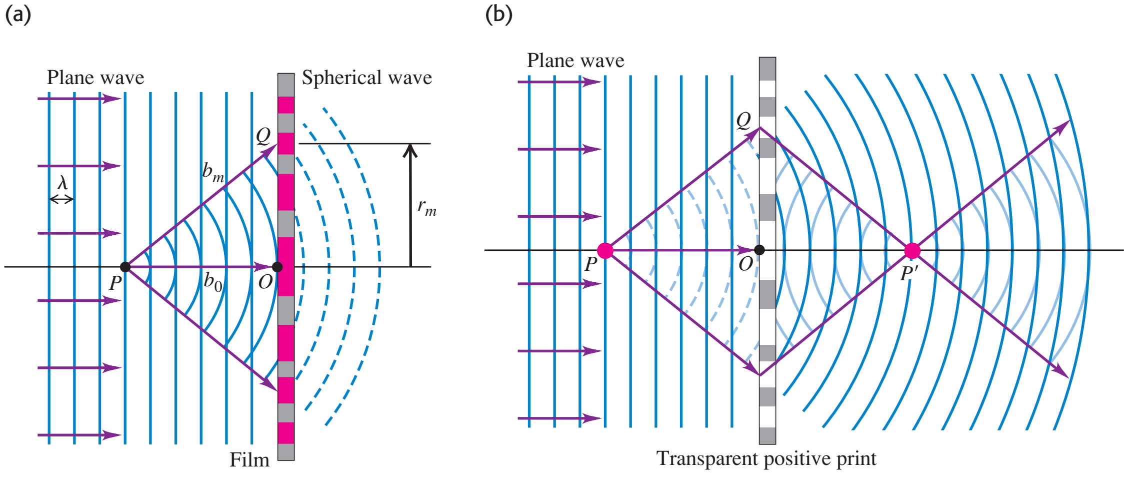
Holography and Interference Patterns

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

36.28 (a) A hologram is the record on film of the interference pattern formed with light from the coherent source and light scattered from the object. (b) Images are formed when light is projected through the hologram. The observer sees the virtual image formed behind the hologram.



36.29 (a) Constructive interference of the plane and spherical waves occurs in the plane of the film at every point Q for which the distance b_m from P is greater than the distance b_0 from P to O by an integral number of wavelengths $m\lambda$. For the point Q shown, $m = 2$. (b) When a plane wave strikes a transparent positive print of the developed film, the diffracted wave consists of a wave converging to P' and then diverging again and a diverging wave that appears to originate at P. These waves form the real and virtual images, respectively.



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

$$r_m^2 = \lambda(2mb_0 + m^2\lambda)$$

Ordinarily, b_0 is very much larger than λ, so we neglect 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.

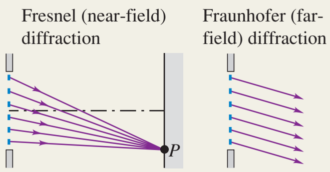
Because of the principle of superposition for waves, what is true for the imaging of a single point is also true for the imaging of any number of points. The film records the superposed interference pattern from the various points, and when light is projected through the film, the various image points 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.

36.30 Two views of the same hologram seen from different angles.

CHAPTER 36 SUMMARY

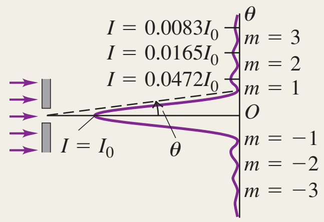
Fresnel and Fraunhofer diffraction: Diffraction occurs when light passes through an aperture or around an edge. When the source and the observer are so far away from the obstructing surface that the outgoing rays can be considered parallel, it is called Fraunhofer diffraction. When the source or the 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.)

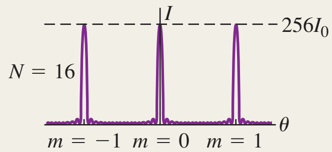
$$\sin\theta = \frac{m\lambda}{a} \qquad (m = \pm 1, \pm 2, \ldots) \tag{36.2}$$

$$I = I_0\left\{\frac{\sin[\pi a(\sin\theta)/\lambda]}{\pi a(\sin\theta)/\lambda}\right\}^2 \tag{36.7}$$



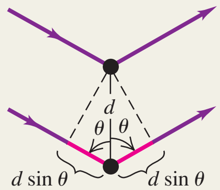
Diffraction gratings: A diffraction grating consists of a large number of thin parallel slits, spaced a distance d apart. The condition for maximum intensity in the interference pattern is the same as for the two-source pattern, but the maxima for the grating are very sharp and narrow. (See Example 36.4.)

$$d\sin\theta = m\lambda \quad (m = 0, \pm 1, \pm 2, \pm 3, \ldots) \tag{36.13}$$



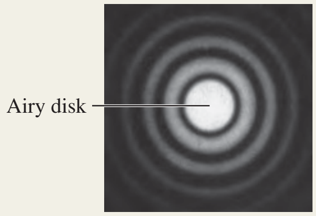
X-ray diffraction: A crystal serves as a three-dimensional diffraction grating for x rays with wavelengths of the same order of magnitude as the spacing between atoms in the crystal. For a set of crystal planes spaced a distance d apart, constructive interference occurs when the angles of incidence and scattering (measured from the crystal planes) are equal and when the Bragg condition [Eq. (36.16)] is satisfied. (See Example 36.5.)

$$2d\sin\theta = m\lambda \qquad (m = 1, 2, 3, \ldots) \tag{36.16}$$



Circular apertures and resolving power: The diffraction pattern from a circular aperture of diameter D consists of a central bright spot, called the Airy disk, and a series of concentric dark and bright rings. Equation (36.17) gives the angular radius θ_1 of the first dark ring, equal to the angular size of the Airy disk. Diffraction sets the ultimate limit on resolution (image sharpness) of optical instruments. According to Rayleigh's criterion, two point objects are just barely resolved when their angular separation θ is given by Eq. (36.17). (See Example 36.6.)

$$\sin\theta_1 = 1.22\frac{\lambda}{D} \tag{36.17}$$



BRIDGING PROBLEM Observing the Expanding Universe

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

See MasteringPhysics® study area for a Video Tutor solution.

IDENTIFY and SET UP

1. You can use the information about the grating to find the wavelength of the H_α line in the galaxy's spectrum.
2. In Section 16.8 we learned about the Doppler effect for electromagnetic radiation: The frequency that we receive from a moving source, such as the galaxy, is different from the frequency that is emitted. Equation (16.30) relates the emitted frequency, the received frequency, and the velocity of the source (the target variable). The equation $c = f\lambda$ relates the frequency f and wavelength λ through the speed of light c.

EXECUTE

3. Find the wavelength of the H_α spectral line in the received light.
4. Rewrite Eq. (16.30) as a formula for the velocity v of the galaxy in terms of the received wavelength and the wavelength emitted by the source.
5. Solve for v. Express it in m/s and as a percentage of c, and decide whether the galaxy is moving toward us or moving away.

EVALUATE

6. Is your answer consistent with the relative sizes of the received wavelength and the emitted wavelength?

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **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 The predominant sound waves used in human speech have wavelengths in the range from 1.0 to 3.0 meters. Using the ideas of diffraction, explain how it is possible to hear a person's voice even when he is facing away from you.

Q36.7 In single-slit diffraction, what is $\sin(\beta/2)$ when $\theta = 0$? In view of your answer, why is the single-slit intensity *not* equal to zero at the center?

Q36.8 A rainbow ordinarily shows a range of colors (see Section 33.4). But if the water droplets that form the rainbow are small enough, the rainbow will appear white. Explain why, using diffraction ideas. How small do you think the raindrops would have to be for this to occur?

Q36.9 Some loudspeaker horns for outdoor concerts (at which the entire audience is seated on the ground) are wider vertically than horizontally. Use diffraction ideas to explain why this is more efficient at spreading the sound uniformly over the audience than either a square speaker horn or a horn that is wider horizontally than vertically. Would this still be the case if the audience were seated at different elevations, as in an amphitheater? Why or why not?

Q36.10 Figure 31.12 (Section 31.2) shows a loudspeaker system. Low-frequency sounds are produced by the *woofer,* which is a speaker with large diameter; the *tweeter,* a speaker with smaller diameter, produces high-frequency sounds. Use diffraction ideas to explain why the tweeter is more effective for distributing high-frequency sounds uniformly over a room than is the woofer.

Q36.11 Information is stored on an audio compact disc, CD-ROM, or DVD disc in a series of pits on the disc. These pits are scanned by a laser beam. An important limitation on the amount of information that can be stored on such a disc is the width of the laser beam. Explain why this should be, and explain how using a shorter-wavelength laser allows more information to be stored on a disc of the same size.

Q36.12 With which color of light can the Hubble Space Telescope see finer detail in a distant astronomical object: red, blue, or ultraviolet? Explain your answer.

Q36.13 At the end of Section 36.4, the following statements were made about an array of N slits. Explain, using phasor diagrams, why each statement is true. (a) A minimum occurs whenever ϕ is an integral multiple of $2\pi/N$, except when ϕ is an integral multiple of 2π (which gives a principal maximum). (b) There are $(N - 1)$ minima between each pair of principal maxima.

Q36.14 Could x-ray diffraction effects with crystals be observed by using visible light instead of x rays? Why or why not?

Q36.15 Why is a diffraction grating better than a two-slit setup for measuring wavelengths of light?

Q36.16 One sometimes sees rows of evenly spaced radio antenna towers. A student remarked that these act like diffraction gratings. What did she mean? Why would one want them to act like a diffraction grating?

Q36.17 If a hologram is made using 600-nm light and then viewed with 500-nm light, how will the images look compared to those observed when viewed with 600-nm light? Explain.

Q36.18 A hologram is made using 600-nm light and then viewed by using white light from an incandescent bulb. What will be seen? Explain.

Q36.19 Ordinary photographic film reverses black and white, in the sense that the most brightly illuminated areas become blackest upon development (hence the term *negative*). Suppose a hologram negative is viewed directly, without making a positive transparency. How will the resulting images differ from those obtained with the positive? Explain.

EXERCISES

Section 36.2 Diffraction from a Single Slit

36.1 •• Monochromatic light from a distant source is incident on a slit 0.750 mm wide. On a screen 2.00 m away, the distance from the central maximum of the diffraction pattern to the first minimum is measured to be 1.35 mm. Calculate the wavelength of the light.

36.2 • Parallel rays of green mercury light with a wavelength of 546 nm pass through a slit covering a lens with a focal length of 60.0 cm. In the focal plane of the lens the distance from the central maximum to the first minimum is 10.2 mm. What is the width of the slit?

36.3 •• Light of wavelength 585 nm falls on a slit 0.0666 mm wide. (a) On a very large and distant screen, how many *totally* dark fringes (indicating complete cancellation) will there be, including both sides of the central bright spot? Solve this problem *without* calculating all the angles! (*Hint:* What is the largest that $\sin\theta$ can be? What does this tell you is the largest that m can be?) (b) At what angle will the dark fringe that is most distant from the central bright fringe occur?

36.4 • Light of wavelength 633 nm from a distant source is incident on a slit 0.750 mm wide, and the resulting diffraction pattern is observed on a screen 3.50 m away. What is the distance between the two dark fringes on either side of the central bright fringe?

36.5 •• Diffraction occurs for all types of waves, including sound waves. High-frequency sound from a distant source with wavelength 9.00 cm passes through a slit 12.0 cm wide. A microphone is placed 8.00 m directly in front of the center of the slit, corresponding to point O in Fig. 36.5a. The microphone is then moved in a direction perpendicular to the line from the center of the slit to point O. At what distances from O will the intensity detected by the microphone be zero?

36.6 • **CP** **Tsunami!** On December 26, 2004, a violent earthquake of magnitude 9.1 occurred off the coast of Sumatra. This quake triggered a huge tsunami (similar to a tidal wave) that killed more than 150,000 people. Scientists observing the wave on the open ocean measured the time between crests to be 1.0 h and the speed of the wave to be 800 km/h. Computer models of the evolution of this enormous wave showed that it bent around the continents and spread to all the oceans of the earth. When the wave reached the gaps between continents, it diffracted between them as through a slit. (a) What was the wavelength of this tsunami? (b) The distance between the southern tip of Africa and northern Antarctica is about 4500 km, while the distance between the southern end of Australia and Antarctica is about 3700 km. As an approximation, we can model this wave's behavior by using Fraunhofer diffraction. Find the smallest angle away from the central maximum for which the waves would cancel after going through each of these continental gaps.

36.7 •• **CP** A series of parallel linear water wave fronts are traveling directly toward the shore at 15.0 cm/s on an otherwise placid lake. A long concrete barrier that runs parallel to the shore at a distance of 3.20 m away has a hole in it. You count the wave crests and observe that 75.0 of them pass by each minute, and you also observe that no waves reach the shore at ± 61.3 cm from the point directly opposite the hole, but waves do reach the shore everywhere within this distance. (a) How wide is the hole in the barrier? (b) At what other angles do you find no waves hitting the shore?

36.8 • Monochromatic electromagnetic radiation with wavelength λ from a distant source passes through a slit. The diffraction pattern is observed on a screen 2.50 m from the slit. If the width of the central maximum is 6.00 mm, what is the slit width a if the wavelength is (a) 500 nm (visible light); (b) 50.0 μm (infrared radiation); (c) 0.500 nm (x rays)?

36.9 •• **Doorway Diffraction.** Sound of frequency 1250 Hz leaves a room through a 1.00-m-wide doorway (see Exercise 36.5). At which angles relative to the centerline perpendicular to the doorway will someone outside the room hear no sound? Use 344 m/s for the speed of sound in air and assume that the source and listener are both far enough from the doorway for Fraunhofer diffraction to apply. You can ignore effects of reflections.

36.10 • **CP** Light waves, for which the electric field is given by $E_y(x, t) = E_{max} \sin[(1.20 \times 10^7\ \mathrm{m}^{-1})x - \omega t]$, 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 •• Parallel rays of light with wavelength 620 nm pass through a slit covering a lens with a focal length of 40.0 cm. The diffraction pattern is observed in the focal plane of the lens, and the distance from the center of the central maximum to the first minimum is 36.5 cm. What is the width of the slit? (*Note:* The angle that locates the first minimum is *not* small.)

36.12 •• Red light of wavelength 633 nm from a helium–neon laser passes through a slit 0.350 mm wide. The diffraction pattern is observed on a screen 3.00 m away. Define the width of a bright fringe as the distance between the minima on either side. (a) What is the width of the central bright fringe? (b) What is the width of the first bright fringe on either side of the central one?

Section 36.3 Intensity in the Single-Slit Pattern

36.13 •• Monochromatic light of wavelength 580 nm passes through a single slit and the diffraction pattern is observed on a screen. Both the source and screen are far enough from the slit for Fraunhofer diffraction to apply. (a) If the first diffraction minima are at $\pm 90.0°$, so the central maximum completely fills the screen, what is the width of the slit? (b) For the width of the slit as calculated in part (a), what is the ratio of the intensity at $\theta = 45.0°$ to the intensity at $\theta = 0$?

36.14 •• Monochromatic light of wavelength $\lambda = 620$ nm from a distant source passes through a slit 0.450 mm wide. The diffraction pattern is observed on a screen 3.00 m from the slit. In terms of the intensity I_0 at the peak of the central maximum, what is the intensity of the light at the screen the following distances from the center of the central maximum: (a) 1.00 mm; (b) 3.00 mm; (c) 5.00 mm?

36.15 •• A slit 0.240 mm wide is illuminated by parallel light rays of wavelength 540 nm. The diffraction pattern is observed on a screen that is 3.00 m from the slit. The intensity at the center of the central maximum ($\theta = 0°$) is 6.00×10^{-6} W/m^2. (a) What is the distance on the screen from the center of the central maximum to the first minimum? (b) What is the intensity at a point on the screen midway between the center of the central maximum and the first minimum?

36.16 • Monochromatic light of wavelength 486 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 ($\theta = 0^{\circ}$) is 4.00×10^{-5} W/m^2. What is the intensity at a point on the screen that corresponds to $\theta = 1.20^{\circ}$.

36.17 •• A single-slit diffraction pattern is formed by monochromatic electromagnetic radiation from a distant source passing through a slit 0.105 mm wide. At the point in the pattern 3.25° from the center of the central maximum, the total phase difference between wavelets from the top and bottom of the slit is 56.0 rad. (a) What is the wavelength of the radiation? (b) What is the intensity at this point, if the intensity at the center of the central maximum is I_0?

36.18 • Consider a single-slit diffraction experiment in which the amplitude of the wave at point O in Fig. 36.5a is E_0. For each of the following cases, draw a phasor diagram like that in Fig. 36.8c and determine *graphically* the amplitude of the wave at the point in question. (*Hint:* Use Eq. (36.6) to determine the value of β for each case.) Compute the intensity and compare to Eq. (36.5). (a) $\sin\theta = \lambda/2a$; (b) $\sin\theta = \lambda/a$; (c) $\sin\theta = 3\lambda/2a$.

36.19 •• 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?

Section 36.4 Multiple Slits

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 is your result different?

36.21 • **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?

36.22 • An interference pattern is produced by eight parallel and equally spaced, narrow slits. There is an interference minimum when the phase difference ϕ between light from adjacent slits is $\pi/4$. The phasor diagram is given in Fig. 36.14b. For which pairs of slits is there totally destructive interference?

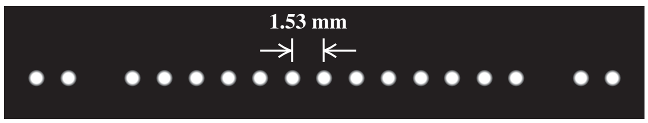
36.23 •• An interference pattern is produced by light of wavelength 580 nm from a distant source incident on two identical parallel slits separated by a distance (between centers) of 0.530 mm. (a) If the slits are very narrow, what would be the angular positions of the first-order and second-order, two-slit, 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)?

36.24 • 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?

36.25 •• An interference pattern is produced by four parallel and equally spaced, narrow slits. By drawing appropriate phasor diagrams, show that 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?

36.26 •• A diffraction experiment involving two thin parallel slits yields the pattern of closely spaced bright and dark fringes shown in Fig. E36.26. Only the central portion of the pattern is shown in the figure. The bright spots are equally spaced at 1.53 mm center to center (except for the missing spots) on a screen 2.50 m from the slits. The light source was a He-Ne laser producing a wavelength of 632.8 nm. (a) How far apart are the two slits? (b) How wide is each one?

Figure **E36.26**



36.27 •• Laser light of wavelength 500.0 nm illuminates two identical slits, producing an interference pattern on a screen 90.0 cm from the slits. The bright bands are 1.00 cm apart, and the third bright bands on either side of the central maximum are missing in the pattern. Find the width and the separation of the two slits.

Section 36.5 The Diffraction Grating

36.28 •• Monochromatic light is at normal incidence on a plane transmission grating. The first-order maximum in the interference pattern is at an angle of 8.94°. What is the angular position of the fourth-order maximum?

36.29 • If a diffraction grating produces its third-order bright band at an angle of 78.4° for light of wavelength 681 nm, find (a) the number of slits per centimeter for the grating and (b) the angular location of the first-order and second-order bright bands. (c) Will there be a fourth-order bright band? Explain.

36.30 • 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)?

36.31 • Visible light passes through a diffraction grating that has 900 slits/cm, and the interference pattern is observed on a screen that is 2.50 m from the grating. (a) Is the angular position of the first-order spectrum small enough for $\sin\theta \approx \theta$ to be a good approximation? (b) In the first-order spectrum, the maxima for two different wavelengths are separated on the screen by 3.00 mm. What is the difference in these wavelengths?

36.32 • 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.)

36.33 • 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?

36.34 • (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.

36.35 • Plane monochromatic waves with wavelength 520 nm are incident normally on a plane transmission grating having 350 slits/mm. Find the angles of deviation in the first, second, and third orders.

36.36 • **Identifying Isotopes by Spectra.** Different isotopes of the same element emit light at slightly different wavelengths. A wavelength in the emission spectrum of a hydrogen atom is 656.45 nm; for deuterium, the corresponding wavelength is 656.27 nm. (a) What minimum number of slits is required to resolve these two wavelengths in second order? (b) If the grating has 500.00 slits/mm, find the angles and angular separation of these two wavelengths in the second order.

36.37 • A typical laboratory diffraction grating has 5.00×10^3 lines/cm, and these lines are contained in a 3.50-cm width of grating. (a) What is the chromatic resolving power of such a grating in the first order? (b) Could this grating resolve the lines of the sodium doublet (see Section 36.5) in the first order? (c) While doing spectral analysis of a star, you are using this grating in the *second* order to resolve spectral lines that are very close to the 587.8002-nm spectral line of iron. (i) For wavelengths longer than the iron line, what is the shortest wavelength you could distinguish from the iron line? (ii) For wavelengths shorter than the iron line, what is the longest wavelength you could distinguish from the iron line? (iii) What is the range of wavelengths you could *not* distinguish from the iron line?

36.38 • The light from an iron arc includes many different wavelengths. Two of these are at $\lambda = 587.9782$ nm and $\lambda = 587.8002$ nm. You wish to resolve these spectral lines in first order using a grating 1.20 cm in length. What minimum number of slits per centimeter must the grating have?

Section 36.6 X-Ray Diffraction

36.39 • 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?

36.40 • If the planes of a crystal are 3.50 Å ($1 \text{ Å} = 10^{-10}\text{ m} =$ 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 15.0°, 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?

36.41 • 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

36.42 • 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.

36.43 • 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?

36.44 •• 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?

36.45 •• 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?

36.46 • **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?

36.47 • **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.

36.48 • A converging lens 7.20 cm in diameter has a focal length of 300 mm. If the resolution is diffraction limited, how far away can an object be if points on it 4.00 mm apart are to be resolved (according to Rayleigh's criterion)? Use $\lambda = 550$ nm.

36.49 •• **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.

36.50 • **Searching for Starspots.** The Hale Telescope on Palomar Mountain in California has a mirror 200 in. (5.08 m) in diameter and it focuses visible light. Given that a large sunspot is about 10,000 mi in diameter, what is the most distant star on which this telescope could resolve a sunspot to see whether other stars have them? (Assume optimal viewing conditions, so that the resolution is diffraction limited.) Are there any stars this close to us, besides our sun?

PROBLEMS

36.51 •• 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?

36.52 •• Suppose the entire apparatus (slit, screen, and space in between) in Exercise 36.4 is immersed in water $(n = 1.333)$. Then what is the distance between the two dark fringes?

36.53 ••• 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 $\mathrm{W/m^2}$. (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.

36.54 •• CP A loudspeaker having a diaphragm that vibrates at 1250 Hz is traveling at 80.0 m/s directly toward a pair of holes in a very large wall in a region for which the speed of sound is 344 m/s. You observe that the sound coming through the openings first cancels at $\pm 11.4°$ with respect to the original direction of the speaker when observed far from the wall. (a) How far apart are the two openings? (b) At what angles would the sound first cancel if the source stopped moving?

36.55 • **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.

36.56 • **Grating Design.** Your boss asks you to design a diffraction grating that will disperse the first-order visible spectrum through an angular range of 21.0° (see Example 36.4 in Section 36.5). (a) What must the number of slits per centimeter be for this grating? (b) At what angles will the first-order visible spectrum begin and end?

36.57 •• 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 $I_0/2$?

36.58 ••• CALC The intensity of light in the Fraunhofer diffraction pattern of a single slit is

$$I = I_0\left(\frac{\sin\gamma}{\gamma}\right)^2$$

where

$$\gamma = \frac{\pi a \sin\theta}{\lambda}$$

(a) Show that the equation for the values of γ at which I is a maximum is $\tan\gamma = \gamma$. (b) Determine the three 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.)

36.59 •• **Angular Width of a Principal Maximum.** Consider N evenly spaced, narrow slits. Use the small-angle approximation $\sin\theta = \theta$ (for θ in radians) to prove the following: For an intensity maximum that occurs at an angle θ, the intensity minima immediately adjacent to this maximum are at angles $\theta + \lambda/Nd$ and $\theta - \lambda/Nd$, so that the angular width of the principal maximum is $2\lambda/Nd$. This is proportional to $1/N$, as we concluded in Section 36.4 on the basis of energy conservation.

36.60 •• 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?

36.61 • **Phasor Diagram for Eight Slits.** An interference pattern is produced by eight equally spaced, narrow slits. Figure 36.14 shows phasor diagrams for the cases in which the phase difference ϕ between light from adjacent slits is $\phi = \pi$, $\phi = \pi/4$, and $\phi = \pi/2$. Each of these cases gives an intensity minimum. The caption for Fig. 36.14 also claims that minima occur for $\phi = 3\pi/4$, $\phi = 5\pi/4$, $\phi = 3\pi/2$, and $\phi = 7\pi/4$. (a) Draw the phasor diagram for each of these four cases, and explain why each diagram proves that there is in fact a minimum. (*Note:* You may find it helpful to use a different colored pencil for each slit!) (b) For each of the four cases $\phi = 3\pi/4$, $\phi = 5\pi/4$, $\phi = 3\pi/2$, and $\phi = 7\pi/4$, for which pairs of slits is there totally destructive interference?

36.62 •• 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.)

36.63 • What is the longest wavelength that can be observed in the third order for a transmission grating having 9200 slits/cm? Assume normal incidence.

36.64 •• (a) Figure 36.16 shows plane waves of light incident *normally* on a diffraction grating. If instead the light strikes the grating at an angle of incidence θ' (measured from the normal), show that the condition for an intensity maximum is *not* Eq. (36.13), but rather

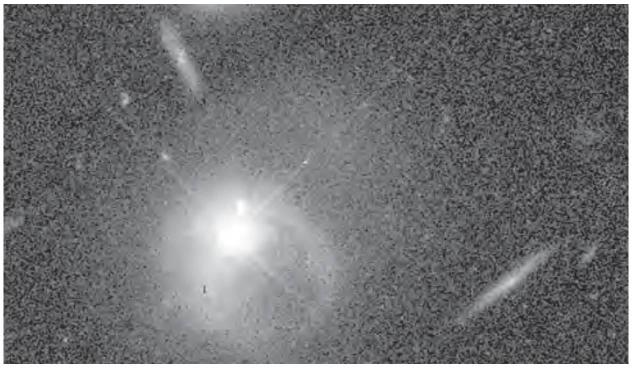
$$d(\sin\theta + \sin\theta') = m\lambda \quad (m = 0, \pm 1, \pm 2, \pm 3, \dots)$$

(b) For the grating described in Example 36.4 (Section 36.5), with 600 slits/mm, find the angles of the maxima corresponding to $m = 0$, 1, and -1 with red light ($\lambda = 650$ nm) for the cases $\theta' = 0$ (normal incidence) and $\theta' = 20.0°$.

36.65 • 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.)

36.66 •• *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. P36.66; 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. P36.66 is thought to be this quasar's host galaxy.) To investigate this model of quasars and to study other exotic astronomical objects, the Russian Space Agency plans to place a radio telescope in an orbit that extends to 77,000 km from the earth. When the signals from this telescope are combined with signals from the ground-based telescopes of the VLBA, the resolution will be that of a single radio telescope 77,000 km in diameter. What is the size of the smallest detail that this arrangement could resolve in quasar 3C 405, which is 7.2×10^8 light-years from earth, using radio waves at a frequency of 1665 MHz? (*Hint:* Use Rayleigh's criterion.) Give your answer in light-years and in kilometers.

Figure **P36.66**



36.67 ••• **Phased-Array Radar.** In one common type of radar installation, a rotating antenna sweeps a radio beam around the sky. But in a *phased-array* radar system, the antennas remain stationary and the beam is swept electronically. To see how this is done, consider an array of N antennas that are arranged along the horizontal x-axis at $x = 0, \pm d, \pm 2d, \ldots, \pm(N-1)d/2$. (The number N is odd.) Each antenna emits radiation uniformly in all directions in the horizontal xy-plane. The antennas all emit radiation coherently, with the same amplitude E_0 and the same wavelength λ. The relative phase δ of the emission from adjacent antennas can be varied, however. If the antenna at $x = 0$ emits a signal that is given by $E_0 \cos \omega t$, as measured at a point next to the antenna, the antenna at $x = d$ emits a signal given by $E_0 \cos(\omega t + \delta)$, as measured at a point next to that antenna. The corresponding quantity for the antenna at $x = -d$ is $E_0 \cos(\omega t - \delta)$; for the antennas at $x = \pm 2d$, it is $E_0 \cos(\omega t \pm 2\delta)$; and so on. (a) If $\delta = 0$, the interference pattern at a distance from the antennas is large compared to d and has a principal maximum at $\theta = 0$ (that is, in the $+y$-direction, perpendicular to the line of the antennas). Show that if $d < \lambda$, this is the *only* principal interference maximum in the angular range $-90° < \theta < 90°$. Hence this principal maximum describes a beam emitted in the direction $\theta = 0$. As described in Section 36.4, if N is large, the beam will have a large intensity and be quite narrow. (b) If $\delta \neq 0$, show that the principal intensity maximum described in part (a) is located at

$$\theta = \arcsin\left(\frac{\delta\lambda}{2\pi d}\right)$$

where δ is measured in radians. Thus, by varying δ from positive to negative values and back again, which can easily be done electronically, the beam can be made to sweep back and forth around $\theta = 0$. (c) A weather radar unit to be installed on an airplane emits radio waves at 8800 MHz. The unit uses 15 antennas in an array 28.0 cm long (from the antenna at one end of the array to the antenna at the other end). What must the maximum and minimum values of δ be (that is, the most positive and most negative values) if the radar beam is to sweep 45° to the left or right of the airplane's direction of flight? Give your answer in radians.

36.68 •• **Underwater Photography.** An underwater camera has a lens of focal length 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?

36.69 •• An astronaut in the space shuttle can just resolve two point sources on earth that are 65.0 m apart. Assume that the resolution is diffraction limited and use Rayleigh's criterion. What is the astronaut's altitude above the earth? Treat his eye as a circular aperture with a diameter of 4.00 mm (the diameter of his pupil), and take the wavelength of the light to be 550 nm. Ignore the effect of fluid in the eye.

36.70 •• 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 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?

36.71 •• A glass sheet is covered by a very thin opaque coating. In the middle of this sheet there is a thin scratch 0.00125 mm thick. The sheet is totally immersed beneath the surface of a liquid. Parallel rays of monochromatic coherent light with wavelength 612 nm in air strike the sheet perpendicular to its surface and pass through the scratch. A screen is placed in the liquid a distance of 30.0 cm away from the sheet and parallel to it. You observe that the first dark fringes on either side of the central bright fringe on the screen are 22.4 cm apart. What is the refractive index of the liquid?

36.72 •• Observing Planets Beyond Our Solar System. NASA is considering a project called *Planet Imager* that would give astronomers the ability to see details on planets orbiting other stars. Using the same principle as the Very Large Array (see Section 36.7), *Planet Imager* will use an array of infrared telescopes spread over thousands of kilometers of space. (Visible light would give even better resolution. Unfortunately, at visible wavelengths, stars are so bright that a planet would be lost in the glare. This is less of a problem at infrared wavelengths.) (a) If *Planet Imager* has an effective diameter of 6000 km and observes infrared radiation at a wavelength of 10 μm, what is the greatest distance at which it would be able to observe details as small as 250 km across (about the size of the greater Los Angeles area) on a planet? Give your answer in light-years (see Appendix E). (*Hint:* Use Rayleigh's criterion.) (b) For comparison, consider the resolution of a single infrared telescope in space that has a diameter of 1.0 m and that observes 10-μm radiation. What is the size of the smallest details that such a telescope could resolve at the distance of the nearest star to the sun, Proxima Centauri, which is 4.22 light-years distant? How does this compare to the diameter of the earth $(1.27 \times 10^4\ \text{km})$? To the average distance from the earth to the sun $(1.50 \times 10^8\ \text{km})$? Would a single telescope of this kind be able to detect the presence of a planet like the earth, in an orbit the size of the earth's orbit, around *any* other star? Explain. (c) Suppose *Planet Imager* is used to observe a planet orbiting the star 70 Virginis, which is 59 light-years from our solar system. A planet (though not an earthlike one) has in fact been detected orbiting this star, not by imaging it directly but by observing the slight "wobble" of the star as both it and the planet orbit their common center of mass. What is the size of the smallest details that *Planet Imager* could hope to resolve on the planet of 70 Virginis? How does this compare to the diameter of the planet, assumed to be comparable to that of Jupiter $(1.38 \times 10^5\ \text{km})$? (Although the planet of 70 Virginis is thought to be at least 6.6 times more massive than Jupiter, its radius is probably not too different from that of Jupiter. The reason is that such large planets are thought to be composed primarily of gases, not rocky material, and hence can be greatly compressed by the mutual gravitational attraction of different parts of the planet.)

CHALLENGE PROBLEMS

36.73 ••• 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).

36.74 ••• 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

$$\begin{aligned} E_P(t) &= E_0\cos(kR - \omega t) + E_0\cos(kR - \omega t + \phi) \\ &\quad + E_0\cos(kR - \omega t + 2\phi) + \cdots \\ &\quad + E_0\cos(kR - \omega t + (N-1)\phi) \end{aligned}$$

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

$$\begin{aligned} &E_0\left(\frac{e^{iN\phi} - 1}{e^{i\phi} - 1}\right)e^{i(kR - \omega t)} \\ &\quad = E_0\left(\frac{e^{iN\phi/2} - e^{-iN\phi/2}}{e^{i\phi/2} - e^{-i\phi/2}}\right)e^{i[kR - \omega t + (N-1)\phi/2]} \end{aligned}$$

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?)

36.75 ••• CALC Intensity Pattern of N Slits, Continued. Part (d) of Challenge Problem 36.74 gives an expression for the intensity in the interference pattern of N identical slits. Use this result to verify the following statements. (a) The maximum intensity in the pattern is $N^2 I_0$. (b) The principal maximum at the center of the pattern extends from $\phi = -2\pi/N$ to $\phi = 2\pi/N$, so its width is inversely proportional to $1/N$. (c) A minimum occurs whenever ϕ is an integral multiple of $2\pi/N$, except when ϕ is an integral multiple of 2π (which gives a principal maximum). (d) There are $(N - 1)$ minima between each pair of principal maxima. (e) Halfway between two principal maxima, the intensity can be no greater than I_0; that is, it can be no greater than $1/N^2$ times the intensity at a principal maximum.

Answers

Chapter Opening Question ?

The shorter wavelength of a Blu-ray scanning laser gives it superior resolving power, so information can be more tightly packed onto a Blu-ray disc than a DVD. See Section 36.7 for details.

Test Your Understanding Questions

36.1 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 of sound, you could hear sounds only from objects that were in plain view.

36.2 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 Answers: (ii) and (iii) If the slit width a is less than the wavelength λ, there are no points in the diffraction pattern at which the intensity is zero (see Fig. 36.10a). The slit width is $0.0100\text{ mm} = 1.00 \times 10^{-5}\text{ m}$, so this condition is satisfied for (ii) $(\lambda = 10.6\ \mu\text{m} = 1.06 \times 10^{-5}\text{ m})$ and (iii) $(\lambda = 1.00\text{ mm} = 1.00 \times 10^{-3}\text{ m})$ but not for (i) $(\lambda = 500\text{ nm} = 5.00 \times 10^{-7}\text{ m})$ or (iv) $(\lambda = 50.0\text{ nm} = 5.00 \times 10^{-8}\text{ m})$.

36.4 Answers: yes; $m_i = \pm5, \pm10, \ldots$ 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 $m_i = \pm5$ and $m_d = \pm2$ (the fifth interference maximum is missing because it coincides with the second diffraction minimum), $m_i = \pm10$ and $m_d = \pm4$ (the tenth interference maximum is missing because it coincides with the fourth diffraction minimum), and so on.

36.5 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 Answer: no The angular position of the mth maximum is given by Eq. (36.16), $2d\sin\theta = m\lambda$. With $d = 0.200\text{ nm}$, $\lambda = 0.100\text{ nm}$, and $m = 5$, this gives $\sin\theta = m\lambda/2d = (5)(0.100\text{ nm})/(2)(0.200\text{ nm}) = 1.25$. Since the sine function can never be greater than 1, this means that there is no solution to this equation and the $m = 5$ maximum does not appear.

36.7 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) $(10\ \mu\text{m})/(2.0\text{ m}) = (1.0 \times 10^{-5}\text{ m})/(2.0\text{ m}) = 5.0 \times 10^{-6}$.

Bridging Problem

Answers: 1.501×10^7 m/s or 5.00% of c; away from us

37 RELATIVITY

? 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. Is there also an upper limit on the *kinetic energy* of a particle?

LEARNING GOALS

By studying this chapter, you will learn:

- The two postulates of Einstein's special theory of relativity, and what motivates these postulates.
- Why different observers can disagree about whether two events are simultaneous.
- How relativity predicts that moving clocks run slow, and that experimental evidence confirms this.
- How the length of an object changes due to the object's motion.
- How the velocity of an object depends on the frame of reference from which it is observed.
- How the theory of relativity modifies the relationship between velocity and momentum.
- How to solve problems involving work and kinetic energy for particles moving at relativistic speeds.
- Some of the key concepts of Einstein's general theory of relativity.

When the year 1905 began, Albert Einstein was an unknown 25-year-old clerk in the Swiss patent office. By the end of that amazing year he had published three 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 third, Einstein introduced his **special theory of relativity,** proposing drastic revisions in the Newtonian concepts of space and time.

The special theory of relativity has made wide-ranging changes in our understanding of nature, but Einstein based it on just two simple postulates. One states that the laws of physics are the same in all inertial frames of reference; the other states that the speed of light in vacuum is the same in all inertial frames. These innocent-sounding propositions have far-reaching implications. Here are three: (1) Events that are simultaneous for one observer may not be simultaneous for another. (2) When two observers moving relative to each other measure a time interval or a length, they may not get the same results. (3) For the conservation principles for momentum and energy to be valid in all inertial systems, Newton's second law and the equations for momentum and kinetic energy have to be revised.

Relativity has important consequences in *all* areas of physics, including electromagnetism, atomic and nuclear physics, and high-energy physics. Although many of the results derived in this chapter may run counter to your intuition, the theory is in solid agreement with experimental observations.

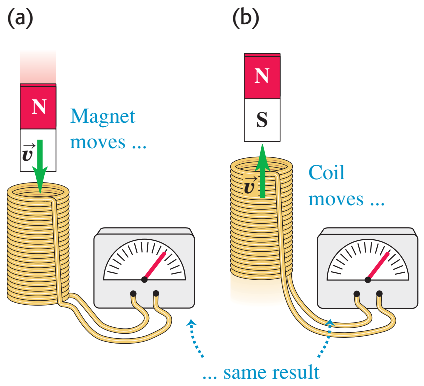
37.1 Invariance of Physical Laws

Let's take a look at the two postulates that make up the special theory of relativity. Both postulates describe what is seen by an observer in an *inertial frame of reference,* which we introduced in Section 4.2. The theory is "special" in the sense that it applies to observers in such special reference frames.

Einstein's First Postulate

Einstein's first postulate, called the **principle of relativity,** states: **The laws of physics are the same in every inertial frame of reference.** If the laws differed, that difference could distinguish one inertial frame from the others or make one frame somehow more "correct" than another. Here are two examples. Suppose you watch two children playing catch with a ball while the three of you are aboard a train moving with constant velocity. Your observations of the motion *of the ball,* no matter how carefully done, can't tell you how fast (or whether) the train is moving. This is because Newton's laws of motion are the same in every inertial frame.

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.



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 Chapter 29, 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, which is within one part in 1000 of the exact value.) As we will see, the speed of light in vacuum plays a central role in the theory of relativity.

Einstein's Second Postulate

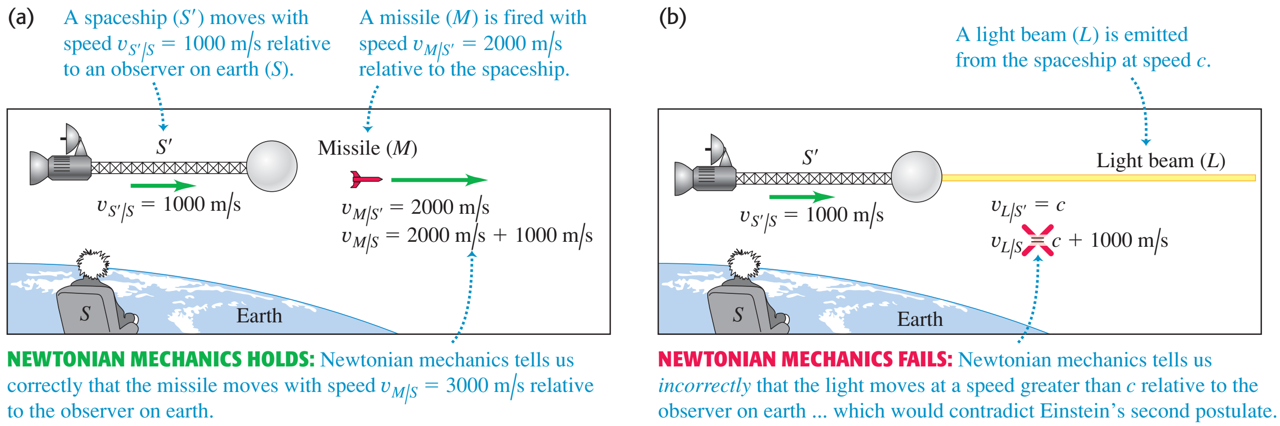
During the 19th century, most physicists believed that light traveled through a hypothetical medium called the *ether,* just as sound waves travel through air. If so, the speed of light measured by observers would depend on their motion relative to the ether and would therefore be different in different directions. The Michelson-Morley experiment, described in Section 35.5, was an effort to detect motion of the earth relative to the ether. Einstein's conceptual leap was to recognize that if Maxwell's equations are valid in all inertial frames, then the speed of light in vacuum should also be the same in all frames and in all directions. In fact, Michelson and Morley detected *no* ether motion across the earth, and the ether concept has been discarded. Although Einstein may not have known about this negative result, it supported his bold hypothesis of the constancy of the speed of light in vacuum.

Einstein's second postulate states: 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, according to Newtonian mechanics, is 3000 m/s.

37.2 (a) Newtonian mechanics makes correct predictions about relatively slow-moving objects; (b) it makes incorrect predictions about the behavior of light.



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:

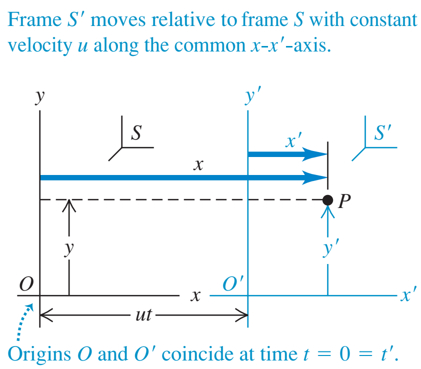
It is impossible for an inertial observer to travel at c, the speed of light in vacuum.

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.

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. For instance, the positive x-direction must be the direction in which the origin O' moves relative to the origin O. In Fig. 37.3 this direction is to the right; if instead O' moves to the left relative to O, you must define the positive x-direction to be to the left.

Now think about how we describe the motion of a particle P. This might be an exploratory vehicle launched from the spacecraft or a pulse of light from a laser. We can describe the *position* of this particle by using the earth coordinates (x, y, z) in S or the spacecraft coordinates (x', y', z') in S'. Figure 37.3 shows that these are simply related by

$$x = x' + ut \qquad y = y' \qquad z = z' \qquad \text{(Galilean coordinate transformation)} \tag{37.1}$$

These equations, based on the familiar Newtonian notions of space and time, are called the **Galilean coordinate transformation.**

If particle P moves in the x-direction, its instantaneous velocity v_x as measured by an observer stationary in S is $v_x = dx/dt$. Its velocity v'_x as measured by an observer stationary in S' is $v'_x = dx'/dt$. We can derive a relationship between v_x and v'_x by taking the derivative with respect to t of the first of Eqs. (37.1):

$$\frac{dx}{dt} = \frac{dx'}{dt} + u$$

Now dx/dt is the velocity v_x measured in S, and dx'/dt is the velocity v'_x measured in S', so we get the *Galilean velocity transformation* for one-dimensional motion:

$$v_x = v'_x + u \qquad \text{(Galilean velocity transformation)} \tag{37.2}$$

Although the notation differs, this result agrees with our discussion of relative velocities in Section 3.5.

Now here's the fundamental problem. Applied to the speed of light in vacuum, Eq. (37.2) says that $c = c' + u$. Einstein's second postulate, supported subsequently by a wealth of experimental evidence, says that $c = c'$. This is a genuine inconsistency, not an illusion, and it demands resolution. If we accept this postulate, we are forced to conclude that Eqs. (37.1) and (37.2) *cannot* be precisely correct, despite our convincing derivation. These equations have to be modified to bring them into harmony with this principle.

The resolution involves some very fundamental modifications in our kinematic concepts. The first idea to be changed is the seemingly obvious assumption that the observers in frames S and S' use the same *time scale,* formally stated as $t = t'$. Alas, we are about to show that this everyday assumption cannot be correct; the two observers *must* have different time scales. We must define the velocity v' in frame S' as $v' = dx'/dt'$, not as dx'/dt; the two quantities are not the same. The difficulty lies in the concept of *simultaneity,* which is our next topic. A careful analysis of simultaneity will help us develop the appropriate modifications of our notions about space and time.

Test Your Understanding of Section 37.1 As a high-speed spaceship flies past you, it fires a strobe light that sends out a pulse of light in all directions. An observer aboard the spaceship measures a spherical wave front that spreads away from the spaceship with the same speed c in all directions. (a) What is the shape of the wave front that *you* measure? (i) spherical; (ii) ellipsoidal, with the longest axis of the ellipsoid along the direction of the spaceship's motion; (iii) ellipsoidal, with the shortest axis of the ellipsoid along the direction of the spaceship's motion; (iv) not enough information is given to decide. (b) Is the wave front centered on the spaceship?

37.2 Relativity of Simultaneity

Measuring times and time intervals involves the concept of **simultaneity.** In a given frame of reference, an **event** is an occurrence that has a definite position and time (Fig. 37.4). When you say that you awoke at seven o'clock, you mean that two events (your awakening and your clock showing 7:00) occurred *simultaneously.* The fundamental problem in measuring time intervals is this: In general, two events that are simultaneous in one frame of reference are *not* simultaneous in a second frame that is moving relative to the first, even if both are inertial frames.

37.4 An event has a definite position and time—for instance, on the pavement directly below the center of the Eiffel Tower at midnight on New Year's Eve.

A Thought Experiment in Simultaneity

This may seem to be contrary to common sense. To illustrate the point, here is a version of one of Einstein's *thought experiments*—mental experiments that follow concepts to their logical conclusions. Imagine a train moving with a speed comparable to c, with uniform velocity (Fig. 37.5). Two lightning bolts strike a passenger car, one near each end. Each bolt leaves a mark on the car and one on the ground at the instant the bolt hits. The points on the ground are labeled A and B in the figure, and the corresponding points on the car are A' and B'. Stanley is stationary on the ground at O, midway between A and B. Mavis is moving with the train at O' in the middle of the passenger car, midway between A' and B'. Both Stanley and Mavis see both light flashes emitted from the points where the lightning strikes.

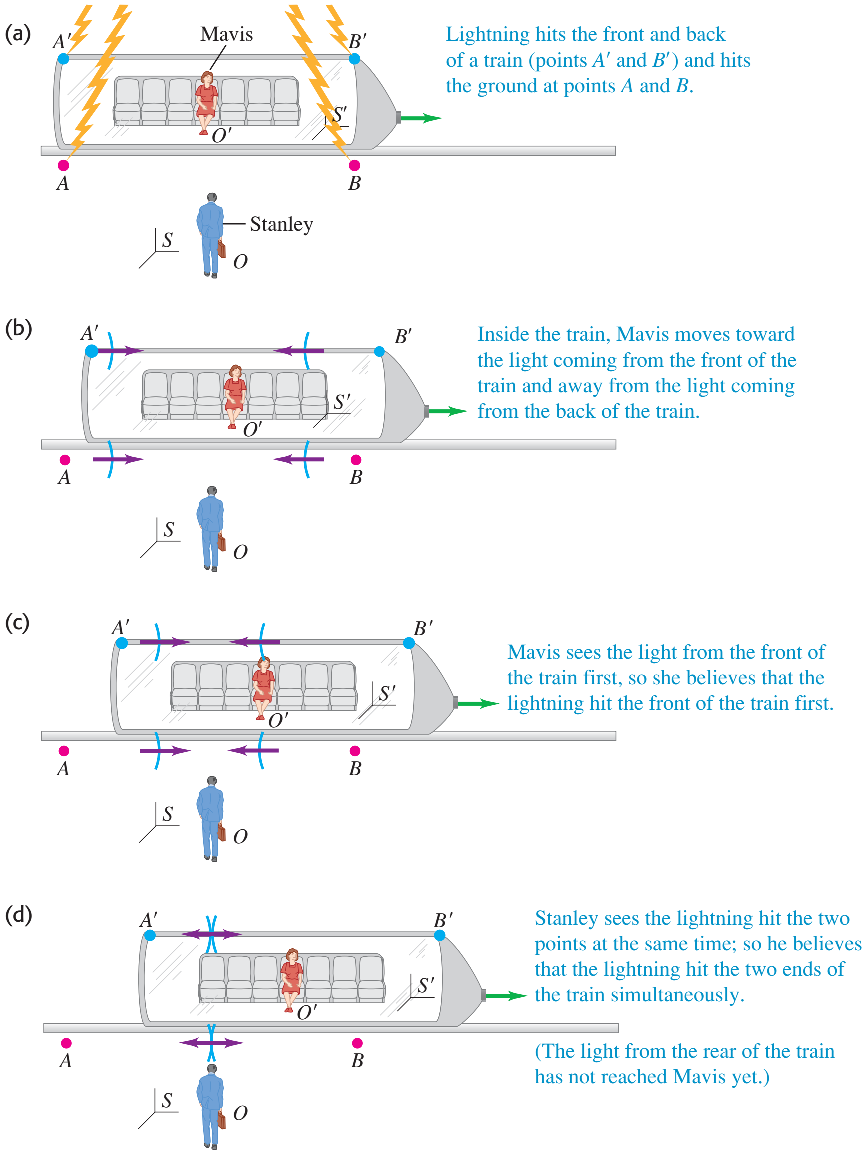
Suppose the two wave fronts from the lightning strikes reach Stanley at O simultaneously. He knows that he is the same distance from B and A, so Stanley concludes that the two bolts struck B and A simultaneously. Mavis agrees that the two wave fronts reached Stanley at the same time, but she disagrees that the flashes were emitted simultaneously.

Stanley and Mavis agree that the two wave fronts do not reach Mavis at the same time. Mavis at O' is moving to the right with the train, so she runs into 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.

37.5 A thought experiment in simultaneity.



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.

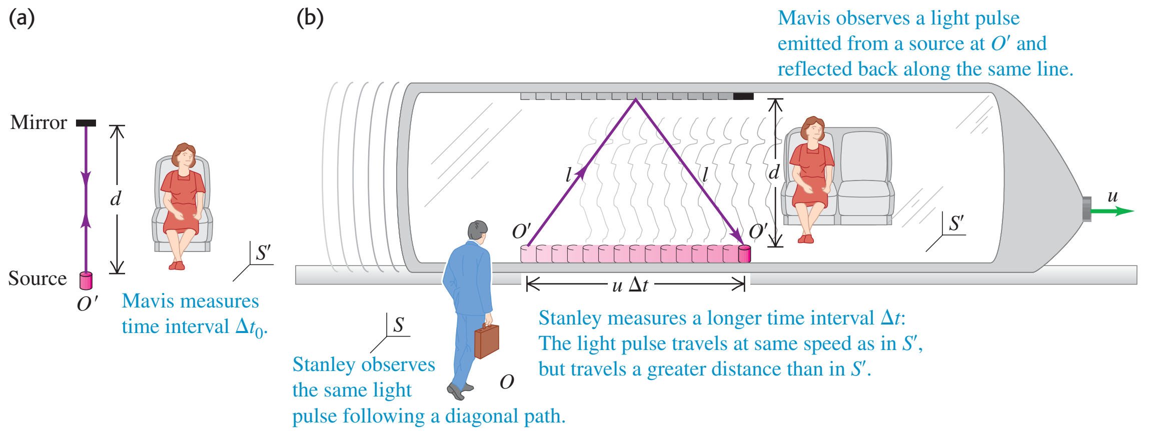
MasteringPHYSICS

ActivPhysics 17.1: Relativity of Time

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 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

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.



$$\Delta t_0 = \frac{2d}{c} \tag{37.3}$$

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 will justify this assumption in the next section. The speed of light is the same for both observers, so the round-trip time measured in S is

$$\Delta t = \frac{2l}{c} = \frac{2}{c}\sqrt{d^2 + \left(\frac{u\ \Delta t}{2}\right)^2} \tag{37.4}$$

We would like to have a relationship between Δt and Δt_0 that is independent of d. To get this, we solve Eq. (37.3) for d and substitute the result into Eq. (37.4), obtaining

$$\Delta t = \frac{2}{c}\sqrt{\left(\frac{c\ \Delta t_0}{2}\right)^2 + \left(\frac{u\ \Delta t}{2}\right)^2} \tag{37.5}$$

Now we square this and solve for Δt; the result is

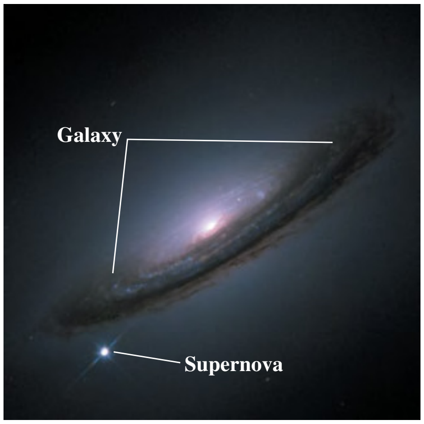
$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}}$$

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

We may generalize this important result. In a particular frame of reference, suppose that two events occur at the same point in space. The time interval between these events, as measured by an observer at rest in this same frame (which we call the *rest frame* of this observer), is Δt_0. Then 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

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.



$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} \quad \text{(time dilation)} \tag{37.6}$$

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) appears so often in relativity that it is given its own symbol γ (the Greek letter gamma):

$$\gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \tag{37.7}$$

In terms of this symbol, we can express the time dilation formula, Eq. (37.6), as

$$\Delta t = \gamma\, \Delta t_0 \quad \text{(time dilation)} \tag{37.8}$$

As a further simplification, u/c is sometimes given the symbol β (the Greek letter beta); then $\gamma = 1/\sqrt{1 - \beta^2}$.

37.8 The quantity $\gamma = 1/\sqrt{1 - u^2/c^2}$ as a function of the relative speed u of two frames of reference.

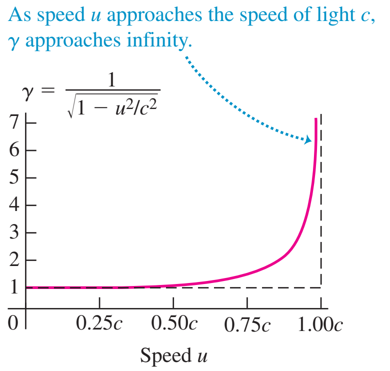


Figure 37.8 shows a graph of γ as a function of the relative speed u of two frames of reference. When u is very small compared to c, u^2/c^2 is much smaller than 1 and γ is very nearly *equal* to 1. In that limit, Eqs. (37.6) and (37.8) approach the Newtonian relationship $\Delta t = \Delta t_0$, corresponding 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.

Proper Time

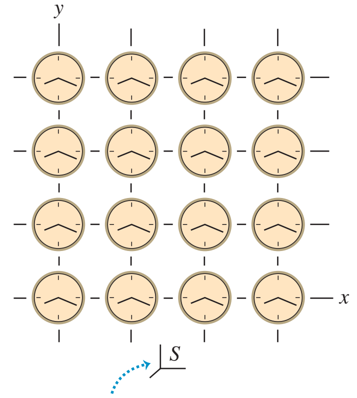
There is only one frame of reference in which a clock is at rest, and there are infinitely many in which it is moving. Therefore the time interval measured between two events (such as two ticks of the clock) that occur at the same point in a particular frame is a more fundamental quantity than the interval between events at different points. We use the term **proper time** to describe the time interval Δt_0 between two events that occur *at the same point.*

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 is able to 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 note that clocks that are synchronized in one frame of reference *are not* in general synchronized in any other frame.)

In thought experiments, it's often helpful to imagine many observers with synchronized clocks at rest at various points in a particular frame of reference. We can picture a frame of reference as a coordinate grid with lots of synchronized clocks distributed around it, as suggested by Fig. 37.9. Only when a clock is moving relative to a given frame of reference do we have to watch for ambiguities of synchronization or simultaneity.

Throughout this chapter we will frequently use phrases like "Stanley *observes* that Mavis passes the point $x = 5.00$ m, $y = 0$, $z = 0$ at time 2.00 s." This means that Stanley is using a grid of clocks in his frame of reference, like the grid shown in Fig. 37.9, to record the time of an event. We could restate the phrase as "When Mavis passes the point at $x = 5.00$ m, $y = 0$, $z = 0$, the clock at that location in Stanley's frame of reference reads 2.00 s." We will avoid using phrases like "Stanley *sees* that Mavis is 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.

37.9 A frame of reference pictured as a coordinate system with a grid of synchronized clocks.



The grid is three dimensional; identical planes of clocks lie in front of and behind the page, connected by grid lines perpendicular to the page.

Problem-Solving Strategy 37.1 Time Dilation

IDENTIFY *the relevant concepts:* The concept of time dilation is used whenever we compare the time intervals between events as measured by observers in different inertial frames of reference.

SET UP *the problem* using the following steps:
1. First decide what two events define the beginning and the end of the time interval. Then identify the two frames of reference in which the time interval is measured.
2. Identify the target variable.

EXECUTE *the solution* as follows:
1. In many problems, the time interval as measured in one frame of reference is the *proper* time Δt_0. This is the time interval between two events in a frame of reference in which the two events occur at the same point in space. In a second frame of reference that has a speed u relative to that first frame, there is a longer time interval Δt between the same two events. In this second frame the two events occur at different points. You will need to decide in which frame the time interval is Δt_0 and in which frame it is Δt.
2. Use Eq. (37.6) or (37.8) to relate Δt_0 and Δt, and then solve for the target variable.

EVALUATE *your answer:* Note that Δt is never smaller than Δt_0, and u is never greater than c. If your results suggest otherwise, you need to rethink your calculation.

Example 37.1 Time dilation at 0.990*c*

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?

SOLUTION

IDENTIFY and SET UP: The muon's lifetime is the time interval between two events: the production of the muon and its subsequent decay. Our target variable is the lifetime in your frame of reference on earth, which we call frame S. We are given the lifetime in a frame S' in which the muon is at rest; this is its *proper* lifetime, $\Delta t_0 = 2.20\ \mu\text{s}$. 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).

Example 37.2 Time dilation at airliner speeds

An airplane flies from San Francisco to New York (about 4800 km, or 4.80×10^6 m) at a steady speed of 300 m/s (about 670 mi/h). How much time does the trip take, as measured by an observer on the ground? By an observer in the plane?

SOLUTION

IDENTIFY and SET UP: Here we're interested in the time interval between the airplane departing from San Francisco and landing in New York. The target variables are the time intervals as measured in the frame of reference of the ground S and in the frame of reference of the airplane S'.

EXECUTE: As measured in S the two events occur at different positions (San Francisco and New York), so the time interval measured by ground observers corresponds to Δt in Eq. (37.6). To find it, we simply divide the distance by the speed $u = 300$ m/s:

$$\Delta t = \frac{4.80 \times 10^6\ \text{m}}{300\ \text{m/s}} = 1.60 \times 10^4\ \text{s} \quad (\text{about } 4\tfrac{1}{2}\text{ hours})$$

In the airplane's frame S', San Francisco and New York passing under the plane occur at the same point (the position of the plane). Hence the time interval in the airplane is a proper time, corresponding to Δt_0 in Eq. (37.6). We have

$$\frac{u^2}{c^2} = \frac{(300\ \text{m/s})^2}{(3.00 \times 10^8\ \text{m/s})^2} = 1.00 \times 10^{-12}$$

From Eq. (37.6),

$$\Delta t_0 = (1.60 \times 10^4\ \text{s})\sqrt{1 - 1.00 \times 10^{-12}}$$

The square root can't be evaluated with adequate precision with an ordinary calculator. But we can approximate it using the binomial theorem (see Appendix B):

$$(1 - 1.00 \times 10^{-12})^{1/2} = 1 - \left(\tfrac{1}{2}\right)(1.00 \times 10^{-12}) + \cdots$$

The remaining terms are of the order of 10^{-24} or smaller and can be discarded. The approximate result for Δt_0 is

$$\Delta t_0 = (1.60 \times 10^4\ \text{s})(1 - 0.50 \times 10^{-12})$$

The proper time Δt_0, measured in the airplane, is very slightly less (by less than one part in 10^{12}) than the time measured on the ground.

EVALUATE: We don't notice such effects in everyday life. But present-day atomic clocks (see Section 1.3) can attain a precision of about one part in 10^{13}. A cesium clock traveling a long distance in an airliner has been used to measure this effect and thereby verify Eq. (37.6) even at speeds much less than c.

Example 37.3 Just when is it proper?

Mavis boards a spaceship and then zips past Stanley on earth at a relative speed of $0.600c$. At the instant she passes him, they both start timers. (a) A short time later Stanley measures that Mavis has traveled 9.00×10^7 m beyond him and is passing a space station. What does Stanley's timer read as she passes the space station? What does Mavis's timer read? (b) Stanley starts to blink just as Mavis flies past him, and Mavis measures that the blink takes 0.400 s from beginning to end. According to Stanley, what is the duration of his blink?

SOLUTION

IDENTIFY and SET UP: This problem involves time dilation for two *different* sets of events measured in Stanley's frame of reference (which we call S) and in Mavis's frame of reference (which we call S'). The two events of interest in part (a) are when Mavis passes Stanley and when Mavis passes the space station; the target variables are the time intervals between these two events as measured in S and in S'. The two events in part (b) are the start and finish of Stanley's blink; the target variable is the time interval between these two events as measured in S.

EXECUTE: (a) The two events, Mavis passing the earth and Mavis passing the space station, occur at different positions in Stanley's frame but at the same position in Mavis's frame. Hence Stanley measures time interval Δt, while Mavis measures the *proper* time Δt_0. As measured by Stanley, Mavis moves at $0.600c = 0.600(3.00 \times 10^8\ \text{m/s}) = 1.80 \times 10^8$ m/s and travels 9.00×10^7 m in time $\Delta t = (9.00 \times 10^7\ \text{m})/(1.80 \times 10^8\ \text{m/s}) = 0.500$ s. From Eq. (37.6), Mavis's timer reads an elapsed time of

$$\Delta t_0 = \Delta t\sqrt{1 - u^2/c^2} = 0.500\ \text{s}\sqrt{1 - (0.600)^2} = 0.400\ \text{s}$$

(b) It is tempting to answer that Stanley's blink lasts 0.500 s in his frame. But this is wrong, because we are now considering a *different* pair of events than in part (a). The start and finish of Stanley's blink occur at the same point in his frame S but at different positions in Mavis's frame S', so the time interval of 0.400 s that she measures between these events is equal to Δt. The duration of the blink measured on Stanley's timer is the proper time Δt_0:

$$\Delta t_0 = \Delta t\sqrt{1 - u^2/c^2} = 0.400\ \text{s}\sqrt{1 - (0.600)^2} = 0.320\ \text{s}$$

EVALUATE: This example illustrates the relativity of simultaneity. In Mavis's frame she passes the space station at the same instant that Stanley finishes his blink, 0.400 s after she passed Stanley. Hence these two events are simultaneous to Mavis in frame S'. But these two events are *not* simultaneous to Stanley in his frame S: According to his timer, he finishes his blink after 0.320 s and Mavis passes the space station after 0.500 s.

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.

Now 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, we recognize 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 inertial 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.

Application Who's the Grandmother?
The answer to this question may seem obvious, but it could depend on which person had traveled to a distant planet at relativistic speeds. Imagine that a 20-year-old woman had given birth to a child and then immediately left on a 100-light-year trip (50 light-years out and 50 light-years back) at 99.5% the speed of light. Because of time dilation for the traveler, only 10 years would pass, and she would be 30 years old when she returned, even though 100 years had passed by for people on earth. Meanwhile, the child she left behind at home could have had a baby 20 years after her departure, and this grandchild would now be 80 years old!

Test Your Understanding of Section 37.3 Samir (who is standing on the ground) starts his stopwatch at the instant that Maria flies past him in her spaceship at a speed of $0.600c$. At the same instant, Maria starts her stopwatch. (a) As measured in Samir's frame of reference, what is the reading on Maria's stopwatch at the instant that Samir's stopwatch reads 10.0 s? (i) 10.0 s; (ii) less than 10.0 s; (iii) more than 10.0 s. (b) As measured in Maria's frame of reference, what is the reading on Samir's stopwatch at the instant that Maria's stopwatch reads 10.0 s? (i) 10.0 s; (ii) less than 10.0 s; (iii) more than 10.0 s.

37.4 Relativity of Length

Not only does the time interval between two events depend on the observer's frame of reference, but the *distance* between two points may also depend on the observer's frame of reference. The concept of simultaneity is involved. Suppose you want to measure the length of a moving car. One way is to have two assistants make marks on the pavement at the positions of the front and rear bumpers. Then you measure the distance between the marks. But your assistants have to make their marks *at the same time.* If one marks the position of the front bumper at one time and the other marks the position of the rear bumper half a second later, you won't get the car's true length. Since we've learned that simultaneity isn't an absolute concept, we have to proceed with caution.

MasteringPHYSICS
ActivPhysics 17.2: Relativity of Length

Lengths Parallel to the Relative Motion

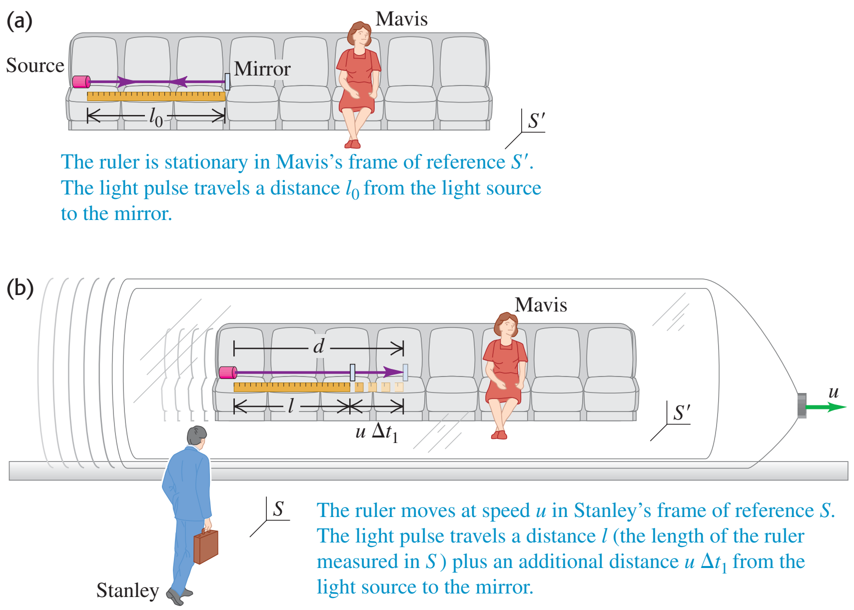
To develop a relationship between lengths that are measured parallel to the direction of motion in various coordinate systems, we consider another thought experiment. We attach a light source to one end of a ruler and a mirror to the other end. The ruler is at rest in reference frame S', and its length in this frame is l_0 (Fig. 37.10a). Then the time Δt_0 required for a light pulse to make the round trip from source to mirror and back is

$$\Delta t_0 = \frac{2l_0}{c} \tag{37.9}$$

This is a proper time interval because departure and return occur at the same point in S'.

In reference frame S the ruler is moving to the right with speed u during this travel of the light pulse (Fig. 37.10b). The length of the ruler in S is l, and the time of travel from source to mirror, as measured in S, is Δt_1. During this interval

37.10 (a) A ruler is at rest in Mavis's frame S'. A light pulse is emitted from a source at one end of the ruler, reflected by a mirror at the other end, and returned to the source position. (b) Motion of the light pulse as measured in Stanley's frame S.



the 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, combining Eqs. (37.14) and (37.15) to eliminate Δt and simplifying, we obtain

$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = \frac{l_0}{\gamma} \qquad \text{(length contraction)} \tag{37.16}$$

[We have used the quantity $\gamma = 1/\sqrt{1 - u^2/c^2}$ defined in Eq. (37.7).] Thus the length l measured in S, in which the ruler is moving, is *shorter* than the length l_0 measured in its rest frame S'.

CAUTION **Length contraction is real** This is *not* an optical illusion! The ruler really is shorter in reference frame S than it is in S'.

A length measured in the frame in which the body is at rest (the rest frame of the body) is called a **proper length;** thus l_0 is a proper length in S', and the length measured in any other frame moving relative to S' is *less than* l_0. This effect is called **length contraction.**

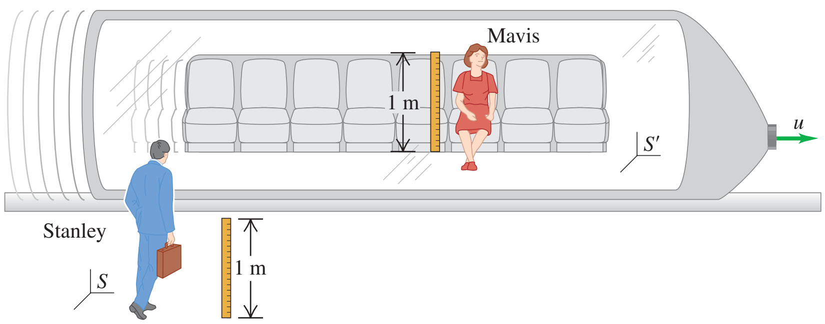
When u is very small in comparison to c, γ approaches 1. Thus in the limit of small speeds we approach the Newtonian relationship $l = l_0$. This and the corresponding result for time dilation show that Eqs. (37.1), the Galilean coordinate transformation, are usually sufficiently accurate for relative speeds much smaller than c. If u is a reasonable fraction of c, however, the quantity $\sqrt{1 - u^2/c^2}$ can be appreciably less than 1. Then l can be substantially smaller than l_0, and the effects of length contraction can be substantial (Fig. 37.11).

37.11 The speed at which electrons traverse the 3-km beam line of the SLAC National Accelerator Laboratory is slower than c by less than 1 cm/s. As measured in the reference frame of such an electron, the beam line (which extends from the top to the bottom of this photograph) is only about 15 cm long!

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.



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.

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?

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?

SOLUTION

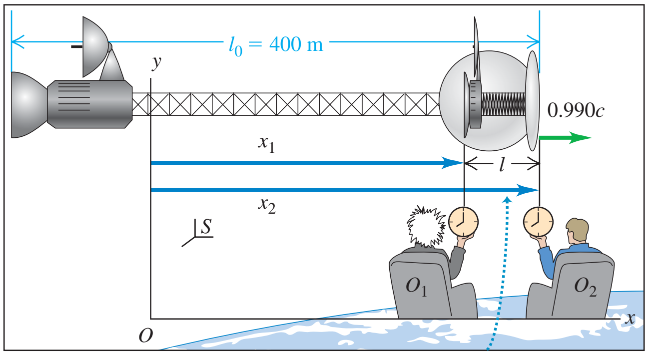
IDENTIFY and SET UP: This problem is about the nose-to-tail length of the spaceship as measured on the spaceship and on earth. This length is along the direction of relative motion (Fig. 37.13), so there will be length contraction. The spaceship's 400-m length is the *proper* length l_0 because it is measured in the frame in which the spaceship is at rest. Our target variable is the length l measured in the earth frame, relative to which the spaceship is moving at $u = 0.990c$.

EXECUTE: From Eq. (37.16), the length in the earth frame is

$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = (400\ \text{m})\sqrt{1 - (0.990)^2} = 56.4\ \text{m}$$

EVALUATE: The spaceship is shorter in a frame in which it is in motion than in a frame in which it is at rest. To measure the length l, two earth observers with synchronized clocks could measure the positions of the two ends of the spaceship simultaneously in the earth's reference frame, as shown in Fig. 37.13. (These two measurements will *not* appear simultaneous to an observer in the spaceship.)

37.13 Measuring the length of a moving spaceship.



The two observers on earth (S) must measure x_2 and x_1 simultaneously to obtain the correct length $l = x_2 - x_1$ in their frame of reference.

Example 37.5 How far apart are the observers?

Observers O_1 and O_2 in Fig. 37.13 are 56.4 m apart on the earth. How far apart does the spaceship crew measure them to be?

SOLUTION

IDENTIFY and SET UP: In this example the 56.4-m distance is the *proper* length l_0. It represents the length of a ruler that extends from O_1 to O_2 and is at rest in the earth frame in which the observers are at rest. Our target variable is the length l of this ruler measured in the spaceship frame, in which the earth and ruler are moving at $u = 0.990c$.

EXECUTE: As in Example 37.4, but with $l_0 = 56.4$ m,

$$l = l_0\sqrt{1 - \frac{u^2}{c^2}} = (56.4 \text{ m})\sqrt{1 - (0.990)^2} = 7.96 \text{ m}$$

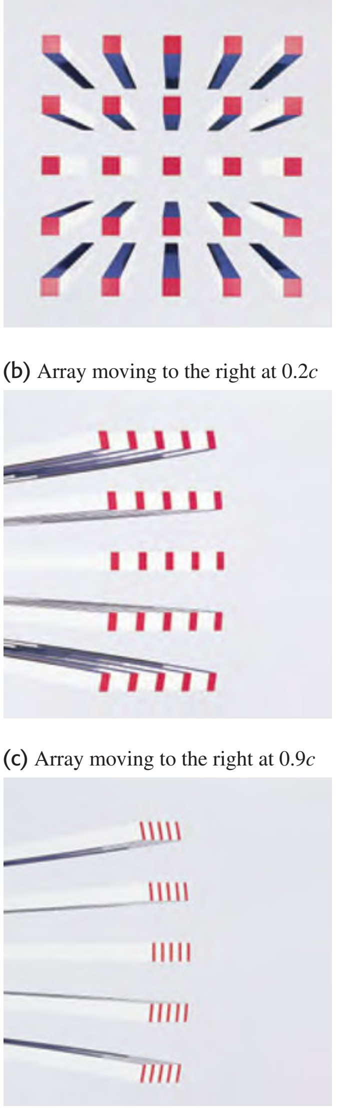
EVALUATE: This answer does *not* say that the crew measures their spaceship to be both 400 m long and 7.96 m long. As measured on earth, the tail of the spacecraft is at the position of O_1 at the same instant that the nose of the spacecraft is at the position of O_2. Hence the length of the spaceship measured on earth equals the 56.4-m distance between O_1 and O_2. But in the spaceship frame O_1 and O_2 are only 7.96 m apart, and the nose (which is 400 m in front of the tail) passes O_2 before the tail passes O_1.

How an Object Moving Near c Would Appear

Let's think a little about the visual appearance of a moving three-dimensional body. If we could see the positions of all points of the body simultaneously, it would appear to shrink only in the direction of motion. But we *don't* see all the points simultaneously; light from points farther from us takes longer to reach us than does light from points near to us, so we see the farther points at the positions they had at earlier times.

Suppose we have a rectangular rod with its faces parallel to the coordinate planes. When we look end-on at the center of the closest face of such a rod at rest, we see only that face. (See the center rod in computer-generated Fig. 37.14a.) But when that rod is moving past us toward the right at an appreciable fraction of the speed of light, we may also see its left side because of the earlier-time effect just described. That is, we can see some points that we couldn't see when the rod was at rest because the rod moves out of the way of the light rays from those points to us. Conversely, some light that can get to us when the rod is at rest is blocked by the moving rod. Because of all this, the rods in Figs. 37.14b and 37.14c appear rotated and distorted.

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 $0.9c$

Test Your Understanding of Section 37.4 A miniature spaceship is flying past you, moving horizontally at a substantial fraction of the speed of light. At a certain instant, you observe that the nose and tail of the spaceship align exactly with the two ends of a meter stick that you hold in your hands. Rank the following distances in order from longest to shortest: (i) the proper length of the meter stick; (ii) the proper length of the spaceship; (iii) the length of the spaceship measured in your frame of reference; (iv) the length of the meter stick measured in the spaceship's frame of reference.

37.5 The Lorentz Transformations

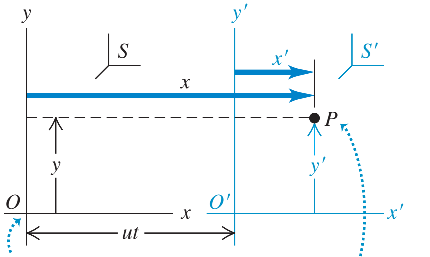
In Section 37.1 we discussed the Galilean coordinate transformation equations, Eqs. (37.1). They relate the coordinates (x, y, z) of a point in frame of reference S to the coordinates (x', y', z') of the point in a second frame S'. The second frame moves with constant speed u relative to S in the positive direction along the common x-x'-axis. This transformation also assumes that the time scale is the same in the two frames of reference, as expressed by the additional relationship $t = t'$. 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

37.15 As measured in frame of reference S, x' is contracted to x'/γ, so $x = ut + x'/\gamma$ and $x' = \gamma(x - ut)$.



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 seen 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

$$x' = -ut' + x\sqrt{1 - \frac{u^2}{c^2}} \tag{37.19}$$

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. We leave the algebraic details for you to work out; the result is

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} \tag{37.20}$$

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 all these transformation equations, we have

$$\begin{aligned} x' &= \frac{x - ut}{\sqrt{1 - u^2/c^2}} = \gamma(x - ut) \\ y' &= y \\ z' &= z \\ t' &= \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} = \gamma(t - ux/c^2) \end{aligned} \quad \text{(Lorentz coordinate transformation)} \tag{37.21}$$

These equations are the *Lorentz coordinate transformation,* the relativistic generalization of the Galilean coordinate transformation, Eqs. (37.1) and $t = t'$. For values of u that approach zero, the radicals in the denominators and γ approach 1, and the ux/c^2 term approaches zero. In this limit, Eqs. (37.21) become identical to Eqs. (37.1) along with $t = t'$. In general, though, both the coordinates and time of an event in one frame depend on its coordinates and time in another frame. *Space and time have become intertwined; we can no longer say that length and time have absolute meanings independent of the frame of reference.* For this reason, we refer to time and the three dimensions of space collectively as a four-dimensional entity called **spacetime,** and we call (x, y, z, t) together the **spacetime coordinates** of an event.

The Lorentz Velocity Transformation

We can use Eqs. (37.21) to derive the relativistic generalization of the Galilean velocity transformation, Eq. (37.2). We consider only one-dimensional motion along the x-axis and use the term "velocity" as being short for the "x-component of the velocity." Suppose that in a time dt a particle moves a distance dx, as measured in frame S. We obtain the corresponding distance dx' and time dt' in S' by taking differentials of Eqs. (37.21):

$$dx' = \gamma(dx - u\,dt)$$
$$dt' = \gamma(dt - u\,dx/c^2)$$

We divide the first equation by the second and then divide the numerator and denominator of the result by dt to obtain

$$\frac{dx'}{dt'} = \frac{\dfrac{dx}{dt} - u}{1 - \dfrac{u}{c^2}\dfrac{dx}{dt}}$$

Now dx/dt is the velocity v_x in S, and dx'/dt' is the velocity v'_x in S', so we finally obtain the relativistic generalization

$$v'_x = \frac{v_x - u}{1 - uv_x/c^2} \quad \text{(Lorentz velocity transformation)} \tag{37.22}$$

When u and v_x are much smaller than c, the denominator in Eq. (37.22) approaches 1, and we approach the nonrelativistic result $v'_x = v_x - u$. The opposite extreme is the case $v_x = c$; then we find

$$v'_x = \frac{c - u}{1 - uc/c^2} = \frac{c(1 - u/c)}{1 - u/c} = c$$

This says that anything moving with velocity $v_x = c$ measured in S also has velocity $v'_x = c$ measured in S', despite the relative motion of the two frames. So Eq. (37.22) is consistent with Einstein's postulate that the speed of light in vacuum is the same in all inertial frames of reference.

The principle of relativity tells us there is no fundamental distinction between the two frames S and S'. Thus the expression for v_x in terms of v'_x must have the same form as Eq. (37.22), with v_x changed to v'_x, and vice versa, and the sign of u reversed. Carrying out these operations with Eq. (37.22), we find

$$v_x = \frac{v'_x + u}{1 + uv'_x/c^2} \quad \text{(Lorentz velocity transformation)} \tag{37.23}$$

This can also be obtained algebraically by solving Eq. (37.22) for v_x. Both Eqs. (37.22) and (37.23) are *Lorentz velocity transformations* for one-dimensional motion.

CAUTION **Use the correct reference frame coordinates** Keep in mind that 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. You should always set up your coordinate system to follow this convention.

When u is less than c, the Lorentz velocity transformations show us that a body moving with a speed less than c in one frame of reference always has a speed less than c in *every other* frame of reference. This is one reason for concluding that no material body may travel with a speed equal to or greater than that of light in vacuum, relative to *any* inertial frame of reference. The relativistic generalizations of energy and momentum, which we will explore later, give further support to this hypothesis.

Problem-Solving Strategy 37.3 Lorentz Transformations

IDENTIFY *the relevant concepts:* The Lorentz *coordinate* transformation equations relate the spacetime coordinates of an event in one inertial reference frame to the coordinates of the same event in a second inertial frame. The Lorentz *velocity* transformation equations relate the velocity of an object in one inertial reference frame to its velocity in a second inertial frame.

SET UP *the problem* using the following steps:

1. Identify the target variable.
2. Define the two inertial frames S and S'. Remember that S' moves relative to S at a constant velocity u in the $+x$-direction.
3. If the coordinate transformation equations are needed, make a list of spacetime coordinates in the two frames, such as x_1, x_1', t_1, t_1', and so on. Label carefully which of these you know and which you don't.
4. In velocity-transformation problems, clearly identify u (the relative velocity of the two frames of reference), v_x (the velocity of the object relative to S), and v_x' (the velocity of the object relative to S').

EXECUTE *the solution* as follows:

1. In a coordinate-transformation problem, use Eqs. (37.21) to solve for the spacetime coordinates of the event as measured in S' in terms of the corresponding values in S. (If you need to solve for the spacetime coordinates in S in terms of the corresponding values in S', you can easily convert the expressions in Eqs. (37.21): Replace all of the primed quantities with unprimed ones, and vice versa, and replace u with $-u$.)
2. In a velocity-transformation problem, use either Eq. (37.22) or Eq. (37.23), as appropriate, to solve for the target variable.

EVALUATE *your answer:* Don't be discouraged if some of your results don't seem to make sense or if they disagree with "common sense." It takes time to develop intuition about relativity; you'll gain it with experience.

Example 37.6 Was it received before it was sent?

Winning an interstellar race, Mavis pilots her spaceship across a finish line in space at a speed of $0.600c$ relative to that line. A "hooray" message is sent from the back of her ship (event 2) at the instant (in her frame of reference) that the front of her ship crosses the line (event 1). She measures the length of her ship to be 300 m. Stanley is at the finish line and is at rest relative to it. When and where does he measure events 1 and 2 to occur?

SOLUTION

IDENTIFY and SET UP: This example involves the Lorentz coordinate transformation. Our derivation of this transformation assumes that the origins of frames S and S' coincide at $t = 0 = t'$. Thus for simplicity we fix the origin of S at the finish line and the origin of S' at the front of the spaceship so that Stanley and Mavis measure event 1 to be at $x = 0 = x'$ and $t = 0 = t'$.

Mavis in S' measures her spaceship to be 300 m long, so she has the "hooray" sent from 300 m behind her spaceship's front at the instant she measures the front to cross the finish line. That is, she measures event 2 at $x' = -300$ m and $t' = 0$.

Our target variables are the coordinate x and time t of event 2 that Stanley measures in S.

EXECUTE: To solve for the target variables, we modify the first and last of Eqs. (37.21) to give x and t as functions of x' and t'. We do so in the same way that we obtained Eq. (37.23) from Eq. (37.22). We remove the primes from x' and t', add primes to x and t, and replace each u with $-u$. The results are

$$x = \gamma(x' + ut') \quad \text{and} \quad t = \gamma(t' + ux'/c^2)$$

From Eq. (37.7), $\gamma = 1.25$ for $u = 0.600c = 1.80 \times 10^8$ m/s. We also substitute $x' = -300$ m, $t' = 0$, $c = 3.00 \times 10^8$ m/s, and $u = 1.80 \times 10^8$ m/s in the equations for x and t to find $x = -375$ m at $t = -7.50 \times 10^{-7}$ s $= -0.750\ \mu$s for event 2.

EVALUATE: Mavis says that the events are simultaneous, but Stanley says that the "hooray" was sent *before* Mavis crossed the finish line. This does not mean that the effect preceded the cause. The fastest that Mavis can send a signal the length of her ship is $300\text{ m}/(3.00 \times 10^8\text{ m/s}) = 1.00\ \mu$s. She cannot send a signal from the front at the instant it crosses the finish line that would cause a "hooray" to be broadcast from the back at the same instant. She would have to send that signal from the front at least $1.00\ \mu$s before then, so she had to slightly anticipate her success.

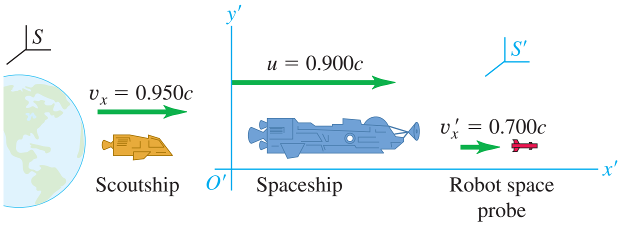
Example 37.7 Relative velocities

(a) A spaceship moving away from the earth at $0.900c$ fires a robot space probe in the same direction as its motion at $0.700c$ relative to the spaceship. What is the probe's velocity relative to the earth? (b) A scoutship is sent to catch up with the spaceship by traveling at $0.950c$ relative to the earth. What is the velocity of the scoutship relative to the spaceship?

SOLUTION

IDENTIFY and SET UP: This example uses the Lorentz velocity transformation. Let the earth and spaceship reference frames be S and S', respectively (Fig. 37.16); their relative velocity is $u = 0.900c$. In part (a) we are given the probe velocity $v_x' = 0.700c$ with respect to S', and the target variable is the velocity v_x of the

37.16 The spaceship, robot space probe, and scoutship.



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:

$$v_x' = \frac{v_x - u}{1 - uv_x/c^2} = \frac{0.950c - 0.900c}{1 - (0.900c)(0.950c)/c^2} = 0.345c$$

EVALUATE: What would the Galilean velocity transformation formula, Eq. (37.2), say? In part (a) we would have found the probe's velocity relative to the earth to be $v_x = v_x' + u = 0.700c + 0.900c = 1.600c$, which is greater than c and hence impossible. In part (b), we would have found the scoutship's velocity relative to the spaceship to be $v_x' = v_x - u = 0.950c - 0.900c = 0.050c$; the relativistically correct value, $v_x' = 0.345c$, is almost seven times greater than the incorrect Galilean value.

Test Your Understanding of Section 37.5 (a) In frame S events P_1 and P_2 occur at the same x-, y-, and z-coordinates, but event P_1 occurs before event P_2. In frame S', which event occurs first? (b) In frame S events P_3 and P_4 occur at the same time t and the same y- and z-coordinates, but event P_3 occurs at a less positive x-coordinate than event P_4. In frame S', which event occurs first?

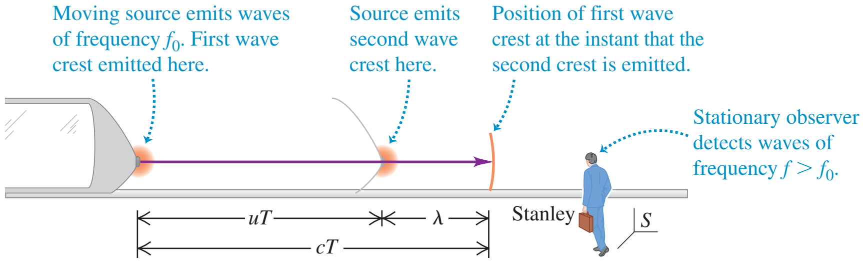
37.6 The Doppler Effect for Electromagnetic Waves

An additional important consequence of relativistic kinematics is the Doppler effect for electromagnetic waves. In our previous discussion of the Doppler effect (see 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



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.

successive crests—that is, the wavelength—is thus $\lambda = (c - u)T$, as measured in Stanley's frame. The frequency that he measures is c/λ. Therefore

$$f = \frac{c}{(c - u)T} \tag{37.24}$$

So far we have followed a pattern similar to that for the Doppler effect for sound from a moving source (see Section 16.8). In that discussion our next step was to equate T to the time T_0 between emissions of successive wave crests by the source. However, due to time dilation it is *not* relativistically correct to equate T to T_0. The time T_0 is measured in the rest frame of the source, so it is a proper time. From Eq. (37.6), T_0 and T are related by

$$T = \frac{T_0}{\sqrt{1 - u^2/c^2}} = \frac{cT_0}{\sqrt{c^2 - u^2}}$$

or, since $T_0 = 1/f_0$,

$$\frac{1}{T} = \frac{\sqrt{c^2 - u^2}}{cT_0} = \frac{\sqrt{c^2 - u^2}}{c} f_0$$

Remember, $1/T$ is not equal to f. We must substitute this expression for $1/T$ into Eq. 37.24 to find f:

$$f = \frac{c}{c - u} \frac{\sqrt{c^2 - u^2}}{c} f_0$$

Using $c^2 - u^2 = (c - u)(c + u)$ gives

$$f = \sqrt{\frac{c + u}{c - u}} f_0 \quad \text{(Doppler effect, electromagnetic waves, source approaching observer)} \tag{37.25}$$

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

$$f = \sqrt{\frac{c - u}{c + u}} f_0 \quad \text{(Doppler effect, electromagnetic waves, source moving away from observer)} \tag{37.26}$$

This agrees with Eq. (16.30), which we quoted previously, with minor notation changes.

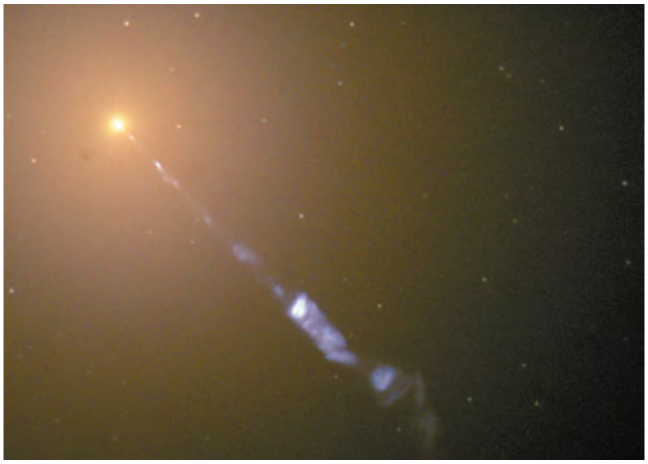
With light, unlike sound, there is no distinction between motion of source and motion of observer; only the *relative* velocity of the two is significant. The last four paragraphs of Section 16.8 discuss several practical applications of the Doppler effect with light and other electromagnetic radiation; we suggest you review those paragraphs now. Figure 37.18 shows one common application.

37.18 This handheld radar gun emits a radio beam of frequency f_0, which in the frame of reference of an approaching car has a higher frequency f given by Eq. (37.25). The reflected beam also has frequency f in the car's frame, but has an even higher frequency f' in the police officer's frame. The radar gun calculates the car's speed by comparing the frequencies of the emitted beam and the doubly Doppler-shifted reflected beam. (Compare Example 16.18 in Section 16.8.)

Example 37.8 A jet from a black hole

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

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).



frequency of 6.66×10^{14} Hz (in the far ultraviolet region of the electromagnetic spectrum; see Fig. 32.4), but in the reference frame of the jet material the light has a frequency of 5.55×10^{13} Hz (in the infrared). What is the speed of the jet material with respect to us?

SOLUTION

IDENTIFY and SET UP: This problem involves the Doppler effect for electromagnetic waves. The frequency we observe is $f = 6.66 \times 10^{14}$ Hz, and the frequency in the frame of the source is $f_0 = 5.55 \times 10^{13}$ Hz. Since $f > f_0$, the jet is approaching us and we use Eq. (37.25) to find the target variable u.

EXECUTE: We need to solve Eq. (37.25) for u. We'll leave it as an exercise for you to show that the result is

$$u = \frac{(f/f_0)^2 - 1}{(f/f_0)^2 + 1}c$$

We have $f/f_0 = (6.66 \times 10^{14}\text{ Hz})/(5.55 \times 10^{13}\text{ Hz}) = 12.0$, so

$$u = \frac{(12.0)^2 - 1}{(12.0)^2 + 1}c = 0.986c$$

EVALUATE: Because the frequency shift is quite substantial, it would have been erroneous to use the approximate expression $\Delta f/f = u/c$. Had you done so, you would have found $u = c(\Delta f/f_0) = c(6.66 \times 10^{14}\text{ Hz} - 5.55 \times 10^{13}\text{ Hz})/(5.55 \times 10^{13}\text{ Hz}) = 11.0c$. This result cannot be correct because the jet material cannot travel faster than light.

37.7 Relativistic Momentum

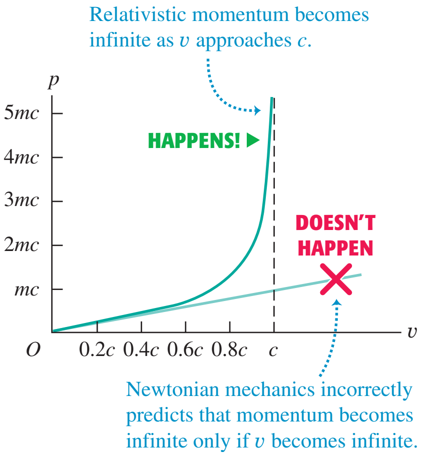
Newton's laws of motion have the same form in all inertial frames of reference. When we use transformations to change from one inertial frame to another, the laws should be *invariant* (unchanging). But we have just learned that the principle of relativity forces us to replace the Galilean transformations with the more general Lorentz transformations. As we will see, this requires corresponding generalizations in the laws of motion and the definitions of momentum and energy.

The principle of conservation of momentum states that *when two bodies interact, the total momentum is constant,* provided that the net external force acting on the bodies in an inertial reference frame is zero (for example, if they form an isolated system, interacting only with each other). If conservation of momentum is a valid physical law, it must be valid in *all* inertial frames of reference. Now, here's the problem: Suppose we look at a collision in one inertial coordinate system S and find that momentum is conserved. Then we use the Lorentz transformation to obtain the velocities in a second inertial system S'. We find that if we use the Newtonian definition of momentum ($\vec{p} = m\vec{v}$), momentum is *not* conserved in the second system! If we are convinced that the principle of relativity and the Lorentz transformation are correct, the only way to save momentum conservation is to generalize the *definition* of momentum.

We won't derive the correct relativistic generalization of momentum, but here is the result. Suppose we measure the mass of a particle to be m when it is at rest relative to us: We often call m the **rest mass.** We will use the term *material particle* for a particle that has a nonzero rest mass. When such a particle has a velocity $\vec{v}$, its **relativistic momentum** $\vec{p}$ is

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} \quad \text{(relativistic momentum)} \qquad (37.27)$$

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.



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.

Relativity, Newton's Second Law, and Relativistic Mass

What about the relativistic generalization of Newton's second law? In Newtonian mechanics the most general form of the second law is

$$\vec{F} = \frac{d\vec{p}}{dt} \tag{37.28}$$

That is, the net force $\vec{F}$ on a particle equals the time rate of change of its momentum. Experiments show that this result is still valid in relativistic mechanics, provided that we use the relativistic momentum given by Eq. 37.27. That is, the relativistically correct generalization of Newton's second law is

$$\vec{F} = \frac{d}{dt}\frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} \tag{37.29}$$

Because momentum is no longer directly proportional to velocity, the rate of change of momentum is no longer directly proportional to the acceleration. As a result, *constant force does not cause constant acceleration.* For example, when the net force and the velocity are both along the x-axis, Eq. 37.29 gives

$$F = \frac{m}{(1 - v^2/c^2)^{3/2}}a \qquad (\vec{F} \text{ and } \vec{v} \text{ along the same line}) \tag{37.30}$$

where a is the acceleration, also along the x-axis. Solving Eq. (37.30) for the acceleration a gives

$$a = \frac{F}{m}\left(1 - \frac{v^2}{c^2}\right)^{3/2}$$

We see that as a particle's speed increases, the acceleration caused by a given force continuously *decreases.* As the speed approaches c, the acceleration approaches zero, no matter how great a force is applied. Thus it is impossible to accelerate a particle with nonzero rest mass to a speed equal to or greater than c. We again see that the speed of light in vacuum represents an ultimate speed limit.

Equation (37.27) for relativistic momentum is sometimes interpreted to mean that a rapidly moving particle undergoes an increase in mass. If the mass at zero velocity (the rest mass) is denoted by m, then the "relativistic mass" m_{rel} is given by

$$m_{\text{rel}} = \frac{m}{\sqrt{1 - v^2/c^2}}$$

Indeed, when we consider the motion of a system of particles (such as rapidly moving ideal-gas molecules in a stationary container), the total rest mass of the system is the sum of the relativistic masses of the particles, not the sum of their rest masses.

However, if blindly applied, the concept of relativistic mass has its pitfalls. As Eq. (37.29) shows, the relativistic generalization of Newton's second law is *not* $\vec{F} = m_{\text{rel}}\vec{a}$, and we will show in Section 37.8 that the relativistic kinetic energy of a particle is *not* $K = \frac{1}{2}m_{\text{rel}}v^2$. The use of relativistic mass has its supporters and detractors, some quite strong in their opinions. We will mostly deal with individual particles, so we will sidestep the controversy and use Eq. (37.27) as the generalized definition of momentum with m as a constant for each particle, independent of its state of motion.

We will use the abbreviation

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

We used this abbreviation in Section 37.3 with v replaced by u, the relative speed of two coordinate systems. Here v is 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

$$\vec{p} = \gamma m\vec{v} \qquad \text{(relativistic momentum)} \tag{37.31}$$

$$F = \gamma^3 ma \qquad (\vec{F} \text{ and } \vec{v} \text{ along the same line}) \tag{37.32}$$

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

An electron (rest mass 9.11×10^{-31} kg, charge -1.60×10^{-19} C) is moving opposite to an electric field of magnitude $E = 5.00 \times 10^5$ N/C. All other forces are negligible in comparison to the electric-field force. (a) Find the magnitudes of momentum and of acceleration at the instants when $v = 0.010c$, $0.90c$, and $0.99c$. (b) Find the corresponding accelerations if a net force of the same magnitude is perpendicular to the velocity.

SOLUTION

IDENTIFY and SET UP: In addition to the expressions from this section for relativistic momentum and acceleration, we need the relationship between electric force and electric field from Chapter 21. In part (a) we use Eq. (37.31) to determine the magnitude of momentum; the force acts along the same line as the velocity, so we use Eq. (37.32) to determine the magnitude of acceleration. In part (b) the force is perpendicular to the velocity, so we use Eq. (37.33) rather than Eq. (37.32).

EXECUTE: (a) For $v = 0.010c$, $0.90c$, and $0.99c$ we have $\gamma = \sqrt{1 - v^2/c^2} = 1.00$, 2.29, and 7.09, respectively. The values of the momentum magnitude $p = \gamma mv$ are

$$p_1 = (1.00)(9.11 \times 10^{-31}\text{ kg})(0.010)(3.00 \times 10^8\text{m/s})$$
$$= 2.7 \times 10^{-24}\text{ kg}\cdot\text{m/s at } v_1 = 0.010c$$

$$p_2 = (2.29)(9.11 \times 10^{-31}\text{ kg})(0.90)(3.00 \times 10^8\text{ m/s})$$
$$= 5.6 \times 10^{-22}\text{ kg}\cdot\text{m/s at } v_2 = 0.90c$$

$$p_3 = (7.09)(9.11 \times 10^{-31}\text{ kg})(0.99)(3.00 \times 10^8\text{ m/s})$$
$$= 1.9 \times 10^{-21}\text{ kg}\cdot\text{m/s at } v_3 = 0.99c$$

From Eq. (21.4), the magnitude of the force on the electron is

$$F = |q|E = (1.60 \times 10^{-19}\text{ C})(5.00 \times 10^5\text{ N/C})$$
$$= 8.00 \times 10^{-14}\text{ N}$$

Continued

From Eq. (37.32), $a = F/\gamma^3 m$. For $v = 0.010c$ and $\gamma = 1.00$,

$$a_1 = \frac{8.00 \times 10^{-14}\ \text{N}}{(1.00)^3(9.11 \times 10^{-31}\ \text{kg})} = 8.8 \times 10^{16}\ \text{m/s}^2$$

The accelerations at the two higher speeds are smaller than the nonrelativistic value by factors of $\gamma^3 = 12.0$ and 356, respectively:

$$a_2 = 7.3 \times 10^{15}\ \text{m/s}^2 \qquad a_3 = 2.5 \times 10^{14}\ \text{m/s}^2$$

(b) From Eq. (37.33), $a = F/\gamma m$ if $\vec{F}$ and $\vec{v}$ are perpendicular. When $v = 0.010c$ and $\gamma = 1.00$,

$$a_1 = \frac{8.00 \times 10^{-14}\ \text{N}}{(1.00)(9.11 \times 10^{-31}\ \text{kg})} = 8.8 \times 10^{16}\ \text{m/s}^2$$

Now the accelerations at the two higher speeds are smaller by factors of $\gamma = 2.29$ and 7.09, respectively:

$$a_2 = 3.8 \times 10^{16}\ \text{m/s}^2 \qquad a_3 = 1.2 \times 10^{16}\ \text{m/s}^2$$

These accelerations are larger than the corresponding ones in part (a) by factors of γ^2.

EVALUATE: Our results in part (a) show that at higher speeds, the relativistic values of momentum differ more and more from the nonrelativistic values calculated from $p = mv$. The momentum at $0.99c$ is more than three times as great as at $0.90c$ because of the increase in the factor γ. Our results also show that the acceleration drops off very quickly as v approaches c.

Test Your Understanding of Section 37.7 According to relativistic mechanics, when you double the speed of a particle, the magnitude of its momentum increases by (i) a factor of 2; (ii) a factor greater than 2; (iii) a factor between 1 and 2 that depends on the mass of the particle.

37.8 Relativistic Work and Energy

When we developed the relationship between work and kinetic energy in Chapter 6, we used Newton's laws of motion. When we generalize 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 applicable relativistic version of Newton's second law. In moving a particle of rest mass m from point x_1 to point x_2,

$$W = \int_{x_1}^{x_2} F\,dx = \int_{x_1}^{x_2} \frac{ma\,dx}{(1 - v^2/c^2)^{3/2}} \tag{37.34}$$

To derive the generalized expression for kinetic energy K as a function of speed v, we would like to convert this to an integral on v. To do this, first remember that the kinetic energy of a particle equals the net work done on it in moving it from rest to the speed v: $K = W$. Thus we let the speeds be zero at point x_1 and v at point x_2. So as not to confuse the variable of integration with the final speed, we change v to v_x in Eq. 37.34. That is, v_x is the varying x-component of the velocity of the particle as the net force accelerates it from rest to a speed v. We also realize 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

$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2 \qquad \text{(relativistic kinetic energy)} \tag{37.36}$$

As v approaches c, the kinetic energy approaches infinity. If Eq. (37.36) is correct, it must also approach the Newtonian expression $K = \frac{1}{2}mv^2$ when v is much smaller than c (Fig. 37.21). To verify this, we expand the radical, using the binomial theorem in the form

$$(1 + x)^n = 1 + nx + n(n-1)x^2/2 + \cdots$$

In our case, $n = -\frac{1}{2}$ and $x = -v^2/c^2$, and we get

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots$$

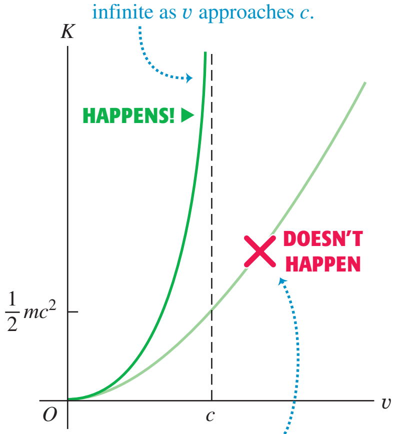
Combining this with $K = (\gamma - 1)mc^2$, we find

$$K = \left(1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots - 1\right)mc^2$$

$$= \frac{1}{2}mv^2 + \frac{3}{8}\frac{mv^4}{c^2} + \cdots \qquad (37.37)$$

When v is much smaller than c, all the terms in the series in Eq. (37.37) except the first are negligibly small, and we obtain the Newtonian expression $\frac{1}{2}mv^2$.

37.21 Graph of the kinetic energy of a particle of rest mass m as a function of speed v. Also shown is the Newtonian prediction, which gives correct results only at speeds much less than c.



Rest Energy and $E = mc^2$

Equation (37.36) for the kinetic energy of a moving particle includes a term $mc^2/\sqrt{1 - v^2/c^2}$ that depends on the motion and a second energy term mc^2 that is independent of the motion. It seems that the kinetic energy of a particle is the difference between some **total energy** E and an energy mc^2 that it has even when it is at rest. Thus we can rewrite Eq. (37.36) as

$$E = K + mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \gamma mc^2 \qquad \text{(total energy of a particle)} \qquad (37.38)$$

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 in fact direct 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 power through nuclear reactions. 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.

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.

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

$$E^2 = (mc^2)^2 + (pc)^2 \qquad \text{(total energy, rest energy, and momentum)} \tag{37.39}$$

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

$$E = pc \qquad \text{(zero rest mass)} \tag{37.40}$$

In fact, zero rest mass particles do exist. Such particles always travel at the speed of light in vacuum. One example is the *photon,* the quantum of electromagnetic radiation (to be discussed in Chapter 38). Photons are emitted and absorbed during changes of state of an atomic or nuclear system when the energy and momentum of the system change.

Example 37.10 Energetic electrons

(a) Find the rest energy of an electron $(m = 9.109 \times 10^{-31}\ \text{kg}$, $q = -e = -1.602 \times 10^{-19}\ \text{C})$ in joules and in electron volts. (b) Find the speed of an electron that has been accelerated by an electric field, from rest, through a potential increase of 20.0 kV or of 5.00 MV (typical of a high-voltage x-ray machine).

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of rest energy, relativistic kinetic energy, and (from Chapter 23) electric potential energy. We use $E = mc^2$ to find the rest energy and Eqs. (37.7) and (37.38) to find the speed that gives the stated total energy.

EXECUTE: (a) The rest energy is

$$\begin{aligned} mc^2 &= (9.109 \times 10^{-31}\ \text{kg})(2.998 \times 10^8\ \text{m/s})^2 \\ &= 8.187 \times 10^{-14}\ \text{J} \end{aligned}$$

From the definition of the electron volt in Section 23.2, 1 eV = 1.602×10^{-19} J. Using this, we find

$$\begin{aligned} mc^2 &= (8.187 \times 10^{-14}\ \text{J})\frac{1\ \text{eV}}{1.602 \times 10^{-19}\ \text{J}} \\ &= 5.11 \times 10^5\ \text{eV} = 0.511\ \text{MeV} \end{aligned}$$

(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:

$$\begin{aligned} E &= \gamma mc^2 = mc^2 + eV_{ba} \quad \text{or} \\ \gamma &= 1 + \frac{eV_{ba}}{mc^2} \end{aligned}$$

An electron accelerated through a potential increase of $V_{ba} =$ 20.0 kV gains 20.0 keV of energy, so for this electron

$$\gamma = 1 + \frac{20.0 \times 10^3\ \text{eV}}{0.511 \times 10^6\ \text{eV}} = 1.039$$

and

$$v = c\sqrt{1 - (1/1.039)^2} = 0.272c = 8.15 \times 10^7\ \text{m/s}$$

Repeating the calculation for $V_{ba} = 5.00$ MV, we find $eV_{ba}/mc^2 = 9.78$, $\gamma = 10.78$, and $v = 0.996c$.

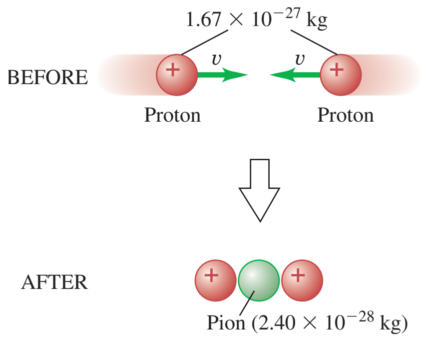
EVALUATE: With $V_{ba} = 20.0$ kV, the added kinetic energy of 20.0 keV is less than 4% of the rest energy of 0.511 MeV, and the final speed is about one-fourth the speed of light. With $V_{ba} =$ 5.00 MV, the added kinetic energy of 5.00 MeV is much greater than the rest energy and the speed is close to c.

CAUTION **Three electron energies** All electrons have *rest* energy 0.511 MeV. An electron accelerated from rest through a 5.00-MeV potential increase has *kinetic* energy 5.00 MeV (we call it a "5.00-MeV electron") and *total* energy 5.51 MeV. Be careful to distinguish these energies from one another.

Example 37.11 A relativistic collision

Two protons (each with mass $M_p = 1.67 \times 10^{-27}$ kg) are initially moving with equal speeds in opposite directions. They continue to exist after a head-on collision that also produces a neutral pion of mass $M_\pi = 2.40 \times 10^{-28}$ kg (Fig. 37.22). If all three particles are at rest after the collision, find the initial speed of the protons. Energy is conserved in the collision.

37.22 In this collision the kinetic energy of two protons is transformed into the rest energy of a new particle, a pion.



SOLUTION

IDENTIFY and SET UP: Relativistic total energy is conserved in the collision, so we can equate the (unknown) total energy of the two protons before the collision to the combined rest energies of the two protons and the pion after the collision. We then use Eq. (37.38) to find the speed of each proton.

EXECUTE: The total energy of each proton before the collision is γMc^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 $(\gamma - 1)Mc^2 = 0.072Mc^2 = (0.072)(938\ \text{MeV}) = 67.5$ MeV. You can verify that the rest energy $M_\pi c^2$ of the pion is twice this, or 135 MeV. All the kinetic energy "lost" in this completely inelastic collision is transformed into the rest energy of the pion.

Test Your Understanding of Section 37.8 A proton is accelerated from rest by a constant force that always points in the direction of the particle's motion. Compared to the amount of kinetic energy that the proton gains during the first meter of its travel, how much kinetic energy does the proton gain during one meter of travel while it is moving at 99% of the speed of light? (i) the same amount; (ii) a greater amount; (iii) a smaller amount.

37.9 Newtonian Mechanics and Relativity

The sweeping changes required by the principle of relativity go to the very roots of Newtonian mechanics, including the concepts of length and time, the equations of motion, and the conservation principles. Thus it may appear that we have destroyed the foundations on which Newtonian mechanics is built. In one sense this is true, yet the Newtonian formulation is still accurate whenever speeds are small in comparison with the speed of light in vacuum. In such cases, time dilation, length contraction, and the modifications of the laws of motion are so small that they are unobservable. In fact, every one of the principles of Newtonian mechanics survives as a special case of the more general relativistic formulation.

The laws of Newtonian mechanics are not *wrong;* they are *incomplete.* They are a limiting case of relativistic mechanics. They are *approximately* correct when all speeds are small in comparison to c, and they become exactly correct in the limit when all speeds approach zero. Thus relativity does not completely destroy the laws of Newtonian mechanics but *generalizes* them. This is a common pattern in the development of physical theory. Whenever a new theory is in partial conflict with an older, established theory, the new must yield the same predictions as the old in areas in which the old theory is supported by experimental evidence. Every new physical theory must pass this test, called the **correspondence principle.**

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 m/s^2. But an alternative interpretation, from her point of view, is that she doesn't fall to the floor because her gravitational interaction with the earth has suddenly been turned off. As long as she remains in the box and it remains in free fall, she cannot tell whether she is indeed in free fall or whether the gravitational interaction has vanished.

A similar problem occurs in a space station in orbit around the earth. Objects in the space station *seem* to be weightless, but without looking outside the station there is no way to determine whether gravity has been turned off or whether the station and all its contents are accelerating toward the center of the earth. Figure 37.23 makes a similar point for a spaceship that is not in free fall but may be accelerating relative to an inertial frame or be at rest on the earth's surface.

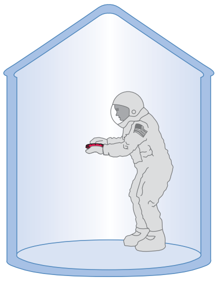
These considerations form the basis of Einstein's **general theory of relativity.** If we cannot distinguish experimentally between a uniform gravitational field at a particular location and a uniformly accelerated reference frame, then there cannot be any real distinction between the two. Pursuing this concept, we may try to represent *any* gravitational field in terms of special characteristics of the coordinate system. This turns out to require even more sweeping revisions of our space-time concepts than did the special theory of relativity. In the general theory of relativity the geometric properties of space are affected by the presence of matter (Fig. 37.24).

The general theory of relativity has passed several experimental tests, including three proposed by Einstein. One test has to do with understanding the rotation of the axes of the planet Mercury's elliptical orbit, called the *precession of the perihelion.* (The perihelion is the point of closest approach to the sun.) A second test concerns the apparent bending of light rays from distant stars when they pass near the sun. The third test is the *gravitational red shift,* the increase in wavelength of light proceeding outward from a massive source. Some details of the general theory are more difficult to test, but this theory has played a central role in investigations of the formation and evolution of stars, black holes, and studies of the evolution of the universe.

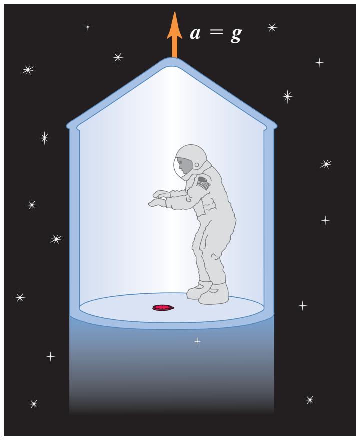
The general theory of relativity may seem to be an exotic bit of knowledge with little practical application. In fact, this theory plays an essential role in the

37.23 Without information from outside the spaceship, the astronaut cannot distinguish situation (b) from situation (c).

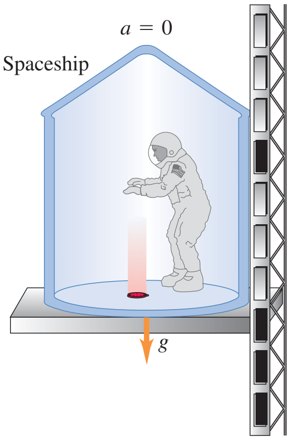
(a) An astronaut is about to drop her watch in a spaceship.



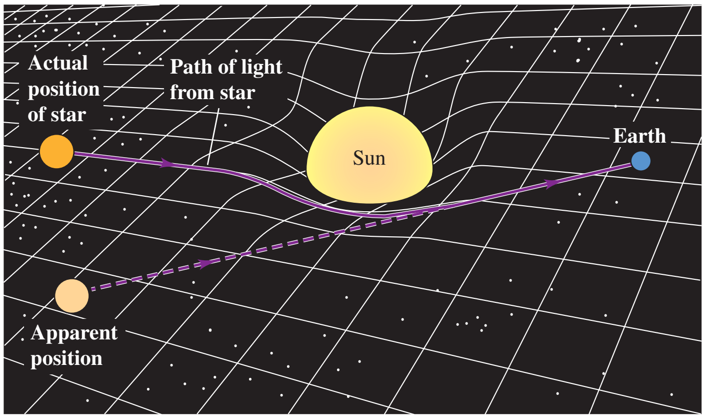
(b) In gravity-free space, the floor accelerates upward at $a = g$ and hits the watch.



(c) On the earth's surface, the watch accelerates downward at $a = g$ and hits the floor.



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°.

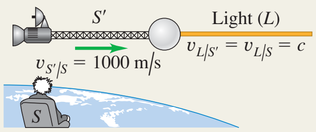


global positioning system (GPS), which makes it possible to determine your position on the earth's surface to within a few meters using a handheld receiver (Fig. 37.25). The heart of the GPS system is a collection of more than two dozen satellites in very precise orbits. Each satellite emits carefully timed radio signals, and a GPS receiver simultaneously detects the signals from several satellites. The receiver then calculates the time delay between when each signal was emitted and when it was received, and uses this information to calculate the receiver's position. To ensure the proper timing of the signals, it's necessary to include corrections due to the special theory of relativity (because the satellites are moving relative to the receiver on earth) as well as the general theory (because the satellites are higher in the earth's gravitational field than the receiver). The corrections due to relativity are small—less than one part in 10^9—but are crucial to the superb precision of the GPS system.

37.25 A GPS receiver uses radio signals from the orbiting GPS satellites to determine its position. To account for the effects of relativity, the receiver must be tuned to a slightly higher frequency (10.23 MHz) than the frequency emitted by the satellites (10.22999999543 MHz).

CHAPTER 37 SUMMARY

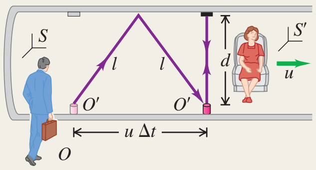
Invariance of physical laws, simultaneity: All of the fundamental laws of physics have the same form in all inertial frames of reference. The speed of light in vacuum is the same in all inertial frames and is independent of the motion of the source. Simultaneity is not an absolute concept; events that 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.)

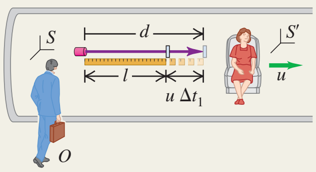
$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - u^2/c^2}} = \gamma\, \Delta t_0 \qquad (37.6), (37.8)$$

$$\gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \qquad (37.7)$$



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.)

$$l = l_0\sqrt{1 - u^2/c^2} = \frac{l_0}{\gamma} \qquad (37.16)$$



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.)

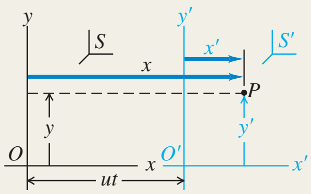
$$x' = \frac{x - ut}{\sqrt{1 - u^2/c^2}} = \gamma(x - ut)$$

$$y' = y \qquad z' = z \qquad (37.21)$$

$$t' = \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} = \gamma(t - ux/c^2)$$

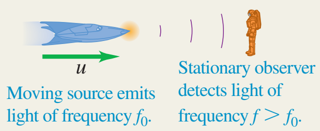
$$v_x' = \frac{v_x - u}{1 - uv_x/c^2} \qquad (37.22)$$

$$v_x = \frac{v_x' + u}{1 + uv_x'/c^2} \qquad (37.23)$$



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.)

$$f = \sqrt{\frac{c + u}{c - u}}\, f_0 \qquad (37.25)$$



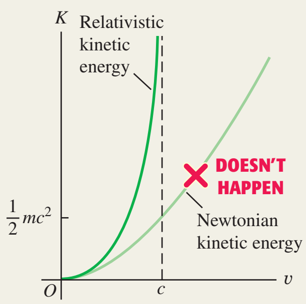
Relativistic momentum and energy: For a particle of rest mass m moving with velocity $\vec{v}$, the relativistic momentum $\vec{p}$ is given by Eq. (37.27) or (37.31) and the relativistic kinetic energy K is given by Eq. (37.36). The total energy E is the sum of the kinetic energy and the rest energy 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.)

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1 - v^2/c^2}} = \gamma m\vec{v} \qquad (37.27), (37.31)$$

$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 = (\gamma - 1)mc^2 \qquad (37.36)$$

$$E = K + mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \gamma mc^2 \qquad (37.38)$$

$$E^2 = (mc^2)^2 + (pc)^2 \qquad (37.39)$$



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

See MasteringPhysics® study area for a Video Tutor solution.

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. Take the x-axis to be the line of motion of the protons, and take the $+x$-direction to be to the right. In the frame in which the left-hand proton is at rest, the right-hand proton has velocity $-0.500c$. In the laboratory frame the two protons have velocities $-\alpha c$ and $+\alpha c$, where α (each proton's laboratory speed as a fraction of c) is our first target variable. Given this we can find the laboratory kinetic energy of each proton.

EXECUTE

3. Write a Lorentz velocity-transformation equation that relates the velocity of the right-hand proton in the laboratory frame to its velocity in the frame of the left-hand proton. Solve this equation for α. (*Hint:* Remember that α cannot be greater than 1. Why?)
4. Use your result from step 3 to find the laboratory kinetic energy of each proton.
5. Find the kinetic energy of the right-hand proton as measured in the frame of the left-hand proton.

EVALUATE

6. How much total kinetic energy must be imparted to the protons by a scientist in the laboratory? If the experiment were to be repeated with one proton stationary, what kinetic energy would have to be given to the other proton for the collision to be equivalent?

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **BIO**: Biosciences problems.

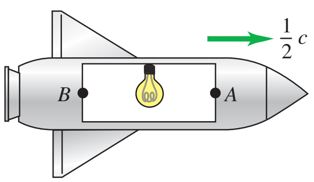
DISCUSSION QUESTIONS

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 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?

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?

EXERCISES

Section 37.2 Relativity of Simultaneity

37.1 • Suppose the two lightning bolts shown in Fig. 37.5a are simultaneous to an observer on the train. Show that they are *not* simultaneous to an observer on the ground. Which lightning strike does the ground observer measure to come first?

Section 37.3 Relativity of Time Intervals

37.2 • The positive muon (μ^+), an unstable particle, lives on average 2.20×10^{-6} s (measured in its own frame of reference) before decaying. (a) If such a particle is moving, with respect to the laboratory, with a speed of $0.900c$, what average lifetime is measured in the laboratory? (b) What average distance, measured in the laboratory, does the particle move before decaying?

37.3 • How fast must a rocket travel relative to the earth so that time in the rocket "slows down" to half its rate as measured by earth-based observers? Do present-day jet planes approach such speeds?

37.4 • A spaceship flies past Mars with a speed of $0.985c$ relative to the surface of the planet. When the spaceship is directly overhead, a signal light on the Martian surface blinks on and then off. An observer on Mars measures that the signal light was on for 75.0 μs. (a) Does the observer on Mars or the pilot on the spaceship measure the proper time? (b) What is the duration of the light pulse measured by the pilot of the spaceship?

37.5 • The negative pion (π^-) is an unstable particle with an average lifetime of 2.60×10^{-8} s (measured in the rest frame of the pion). (a) If the pion is made to travel at very high speed relative to a laboratory, its average lifetime is measured in the laboratory to be 4.20×10^{-7} s. Calculate the speed of the pion expressed as a fraction of c. (b) What distance, measured in the laboratory, does the pion travel during its average lifetime?

37.6 •• As you pilot your space utility vehicle at a constant speed toward the moon, a race pilot flies past you in her spaceracer at a constant speed of $0.800c$ relative to you. At the instant the spaceracer passes you, both of you start timers at zero. (a) At the instant when you measure that the spaceracer has traveled 1.20×10^8 m past you, what does the race pilot read on her timer? (b) When the race pilot reads the value calculated in part (a) on her timer, what does she measure to be your distance from her? (c) At the instant when the race pilot reads the value calculated in part (a) on her timer, what do you read on yours?

37.7 •• A spacecraft flies away from the earth with a speed of 4.80×10^6 m/s relative to the earth and then returns at the same speed. The spacecraft carries an atomic clock that has been carefully synchronized with an identical clock that remains at rest on earth. The spacecraft returns to its starting point 365 days (1 year) later, as measured by the clock that remained on earth. What is the difference in the elapsed times on the two clocks, measured in hours? Which clock, the one in the spacecraft or the one on earth, shows the shorter elapsed time?

37.8 • An alien spacecraft is flying overhead at a great distance as you stand in your backyard. You see its searchlight blink on for 0.190 s. The first officer on the spacecraft measures that the searchlight is on for 12.0 ms. (a) Which of these two measured times is the proper time? (b) What is the speed of the spacecraft relative to the earth expressed as a fraction of the speed of light c?

Section 37.4 Relativity of Length

37.9 • A spacecraft of the Trade Federation flies past the planet Coruscant at a speed of $0.600c$. A scientist on Coruscant measures the length of the moving spacecraft to be 74.0 m. The spacecraft later lands on Coruscant, and the same scientist measures the length of the now stationary spacecraft. What value does she get?

37.10 • A meter stick moves past you at great speed. Its motion relative to you is parallel to its long axis. If you measure the length of the moving meter stick to be 1.00 ft (1 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 • An unstable particle is created in the upper atmosphere from a cosmic ray and travels straight down toward the surface of the earth with a speed of $0.99540c$ relative to the earth. A scientist at rest on the earth's surface measures that the particle is created at an altitude of 45.0 km. (a) As measured by the scientist, how much time does it take the particle to travel the 45.0 km to the surface of the earth? (b) Use the length-contraction formula to calculate the distance from where the particle is created to the surface of the earth as measured in the particle's frame. (c) In the particle's frame, how much time does it take the particle to travel from where it is created to the surface of the earth? Calculate this time both by the time dilation formula and from the distance calculated in part (b). Do the two results agree?

37.13 • As measured by an observer on the earth, a spacecraft runway on earth has a length of 3600 m. (a) What is the length of the runway as measured by a pilot of a spacecraft flying past at a speed of 4.00×10^7 m/s relative to the earth? (b) An observer on earth measures the time interval from when the spacecraft is directly over one end of the runway until it is directly over the other end. What result does she get? (c) The pilot of the spacecraft measures the time it takes him to travel from one end of the runway to the other end. What value does he get?

37.14 • A rocket ship flies past the earth at 85.0% 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 the rocket 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 earth measure for his height? (b) If the earth-based person had measured 2.00 m, what would the doctor in the spaceship have measured for the astronaut's height? Is this a reasonable height? (c) Suppose the astronaut in part (a) gets up after the examination and stands with his body perpendicular to the direction of motion. What would the doctor in the rocket and the observer on earth measure for his height now?

Section 37.5 The Lorentz Transformations

37.15 • An observer in frame S' is moving to the right $(+x\text{-direction})$ 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 extraterrestrial spaceship is moving away from the earth after an unpleasant encounter with its inhabitants. As it departs, the spaceship fires a missile toward the earth. An observer on earth measures that the spaceship is moving away with a speed of $0.600c$. An observer in the spaceship measures that the missile is moving away from him at a speed of $0.800c$. As measured by an observer on earth, how fast is the missile approaching the earth?

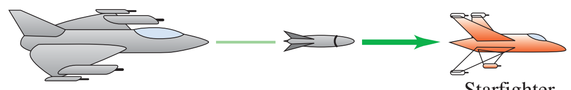
37.19 •• Two particles are created in a high-energy accelerator and move off in opposite directions. The speed of one particle, as measured in the laboratory, is $0.650c$, and the speed of each particle relative to the other is $0.950c$. What is the speed of the second particle, as measured in the laboratory?

37.20 •• Two particles in a high-energy accelerator experiment are approaching each other head-on, each with a speed of $0.9520c$ as measured in the laboratory. What is the magnitude of the velocity of one particle relative to the other?

37.21 •• Two particles in a high-energy accelerator experiment approach each other head-on with a relative speed of $0.890c$. Both particles travel at the same speed as measured in the laboratory. What is the speed of each particle, as measured in the laboratory?

37.22 •• An 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. E37.22). (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 **E37.22**



37.23 • An imperial spaceship, moving at high speed relative to the planet Arrakis, fires a rocket toward the planet with a speed of $0.920c$ relative to the spaceship. An observer on Arrakis measures that the rocket is approaching with a speed of $0.360c$. What is the speed of the spaceship relative to Arrakis? Is the spaceship moving toward or away from Arrakis?

Section 37.6 The Doppler Effect for Electromagnetic Waves

37.24 • Electromagnetic radiation from a star is observed with an earth-based telescope. The star is moving away from the earth with a speed of $0.600c$. If the radiation has a frequency of 8.64×10^{14} Hz in the rest frame of the star, what is the frequency measured by an observer on earth?

37.25 • **Tell It to the Judge.** (a) How fast must you be approaching a red traffic light $(\lambda = 675 \text{ nm})$ for it to appear yellow $(\lambda = 575 \text{ nm})$? 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.

37.26 • 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

37.27 • A proton has momentum with magnitude p_0 when its speed is $0.400c$. In terms of p_0, what is the magnitude of the proton's momentum when its speed is doubled to $0.800c$?

37.28 • **When Should You Use Relativity?** As you have seen, relativistic calculations usually involve the quantity γ. When γ is appreciably greater than 1, we must use relativistic formulas instead of Newtonian ones. For what speed v (in terms of c) is the value of γ (a) 1.0% greater than 1; (b) 10% greater than 1; (c) 100% greater than 1?

37.29 • (a) At what speed is the momentum of a particle twice as great as the result obtained from the nonrelativistic expression mv? Express your answer in terms of the speed of light. (b) A force is applied to a particle along its direction of motion. At what speed is the magnitude of force required to produce a given acceleration twice as great as the force required to produce the same acceleration when the particle is at rest? Express your answer in terms of the speed of light.

37.30 • As measured in an earth-based frame, a proton is moving in the $+x$-direction at a speed of 2.30×10^8 m/s. (a) What force (magnitude and direction) is required to produce an acceleration in the $-x$-direction that has magnitude $2.30 \times 10^8 \text{ m/s}^2$? (b) What magnitude of acceleration does the force calculated in part (a) give to a proton that is initially at rest?

37.31 • An electron is acted upon by a force of 5.00×10^{-15} N due to an electric field. Find the acceleration this force produces in each case: (a) The electron's speed is 1.00 km/s. (b) The electron's speed is 2.50×10^{8} m/s and the force is parallel to the velocity.

37.32 • **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.

Section 37.8 Relativistic Work and Energy

37.33 •• What is the speed of a particle whose kinetic energy is equal to (a) its rest energy and (b) five times its rest energy?

37.34 • 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.)

37.35 • A proton (rest mass 1.67×10^{-27} kg) has total energy that is 4.00 times its rest energy. What are (a) the kinetic energy of the proton; (b) the magnitude of the momentum of the proton; (c) the speed of the proton?

37.36 •• (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?

37.37 • **CP** (a) By what percentage does your rest mass increase when you climb 30 m to the top of a ten-story building? Are you aware of this increase? Explain. (b) By how many grams does the mass of a 12.0-g spring with force constant 200 N/cm change when you compress it by 6.0 cm? Does the mass increase or decrease? Would you notice the change in mass if you were holding the spring? Explain.

37.38 • A 60.0-kg person is standing at rest on level ground. How fast would she have to run to (a) double her total energy and (b) increase her total energy by a factor of 10?

37.39 • **An Antimatter Reactor.** When a particle meets its antiparticle, they annihilate each other and their mass is converted to light energy. The United States uses approximately 1.0×10^{20} J of energy per year. (a) If all this energy came from a futuristic antimatter reactor, how much mass of matter and antimatter fuel would be consumed yearly? (b) If this fuel had the density of iron ($7.86\ \text{g/cm}^3$) and were stacked in bricks to form a cubical pile, how high would it be? (Before you get your hopes up, antimatter reactors are a *long* way in the future—if they ever will be feasible.)

37.40 •• 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?

37.41 • A particle has rest mass 6.64×10^{-27} kg and momentum $2.10 \times 10^{-18}\ \text{kg} \cdot \text{m/s}$. (a) What is the total energy (kinetic plus rest energy) of the particle? (b) What is the kinetic energy of the particle? (c) What is the ratio of the kinetic energy to the rest energy of the particle?

37.42 •• A 0.100-μg speck of dust is accelerated from rest to a speed of $0.900c$ by a constant 1.00×10^{6} N force. (a) If the nonrelativistic mechanics is used, how far does the object travel to reach its final speed? (b) Using the correct relativistic treatment of Section 37.8, how far does the object travel to reach its final speed? (c) Which distance is greater? Why?

37.43 • 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.

37.44 • What is the kinetic energy of a proton moving at (a) $0.100c$; (b) $0.500c$; (c) $0.900c$? How much work must be done to (d) increase the proton's speed from $0.100c$ to $0.500c$ and (e) increase the proton's speed from $0.500c$ to $0.900c$? (f) How do the last two results compare to results obtained in the nonrelativistic limit?

37.45 • (a) Through what potential difference does an electron have to be accelerated, starting from rest, to achieve a speed of $0.980c$? (b) What is the kinetic energy of the electron at this speed? Express your answer in joules and in electron volts.

37.46 • **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).

37.47 • 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 lbs)? (b) At this rate, how long would it take the sun to use up all its mass?

PROBLEMS

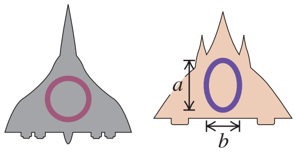
37.48 • Inside a spaceship flying past the earth at three-fourths the speed of light, a pendulum is swinging. (a) If each swing takes 1.50 s as measured by an astronaut performing an experiment inside the spaceship, how long will the swing take as measured by a person at mission control on earth who is watching the experiment? (b) If each swing takes 1.50 s as measured by a person at mission control on earth, how long will it take as measured by the astronaut in the spaceship?

37.49 • After being produced in a collision between elementary particles, a positive pion (π^+) must travel down a 1.90-km-long tube to reach an experimental area. A π^+ particle has an average lifetime (measured in its rest frame) of 2.60×10^{-8} s; the π^+ we are considering has this lifetime. (a) How fast must the π^+ travel if it is not to decay before it reaches the end of the tube? (Since u will be very close to c, write $u = (1 - \Delta)c$ and give your answer in terms of Δ rather than u.) (b) The π^+ has a rest energy of 139.6 MeV. What is the total energy of the π^+ at the speed calculated in part (a)?

37.50 •• 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?

37.51 ••• 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. P37.51). 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 **P37.51**



37.52 •• 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?

37.53 • A particle is said to be *extremely relativistic* when its kinetic energy is much greater than its rest energy. (a) What is the speed of a particle (expressed as a fraction of c) such that the total energy is ten times the rest energy? (b) What is the percentage difference between the left and right sides of Eq. (37.39) if $(mc^2)^2$ is neglected for a particle with the speed calculated in part (a)?

37.54 •• **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.)

37.55 • **The Large Hadron Collider (LHC).** Physicists and engineers from around the world have come together to build the largest accelerator in the world, the Large Hadron Collider (LHC) at the CERN Laboratory in Geneva, Switzerland. The machine will accelerate protons to kinetic energies of 7 TeV in an underground ring 27 km in circumference. (For the latest news and more information on the LHC, visit www.cern.ch.) (a) What speed v will protons reach in the LHC? (Since 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 protons in terms of their rest mass.

37.56 • **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?

37.57 • **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?

37.58 •• A photon with energy E is emitted by an atom with mass m, which recoils in the opposite direction. (a) Assuming that the motion of the atom can be treated nonrelativistically, compute the recoil speed of the atom. (b) From the result of part (a), show that the recoil speed is much less than c whenever E is much less than the rest energy mc^2 of the atom.

37.59 •• In an experiment, two protons are shot directly toward each other, each moving at half the speed of light relative to the laboratory. (a) What speed does one proton measure for the other proton? (b) What would be the answer to part (a) if we used only nonrelativistic Newtonian mechanics? (c) What is the kinetic energy of each proton as measured by (i) an observer at rest in the laboratory and (ii) an observer riding along with one of the protons? (d) What would be the answers to part (c) if we used only nonrelativistic Newtonian mechanics?

37.60 •• Two protons are moving away from each other. In the frame of each proton, the other proton has a speed of $0.600c$. What does an observer in the rest frame of the earth measure for the speed of each proton?

37.61 •• Frame S' has an x-component of velocity u relative to frame S, and at $t = t' = 0$ the two frames coincide (see Fig. 37.3). A light pulse with a spherical wave front is emitted at the origin of S' at time $t' = 0$. Its distance x' from the origin after a time t' is given by $x'^2 = c^2t'^2$. Use the Lorentz coordinate transformation to transform this equation to an equation in x and t, and show that the result is $x^2 = c^2t^2$; that is, the motion appears exactly the same in frame of reference S as it does in S'; the wave front is observed to be spherical in both frames.

37.62 • In certain radioactive beta decay processes, the beta particle (an electron) leaves the atomic nucleus with a speed of 99.95% the speed of light relative to the decaying nucleus. If this nucleus is moving at 75.00% the speed of light in the laboratory reference frame, find the speed of the emitted electron relative to the laboratory reference frame if the electron is emitted (a) in the same direction that the nucleus is moving and (b) in the opposite direction from the nucleus's velocity. (c) In each case in parts (a) and (b), find the kinetic energy of the electron as measured in (i) the laboratory frame and (ii) the reference frame of the decaying nucleus.

37.63 •• **CALC** A particle with mass m accelerated from rest by a constant force F will, according to Newtonian mechanics, continue to accelerate without bound; that is, as $t \to \infty$, $v \to \infty$. Show that according to relativistic mechanics, the particle's speed approaches c as $t \to \infty$. [*Note:* A useful integral is $\int (1 - x^2)^{-3/2} dx = x/\sqrt{1 - x^2}$.]

37.64 •• 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 second frame S' moving relative to S, the second event is observed to occur 2.35 s after the first. What is the difference between the positions of the two events as measured in S'?

37.65 ••• Two events observed in a frame of reference S have positions and times given by (x_1, t_1) and (x_2, t_2), respectively. (a) Frame S' moves along the x-axis just fast enough that the two events occur at the same position in S'. Show that in S', the time interval $\Delta t'$ between the two events is given by

$$\Delta t' = \sqrt{(\Delta t)^2 - \left(\frac{\Delta x}{c}\right)^2}$$

where $\Delta x = x_2 - x_1$ and $\Delta t = t_2 - t_1$. Hence show that if $\Delta x > c\,\Delta t$, there is *no* frame S' in which the two events occur at the same point. The interval $\Delta t'$ is sometimes called the *proper time interval* for the events. Is this term appropriate? (b) Show that if $\Delta x > c\,\Delta t$, there is a different frame of reference S' in which the two events occur *simultaneously.* Find the distance between the two events in S'; express your answer in terms of Δx, Δt, and c. This distance is sometimes called a *proper length.* Is this term appropriate? (c) Two events are observed in a frame of reference S' to occur simultaneously at points separated by a distance of 2.50 m. In a second frame S moving relative to S' along the line joining the two points in S', the two events appear to be separated by 5.00 m. What is the time interval between the events as measured in S? [*Hint:* Apply the result obtained in part (b).]

37.66 •• Albert in Wonderland. Einstein and Lorentz, being avid tennis players, play a fast-paced game on a court where they stand 20.0 m from each other. Being very skilled players, they play without a net. The tennis ball has mass 0.0580 kg. You can ignore gravity and assume that the ball travels parallel to the ground as it travels between the two players. Unless otherwise specified, all measurements are made by the two men. (a) Lorentz serves the ball at 80.0 m/s. What is the ball's kinetic energy? (b) Einstein slams a return at 1.80×10^8 m/s. What is the ball's kinetic energy? (c) During Einstein's return of the ball in part (a), a white rabbit runs beside the court in the direction from Einstein to Lorentz. The rabbit has a speed of 2.20×10^8 m/s relative to the two men. What is the speed of the rabbit relative to the ball? (d) What does the rabbit measure as the distance from Einstein to Lorentz? (e) How much time does it take for the rabbit to run 20.0 m, according to the players? (f) The white rabbit carries a pocket watch. He uses this watch to measure the time (as he sees it) for the distance from Einstein to Lorentz to pass by under him. What time does he measure?

37.67 • 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?

37.68 •• 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?)

37.69 • Space Travel? Travel to the stars requires hundreds or thousands of years, even at the speed of light. Some people have suggested that we can get around this difficulty by accelerating the rocket (and its astronauts) to very high speeds so that they will age less due to time dilation. The fly in this ointment is that it takes a great deal of energy to do this. Suppose you want to go to the immense red giant Betelgeuse, which is about 500 light-years away. (A light-year is the distance that light travels in a year.) You plan to travel at constant speed in a 1000-kg rocket ship (a little over a ton), which, in reality, is far too small for this purpose. In each case that follows, calculate the time for the trip, as measured by people on earth and by astronauts in the rocket ship, the energy needed in joules, and the energy needed as a percentage of U.S. yearly use (which is 1.0×10^{20} J). For comparison, arrange your results in a table showing v_{rocket}, t_{earth}, t_{rocket}, E (in J), and E (as % of U.S. use). The rocket ship's speed is (a) 0.50c; (b) 0.99c; (c) 0.9999c. On the basis of your results, does it seem likely that any government will invest in such high-speed space travel any time soon?

37.70 •• 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)?

37.71 • CP In a particle accelerator a proton moves with constant speed $0.750c$ in a circle of radius 628 m. What is the net force on the proton?

37.72 •• 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?

CHALLENGE PROBLEMS

37.73 ••• CALC Lorentz Transformation for Acceleration. Using a method analogous to the one in the text to find the Lorentz transformation formula for velocity, we can find the Lorentz transformation for *acceleration*. Let frame S' have a constant x-component of velocity u relative to frame S. An object moves relative to frame S along the x-axis with instantaneous velocity v_x and instantaneous acceleration a_x. (a) Show that its instantaneous acceleration in frame S' is

$$a'_x = a_x\left(1 - \frac{u^2}{c^2}\right)^{3/2}\left(1 - \frac{uv_x}{c^2}\right)^{-3}$$

[*Hint:* Express the acceleration in S' as $a'_x = dv'_x/dt'$. Then use Eq. (37.21) to express dt' in terms of dt and dx, and use Eq. (37.22) to express dv'_x in terms of u and dv_x. The velocity of the object in S is $v_x = dx/dt$.] (b) Show that the acceleration in frame S can be expressed as

$$a_x = a'_x\left(1 - \frac{u^2}{c^2}\right)^{3/2}\left(1 + \frac{uv'_x}{c^2}\right)^{-3}$$

where $v'_x = dx'/dt'$ is the velocity of the object in frame S'.

37.74 ••• CALC A Realistic Version of the Twin Paradox. A rocket ship leaves the earth on January 1, 2100. Stella, one of a pair of twins born in the year 2075, pilots the rocket (reference frame S'); the other twin, Terra, stays on the earth (reference frame S). The rocket ship has an acceleration of constant magnitude g in its own reference frame (this makes the pilot feel at home, since it simulates the earth's gravity). The path of the rocket ship is a straight line in the $+x$-direction in frame S. (a) Using the results of Challenge Problem 37.73, show that in Terra's earth frame S, the rocket's acceleration is

$$\frac{du}{dt} = g\left(1 - \frac{u^2}{c^2}\right)^{3/2}$$

where u is the rocket's instantaneous velocity in frame S. (b) Write the result of part (a) in the form $dt = f(u)\,du$, where $f(u)$ is a function of u, and integrate both sides. (*Hint:* Use the integral given in Problem 37.63.) Show that in Terra's frame, the time when Stella attains a velocity v_{1x} is

$$t_1 = \frac{v_{1x}}{g\sqrt{1 - v_{1x}^2/c^2}}$$

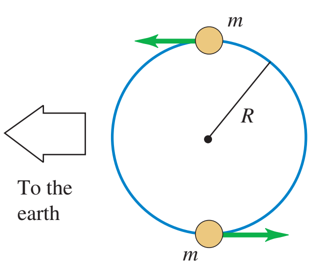
(c) Use the time dilation formula to relate dt and dt' (infinitesimal time intervals measured in frames S and S', respectively). Combine this result with the result of part (a) and integrate as in part (b) to show the following: When Stella attains a velocity v_{1x} relative to Terra, the time t_1' that has elapsed in frame S' is

$$t_1' = \frac{c}{g}\text{arctanh}\left(\frac{v_{1x}}{c}\right)$$

Here arctanh is the inverse hyperbolic tangent. (*Hint:* Use the integral given in Challenge Problem 5.124.) (d) Combine the results of parts (b) and (c) to find t_1 in terms of t_1', g, and c alone. (e) Stella accelerates in a straight-line path for five years (by her clock), slows down at the same rate for five years, turns around, accelerates for five years, slows down for five years, and lands back on the earth. According to Stella's clock, the date is January 1, 2120. What is the date according to Terra's clock?

37.75 ••• 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 P37.75 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 **P37.75**



37.76 ••• 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'.

37.77 ••• 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.)

Answers

Chapter Opening Question ?

No. While the speed of light c is the ultimate "speed limit" for any particle, there is *no* upper limit on a particle's kinetic energy (see Fig. 37.21). As the speed approaches c, a small increase in speed corresponds to a large increase in kinetic energy.

Test Your Understanding Questions

37.1 Answers: (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 is *not* centered on the current position of the spaceship; rather, it is centered on the point P where the spaceship was located at the instant that it emitted the light pulse. For example, suppose the spaceship is moving at speed $c/2$. When your watch shows that a time t has elapsed since the pulse of light was emitted, your measurements will show that the wave front is a sphere of radius ct centered on P and that the spaceship is a distance $ct/2$ from P.

37.2 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 Answers: (a) (ii), (b) (ii) The statement that moving clocks run slow refers to any clock that is moving relative to an observer. Maria and her stopwatch are moving relative to Samir, so Samir measures Maria's stopwatch to be running slow and to have ticked off fewer seconds than his own stopwatch. Samir and his stopwatch are moving relative to Maria, so she likewise measures Samir's stopwatch to be running slow. Each observer's measurement is correct for his or her own frame of reference. *Both* observers conclude that a moving stopwatch runs slow. This is consistent with the principle of relativity (see Section 37.1), which states that the laws of physics are the same in all inertial frames of reference.

37.4 Answer: (ii), (i) and (iii) (tie), (iv) You measure the rest length of the stationary meter stick and the contracted length of the moving spaceship to both be 1 meter. The rest length of the spaceship is greater than the contracted length that you measure, and so must be greater than 1 meter. A miniature observer on board the spaceship would measure a contracted length for the meter stick of less than 1 meter. Note that in your frame of reference the nose and tail of the spaceship can simultaneously align with the two ends of the meter stick, since in your frame of reference they have the same length of 1 meter. In the spaceship's frame these two alignments cannot happen simultaneously because the meter stick is shorter than the spaceship. Section 37.2 tells us that this shouldn't be a surprise; two events that are simultaneous to one observer may not be simultaneous to a second observer moving relative to the first one.

37.5 Answers: (a) P_1, (b) P_4 (a) The last of Eqs. (37.21) tells us the times of the two events in S': $t'_1 = \gamma(t_1 - ux_1/c^2)$ and $t'_2 = \gamma(t_2 - ux_2/c^2)$. In frame S the two events occur at the same x-coordinate, so $x_1 = x_2$, and event P_1 occurs before event P_2, so $t_1 < t_2$. Hence you can see that $t'_1 < t'_2$ and event P_1 happens before P_2 in frame S', too. This says that if event P_1 happens before P_2 in a frame of reference S where the two events occur at the same position, then P_1 happens before P_2 in any other frame moving relative to S. (b) In frame S the two events occur at different x-coordinates such that $x_3 < x_4$, and events P_3 and P_4 occur at the same time, so $t_3 = t_4$. Hence you can see that $t'_3 = \gamma(t_3 - ux_3/c^2)$ is greater than $t'_4 = \gamma(t_4 - ux_4/c^2)$, so event P_4 happens before P_3 in frame S'. This says that even though the two events are simultaneous in frame S, they need not be simultaneous in a frame moving relative to S.

37.7 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 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. (It's true that as the proton approaches the ultimate speed limit of c, the increase in the proton's *speed* is less and less with each subsequent meter of travel. That's not what the question is asking, however.)

Bridging Problem

Answers: (a) $0.268c$ **(b)** 35.6 MeV **(c)** 145 MeV

38 PHOTONS: LIGHT WAVES BEHAVING AS PARTICLES

? This plastic surgeon is using two light sources: a headlamp that emits a beam of visible light and a handheld laser that emits infrared light. The light from both sources is emitted in the form of packets of energy called photons. For which source are the photons more energetic: the headlamp or the laser?

LEARNING GOALS

By studying this chapter, you will learn:

- How experiments involving the photoelectric effect and x rays pointed the way to a radical reinterpretation of the nature of light.
- How Einstein's photon picture of light explains the photoelectric effect.
- How experiments with x rays and gamma rays helped confirm the photon picture of light.
- How the wave and particle pictures of light complement each other.
- How the Heisenberg uncertainty principle imposes fundamental limits on what can be measured.

In Chapter 32 we saw how Maxwell, Hertz, and others established firmly that light is an electromagnetic wave. Interference, diffraction, and polarization, discussed in Chapters 35 and 36, further demonstrate this *wave nature* of light.

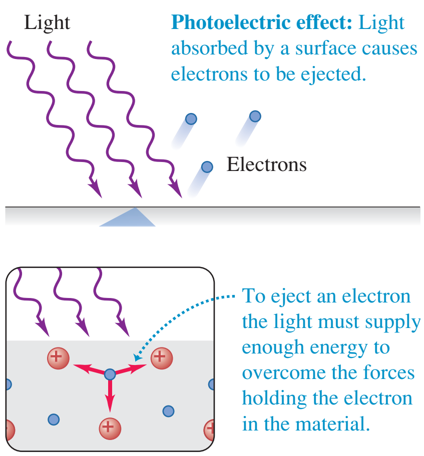
When we look more closely at the emission, absorption, and scattering of electromagnetic radiation, however, we discover a completely different aspect of light. We find that the energy of an electromagnetic wave is *quantized;* it is emitted and absorbed in particle-like packages of definite energy, called *photons.* The energy of a single photon is proportional to the frequency of the radiation.

We'll find that light and other electromagnetic radiation exhibits *wave–particle duality:* Light acts sometimes like waves and sometimes like particles. Interference and diffraction demonstrate wave behavior, while emission and absorption of photons demonstrate the particle behavior. This radical reinterpretation of light will lead us in the next chapter to no less radical changes in our views of the nature of matter.

38.1 Light Absorbed as Photons: The Photoelectric Effect

A phenomenon that gives insight into the nature of light is the **photoelectric effect,** in which a material emits electrons from its surface when illuminated (Fig. 38.1). To escape from the surface, an electron must absorb enough energy from the incident light to overcome the attraction of positive ions in the material. These attractions constitute a potential-energy barrier; the light supplies the "kick" that enables the electron to escape.

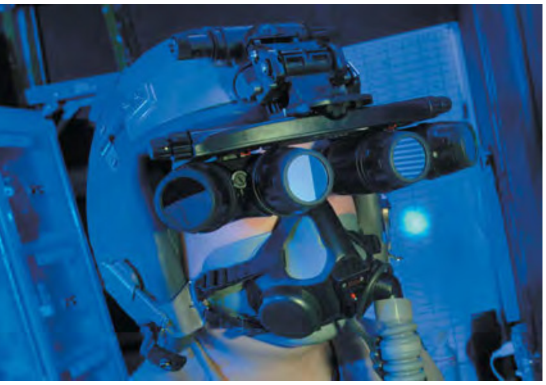
38.1 The photoelectric effect.



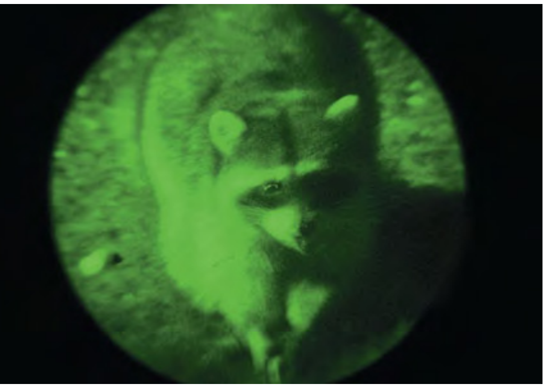
The photoelectric effect has a number of applications. Digital cameras and night-vision scopes use it to convert light energy into an electric signal that is

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)



MasteringPHYSICS

PhET: Photoelectric Effect
ActivPhysics 17.3: Photoelectric Effect

reconstructed into an image (Fig. 38.2). On the moon, sunlight striking the surface 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 do we expect the photocurrent to 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* expect:

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*.

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.

These results directly contradict Maxwell's description of light as an electromagnetic wave. A solution to this dilemma was provided by Albert Einstein in 1905. His proposal involved nothing less than a new picture of the nature of light.

Einstein's Photon Explanation

Einstein made the radical postulate that a beam of light consists of small packages of energy called **photons** or *quanta*. This postulate was an extension of an idea developed five years earlier by Max Planck to explain the properties of blackbody radiation, which we discussed in Section 17.7. (We'll explore Planck's ideas in Section 39.5.) In Einstein's picture, the energy E of an individual photon is equal to a constant h times the photon frequency f. From the relationship $f = c/\lambda$ for electromagnetic waves in vacuum, we have

$$E = hf = \frac{hc}{\lambda} \qquad \text{(energy of a photon)} \tag{38.2}$$

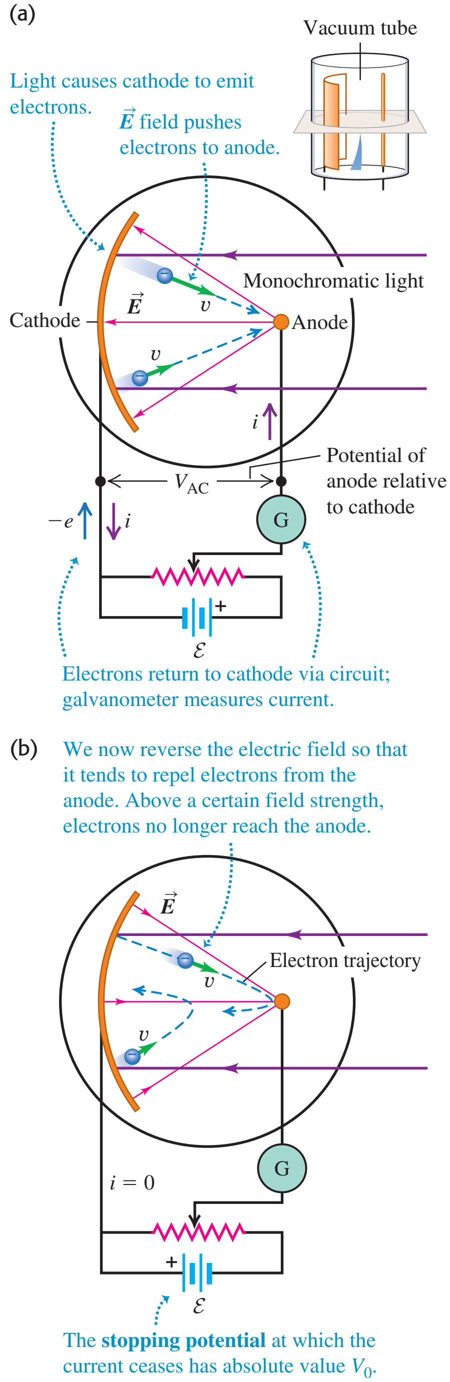
where h is a universal constant called **Planck's constant.** The numerical value of this constant, to the accuracy known at present, is

$$h = 6.62606896(33) \times 10^{-34}\ \text{J} \cdot \text{s}$$

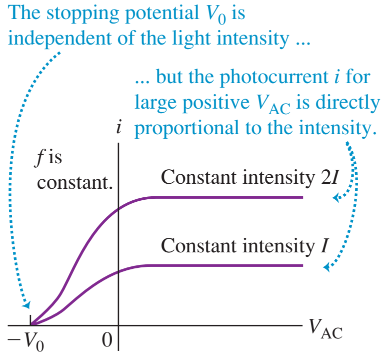
CAUTION **Photons are not "particles" in the usual sense** It's common to envision photons as miniature billiard balls or pellets. While that's a convenient mental picture, it's not very accurate. For one thing, billiard balls and bullets have a rest mass and travel slower than the speed of light c, while photons travel at the speed of light and have *zero* rest mass. For another thing, photons have wave aspects (frequency and wavelength) that are easy to observe. The fact is that 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 the dual personality of photons in more detail in Section 38.4.

In Einstein's picture, an individual photon arriving at the surface in Fig. 38.1a or 38.2 is absorbed by a single electron. This energy transfer is an all-or-nothing process, in contrast to the continuous transfer of energy in the wave theory of

38.3 An experiment testing whether the photoelectric effect is consistent with the wave model of light.



38.4 Photocurrent i for a constant light frequency f as a function of the potential V_{AC} of the anode with respect to the cathode.



38.5 Stopping potential as a function of frequency for a particular cathode material.

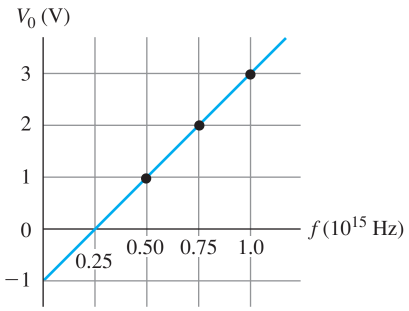
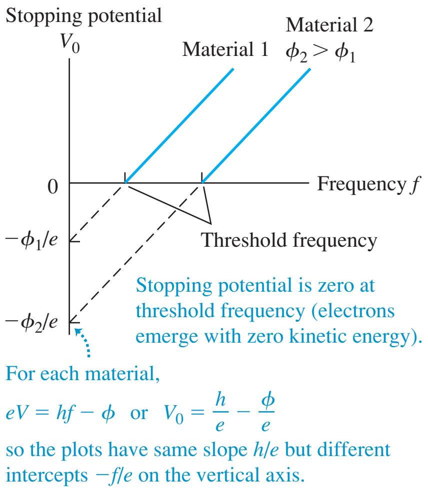


Table 38.1 Work Functions of Several Elements

Element	Work Function (eV)
Aluminum	4.3
Carbon	5.0
Copper	4.7
Gold	5.1
Nickel	5.1
Silicon	4.8
Silver	4.3
Sodium	2.7

38.6 Stopping potential as a function of frequency for two cathode materials having different work functions ϕ.



light; the electron gets all of the photon's energy or none at all. The electron can escape from the surface only if the energy it acquires is greater than the work function ϕ. Thus photoelectrons will be ejected only if $hf > \phi$, or $f > \phi/h$. Einstein's postulate therefore explains why the photoelectric effect occurs only for frequencies greater than a minimum threshold frequency. This postulate is also consistent with the observation that greater intensity causes a greater photocurrent (Fig. 38.4). Greater intensity at a particular frequency means a greater number of photons per second absorbed, and thus a greater number of electrons emitted per second and a greater photocurrent.

Einstein's postulate also explains why there is no delay between illumination and the emission of photoelectrons. As soon as photons of sufficient energy strike the surface, electrons can absorb them and be ejected.

Finally, Einstein's postulate explains why the stopping potential for a given surface depends only on the light frequency. Recall that ϕ is the *minimum* energy needed to remove an electron from the surface. Einstein applied conservation of energy to find that the *maximum* kinetic energy $K_{max} = \frac{1}{2}mv_{max}{}^2$ for an emitted electron is the energy hf gained from a photon minus the work function ϕ:

$$K_{max} = \tfrac{1}{2}mv_{max}{}^2 = hf - \phi \tag{38.3}$$

Substituting $K_{max} = eV_0$ from Eq. (38.1), we find

$$eV_0 = hf - \phi \qquad \text{(photoelectric effect)} \tag{38.4}$$

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 explains a number of other phenomena in which light is absorbed. One example is a *suntan*, which is caused when the energy in sunlight triggers a chemical reaction in skin cells that leads to increased production of the pigment melanin. This reaction can occur only if a specific molecule in the cell absorbs a certain minimum amount of energy. A short-wavelength ultraviolet photon has enough energy to trigger the reaction, but a longer-wavelength visible-light photon does not. Hence ultraviolet light causes tanning, while visible light cannot.

Photon Momentum

Einstein's photon concept applies to *all* regions of the electromagnetic spectrum, including radio waves, x rays, and so on. A photon of any electromagnetic radiation with 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 also have momentum, even if it has no rest mass. Photons have zero rest mass. As we saw in Eq. (37.40), a particle with zero rest mass and energy E has momentum with magnitude p given by $E = pc$. Thus the wavelength λ of a photon and the magnitude of its momentum p are related simply by

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda} \qquad \text{(momentum of a photon)} \qquad (38.5)$$

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: $1\text{ eV} = 1.602 \times 10^{-19}$ 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 $\lambda = 600$ nm and $f = 5 \times 10^{14}$ Hz has an energy E of about 2 eV, or about 3×10^{-19} J.

Example 38.1 Laser-pointer photons

A laser pointer with a power output of 5.00 mW emits red light ($\lambda = 650$ nm). (a) What is the magnitude of the momentum of each photon? (b) How many photons does the laser pointer emit each second?

SOLUTION

IDENTIFY and SET UP: This problem involves the ideas of (a) photon momentum and (b) photon energy. In part (a) we'll use Eq. (38.5) and the given wavelength to find the magnitude of each photon's momentum. In part (b), Eq. (38.2) gives the energy per photon, and the power output tells us the energy emitted per second. We can combine these quantities to calculate the number of photons emitted per second.

EXECUTE: (a) We have $\lambda = 650\text{ nm} = 6.50 \times 10^{-7}$ m, so from Eq. (38.5) the photon momentum is

$$p = \frac{h}{\lambda} = \frac{6.626 \times 10^{-34}\text{ J}\cdot\text{s}}{6.50 \times 10^{-7}\text{ m}}$$
$$= 1.02 \times 10^{-27}\text{ kg}\cdot\text{m/s}$$

(Recall that $1\text{ J} = 1\text{ kg}\cdot\text{m}^2/\text{s}^2$.)

(b) From Eq. (38.2), the energy of a single photon is

$$E = pc = (1.02 \times 10^{-27}\text{ kg}\cdot\text{m/s})(3.00 \times 10^8\text{ m/s})$$
$$= 3.06 \times 10^{-19}\text{ J} = 1.91\text{ eV}$$

The laser pointer emits energy at the rate of 5.00×10^{-3} J/s, so it emits photons at the rate of

$$\frac{5.00 \times 10^{-3}\text{ J/s}}{3.06 \times 10^{-19}\text{ J/photon}} = 1.63 \times 10^{16}\text{ photons/s}$$

EVALUATE: The result in part (a) is very small; a typical oxygen molecule in room-temperature air has 2500 times more momentum. As a check on part (b), we can calculate the photon energy using Eq. (38.2):

$$E = hf = \frac{hc}{\lambda} = \frac{(6.626 \times 10^{-34}\text{ J}\cdot\text{s})(3.00 \times 10^8\text{ m/s})}{6.50 \times 10^{-7}\text{ m}}$$
$$= 3.06 \times 10^{-19}\text{ J} = 1.91\text{ eV}$$

Our result in part (b) shows that a huge number of photons leave the laser pointer each second, each of which has an infinitesimal amount of energy. Hence the discreteness of the photons isn't noticed, and the radiated energy appears to be a continuous flow.

Example 38.2 A photoelectric-effect experiment

While conducting a photoelectric-effect experiment with light of a certain frequency, you find that a reverse potential difference of 1.25 V is required to reduce the current to zero. Find (a) the maximum kinetic energy and (b) the maximum speed of the emitted photoelectrons.

SOLUTION

IDENTIFY and SET UP: The value of 1.25 V is the stopping potential V_0 for this experiment. We'll use this in Eq. (38.1) to find the maximum photoelectron kinetic energy K_{max}, and from this we'll find the maximum photoelectron speed.

EXECUTE: (a) From Eq. (38.1),

$$K_{max} = eV_0 = (1.60 \times 10^{-19}\ \mathrm{C})(1.25\ \mathrm{V}) = 2.00 \times 10^{-19}\ \mathrm{J}$$

(Recall that 1 V = 1 J/C.) In terms of electron volts,

$$K_{max} = eV_0 = e(1.25\ \mathrm{V}) = 1.25\ \mathrm{eV}$$

since the electron volt (eV) is the magnitude of the electron charge e times one volt (1 V).

(b) From $K_{max} = \frac{1}{2}mv_{max}{}^2$ we get

$$v_{max} = \sqrt{\frac{2K_{max}}{m}} = \sqrt{\frac{2(2.00 \times 10^{-19}\ \mathrm{J})}{9.11 \times 10^{-31}\ \mathrm{kg}}}$$
$$= 6.63 \times 10^5\ \mathrm{m/s}$$

EVALUATE The value of v_{max} is about 0.2% of the speed of light, so we are justified in using the nonrelativistic expression for kinetic energy. (An equivalent justification is that the electron's 1.25-eV kinetic energy is much less than its rest energy $mc^2 = 0.511\ \mathrm{MeV} = 5.11 \times 10^5\ \mathrm{eV}$.)

Example 38.3 Determining ϕ and h experimentally

For a particular cathode material in a photoelectric-effect experiment, you measure stopping potentials $V_0 = 1.0$ V for light of wavelength $\lambda = 600$ nm, 2.0 V for 400 nm, and 3.0 V for 300 nm. Determine the work function ϕ for this material and the implied value of Planck's constant h.

SOLUTION

IDENTIFY and SET UP: This example uses the relationship among stopping potential V_0, frequency f, and work function ϕ in the photoelectric effect. According to Eq. (38.4), a graph of V_0 versus f should be a straight line as in Fig. 38.5 or 38.6. Such a graph is completely determined by its slope and the value at which it intercepts the vertical axis; we will use these to determine the values of the target variables ϕ and h.

EXECUTE: We rewrite Eq. (38.4) as

$$V_0 = \frac{h}{e}f - \frac{\phi}{e}$$

In this form we see that the slope of the line is h/e and the vertical-axis intercept (corresponding to $f = 0$) is $-\phi/e$. The frequencies, obtained from $f = c/\lambda$ and $c = 3.00 \times 10^8$ m/s, are 0.50×10^{15} Hz, 0.75×10^{15} Hz, and 1.0×10^{15} Hz, respectively. From a graph of these data (see Fig. 38.6), we find

$$-\frac{\phi}{e} = \text{vertical intercept} = -1.0\ \mathrm{V}$$
$$\phi = 1.0\ \mathrm{eV} = 1.6 \times 10^{-19}\ \mathrm{J}$$

and

$$\text{Slope} = \frac{\Delta V_0}{\Delta f} - \frac{3.0\ \mathrm{V} - (-1.0\ \mathrm{V})}{1.00 \times 10^{15}\ \mathrm{s}^{-1} - 0} = 4.0 \times 10^{-15}\ \mathrm{J \cdot s/C}$$
$$h = \text{slope} \times e = (4.0 \times 10^{-15}\ \mathrm{J \cdot s/C})(1.60 \times 10^{-19}\ \mathrm{C})$$
$$= 6.4 \times 10^{-34}\ \mathrm{J \cdot s}$$

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.

Application Sterilizing with High-Energy Photons
One technique for killing harmful microorganisms is to illuminate them with ultraviolet light with a wavelength shorter than 254 nm. If a photon of such short wavelength strikes a DNA molecule within a microorganism, the energy of the photon is great enough to break the bonds within the molecule. This renders the microorganism unable to grow or reproduce. Such ultraviolet germicidal irradiation is used for medical sanitation, to keep laboratories sterile (as shown here), and to treat both drinking water and wastewater.

Test Your Understanding of Section 38.1 Silicon films become better electrical conductors when illuminated by photons with energies of 1.14 eV or greater, an effect called *photoconductivity*. Which of the following wavelengths of electromagnetic radiation can cause photoconductivity in silicon films? (i) ultraviolet light with $\lambda = 300$ nm; (ii) red light with $\lambda = 600$ nm; (iii) infrared light with $\lambda = 1200$ nm.

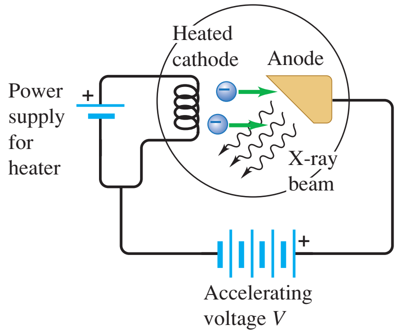
38.2 Light Emitted as Photons: X-Ray Production

The photoelectric effect provides convincing evidence that light is *absorbed* in the form of photons. For physicists to accept Einstein's radical photon concept, however, it was also necessary to show that light is *emitted* as photons. An experiment that demonstrates this convincingly is the inverse of the photoelectric effect: Instead of releasing electrons from a surface by shining electromagnetic radiation on it, we cause a surface to emit radiation—specifically, *x rays*—by bombarding it with fast-moving electrons.

X-Ray Photons

X rays were first produced in 1895 by the German physicist Wilhelm Röntgen, using an apparatus similar in principle to the setup shown in Fig. 38.7. Electrons are released from the cathode by *thermionic emission,* in which the escape energy is supplied by heating the cathode to a very high temperature. (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.

38.7 An apparatus used to produce x rays, similar to Röntgen's 1895 apparatus.



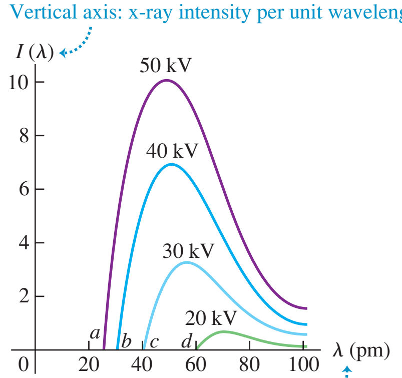
The anode produces x rays in part simply by slowing the electrons abruptly. (Recall from Section 32.1 that accelerated charges emit electromagnetic waves.) This process is called *bremsstrahlung* (German for "braking radiation"). Because the electrons undergo accelerations of very great magnitude, they 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 using the same cathode and anode with four different accelerating voltages. We see that *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 potential difference V_{AC}, the higher the maximum frequency and the shorter the minimum wavelength.

38.8 The continuous spectrum of x rays produced when a tungsten target is struck by electrons accelerated through a voltage V_{AC}. The curves represent different values of V_{AC}; points a, b, c, and d show the minimum wavelength for each voltage.



The wave model of electromagnetic radiation cannot explain these experimental results. But we can readily understand them 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,

$$eV_{AC} = hf_{max} = \frac{hc}{\lambda_{min}} \qquad \text{(bremsstrahlung)} \qquad (38.6)$$

(In this equation we neglect the work function of the target anode and the initial kinetic energy of the electrons "boiled off" from the cathode. These energies are very small compared to the kinetic energy eV_{AC} gained due to the potential difference.) If only a portion of an electron's kinetic energy goes into producing a photon, the photon energy will be less than eV_{AC} and the wavelength will be greater than λ_{min}. As further support for the photon model, the measured values for λ_{min} for different values of eV_{AC} (see Fig. 38.8) agree with Eq. (38.6). Note that according to Eq. (38.6), the maximum frequency and minimum wavelength in the bremsstrahlung process do not depend on the target material; this also agrees with experiment. So we can conclude that the photon picture of electromagnetic radiation is valid for the *emission* as well as the absorption of radiation.

The apparatus shown in Fig. 38.7 can also produce x rays by a second process in which electrons transfer their kinetic energy partly or completely to individual atoms within the target. It turns out that this process not only is consistent with the photon model of electromagnetic radiation, but also provides insight into the structure of atoms. We'll return to this process in Section 41.5.

Example 38.4 Producing x rays

Electrons in an x-ray tube accelerate through a potential difference of 10.0 kV before striking a target. If an electron produces one photon on impact with the target, what is the minimum wavelength of the resulting x rays? Find the answer by expressing energies in both SI units and electron volts.

SOLUTION

IDENTIFY and SET UP: To produce an x-ray photon with minimum wavelength and hence maximum energy, all of the electron's kinetic energy must go into producing a single x-ray photon. We'll use Eq. (38.6) to determine the wavelength.

EXECUTE: From Eq. (38.6), using SI units we have

$$\lambda_{min} = \frac{hc}{eV_{AC}} = \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})(3.00 \times 10^{8}\ \mathrm{m/s})}{(1.602 \times 10^{-19}\ \mathrm{C})(10.0 \times 10^{3}\ \mathrm{V})}$$
$$= 1.24 \times 10^{-10}\ \mathrm{m} = 0.124\ \mathrm{nm}$$

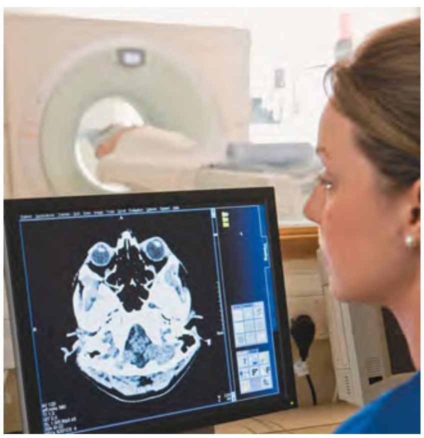
Using electron volts, we have

$$\lambda_{min} = \frac{hc}{eV_{AC}} = \frac{(4.136 \times 10^{-15}\ \mathrm{eV \cdot s})(3.00 \times 10^{8}\ \mathrm{m/s})}{e(10.0 \times 10^{3}\ \mathrm{V})}$$
$$= 1.24 \times 10^{-10}\ \mathrm{m} = 0.124\ \mathrm{nm}$$

In the second calculation, the "e" for the magnitude of the electron charge cancels the "e" in the unit "eV," because the electron volt (eV) is the magnitude of the electron charge e times one volt (1 V).

EVALUATE: To check our result, recall from Example 38.1 that a 1.91-eV photon has a wavelength of 650 nm. Here the electron energy, and therefore the x-ray photon energy, is 10.0×10^{3} eV = 10.0 keV, about 5000 times greater than in Example 38.1, and the wavelength is about $\frac{1}{5000}$ as great as in Example 38.1. This makes sense, since wavelength and photon energy are inversely proportional.

38.9 This radiologist is operating a CT scanner (seen through the window) from a separate room to avoid repeated exposure to x rays.



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) or a piece of photographic film. 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.

X rays cause damage to living tissues. As x-ray photons are absorbed in tissues, their energy breaks molecular bonds and creates highly reactive free radicals (such as neutral H and OH), which in turn can disturb the molecular structure of proteins and especially genetic material. Young and rapidly growing cells are particularly susceptible, which is why x rays are useful for selective destruction of cancer cells. Conversely, however, a cell may be damaged by radiation but survive, continue dividing, and produce generations of defective cells; thus x rays can *cause* cancer.

Even when the organism itself shows no apparent damage, excessive exposure to x rays can cause changes in the organism's reproductive system that will affect its offspring. A careful assessment of the balance between risks and benefits of radiation exposure is essential in each individual case.

Test Your Understanding of Section 38.2 In the apparatus shown in Fig. 38.7, suppose you increase the number of electrons that are emitted from the cathode per second while keeping the potential difference V_{AC} the same. How will this affect the intensity I and minimum wavelength λ_{min} of the emitted x rays? (i) I and λ_{min} will both increase; (ii) I will increase but λ_{min} will be unchanged; (iii) I will increase but λ_{min} will decrease; (iv) I will remain the same but λ_{min} will decrease; (v) none of these.

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, while 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 can pass relatively easily through soft tissue.

38.3 Light Scattered as Photons: Compton Scattering and Pair Production

The final aspect of light that we must test against Einstein's photon model is its behavior after the light is produced and before it is eventually absorbed. We can do this by considering the *scattering* of light. As we discussed in Section 33.6, scattering is what happens when light bounces off particles such as molecules in the air.

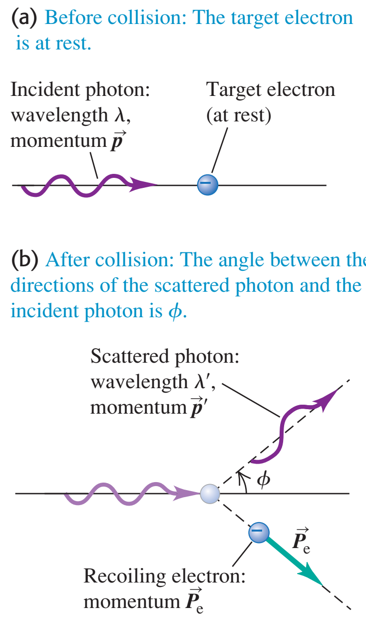
Compton Scattering

Let's see what Maxwell's wave model and Einstein's photon model predict for how light behaves when it undergoes scattering by a single electron, such as an individual electron within an atom.

Wave-Model Prediction: In the wave description, scattering would be a process of absorption and re-radiation. Part of the energy of the light wave would be absorbed by the electron, which would oscillate in response to the oscillating electric field of the wave. The oscillating electron would act like a miniature antenna (see Section 32.1), re-radiating its acquired energy as *scattered* waves in a variety of directions. The frequency at which the electron oscillates would be the same as the frequency of the incident light, and the re-radiated light would have the same frequency as the oscillations of the electron. So, *in the wave model, the scattered light and incident light have the same frequency and same wavelength.*

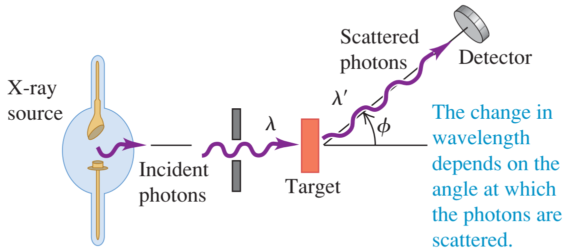
Photon-Model Prediction: In the photon model we imagine the scattering process as a collision of two *particles,* the incident photon and an electron that is initially at rest (Fig. 38.10a). The incident photon would give up part of its energy and momentum to the electron, which recoils as a result of this impact. The scattered photon that remains can fly off at a variety of angles ϕ with respect to the incident direction, but it has less energy and less momentum than the incident photon (Fig. 38.10b). The energy and momentum of a photon are given by $E = hf = hc/\lambda$ (Eq. 38.2) and $p = hf/c = h/\lambda$ (Eq. 38.5). Therefore, *in the photon model, the scattered light has a lower frequency f and longer wavelength λ than the incident light.*

38.10 The photon model of light scattering by an electron.



MasteringPHYSICS
ActivPhysics 17.4: Compton Scattering

38.11 A Compton-effect experiment.



The definitive experiment that tested these predictions of the wave and photon models was carried out in 1922 by the American physicist Arthur H. Compton. In his experiment Compton aimed a beam of x rays at a solid target and measured the wavelength of the radiation scattered from the target (Fig. 38.11). He 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, if the scattered radiation emerges at an angle ϕ with respect to the incident direction, as shown in Fig. 38.11, and if λ and λ' are the wavelengths of the incident and scattered radiation, respectively, Compton found that

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\phi) \qquad \text{(Compton scattering)} \qquad (38.7)$$

where m is the electron rest mass. 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 $\vec{p}$, with magnitude p and energy pc. The scattered photon has momentum $\vec{p}\,'$, with magnitude p' and energy $p'c$. The electron is initially at rest, so its initial momentum is zero and its initial energy is its rest energy mc^2. The final electron momentum $\vec{P}_e$ has magnitude P_e, and the final electron energy is given by $E_e{}^2 = (mc^2)^2 + (P_e c)^2$. Then energy conservation gives us the relationship

$$pc + mc^2 = p'c + E_e$$

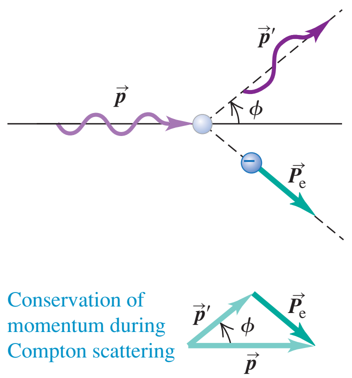
Rearranging, we find

$$(pc - p'c + mc^2)^2 = E_e{}^2 = (mc^2)^2 + (P_e c)^2 \qquad (38.8)$$

We can eliminate the electron momentum $\vec{P}_e$ from Eq. (38.8) by using momentum conservation. From Fig. 38.12 we see that $\vec{p} = \vec{p}\,' + \vec{P}_e$, or

$$\vec{P}_e = \vec{p} - \vec{p}\,' \qquad (38.9)$$

38.12 Vector diagram showing conservation of momentum in Compton scattering.



By taking the scalar product of each side of Eq. (38.9) with itself, we find

$$P_e^{\,2} = p^2 + p'^2 - 2pp'\cos\phi \tag{38.10}$$

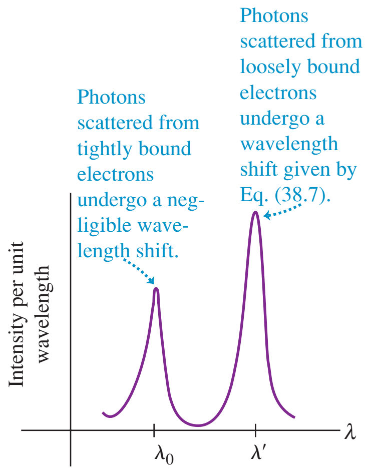
We now substitute this expression for $P_e^{\,2}$ into Eq. (38.8) and multiply out the left side. We divide out a common factor c^2; several terms cancel, and when the resulting equation is divided through by (pp'), the result is

$$\frac{mc}{p'} - \frac{mc}{p} = 1 - \cos\phi \tag{38.11}$$

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, labeled 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 the m in Eq. (38.7) is the mass of the entire atom rather than of a single electron. The resulting wavelength shifts are negligible.

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

You use 0.124-nm x-ray photons in a Compton-scattering experiment. (a) At what angle is the wavelength of the scattered x rays 1.0% longer than that of the incident x rays? (b) At what angle is it 0.050% longer?

SOLUTION

IDENTIFY and SET UP: We'll use the relationship between scattering angle and wavelength shift in the Compton effect. In each case our target variable is the angle ϕ (see Fig. 38.10b). We solve for ϕ using Eq. (38.7).

EXECUTE: (a) In Eq. (38.7) we want $\Delta\lambda = \lambda' - \lambda$ to be 1.0% of 0.124 nm, so $\Delta\lambda = 0.00124\text{ nm} = 1.24 \times 10^{-12}$ m. Using the value $h/mc = 2.426 \times 10^{-12}$ m, we find

$$\Delta\lambda = \frac{h}{mc}(1 - \cos\phi)$$

$$\cos\phi = 1 - \frac{\Delta\lambda}{h/mc} = 1 - \frac{1.24 \times 10^{-12}\text{ m}}{2.426 \times 10^{-12}\text{ m}} = 0.4889$$

$$\phi = 60.7°$$

(b) For $\Delta\lambda$ to be 0.050% of 0.124 nm, or 6.2×10^{-14} m,

$$\cos\phi = 1 - \frac{6.2 \times 10^{-14}\text{ m}}{2.426 \times 10^{-12}\text{ m}} = 0.9744$$

$$\phi = 13.0°$$

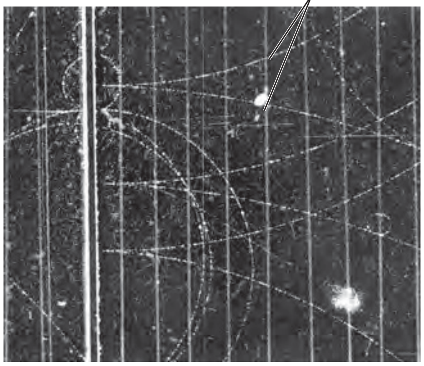
EVALUATE: Our results show that smaller scattering angles give smaller wavelength shifts. Thus in a grazing collision the photon energy loss and the electron recoil energy are smaller than when the scattering angle is larger. This is just what we would expect for an elastic collision, whether between a photon and an electron or between two billiard balls.

Pair Production

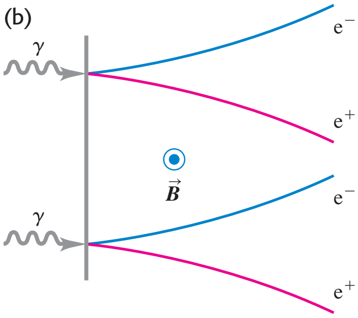
Another effect that can be explained only with the photon picture involves *gamma rays,* the shortest-wavelength and highest-frequency variety of electromagnetic radiation. If a gamma-ray photon of sufficiently short wavelength is fired at a target, it may not scatter. Instead, as depicted in Fig. 38.14, it may disappear completely and be replaced by two new particles: an electron and a **positron** (a particle that has the same rest mass m as an electron but has a positive charge $+e$ rather than the negative charge $-e$ of the electron). This process, called **pair production,** was first observed by the physicists Patrick Blackett and Giuseppe Occhialini in 1933. The electron and positron have to be produced in pairs in order to conserve electric charge: The incident photon has zero charge, and the electron–positron pair has net charge $(-e) + (+e) = 0$. Enough energy must be available to account for the rest energy $2mc^2$ of the two particles. To four significant figures, this minimum energy is

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).

(a)



(b)



$$E_{\min} = 2mc^2 = 2(9.109 \times 10^{-31}\ \text{kg})(2.998 \times 10^8\ \text{m/s})^2$$
$$= 1.637 \times 10^{-13}\ \text{J} = 1.022\ \text{MeV}$$

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

$$\lambda_{\max} = \frac{hc}{E_{\min}} = \frac{(6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^8\ \text{m/s})}{1.637 \times 10^{-13}\ \text{J}}$$
$$= 1.213 \times 10^{-12}\ \text{m} = 1.213 \times 10^{-3}\ \text{nm} = 1.213\ \text{pm}$$

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

An electron and a positron, initially far apart, move toward each other with the same speed. They collide head-on, annihilating each other and producing two photons. Find the energies, wavelengths, and frequencies of the photons if the initial kinetic energies of the electron and positron are (a) both negligible and (b) both 5.000 MeV. The electron rest energy is 0.511 MeV.

SOLUTION

IDENTIFY and SET UP: Just as in the elastic collisions we studied in Chapter 8, both momentum and energy are conserved in pair annihilation. The electron and positron are initially far apart, so the initial electric potential energy is zero and the initial energy is the sum of the particle kinetic and rest energies. The final energy is the sum of the photon energies. The total initial momentum is zero; the total momentum of the two photons must likewise be zero. We find the photon energy E using conservation of energy, conservation of momentum, and the relationship $E = pc$ (see Section 38.1). We then calculate the wavelengths and frequencies using $E = hc/\lambda = hf$.

EXECUTE: If the total momentum of the two photons is to be zero, their momenta must have equal magnitudes p and opposite directions. From $E = pc = hc/\lambda = hf$, the two photons must also have the same energy E, wavelength λ, and frequency f.

Before the collision the energy of each electron is $K + mc^2$, where K is its kinetic energy and $mc^2 = 0.511$ MeV. Conservation of energy then gives

$$(K + mc^2) + (K + mc^2) = E + E$$

Hence the energy of each photon is $E = K + mc^2$.

(a) In this case the electron kinetic energy K is negligible compared to its rest energy mc^2, so each photon has energy $E = mc^2 = 0.511$ MeV. The corresponding photon wavelength and frequency are

$$\lambda = \frac{hc}{E} = \frac{(4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(3.00 \times 10^8\ \text{m/s})}{0.511 \times 10^6\ \text{eV}}$$
$$= 2.43 \times 10^{-12}\ \text{m} = 2.43\ \text{pm}$$
$$f = \frac{E}{h} = \frac{0.511 \times 10^6\ \text{eV}}{4.136 \times 10^{-15}\ \text{eV}\cdot\text{s}} = 1.24 \times 10^{20}\ \text{Hz}$$

(b) In this case $K = 5.000$ MeV, so each photon has energy $E = 5.000\ \text{MeV} + 0.511\ \text{MeV} = 5.511$ MeV. Proceeding as in part (a), you can show that the photon wavelength is 0.2250 pm and the frequency is 1.333×10^{21} Hz.

EVALUATE: As a check, recall from Example 38.1 that a 650-nm visible-light photon has energy 1.91 eV and frequency 4.62×10^{14} Hz. The photon energy in part (a) is about 2.5×10^5 times greater. As expected, the photon's wavelength is shorter and its frequency higher than those for a visible-light photon by the same factor. You can check the results for part (b) in the same way.

Test Your Understanding of Section 38.3 If you used visible-light photons in the experiment shown in Fig. 38.11, would the photons undergo a wavelength shift due to the scattering? If so, is it possible to detect the shift with the human eye?

38.4 Wave–Particle Duality, Probability, and Uncertainty

We have studied many examples of the behavior of light and other electromagnetic radiation. Some, including the interference and diffraction effects described in Chapters 35 and 36, demonstrate conclusively the *wave* nature of light. Others, the subject of the present chapter, point with equal force to the *particle* nature of light. At first glance these two aspects seem to be in direct conflict. How can light be a wave and a particle at the same time?

We can find the answer to this apparent wave–particle conflict in the **principle of complementarity,** first stated by the Danish physicist Niels Bohr in 1928. The wave descriptions and the particle descriptions are complementary. That is, we need both to complete our model of nature, but we will never need to use both at the same time to describe a single part of an occurrence.

Diffraction and Interference in the Photon Picture

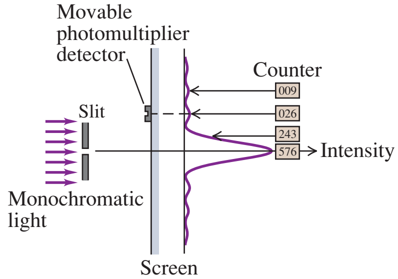
Let's start by considering again the diffraction pattern for a single slit, which we analyzed in Sections 36.2 and 36.3. Instead of recording the pattern on a digital camera chip or photographic film, we use a detector called a *photomultiplier* that can actually detect individual photons. Using the setup shown in Fig. 38.15, we place the photomultiplier at various positions for equal time intervals, count the photons at each position, and plot out the intensity distribution.

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).

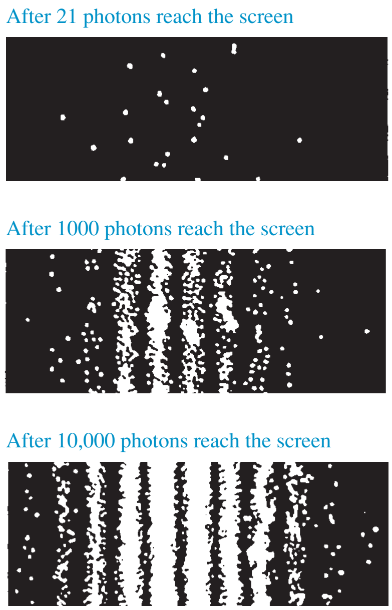
But suppose we now reduce the intensity to such a low level that only a few photons per second pass through the slit. We now record a series of discrete strikes, each representing a single photon. While we *cannot predict* where any given photon will strike, over time the accumulating strikes build up 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.

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,

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.



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.



the fact that the photomultiplier detects faint light as a sequence of individual "spots" can't be explained in wave terms.

MasteringPHYSICS®

PhET: Fourier: Making Waves
PhET: Quantum Wave Interference
ActivPhysics 17.6: Uncertainty Principle

Probability and Uncertainty

Although photons have energy and momentum, they are nonetheless very different from the particle model we used for Newtonian mechanics in Chapters 4 through 8. The Newtonian particle model treats an object as a point mass. We can describe the location and state of motion of such a particle at any instant with three spatial coordinates and three components of momentum, and we can then predict the particle's future motion. This model doesn't work at all for photons, however: We *cannot* treat a photon as a point object. This is because there are fundamental limitations on the precision with which we can simultaneously determine the position and momentum of a photon. Many aspects of a photon's behavior can be stated only in terms of *probabilities.* (In Chapter 39 we will find that the non-Newtonian ideas we develop for photons in this section also apply to particles such as electrons.)

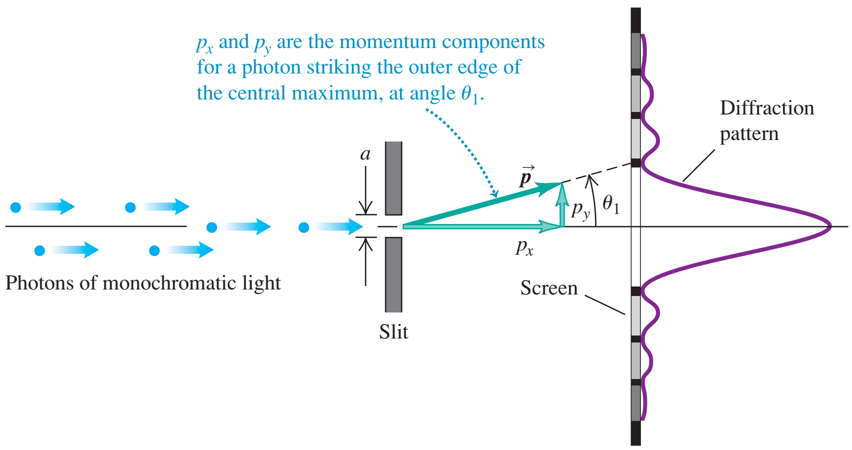
To get more insight into the problem of measuring a photon's position and momentum simultaneously, let's look again at the single-slit diffraction of light (Fig. 38.17). Suppose the wavelength λ is much less than the slit width a. Then most (85%) of the photons go into the central maximum of the diffraction pattern, and the remainder go into other parts of the pattern. We use θ_1 to denote the angle between the central maximum and the first minimum. Using Eq. (36.2) with $m = 1$, we find that θ_1 is given by $\sin\theta_1 = \lambda/a$. Since we assume $\lambda = a$, 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

38.17 Interpreting single-slit diffraction in terms of photon momentum.



$$p_y = p_x\theta_1 \tag{38.13}$$

Substituting Eq. (38.12), $\theta_1 = \lambda/a$, into Eq. (38.13) gives

$$p_y = p_x\frac{\lambda}{a} \tag{38.14}$$

Equation (38.14) says that for the 85% of the photons that strike the detector within the central maximum (that is, at angles between $-\lambda/a$ and $+\lambda/a$), the y-component of momentum is spread out over a range from $-p_x\lambda/a$ to $+p_x\lambda/a$. Now let's consider *all* the photons that pass through the slit and strike the screen. Again, they may hit above or below the center of the pattern, so their component p_y may be positive or negative. However the symmetry of the diffraction pattern shows us the average value $(p_y)_{\text{av}} = 0$. There will be an *uncertainty* Δp_y in the y-component of momentum at least as great as $p_x\lambda/a$. That is,

$$\Delta p_y \geq p_x\frac{\lambda}{a} \tag{38.15}$$

The narrower the slit width a, the broader is the diffraction pattern and the greater is the uncertainty in the y-component of momentum p_y.

The photon wavelength λ is related to the momentum p_x by Eq. (38.5), which we can rewrite as $\lambda = h/p_x$. Using this relationship in Eq. (38.15) and simplifying, we find

$$\Delta p_y \geq p_x\frac{h}{p_x a} = \frac{h}{a}$$

$$\Delta p_y a \geq h \tag{38.16}$$

What does Eq. (38.16) mean? The slit width a represents an uncertainty in the y-component of the *position* of a photon as it passes through the slit. We don't know exactly *where* in the slit each photon passes through. So both the y-position and the y-component of momentum have uncertainties, and the two uncertainties are related by Eq. (38.16). We can reduce the *momentum* uncertainty Δp_y only by reducing the width of the diffraction pattern. To do this, we have to increase the slit width a, which increases the *position* uncertainty. Conversely, when we *decrease* the position uncertainty by narrowing the slit, the diffraction pattern broadens and the corresponding momentum uncertainty *increases.*

You may protest that it doesn't seem to be consistent with common sense for a photon not to have a definite position and momentum. We reply that 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 those standard-deviation uncertainties are found to be related in general by the inequality

$$\Delta x\,\Delta p_x \geq \hbar/2 \quad \text{(Heisenberg uncertainty principle for position and momentum)} \tag{38.17}$$

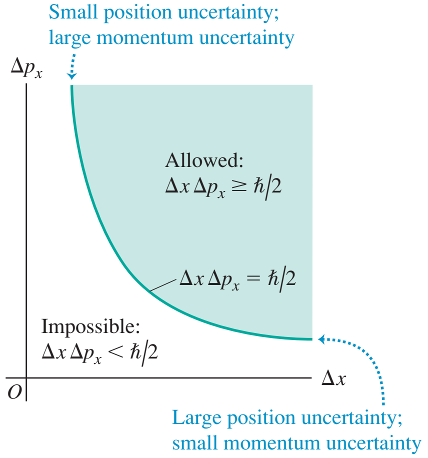
In this expression the quantity $\hbar$ (pronounced "h-bar") is Planck's constant divided by 2π:

$$\hbar = \frac{h}{2\pi} = 1.054571628(53) \times 10^{-34}\ \mathrm{J \cdot s}$$

We will use this quantity frequently to avoid writing a lot of factors of 2π in later equations.

CAUTION **h versus h-bar** It's common for students to plug in the value of h when what they really wanted was $\hbar = h/2\pi$, or vice versa. Be careful not to make the same mistake, or you'll find yourself wondering why your answer is off by a factor of 2π!

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 will see in Chapter 39 that electrons and other subatomic particles also have a wave aspect, and the same uncertainty principle applies to them as well.

It is tempting to suppose that we could get greater precision by using more sophisticated detectors of position and momentum. This turns out not to be possible. To detect a particle, the detector must *interact* with it, and this interaction 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

$$p_x = \frac{h}{\lambda} = \frac{h}{2\pi}\frac{2\pi}{\lambda} = \hbar k \qquad \text{(photon momentum in terms of wave number)} \tag{38.19a}$$

$$E = hf = \frac{h}{2\pi} 2\pi f = \hbar\omega \quad \text{(photon energy in terms of angular frequency)} \qquad (38.19b)$$

Using Eqs. (38.19) in Eq. (38.18), we can rewrite our photon wave equation as

$$E_y(x, t) = A \sin[(p_x x - Et)/\hbar] \quad \text{(wave function for a photon with } x\text{-momentum } p_x \text{ and energy } E) \qquad (38.20)$$

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

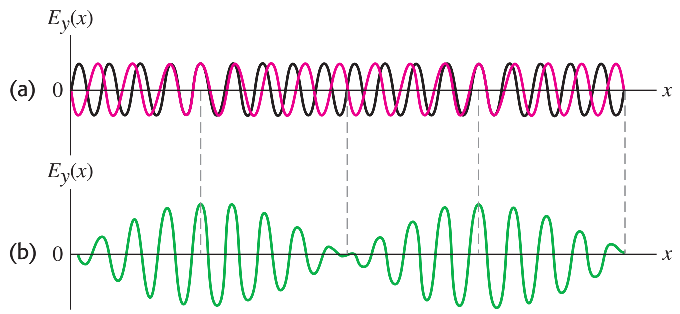
$$E_y(x, t) = A_1 \sin[(p_{1x} x - E_1 t)/\hbar] + A_2 \sin[(p_{2x} x - E_2 t)/\hbar] \qquad (38.21)$$

Consider what this wave function looks like at a particular instant of time, say $t = 0$. At this instant Eq. (38.21) becomes

$$E_y(x, t = 0) = A_1 \sin(p_{1x} x/\hbar) + A_2 \sin(p_{2x} x/\hbar) \qquad (38.22)$$

Figure 38.19a is a graph of the individual wave functions at $t = 0$ for the case $A_2 = -A_1$, and Fig. 38.19b graphs the combined wave function $E_y(x, t = 0)$ given by Eq. (38.22). We saw something very similar to Fig. 38.19b in our discussion of beats in Section 16.7: When we superimposed two sinusoidal waves with slightly different frequencies (see Fig. 16.24), 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.

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.



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. 16.24). 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 sinousoidal wave like that in Eq. (38.20) that *does* have a definite energy E but that has the same amplitude at all times, we have no idea when the photon will appear at $x = 0$. So the better we know the photon's energy, the less certain we are of when we will observe the photon.

Just as for the momentum–position uncertainty principle, we can write a mathematical expression for the uncertainty principle that relates energy and time. In fact, except for an overall minus sign, Eq. (38.23) is identical to Eq. (38.22) if we replace the x-momentum p_x by energy E and the position x by time t. This tells us that in the momentum–position uncertainty relation, Eq. (38.17), we can replace the momentum uncertainty Δp_x with the energy uncertainty ΔE and replace the position uncertainty Δx with the time uncertainty Δt. The result is

$$\Delta t \, \Delta E \geq \hbar/2 \quad \text{(Heisenberg uncertainty principle for energy and time)} \tag{38.24}$$

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

Many varieties of lasers emit light in the form of pulses rather than a steady beam. A tellurium–sapphire laser can produce light at a wavelength of 800 nm in ultrashort pulses that last only 4.00×10^{-15} s (4.00 femtoseconds, or 4.00 fs). The energy in a single pulse produced by one such laser is $2.00\ \mu\text{J} = 2.00 \times 10^{-6}$ J, and the pulses propagate in the positive x-direction. Find (a) the frequency of the light; (b) the energy and minimum energy uncertainty of a single photon in the pulse; (c) the minimum frequency uncertainty of the light in the pulse; (d) the spatial length of the pulse, in meters and as a multiple of the wavelength; (e) the momentum and minimum momentum uncertainty of a single photon in the pulse; and (f) the approximate number of photons in the pulse.

SOLUTION

IDENTIFY and SET UP: It's important to distinguish between the light pulse as a whole (which contains a very large number of photons) and an individual photon within the pulse. The 5.00-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:

$$\Delta x = c\Delta t = (3.00 \times 10^{8}\ \text{m/s})(4.00 \times 10^{-15}\ \text{s})$$
$$= 1.20 \times 10^{-6}\ \text{m}$$

$$\Delta x = \frac{1.20 \times 10^{-6}\ \text{m}}{8.00 \times 10^{-7}\ \text{m/wavelength}} = 1.50\ \text{wavelengths}$$

This justifies the term *ultrashort.* The pulse is less than two wavelengths long!

(e) From Eq. (38.5), the momentum of an average photon in the pulse is

$$p_x = \frac{E}{c} = \frac{2.48 \times 10^{-19}\ \text{J}}{3.00 \times 10^{8}\ \text{m/s}} = 8.28 \times 10^{-28}\ \text{kg}\cdot\text{m/s}$$

The spatial uncertainty is $\Delta x = 1.20 \times 10^{-6}$ m. From Eq. (38.17) minimum momentum uncertainty corresponds to $\Delta x\,\Delta p_x = \hbar/2$, so

$$\Delta p_x = \frac{\hbar}{2\Delta x} = \frac{1.055 \times 10^{-34}\ \text{J}\cdot\text{s}}{2(1.20 \times 10^{-6}\ \text{m})} = 4.40 \times 10^{-29}\ \text{kg}\cdot\text{m/s}$$

This is 5.3% of the average photon momentum p_x. An individual photon within the pulse can have a momentum that is 5.3% greater or less than the average.

(f) To estimate the number of photons in the pulse, we divide the total pulse energy by the average photon energy:

$$\frac{2.00 \times 10^{-6}\ \text{J/pulse}}{2.48 \times 10^{-19}\ \text{J/photon}} = 8.06 \times 10^{12}\ \text{photons/pulse}$$

The energy of an individual photon is uncertain, so this is the *average* number of photons per pulse.

EVALUATE: The percentage uncertainties in energy and momentum are large because this laser pulse is so short. If the pulse were longer, both Δt and Δx would be greater and the corresponding uncertainties in photon energy and photon momentum would be smaller.

Our calculation in part (f) shows an important distinction between photons and other kinds of particles. In principle it is possible to make an exact count of the number of electrons, protons, and neutrons in an object such as this book. If you repeated the count, you would get the same answer as the first time. By contrast, if you counted the number of photons in a laser pulse you would *not* necessarily get the same answer every time! The uncertainty in photon energy means that on each count there could be a different number of photons whose individual energies sum to 2.00×10^{-6} J. That's yet another of the many strange properties of photons.

Test Your Understanding of Section 38.4 Through which of the following angles is a photon of wavelength λ most likely to be deflected after passing through a slit of width a? Assume that λ is much less than a. (i) $\theta = \lambda/a$; (ii) $\theta = 3\lambda/2a$; (iii) $\theta = 2\lambda/a$; (iv) $\theta = 3\lambda/a$; (v) not enough information given to decide.

CHAPTER 38 SUMMARY

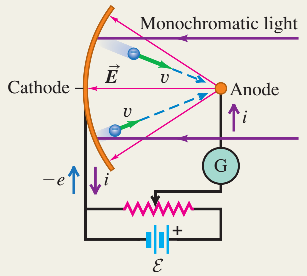
Photons: Electromagnetic radiation behaves as both waves and particles. The energy in an electromagnetic wave is carried in units called photons. The energy E of one photon is proportional to the wave frequency f and inversely proportional to the wavelength λ, and is proportional to a universal quantity h called Planck's constant. The momentum of a photon has magnitude E/c. (See Example 38.1.)

$$E = hf = \frac{hc}{\lambda} \quad (38.2)$$

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda} \quad (38.5)$$

The photoelectric effect: In the photoelectric effect, a surface can eject an electron by absorbing a photon whose energy hf is greater than or equal to the work function ϕ of the material. The stopping potential V_0 is the voltage required to stop a current of ejected electrons from reaching an anode. (See Examples 38.2 and 38.3.)

$$eV_0 = hf - \phi \quad (38.4)$$



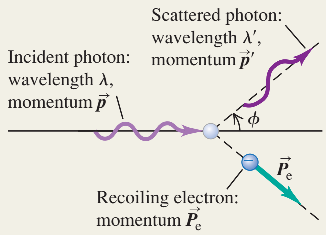
Photon production, photon scattering, and pair production: X rays can be produced when electrons accelerated to high kinetic energy across a potential increase V_{AC} strike a target. The photon model explains why the maximum frequency and minimum wavelength produced are given by Eq. (38.6). (See Example 38.4.) In Compton scattering a photon transfers some of its energy and momentum to an electron with which it collides. For free electrons (mass m), the wavelengths of incident and scattered photons are related to the photon scattering angle ϕ by Eq. (38.7). (See Example 38.5.) In pair production a photon of sufficient energy can disappear and be replaced by an electron–positron pair. In the inverse process, an electron and a positron can annihilate and be replaced by a pair of photons. (See Example 38.6.)

$$eV_{AC} = hf_{max} = \frac{hc}{\lambda_{min}} \quad (38.6)$$

(bremsstrahlung)

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\phi) \quad (38.7)$$

(Compton scattering)



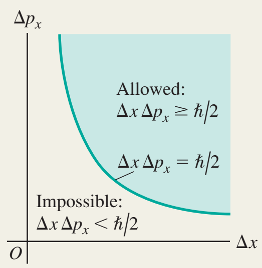
The Heisenberg uncertainty principle: It is impossible to determine both a photon's position and its momentum at the same time to arbitrarily high precision. The precision of such measurements for the x-components is limited by the Heisenberg uncertainty principle, Eq. (38.17); there are corresponding relationships for the y- and z-components. The uncertainty ΔE in the energy of a state that is occupied for a time Δt is given by Eq. (38.24). In these expressions, $\hbar = h/2\pi$. (See Example 38.7.)

$$\Delta x \, \Delta p_x \geq \hbar/2 \quad (38.17)$$

(Heisenberg uncertainty principle for position and momentum)

$$\Delta t \, \Delta E \geq \hbar/2 \quad (38.24)$$

(Heisenberg uncertainty principle for energy and time)



BRIDGING PROBLEM Compton Scattering and Electron Recoil

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

See MasteringPhysics® study area for a Video Tutor solution.

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.
2. Which key equation can be used to find the incident photon wavelength? What is the photon scattering angle ϕ in this problem?

EXECUTE

3. Use the equation you selected in step 2 to find the wavelength of the incident photon.
4. Use momentum conservation and your result from step 3 to find the momentum of the recoiling electron. (*Hint:* All of the momentum vectors are along the same line, but not all point in the same direction. Be careful with signs.)
5. Find the speed of the recoiling electron from your result in step 4. (*Hint:* Assume that the electron is nonrelativistic, so you can use the relationship between momentum and speed from Chapter 8. This is acceptable if the speed of the electron is less than about $0.1c$. Is it?)
6. Use your result from step 4 or step 5 to find the electron kinetic energy.

EVALUATE

7. You can check your answer in step 6 by finding the difference between the energies of the incident and scattered photons. Is your result consistent with conservation of energy?

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **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) A proton is moving at a speed much slower than the speed of light. It has kinetic energy K_1 and momentum p_1. If the momentum of the proton is doubled, so $p_2 = 2p_1$, how is its new kinetic energy K_2 related to K_1? (b) A photon with energy E_1 has momentum p_1. If another photon has momentum p_2 that is twice p_1, how is the energy E_2 of the second photon related to E_1?

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 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.

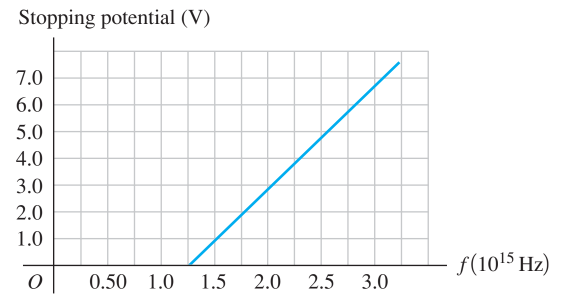
38.4 • BIO A laser used to weld detached retinas emits light with a wavelength of 652 nm in pulses that are 20.0 ms in duration. The average power during each pulse is 0.600 W. (a) How much energy is in each pulse in joules? In electron volts? (b) What is the energy of one photon in joules? In electron volts? (c) How many photons are in each pulse?

38.5 • A 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.

38.6 • A photon has momentum of magnitude 8.24×10^{-28} kg·m/s. (a) What is the energy of this photon? Give your answer in joules and in electron volts. (b) What is the wavelength of this photon? In what region of the electromagnetic spectrum does it lie?

38.7 • The graph in Fig. E38.7 shows the stopping potential as a function of the frequency of the incident light falling on a metal surface. (a) Find the photoelectric work function for this metal. (b) What value of Planck's constant does the graph yield? (c) Why does the graph *not* extend below the x-axis? (d) If a different metal were used, which characteristics of the graph would you expect to be the same and which ones would be different?

Figure **E38.7**



38.8 • The photoelectric threshold wavelength of a tungsten surface is 272 nm. Calculate the maximum kinetic energy of the electrons ejected from this tungsten surface by ultraviolet radiation of frequency 1.45×10^{15} Hz. Express the answer in electron volts.

38.9 •• 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.

38.10 •• What would the minimum work function for a metal have to be for visible light (380–750 nm) to eject photoelectrons?

38.11 •• 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?

38.12 •• The photoelectric work function of potassium is 2.3 eV. If light having a wavelength of 250 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.

38.13 • 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

38.14 • 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? (Modern televisions contain shielding to stop these x rays.)

38.15 • Protons are accelerated from rest by a potential difference of 4.00 kV and strike a metal target. If a proton produces one photon on impact, what is the minimum wavelength of the resulting x rays? How does your answer compare to the minimum wavelength if 4.00-keV electrons are used instead? Why do x-ray tubes use electrons rather than protons to produce x rays?

38.16 •• (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

38.17 • 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?

38.18 • X rays are produced in a tube operating at 18.0 kV. After emerging from the tube, x rays with the minimum wavelength produced strike a target and are Compton-scattered 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)?

38.19 •• 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?

38.20 • A beam of x rays with wavelength 0.0500 nm is Compton-scattered by the electrons in a sample. At what angle from the incident beam should you look to find x rays with a wavelength of (a) 0.0542 nm; (b) 0.0521 nm; (c) 0.0500 nm?

38.21 •• 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.

38.22 •• 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?

38.23 •• X rays with an initial wavelength of 0.900×10^{-10} m undergo Compton scattering. For what scattering angle is the wavelength of the scattered x rays greater by 1.0% than that of the incident x rays?

38.24 •• 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?

38.25 • 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

38.26 • 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%?

38.27 • 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?

38.28 • 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.

PROBLEMS

38.29 • **Exposing Photographic Film.** The light-sensitive compound on most photographic films is silver bromide, AgBr. A film is "exposed" when the light energy absorbed dissociates this molecule into its atoms. (The actual process is more complex, but the quantitative result does not differ greatly.) The energy of dissociation of AgBr is 1.00×10^5 J/mol. For a photon that is just able to dissociate a molecule of silver bromide, find (a) the photon energy in electron volts; (b) the wavelength of the photon; (c) the frequency of the photon. (d) What is the energy in electron volts of a photon having a frequency of 100 MHz? (e) Light from a firefly can expose photographic film, but the radiation from an FM station broadcasting 50,000 W at 100 MHz cannot. Explain why this is so.

38.30 •• (a) If the average frequency emitted by a 200-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 square centimeter per second if the light is emitted uniformly in all directions?

38.31 • When a certain photoelectric surface is illuminated with light of different wavelengths, the following stopping potentials are observed:

Wavelength (nm)	Stopping potential (V)
366	1.48
405	1.15
436	0.93
492	0.62
546	0.36
579	0.24

Plot the stopping potential on the vertical axis against the frequency of the light on the horizontal axis. Determine (a) the threshold frequency; (b) the threshold wavelength; (c) the photoelectric work function of the material (in electron volts); (d) the value of Planck's constant h (assuming that the value of e is known).

38.32 • 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)?

38.33 •• 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 ($4190\ \mathrm{J/kg \cdot K}$, 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?

38.34 • The photoelectric work functions for particular samples of certain metals are as follows: cesium, 2.1 eV; copper, 4.7 eV;

potassium, 2.3 eV; and zinc, 4.3 eV. (a) What is the threshold wavelength for each metal surface? (b) Which of these metals could *not* emit photoelectrons when irradiated with visible light (380–750 nm)?

38.35 •• An incident x-ray photon of wavelength 0.0900 nm is scattered in the backward direction from a free electron that is initially at rest. (a) What is the magnitude of the momentum of the scattered photon? (b) What is the kinetic energy of the electron after the photon is scattered?

38.36 •• **CP** A photon with wavelength $\lambda = 0.0900$ 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?

38.37 •• **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 the collision with the photon?

38.38 •• **CP** An x-ray tube is operating at voltage V and current I. (a) If only a fraction p of the electric power supplied is converted into x rays, at what rate is energy being delivered to the target? (b) If the target has mass m and specific heat c (in $\mathrm{J/kg \cdot K}$), at what average rate would its temperature rise if there were no thermal losses? (c) Evaluate your results from parts (a) and (b) for an x-ray tube operating at 18.0 kV and 60.0 mA that converts 1.0% of the electric power into x rays. Assume that the 0.250-kg target is made of lead ($c = 130\ \mathrm{J/kg \cdot K}$). (d) What must the physical properties of a practical target material be? What would be some suitable target elements?

38.39 •• 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.)

38.40 •• (a) Derive an expression for the total shift in photon wavelength after two successive Compton scatterings from electrons at rest. The photon is scattered by an angle θ_1 in the first scattering and by θ_2 in the second. (b) In general, is the total shift in wavelength produced by two successive scatterings of an angle $\theta/2$ the same as by a single scattering of θ? If not, are there any specific values of θ, other than $\theta = 0°$, for which the total shifts are the same? (c) Use the result of part (a) to calculate the total wavelength shift produced by two successive Compton scatterings of 30.0° each. Express your answer in terms of h/mc. (d) What is the wavelength shift produced by a single Compton scattering of 60.0°? Compare to the answer in part (c).

38.41 •• 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?

38.42 •• 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.

38.43 •• (a) Calculate the maximum increase in photon wavelength that can occur during Compton scattering. (b) What is the energy (in electron volts) of the lowest-energy x-ray photon for which Compton scattering could result in doubling the original wavelength?

CHALLENGE PROBLEM

38.44 ••• 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 ($\lambda = 10.6\ \mu\mathrm{m}$) collides head-on with a beam of electrons, each of total energy $E = 10.0$ GeV ($1\ \mathrm{GeV} = 10^9$ 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?

Answers

Chapter Opening Question ?

The energy of a photon E is inversely proportional to its wavelength λ: The shorter the wavelength, the more energetic is the photon. Since visible light has shorter wavelengths than infrared light, the headlamp emits photons of greater energy. However, the light from the infrared laser is far more *intense* (delivers much more energy per second per unit area to the patient's skin) because it emits many more photons per second than does the headlamp and concentrates them onto a very small spot.

Test Your Understanding Questions

38.1 Answers: (i) and (ii) From Eq. (38.2), a photon of energy $E = 1.14\,\text{eV}$ has wavelength $\lambda = hc/E = (4.136 \times 10^{-15}\ \text{eV}) \cdot (3.00 \times 10^{8}\ \text{m/s})/(1.14\,\text{eV}) = 1.09 \times 10^{-6}\ \text{m} = 1090\ \text{nm}$. 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.

38.2 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. Each electron produces at most one photon, so increasing the number of electrons per second causes an increase in the number of x-ray photons emitted per second (that is, the x-ray intensity).

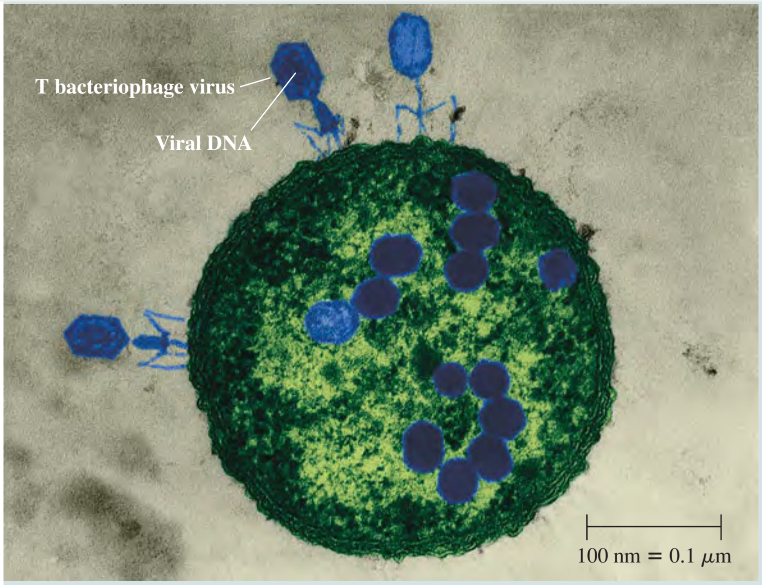
38.3 Answers: yes, no Equation (38.7) shows that the wavelength shift $\Delta\lambda = \lambda' - \lambda$ depends only on the photon scattering angle ϕ, not on the wavelength of the incident photon. So a visible-light photon scattered through an angle ϕ undergoes the same wavelength shift as an x-ray photon. Equation (38.7) also shows that this shift is of the order of $h/mc = 2.426 \times 10^{-12}\ \text{m} = 0.002426\ \text{nm}$. 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 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.

Bridging Problem

Answers: **(a)** 0.0781 nm
(b) $1.65 \times 10^{-23}\ \text{kg}\cdot\text{m/s}$, $1.81 \times 10^{7}\ \text{m/s}$
(c) $1.49 \times 10^{-16}\ \text{J}$

39 PARTICLES BEHAVING AS WAVES



? 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 using a beam of electrons rather than a light beam. What properties of electrons make them useful for imaging such fine details?

LEARNING GOALS

By studying this chapter, you will learn:

- De Broglie's proposal that electrons, protons, and other particles can behave like waves.
- How electron diffraction experiments provided evidence for de Broglie's ideas.
- How electron microscopes can provide much higher magnification than visible-light microscopes.
- How physicists discovered the atomic nucleus.
- How Bohr's model of electron orbits explained the spectra of hydrogen and hydrogenlike atoms.
- How a laser operates.
- How the idea of energy levels, coupled with the photon model of light, explains the spectrum of light emitted by a hot, opaque object.
- What the uncertainty principle tells us about the nature of the atom.

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? As we will discover, the answer is a resounding yes. Electrons can be made to interfere and diffract just like other kinds of waves. We will see that 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 profoundly unstable, are able to exist. In this chapter we'll use the wave nature of matter to 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 and nobleman, Prince Louis de Broglie (pronounced "de broy"; Fig. 39.1), 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 and protons, 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 as for a photon, as expressed by Eq. (38.5): $\lambda = h/p$. The **de Broglie wavelength** of a particle is then

$$\lambda = \frac{h}{p} = \frac{h}{mv} \qquad \text{(de Broglie wavelength of a particle)} \qquad (39.1)$$

where h is Planck's constant. If the particle's speed is an appreciable fraction of the speed of light c, we use Eq. (37.27) to replace mv in Eq. (39.1) with $\gamma mv = mv/\sqrt{1 - v^2/c^2}$. The frequency f, according to de Broglie, is also related to the particle's energy E in the same way as for a photon—namely,

$$E = hf \qquad (39.2)$$

Thus in de Broglie's hypothesis, the relationships of wavelength to momentum and of frequency to energy are exactly the same for free particles as for photons.

CAUTION **Not all photon equations apply to particles with mass** Be careful when applying the relationship $E = hf$ to particles with nonzero rest mass, such as electrons and protons. Unlike a photon, they do *not* travel at speed c, so the equations $f = c/\lambda$ and $E = pc$ do *not* apply to them!

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.

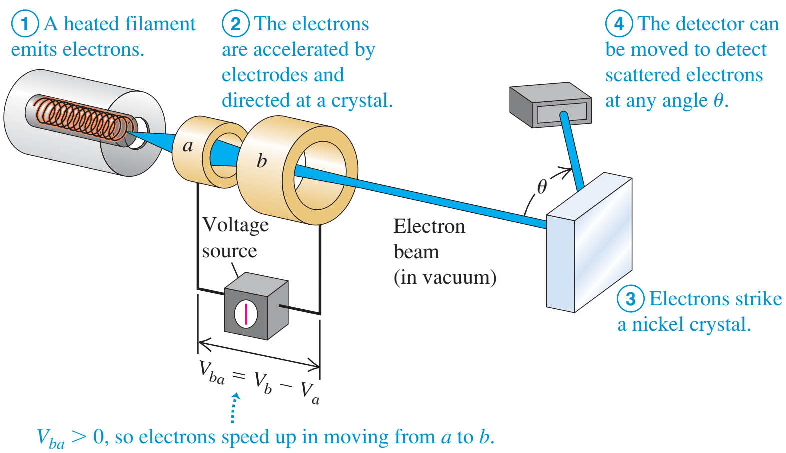
Observing the Wave Nature of Electrons

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 of de Broglie's publication of his ideas, they 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.

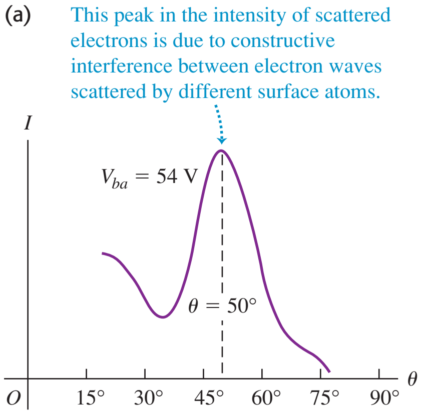
MasteringPHYSICS

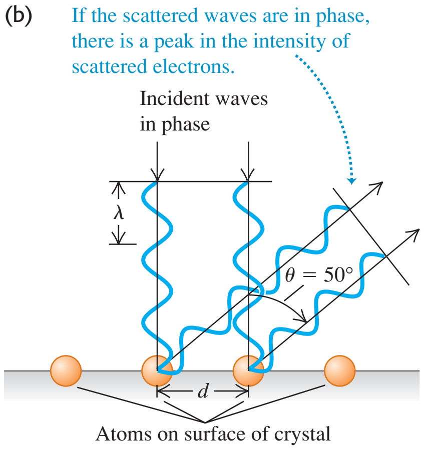
PhET: Davisson-Germer: Electron Diffraction
ActivPhysics 17.5: Electron Interference



39.2 An apparatus similar to that used by Davisson and Germer to discover electron diffraction.

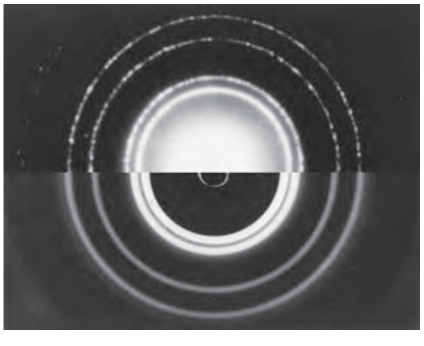
39.3 (a) Intensity of the scattered electron beam in Fig. 39.2 as a function of the scattering angle θ. (b) Electron waves scattered from two adjacent atoms interfere constructively when $d\sin\theta = m\lambda$. In the case shown here, $\theta = 50°$ and $m = 1$.





39.4 X-ray and electron diffraction. The upper half of the photo shows the diffraction pattern for 71-pm x rays passing through aluminum foil. The lower half, with a different scale, shows the diffraction pattern for 600-eV electrons from aluminum. The similarity shows that electrons undergo the same kind of diffraction as x rays.

Top: x-ray diffraction



Bottom: electron diffraction

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, almost hot enough to melt it. Unknown to them, this had the effect of creating large regions within the nickel with crystal planes that were continuous over the width of the electron beam. From the perspective of the electrons, the sample looked like a *single* crystal of nickel.

When the observations were repeated with this sample, the results were quite different. Now strong maxima in the intensity of the reflected electron beam occurred at specific angles (Fig. 39.3a), in contrast to the smooth variation of intensity with angle that Davisson and Germer had observed before the accident. The angular positions of the maxima depended on the accelerating voltage V_{ba} used to produce the electron beam. Davisson and Germer were familiar with de Broglie's hypothesis, and they noticed the similarity of this behavior to x-ray diffraction. This was not the effect they had been looking for, but they immediately recognized that the electron beam was being *diffracted.* They had discovered a very direct experimental confirmation of the wave hypothesis.

Davisson and Germer could determine the speeds of the electrons from the accelerating voltage, so they could compute the de Broglie wavelength from Eq. (39.1). If an electron is accelerated from rest at point a to point b through a potential increase $V_{ba} = V_b - V_a$ as shown in Fig. 39.2, the work done on the electron eV_{ba} equals its kinetic energy K. Using $K = (\frac{1}{2})mv^2 = p^2/2m$ for a nonrelativistic particle, we have

$$eV_{ba} = \frac{p^2}{2m} \qquad p = \sqrt{2meV_{ba}}$$

We substitute this into Eq. (39.1), the expression for the de Broglie wavelength of the electron:

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2meV_{ba}}} \qquad \text{(de Broglie wavelength of an electron)} \quad (39.3)$$

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

$$d\sin\theta = m\lambda \qquad (m = 1, 2, 3, \dots) \quad (39.4)$$

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, using the de Broglie wavelength given by Eq. (39.3), agreed with the observed values (Fig. 39.3a). Thus the accidental discovery of **electron diffraction** was the first direct evidence confirming de Broglie's hypothesis.

In 1928, just a year after the Davisson–Germer discovery, the English physicist G. P. Thomson carried out electron-diffraction experiments using a thin, polycrystalline, metallic foil as a target. Debye and Sherrer had used a similar technique several years earlier to study x-ray diffraction from polycrystalline specimens. In these experiments the beam passes *through* the target rather than being reflected from it. Because of the random orientations of the individual microscopic crystals in the foil, the diffraction pattern consists of intensity maxima forming rings around the direction of the incident beam. Thomson's results again confirmed the de Broglie relationship. Figure 39.4 shows both x-ray and electron diffraction patterns for a polycrystalline aluminum foil. (G. P. Thomson was the son of J. J. Thomson, who 31 years earlier discovered the electron. Davisson and the younger Thomson shared the 1937 Nobel Prize in physics for their discoveries.)

Additional diffraction experiments were soon carried out in many laboratories using not only electrons but also various ions and low-energy neutrons. All of these are in agreement with de Broglie's bold predictions. Thus the wave nature of particles, so strange in 1924, became firmly established in the years that followed.

Problem-Solving Strategy 39.1 Wavelike Properties of Particles

IDENTIFY *the relevant concepts:* Particles have wavelike properties. A particle's (de Broglie) wavelength is inversely proportional to its momentum, and its frequency is proportional to its energy.

SET UP *the problem:* Identify the target variables and decide which equations you will use to calculate them.

EXECUTE *the solution* as follows:

1. Use Eq. (39.1) to relate a particle's momentum p to its wavelength λ; use Eq. (39.2) to relate its energy E to its frequency f.
2. Nonrelativistic kinetic energy may be expressed as either $K = \frac{1}{2}mv^2$ or (because $p = mv$) $K = p^2/2m$. The latter form is useful in calculations involving the de Broglie wavelength.
3. You may express energies in either joules or electron volts, using $h = 6.626 \times 10^{-34}\ \text{J}\cdot\text{s}$ or $h = 4.136 \times 10^{-15}\ \text{eV}\cdot\text{s}$ as appropriate.

EVALUATE *your answer:* To check numerical results, it helps to remember some approximate orders of magnitude. Here's a partial list:

Size of an atom: $10^{-10}\ \text{m} = 0.1\ \text{nm}$

Mass of an atom: $10^{-26}\ \text{kg}$

Mass of an electron: $m = 10^{-30}\ \text{kg}$; $mc^2 = 0.511\ \text{MeV}$

Electron charge magnitude: $10^{-19}\ \text{C}$

kT at room temperature: $\frac{1}{40}\ \text{eV}$

Difference between energy levels of an atom (to be discussed in Section 39.3): 1 to 10 eV

Speed of an electron in the Bohr model of a hydrogen atom (to be discussed in Section 39.3): $10^6\ \text{m/s}$

Example 39.1 An electron-diffraction experiment

In an electron-diffraction experiment using an accelerating voltage of 54 V, an intensity maximum occurs for $\theta = 50°$ (see Fig. 39.3a). X-ray diffraction indicates that the atomic spacing in the target is $d = 2.18 \times 10^{-10}\ \text{m} = 0.218\ \text{nm}$. The electrons have negligible kinetic energy before being accelerated. Find the electron wavelength.

SOLUTION

IDENTIFY, SET UP, and EXECUTE: We'll determine λ from both de Broglie's equation, Eq. (39.3), and the diffraction equation, Eq. (39.4). From Eq. (39.3),

$$\lambda = \frac{6.626 \times 10^{-34}\ \text{J}\cdot\text{s}}{\sqrt{2(9.109 \times 10^{-31}\ \text{kg})(1.602 \times 10^{-19}\ \text{C})(54\ \text{V})}}$$
$$= 1.7 \times 10^{-10}\ \text{m} = 0.17\ \text{nm}$$

Alternatively, using Eq. (39.4) and assuming $m = 1$,

$$\lambda = d\sin\theta = (2.18 \times 10^{-10}\ \text{m})\sin 50° = 1.7 \times 10^{-10}\ \text{m}$$

EVALUATE: The two numbers agree within the accuracy of the experimental results, which gives us an excellent check on our calculations. Note that this electron wavelength is less than the spacing between the atoms.

Example 39.2 Energy of a thermal neutron

Find the speed and kinetic energy of a neutron ($m = 1.675 \times 10^{-27}\ \text{kg}$) with de Broglie wavelength $\lambda = 0.200\ \text{nm}$, a typical interatomic spacing in crystals. Compare this energy with the average translational kinetic energy of an ideal-gas molecule at room temperature ($T = 20°\text{C} = 293\ \text{K}$).

SOLUTION

IDENTIFY and SET UP: This problem uses the relationships between particle speed and wavelength, between particle speed and kinetic energy, and between gas temperature and the average kinetic energy of a gas molecule. We'll find the neutron speed v 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

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}(1.675 \times 10^{-27}\ \text{kg})(1.98 \times 10^3\ \text{m/s})^2$$
$$= 3.28 \times 10^{-21}\ \text{J} = 0.0205\ \text{eV}$$

Continued

From Eq. (18.16), the average translational kinetic energy of an ideal-gas molecule at $T = 293$ K is

$$\tfrac{1}{2}m(v^2)_{\text{av}} = \tfrac{3}{2}kT = \tfrac{3}{2}(1.38 \times 10^{-23}\ \text{J/K})(293\ \text{K})$$
$$= 6.07 \times 10^{-21}\ \text{J} = 0.0379\ \text{eV}$$

The two energies are comparable in magnitude, which is why a neutron with kinetic energy in this range is called a *thermal neutron.*

Diffraction of thermal neutrons is used to study crystal and molecular structure in the same way as x-ray diffraction. Neutron diffraction has proved to be especially useful in the study of large organic molecules.

EVALUATE: Note that the calculated neutron speed is much less than the speed of light. This justifies our use of the nonrelativistic form of Eq. (39.1).

De Broglie Waves and the Macroscopic World

If the de Broglie picture is correct and matter has wave aspects, you might wonder why we don't see these aspects in everyday life. As an example, we know that waves diffract when sent through a single slit. Yet when we walk through a doorway (a kind of single slit), we don't worry about our body diffracting!

The principal reason we don't see these effects on human scales is that Planck's constant h has such a minuscule value. As a result, the de Broglie wavelengths of even the smallest ordinary objects that you can see are extremely small, and the wave effects are unimportant. For instance, what is the wavelength of a falling grain of sand? If the grain's mass is 5×10^{-10} kg and its diameter is $0.07\ \text{mm} = 7 \times 10^{-5}$ m, it will fall in air with a terminal speed of about 0.4 m/s. The magnitude of its momentum is then $p = mv = (5 \times 10^{-10}\ \text{kg}) \times (0.4\ \text{m/s}) = 2 \times 10^{-10}\ \text{kg} \cdot \text{m/s}$. 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 neglected. 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 neglect 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)?

SOLUTION

IDENTIFY, SET UP, and EXECUTE: We can use the same concepts we used to understand the Davisson–Germer experiment. The accelerating voltage is the quantity V_{ba} in Eq. (39.3). Rewrite this equation to solve for V_{ba}:

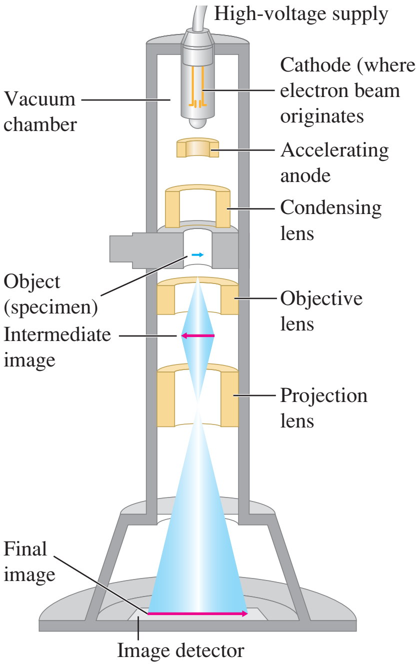
$$\begin{aligned} V_{ba} &= \frac{h^2}{2me\lambda^2} \\ &= \frac{(6.626 \times 10^{-34}\ \mathrm{J \cdot s})^2}{2(9.109 \times 10^{-31}\ \mathrm{kg})(1.602 \times 10^{-19}\ \mathrm{C})(10 \times 10^{-12}\ \mathrm{m})^2} \\ &= 1.5 \times 10^4\ \mathrm{V} = 15{,}000\ \mathrm{V} \end{aligned}$$

EVALUATE: It is easy to attain 15-kV accelerating voltages from 120-V or 240-V line voltage using a step-up transformer (Section 31.6) and a rectifier (Section 31.1). The accelerated electrons have kinetic energy 15 keV; since the electron rest energy is 0.511 MeV $=$ 511 keV, these electrons are indeed nonrelativistic.

Types of Electron Microscope

Figure 39.5 shows the design of a *transmission electron microscope,* in which electrons actually pass through the specimen being studied. The specimen to be viewed can be no more than 10 to 100 nm thick so the electrons are not slowed appreciably as they pass through. The electrons used in a transmission electron microscope are emitted from a hot cathode and accelerated by a potential difference, typically 40 to 400 kV. They then pass through a condensing "lens" that uses magnetic fields to focus the electrons into a parallel beam before they pass through the specimen. The beam then passes through two more magnetic lenses: an objective lens that forms an intermediate image of the specimen and a projection lens that produces a final real image of the intermediate image. The objective and projection lenses play the roles of the objective and eyepiece lenses, respectively, of a compound optical microscope (see Section 34.8). The final image is projected onto a fluorescent screen for viewing or photographing. The entire apparatus, including the specimen, must be enclosed in a vacuum container; otherwise, electrons would scatter off air molecules and muddle the image. The image that opens this chapter was made with a transmission electron microscope.

39.5 Schematic diagram of a transmission electron microscope (TEM).



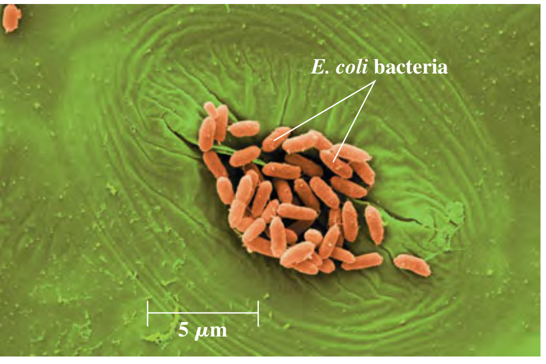
We might think that when the electron wavelength is 0.01 nm (as in Example 39.3), the resolution would also be about 0.01 nm. In fact, it is seldom better than 0.1 nm, in part because the focal length of a magnetic lens depends on the electron speed, which is never exactly the same for all electrons in the beam.

An important variation is the *scanning electron microscope*. The electron beam is focused to a very fine line and scanned across the specimen. The beam knocks additional electrons off the specimen wherever it hits. These ejected electrons are collected by an anode that is kept at a potential a few hundred volts positive with respect to the specimen. The current of ejected electrons flowing to the collecting anode varies as the microscope beam sweeps across the specimen. The varying strength of the current is then used to create a "map" of the scanned specimen, and this map forms a greatly magnified image of the specimen.

This scheme has several advantages. The specimen can be thick because the beam does not need to pass through it. Also, the knock-off electron production depends on the *angle* at which the beam strikes the surface. Thus scanning electron micrographs have an appearance that is much more three-dimensional than conventional visible-light micrographs (Fig. 39.6). The resolution is typically of the order of 10 nm, not as good as a transmission electron microscope but still much finer than the best optical microscopes.

Test Your Understanding of Section 39.1 (a) A proton has a slightly smaller mass than a neutron. Compared to the neutron described in Example 39.2, would a proton of the same wavelength have (i) more kinetic energy; (ii) less kinetic energy; or (iii) the same kinetic energy? (b) Example 39.1 shows that to give electrons a wavelength of 1.7×10^{-10} m, they must be accelerated from rest through a voltage of 54 V and so acquire a kinetic energy of 54 eV. Does a photon of this same energy also have a wavelength of 1.7×10^{-10} m?

39.6 This scanning electron microscope image shows *Escherichia coli* bacteria crowded into a stoma, or respiration opening, on the surface of a lettuce leaf. (False color has been added.) If not washed off before the lettuce is eaten, these bacteria can be a health hazard. The *transmission* electron micrograph that opens this chapter shows a greatly magnified view of the surface of an *E. coli* bacterium.

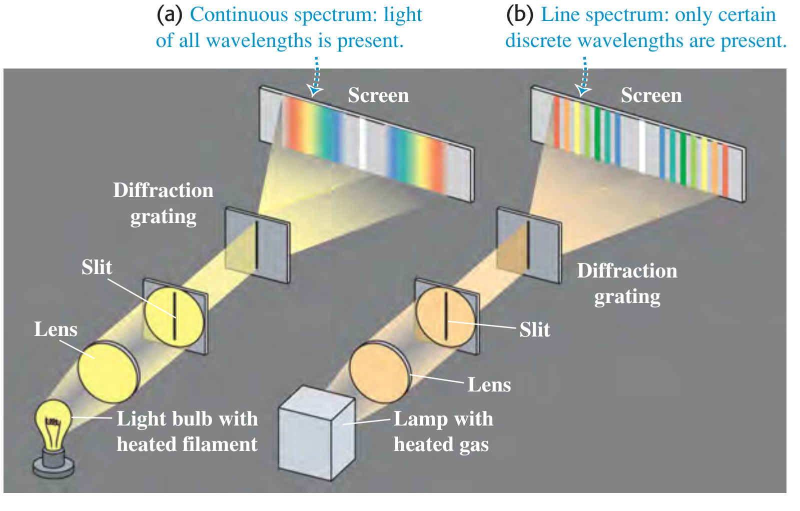


39.2 The Nuclear Atom and Atomic Spectra

Every neutral atom contains at least one electron. How does the wave aspect of electrons affect atomic structure? As we will see, it is crucial for understanding not only the structure of atoms but also how they interact with light. Historically, the quest to understand the nature of the atom was intimately linked with both the idea that electrons have wave characteristics and the notion that light has particle characteristics. Before we explore how these ideas shaped atomic theory, it's useful to look at what was known about atoms—as well as what remained mysterious—by the first decade of the twentieth century.

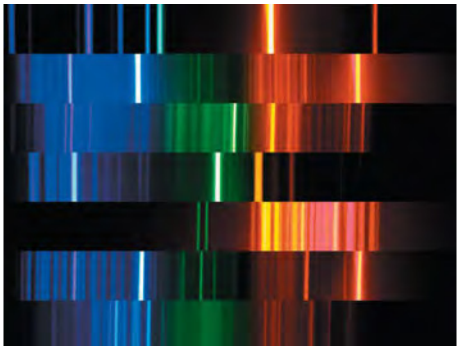
Line Spectra

Everyone knows that heated materials emit light, and that 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 an advertising 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.



39.7 (a) Continuous spectrum produced by a glowing light bulb filament.
(b) Emission line spectrum emitted by a lamp containing a heated gas.

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 *seeing* an atom using that light. But we can still describe how the mass and electric charge are distributed throughout the volume of the atom.

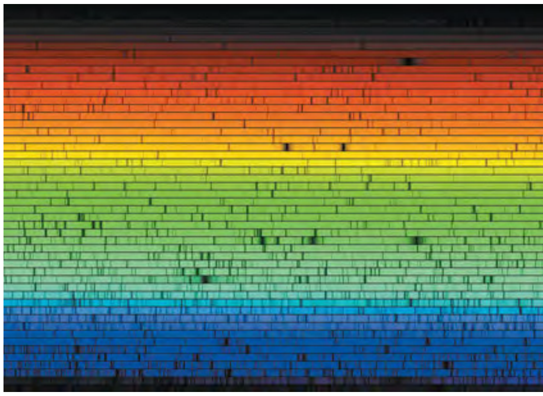
Here's where things stood in 1910. In 1897 the English physicist J. J. Thomson (Nobel Prize 1906) had discovered the electron and measured its charge-to-mass ratio e/m. By 1909, the American physicist Robert Millikan (Nobel Prize 1923) had made the first measurements of the electron charge $-e$. These 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.)

Application Using Spectra to Analyze an Interstellar Gas Cloud

The light from this glowing gas cloud—located in the Small Magellanic Cloud, a small satellite galaxy of the Milky Way some 200,000 light-years (1.9×10^{18} km) from earth—has an emission line spectrum. Despite its immense distance, astronomers can tell that this cloud is composed mostly of hydrogen because its spectrum is dominated by red light at a wavelength of 656.3 nm, a wavelength emitted by hydrogen and no other element.

39.9 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.



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.

PhET: Rutherford Scattering
ActivPhysics 19.1: Particle Scattering

Rutherford's Exploration of the Atom

The first experiments designed to test Thomson's model by probing the interior structure of the atom were carried out in 1910–1911 by Ernest Rutherford (Fig. 39.10) and two of his students, Hans Geiger and Ernest Marsden, at the University of Manchester in England. These experiments consisted of shooting a beam of charged particles at thin foils of various elements and observing how the foil deflected the particles.

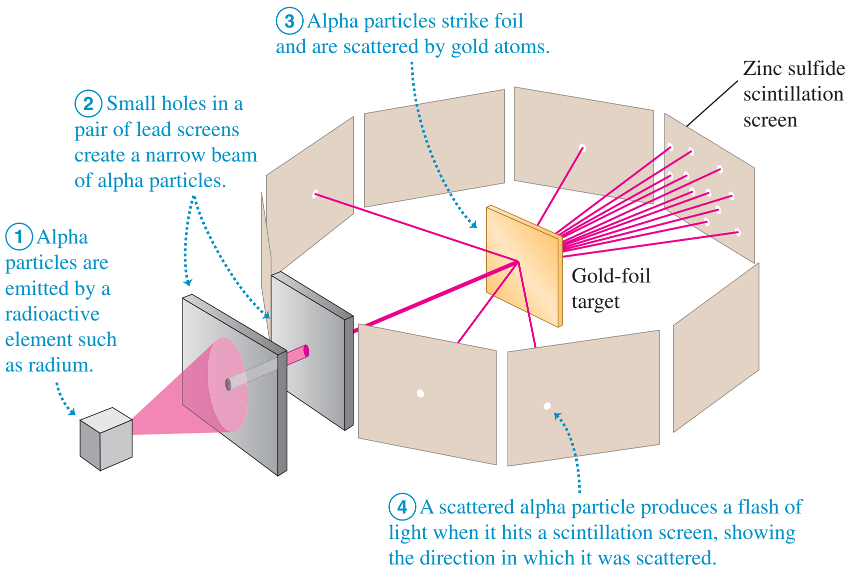
The particle accelerators now in common use in laboratories had not yet been invented, and Rutherford's projectiles were *alpha particles* emitted from naturally radioactive elements. The nature of these alpha particles was not completely understood, but it was known that they are ejected from unstable nuclei with speeds of the order of 10^7 m/s, are positively charged, and can travel several centimeters through air or 0.1 mm or so through solid matter before they are brought to rest by collisions.

Figure 39.11 is a schematic view of Rutherford's experimental setup. A radioactive substance at the left emits alpha particles. Thick lead screens stop all particles except those in a narrow beam. The beam passes through the foil target (consisting of gold, silver, or copper) and strikes screens coated with zinc sulfide, creating a momentary flash, or *scintillation*. Rutherford and his students counted the numbers of particles deflected through various angles.

The atoms in a metal foil are packed together like marbles in a box (not spaced apart). Because the particle beam passes through the foil, the alpha particles must pass through the interior of atoms. Within an atom, the charged alpha particle will interact with the electrons and the positive charge. (Because the *total* charge of the atom is zero, alpha particles feel little electrical 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). The results of the Rutherford experiments were *very* different

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.



from the Thomson prediction. Some alpha particles were scattered by nearly 180°—that is, almost straight backward (Fig. 39.12b). Rutherford later wrote:

> **It was quite the most incredible event that ever happened to me in my life. It was almost as incredible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you.**

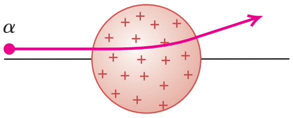
Clearly the Thomson model was wrong and a new model was needed. Suppose the positive charge, instead of being distributed through a sphere with atomic dimensions (of the order of 10^{-10} m), is all concentrated in a much *smaller* volume. Then it would act like a point charge down to much smaller distances. The maximum electric field repelling the alpha particle would be much larger, and the amazing large-angle scattering that Rutherford observed could occur. Rutherford developed this model and called the concentration of positive charge the **nucleus.** He again computed the numbers of particles expected to be scattered through various angles. Within the accuracy of his experiments, the computed and measured results agreed, down to distances of the order of 10^{-14} m. His experiments therefore established that the atom does have a nucleus—a very small, very dense structure, no larger than 10^{-14} m in diameter. The nucleus occupies only about 10^{-12} of the total volume of the atom or less, but it contains *all* the positive charge and at least 99.95% of the total mass of the atom.

Figure 39.13 shows a computer simulation of alpha particles with a kinetic energy of 5.0 MeV being scattered from a gold nucleus of radius 7.0×10^{-15} m (the actual value) and from a nucleus with a hypothetical radius ten times larger. In the second case there is *no* large-angle scattering. The presence of large-angle scattering in Rutherford's experiments thus attested to the small size of the nucleus.

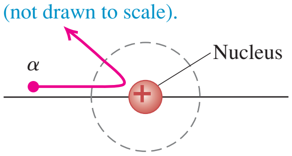
Later experiments showed that all nuclei are composed of positively charged protons (discovered in 1918) and electrically neutral neutrons (discovered in 1930). For example, the gold atoms in Rutherford's experiments have 79 protons and 118 neutrons. In fact, an alpha particle is itself the nucleus of a helium atom, with two protons and two neutrons. It is much more massive than an electron but only about 2% as massive as a gold nucleus, which helps explain why alpha particles are scattered by gold nuclei but not by electrons.

39.12 A comparison of Thomson's and Rutherford's models of the atom.

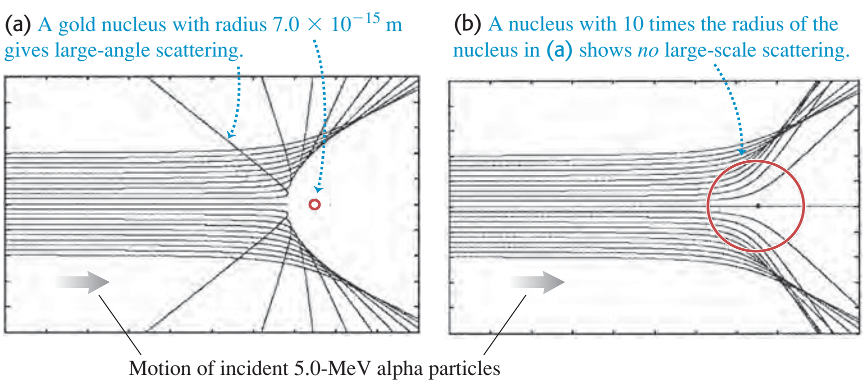
(a) Thomson's model of the atom: An alpha particle is scattered through only a small angle.



(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).



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.

Continued

SOLUTION

IDENTIFY: The repulsive electric force exerted by the gold nucleus makes the 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

$$\begin{aligned} U_2 &= \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r} \\ &= (9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2})\frac{(2)(79)(1.60 \times 10^{-19}\ \mathrm{C})^2}{5.0 \times 10^{-14}\ \mathrm{m}} \\ &= 7.3 \times 10^{-13}\ \mathrm{J} = 4.6 \times 10^6\ \mathrm{eV} = 4.6\ \mathrm{MeV} \end{aligned}$$

By energy conservation $K_1 + U_1 = K_2 + U_2$, so $K_1 = K_2 + U_2 - U_1 = 0 + 4.6\ \mathrm{MeV} - 0 = 4.6\ \mathrm{MeV}$. Thus, to approach within 5.0×10^{-14} m, the alpha particle must have initial kinetic energy $K_1 = 4.6$ MeV.

EVALUATE: Alpha particles emitted from naturally occurring radioactive elements typically have energies in the range 4 to 6 MeV. For example, the common isotope of radium, $^{226}\mathrm{Ra}$, emits an alpha particle with energy 4.78 MeV.

Was it valid to assume that the gold nucleus remains at rest? To find out, note that when the alpha particle stops momentarily, all of its initial momentum has been transferred to the gold nucleus. An alpha particle has a mass $m_\alpha = 6.64 \times 10^{-27}$ kg; if its initial kinetic energy $K_1 = \frac{1}{2}mv_1^2$ is 7.3×10^{-13} J, you can show that its initial speed is $v_1 = 1.5 \times 10^7$ m/s and its initial momentum is $p_1 = m_\alpha v_1 = 9.8 \times 10^{-20}\ \mathrm{kg \cdot m/s}$. A gold nucleus (mass $m_{\mathrm{Au}} = 3.27 \times 10^{-25}$ kg) with this much momentum has a much slower speed $v_{\mathrm{Au}} = 3.0 \times 10^5$ m/s and kinetic energy $K_{\mathrm{Au}} = \frac{1}{2}mv_{\mathrm{Au}}^2 = 1.5 \times 10^{-14}\ \mathrm{J} = 0.092$ MeV. This *recoil kinetic energy* of the gold nucleus is only 2% of the total energy in this situation, so we are justified in neglecting it.

The Failure of Classical Physics

Rutherford's discovery of the atomic nucleus raised a serious question: What prevented the negatively charged electrons from falling into the positively charged nucleus due to the strong electrostatic attraction? Rutherford suggested that perhaps the electrons *revolve* in orbits about the nucleus, just as the planets revolve around the sun.

But according to classical electromagnetic theory, any accelerating electric charge (either oscillating or revolving) radiates electromagnetic waves. An example is the radiation from an oscillating point charge that we depicted in Fig. 32.3 (Section 32.1). An electron orbiting inside an atom would always have a centripetal acceleration toward the nucleus, and so should be emitting radiation *at all times*. The energy of an orbiting electron should therefore decrease continuously, its orbit should become smaller and smaller, and it should spiral into the nucleus within a fraction of a second (Fig. 39.14). Even worse, according to classical theory the *frequency* of the electromagnetic waves emitted should equal the frequency of revolution. As the electrons radiated energy, their angular speeds would change continuously, and they would emit a *continuous* spectrum (a mixture of all frequencies), not the *line* spectrum actually observed.

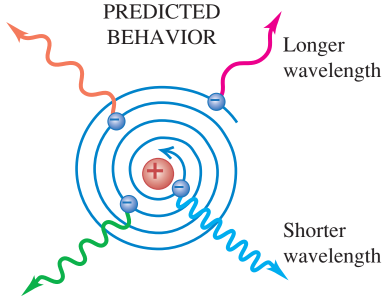
Thus Rutherford's model of electrons orbiting the nucleus, which is based on Newtonian mechanics and classical electromagnetic theory, makes three entirely *wrong* predictions about atoms: They should emit light continuously, they should be unstable, and the light they emit should have a continuous spectrum. Clearly a radical reappraisal of physics on the scale of the atom was needed. In the next section we will see the bold idea that led to an new understanding of the atom, and see how this idea meshes with de Broglie's no less bold notion that electrons have wave attributes.

39.14 Classical physics makes predictions about the behavior of atoms that do not match reality.

ACCORDING TO CLASSICAL PHYSICS:

- An orbiting electron is accelerating, so it should radiate electromagnetic waves.
- The waves would carry away energy, so the electron should lose energy and spiral inward.
- The electron's angular speed would increase as its orbit shrank, so the frequency of the radiated waves should increase.

Thus, classical physics says that atoms should collapse within a fraction of a second and should emit light with a continuous spectrum as they do so.



IN FACT:

- Atoms are stable.
- They emit light only when excited, and only at specific frequencies (as a line spectrum).

Test Your Understanding of Section 39.2 Suppose you repeated Rutherford's scattering experiment using a thin sheet of solid hydrogen in place of the gold foil. (Hydrogen is a solid at temperatures below 14.0 K.) The nucleus of a hydrogen atom is a single proton, with about one-fourth the mass of an alpha particle. Compared to the original experiment with gold foil, would you expect the alpha particles in this experiment to undergo (i) more large-angle scattering; (ii) the same amount of large-angle scattering; or (iii) less large-angle scattering?

39.3 Energy Levels and the Bohr Model of the Atom

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.

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.

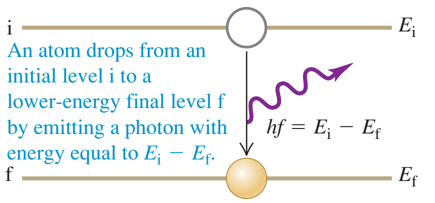
Photon Emission and Absorption by Atoms

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). If E_i is the initial energy of the atom before such a transition, E_f is its final energy after the transition, and the photon's energy is $hf = hc/\lambda$, then conservation of energy gives

$$hf = \frac{hc}{\lambda} = E_i - E_f \qquad \text{(energy of emitted photon)} \tag{39.5}$$

39.16 An excited atom emitting a photon.



For example, an excited lithium atom emits red light with wavelength $\lambda = 671\ \text{nm}$. 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 $E_i - E_f = 1.85\ \text{eV}$.

The emission line spectra shown in 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 energy levels of the atom.

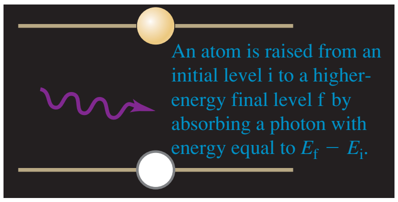
CAUTION **Producing a line spectrum** The lines of an emission line spectrum, such as the helium spectrum shown at the top of Fig. 39.8, are *not* all produced by a single atom. The sample of helium gas that produced the spectrum in Fig. 39.8 contained a large number of helium atoms; these were excited in an electric discharge tube to various energy levels. The spectrum of the gas shows the light emitted from all the different transitions that occurred in different atoms of the sample.

MasteringPHYSICS

PhET: Models of the Hydrogen Atom
ActivPhysics 18.1: The Bohr Model

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.

39.17 An atom absorbing a photon. (Compare with Fig. 39.16.)



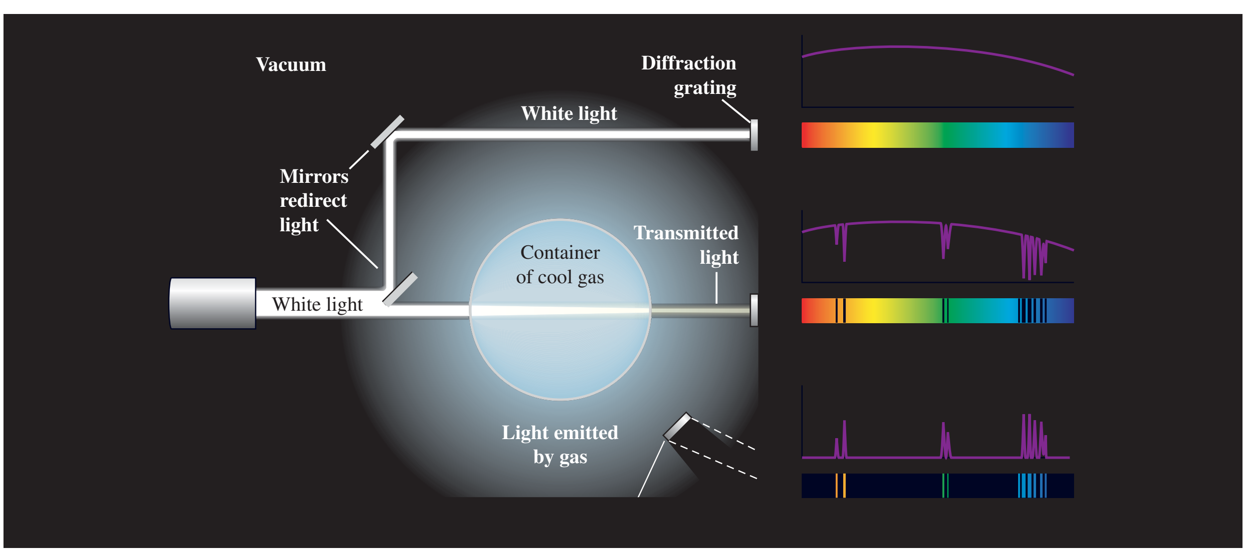
Collisions are not the only way that an atom's energy can be raised from one level to a higher level. If an atom initially in the lower energy level in Fig. 39.16 is struck by a photon with just the right amount of energy, the photon can be *absorbed* and the atom will end up in the higher level (Fig. 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.

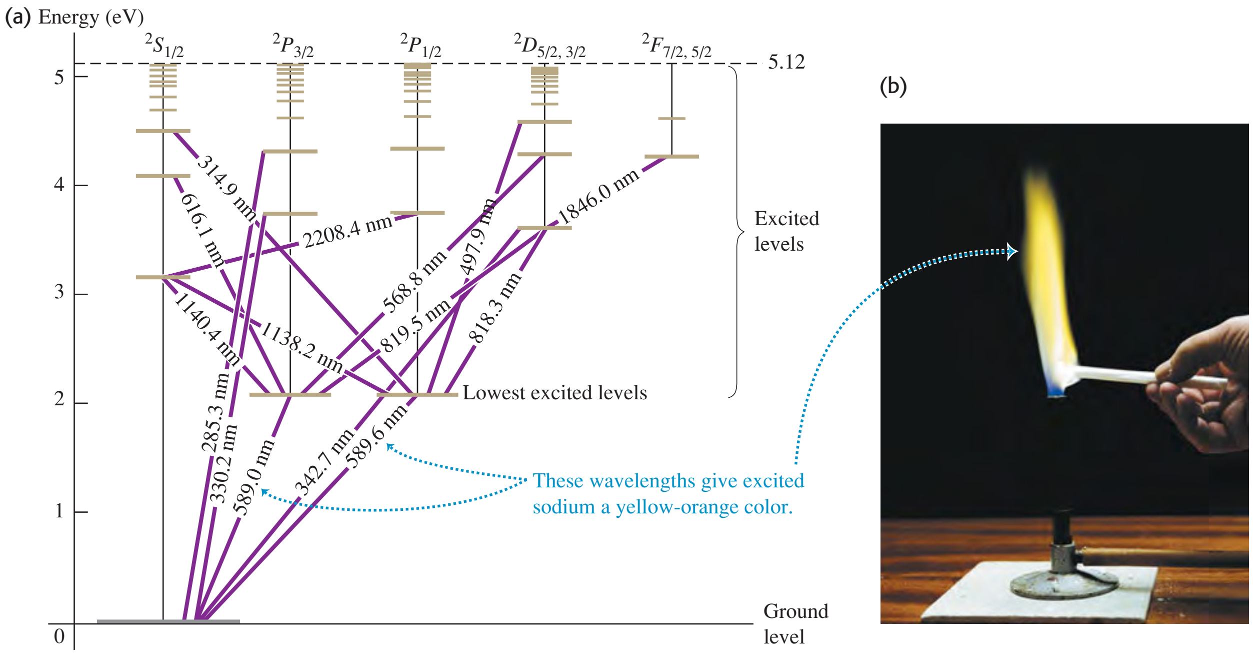
Note that a lithium atom *cannot* absorb a photon with a slightly longer wavelength (say, 672 nm) or one with a slightly shorter wavelength (say, 670 nm). That's because these photons have, respectively, slightly too little or slightly too much energy to raise the atom's energy from one level to the 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 shows some of the energy levels for a sodium atom. You may have noticed the yellow-orange light emitted by sodium

39.18 When a beam of white light with a continuous spectrum passes through a cool gas, the transmitted light has an absorption spectrum. The absorbed light energy excites the gas and causes it to emit light of its own, which has an emission spectrum.





39.19 (a) Energy levels of the sodium atom relative to the ground level. Numbers on the lines between levels are wavelengths of the light emitted or absorbed during transitions between those levels. The column labels, such as $^2S_{1/2}$, refer to some 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.

vapor street lights. Sodium atoms emit this characteristic yellow-orange light with wavelengths 589.0 and 589.6 nm when they make transitions from the two closely spaced levels labeled *lowest excited levels* to the ground level. A standard test for the presence of sodium compounds is to look for this yellow-orange light from a sample placed in a flame (Fig. 39.19b).

Example 39.5 Emission and absorption spectra

A hypothetical atom (Fig. 39.20a) has energy levels at 0.00 eV (the ground level), 1.00 eV, and 3.00 eV. (a) What are the frequencies and wavelengths of the spectral lines this atom can emit when excited? (b) What wavelengths can this atom absorb if it is in its ground level?

SOLUTION

IDENTIFY and SET UP: Energy is conserved when a photon is emitted or absorbed. In each transition the photon energy is equal to the difference between the energies of the levels involved in the transition.

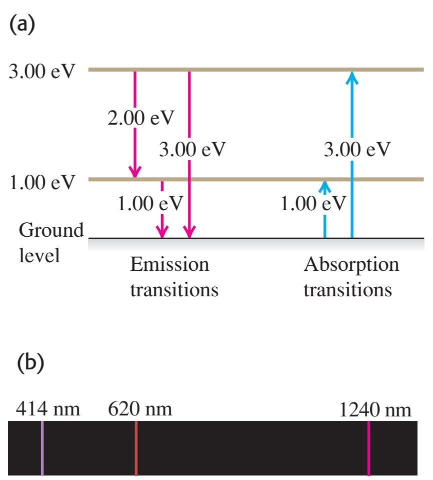
EXECUTE: (a) The possible energies of emitted photons are 1.00 eV, 2.00 eV, and 3.00 eV. For 1.00 eV, Eq. (39.2) gives

$$f = \frac{E}{h} = \frac{1.00\text{ eV}}{4.136 \times 10^{-15}\text{ eV}\cdot\text{s}} = 2.42 \times 10^{14}\text{ Hz}$$

For 2.00 eV and 3.00 eV, $f = 4.84 \times 10^{14}$ Hz and 7.25×10^{14} Hz, respectively. For 1.00-eV photons,

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8\text{ m/s}}{2.42 \times 10^{14}\text{ Hz}} = 1.24 \times 10^{-6}\text{ m} = 1240\text{ nm}$$

39.20 (a) Energy-level diagram for the hypothetical atom, showing the possible transitions for emission from excited levels and for absorption from the ground level. (b) Emission spectrum of this hypothetical atom.



Continued

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 state 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.

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).

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.

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.

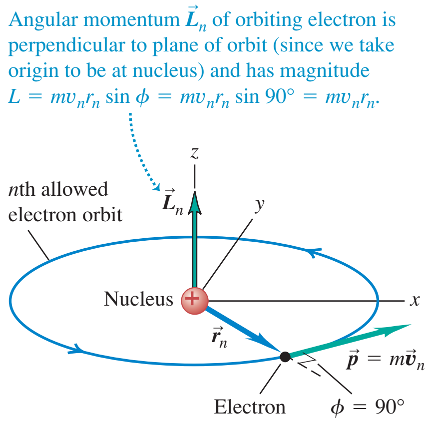
Electron Waves and the Bohr Model of Hydrogen

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. Let's look at the ideas behind the **Bohr model** of the hydrogen atom.

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 written as $\text{kg} \cdot \text{m}^2/\text{s}$.) Let's number the orbits by an integer n, where $n = 1, 2, 3, \ldots$, and call the radius of orbit n r_n and the speed of the electron in that orbit v_n. The value of n for each orbit is called the **principal quantum number** for the orbit. From Section 10.5, Eq. (10.28), the magnitude of the angular momentum of an electron of mass m in such an orbit is $L_n = mv_nr_n$ (Fig. 39.21). So Bohr's argument led to

$$L_n = mv_nr_n = n\frac{h}{2\pi} \qquad \text{(quantization of angular momentum)} \qquad (39.6)$$



39.21 Calculating the angular momentum of an electron in a circular orbit around an atomic nucleus.

Instead of going through Bohr's argument to justify Eq. (39.6), we can use de Broglie's picture of electron waves. Rather then visualizing the orbiting electron as a particle moving around the nucleus in a circular path, think of it as a sinusoidal *standing wave* with wavelength λ that extends around the circle. A standing wave on a string transmits no energy (see Section 15.7), and electrons in Bohr's orbits radiate no energy. For the wave to "come out even" and join onto itself smoothly, the circumference of this circle must include some *whole number* of wavelengths, as Fig. 39.22 suggests. Hence for an orbit with radius r_n and circumference $2\pi r_n$, we must have $2\pi r_n = n\lambda_n$, where λ_n is the wavelength and $n = 1, 2, 3, \ldots$. According to the de Broglie relationship, Eq. (39.1), the wavelength of a particle with rest mass m moving with nonrelativistic speed v_n is $\lambda_n = h/mv_n$. Combining $2\pi r_n = n\lambda_n$ and $\lambda_n = h/mv_n$, we find $2\pi r_n = nh/mv_n$ or

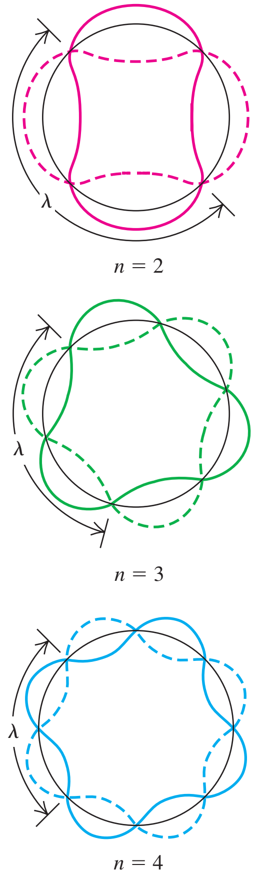
$$mv_nr_n = n\frac{h}{2\pi}$$

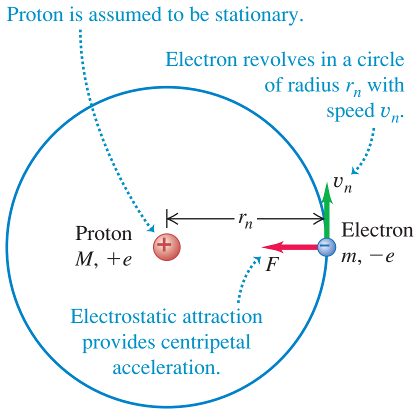
This is the same as Bohr's result, Eq. (39.6). Thus a wave picture of the electron leads naturally to the quantization of the electron's angular momentum.

Now let's consider a model of the hydrogen atom that is Newtonian in spirit but incorporates this quantization assumption (Fig. 39.23). This atom consists of a single electron with mass m and charge $-e$ in a circular orbit around a single proton with charge $+e$. The proton is nearly 2000 times as massive as the electron, so we can assume that the proton does not move. We learned in Section 5.4 that when a particle with mass m moves with speed v_n in a circular orbit with radius r_n, its centripetal (inward) acceleration is v_n^2/r_n. According to Newton's second law, a radially inward net force with magnitude $F = mv_n^2/r_n$ is needed to cause this acceleration. We discussed in Section 12.4 how the gravitational attraction provides that inward force for satellite orbits. In hydrogen the force F is provided by the electrical attraction between the positive proton and the negative electron. From Coulomb's law, Eq. (21.2),

$$F = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n^2}$$

39.22 These diagrams show the idea of fitting a standing electron wave around a circular orbit. For the wave to join onto itself smoothly, the circumference of the orbit must be an integral number n of wavelengths.





39.23 The Bohr model of the hydrogen atom.

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

$$r_n = \epsilon_0 \frac{n^2h^2}{\pi me^2} \quad \text{(orbital radii in the Bohr model)} \tag{39.8}$$

$$v_n = \frac{1}{\epsilon_0}\frac{e^2}{2nh} \quad \text{(orbital speeds in the Bohr model)} \tag{39.9}$$

Equation (39.8) shows that the orbit radius r_n is proportional to n^2, so the smallest orbit radius corresponds to $n = 1$. We'll denote this minimum radius, called the *Bohr radius,* as a_0:

$$a_0 = \epsilon_0 \frac{h^2}{\pi me^2} \quad \text{(Bohr radius)} \tag{39.10}$$

Then we can rewrite Eq. (39.8) as

$$r_n = n^2 a_0 \tag{39.11}$$

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 39.29); the result is $v_1 = 2.19 \times 10^6$ m/s. This is less than 1% of the speed of light, so relativistic considerations aren't significant.

Hydrogen Energy Levels in the Bohr Model

We can now use Eqs. (39.8) and (39.9) to find the kinetic and potential energies K_n and U_n when the electron is in the orbit with quantum number n:

$$K_n = \tfrac{1}{2}mv_n^{\,2} = \frac{1}{\epsilon_0^{\,2}}\frac{me^4}{8n^2h^2} \quad \text{(kinetic energies in the Bohr model)} \tag{39.12}$$

$$U_n = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_n} = -\frac{1}{\epsilon_0^{\,2}}\frac{me^4}{4n^2h^2} \quad \text{(potential energies in the Bohr model)} \tag{39.13}$$

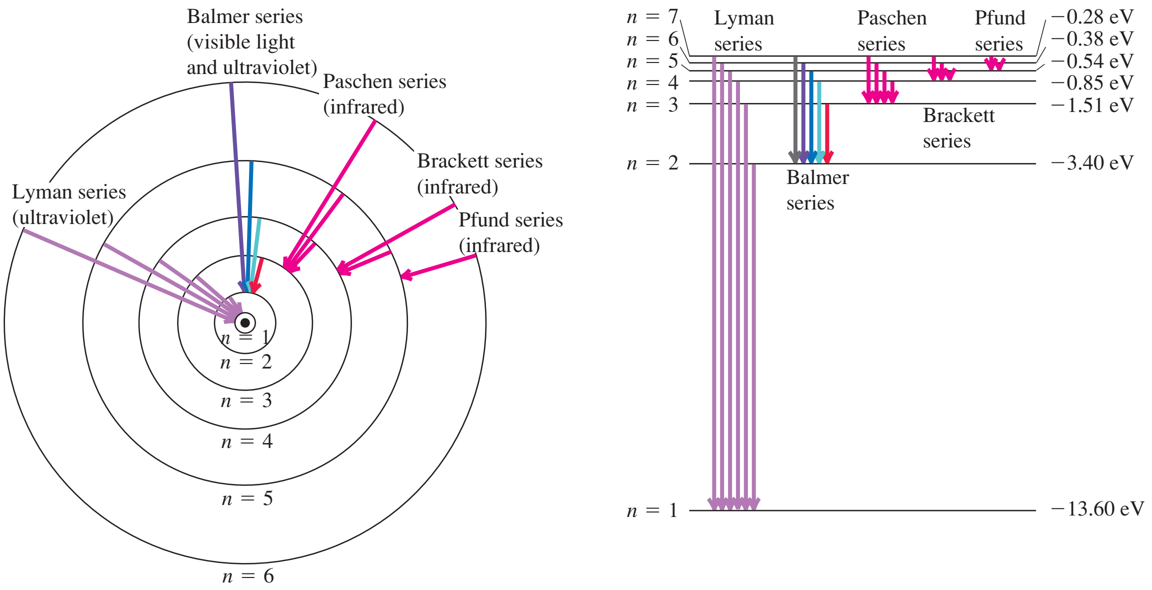
The potential energy has a negative sign because we have taken the electric potential energy to be zero when the electron is infinitely far from the nucleus. We are interested only in the *differences* in energy between orbits, so the reference position doesn't matter. The total energy E_n is the sum of the kinetic and potential energies:

$$E_n = K_n + U_n = -\frac{1}{\epsilon_0^{\,2}}\frac{me^4}{8n^2h^2} \quad \text{(total energies in the Bohr model)} \tag{39.14}$$

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



Since E_n in Eq. (39.14) has a different value for each n, you can see that this equation gives the *energy levels* of the hydrogen atom in the Bohr model. Each distinct orbit corresponds to a distinct energy level.

Figure 39.24 depicts the orbits and energy levels. We label the possible energy levels of the atom by values of the quantum number n. For each value of n there are corresponding values of orbit radius r_n, speed v_n, angular momentum $L_n = nh/2\pi$, and total energy E_n. The energy of the atom is least when $n = 1$ and E_n has its most negative value. This is the *ground level* of the hydrogen atom; it is the level with the smallest orbit, of radius a_0. For $n = 2, 3, \ldots,$ the absolute value of E_n is smaller and the energy is progressively larger (less negative).

Figure 39.24 also shows some of the possible transitions from one electron orbit to an orbit of lower energy. Consider a transition from orbit n_U (for "upper") to a smaller orbit n_L (for "lower"), with $n_L < n_U$—or, equivalently, from *level* n_U to a lower *level* n_L. Then the energy hc/λ of the emitted photon of wavelength λ is equal to $E_{n_U} - E_{n_L}$. Before we use this relationship to solve for λ, it's convenient to rewrite Eq. (39.14) for the energies as

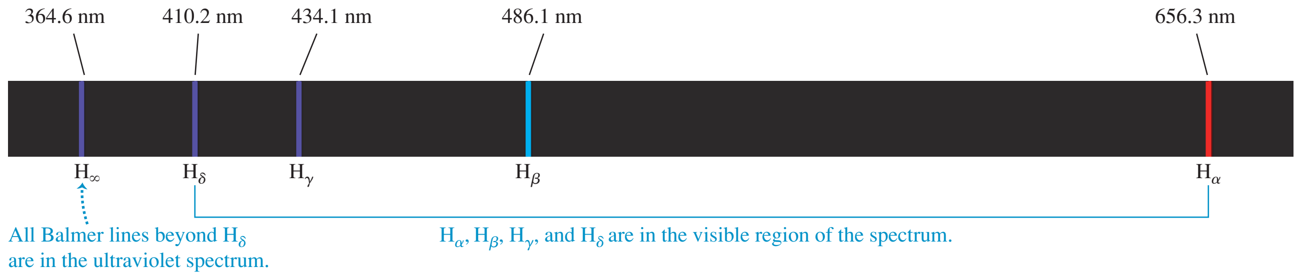
$$E_n = -\frac{hcR}{n^2}, \quad \text{where} \quad R = \frac{me^4}{8\epsilon_0^2 h^3 c} \quad \text{(total energies in the Bohr model)} \tag{39.15}$$

The quantity R in Eq. (39.15) is called the **Rydberg constant** (named for the Swedish physicist Johannes Rydberg, who did pioneering work on the hydrogen spectrum). When we substitute the numerical values of the fundamental physical constants m, c, e, h, and ϵ_0, all of which can be determined quite independently of the Bohr theory, we find that $R = 1.097 \times 10^7\ \text{m}^{-1}$. Now we solve for the wavelength of the photon emitted in a transition from level n_U to level n_L:

$$\frac{hc}{\lambda} = E_{n_U} - E_{n_L} = \left(-\frac{hcR}{n_U^2}\right) - \left(-\frac{hcR}{n_L^2}\right) = hcR\left(\frac{1}{n_L^2} - \frac{1}{n_U^2}\right)$$

$$\frac{1}{\lambda} = R\left(\frac{1}{n_L^2} - \frac{1}{n_U^2}\right) \quad \text{(hydrogen wavelengths in the Bohr model, } n_L < n_U\text{)} \tag{39.16}$$

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.



PhET: Neon Lights and Other Discharge Lamps
ActivPhysics 18.2: Spectroscopy

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) are all in the visible and ultraviolet parts of the electromagnetic spectrum. These wavelengths are collectively called the *Balmer series* (Fig. 39.25). 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 loses energy and a photon is emitted. But as we discussed previously, the wavelengths of those photons that an atom can *absorb* are the same as those that it can emit. For example, a hydrogen atom in the $n = 2$ level can absorb a 656.3-nm photon and end up in the $n = 3$ level.

One additional test of the Bohr model is its predicted value of the *ionization energy* of the hydrogen atom. This is the energy required to remove the electron completely from the atom. Ionization corresponds to a transition from the ground level $(n = 1)$ to an infinitely large orbit radius $(n = \infty)$, so the energy that must be added to the atom is $E_\infty - E_1 = 0 - E_1 = -E_1$ (recall that E_1 is negative).

Substituting the constants from Appendix F into Eq. (39.15) gives an ionization energy of 13.606 eV. The ionization energy can also be measured directly; the result is 13.60 eV. These two values agree within 0.1%.

Example 39.6 Exploring the Bohr model

Find the kinetic, potential, and total energies of the hydrogen atom in the first excited level, and find the wavelength of the photon emitted in a transition from that level to the ground level.

SOLUTION

IDENTIFY and SET UP: This problem uses the ideas of the Bohr model. We use simplified versions of Eqs. (39.12), (39.13), and (39.14) to find the energies of the atom, and Eq. (39.16), $hc/\lambda = E_{n_U} - E_{n_L}$, to find the photon wavelength λ in a transition from $n_U = 2$ (the first excited level) to $n_L = 1$ (the ground level).

EXECUTE: We could evaluate Eqs. (39.12), (39.13), and (39.14) for the nth level by substituting the values of m, e, ϵ_0, and h. But we can simplify the calculation by comparing with Eq. (39.15), which shows that the constant $me^4/8\epsilon_0^2h^2$ that appears in Eqs. (39.12), (39.13), and (39.14) is equal to hcR:

$$\begin{aligned}\frac{me^4}{8\epsilon_0^2h^2} &= hcR\\ &= (6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^8\ \text{m/s})\\ &\quad \times (1.097 \times 10^7\ \text{m}^{-1})\\ &= 2.179 \times 10^{-18}\ \text{J} = 13.60\ \text{eV}\end{aligned}$$

This allows us to rewrite Eqs. (39.12), (39.13), and (39.14) as

$$K_n = \frac{13.60\ \text{eV}}{n^2} \qquad U_n = \frac{-27.20\ \text{eV}}{n^2} \qquad E_n = \frac{-13.60\ \text{eV}}{n^2}$$

For the first excited level ($n = 2$), we have $K_2 = 3.40$ eV, $U_2 = -6.80$ eV, and $E_2 = -3.40$ eV. For the ground level ($n = 1$), $E_1 = -13.60$ eV. The energy of the emitted photon is then $E_2 - E_1 = -3.40\ \text{eV} - (-13.60\ \text{eV}) = 10.20$ eV, and

$$\begin{aligned}\lambda &= \frac{hc}{E_2 - E_1} = \frac{(4.136 \times 10^{-15}\ \text{eV}\cdot\text{s})(3.00 \times 10^8\ \text{m/s})}{10.20\ \text{eV}}\\ &= 1.22 \times 10^{-7}\ \text{m} = 122\ \text{nm}\end{aligned}$$

This is the wavelength of the Lyman-alpha (L_α) line, the longest-wavelength line in the Lyman series of ultraviolet lines in the hydrogen spectrum (see Fig. 39.24).

EVALUATE: The total mechanical energy for any level is negative and is equal to one-half the potential energy. We found the same energy relationship for Newtonian satellite orbits in Section 12.4. The situations are similar because both the electrostatic and gravitational forces are inversely proportional to $1/r^2$.

Nuclear Motion and the Reduced Mass of an Atom

The Bohr model is so successful that we can justifiably ask why its predictions for the wavelengths and ionization energy of hydrogen differ from the measured values by about 0.1%. The explanation is that we assumed that the nucleus (a proton) remains at rest. However, as Fig. 39.26 shows, the proton and electron *both* revolve in circular orbits about their common center of mass (see Section 8.5). It turns out that we can take this motion into account very simply by using in Bohr's equations not the electron rest mass m but a quantity called the **reduced mass** m_r of the system. For a system composed of two bodies of masses m_1 and m_2, the reduced mass is

$$m_r = \frac{m_1 m_2}{m_1 + m_2} \tag{39.17}$$

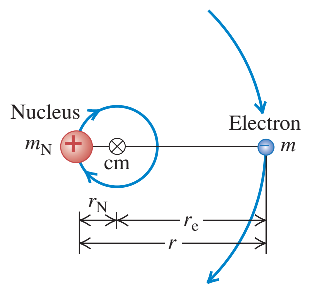
For ordinary hydrogen we let m_1 equal m and m_2 equal the proton mass, $m_p = 1836.2m$. Thus the proton–electron system of ordinary hydrogen has a reduced mass of

$$m_r = \frac{m(1836.2m)}{m + 1836.2m} = 0.99946m$$

When this value is used instead of the electron mass m in the Bohr equations, the predicted values agree very well with the measured values.

In an atom of deuterium, also called *heavy hydrogen,* the nucleus is not a single proton but a proton and a neutron bound together to form a composite body called the *deuteron.* The reduced mass of the deuterium atom turns out to be $0.99973m$. Equations (39.15) and (39.16) (with m replaced by m_r) show that all wavelengths are inversely proportional to m_r. Thus the wavelengths

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 $r_e/1836.2$.



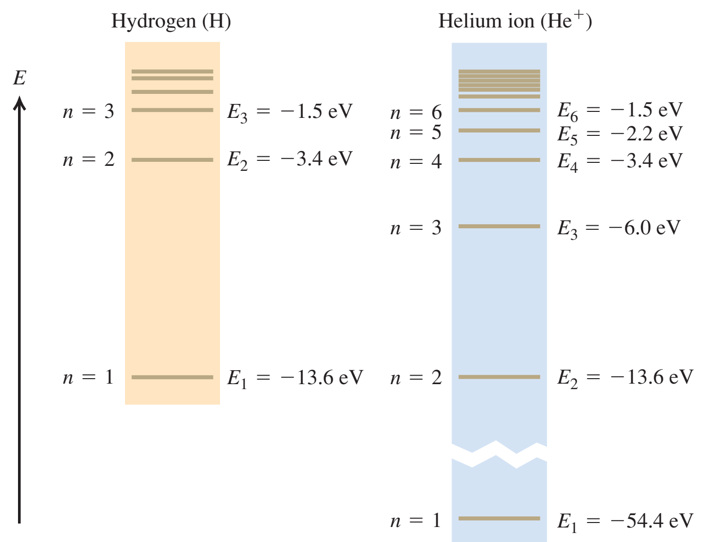
of the deuterium spectrum should be those of hydrogen divided by $(0.99973m)/(0.99946m) = 1.00027$. This is a small effect but well within the precision of modern spectrometers. This small wavelength shift led the American scientist Harold Urey to the discovery of deuterium in 1932, an achievement that 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. In particular, 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. We invite you to verify these statements. 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 $n = 1000$. In accordance with 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 same size as a small grain of sand.

Although the Bohr model predicted the energy levels of the hydrogen atom correctly, it raised as many questions as it answered. It combined elements of classical physics with new postulates that were inconsistent with classical ideas. The model provided no insight into what happens during a transition from one orbit to another; the angular speeds of the electron motion were not in general the angular frequencies of the emitted radiation, a result that is contrary to classical electrodynamics. Attempts to extend the model to atoms with two or more electrons were not successful. An electron moving in one of Bohr's circular orbits forms a current loop and should produce a magnetic dipole moment (see Section 27.7). However, a hydrogen atom in its ground level has *no* magnetic moment due to orbital motion. In Chapters 40 and 41 we will find that an even more radical departure from classical concepts was needed before the understanding of atomic structure could progress further.



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.

39.4 The Laser

MasteringPHYSICS®
PhET: Lasers

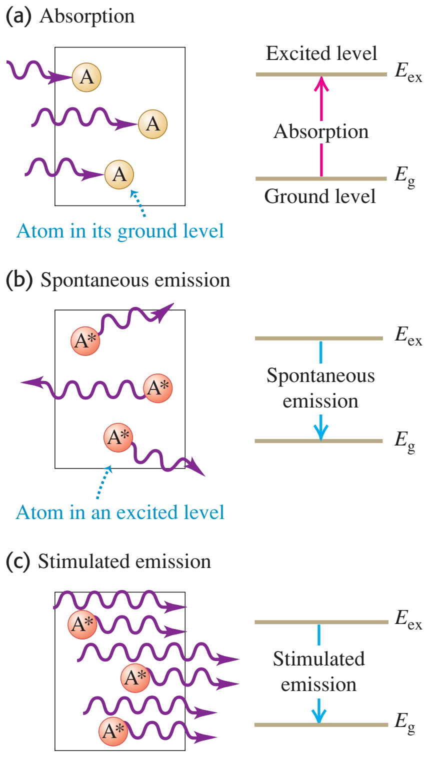
The **laser** is a light source that produces a beam of highly coherent and very nearly monochromatic light as a result of cooperative emission from many atoms. The name "laser" is an acronym for "light amplification by stimulated emission of radiation." We can understand the principles of laser operation from what we have learned about atomic energy levels and photons. To do this we'll have to introduce two new concepts: *stimulated emission* and *population inversion.*

Spontaneous and Stimulated Emission

Consider a gas of atoms in a transparent container. Each atom is initially in its ground level of energy E_g and also has an excited level of energy E_{ex}. If we shine light of frequency f on the container, an atom can absorb one of the photons provided the photon energy $E = hf$ equals the energy difference $E_{ex} - E_g$ between the levels. Figure 39.28a shows this process, in which three atoms A each absorb a photon and go into the excited level. Some time later, the excited atoms (which we denote as A*) return to the ground level by each emitting a photon with the same frequency as the one originally absorbed (Fig. 39.28b). This process is called **spontaneous emission.** The direction and phase of the spontaneously emitted photons 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.

39.28 Three processes in which atoms interact with light.



To discuss stimulated emission from atoms in excited levels, we need to know something about how many atoms are in each of the various energy levels. First, we need to make the distinction between the terms *energy level* and *state.* A system may have more than one way to attain a given energy level; each different way is a different **state.** For instance, there are two ways of putting an ideal unstretched spring in a given energy level. Remembering that the spring potential energy is $U = \frac{1}{2}kx^2$, we could compress the spring by $x = -b$ or we could stretch it by $x = +b$ to get the same $U = \frac{1}{2}kb^2$. The Bohr model had only one state in each energy level, but we will find in Chapter 41 that the hydrogen atom (Fig. 39.24b) actually has two states in its -13.60-eV ground level, eight states in its -3.40-eV first exited 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 Boltzmann's 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, as we should expect. 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 (the temperature of the filament in an incandescent light bulb),

$$\frac{E_{ex} - E_g}{kT} = \frac{3.2 \times 10^{-19}\text{ J}}{(1.38 \times 10^{-23}\text{ J/K})(3000\text{ K})} = 7.73$$

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.

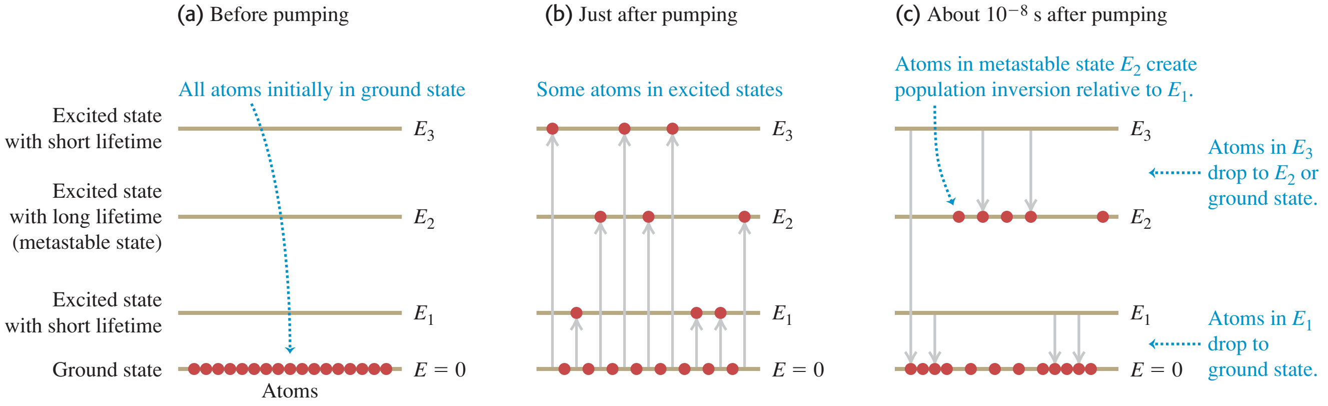
Enhancing Stimulated Emission: Population Inversions

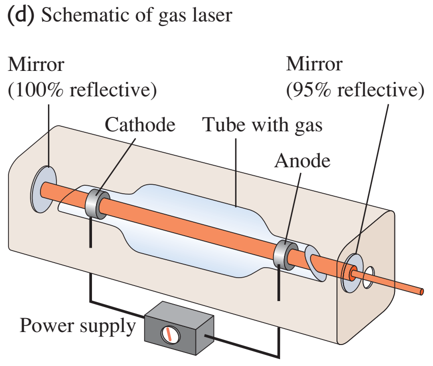
To make a laser, we need to promote stimulated emission by increasing the number of atoms in excited states. Can we do that simply by illuminating the container with radiation of frequency $f = E/h$ corresponding to the energy difference $E = E_{ex} - E_g$, as in Fig. 39.28a? Some of the atoms absorb 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 the number of atoms in a higher-energy state is greater than the number 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. It turns out that we can achieve a population inversion by starting with atoms that have the right kinds of excited states. Figure 39.29a shows an energy-level diagram for such an atom with a ground state and *three* excited states of energies E_1, E_2, and E_3. A laser that uses a material with energy levels like these is called a *four-level laser*. For the laser action to work, the states of energies E_1 and E_3 must have ordinary short lifetimes of about 10^{-8} s, while the state of energy E_2 must have an unusually long lifetime of 10^{-3} s or so. Such a long-lived **metastable state** can occur if, for instance, there are restrictions imposed by conservation of angular momentum that hinder photon emission from this state. (We'll discuss these restrictions in Chapter 41.) The metastable state is the one that we want to populate.

To produce a population inversion, we *pump* the material to excite the atoms out of the ground state into the states of energies E_1, E_2, and E_3 (Fig. 39.29b). If the atoms are in a gas, this pumping can be done by inserting two electrodes into the gas container. When a burst of sufficiently high voltage is applied to the electrodes, an electric discharge occurs. Collisions between ionized atoms and electrons carrying the discharge current then excite the atoms to various energy states. Within about 10^{-8} s the atoms that are excited to states E_1 and E_3 undergo spontaneous photon emission, so these states end up depopulated. But atoms "pile up" in the metastable state with energy E_2. The number of atoms in the metastable state is *less* than the number in the ground state, but is *much greater* than in the nearly unoccupied state of energy E_1. Hence there is a population inversion of state E_2 relative to state E_1 (Fig. 39.29c). You can see why we need

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.





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 21.2), 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.

MasteringPHYSICS

PhET: Blackbody Spectrum
PhET: The Greenhouse Effect

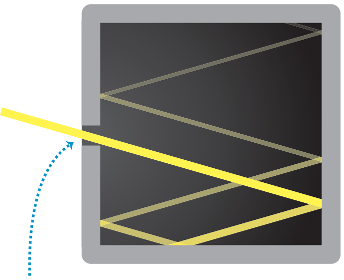
39.5 Continuous Spectra

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.

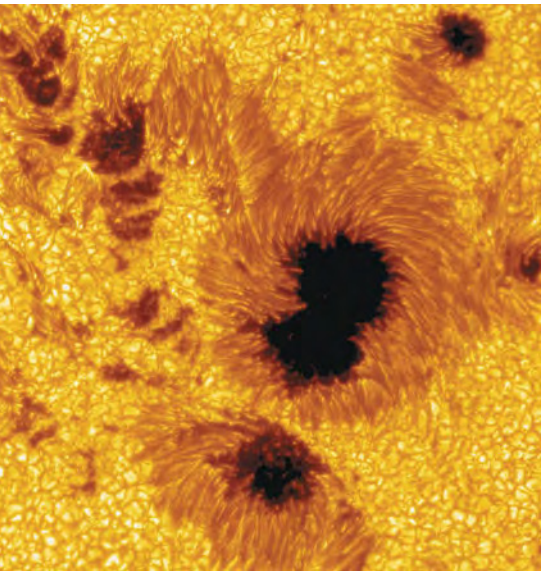
39.30 A hollow box with a small aperture behaves like a blackbody. When the box is heated, the electromagnetic radiation that emerges from the aperture has a blackbody spectrum.



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.

39.31 This close-up view of the sun's surface shows two dark sunspots. Their temperature is about 4000 K, while the surrounding solar material is at $T =$ 5800 K. From the Stefan–Boltzmann law, the intensity from a given area of sunspot is only $(4000\text{ K}/5800\text{ K})^4 = 0.23$ as great as the intensity from the same area of the surrounding material—which is why sunspots appear dark.



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). We studied this relationship in Section 17.7 during our study of heat-transfer mechanisms. This total intensity I emitted at absolute temperature T is given by the **Stefan–Boltzmann law:**

$$I = \sigma T^4 \qquad \text{(Stefan–Boltzmann law for a blackbody)} \tag{39.19}$$

where σ is a fundamental physical constant called the *Stefan–Boltzmann constant.* In SI units,

$$\sigma = 5.670400(40) \times 10^{-8} \frac{\text{W}}{\text{m}^2 \cdot \text{K}^4}$$

Second, the intensity is not uniformly distributed over all wavelengths. Its distribution can be measured and described by the intensity per wavelength interval $I(\lambda)$, called the *spectral emittance.* Thus $I(\lambda)\, d\lambda$ is the intensity corresponding to wavelengths in the interval from λ to $\lambda + d\lambda$. The *total* intensity I, given by Eq. (39.19), is the *integral* of the distribution function $I(\lambda)$ over all wavelengths, which equals the area under the $I(\lambda)$ versus λ curve:

$$I = \int_0^\infty I(\lambda)\, d\lambda \tag{39.20}$$

CAUTION **Spectral emittance vs. intensity** Although we use the symbol $I(\lambda)$ for spectral emittance, keep in mind that spectral emittance is *not* the same thing as intensity I. Intensity is power per unit area, with units W/m^2. Spectral emittance is power per unit area *per unit wavelength interval*, with units W/m^3.

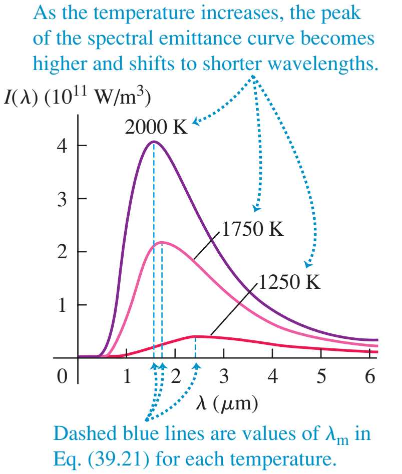
Figure 39.32 shows the measured spectral emittances $I(\lambda)$ for blackbody radiation at three different temperatures. Each has a peak wavelength λ_m at which the emitted intensity per wavelength interval is largest. Experiment shows that λ_m is inversely proportional to T, so their product is constant. This observation is called the **Wien displacement law.** The experimental value of the constant is 2.90×10^{-3} m·K:

$$\lambda_m T = 2.90 \times 10^{-3}\ \mathrm{m \cdot K} \qquad \text{(Wien displacement law)} \qquad (39.21)$$

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 a body that glows yellow is hotter and brighter than one of the same size that glows red.

Third, experiments show that the *shape* of the distribution function is the same for all temperatures. We can make a curve for one temperature fit any other temperature by simply changing the scales on the graph.

39.32 These graphs show the spectral emittance $I(\lambda)$ for radiation from a blackbody at three different temperatures.



Rayleigh and the "Ultraviolet Catastrophe"

During the last decade of the 19th century, many attempts were made to derive these empirical results about blackbody radiation from basic principles. In one attempt, the English physicist Lord Rayleigh considered the light enclosed within a rectangular box like that shown in Fig. 39.30. Such a box, he reasoned, has a series of possible *normal modes* for electromagnetic waves, as we discussed in Section 32.5. It also seemed reasonable to assume that the distribution of energy among the various modes would be given by the equipartition principle (see Section 18.4), which had been used successfully in the analysis of heat capacities.

Including both the electric- and magnetic-field energies, Rayleigh assumed that the total energy of each normal mode was equal to kT. Then by computing the *number* of normal modes corresponding to a wavelength interval $d\lambda$, Rayleigh calculated the expected distribution of wavelengths in the radiation within the box. Finally, he computed the predicted intensity distribution $I(\lambda)$ for the radiation emerging from the hole. His result was quite simple:

$$I(\lambda) = \frac{2\pi ckT}{\lambda^4} \qquad \text{(Rayleigh's calculation)} \qquad (39.22)$$

At large wavelengths this formula agrees quite well with the experimental results shown in Fig. 39.32, but there is serious disagreement at small wavelengths. The experimental curves in Fig. 39.32 fall toward zero at small λ. By contrast, Rayleigh's prediction in Eq. (39.22) goes in the opposite direction, approaching infinity at $1/\lambda^4$, 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. Planck assumed 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 turned out to be 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 to explain the photoelectric effect (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.

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.

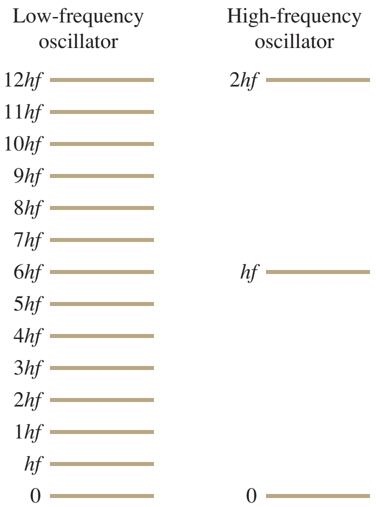


Figure 39.33 shows energy-level diagrams for two of the oscillators that Planck envisioned in the walls of the rectangular box, one with a low frequency and the other with a high frequency. The spacing in energy between adjacent levels is hf. This spacing is small for the low-frequency oscillator that 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 of wavelength $\lambda = 3.00\ \mu\text{m}$, we can show that $hf = hc/\lambda = 0.413$ eV; for a higher-frequency oscillator that emits photons of wavelength $\lambda = 0.500\ \mu\text{m}$, $hf = hc/\lambda = 2.48$ eV. For these two cases Eq. (39.23) gives

$$\frac{n_1}{n_0} = e^{-hf/kT} = 0.0909 \text{ for } \lambda = 3.00\ \mu\text{m}$$

$$\frac{n_1}{n_0} = e^{-hf/kT} = 5.64 \times 10^{-7} \text{ for } \lambda = 0.500\ \mu\text{m}$$

The value for $\lambda = 3.00\ \mu\text{m}$ 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:

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5(e^{hc/\lambda kT} - 1)} \qquad \text{(Planck radiation law)} \tag{39.24}$$

where h is Planck's constant, c is the speed of light, k is Boltzmann's constant, T is the absolute temperature, and λ is the wavelength. 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

$$\lambda_{\text{m}} = \frac{hc}{4.965kT} \tag{39.25}$$

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 39.67). 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 as a combination of other fundamental constants:

$$\sigma = \frac{2\pi^5 k^4}{15c^2h^3} \tag{39.28}$$

You should substitute the values of k, c, and h from Appendix F and verify that you obtain the value $\sigma = 5.6704 \times 10^{-8}\ \text{W/m}^2 \cdot \text{K}^4$ 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 at large values 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 $x = 1$). 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

To a good approximation, the sun's surface is a blackbody with a surface temperature of 5800 K. (We are ignoring the absorption produced by the sun's atmosphere, shown in Fig. 39.9.) (a) At what wavelength does the sun emit most strongly? (b) What is the total radiated power per unit surface area?

SOLUTION

IDENTIFY and SET UP: Our target variables are the peak-intensity wavelength λ_m and the radiated power per area I. Hence we'll use the Wien displacement law, Eq. (39.21) (which relates λ_m to the blackbody temperature T), and the Stefan–Boltzmann law, Eq. (39.19) (which relates I to T).

EXECUTE: (a) From Eq. (39.21),

$$\lambda_m = \frac{2.90 \times 10^{-3}\ \text{m}\cdot\text{K}}{T} = \frac{2.90 \times 10^{-3}\ \text{m}\cdot\text{K}}{5800\ \text{K}}$$
$$= 0.500 \times 10^{-6}\ \text{m} = 500\ \text{nm}$$

(b) From Eq. (39.19),

$$I = \sigma T^4 = (5.67 \times 10^{-8}\ \text{W/m}^2\cdot\text{K}^4)(5800\ \text{K})^4$$
$$= 6.42 \times 10^7\ \text{W/m}^2 = 64.2\ \text{MW/m}^2$$

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 $I = 64.2\ \text{MW/m}^2$ found in part (b) is the intensity at the *surface* of the sun, 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 $[(6.96 \times 10^8\ \text{m})/(1.50 \times 10^{11}\ \text{m})]^2 = 2.15 \times 10^{-5}$ to the still-impressive $1.4\ \text{kW/m}^2$.

Example 39.8 A slice of sunlight

Find the power per unit area radiated from the sun's surface in the wavelength range 600.0 to 605.0 nm.

SOLUTION

IDENTIFY and SET UP: This question concerns the power emitted by a blackbody over a narrow range of wavelengths, and so involves the spectral emittance $I(\lambda)$ given by the Planck radiation law, Eq. (39.24). This requires that we find the area under the $I(\lambda)$ curve between 600.0 and 605.0 nm. We'll *approximate* this area as the product of the height of the curve at the median wavelength $\lambda = 602.5$ nm and the width of the interval, $\Delta\lambda = 5.0$ nm. From Example 39.7, $T = 5800$ K.

EXECUTE: To obtain the height of the $I(\lambda)$ curve at $\lambda = 602.5\ \text{nm} = 6.025 \times 10^{-7}$ m, we first evaluate the quantity $hc/\lambda kT$ in Eq. (39.24) and then substitute the result into Eq. (39.24):

$$\frac{hc}{\lambda kT} = \frac{(6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^8\ \text{m/s})}{(6.025 \times 10^{-7}\ \text{m})(1.381 \times 10^{-23}\ \text{J/K})(5800\ \text{K})} = 4.116$$

$$I(\lambda) = \frac{2\pi(6.626 \times 10^{-34}\ \text{J}\cdot\text{s})(2.998 \times 10^8\ \text{m/s})^2}{(6.025 \times 10^{-7}\ \text{m})^5(e^{4.116} - 1)}$$
$$= 7.81 \times 10^{13}\ \text{W/m}^3$$

The intensity in the 5.0-nm range from 600.0 to 605.0 nm is then approximately

$$I(\lambda)\Delta\lambda = (7.81 \times 10^{13}\ \text{W/m}^3)(5.0 \times 10^{-9}\ \text{m})$$
$$= 3.9 \times 10^5\ \text{W/m}^2 = 0.39\ \text{MW/m}^2$$

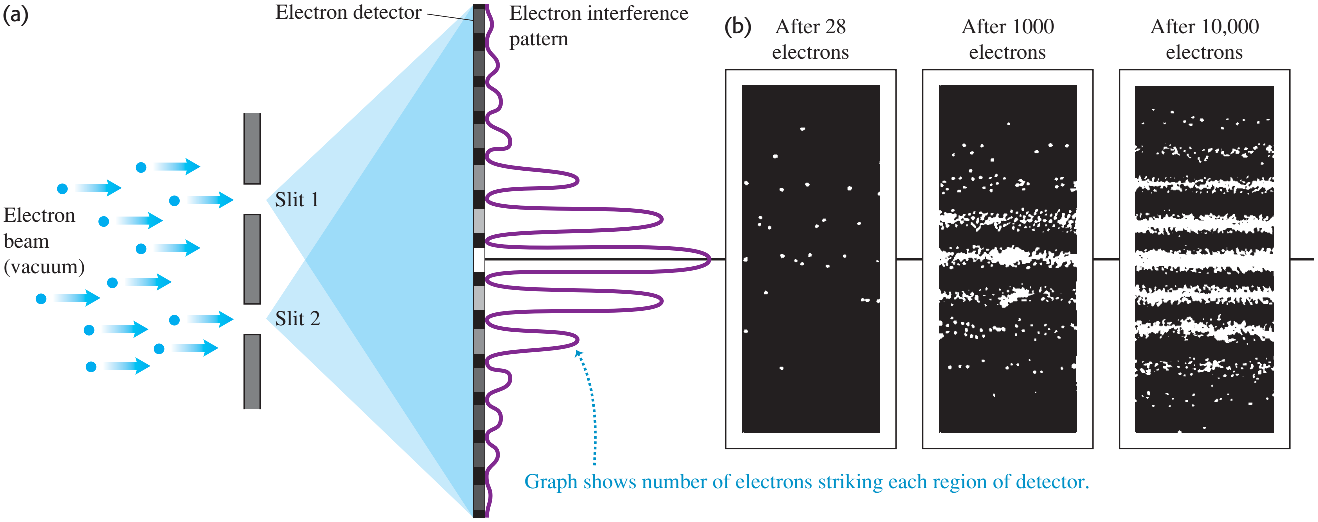
EVALUATE: In part (b) of Example 39.7, we found the power radiated per unit area by the sun at *all* wavelengths to be $I = 64.2\ \text{MW/m}^2$; here we have found that the power radiated per unit area in the wavelength range from 600 to 605 nm is $I(\lambda)\Delta\lambda = 0.39\ \text{MW/m}^2$, about 0.6% of the total.

Test Your Understanding of Section 39.5 (a) Does a blackbody at 2000 K emit x rays? (b) Does it emit radio waves?

39.6 The Uncertainty Principle Revisited

The discovery of the dual wave–particle nature of matter forces us to reevaluate the kinematic language we use to describe the position and motion of a particle. In classical Newtonian mechanics we think of a particle as a point. We can describe its location and state of motion at any instant with three spatial coordinates and three components of velocity. But because matter also has a wave aspect, when we look at the behavior on a small enough scale—comparable to the de Broglie wavelength of the particle—we can no longer use the Newtonian description. Certainly no Newtonian particle would undergo diffraction like electrons do (Section 39.1).

39.34 (a) A two-slit interference experiment for electrons. (b) The interference pattern after 28, 1000, and 10,000 electrons.



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, just as we did for light in Section 38.4. (The electron experiment has to be done in vacuum so that the electrons don't collide with air molecules.) What kind of pattern appears on the detector on the other side of the slits? The answer is: *exactly the same* kind of interference pattern we saw for photons in Section 38.4! Moreover, the principle of complementarity, which we introduced in Section 38.4, tells us that we cannot apply the wave and particle models simultaneously to describe any single element of this experiment. Thus we *cannot* predict exactly where in the pattern (a wave phenomenon) any individual electron (a particle) will land. We can't even ask which slit an individual electron passes through. If we tried to look at where the electrons were going by shining a light on them—that is, by scattering photons off them—the electrons would recoil, which would modify their motions so that the two-slit interference pattern would not appear.

CAUTION **Electron two-slit interference is not interference between two electrons** It's a common misconception that the pattern in Fig. 39.34b is due to the interference between *two* electron waves, each representing an electron passing through one slit. To show that this cannot be the case, we can send just one electron at a time through the apparatus. It makes no difference; we end up with the same interference pattern. In a sense, each electron wave interferes with itself.

The Heisenberg Uncertainty Principles for Matter

Just as electrons and photons show the same behavior in a two-slit interference experiment, electrons and other forms of matter obey the same Heisenberg uncertainty principles as photons do:

$$\begin{aligned} \Delta x \Delta p_x &\geq \hbar/2 \\ \Delta y \Delta p_y &\geq \hbar/2 \\ \Delta z \Delta p_z &\geq \hbar/2 \end{aligned} \quad \text{(Heisenberg uncertainty principle for position and momentum)} \tag{39.29}$$

$$\Delta t\, \Delta E \geq \hbar/2 \quad \text{(Heisenberg uncertainty principle for energy and time interval)} \tag{39.30}$$

In these equations $\hbar = h/2\pi = 1.055 \times 10^{-34}\,\mathrm{J \cdot s}$. 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.

Example 39.9 The uncertainty principle: position and momentum

An electron is confined within a region of width 5.000×10^{-11} m (roughly the Bohr radius). (a) Estimate the minimum uncertainty in the x-component of the electron's momentum. (b) What is the kinetic energy of an electron with this magnitude of momentum? Express your answer in both joules and electron volts.

SOLUTION

IDENTIFY and SET UP: This problem uses the Heisenberg uncertainty principle for position and momentum and the relationship between a particle's momentum and its kinetic energy. The electron could be anywhere within the region, so we take $\Delta x = 5.000 \times 10^{-11}$ m as its position uncertainty. We then find the momentum uncertainty Δp_x using Eq. (39.29) and the kinetic energy using 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 $\hbar$. Hence

$$\Delta p_x = \frac{\hbar}{2\Delta x} = \frac{1.055 \times 10^{-34}\,\mathrm{J \cdot s}}{2(5.000 \times 10^{-11}\,\mathrm{m})} = 1.055 \times 10^{-24}\,\mathrm{J \cdot s/m}$$
$$= 1.055 \times 10^{-24}\,\mathrm{kg \cdot m/s}$$

(b) We can rewrite the nonrelativistic expression for kinetic energy as

$$K = \tfrac{1}{2}mv^2 = \frac{(mv)^2}{2m} = \frac{p^2}{2m}$$

Hence an electron with a magnitude of momentum equal to Δp_x from part (a) has kinetic energy

$$K = \frac{p^2}{2m} = \frac{(1.055 \times 10^{-24}\,\mathrm{kg \cdot m/s})^2}{2(9.11 \times 10^{-31}\,\mathrm{kg})}$$
$$= 6.11 \times 10^{-19}\,\mathrm{J} = 3.81\,\mathrm{eV}$$

EVALUATE: This energy is typical of electron energies in atoms. This agreement suggests that the uncertainty principle is deeply involved in atomic structure.

A similar calculation explains why electrons in atoms do not fall into the nucleus. If an electron were confined to the interior of a nucleus, its position uncertainty would be $\Delta x \approx 10^{-14}$ m. This would give the electron a momentum uncertainty about 5000 times greater than that of the electron in this example, and a kinetic energy so great that the electron would immediately be ejected from the nucleus.

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 level, emitting a photon with wavelength 589.0 nm and energy 2.105 eV. What is the uncertainty in energy of that excited state? What is the wavelength spread of the corresponding spectral line?

SOLUTION

IDENTIFY and SET UP: This problem uses 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 level by replacing the $\geq$ sign in Eq. (39.30) with an equals sign and solving for ΔE.

EXECUTE: From Eq. (39.30),

$$\Delta E = \frac{\hbar}{2\Delta t} = \frac{1.055 \times 10^{-34}\,\mathrm{J \cdot s}}{2(1.6 \times 10^{-8}\,\mathrm{s})}$$
$$= 3.3 \times 10^{-27}\,\mathrm{J} = 2.1 \times 10^{-8}\,\mathrm{eV}$$

The atom remains in the ground level indefinitely, so that level has *no* associated energy uncertainty. The fractional uncertainty of the *photon* energy is therefore

$$\frac{\Delta E}{E} = \frac{2.1 \times 10^{-8}\,\mathrm{eV}}{2.105\,\mathrm{eV}} = 1.0 \times 10^{-8}$$

You can use some simple calculus and the relation $E = hc/\lambda$ 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

$$\Delta\lambda = \lambda\frac{\Delta E}{E} = (589.0\,\mathrm{nm})(1.0 \times 10^{-8}) = 0.0000059\,\mathrm{nm}$$

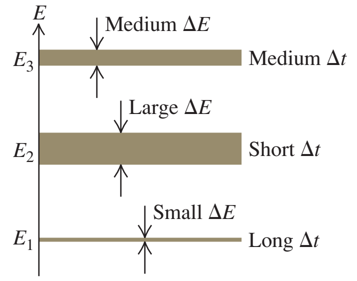
EVALUATE: This irreducible uncertainty $\Delta\lambda$ is called the *natural line width* of this particular spectral line. Though very small, it is within the limits of resolution of present-day spectrometers. Ordinarily, the natural line width is much smaller than the line width arising from other causes such as the Doppler effect and collisions among the rapidly moving atoms.

The Uncertainty Principle and the Limits of the Bohr Model

We saw in Section 39.3 that the Bohr model of the hydrogen atom was tremendously successful. However, the Heisenberg uncertainty principle for position and momentum shows that this model *cannot* be a correct description of how an electron in an atom behaves. Figure 39.22 shows that in the Bohr model as interpreted by de Broglie, an electron wave moves in a plane around the nucleus. Let's call this the xy-plane, so the z-axis is perpendicular to the plane. Hence the Bohr model says that an electron is always found at $z = 0$, and its z-momentum p_z is always zero (the electron does not move out of the xy-plane). But this implies that there are *no* uncertainties in either z or p_z, so $\Delta z = 0$ and $\Delta p_z = 0$. This directly contradicts Eq. (39.29), which says that the product $\Delta z \Delta p_z$ must be greater than or equal to $\hbar$.

This conclusion isn't too surprising, since the electron in the Bohr model is a mix of particle and wave ideas (the electron moves in an orbit like a miniature planet, but has a wavelength). To get an accurate picture of how electrons behave inside an atom and elsewhere, we need a description that is based *entirely* on the electron's wave properties. Our goal in Chapter 40 will be to develop this description, which we call *quantum mechanics*. To do this we'll introduce the *Schrödinger equation*, the fundamental equation that describes the dynamics of matter waves. This equation, as we will see, is as fundamental to quantum mechanics as Newton's laws are to classical mechanics or as Maxwell's equations are to electromagnetism.

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).



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.

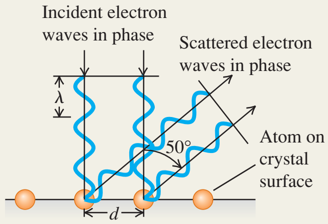
SUMMARY

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.)

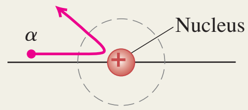
$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{39.1}$$

$$E = hf \tag{39.2}$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2meV_{ba}}} \tag{39.3}$$

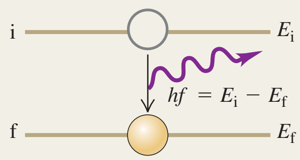


The nuclear atom: The Rutherford scattering experiments show that most of an atom's mass and all of its positive charge are concentrated in a tiny, dense nucleus at the center of the atom. (See Example 39.4.)



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.)

$$hf = \frac{hc}{\lambda} = E_\mathrm{i} - E_\mathrm{f} \tag{39.5}$$



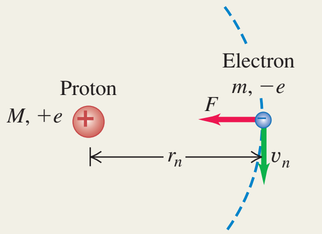
The Bohr model: In the Bohr model of the hydrogen atom, the permitted values of angular momentum are integral multiples of $h/2\pi$. The integer multiplier n is called the principal quantum number for the level. The orbital radii are proportional to n^2 and the orbital speeds are proportional to $1/n$. The energy levels of the hydrogen atom are given by Eq. (39.15), where R is the Rydberg constant. (See Example 39.6.)

$$L_n = mv_n r_n = n\frac{h}{2\pi}$$
$$(n = 1, 2, 3, \ldots) \tag{39.6}$$

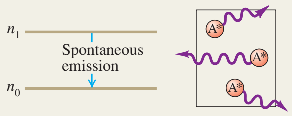
$$r_n = \epsilon_0 \frac{n^2h^2}{\pi m e^2} = n^2 a_0 \tag{39.8}$$
$$= n^2(5.29 \times 10^{-11}\ \mathrm{m}) \tag{39.10}$$

$$v_n = \frac{1}{\epsilon_0}\frac{e^2}{2nh} = \frac{2.19 \times 10^6\ \mathrm{m/s}}{n} \tag{39.9}$$

$$E_n = -\frac{hcR}{n^2} = -\frac{13.60\ \mathrm{eV}}{n^2} \tag{39.15}$$
$$(n = 1, 2, 3, \ldots)$$



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.

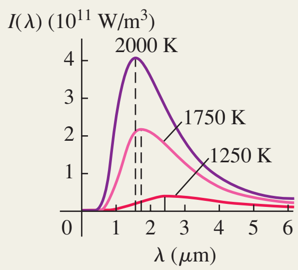


Blackbody radiation: The total radiated intensity (average power radiated per area) from a blackbody surface is proportional to the fourth power of the absolute temperature T. The quantity $\sigma = 5.67 \times 10^{-8}\ \mathrm{W/m^2 \cdot K^4}$ is called the Stefan–Boltzmann constant. The wavelength 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)

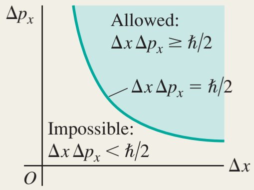
$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5(e^{hc/\lambda kT} - 1)}$$
(Planck radiation law) (39.24)



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). (See Examples 39.9 and 39.10.)

$$\Delta x \Delta p_x \geq \hbar/2$$
$$\Delta y \Delta p_y \geq \hbar/2 \qquad \text{(Heisenberg uncertainty principle for position and momentum)} \qquad (39.29)$$
$$\Delta z \Delta p_z \geq \hbar/2$$

$$\Delta t\, \Delta E \geq \hbar/2 \qquad \text{(Heisenberg uncertainty principle for energy and time interval)} \qquad (39.30)$$



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?

SOLUTION GUIDE

See MasteringPhysics® study area for a Video Tutor solution.

39.36 The Rosette Nebula.

IDENTIFY and SET UP

1. To solve this problem you need to use your knowledge of both blackbody radiation (Section 39.5) and the Bohr model of the hydrogen atom (Section 39.3).
2. In part (a) the target variable is the wavelength at which the star emits most strongly; in part (b) the target variable is a principal quantum number, and in part (c) it is the de Broglie wavelength of an electron in the $n = 2$ and $n = 3$ Bohr orbits (see Fig. 39.24). Select the equations you will need to find the target variables. (*Hint:* In Section 39.5 you learned how to find the energy change involved in a transition between two given levels of a hydrogen atom. Part (b) is a variation on this: You are to find the final level in a transition that starts in the $n = 1$ level and involves the absorption of a photon of a given wavelength and hence a given energy.)

EXECUTE

3. Use the Wien displacement law to find the wavelength at which the star has maximum spectral emittance. In what part of the electromagnetic spectrum is this wavelength?
4. Use your result from step 3 to find the range of wavelengths in which the star radiates most of its energy. Which end of this range corresponds to a photon with the greatest energy?
5. Write an expression for the wavelength of a photon that must be absorbed to cause an electron transition from the ground level ($n = 1$) to a higher level n. Solve for the value of n that corresponds to the highest-energy photon in the range you calculated in step 4. (*Hint:* Remember than n must be an integer.)
6. Find the electron wavelengths that correspond to the $n = 2$ and $n = 3$ orbits shown in Fig. 39.22.

EVALUATE

7. Check your result in step 5 by calculating the wavelength needed to excite a hydrogen atom from the ground level into the level *above* the highest-energy level that you found in step 5. Is it possible for light in the range of wavelengths you found in step 4 to excite hydrogen atoms from the ground level into this level?
8. How do the electron wavelengths you found in step 6 compare to the wavelength of a *photon* emitted in a transition from the $n = 3$ level to the $n = 2$ level?

Problems

For instructor-assigned homework, go to www.masteringphysics.com

•, ••, •••: Problems of increasing difficulty. **CP**: Cumulative problems incorporating material from earlier chapters. **CALC**: Problems requiring calculus. **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 $\pm 3e$ instead of just $+e$. How are the energy levels related to those of hydrogen? How is the *radius* of the ion in the ground level related to that of the hydrogen atom? Explain.

Q39.7 The emission of a photon by an isolated atom is a recoil process in which momentum is conserved. Thus Eq. (39.5) should include a recoil kinetic energy K_r for the atom. Why is this energy negligible in that equation?

Q39.8 How might the energy levels of an atom be measured directly—that is, without recourse to analysis of spectra?

Q39.9 Elements in the gaseous state emit line spectra with well-defined wavelengths. But hot solid bodies always emit a continuous spectrum—that is, a continuous smear of wavelengths. Can you account for this difference?

Q39.10 As a body is heated to a very high temperature and becomes self-luminous, the apparent color of the emitted radiation shifts from red to yellow and finally to blue as the temperature increases. Why does the color shift? What other changes in the character of the radiation occur?

Q39.11 The peak-intensity wavelength of red dwarf stars, which have surface temperatures around 3000 K, is about 1000 nm, which is beyond the visible spectrum. So why are we able to see these stars, and why do they appear red?

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 • In the Bohr model of the hydrogen atom, what is the de Broglie wavelength for the electron when it is in (a) the $n = 1$