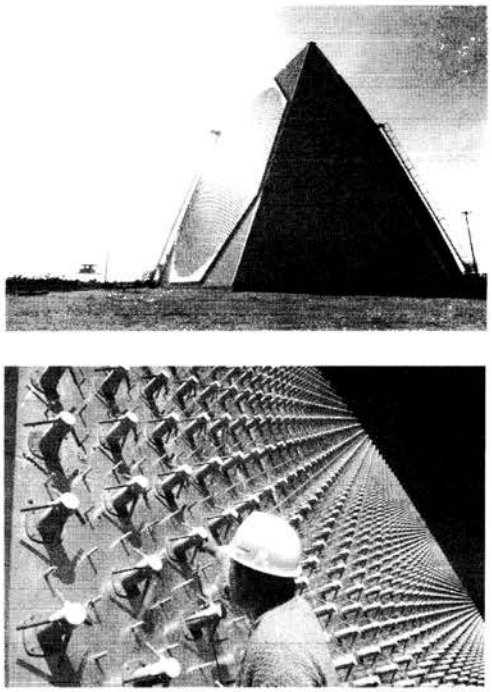


Figure 4.16 Plane waves enter from the left and are reflected off to the right. The reflected wavefront $\overline{CD}$ is formed of waves scattered by the atoms on the surface from A to D. Just as the first wavelet arrives at C from A, the atom at D emits, and the wavefront along $\overline{CD}$ is completed.



A modern phased-array radar system. The field of individual small antennas behaves very much like the atoms on a smooth surface. By introducing a proper phase shift between adjacent rows the antenna can "look" in any direction. A reflecting surface has a similar phase shift determined by θ_i as the incident wave sweeps over the array of atoms. (Courtesy Raytheon Corp.)

that the wavelet emitted from A will arrive at C in-phase with the wavelet just being emitted from D (as it is stimulated by B), as long as the distances $\overline{AC}$ and $\overline{BD}$ are equal. In other words, if all the wavelets emitted from all the surface scatterers are to overlap in-phase and form a single reflected plane wave, it must be that $\overline{AC} = \overline{BD}$. Then, since the two triangles have a common hypotenuse

$$\frac{\sin \theta_i}{\overline{BD}} = \frac{\sin \theta_r}{\overline{AC}}$$

All the waves travel in the incident medium with the same speed v_i. It follows that in the time (Δt) it takes for point B on the wavefront to reach point D on the surface, the wavelet emitted from A reaches point C. In other words, $\overline{BD} = v_i \Delta t = \overline{AC}$, and so from the above equation, $\sin \theta_i = \sin \theta_r$, which means that

$$\theta_i = \theta_r \tag{4.3}$$

The angle-of-incidence equals the angle-of-reflection. This equation is the first part of the **Law of Reflection**. It initially appeared in the book *Catoptrics*, which was purported to have been written by Euclid. We say that a beam is *normally incident* when $\theta_i = 0°$, in which case $\theta_r = 0°$ and for a mirror the beam reflects back on itself. Similarly, *glancing incidence* corresponds to $\theta_i \approx 90°$ and perforce $\theta_r \approx 90°$.

Rays

Drawing wavefronts can get things a bit cluttered, so we introduce another convenient scheme for visualizing the progression of light. The imagery of antiquity was in terms of straight-line streams of light, a notion that got into Latin as "radii" and reached English as "rays." ***A ray is a line drawn in space corresponding to the direction of flow of radiant energy***. It is a mathematical construct and not a physical entity. In a medium that is uniform (homogeneous), rays are straight. If the medium behaves in the same manner in every direction (isotropic), ***the rays are perpendicular to the wavefronts***. Thus for a point source emitting spherical waves, the rays, which are perpendicular to them, point radially outward from the source. Similarly, the rays associated with plane waves are all parallel. Rather than sketching bundles of rays, we can simply draw one incident ray and one reflected ray (Fig. 4.17*a*). *All the angles are now measured from the perpendicular (or*

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

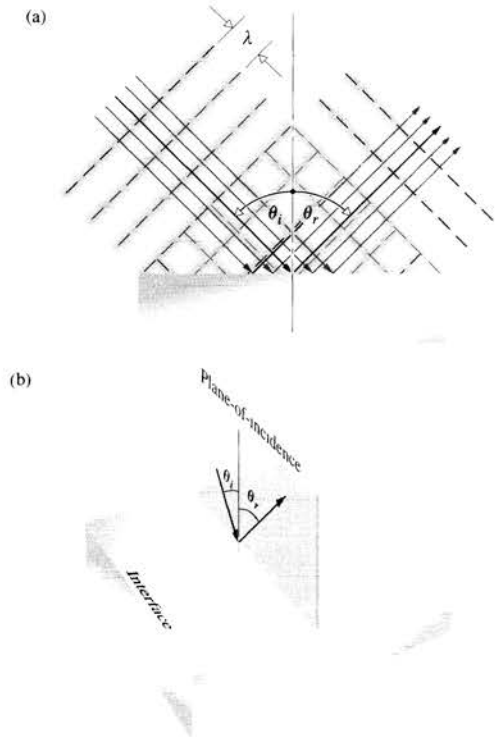


Figure 4.17 (*a*) Select one ray to represent the beam of plane waves. Both the angle-of-incidence θ_i and the angle-of-reflection θ_r are measured from a perpendicular drawn to the reflecting surface. (*b*) The incident ray and the reflected ray define the *plane-of-incidence*, perpendicular to the reflecting surface.

normal) to the surface, and θ_i and θ_r have the same numerical values as before (Fig. 4.16).

The ancient Greeks knew the Law of Reflection. It can be deduced by observing the behavior of a flat mirror, and nowadays that observation can be done most simply with a flashlight or, even better, a low-power laser. The second part of the Law of Reflection maintains that ***the incident ray, the perpendicular to the surface, and the reflected ray all lie in a plane*** called the **plane-of-incidence** (Fig. 4.17*b*)—this is a three-dimensional business. Try to hit some target in a room with a flashlight beam by reflecting it off a stationary mirror, and the importance of this second part of the law becomes obvious!

Figure 4.18*a* shows a beam of light incident upon a reflecting surface that is smooth (one for which any irregularities are small compared to a wavelength). In that case, the light re-

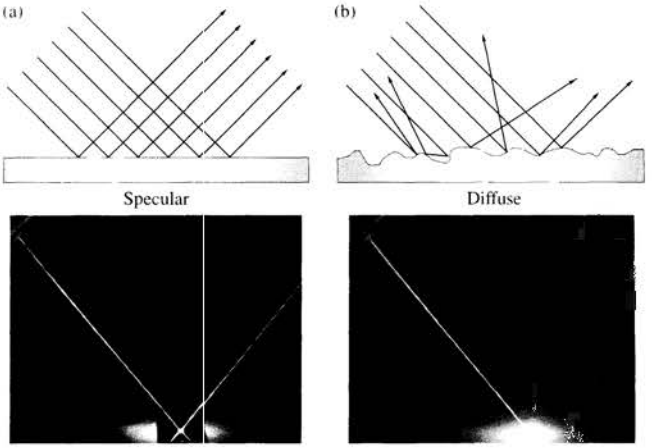
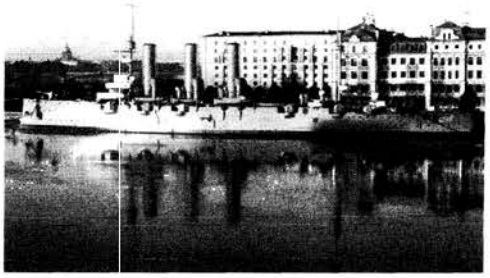


Figure 4.18 (*a*) Specular reflection. (*b*) Diffuse reflection. (Photos courtesy Donald Dunitz.)

emitted by millions upon millions of atoms will combine to form a single well-defined beam in a process called **specular reflection** (from the word for a common mirror alloy in ancient times, speculum). Provided the ridges and valleys are small compared to λ, the scattered wavelets will still arrive more or less in-phase when $\theta_i = \theta_r$. This is the situation assumed in Figs. 4.13, 4.15, 4.16, and 4.17. On the other hand, when the surface is rough in comparison to λ, although the angle-of-incidence will equal the angle-of-reflection for each ray, the whole lot of rays will emerge every which way, constituting what is called **diffuse reflection** (see photo). Both of these conditions are extremes; the reflecting behavior of most

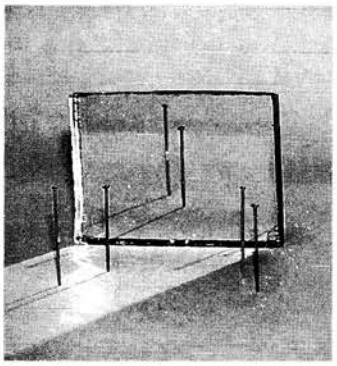


The cruiser *Aurora*, which played a key role in the Communist Revolution (1917), docked in St. Petersburg. Where the water is still, the reflection is specular. The image blurs where the water is rough and the reflection diffuse.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



The F-117A Stealth fighter has an extremely small radar profile, that is, it returns very little of the incoming microwaves back to the station that sent them. That's accomplished mostly by constructing the aircraft with flat tilted-planes that use the Law of Reflection to scatter the radar waves away from their source. One wants to avoid $\theta_i = \theta_r \approx 0$.



By placing a pair of pins in front of a flat mirror and aligning their images with another pair of pins, you can easily verify that $\theta_i = \theta_r$. (Photo by E.H.)

surfaces lies somewhere between them. Thus, although the paper of this page was deliberately manufactured to be a fairly diffuse scatterer, the cover of the book reflects in a manner that is somewhere between diffuse and specular.

4.4 Refraction

Figure 4.13 shows a beam of light impinging on an interface at some angle ($\theta_i \neq 0$). The interface corresponds to a major inhomogeniety, and the atoms that compose it scatter light both backward, as the reflected beam, and forward, as the transmitted beam. The fact that the incident rays are bent or "turned out of their way," as Newton put it, is called **refraction**.

Examine the transmitted or refracted beam. Speaking classically, each energized molecule on the interface radiates wavelets into the glass that expand out at speed c. These can be imagined as combining into a secondary wave that then recombines with the unscattered remainder of the primary wave, to form the net transmitted wave. The process continues over and over again as the wave advances in the transmitting medium.

However we visualize it, immediately on entering the transmitting medium, there is a single net field, a single net wave. As we have seen, this transmitted wave usually propagates with an effective speed $v_t < c$. It's essentially as if the atoms at the interface scattered "slow wavelets" into the glass that combine to form the "slow transmitted wave." We'll come back to this imagery when we talk about Huygens's Principle. In any event, because the cooperative phenomenon known as the transmitted electromagnetic wave *is* slower than the incident electromagnetic wave, the transmitted wavefronts are refracted, displaced (turned with respect to the incident wavefronts), and the beam bends.

4.4.1 The Law of Refraction

Figure 4.19 picks up where we left off with Figs. 4.13 and 4.16. The diagram depicts several wavefronts, all shown at a single instant in time. Remember that each wavefront is a surface of constant phase, and, to the degree that the phase of the

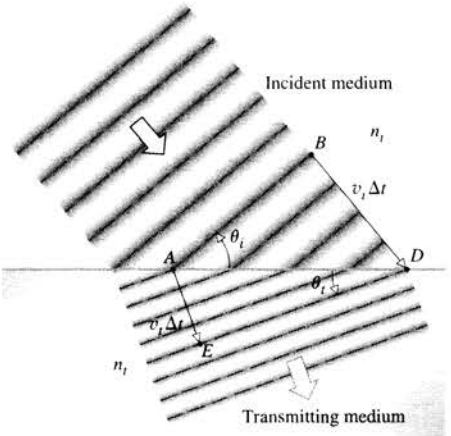


Figure 4.19 The refraction of waves. The atoms in the region of the surface of the transmitting medium reradiate wavelets that combine constructively to form a refracted beam.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

net field is retarded by the transmitting medium, each wavefront is held back, as it were. The wavefronts "bend" as they cross the boundary because of the speed change. Alternatively, we can envision Fig. 4.19 as a multiple-exposure picture of a single wavefront showing it after successive equal intervals of time. Notice that in the time Δt, which it takes for point B on a wavefront (traveling at speed v_i) to reach point D, the transmitted portion of that same wavefront (traveling at speed v_t) has reached point E. If the glass ($n_t = 1.5$) is immersed in an incident medium that is vacuum ($n_i = 1$) or air ($n_i = 1.000\,3$) or anything else where $n_t > n_i$, $v_t < v_i$ and $\overline{AE} < \overline{BD}$, the wavefront bends. The refracted wavefront extends from E to D, making an angle with the interface of θ_t. As before, the two triangles ABD and AED in Fig. 4.19 share a common hypotenuse ($\overline{AD}$), and so

$$\frac{\sin \theta_i}{\overline{BD}} = \frac{\sin \theta_t}{\overline{AE}}$$

where $\overline{BD} = v_i \, \Delta t$ and $\overline{AE} = v_t \, \Delta t$. Hence

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

Multiply both sides by c, and since $n_i = c/v_i$ and $n_t = c/v_t$

$$n_i \sin \theta_i = n_t \sin \theta_t \tag{4.4}$$

This equation is the first portion of the **Law of Refraction**, also known as **Snell's Law** after the man who proposed it (1621), Willebrord Snel van Royen (1591–1626). Snel's analysis has been lost, but contemporary accounts follow the treatment shown in Fig. 4.20. What was found through observation was that the bending of the rays could be quantified via

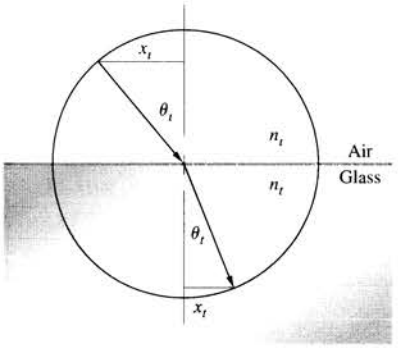


Figure 4.20 Descartes's arrangement for deriving the Law of Refraction.

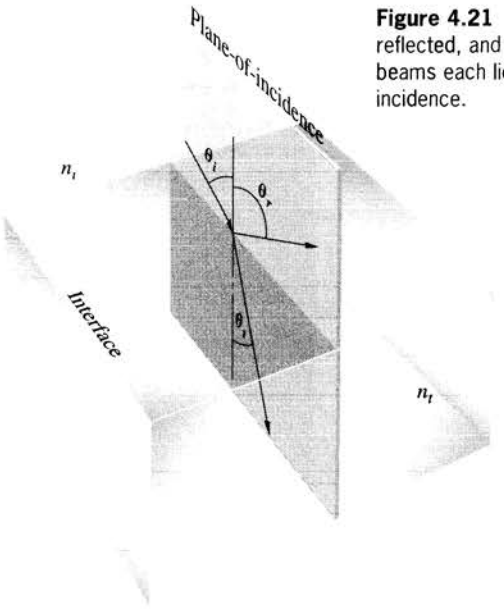


Figure 4.21 The incident, reflected, and transmitted beams each lie in the plane-of-incidence.

the ratio of x_i to x_t which was constant for all θ_i. That constant was naturally enough called the *index of refraction*. In other words,

$$\frac{x_i}{x_t} \equiv n_t$$

and in air that's equivalent to Eq. (4.4). We now know that the Englishman Thomas Harriot had come to the same conclusion before 1601, but he kept it to himself.

At first, the indices of refraction were simply experimentally determined constants of the physical media. Later, Newton was actually able to derive Snell's Law using his own corpuscular theory. By then, the significance of n as a measure of the speed of light was evident. Still later, Snell's Law was shown to be a natural consequence of Maxwell's Electromagnetic Theory (p. 112).

It is again convenient to transform the diagram into a ray representation (Fig. 4.21) wherein all the angles are measured from the perpendicular. Along with Eq. (4.4), there goes the understanding that ***the incident, reflected, and refracted rays all lie in the plane-of-incidence***. In other words, the respective unit propagation vectors $\hat{\mathbf{k}}_i$, $\hat{\mathbf{k}}_r$, and $\hat{\mathbf{k}}_t$ are coplanar (Fig. 4.22).

When $n_i < n_t$ (that is, when the light is initially traveling within the lower-index medium), it follows from Snell's Law that $\sin \theta_i > \sin \theta_t$, and since the same function is everywhere positive between 0° and 90°, then $\theta_i > \theta_t$. Rather than going

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

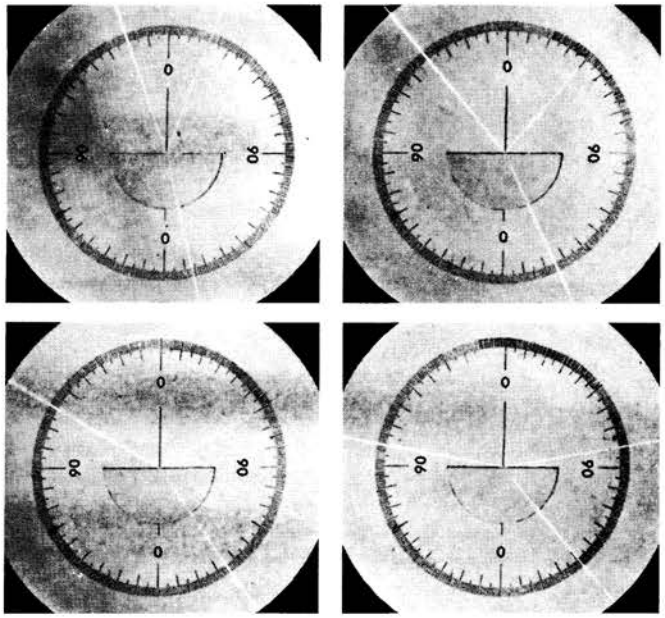


Figure 4.22 Refraction at various angles of incidence. Notice that the bottom surface is cut circular so that the transmitted beam within the glass always lies along a radius and is normal to the lower surface in every case. (Photos courtesy *PSSC College Physics*, D. C. Heath & Co., 1968.)

straight through, ***the ray entering a higher-index medium bends toward the normal*** (Fig. 4.23*a*). The reverse is also true (Fig. 4.23*b*); that is, ***on entering a medium having a lower index, the ray, rather than going straight through, will bend away from the normal*** (see photo). Notice that this implies that the rays will traverse the same path going either way, into or out of either medium. The arrows can be reversed and the resulting picture is still true.

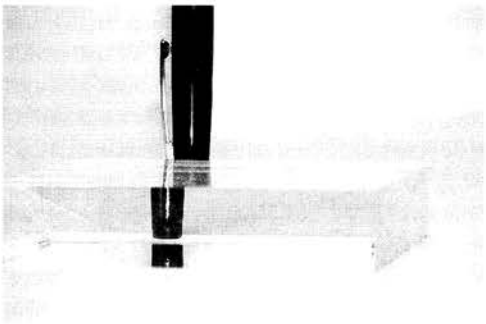
Snell's Law can be rewritten in the form

$$\frac{\sin\theta_i}{\sin\theta_t} = n_{ti} \tag{4.5}$$

where $n_{ti} \equiv n_t/n_i$ is the **relative index of refraction** of the two media.

Let $\hat{\mathbf{u}}_n$ be a unit vector normal to the interface pointing in the direction from the incident to the transmitting medium (Fig. 4.24). As you will have the opportunity to prove in Problem 4.29, the complete statement of the Law of Refraction can be written vectorially as

$$n_i(\hat{\mathbf{k}}_i \times \hat{\mathbf{u}}_n) = n_t(\hat{\mathbf{k}}_t \times \hat{\mathbf{u}}_n) \tag{4.6}$$



The image of a pen seen through a thick block of clear plastic. The displacement of the image arises from the refraction of light toward the normal at the air-plastic interface. If this arrangement is set up with a narrow object (e.g., an illuminated slit) and the angles are carefully measured, you can confirm Snell's Law directly. (Photo by E.H.)

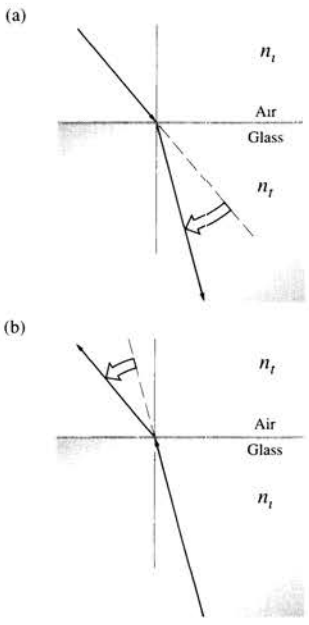


Figure 4.23 The bending of rays at an interface. (a) When a beam of light enters a more optically dense medium, one with a greater index of refraction ($n_i < n_t$), it bends toward the perpendicular. (b) When a beam goes from a more dense to a less dense medium ($n_i > n_t$), it bends away from the perpendicular.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

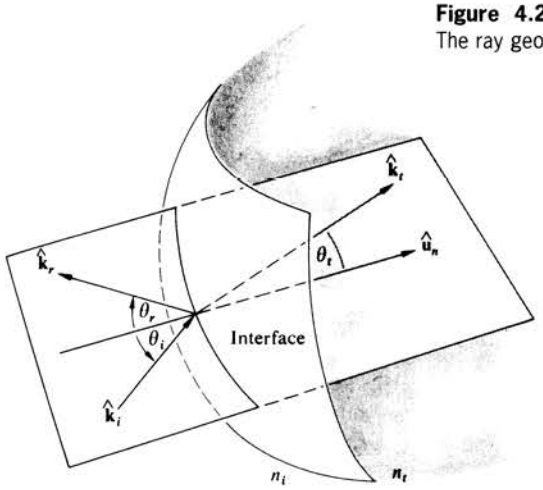


Figure 4.24
The ray geometry.

or alternatively,

$$n_t\hat{\mathbf{k}}_t - n_i\hat{\mathbf{k}}_i = (n_t \cos\theta_t - n_i \cos\theta_i)\,\hat{\mathbf{u}}_n \qquad (4.7)$$

Fig. 4.19 illustrates the three important changes that occur in the beam traversing the interface. (1) It changes direction. Because the leading portion of the wavefront in the glass slows down, the part still in the air advances more rapidly, sweeping past and bending the wave toward the normal. (2) The beam in the glass has a broader cross section than the beam in the air; hence, the transmitted energy is spread thinner. (3) The wavelength decreases because the frequency is unchanged while the speed decreases; $\lambda = v/\nu = c/n\nu$ and

$$\lambda = \frac{\lambda_0}{n} \qquad (4.8)$$

This latter notion suggests that the color aspect of light is better thought of as associated with its frequency (or energy, $\mathscr{E} = h\nu$) than its wavelength, since the wavelength changes with the medium through which the light moves. Color is so much a physio-psychological phenomenon (p. 131) that it must be treated rather gingerly. Still, even though it's a bit simplistic, it's useful to remember that blue photons are more energetic than red photons. When we talk about wavelengths and colors, we should always be referring to **vacuum wavelengths** (henceforth to be given as λ_0).

In all the situations treated thus far, it was assumed that the reflected and refracted beams always had the same frequency as the incident beam, and that's ordinarily a reasonable assumption. Light of frequency ν impinges on a medium and presumably drives the molecules into simple harmonic motion. That's certainly the case when the amplitude of the vibration is fairly small, as it is when the electric field driving the molecules is small. The E-field for bright sunlight is only about 1000 V/m (while the B-field is less than a tenth of the Earth's surface field). This isn't very large compared to the fields keeping a crystal together, which are of the order of 10^{11} V/m—just about the same magnitude as the cohesive field holding the electron in an atom. We can usually expect the oscillators to vibrate in simple harmonic motion, and so the frequency will remain constant—the medium will ordinarily respond linearly. That will not be true, however, if the incident beam has an exceedingly large-amplitude E-field, as can be the case with a high-power laser. So driven, at some frequency ν the medium can behave in a nonlinear fashion, resulting in reflection and refraction of harmonics (2ν, 3ν, etc.) in addition to ν. Nowadays, second-harmonic generators (p. 641) are available commercially. You shine red light (694.3 nm) into an appropriately oriented transparent nonlinear crystal (of, for example, potassium dihydrogen phosphate, KDP, or ammonium dihydrogen phosphate, ADP) and out will come a beam of UV (347.15 nm).

One feature of the above treatment merits some further discussion. It was reasonably assumed that each point on the interface in Fig. 4.13 coincides with a particular point on each of the incident, reflected, and transmitted waves. In other words, there is a fixed phase relationship between each of the



Rays from the submerged portion of the pencil bend on leaving the water as they rise toward the viewer. (Photo by E.H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

waves at all points along the interface. As the incident front sweeps across the interface, every point on it in contact with the interface is also a point on both a corresponding reflected front and a corresponding transmitted front. This situation is known as **wavefront continuity**, and it will be justified in a more mathematically rigorous treatment in Section 4.6.1. Interestingly, Sommerfeld* has shown that the laws of reflection and refraction (independent of the kind of wave involved) can be derived directly from the requirement of wavefront continuity and the solution to Problem 4.26 demonstrates as much.

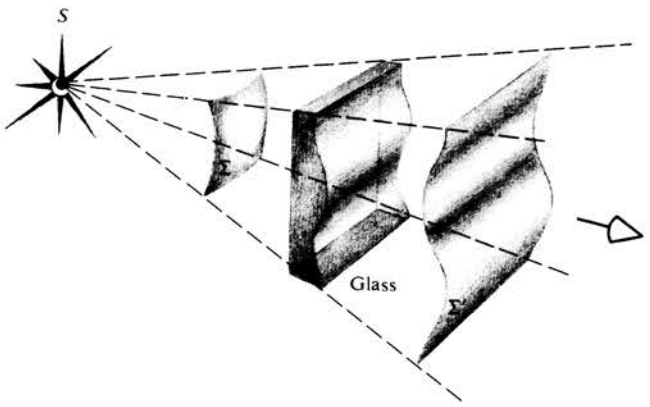


Figure 4.25 Distortion of a portion of a wavefront on passing through a material of nonuniform thickness.

4.4.2 Huygens's Principle

Suppose that light passes through a nonuniform sheet of glass, as in Fig. 4.25, so that the wavefront Σ is distorted. How can we determine its new form Σ'? Or for that matter, what will Σ' look like at some later time, if it is allowed to continue unobstructed?

A preliminary step toward the solution of this problem appeared in print in 1690 in the work entitled *Traité de la Lumière*, which had been written 12 years earlier by the Dutch physicist Christiaan Huygens. It was there that he enunciated what has since become known as **Huygens's Principle:** ***every point on a propagating wavefront serves as the source of spherical secondary wavelets, such that the wavefront at some later time is the envelope of these wavelets.***

A further crucial point is that ***if the propagating wave has a frequency ν, and is transmitted through the medium at a speed v_t, then the secondary wavelets have that same frequency and speed.*** Huygens was a brilliant scientist, and this is the basis of a remarkably insightful, though quite naive, scattering theory. It's a very early treatment and naturally has several shortcomings, one of which is that it doesn't overtly incorporate the concept of interference and perforce cannot deal with lateral scattering. Moreover, the idea that the secondary wavelets propagate at a speed determined by the medium (a speed that may even be anisotropic, e.g., p. 340) is a happy guess. Nonetheless, Huygens's Principle can be used to arrive at Snell's Law in a way that's similar to the treatment that led to Eq. (4.4). It's probably best not to fuss over the physical details (such as how to rationalize propagation in vacuum) and just use the Principle as a tool—a highly useful fiction that works. After all, if Einstein is right, there are only scattered photons; the wavelets themselves are a theoretical construct.

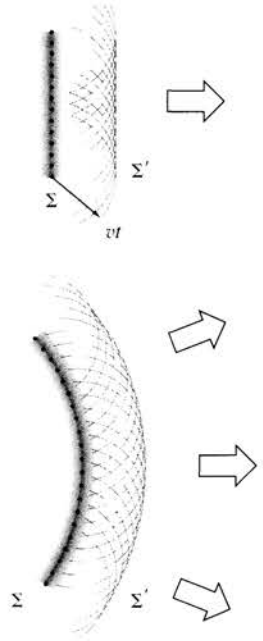


Figure 4.26 According to Huygens's Principle, a wave propagates as if the wavefront were composed of an array of point sources, each emitting a spherical wave.

*A. Sommerfeld, *Optics*, p. 151. See also J. J. Sein, *Am. J. Phys.* **50**, 180 (1982).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

If the medium is homogeneous, the wavelets may be constructed with finite radii, whereas if it is inhomogeneous, the wavelets must have infinitesimal radii. Figure 4.26 should make this fairly clear; it shows a view of a wavefront Σ, as well as a number of spherical secondary wavelets, which, after a time t, have propagated out to a radius of vt. The envelope of all these wavelets is then asserted to correspond to the advanced wave Σ'. It is easy to visualize the process in terms of mechanical vibrations of an elastic medium. Indeed, this is the way that Huygens envisioned it within the context of an all-pervading aether, as is evident from his comment:

> We have still to consider, in studying the spreading out of these waves, that each particle of matter in which a wave proceeds not only communicates its motion to the next particle to it, which is on the straight line drawn from the luminous point, but that it also necessarily gives a motion to all the others which touch it and which oppose its motion. The result is that around each particle there arises a wave of which this particle is a center.

Fresnel, in the 1800s, successfully modified Huygens's Principle mathematically adding in the concept of interference. A little later on, Kirchhoff showed that the *Huygens–Fresnel Principle* was a direct consequence of the differential wave equation [Eq. (2.60)], thereby putting it on a firm mathematical base. That there was a need for a reformulation of the principle is evident from Fig. 4.26, where we deceptively only drew hemispherical wavelets.* Had we drawn them as spheres, there would have been a *backwave* moving toward the source—something that is not observed. Since this difficulty was taken care of theoretically by Fresnel and Kirchhoff, we need not be disturbed by it.

4.4.3 Light Rays and Normal Congruence

In practice, one can produce very narrow *beams* or *pencils* of light (e.g., a laserbeam), and we might imagine a ray to be the unattainable limit on the narrowness of such a beam. Bear in mind that in an *isotropic medium* (i.e., one whose properties are the same in all directions) ***rays are orthogonal trajectories of the wavefronts***. That is to say, *they are lines normal to the wavefronts at every point of intersection.* Evidently, *in such a medium a ray is parallel to the propagation vector* $\vec{\mathbf{k}}$. As you might suspect, this is not true in *anisotropic* substances, which we will consider later (see Section 8.4.1). *Within homogeneous isotropic materials, rays will be straight lines*, since by symmetry they cannot bend in any preferred direction, there being none. Moreover, because the speed of propagation is identical in all directions within a given medium, the spatial separation between two wavefronts, measured along rays, must be the same everywhere.* Points where a single ray intersects a set of wavefronts are called *corresponding points*, for example, A, A', and A'' in Fig. 4.27. *Evidently, the separation in time between any two corresponding points on any two sequential wavefronts is identical.* If wavefront Σ is transformed into Σ'' after a time t'', the distance between corresponding points on any and all rays will be traversed in that same time t''. This will be true even if the wavefronts pass from one homogeneous isotropic medium into another. This just means that each point on Σ can be imagined as following the path of a ray to arrive at Σ'' in the time t''.

If a group of rays is such that we can find a surface that is orthogonal to each and every one of them, they are said to form a **normal congruence**. For example, the rays emanating from a point source are perpendicular to a sphere centered at the source and consequently form a normal congruence.

We can now briefly consider a scheme that will also allow us to follow the progress of light through various isotropic media. The basis for this approach is the ***Theorem of Malus and Dupin*** (introduced in 1808 by E. Malus and modified in

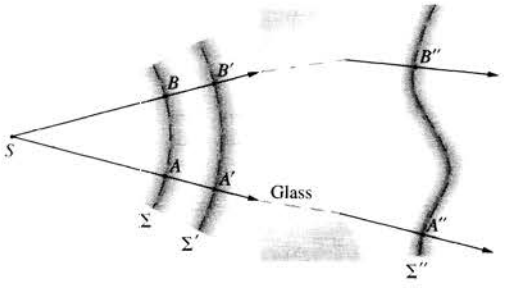


Figure 4.27 Wavefronts and rays.

*See E. Hecht, *Phys. Teach.* **18**, 149 (1980).

*When the material is inhomogeneous or when there is more than one medium involved, it will be the *optical path length* (see Section 4.5) between the two wavefronts that is the same.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

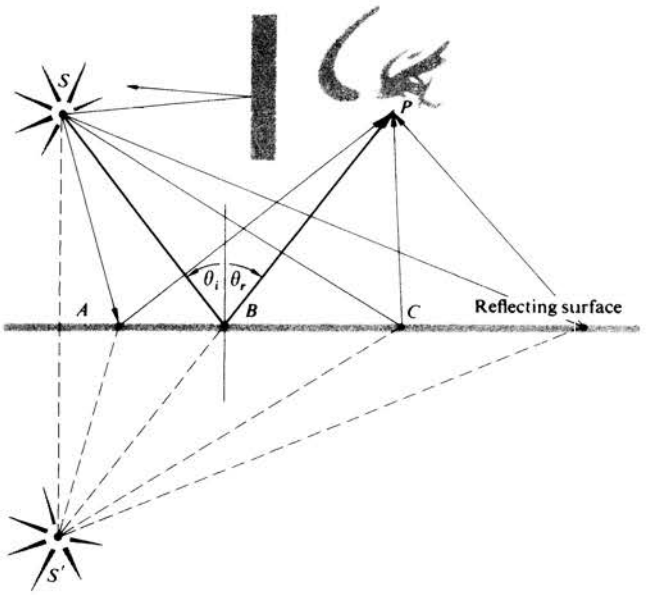


Figure 4.28 Minimum path from the source S to the observer's eye at P.

1816 by C. Dupin), according to which ***a group of rays will preserve its normal congruence after any number of reflections and refractions*** (as in Fig. 4.27). From our present vantage point of the wave theory, this is equivalent to the statement that rays remain orthogonal to wavefronts throughout all propagation processes in isotropic media. As shown in Problem 4.28, the theorem can be used to derive the Law of Reflection as well as Snell's Law. It is often most convenient to carry out a ray trace through an optical system and then reconstruct the wavefronts using the idea of equal transit times between corresponding points and the orthogonality of the rays and wavefronts.

4.5 Fermat's Principle

The laws of reflection and refraction, and indeed the manner in which light propagates in general, can be viewed from an entirely different and intriguing perspective afforded us by **Fermat's Principle**. The ideas that will unfold presently have had a tremendous influence on the development of physical thought in and beyond the study of Classical Optics.

Hero of Alexandria, who lived some time between 150 B.C.E. and A.D. 250, was the first to propose what has since become known as a *variational principle*. In his treatment of reflection, he asserted that *the path taken by light in going from some point S to a point P via a reflecting surface was the shortest possible one*. This can be seen rather easily in Fig. 4.28, which depicts a point source S emitting a number of rays that are then "reflected" toward P. Presumably, only one of these paths will have any physical reality. If we draw the rays as if they emanated from S' (the image of S), none of the distances to P will have been altered (i.e., $SAP = S'AP$, $SBP = S'BP$, etc.). But obviously the straight-line path $S'BP$, which corresponds to $\theta_i = \theta_r$, is the shortest possible one. The same kind of reasoning (Problem 4.31) makes it evident that points S, B, and P must lie in what has previously been defined as the plane-of-incidence. For over fifteen hundred years Hero's curious observation stood alone, until in 1657 Fermat propounded his celebrated *Principle of Least Time*, which encompassed both reflection and refraction. A beam of light traversing an interface does not take a straight line or *minimum spatial path* between a point in the incident medium and one in the transmitting medium. Fermat consequently reformulated Hero's statement to read: *the actual path between two points taken by a beam of light is the one that is traversed in the least time*. As we shall see, even this form of the statement is incomplete and a bit erroneous at that. For the moment then, let us embrace it but not passionately.

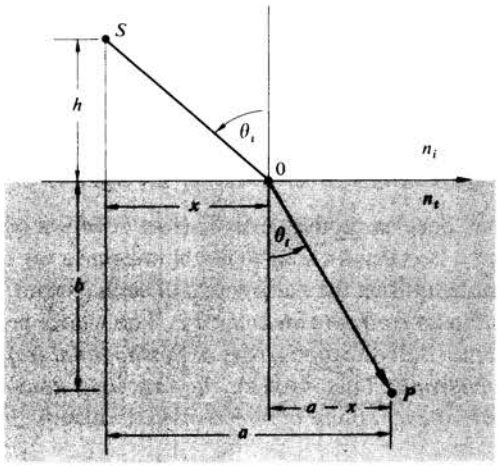


Figure 4.29 Fermat's Principle applied to refraction.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

As an example of the application of the principle to the case of refraction, refer to Fig. 4.29, where we minimize t, the transit time from S to P, with respect to the variable x. In other words, changing x shifts point O, changing the ray from S to P. The smallest transit time will then presumably coincide with the actual path. Hence

$$t = \frac{\overline{SO}}{v_t} + \frac{\overline{OP}}{v_t}$$

or

$$t = \frac{(h^2 + x^2)^{1/2}}{v_i} + \frac{[b^2 + (a - x)^2]^{1/2}}{v_t}$$

To minimize $t(x)$ with respect to variations in x, we set $dt/dx = 0$, that is,

$$\frac{dt}{dx} = \frac{x}{v_i(h^2 + x^2)^{1/2}} + \frac{-(a - x)}{v_t[b^2 + (a - x)^2]^{1/2}} = 0$$

Using the diagram, we can rewrite the expression as

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

which is no less than Snell's Law (Eq. 4.4). If a beam of light is to advance from S to P in the least possible time, it must comply with the Law of Refraction.

Suppose that we have a stratified material composed of m layers, each having a different index of refraction, as in Fig. 4.30. The transit time from S to P will then be

$$t = \frac{s_1}{v_1} + \frac{s_2}{v_2} + \cdots + \frac{s_m}{v_m}$$

or

$$t = \sum_{i=1}^{m} s_i/v_i$$

where s_i and v_i are the path length and speed, respectively, associated with the ith contribution. Thus

$$t = \frac{1}{c}\sum_{i=1}^{m} n_i s_i \tag{4.9}$$

in which the summation is known as the **optical path length** (OPL) traversed by the ray, in contrast to the spatial path length $\sum_{i=1}^{m} s_i$. Clearly, for an inhomogeneous medium where n is a function of position, the summation must be changed to an integral:

$$OPL = \int_S^P n(s)\, ds \tag{4.10}$$

The optical path length corresponds to the distance in vacuum

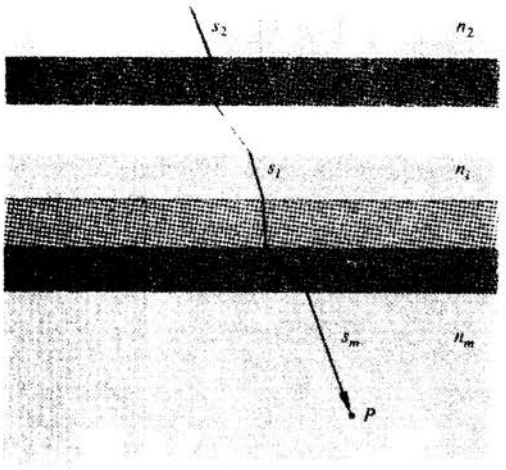


Figure 4.30 A ray propagating through a layered material.

equivalent to the distance traversed (s) in the medium of index n. That is, the two will correspond to the same number of wavelengths, $(OPL)/\lambda_0 = s/\lambda$, and the same phase change as the light advances.

Inasmuch as $t = (OPL)/c$, we can restate Fermat's Principle: *light, in going from point S to P, traverses the route having the smallest optical path length.*

Fermat and Mirages

When light rays from the Sun pass through the inhomogeneous atmosphere of the Earth, as shown in Fig. 4.31, they bend so as to traverse the lower, denser regions as abruptly as

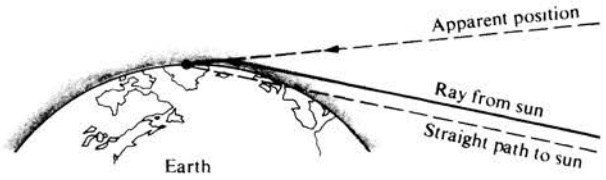


Figure 4.31 The bending of rays through inhomogeneous media. Because the rays bend as they pass through the atmosphere the Sun appears higher in the sky.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

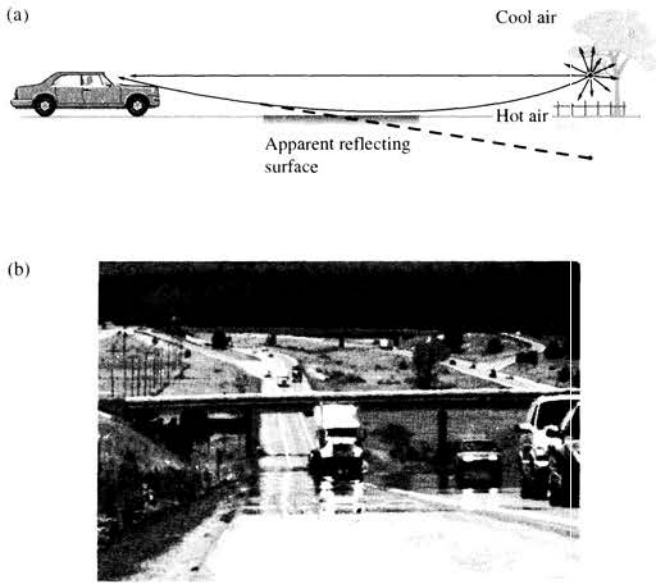


Figure 4.32 (*a*) At very low angles the rays appear to be coming from beneath the road as if reflected in a puddle. (*b*) A photo of this puddle effect (Photo courtesy Matt Malloy and Dan MacIsaac, NAU Physics & Astronomy.)

possible, minimizing the *OPL*. Ergo, one can still see the Sun after it has actually passed below the horizon.

In the same way, a road viewed at a glancing angle, as in Fig. 4.32, appears to reflect the environs as if it were covered with a sheet of water. The air near the roadway is warmer and less dense than that farther above it. It was established experimentally by Gladstone and Dale that for a gas of density ρ

$$(n - 1) \propto \rho$$

It follows from the Ideal Gas Law that at a fixed pressure, since $\rho \propto P/T$, $(n - 1) \propto 1/T$; the hotter the road, the lower the index of refraction of the air immediately above it.

According to Fermat's Principle, a ray leaving a branch in Fig. 4.32*a* heading somewhat downward would take a route that minimized the *OPL*. Such a ray would bend upward, passing through more of the less dense air than if it had traveled straight. To appreciate how that works, imagine the air divided into an infinite number of infinitesimally thin constant-n horizontal layers. A ray passing from layer to layer would bend (via Snell's Law) slightly upward at each interface (much as in Fig. 4.29 held upside down with the ray run backwards). Of course, if the ray comes down nearly vertically it makes a small angle-of-incidence at each interface between layers, only bends slightly, and soon strikes the ground where no one will "see" it.

On the other hand, it is possible that a light ray that comes in at a shallow enough angle to begin with could ultimately approach an interface at glancing incidence (p. 116). It would then be completely reflected (p. 122), thereby starting its climb back up into the more dense air (much as in Fig. 4.29 held upside down with the ray run forwards).

Any viewer, off on the left in Fig. 4.32, who receives these bent rays, naturally projects them straight backward as if they were reflected from a mirrored surface. Depending on where you stand, you'll see a different mirage puddle, but it will always be far from you and so will always disappear as you approach it. The effect is particularly easy to view on long modern highways. The only requirement is that you look at the road at near glancing incidence, because the rays bend very gradually.*

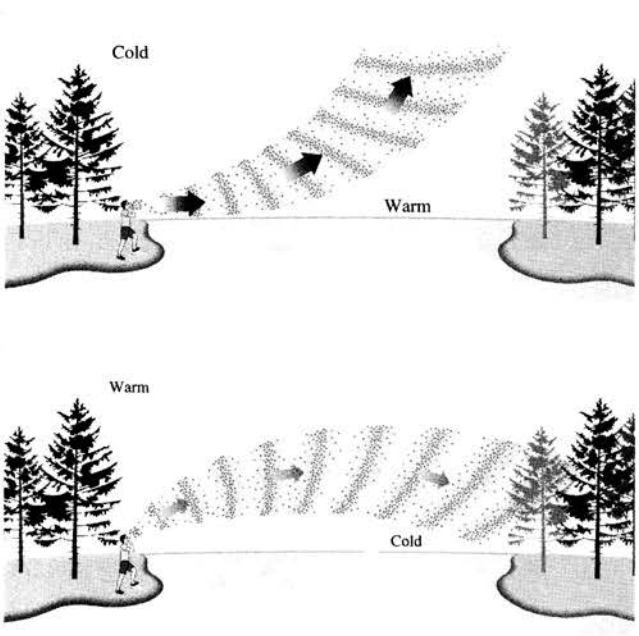


Figure 4.33 The puddle mirage can be understood via waves; the speed, and therefore the wavelength, increase in the less dense medium. That bends the wavefronts and the rays. The same effect is common with sound waves; (*a*) when the surface air is cold, sounds can be heard much farther than normal. (*b*) And when it's warm, sounds seem to vanish into the air.

*See for example T. Kosa and P. Palffy-Muhoray, "Mirage mirror on the wall," *Am. J. Phys.* **68** (12), 1120 (2000).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

The same effect is well known as it applies to sound. Figure 4.33 depicts the alternative understanding in terms of waves. The wavefronts bend because of temperature-induced changes in speed and therefore in wavelength. (The speed of sound is proportional to the square root of the temperature.) The noises of people on a hot beach climb up and away, and the place can seem strangely quiet. The opposite occurs in the evening when the ground cools before the upper air and distant sounds can clearly be heard.

The Modern Formulation of Fermat's Principle

The original statement of Fermat's *Principle of Least Time* has some serious failings and is in need of alteration. To that end, recall that if we have a function, say $f(x)$, we can determine the specific value of the variable x that causes $f(x)$ to have a *stationary* value by setting $df/dx = 0$ and solving for x. By a stationary value, we mean one for which the slope of $f(x)$ versus x is zero or equivalently where the function has a maximum ⌒, minimum ⌣, or a point of inflection with a horizontal tangent ⌁.

Fermat's Principle in its modern form reads: ***a light ray in going from point S to point P must traverse an optical path length that is stationary with respect to variations of that path***. In essence what that means is that the curve of the *OPL* versus x will have a somewhat flattened region in the vicinity of where the slope goes to zero. The zero-slope point corresponds to the actual path taken. In other words, the *OPL* for the true trajectory will equal, to a first approximation, the *OPL* of paths immediately adjacent to it.* For example, in a situation where the *OPL* is a minimum, as with the refraction illustrated in Fig. 4.29, the *OPL* curve will look something like

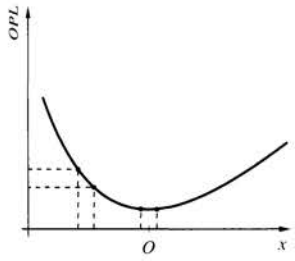


Figure 4.34 In the situation shown in Fig. 4.29 the actual location of point *O* corresponds to a path of minimum *OPL*.

*The first derivative of the *OPL* vanishes in its Taylor series expansion, since the path is stationary.

Fig. 4.34. A small change in x in the vicinity of O has little effect on the *OPL*, but a similar change in x anywhere well away from O results in a substantial change in *OPL*. Thus there will be many paths neighboring the actual one that would take nearly the same time for the light to traverse. This latter insight makes it possible to begin to understand how light manages to be so clever in its meanderings.

Suppose that a beam of light advances through a homogeneous isotropic medium (Fig. 4.35) so that a ray passes from points S to P. Atoms within the material are driven by the incident disturbance, and they reradiate in all directions. Wavelets progressing along paths in the immediate vicinity of a stationary straight-line path will reach P by routes that differ only slightly in *OPL* (as with group-I in Fig. 4.35*b*). They will therefore arrive nearly in-phase and reinforce each other. Think of each wavelet represented by a tiny phasor that rotates once around as the wave advances one wavelength (p. 23) along any ray path. Because the *OPL*s are all about the same, the phasors at P all point in more or less the same direction, and even though they're all small they combine to make the dominant contribution.

Wavelets taking other paths far from the stationary one (as with group-II in Fig. 4.35*b*) will arrive at P appreciably out-of-phase with each other and will therefore tend to cancel. In oth-

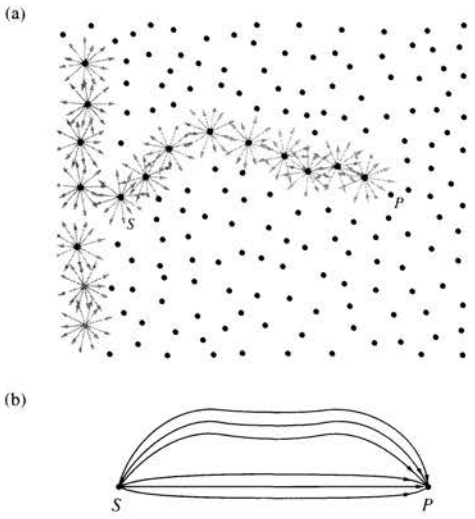


Figure 4.35 (*a*) Light can presumably take any number of paths from *S* to *P*, but it apparently takes only the one that corresponds to a stationary *OPL*. All other routes effectively cancel out. (*b*) For example, if some light takes each of the three upper paths in the diagram, it arrives at *P* with three very different phases and interferes more or less destructively.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

er words, there will be large angles between the little phasors; placed tip-to-tail they'll spiral around producing only a tiny net contribution. Keep in mind that we've just drawn three ray paths—the argument would be better made with millions of them in each group.

We can conclude that energy will effectively propagate along the ray from S to P that satisfies Fermat's Principle. And this is true whether we're talking about interfering electromagnetic waves or photon probability amplitudes (p. 138).

We can expect that this same logic holds for all propagation processes,* such as, for example, reflection from a plane mirror (Fig. 4.28). There, spherical waves leaving S sweep across the entire mirror, and yet an observer at P sees a well defined point source and not a great blotch of light covering the whole surface. Only rays for which $\theta_i \approx \theta_r$ (as with group-I in Fig. 4.36) have a stationary OPL; the associated wavelets will arrive at P nearly in-phase and reinforce each other. All other rays (e.g., group-II in Fig. 4.36) will make negligible contributions to the energy reaching P.

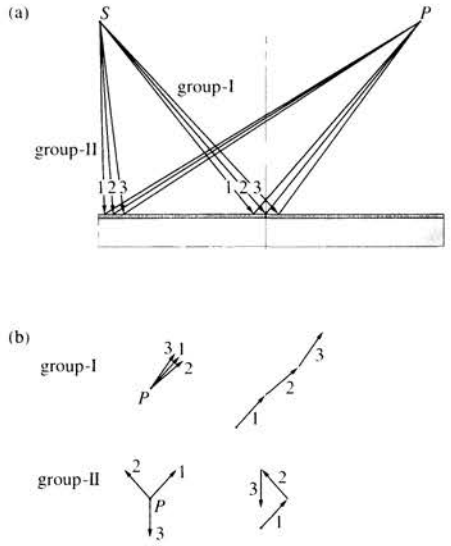


Figure 4.36 Rays reflecting off a plane mirror. Only those in group-I for which the OPL is stationary will correspond to waves that arrive at point-P more or less in-phase. There phasors will add along an almost straight line, producing a substantial resultant wave amplitude (going from the tail of 1 to the tip of 3). The phasors for group-II have large phase angle differences and so when added they essentially spiral around producing a very small resultant wave amplitude (going from the tail of 1 to the tip of 3). Of course, we should really be drawing millions of very tiny phasors in each group and not just three relatively large ones.

*We'll come back to these ideas when we consider QED in this chapter and the Fresnel zone plate in Chapter 10.

Stationary Paths

To see that the OPL for a ray need not always be a minimum, examine Fig. 4.37, which depicts a segment of a hollow three-dimensional ellipsoidal mirror. If the source S and the observer P are at the foci of the ellipsoid, then by definition the length SQP will be constant, regardless of where on the perimeter Q happens to be. It is also a geometrical property of the ellipse that $\theta_i = \theta_r$ for any location of Q. All optical paths from S to P via a reflection are therefore precisely equal. None is a minimum, and the OPL is clearly stationary with respect to variations. Rays leaving S and striking the mirror will arrive at the focus P. From another viewpoint we can say that radiant energy emitted by S will be scattered by electrons in the mirrored surface such that the wavelets will substantially reinforce each other only at P, where they have traveled the same distance and have the same phase. In any case, if a plane mirror was tangent to the ellipse at Q, the exact same path SQP traversed by a ray would then be a relative minimum. At the other extreme, if the mirrored surface conformed to a curve lying within the ellipse, like the dashed one shown, that same ray along SQP would now negotiate a relative maximum OPL. This is true even though other unused paths (where $\theta_i \neq \theta_r$) would actually be shorter (i.e., apart from inadmissible curved paths). Thus in all cases the rays travel a stationary OPL in accord with the reformulated Fermat's Principle. Note that since the Principle speaks only about the path and not the direction along it, a ray going from P to S will trace the same route as one from S to P. This is the very useful *Principle of Reversibility*.

Fermat's achievement stimulated a great deal of effort to supersede Newton's laws of mechanics with a similar variational formulation. The work of many men, notably Pierre de Maupertuis (1698–1759) and Leonhard Euler, finally led to the mechanics of Joseph Louis Lagrange (1736–1813) and hence to the *Principle of Least Action*, formulated by William Rowan Hamilton (1805–1865). The striking similarity between the principles of Fermat and Hamilton played an important part in Schrödinger's development of Quantum Mechanics. In 1942 Richard Phillips Feynman (1918–1988) showed that Quantum Mechanics can be fashioned in an alternative way using a variational approach. The continuing evo-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

lution of variational principles brings us back to Optics via the modern formalism of Quantum Optics.

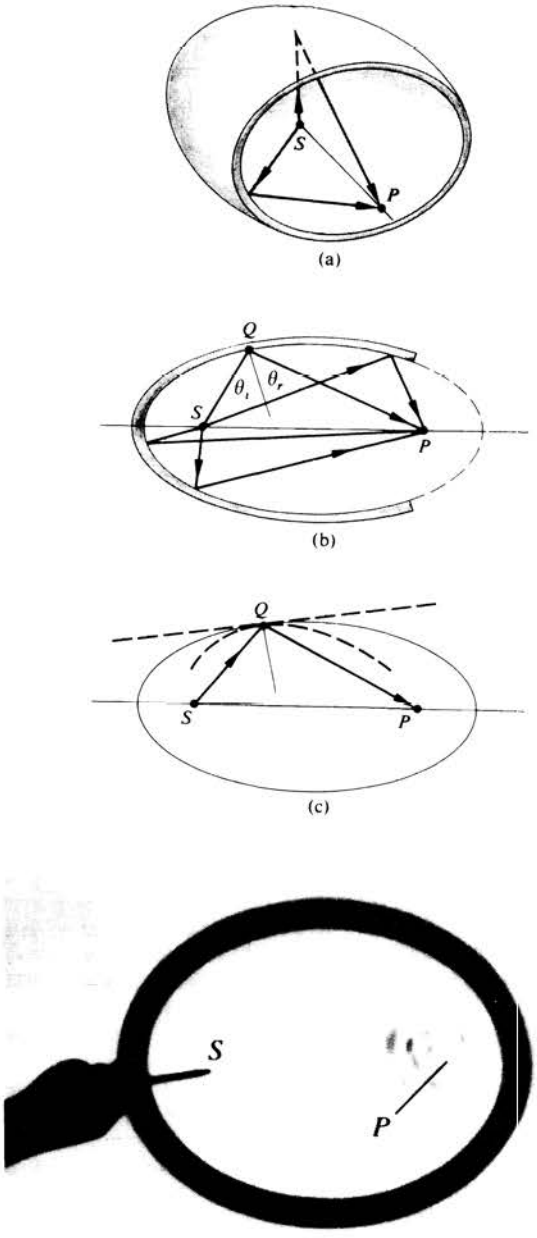


Figure 4.37 Reflection off an ellipsoidal surface. Observe the reflection of waves using a frying pan filled with water. Even though these are usually circular, it is well worth playing with. (Photo courtesy *PSSC College Physics*, D. C. Heath & Co., 1968.)

Fermat's Principle is not so much a computational device as it is a concise way of thinking about the propagation of light. It is a statement about the grand scheme of things without any concern for the contributing mechanisms, and as such it will yield insights under a myriad of circumstances.

4.6 The Electromagnetic Approach

Thus far, we have studied reflection and refraction from the perspectives of Scattering Theory, the Theorem of Malus and Dupin, and Fermat's Principle. Yet another and even more powerful approach is provided by Electromagnetic Theory. Unlike the previous techniques, which say nothing about the incident, reflected, and transmitted radiant flux densities (i.e., I_i, I_r, I_t, respectively), Electromagnetic Theory treats these within the framework of a far more complete description.

4.6.1 Waves at an Interface

Suppose that the incident monochromatic lightwave is planar, so that it has the form

$$\vec{\mathbf{E}}_i = \vec{\mathbf{E}}_{0i} \exp\left[i(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t)\right] \tag{4.11}$$

or, more simply,

$$\vec{\mathbf{E}}_i = \vec{\mathbf{E}}_{0i} \cos\left(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t\right) \tag{4.12}$$

Assume that $\vec{\mathbf{E}}_{0i}$ is constant in time; that is, the wave is linearly or plane polarized. We'll find in Chapter 8 that any form of light can be represented by two orthogonal linearly polarized waves, so that this doesn't actually represent a restriction. Note that just as the origin in time, $t = 0$, is arbitrary, so too is the origin O in space, where $\vec{\mathbf{r}} = 0$. Thus, making no assumptions about their directions, frequencies, wavelengths, phases, or amplitudes, we can write the reflected and transmitted waves as

$$\vec{\mathbf{E}}_r = \vec{\mathbf{E}}_{0r} \cos\left(\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} - \omega_r t + \varepsilon_r\right) \tag{4.13}$$

and

$$\vec{\mathbf{E}}_t = \vec{\mathbf{E}}_{0t} \cos\left(\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} - \omega_t t + \varepsilon_t\right) \tag{4.14}$$

Here ε_r and ε_t are *phase constants* relative to $\vec{\mathbf{E}}_i$ and are introduced because the position of the origin is not unique. Figure 4.38 depicts the waves in the vicinity of the planar interface between two homogeneous lossless dielectric media of indices n_i and n_t.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

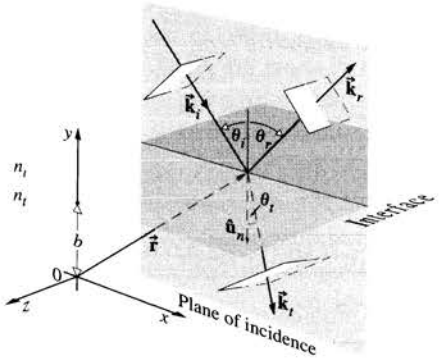


Figure 4.38 Plane waves incident on the boundary between two homogeneous, isotropic, lossless dielectric media.

The laws of Electromagnetic Theory (Section 3.1) lead to certain requirements that must be met by the fields, and they are referred to as the *boundary conditions*. Specifically, one of these is that the component of the electric field $\vec{\mathbf{E}}$ that is tangent to the interface must be continuous across it (the same is true for $\vec{\mathbf{H}}$). In other words, the total tangential component of $\vec{\mathbf{E}}$ on one side of the surface must equal that on the other (Problem 4.37). Thus, since $\hat{\mathbf{u}}_n$ is the unit vector normal to the interface, regardless of the direction of the electric field within the wavefront, the cross-product of it with $\hat{\mathbf{u}}_n$ will be perpendicular to $\hat{\mathbf{u}}_n$ and therefore tangent to the interface. Hence

$$\hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_i + \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_r = \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_t \tag{4.15}$$

or

$$\begin{aligned}\hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_{0i} \cos\,(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}}\, \omega_i t) \\ + \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_{0r} \cos\,(\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} - \omega_r t + \varepsilon_r) \\ = \hat{\mathbf{u}}_n \times \vec{\mathbf{E}}_{0t} \cos\,(\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} - \omega_t t + \varepsilon_t)\end{aligned} \tag{4.16}$$

This relationship must obtain at any instant in time and at any point on the interface ($y = b$). Consequently, $\vec{\mathbf{E}}_i$, $\vec{\mathbf{E}}_r$, and $\vec{\mathbf{E}}_t$ must have precisely the same functional dependence on the variables t and r, which means that

$$\begin{aligned}(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}} - \omega_i t)|_{y=b} = (\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} - \omega_r t + \varepsilon_r)|_{y=b} \\ = (\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} - \omega_t t + \varepsilon_t)|_{y=b}\end{aligned} \tag{4.17}$$

With this as the case, the cosines in Eq. (4.16) cancel, leaving an expression independent of t and r, as indeed it must be. Inasmuch as this has to be true for all values of time, the coefficients of t must be equal, to wit

$$\omega_i = \omega_r = \omega_t \tag{4.18}$$

Recall that the electrons within the media are undergoing (linear) forced vibrations at the frequency of the incident wave. Whatever light is scattered has that same frequency. Furthermore,

$$(\vec{\mathbf{k}}_i \cdot \vec{\mathbf{r}})|_{y=b} = (\vec{\mathbf{k}}_r \cdot \vec{\mathbf{r}} + \varepsilon_r)|_{y=b} = (\vec{\mathbf{k}}_t \cdot \vec{\mathbf{r}} + \varepsilon_t)|_{y=b} \tag{4.19}$$

wherein $\vec{\mathbf{r}}$ terminates on the interface. The values of ε_r and ε_t correspond to a given position of O, and thus they allow the relation to be valid regardless of that location. (For example, the origin might be chosen such that $\vec{\mathbf{r}}$ was perpendicular to $\vec{\mathbf{k}}_i$ but not to $\vec{\mathbf{k}}_r$ or $\vec{\mathbf{k}}_t$.) From the first two terms we obtain

$$[(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r) \cdot \vec{\mathbf{r}}]_{y=b} = \varepsilon_r \tag{4.20}$$

Recalling Eq. (2.43), this expression simply says that the endpoint of $\vec{\mathbf{r}}$ sweeps out a plane (which is of course the interface) perpendicular to the vector $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$. To phrase it slightly differently, $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$ is parallel to $\hat{\mathbf{u}}_n$. Notice, however, that since the incident and reflected waves are in the same medium, $k_i = k_r$. From the fact that $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$ has no component in the plane of the interface, that is, $\hat{\mathbf{u}}_n \times (\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r) = 0$, we conclude that

$$k_i \sin\,\theta_i = k_r \sin\,\theta_r$$

Hence we have the Law of Reflection; that is,

$$\theta_i = \theta_r$$

Furthermore, since $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_r)$ is parallel to $\hat{\mathbf{u}}_n$ all three vectors, $\vec{\mathbf{k}}_i$, $\vec{\mathbf{k}}_r$, and $\hat{\mathbf{u}}_n$, are in the same plane, the plane-of-incidence. Again, from Eq. (4.19)

$$[(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_t) \cdot \vec{\mathbf{r}}]_{y=b} = \varepsilon_t \tag{4.21}$$

and therefore $(\vec{\mathbf{k}}_i - \vec{\mathbf{k}}_t)$ is also normal to the interface. Thus $\vec{\mathbf{k}}_i$, $\vec{\mathbf{k}}_r$, $\vec{\mathbf{k}}_t$, and $\hat{\mathbf{u}}_n$ are all coplanar. As before, the tangential components of $\vec{\mathbf{k}}_i$ and $\vec{\mathbf{k}}_t$ must be equal, and consequently

$$k_i \sin\,\theta_i = k_t \sin\,\theta_t \tag{4.22}$$

But because $\omega_i = \omega_t$, we can multiply both sides by c/ω_i to get

$$n_i \sin\,\theta_i = n_t \sin\,\theta_t$$

which is Snell's Law. Finally, if we had chosen the origin O to

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

be in the interface, it is evident from Eqs. (4.20) and (4.21) that ε_r and ε_t would both have been zero. That arrangement, though not as instructive, is certainly simpler, and we'll use it from here on.

4.6.2 The Fresnel Equations

We have just found the relationship that exists among the phases of $\vec{\mathbf{E}}_i(\vec{r}, t)$, $\vec{\mathbf{E}}_r(\vec{r}, t)$, and $\vec{\mathbf{E}}_t(\vec{r}, t)$ at the boundary. There is still an interdependence shared by the amplitudes $\vec{\mathbf{E}}_{0i}$, $\vec{\mathbf{E}}_{0r}$, and $\vec{\mathbf{E}}_{0t}$, which can now be evaluated. To that end, suppose that a plane monochromatic wave is incident on the planar surface separating two isotropic media. Whatever the polarization of the wave, we shall resolve its $\vec{\mathbf{E}}$- and $\vec{\mathbf{B}}$-fields into components parallel and perpendicular to the plane-of-incidence and treat these constituents separately.

Case 1: $\vec{\mathbf{E}}$ perpendicular to the plane-of-incidence. Assume that $\vec{\mathbf{E}}$ is perpendicular to the plane-of-incidence and that $\vec{\mathbf{B}}$ is parallel to it (Fig. 4.39). Recall that $E = vB$, so that

$$\hat{\mathbf{k}} \times \vec{\mathbf{E}} = v\vec{\mathbf{B}} \tag{4.23}$$

and

$$\hat{\mathbf{k}} \cdot \vec{\mathbf{E}} = 0 \tag{4.24}$$

(i.e., $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and the unit propagation vector $\hat{\mathbf{k}}$ form a right-handed system). Again, making use of the continuity of the tangential components of the $\vec{\mathbf{E}}$-field, we have at the boundary at any time and any point

$$\vec{\mathbf{E}}_{0i} + \vec{\mathbf{E}}_{0r} = \vec{\mathbf{E}}_{0t} \tag{4.25}$$

where the cosines cancel. Realize that the field vectors as shown really ought to be envisioned at $y = 0$ (i.e., at the surface), from which they have been displaced for the sake of clarity. Note too that although $\vec{\mathbf{E}}_r$ and $\vec{\mathbf{E}}_t$ must be normal to the plane-of-incidence by symmetry, *we are guessing that they point outward* at the interface when $\vec{\mathbf{E}}_i$ does. The directions of the $\vec{\mathbf{B}}$-fields then follow from Eq. (4.23).

We will need to invoke another of the boundary conditions in order to get one more equation. The presence of material substances that become electrically polarized by the wave has a definite effect on the field configuration. Thus, although the tangential component of $\vec{\mathbf{E}}$ is continuous across the boundary, its normal component is not. Instead, the normal component of the product $\epsilon\vec{\mathbf{E}}$ is the same on either side of the interface. Similarly, the normal component of $\vec{\mathbf{B}}$ is continuous, as is the tangential component of $\mu^{-1}\vec{\mathbf{B}}$. Here the magnetic effect of the two media appears via their permeabilities μ_i and μ_t. This boundary condition will be the simplest to use, particularly as applied to reflection from the surface of a conductor.* Thus the continuity of the tangential component of $\vec{\mathbf{B}}/\mu$ requires that

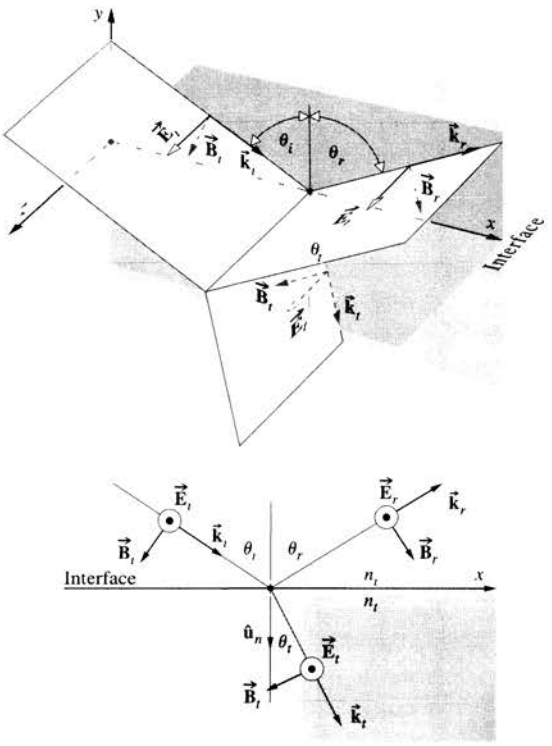


Figure 4.39 An incoming wave whose $\vec{\mathbf{E}}$-field is normal to the plane-of-incidence.

$$-\frac{B_i}{\mu_i}\cos\theta_i + \frac{B_r}{\mu_i}\cos\theta_r = -\frac{B_t}{\mu_t}\cos\theta_t \tag{4.26}$$

where the left and right sides are the total magnitudes of $\vec{\mathbf{B}}/\mu$ parallel to the interface in the incident and transmitting media, respectively. The positive direction is that of increasing x, so

*In keeping with our intent to use only the $\vec{\mathbf{E}}$- and $\vec{\mathbf{B}}$-fields, at least in the early part of this exposition, we have avoided the usual statements in terms of $\vec{\mathbf{H}}$, where

$$\vec{\mathbf{H}} = \mu^{-1}\vec{\mathbf{B}} \tag{A1.14}$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

that the scalar components of $\vec{\mathbf{B}}_i$ and $\vec{\mathbf{B}}_t$ appear with minus signs. From Eq. (4.23) we have

$$B_i = E_i/v_i \tag{4.27}$$

$$B_r = E_r/v_r \tag{4.28}$$

and

$$B_t = E_t/v_t \tag{4.29}$$

Since $v_i = v_r$ and $\theta_i = \theta_r$, Eq. (4.26) can be written as

$$\frac{1}{\mu_i v_i}(E_i - E_r)\cos\theta_i = \frac{1}{\mu_t v_t} E_t \cos\theta_t \tag{4.30}$$

Making use of Eqs. (4.12), (4.13), and (4.14) and remembering that the cosines therein equal one another at $y = 0$, we obtain

$$\frac{n_i}{\mu_i}(E_{0i} - E_{0r})\cos\theta_i = \frac{n_t}{\mu_t} E_{0t}\cos\theta_t \tag{4.31}$$

Combined with Eq. (4.25), this yields

$$\left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{\dfrac{n_i}{\mu_i}\cos\theta_i - \dfrac{n_t}{\mu_t}\cos\theta_t}{\dfrac{n_i}{\mu_i}\cos\theta_i + \dfrac{n_t}{\mu_t}\cos\theta_t} \tag{4.32}$$

and

$$\left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2\dfrac{n_i}{\mu_i}\cos\theta_i}{\dfrac{n_i}{\mu_i}\cos\theta_i + \dfrac{n_t}{\mu_t}\cos\theta_t} \tag{4.33}$$

The $\perp$ subscript serves as a reminder that we are dealing with the case in which $\vec{\mathbf{E}}$ is perpendicular to the plane-of-incidence. These two expressions, *which are completely general statements applying to any linear, isotropic, homogeneous media*, are two of the **Fresnel Equations**. Most often one deals with dielectrics for which $\mu_i \approx \mu_t \approx \mu_0$; consequently, the common form of these equations is simply

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{n_i\cos\theta_i - n_t\cos\theta_t}{n_i\cos\theta_i + n_t\cos\theta_t} \tag{4.34}$$

and

$$t_\perp \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2n_i\cos\theta_i}{n_i\cos\theta_i + n_t\cos\theta_t} \tag{4.35}$$

Here $r_\perp$ denotes the **amplitude reflection coefficient**, and $t_\perp$ is the **amplitude transmission coefficient**.

Case 2: $\vec{\mathbf{E}}$ parallel to the plane-of-incidence. A similar pair of equations can be derived when the incoming $\vec{\mathbf{E}}$-field lies in the plane-of-incidence, as shown in Fig. 4.40. Continuity of the tangential components of $\vec{\mathbf{E}}$ on either side of the boundary leads to

$$E_{0i}\cos\theta_i - E_{0r}\cos\theta_r = E_{0t}\cos\theta_t \tag{4.36}$$

In much the same way as before, continuity of the tangential components of $\vec{\mathbf{B}}/\mu$ yields

$$\frac{1}{\mu_i v_i} E_{0i} + \frac{1}{\mu_r v_r} E_{0r} = \frac{1}{\mu_t v_t} E_{0t} \tag{4.37}$$

Using the fact that $\mu_i = \mu_r$ and $\theta_i = \theta_r$, we can combine these formulas to obtain two more of the *Fresnel Equations*:

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\parallel = \frac{\dfrac{n_t}{\mu_t}\cos\theta_i - \dfrac{n_i}{\mu_i}\cos\theta_t}{\dfrac{n_i}{\mu_i}\cos\theta_i + \dfrac{n_t}{\mu_t}\cos\theta_t} \tag{4.38}$$

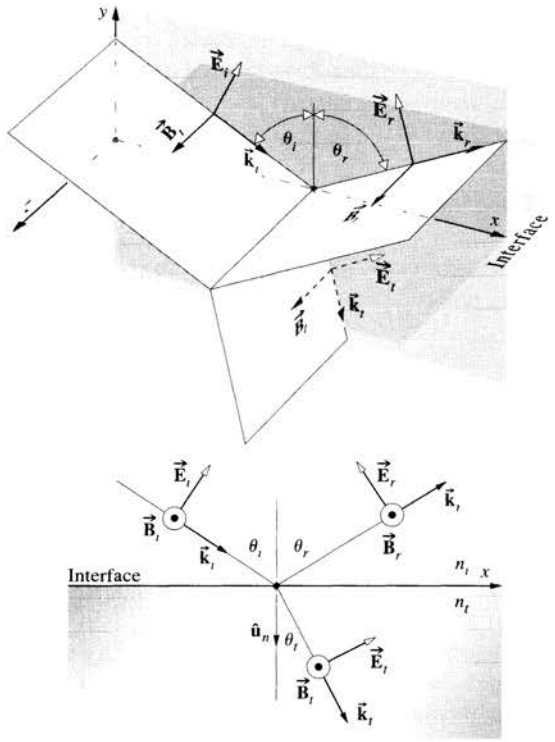


Figure 4.40 An incoming wave whose $\vec{\mathbf{E}}$-field is in the plane-of-incidence.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

and

$$t_{\parallel} \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_{\parallel} = \frac{2\,\dfrac{n_i}{\mu_i}\cos\theta_i}{\dfrac{n_i}{\mu_i}\cos\theta_t + \dfrac{n_t}{\mu_t}\cos\theta_i} \tag{4.39}$$

When both media forming the interface are dielectrics that are essentially "nonmagnetic" (p. 66), the amplitude coefficients become

$$r_{\parallel} = \frac{n_t\cos\theta_i - n_i\cos\theta_t}{n_i\cos\theta_t + n_t\cos\theta_i} \tag{4.40}$$

and

$$t_{\parallel} = \frac{2n_i\cos\theta_i}{n_i\cos\theta_t + n_t\cos\theta_i} \tag{4.41}$$

One further notational simplification can be made using Snell's Law, whereupon the Fresnel Equations for dielectric media become (Problem 4.39)

$$r_{\perp} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \tag{4.42}$$

$$r_{\parallel} = +\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \tag{4.43}$$

$$t_{\perp} = +\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i + \theta_t)} \tag{4.44}$$

$$t_{\parallel} = +\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i + \theta_t)\cos(\theta_i - \theta_t)} \tag{4.45}$$

A note of caution must be introduced here. Bear in mind that the directions (or more precisely, the phases) of the fields in Figs. 4.39 and 4.40 were selected rather arbitrarily. For example, in Fig. 4.39 we could have assumed that $\vec{\mathbf{E}}_r$ pointed inward, whereupon $\vec{\mathbf{B}}_r$ would have had to be reversed as well. Had we done that, the sign of $r_{\perp}$ would have turned out to be positive, leaving the other amplitude coefficients unchanged. The signs appearing in Eqs. (4.42) through (4.45), which are positive except for the first, correspond to the particular set of field directions selected. The minus sign in Eq. (4.42), as we will see, just means that we didn't guess correctly concerning $\vec{\mathbf{E}}_r$ in Fig. 4.39. Nonetheless, be aware that the literature is not standardized, and all possible sign variations have been labeled the *Fresnel Equations*. To avoid confusion *they must be related to the specific field directions from which they were derived.*

4.6.3 Interpretation of the Fresnel Equations

This section examines the physical implications of the Fresnel Equations. In particular, we are interested in determining the fractional amplitudes and flux densities that are reflected and refracted. In addition we shall be concerned with any possible phase shifts that might be incurred in the process.

Amplitude Coefficients

Let's briefly examine the form of the amplitude coefficients over the entire range of θ_i values. At nearly normal incidence ($\theta_i \approx 0$) the tangents in Eq. (4.43) are essentially equal to sines, in which case

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \left[\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}\right]_{\theta_i=0}$$

We will come back to the physical significance of the minus sign presently. After expanding the sines and using Snell's Law, this expression becomes

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \left[\frac{n_t\cos\theta_i - n_i\cos\theta_t}{n_t\cos\theta_i + n_i\cos\theta_t}\right]_{\theta_i=0} \tag{4.46}$$

which follows as well from Eqs. (4.34) and (4.40). In the limit, as θ_i goes to 0, $\cos\theta_i$ and $\cos\theta_t$ both approach one, and consequently

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \frac{n_t - n_i}{n_t + n_i} \tag{4.47}$$

This equality of the reflection coefficients arises because the plane-of-incidence is no longer specified when $\theta_t = 0$. Thus, for example, at an air ($n_i = 1$) glass ($n_t = 1.5$) interface at nearly normal incidence, the amplitude reflection coefficients equal ± 0.2. (See Problem 4.45.)

When $n_t > n_i$ it follows from Snell's Law that $\theta_i > \theta_t$, and $r_{\perp}$ is negative for all values of θ_i (Fig. 4.41). In contrast, Eq. (4.43) tells us that $r_{\parallel}$ starts out positive at $\theta_i = 0$ and decreases gradually until it equals zero when $(\theta_i + \theta_t) = 90°$, since there $\tan\pi/2$ is infinite. The particular value of the incident angle for which this occurs is denoted by θ_p and referred to as the **polarization angle** (see Section 8.6.1). Notice that $r_{\parallel} \to 0$ at θ_p, just when the phase shifts 180°. That means we won't see the $\vec{\mathbf{E}}$-field do any flipping when θ_i approaches θ_p from either side. As θ_i increases beyond θ_p, $r_{\parallel}$ becomes progressively more negative, reaching -1.0 at 90°.

If you place a single sheet of glass, a microscope slide, on this page and look straight down into it ($\theta_i = 0$), the region

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

beneath the glass will seem decidedly grayer than the rest of the paper, because the slide will reflect at both its interfaces, and the light reaching and returning from the paper will be diminished appreciably. Now hold the slide near your eye and again view the page through it as you tilt it, increasing θ_i. The amount of light reflected will increase, and it will become more difficult to see the page through the glass. When $\theta_i \approx 90°$ the slide will look like a perfect mirror as the reflection coefficients (Fig. 4.41) go to -1.0. Even a rather poor surface (see photo), such as the cover of this book, will be mirrorlike at glancing incidence. Hold the book horizontally at the level of the middle of your eye and face a bright light; you will see the source reflected rather nicely in the cover. This suggests that even X-rays could be mirror-reflected at glancing incidence (p. 242), and modern X-ray telescopes are based on that very fact.

At normal incidence Eqs. (4.35) and (4.41) lead rather straightforwardly to

$$[t_\parallel]_{\theta_i=0} = [t_\perp]_{\theta_i=0} = \frac{2n_i}{n_i + n_t} \tag{4.48}$$

It will be shown in Problem 4.50 that the expression

$$t_\perp + (-r_\perp) = 1 \tag{4.49}$$

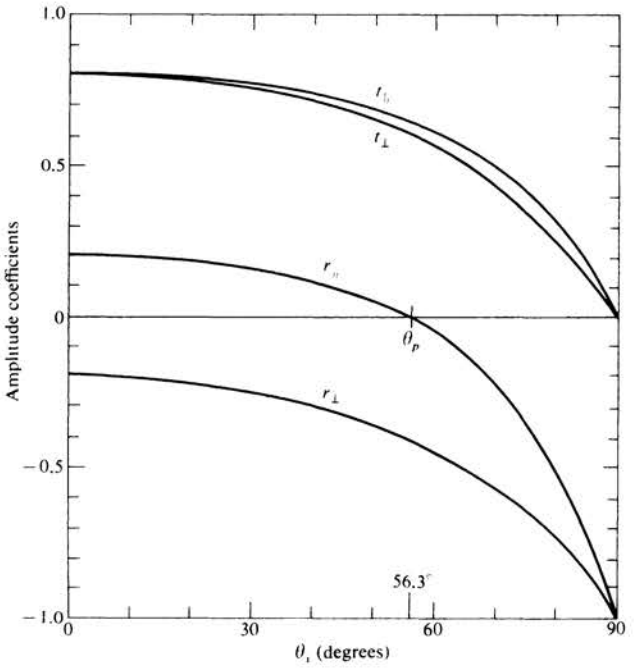


Figure 4.41 The amplitude coefficients of reflection and transmission as a function of incident angle. These correspond to external reflection $n_t > n_i$ at an air–glass interface ($n_{ti} = 1.5$).



At near-glancing incidence the walls and floor are mirrorlike—this despite the fact that the surfaces are rather poor reflectors at $\theta_i = 0°$. (Photo by E.H.)

holds for all θ_i, whereas

$$t_\parallel + r_\parallel = 1 \tag{4.50}$$

is true only at normal incidence.

The foregoing discussion, for the most part, was restricted to the case of **external reflection** (i.e., $n_t > n_i$). The opposite situation of **internal reflection**, in which the incident medium is the more dense ($n_i > n_t$), is of interest as well. In that instance $\theta_t > \theta_i$, and $r_\perp$, as described by Eq. (4.42), will always be positive. Figure 4.42 shows that $r_\perp$ increases from its initial value [Eq. (4.47)] at $\theta_i = 0$, reaching $+1$ at what is called the **critical angle**, θ_c. Specifically, θ_c is the special value of the incident angle (p. 122) for which $\theta_t = \pi/2$. Likewise, $r_\parallel$ starts off negatively [Eq. (4.47)] at $\theta_i = 0$ and thereafter increases, reaching $+1$ at $\theta_i = \theta_c$, as is evident from the Fresnel Equation (4.40). Again, $r_\parallel$ passes through zero at the *polarization angle* θ_p'. It is left for Problem 4.66 to show that the polarization angles θ_p' and θ_p for internal and external reflection at the interface between the same media are simply the complements of each other. We will return to internal reflection in Section 4.7, where it will be shown that $r_\perp$ and $r_\parallel$ are complex quantities for $\theta_i > \theta_c$.

Phase Shifts

It should be evident from Eq. (4.42) that $r_\perp$ is negative regardless of θ_i when $n_t > n_i$. Yet we saw earlier that had we chosen $[\vec{\mathbf{E}}_r]_\perp$ in Fig. 4.37 to be in the opposite direction, the first Fres-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

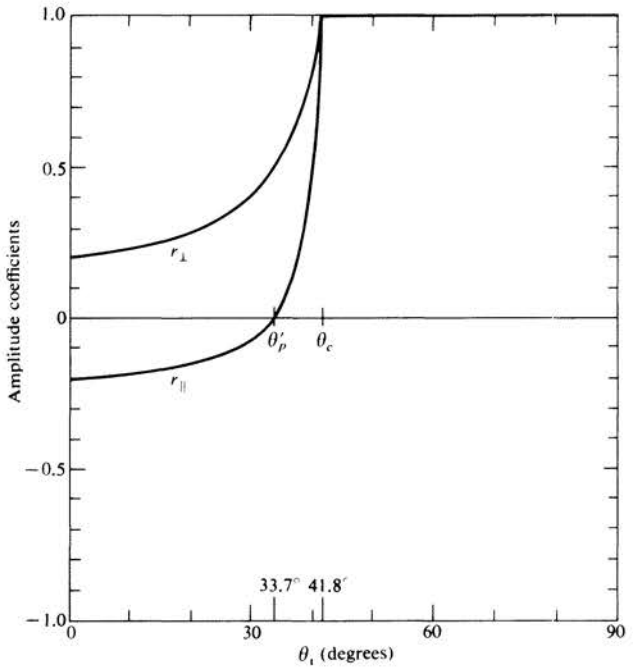


Figure 4.42 The amplitude coefficients of reflection as a function of incident angle. These correspond to internal reflection $n_t < n_i$ at an air–glass interface ($n_{ti} = 1/1.5$).

nel Equation (4.42) would have changed signs, causing $r_\perp$ to become a positive quantity. The sign of $r_\perp$ is associated with the relative directions of $[\vec{\mathbf{E}}_{0i}]_\perp$ and $[\vec{\mathbf{E}}_{0r}]_\perp$. Bear in mind that a reversal of $[\vec{\mathbf{E}}_{0r}]_\perp$ is tantamount to introducing a phase shift, $\Delta\varphi_\perp$, of π radians into $[\vec{\mathbf{E}}_r]_\perp$. Hence at the boundary $[\vec{\mathbf{E}}_t]_\perp$ and $[\vec{\mathbf{E}}_r]_\perp$ will be antiparallel and therefore π out-of-phase with each other, as indicated by the negative value of $r_\perp$. When we consider components normal to the plane-of-incidence, there is no confusion as to whether two fields are in-phase or π radians out-of-phase: if parallel, they're in-phase; if antiparallel, they're π out-of-phase. In summary, then, ***the component of the electric field normal to the plane-of-incidence undergoes a phase shift of π radians upon reflection when the incident medium has a lower index than the transmitting medium***. Similarly, $t_\perp$ and $t_\parallel$ are always positive and $\Delta\varphi = 0$. *Furthermore, when $n_i > n_t$ no phase shift in the normal component results on reflection, that is, $\Delta\varphi_\perp = 0$ so long as $\theta_i < \theta_c$.*

Things are a bit less obvious when we deal with $[\vec{\mathbf{E}}_i]_\parallel$, $[\vec{\mathbf{E}}_r]_\parallel$, and $[\vec{\mathbf{E}}_t]_\parallel$. It now becomes necessary to define more explicitly what is meant by *in-phase*, since the field vectors are coplanar but generally not colinear. The field directions were chosen in Figs. 4.39 and 4.40 such that if you looked down any one of the propagation vectors toward the direction from which the light was coming, $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{k}}$ would appear to have the same relative orientation whether the ray was incident, reflected, or transmitted. We can use this as the required condition for two **E**-fields to be in-phase. Equivalently, but more simply, ***two fields in the incident plane are in-phase if their y-components are parallel and are out-of-phase if the components are antiparallel***. Notice that when two $\vec{\mathbf{E}}$-fields are out-of-phase so too are their associated $\vec{\mathbf{B}}$-fields and vice versa. With this definition we need only look at the vectors normal to the plane-of-incidence, whether they be $\vec{\mathbf{E}}$ or $\vec{\mathbf{B}}$, to determine the relative phase of the accompanying fields in the incident plane. Thus in Fig. 4.43*a* $\vec{\mathbf{E}}_i$ and $\vec{\mathbf{E}}_t$ are in-phase, as are $\vec{\mathbf{B}}_i$ and $\vec{\mathbf{B}}_t$, whereas $\vec{\mathbf{E}}_i$ and $\vec{\mathbf{E}}_r$ are out-of-phase, along with $\vec{\mathbf{B}}_i$ and $\vec{\mathbf{B}}_r$. Similarly, in Fig. 4.43*b* $\vec{\mathbf{E}}_i$, $\vec{\mathbf{E}}_r$, and $\vec{\mathbf{E}}_t$ are in-phase, as are $\vec{\mathbf{B}}_i$, $\vec{\mathbf{B}}_r$, and $\vec{\mathbf{B}}_t$.

Now, the amplitude reflection coefficient for the parallel component is given by

$$r_\parallel = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t}$$

which is positive ($\Delta\varphi_\parallel = 0$) as long as

$$n_t \cos\theta_i - n_i \cos\theta_t > 0$$

that is, if

$$\sin\theta_t \cos\theta_t - \cos\theta_t \sin\theta_t > 0$$

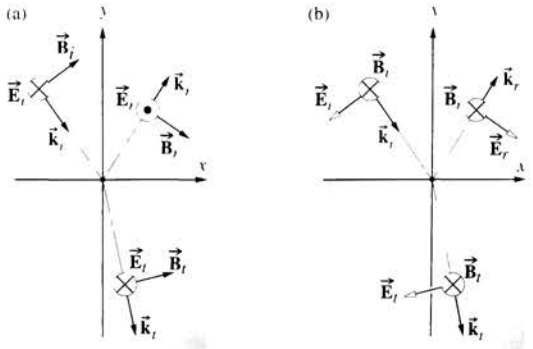


Figure 4.43 Field orientations and phase shifts.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

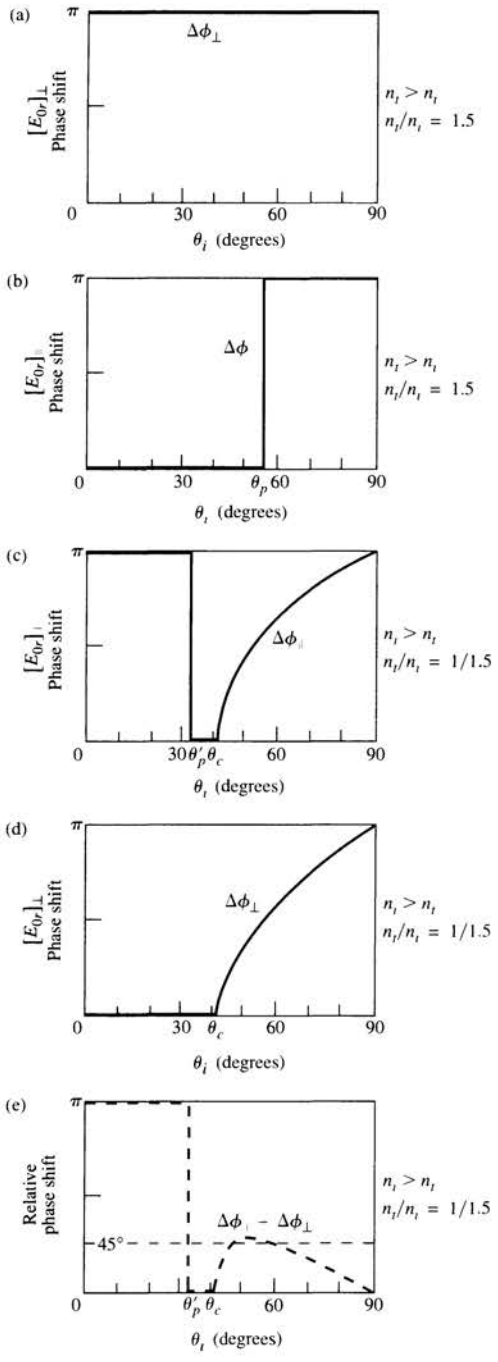


Figure 4.44 Phase shifts for the parallel and perpendicular components of the $\vec{\mathbf{E}}$-field corresponding to internal and external reflection.

or equivalently

$$\sin(\theta_i - \theta_t)\cos(\theta_t + \theta_t) > 0 \tag{4.51}$$

This will be the case for $n_i < n_t$ if

$$(\theta_i + \theta_t) < \pi/2 \tag{4.52}$$

and for $n_i > n_t$ when

$$(\theta_i + \theta_t) > \pi/2 \tag{4.53}$$

Thus when $n_i < n_t$, $[\vec{\mathbf{E}}_{0r}]_\parallel$ and $[\vec{\mathbf{E}}_{0i}]_\parallel$ will be in-phase ($\Delta\varphi_\parallel = 0$) until $\theta_i = \theta_p$ and out-of-phase by π radians thereafter. The transition is not actually discontinuous, since $[\vec{\mathbf{E}}_{0r}]_\parallel$ goes to zero at θ_p. In contrast, for internal reflection $r_\parallel$ is negative until θ'_p, which means that $\Delta\varphi_\parallel = \pi$. From θ'_p to θ_c, $r_\parallel$ is positive and $\Delta\varphi_\parallel = 0$. Beyond θ_c, $r_\parallel$ becomes complex, and $\Delta\varphi_\parallel$ gradually increases to π at $\theta_i = 90°$.

Figure 4.44, which summarizes these conclusions, will be of continued use to us. The actual functional form of $\Delta\varphi_\parallel$ and $\Delta\varphi_\perp$ for internal reflection in the region where $\theta_i > \theta_c$ can be found in the literature,* but the curves depicted here will suffice for our purposes. Figure 4.44*e* is a plot of the relative phase shift between the parallel and perpendicular components, that is, $\Delta\varphi_\parallel - \Delta\varphi_\perp$. It is included here because it will be useful later on (e.g., when we consider polarization effects). Finally, many of the essential features of this discussion are illustrated in Figs. 4.45 and 4.46. The amplitudes of the reflected vectors are in accord with those of Figs. 4.41 and 4.42 (for an air–glass interface), and the phase shifts agree with those of Fig. 4.44.

Many of these conclusions can be verified with the simplest experimental equipment, namely, two linear polarizers, a piece of glass, and a small source, such as a flashlight or high-intensity lamp. By placing one polarizer in front of the source (at 45° to the plane-of-incidence), you can easily duplicate the conditions of Fig. 4.45. For example, when $\theta_i = \theta_p$ (Fig. 4.45*b*) no light will pass through the second polarizer if its transmission axis is parallel to the plane-of-incidence. In comparison, at near-glancing incidence the reflected beam will vanish when the axes of the two polarizers are almost normal to each other.

*Born and Wolf, *Principles of Optics*, p. 49.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

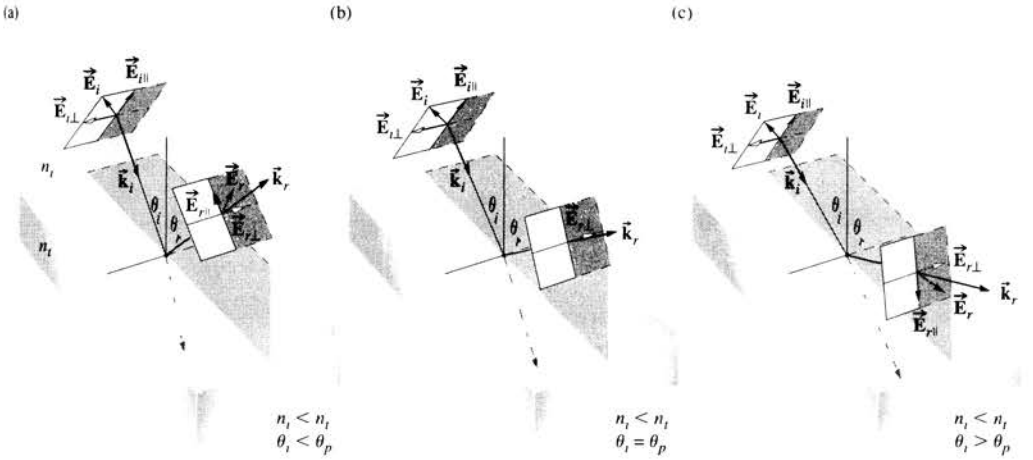


Figure 4.45 The reflected $\vec{\mathbf{E}}$-field at various angles concomitant with external reflection.

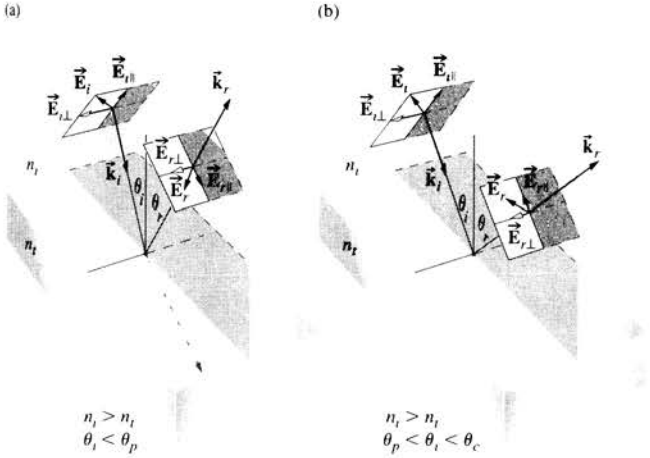


Figure 4.46 The reflected $\vec{\mathbf{E}}$-field at various angles concomitant with internal reflection.

Reflectance and Transmittance

Consider a circular beam of light incident on a surface, as shown in Fig. 4.47, such that there is an illuminated spot of area A. Recall that the power per unit area crossing a surface in vacuum whose normal is parallel to $\vec{\mathbf{S}}$, the Poynting vector, is given by

$$\vec{\mathbf{S}} = c^2\epsilon_0\vec{\mathbf{E}} \times \vec{\mathbf{B}} \qquad [3.40]$$

Furthermore, the radiant flux density (W/m^2) or irradiance is

$$I = \langle S \rangle_T = \frac{c\epsilon_0}{2} E_0^2 \qquad [3.44]$$

This is the average energy per unit time crossing a unit area normal to $\vec{\mathbf{S}}$ (in isotropic media $\vec{\mathbf{S}}$ is parallel to $\vec{\mathbf{k}}$). In the case at hand (Fig. 4.47), let I_i, I_r, and I_t be the incident, reflected, and transmitted flux densities, respectively. The cross-sectional areas of the incident, reflected, and transmitted beams are, respectively, $A\cos\theta_i$, $A\cos\theta_r$, and $A\cos\theta_t$. Accordingly, the incident power is $I_iA\cos\theta_i$; this is the energy per unit time flowing in the incident beam, and it's therefore the power

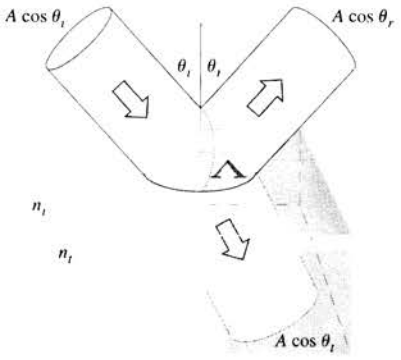


Figure 4.47 Reflection and transmission of an incident beam.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

arriving on the surface over A. Similarly, $I_r A \cos \theta_r$ is the power in the reflected beam, and $I_t A \cos \theta_t$ is the power being transmitted through A. We define the **reflectance** R to be the ratio of the reflected power (or flux) to the incident power:

$$R \equiv \frac{I_r A \cos \theta_r}{I_i A \cos \theta_i} = \frac{I_r}{I_i} \tag{4.54}$$

In the same way, the **transmittance** T is defined as the ratio of the transmitted to the incident flux and is given by

$$T \equiv \frac{I_t \cos \theta_t}{I_i \cos \theta_i} \tag{4.55}$$

The quotient I_r/I_i equals $(v_r \epsilon_r E_{0r}^2/2)/(v_i \epsilon_i E_{0i}^2/2)$, and since the incident and reflected waves are in the same medium, $v_r = v_i$, $\epsilon_r = \epsilon_i$, and

$$R = \left(\frac{E_{0r}}{E_{0i}}\right)^2 = r^2 \tag{4.56}$$

In like fashion (assuming $\mu_i = \mu_t = \mu_0$),

$$T = \frac{n_t \cos \theta_t}{n_i \cos \theta_i}\left(\frac{E_{0t}}{E_{0i}}\right)^2 = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right) t^2 \tag{4.57}$$

where use was made of the fact that $\mu_0 \epsilon_t = 1/v_t^2$ and $\mu_0 v_t \epsilon_t = n_t/c$. Notice that at normal incidence, which is a situation of great practical interest, $\theta_t = \theta_i = 0$, and the transmittance [Eq. (4.55)], like the reflectance [Eq. (4.54)], is then simply the ratio of the appropriate irradiances. Since $R = r^2$, we need not worry about the sign of r in any particular formulation, and that makes reflectance a convenient notion. Observe that in Eq. (4.57) T is not simply equal to t^2, for two reasons. First, the ratio of the indices of refraction must be there, since the speeds at which energy is transported into and out of the interface are different, in other words, $I \propto v$, from Eq. (3.47). Second, the cross-sectional areas of the incident and refracted beams are different. The energy flow per unit area is affected accordingly, and that manifests itself in the presence of the ratio of the cosine terms.

Let's now write an expression representing the conservation of energy for the configuration depicted in Fig. 4.47. In other words, the total energy flowing into area A per unit time must equal the energy flowing outward from it per unit time:

$$I_i A \cos \theta_i = I_r A \cos \theta_i + I_t A \cos \theta_t \tag{4.58}$$

When both sides are multiplied by c, this expression becomes

$$n_i E_{0i}^2 \cos \theta_i = n_i E_{0r}^2 \cos \theta_i + n_i E_{0t}^2 \cos \theta_t$$

or

$$1 = \left(\frac{E_{0r}}{E_{0i}}\right)^2 + \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right)\left(\frac{E_{0t}}{E_{0i}}\right)^2 \tag{4.59}$$

But this is simply

$$R + T = 1 \tag{4.60}$$

where there was no absorption. It is convenient to use the component forms, that is,

$$R_\perp = r_\perp^2 \tag{4.61}$$

$$R_\parallel = r_\parallel^2 \tag{4.62}$$

$$T_\perp = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right) t_\perp^2 \tag{4.63}$$

and

$$T_\parallel = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i}\right) t_\parallel^2 \tag{4.64}$$

which are illustrated in Fig. 4.48. Furthermore, it can be shown (Problem 4.71) that

$$R_\parallel + T_\parallel = 1 \tag{4.65}$$

and

$$R_\perp + T_\perp = 1 \tag{4.66}$$

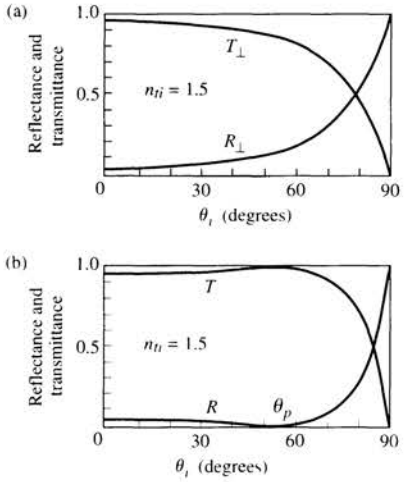


Figure 4.48 Reflectance and transmittance versus incident angle.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Looking down into a puddle (that's melting snow on the right) we see a reflection of the surrounding trees. At normal incidence water reflects about 2% of the light. As the viewing angle increases here its about 40° that percentage increases. (Photo by E.H.)

When $\theta_i = 0$, the incident plane becomes undefined, and any distinction between the parallel and perpendicular components of R and T vanishes. In this case Eqs. (4.61) through (4.64), along with (4.47) and (4.48), lead to

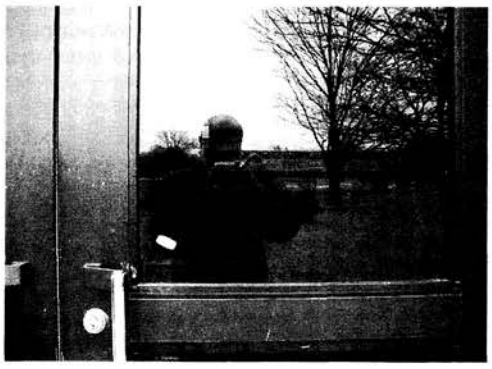
$$R = R_{\parallel} = R_{\perp} = \left(\frac{n_t - n_i}{n_t + n_i}\right)^2 \tag{4.67}$$

and

$$T = T_{\parallel} = T_{\perp} = \frac{4n_t n_i}{(n_t + n_i)^2} \tag{4.68}$$

Thus 4% of the light incident normally on an air–glass ($n_g = 1.5$) interface will be reflected back, whether internally, $n_i > n_t$, or externally, $n_i < n_t$ (Problem 4.72). This will be of concern to anyone who is working with a complicated lens system, which might have 10 or 20 such air–glass boundaries. Indeed, if you look perpendicularly into a stack of about 50 microscope slides (cover-glass sliders are much thinner and easier to handle in large quantities), most of the light will be reflected. The stack will look very much like a mirror (see photo). Roll up a thin sheet of clear plastic into a multiturned cylinder and it too will look like shiny metal. The many interfaces produce a large number of closely spaced *specular* reflections that send much of the light back into the incident medium, more or less, as if it had undergone a single frequency-independent reflection. A smooth gray-metal surface does pretty much the same thing —it has a large, frequency-independent specular reflectance—and looks shiny (that's what "shiny" is). If the reflection is diffuse, the surface will appear gray or even white if the reflectance is large enough.

Figure 4.49 is a plot of the reflectance at a single interface, assuming normal incidence for various transmitting media in air. Figure 4.50 depicts the corresponding dependence of the transmittance at normal incidence on the number of interfaces and the index of the medium. Of course, this is why you can't see through a roll of "clear" smooth-surfaced plastic tape, and it's also why the many elements in a periscope must be coated with antireflection films (Section 9.9.2).



At near normal incidence about 4% of the light is reflected back off each air-glass interface. Here because it's a lot brighter outside than inside the building, you have no trouble seeing the photographer. (Photo by E.H.)



Near normal reflection off a stack of microscope slides. You can see the image of the camera that took the picture. (Photo by E.H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

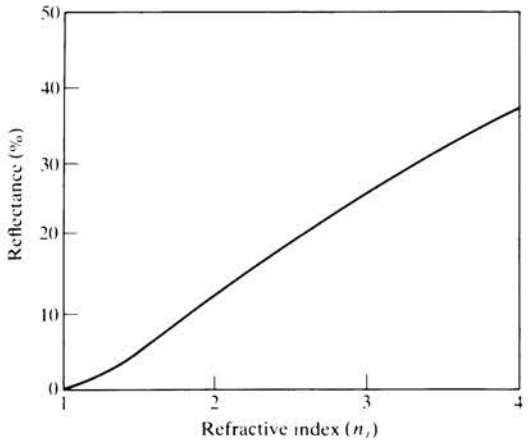


Figure 4.49 Reflectance at normal incidence in air ($n_i = 1.0$) at a single interface.

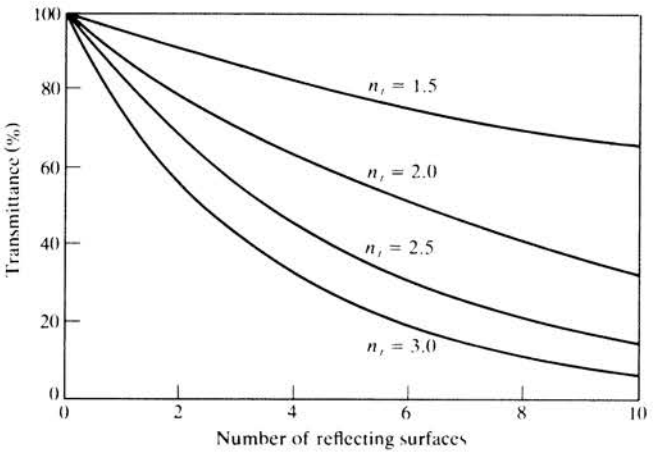


Figure 4.50 Transmittance through a number of surfaces in air ($n_i = 1.0$) at normal incidence.

4.7 Total Internal Reflection

In the previous section it was evident that something rather interesting was happening in the case of internal reflection ($n_i > n_t$) when θ_i was equal to or greater than θ_c, the so-called **critical angle**. Let's now return to that situation for a closer look. Suppose that we have a source embedded in an optically dense medium, and we allow θ_i to increase gradually, as indicated in Fig. 4.51. We know from the preceding section (Fig. 4.42) that $r_\parallel$ and $r_\perp$ increase with increasing θ_i, and therefore $t_\parallel$ and $t_\perp$ both decrease. Moreover $\theta_t > \theta_i$, since

$$\sin \theta_t = \frac{n_t}{n_i} \sin \theta_t$$

and $n_t > n_t$, in which case $n_{ti} < 1$. Thus as θ_i becomes larger, the transmitted ray gradually approaches tangency with the boundary, and as it does more and more of the available energy appears in the reflected beam. Finally, when $\theta_t = 90°$, $\sin \theta_t = 1$ and

$$\sin \theta_c = n_{ti} \tag{4.69}$$

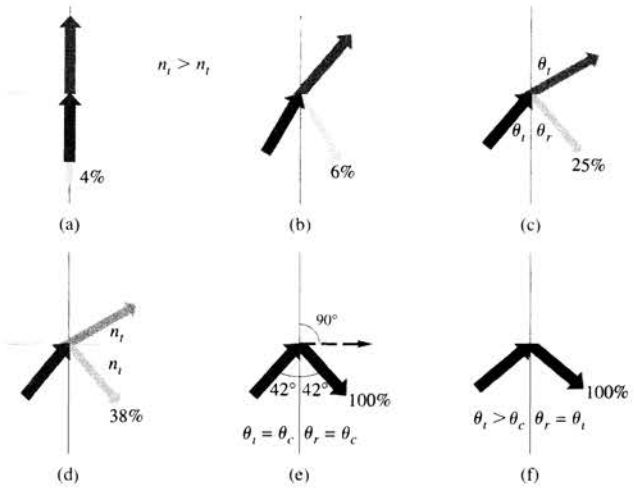
As noted earlier, *the critical angle is that special value of* θ_i *for which* $\theta_t = 90°$. The larger n_i is, the smaller n_{ti} is, and the smaller θ_c is. For incident angles greater than or equal to θ_c, all the incoming energy is reflected back into the incident medium in the process known as **total internal reflection** (see photo).

It should be stressed that the transition from the conditions depicted in Fig. 4.51*a* to those of 4.51*d* takes place without any discontinuities. As θ_i becomes larger, the reflected beam grows stronger and stronger while the transmitted beam grows weaker, until the latter vanishes and the former carries off all the energy at $\theta_r = \theta_c$. It's an easy matter to observe the diminution of the transmitted beam as θ_i is made larger. Just place a glass microscope slide on a printed page, this time blocking out any specularly reflected light. At $\theta_i \approx 0$, θ_t is roughly zero, and the page as seen through the glass is fairly bright and clear. But if you move your head, allowing θ_t (the angle at which you view the interface) to increase, the region



Notice that you can't see the two front flames through the water along a bright horizontal band. That's due to total internal reflection. Look at the bottom of a drinking galss through its side. Now add a few inches of water. What happens? (Photo by E.H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



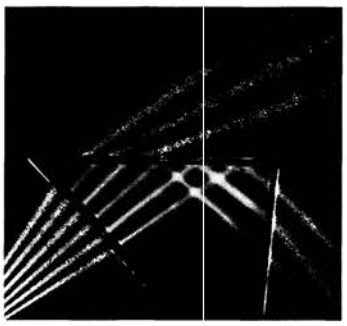


Figure 4.51 Internal reflection and the critical angle. (Photo courtesy of Educational Service, Inc.)

of the printed page covered by the glass will appear darker and darker, indicating that T has indeed been markedly reduced.

The critical angle for our air–glass interface is roughly 42° (see Table 4.2). Consequently, a ray incident normally on the left face of either of the prisms in Fig. 4.52 will have a $\theta_i > 42°$ and therefore be internally reflected. This is a convenient way to reflect nearly 100% of the incident light without having to worry about the deterioration that can occur with metallic surfaces (see photo).

Another useful way to view the situation is via Fig. 4.53, which shows a simplified representation of scattering off atomic oscillators. We know that the net effect of the presence of the homogeneous isotropic media is to alter the speed of the light from c to v_i and v_t, respectively (p. 92). The resultant wave is the superposition of these wavelets propagating at the appropriate speeds. In Fig. 4.53*a* an incident wave results in the emission of wavelets successively from scattering centers A and B. These overlap to form the transmitted wave. The reflected wave, which comes back down into the incident medium as usual ($\theta_i = \theta_r$), is not shown. In a time t the incident front travels a distance $v_i t = \overline{CB}$, while the transmitted front moves a distance $v_t t = \overline{AD} > \overline{CB}$. Since one wave moves from A to E in the same time that the other moves from

TABLE 4.2 Critical Angles

n_{tt}	θ_c (degrees)	θ_c (radians)	n_{tt}	θ_c (degrees)	θ_c (radians)
1.30	50.284 9	0.877 6	1.50	41.810 3	0.729 7
1.31	49.761 2	0.868 5	1.51	41.471 8	0.723 8
1.32	49.250 9	0.859 6	1.52	41.139 5	0.718 0
1.33	48.753 5	0.850 9	1.53	40.813 2	0.712 3
1.34	48.268 2	0.842 4	1.54	40.492 7	0.706 7
1.35	47.794 6	0.834 2	1.55	40.177 8	0.701 2
1.36	47.332 1	0.826 1	1.56	39.868 3	0.695 8
1.37	46.880 3	0.818 2	1.57	39.564 2	0.690 5
1.38	46.438 7	0.810 5	1.58	39.265 2	0.685 3
1.39	46.007 0	0.803 0	1.59	38.971 3	0.680 2
1.40	45.584 7	0.795 6	1.60	38.682 2	0.675 1
1.41	45.171 5	0.788 4	1.61	38.397 8	0.670 2
1.42	44.767 0	0.781 3	1.62	38.118 1	0.665 3
1.43	44.370 9	0.774 4	1.63	37.842 8	0.660 5
1.44	43.983 0	0.767 6	1.64	37.571 9	0.655 8
1.45	43.602 8	0.761 0	1.65	37.305 2	0.651 1
1.46	43.230 2	0.754 5	1.66	37.042 7	0.646 5
1.47	42.864 9	0.748 1	1.67	36.784 2	0.642 0
1.48	42.506 6	0.741 9	1.68	36.529 6	0.637 6
1.49	42.155 2	0.735 7	1.69	36.278 9	0.633 2

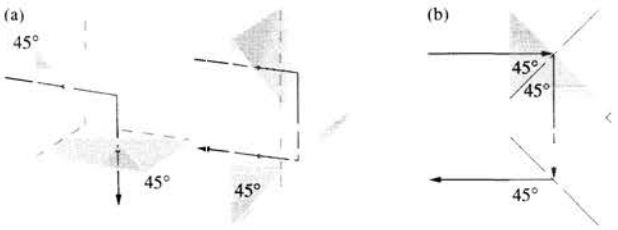


Figure 4.52 Total internal reflection.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

C to B, and since they have the same frequency and period, they must change phase by the same amount in the process. Thus the disturbance at point E must be in-phase with that at point B; both of these points must be on the same transmitted wavefront (remember Section 4.4.2).

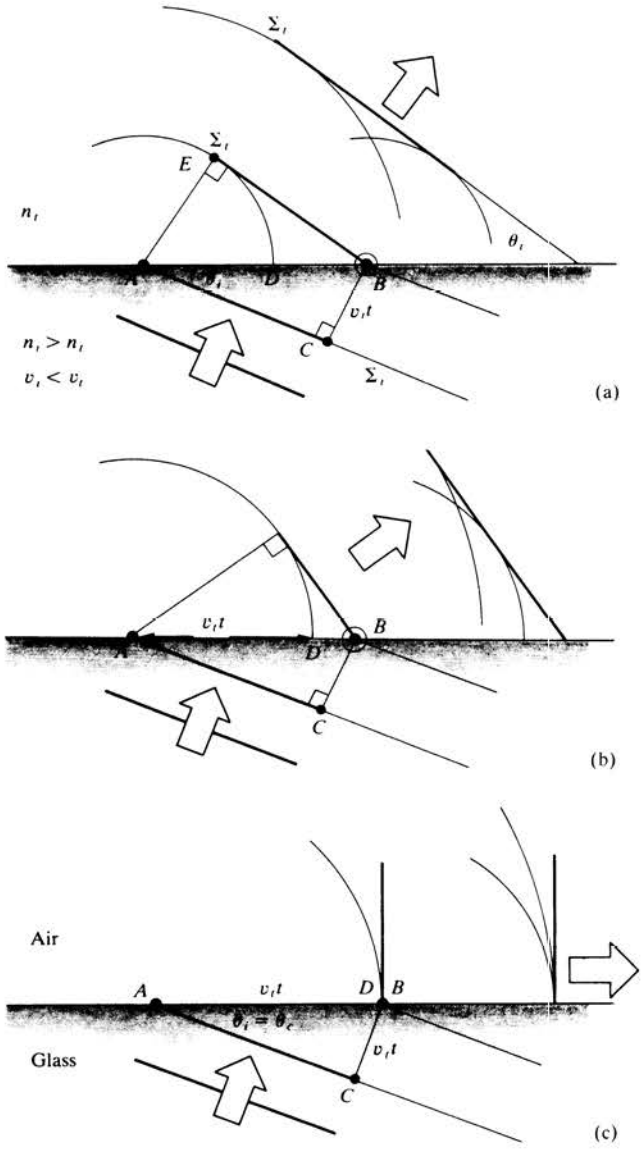


Figure 4.53 An examination of the transmitted wave in the process of total internal reflection from a scattering perspective. Here we keep θ_i and n_i constant and in successive parts of the diagram decrease n_t, thereby increasing v_t. The reflected wave ($\theta_r = \theta_i$) is not drawn.



The prism behaves like a mirror and reflects a portion of the pencil (reversing the lettering on it). The operating process is total internal reflection. (Photo by E.H.)

It can be seen that the greater v_t is in comparison to v_i, the more tilted the transmitted front will be (i.e., the larger θ_t will be). That much is depicted in Fig. 4.53*b*, where n_{ti} has been taken to be smaller by assuming n_t to be smaller. The result is a higher speed v_t, increasing $\overline{AD}$ and causing a greater transmission angle. In Fig. 4.53*c* a special case is reached: $\overline{AD} = \overline{AB} = v_t t$, and the wavelets will overlap in-phase *only along the line of the interface*, $\theta_t = 90°$. From triangle ABC, $\sin \theta_i = v_i t / v_t t = n_t / n_i$, which is Eq. (4.69). For the two given media (i.e., for the particular value of n_{ti}), the direction in which the scattered wavelets will add constructively in the transmitting medium is along the interface. The resulting disturbance ($\theta_t = 90°$) is known as a *surface wave*.

4.7.1 The Evanescent Wave

Because the frequency of X-rays is higher than the resonance frequencies of the atoms of the medium, Eq. (3.70) suggests, and experiments confirm, that the index of refraction of X-rays is less than 1.0. Thus the wave velocity of X-rays (i.e., the phase speed) in matter exceeds its value (c) in vacuum, although it usually does so by less than 1 part in 10 000, even in the densest solids. When X-rays traveling in air enter a dense material like glass, the beam bends ever so slightly *away* from the normal rather than toward it. With the above discussion of total internal reflection in mind, we should expect that X-rays will be **totally "externally" reflected** when, for example, $n_i = n_{air}$ and $n_t = n_{glass}$. This is the way it's often spoken of in the literature, but that's a misnomer; since for X-rays $n_{air} > n_{glass}$ and therefore $n_i > n_t$ (even though glass is physically more dense than air), the process is

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

actually still *internal* reflection. In any event, because n_t is less than, but very nearly equal to, 1 the index ratio $n_{ti} \approx 1$ and $\theta_t \approx 90°$.

In 1923 A. H. Compton reasoned that even though X-rays incident on a sample at ordinary angles are not specularly reflected, they should be totally "externally" reflected at glancing incidence. He shined 0.128 nm X-rays on a glass plate and got a critical angle of about 10 minutes of arc (0.167°) measured with respect to the surface. That yielded an index of refraction for glass that differed from 1 by -4.2×10^{-6}.

We'll come back to some important practical applications of both total internal and total "external" reflection later on (p. 200).

If we assume in the case of total internal reflection that there is no transmitted wave, it becomes impossible to satisfy the boundary conditions using only the incident and reflected waves—things are not at all as simple as they might seem. Furthermore, we can reformulate Eqs. (4.34) and (4.40) (Problem 4.75) such that

$$r_\perp = \frac{\cos\theta_i - (n_{ti}^2 - \sin^2\theta_i)^{1/2}}{\cos\theta_i + (n_{ti}^2 - \sin^2\theta_i)^{1/2}} \tag{4.70}$$

and

$$r_\parallel = \frac{n_{ti}^2\cos\theta_i - (n_{ti}^2 - \sin^2\theta_i)^{1/2}}{n_{ti}^2\cos\theta_i + (n_{ti}^2 - \sin^2\theta_i)^{1/2}} \tag{4.71}$$

Since $\sin\theta_c = n_{ti}$ when $\theta_i > \theta_c$, $\sin\theta_i > n_{ti}$, and both $r_\perp$ and $r_\parallel$ become complex quantities. Despite this (Problem 4.76), $r_\perp r_\perp^* = r_\parallel r_\parallel^* = 1$ and $R = 1$, which means that $I_r = I_i$ and $I_t = 0$. Thus, although there must be a transmitted wave, it cannot, on the average, carry energy across the boundary. We shall not perform the complete and rather lengthy computation needed to derive expressions for all the reflected and transmitted fields, but we can get an appreciation of what's happening in the following way. The wavefunction for the transmitted electric field is

$$\vec{\mathbf{E}}_t = \vec{\mathbf{E}}_{0t}\exp i(\vec{\mathbf{k}}_t\cdot\vec{\mathbf{r}} - \omega t)$$

where

$$\vec{\mathbf{k}}_t\cdot\vec{\mathbf{r}} = k_{tx}x + k_{ty}y$$

there being no z-component of $\vec{\mathbf{k}}$. But

$$k_{tx} = k_t\sin\theta_t$$

and

$$k_{ty} = k_t\cos\theta_t$$

as seen in Fig. 4.54. Once again using Snell's Law,

$$k_t\cos\theta_t = \pm k_t\left(1 - \frac{\sin^2\theta_i}{n_{ti}^2}\right)^{1/2} \tag{4.72}$$

or, since we are concerned with the case where $\sin\theta_i > n_{ti}$,

$$k_{ty} = \pm ik_t\left(\frac{\sin^2\theta_i}{n_{ti}^2} - 1\right)^{1/2} \equiv \pm i\beta$$

and

$$k_{tx} = \frac{k_t}{n_{ti}}\sin\theta_i$$

Hence

$$\vec{\mathbf{E}}_t = \vec{\mathbf{E}}_{0t}e^{\mp\beta y}e^{i(k_t x\sin\theta_i/n_{ti} - \omega t)} \tag{4.73}$$

Neglecting the positive exponential, which is physically untenable, we have a wave whose amplitude drops off exponentially as it penetrates the less dense medium. The disturbance advances in the x-direction as a surface or **evanescent wave**. Notice that the wavefronts or surfaces of constant phase (parallel to the yz-plane) are perpendicular to the surfaces of constant amplitude (parallel to the xz-plane), and as such the wave is *inhomogeneous* (p. 26). Its amplitude decays rapidly in the y-direction, becoming negligible at a distance into the second medium of only a few wavelengths.

If you are still concerned about the conservation of energy, a more extensive treatment would have shown that energy actually circulates back and forth across the interface, resulting on the average in a zero net flow through the boundary into the second medium. Yet one puzzling point remains, inas-

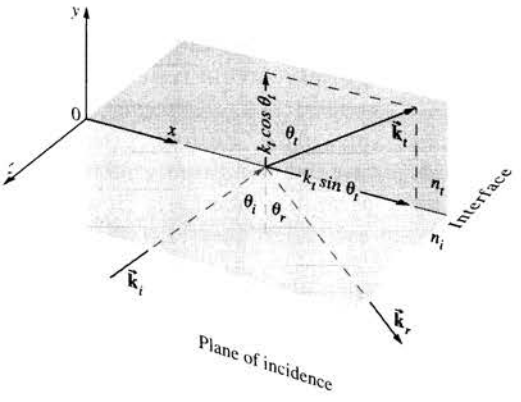


Figure 4.54 Propagation vectors for internal reflection.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

much as there is still a bit of energy to be accounted for, namely, that associated with the evanescent wave that moves along the boundary in the plane-of-incidence. Since this energy could not have penetrated into the less dense medium under the present circumstances (so long as $\theta_i \geq \theta_c$), we must look elsewhere for its source. Under actual experimental conditions the incident beam would have a finite cross section and therefore would obviously differ from a true plane wave. This deviation gives rise (via diffraction) to a slight transmission of energy across the interface, which is manifested in the evanescent wave.

Incidentally, it is clear from (*c*) and (*d*) in Fig. 4.44 that the incident and reflected waves (except at $\theta_i = 90°$) do not differ in phase by π and cannot therefore cancel each other. It follows from the continuity of the tangential component of $\vec{\mathbf{E}}$ that there must be an oscillatory field in the less dense medium, with a component parallel to the interface having a frequency ω (i.e., the evanescent wave).

The exponential decay of the surface wave, or *boundary wave*, as it is also sometimes called, has been confirmed experimentally at optical frequencies.*

Imagine that a beam of light traveling within a block of glass is internally reflected at a boundary. Presumably, if you pressed another piece of glass against the first, the air–glass interface could be made to vanish, and the beam would then propagate onward undisturbed. Furthermore, you might expect this transition from total to no reflection to occur gradually as the air film thinned out. In much the same way, if you hold a drinking glass or a prism, you can see the ridges of your fingerprints in a region that, because of total internal reflection, is otherwise mirrorlike. In more general terms, when the evanescent wave extends with appreciable amplitude across the rare medium into a nearby region occupied by a higher-index material, energy may flow through the gap in what is known as **frustrated total internal reflection** (FTIR). The evanescent wave, having traversed the gap, is still strong enough to drive electrons in the "frustrating" medium; they in turn will generate a wave that significantly alters the field configuration, thereby permitting energy to flow. Figure 4.55 is a schematic representation of FTIR: the width of the lines depicting the wavefronts decreases across the gap as a reminder that the amplitude of the field behaves in the same way. The process as a whole is remarkably similar to the quantum-mechanical phenomenon of *barrier penetration* or *tunneling*, which has numerous applications in contemporary physics.

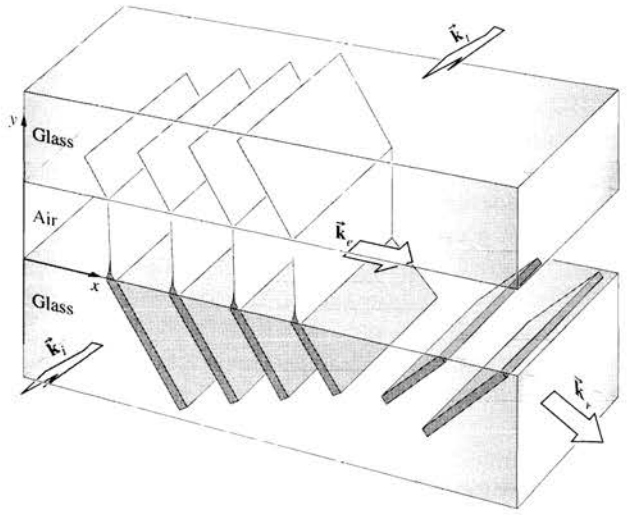
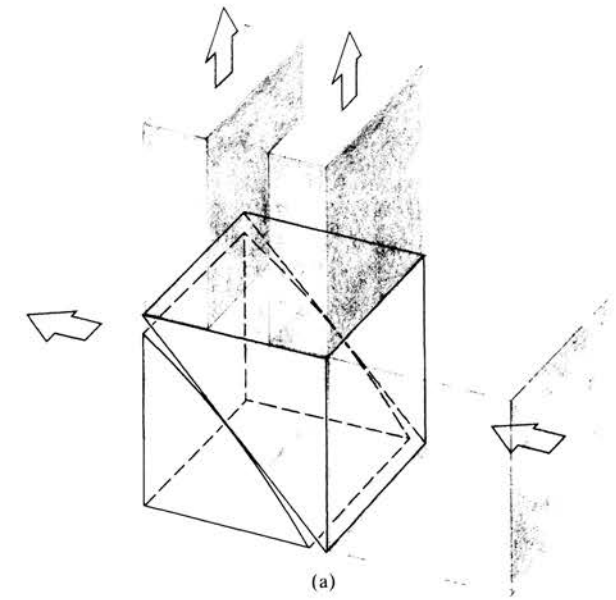


Figure 4.55 Frustrated total internal reflection.

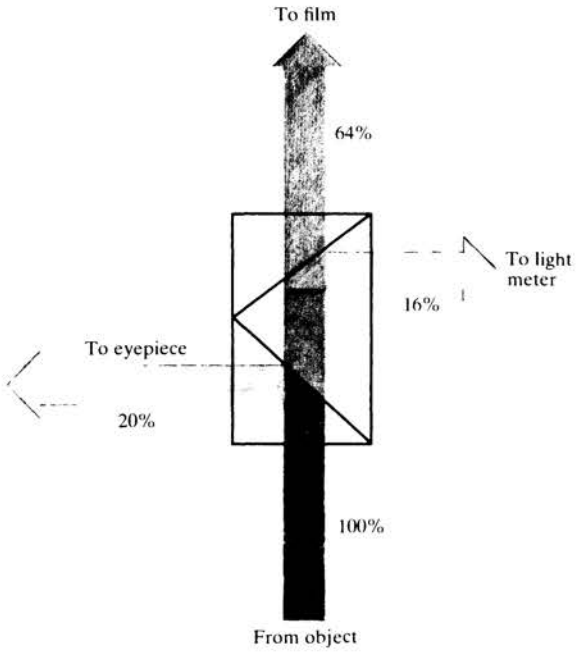
One can demonstrate FTIR with the prism arrangement of Fig. 4.56 in a manner that is fairly self-evident. Moreover, if the hypotenuse faces of both prisms are made planar and parallel, they can be positioned so as to transmit and reflect any desired fraction of the incident flux density. Devices that perform this function are known as ***beamsplitters***. A *beamsplitter cube* can be made rather conveniently by using a thin, low-index transparent film as a precision spacer. Low-loss reflectors whose transmittance can be controlled by frustrating internal reflection are of considerable practical interest. FTIR can also be observed in other regions of the electromagnetic spectrum. Three-centimeter microwaves are particularly easy to work with, inasmuch as the evanescent wave will extend roughly 10^5 times farther than it would at optical frequencies. One can duplicate the above optical experiments with solid prisms made of paraffin or hollow ones of acrylic plastic filled with kerosene or motor oil. Any one of these would have an index of about 1.5 for 3-cm waves. It then becomes an easy matter to measure the dependence of the field amplitude on *y*.

*Take a look at the fascinating article by K. H. Drexhage, "Monomolecular layers and light." *Sci. Am.* **222**, 108 (1970).

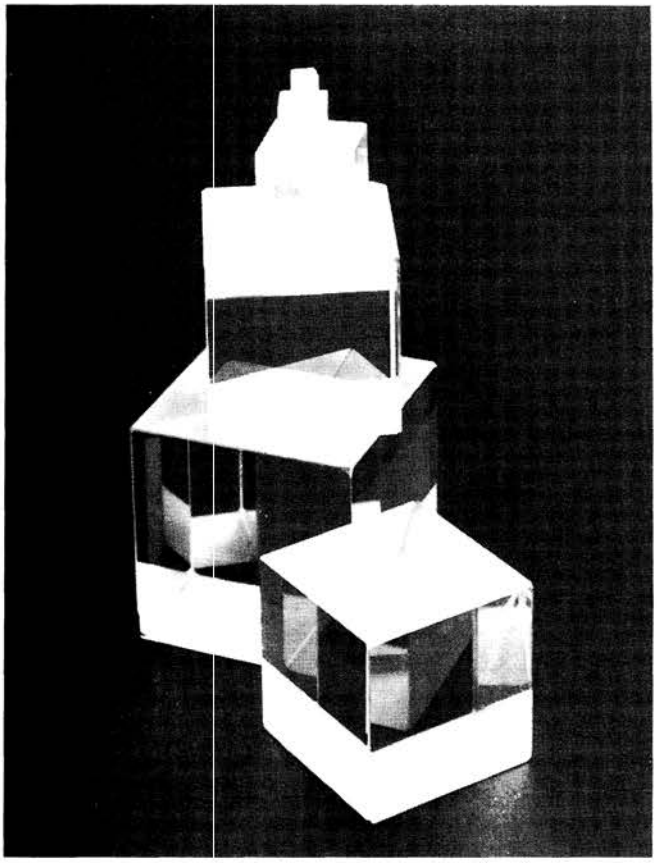
Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(a)



(b)



(c)

Figure 4.56 (*a*) A beamsplitter utilizing FTIR. (*b*) A typical modern application of FTIR: a conventional beamsplitter arrangement used to take photographs through a microscope. (*c*) Beamsplitter cubes. (Photo courtesy Melles Griot.)

4.8 Optical Properties of Metals

The characteristic feature of conducting media is the presence of a number of free electric charges (free in the sense of being unbound, i.e., able to circulate within the material). For metals these charges are of course electrons, and their motion constitutes a current. The current per unit area resulting from the application of a field $\vec{\mathbf{E}}$ is related by means of Eq. (A1.15) to the conductivity of the medium σ. For a dielectric there are no

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

free or conduction electrons and $\sigma = 0$, whereas for metals σ is nonzero and finite. In contrast, an idealized "perfect" conductor would have an infinite conductivity. This is equivalent to saying that the electrons, driven into oscillation by a harmonic wave, would simply follow the field's alternations. There would be no restoring force, no natural frequencies, and no absorption, only re-emission. In real metals the conduction electrons undergo collisions with the thermally agitated lattice or with imperfections and in so doing irreversibly convert electromagnetic energy into joule heat. The absorption of radiant energy by a material is a function of its conductivity.

Waves in a Metal

If we visualize the medium as continuous, Maxwell's Equations lead to

$$\frac{\partial^2 \vec{\mathbf{E}}}{\partial x^2} + \frac{\partial^2 \vec{\mathbf{E}}}{\partial y^2} + \frac{\partial^2 \vec{\mathbf{E}}}{\partial z^2} = \mu\epsilon \frac{\partial^2 \vec{\mathbf{E}}}{\partial t^2} + \mu\sigma \frac{\partial \vec{\mathbf{E}}}{\partial t} \qquad (4.74)$$

which is Eq. (A1.21) in Cartesian coordinates. The last term, $\mu\sigma\, \partial\vec{\mathbf{E}}/\partial t$, is a first-order time derivative, like the damping force in the oscillator model (p. 71). The time rate-of-change of $\vec{\mathbf{E}}$ generates a voltage, currents circulate, and since the material is resistive, light is converted to thermal energy—ergo absorption. This expression can be reduced to the unattenuated wave equation, if the permittivity is reformulated as a complex quantity. This in turn leads to a complex index of refraction, which, as we saw earlier (p. 71), is tantamount to absorption. We then need only substitute the complex index

$$\tilde{n} = n_R - in_I \qquad (4.75)$$

(where the real and imaginary indices n_R and n_I are both real numbers) into the corresponding solution for a nonconducting medium. Alternatively, we can utilize the wave equation and appropriate boundary conditions to yield a specific solution. In either event, it is possible to find a simple sinusoidal plane-wave solution applicable within the conductor. Such a wave propagating in the y-direction is ordinarily written as

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 \cos(\omega t - ky)$$

or as a function of n,

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 \cos \omega(t - \tilde{n}y/c)$$

but here the refractive index must be taken as complex. Writing the wave as an exponential and using Eq. (4.75) yields

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 e^{(-\omega n_I y/c)} e^{i\omega(t - n_R y/c)} \qquad (4.76)$$

or

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_0 e^{-\omega n_I y/c} \cos \omega(t - n_R y/c) \qquad (4.77)$$

The disturbance advances in the y-direction with a speed c/n_R, precisely as if n_R were the more usual index of refraction. As the wave progresses into the conductor, its amplitude, $\vec{\mathbf{E}}_0 \exp(-\omega n_I y/c)$, is exponentially attenuated. Inasmuch as irradiance is proportional to the square of the amplitude, we have

$$I(y) = I_0 e^{-\alpha y} \qquad (4.78)$$

where $I_0 = I(0)$; that is, I_0 is the irradiance at $y = 0$ (the interface), and $\alpha \equiv 2\omega n_I/c$ is called the *absorption coefficient* or (even better) the **attenuation coefficient**. The flux density will drop by a factor of $e^{-1} = 1/2.7 \approx \frac{1}{3}$ after the wave has propagated a distance $y = 1/\alpha$, known as the **skin** or **penetration depth**. For a material to be transparent, the penetration depth must be large in comparison to its thickness. The penetration depth for metals, however, is exceedingly small. For example, copper at ultraviolet wavelengths ($\lambda_0 \approx 100$ nm) has a miniscule penetration depth, about 0.6 nm, while it is still only about 6 nm in the infrared ($\lambda_0 \approx 10\,000$ nm). This accounts for the generally observed opacity of metals, which nonetheless can become partly transparent when formed into extremely thin films (e.g., in the case of partially silvered two-way mirrors). The familiar metallic sheen of conductors corresponds to a high reflectance, which exists because the incident wave cannot effectively penetrate the material. Relatively few electrons in the metal "see" the transmitted wave, and therefore, although each absorbs strongly, little total energy is dissipated by them. Instead, most of the incoming energy reappears as the reflected wave. The majority of metals, including the less common ones (e.g., sodium, potassium, cesium, vanadium, niobium, gadolinium, holmium, yttrium, scandium, and osmium) have a silvery gray appearance like that of aluminum, tin, or steel. They reflect almost all the incident light (roughly 85 to 95%) regardless of wavelengths and are therefore essentially colorless.

Equation (4.77) is certainly reminiscent of Eq. (4.73) and FTIR. In both cases there is an exponential decay of the amplitude. Moreover, a complete analysis would show that the transmitted waves are not strictly transverse, there being a component of the field in the direction of propagation in both instances.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

The representation of metal as a continuous medium works fairly well in the low-frequency, long-wavelength domain of the infrared. Yet we certainly might expect that as the wavelength of the incident beam decreased the actual granular nature of matter would have to be reckoned with. Indeed, the continuum model shows large discrepancies from experimental results at optical frequencies. And so we again turn to the classical atomistic picture initially formulated by Hendrik Lorentz, Paul Karl Ludwig Drude (1863–1906), and others. This simple approach will provide qualitative agreement with the experimental data, but the ultimate treatment requires quantum theory.

The Dispersion Equation

Envision the conductor as an assemblage of driven, damped oscillators. Some correspond to free electrons and will therefore have zero restoring force, whereas others are bound to the atom, much like those in the dielectric media of Section 3.5.1. The conduction electrons are, however, the predominant contributors to the optical properties of metals. Recall that the displacement of a vibrating electron was given by

$$x(t) = \frac{q_e/m_e}{(\omega_0^2 - \omega^2)} E(t) \qquad [3.66]$$

With no restoring force, $\omega_0 = 0$, the displacement is opposite in sign to the driving force $q_e E(t)$ and therefore 180° out-of-phase with it. This is unlike the situation for transparent dielectrics, where the resonance frequencies are above the visible and the electrons oscillate in-phase with the driving force (Fig. 4.57). Free electrons oscillating out-of-phase with the incident light will reradiate wavelets that tend to cancel the incoming disturbance. The effect, as we have already seen, is a rapidly decaying refracted wave.

Assuming that the average field experienced by an electron moving about within a conductor is just the applied field $\vec{\mathbf{E}}(t)$, we can extend the dispersion equation of a rare medium [Eq. (3.72)] to read

$$n^2(\omega) = 1 + \frac{Nq_e^2}{\epsilon_0 m_e}\left[\frac{f_e}{-\omega^2 + i\gamma_e\omega} + \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 + i\gamma_j\omega}\right] \qquad (4.79)$$

The first bracketed term is the contribution from the free electrons, wherein N is the number of atoms per unit volume. Each of these has f_e conduction electrons, which have no natural frequencies. The second term arises from the bound electrons and is identical to Eq. (3.72). It should be noted that if a metal has a particular color, it indicates that the atoms are partaking of selective absorption by way of the bound electrons, in addition to the general absorption characteristic of the free electrons. Recall that a medium that is very strongly absorbing at a given frequency doesn't actually absorb much of the incident light at that frequency but rather *selectively reflects* it. Gold and copper are reddish yellow because n_f increases with wavelength, and the larger values of λ are reflected more strongly. Thus, for example, gold should be fairly opaque to the longer visible wavelengths. Consequently, under white light, a gold foil less than roughly 10^{-6} m thick will indeed transmit predominantly greenish blue light.

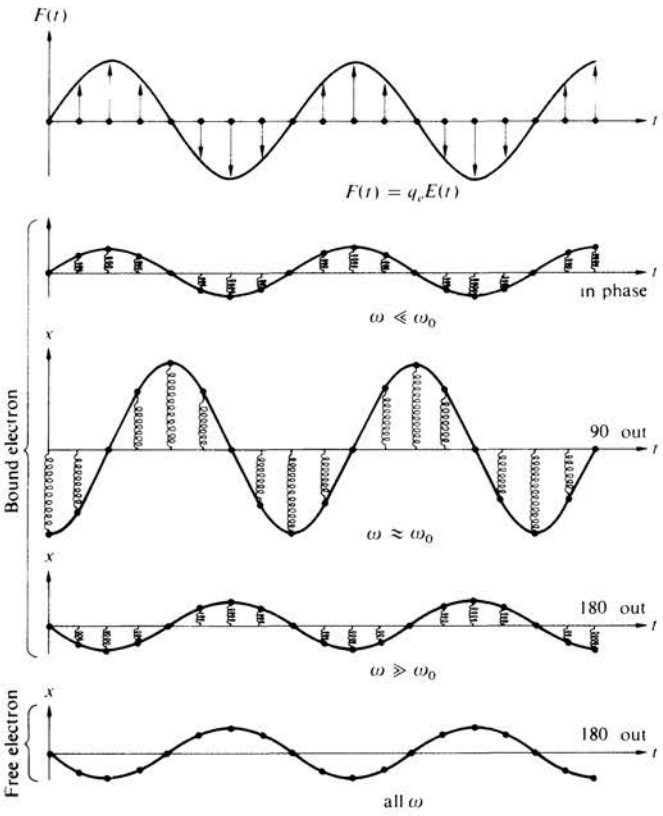


Figure 4.57 Oscillations of bound and free electrons.

We can get a rough idea of the response of metals to light by making a few simplifying assumptions. Accordingly, neglect the bound electron contribution and assume that γ_e is

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

also negligible for very large ω, whereupon

$$n^2(\omega) = 1 - \frac{Nq^2{}_e}{\epsilon_0 m_e \omega^2} \tag{4.80}$$

The latter assumption is based on the fact that at high frequencies the electrons will undergo a great many oscillations between each collision. Free electrons and positive ions within a metal may be thought of as a plasma whose density oscillates at a natural frequency ω_p, the **plasma frequency**. This in turn can be shown to equal $(Nq_e^2/\epsilon_0 m_e)^{1/2}$, and so

$$n^2(\omega) = 1 - (\omega_p/\omega)^2 \tag{4.81}$$

The plasma frequency serves as a critical value below which the index is complex and the penetrating wave drops off exponentially [Eq. (4.77)] from the boundary; at frequencies above ω_p, n is real, absorption is small, and the conductor is transparent. In the latter circumstance n is less than 1, as it was for dielectrics at very high frequencies (v can be greater than c—see p. 72). Hence we can expect metals in general to be fairly transparent to X-rays. Table 4.3 lists the plasma frequencies for some of the alkali metals that are transparent even to ultraviolet.

The index of refraction for a metal will usually be complex, and the impinging wave will suffer absorption in an amount that is frequency dependent. For example, the outer visors on the Apollo space suits were overlaid with a very thin film of gold (see photo). The coating reflected about 70% of the incident light and was used under bright conditions, such as low and forward Sun angles. It was designed to decrease the thermal load on the cooling system by strongly reflecting radiant energy in the infrared while still transmitting adequately in the visible. Inexpensive metal-coated sunglasses which are quite similar in principle are also available commercially, and they're well worth having just to experiment with.



Edwin Aldrin Jr. at Tranquility Base on the Moon. The photographer, Neil Armstrong, is reflected in the gold-coated visor. (Photo courtesy NASA.)

The ionized upper atmosphere of the Earth contains a distribution of free electrons that behave very much like those confined within a metal. The index of refraction of such a medium will be real and less than 1 for frequencies above ω_p. In July of 1965 the *Mariner IV* spacecraft made use of this effect to examine the ionosphere of the planet Mars, 216 million kilometers from Earth.*

If we wish to communicate between two distant terrestrial points, we might bounce low-frequency waves off the Earth's ionosphere. To speak to someone on the Moon, however, we should use high-frequency signals, to which the ionosphere would be transparent.

TABLE 4.3 Critical Wavelengths and Frequencies for Some Alkali Metals

Metal	λ_p (observed) nm	λ_p (calculated) nm	$\nu_p = c/\lambda_p$ (observed) Hz
Lithium (Li)	155	155	1.94×10^{15}
Sodium (Na)	210	209	1.43×10^{15}
Potassium (K)	315	287	0.95×10^{15}
Rubidium (Rb)	340	322	0.88×10^{15}

Reflection from a Metal

Imagine that a plane wave initially in air impinges on a conducting surface. The transmitted wave advancing at some angle to the normal will be inhomogeneous. But if the conductivity of the medium is increased, the wavefronts will become aligned with the surfaces of constant amplitude, whereupon $\vec{\mathbf{k}}_t$ and $\hat{\mathbf{u}}_n$ will approach parallelism. In other

*R. Von Eshelman, *Sci. Am.* **220**, 78 (1969).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

words, in a good conductor the transmitted wave propagates in a direction normal to the interface regardless of θ_i.

Let's now compute the reflectance, $R = I_r/I_i$, for the simplest case of normal incidence on a metal. Taking $n_i = 1$ and $n_t = \tilde{n}$ (i.e., the complex index), we have from Eq. (4.47) that

$$R = \left(\frac{\tilde{n} - 1}{\tilde{n} + 1}\right)\left(\frac{\tilde{n} - 1}{\tilde{n} + 1}\right)^* \tag{4.82}$$

and therefore, since $\tilde{n} = n_R - in_I$,

$$R = \frac{(n_R - 1)^2 + n_I^2}{(n_R + 1)^2 + n_I^2} \tag{4.83}$$

If the conductivity of the material goes to zero, we have the case of a dielectric, whereupon in principle the index is real ($n_I = 0$), and the attenuation coefficient, α, is zero. Under those circumstances, the index of the transmitting medium n_t is n_R, and the reflectance [Eq. (4.83)] becomes identical with that of Eq. (4.67). If instead n_I is large while n_R is comparatively small, R in turn becomes large (Problem 4.81). In the unattainable limit where $\tilde{n}$ is purely imaginary, 100% of the incident flux density would be reflected ($R = 1$). Notice that it is possible for the reflectance of one metal to be greater than that of another even though its n_I is smaller. For example, at $\lambda_0 = 589.3$ nm the parameters associated with solid sodium are roughly $n_R = 0.04$, $n_I = 2.4$, and $R = 0.9$; and those for bulk tin are $n_R = 1.5$, $n_I = 5.3$, and $R = 0.8$; whereas for a gallium single crystal $n_R = 3.7$, $n_I = 5.4$, and $R = 0.7$.

The curves of $R_\parallel$ and $R_\perp$ for oblique incidence shown in Fig. 4.58 are somewhat typical of absorbing media. Thus, although R at $\theta_i = 0$ is about 0.5 for gold, as opposed to nearly 0.9 for silver in white light, the two metals have reflectances

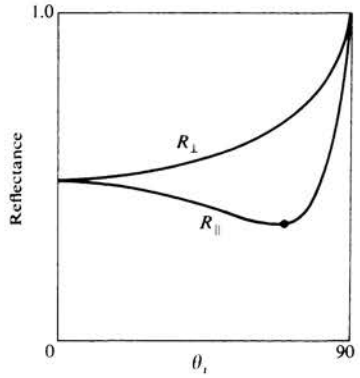


Figure 4.58 Typical reflectance for a linearly polarized beam of white light incident on an absorbing medium.

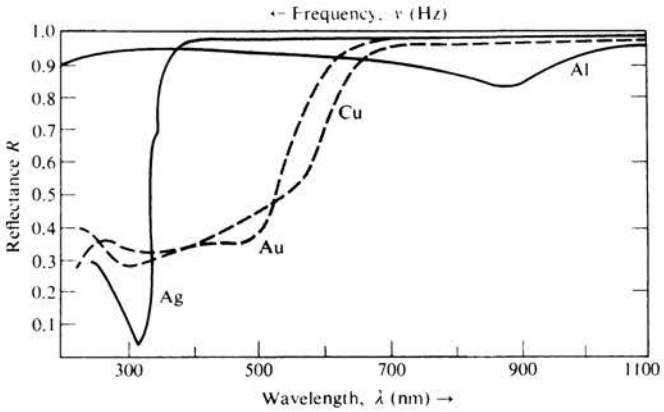


Figure 4.59 Reflectance versus wavelength for silver, gold, copper, and aluminum.

that are quite similar in shape, approaching 1.0 at $\theta_i = 90°$. Just as with dielectrics (Fig. 4.48), $R_\parallel$ drops to a minimum at what is now called the *principal angle-of-incidence*, but here that minimum is nonzero. Figure 4.59 illustrates the spectral reflectance at normal incidence for a number of evaporated metal films under ideal conditions. Observe that although gold transmits fairly well in and below the green region of the spectrum, silver, which is highly reflective across the visible, becomes transparent in the ultraviolet at about 316 nm.

Phase shifts arising from reflection off a metal occur in both components of the field (i.e., parallel and perpendicular to the plane-of-incidence). These are generally neither 0 nor π, with a notable exception at $\theta_i = 90°$, where, just as with a dielectric, both components shift phase by 180° on reflection.

4.9 Familiar Aspects of the Interaction of Light and Matter

Let's now examine some of the phenomena that paint the everyday world in a marvel of myriad colors.

As we saw earlier (p. 77), light that contains a roughly equal amount of every frequency in the visible region of the spectrum is perceived as white. A broad source of white light (whether natural or artificial) is one for which every point on its surface can be imagined as sending out a stream of light of every visible frequency. Given that we evolved on this planet, it's not surprising that a source appears white when its emis-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

sion spectrum resembles that of the Sun. Similarly, a reflecting surface that accomplishes essentially the same thing will also appear white: a highly reflecting, frequency-independent, *diffusely* scattering object will be perceived as white under white light illumination.

Although water is essentially transparent, water vapor appears white, as does ground glass. The reason is simple enough—if the grain size is small but larger than the wavelengths involved, light will enter each transparent particle, be reflected and refracted, and emerge. There will be no distinction among any of the frequency components, so the reflected light reaching the observer will be white (p. 77). This is the mechanism accountable for the whiteness of things like sugar, salt, paper, cloth, clouds, talcum powder, snow, and paint, each grain or fiber of which is actually transparent.

Similarly, a wadded-up piece of crumpled clear plastic wrap will appear whitish, as will an ordinarily transparent material filled with small air bubbles (e.g., shaving cream or beaten egg white). Even though we usually think of paper, talcum powder, and sugar as each consisting of some sort of opaque white substance, it's an easy matter to dispel that misconception. Cover a printed page with a few of these materials (a sheet of white paper, some grains of sugar, or talcum) and illuminate it from behind. You'll have little difficulty seeing through them. In the case of white paint, one simply suspends colorless transparent particles, such as the oxides of zinc, titanium, or lead, in an equally transparent vehicle, for example, linseed oil or acrylics. Obviously, if the particles and vehicle have the same index of refraction, there will not be any reflections at the grain boundaries. The particles will simply disappear into the conglomeration, which itself remains clear. In contrast, if the indices are markedly different, there will be a good deal of reflection at all wavelengths (Problem 4.74), and the paint will appear white and opaque [take another look at Eq. (4.67)]. To color paint one need only dye the particles so that they absorb all frequencies except the desired range.

Carrying the logic in the reverse direction, if we reduce the relative index, $n_{\prime\prime}$, at the grain or fiber boundaries, the particles of material will reflect less, thereby decreasing the overall whiteness of the object. Consequently, a wet white tissue will have a grayish, more transparent look. Wet talcum powder loses its sparkling whiteness, becoming a dull gray, as does wet white cloth. In the same way, a piece of dyed fabric soaked in a clear liquid (e.g., water, gin, or benzene) will lose its whitish haze and become much darker, the colors then being deep and rich like those of a still-wet watercolor painting.

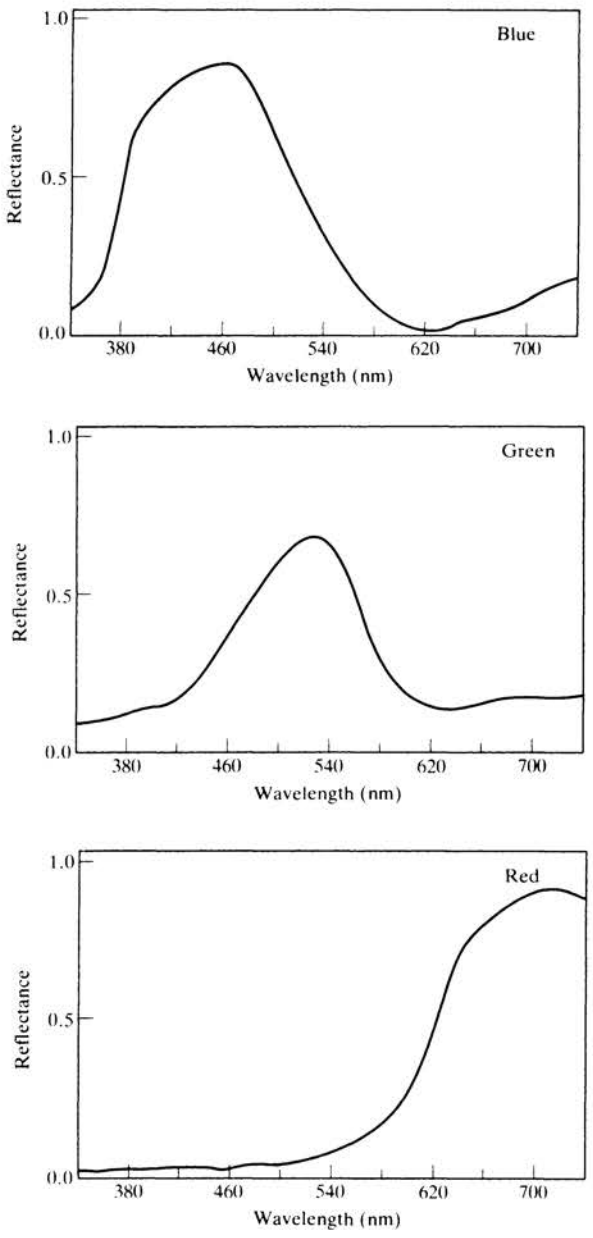


Figure 4.60 Reflection curves for blue, green, and red pigments. These are typical, but there is a great deal of possible variation among the colors.

A diffusely reflecting surface that absorbs somewhat—uniformly across the spectrum—will reflect a bit less than a white surface and so appear mat gray. The less it reflects, the darker

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

the gray, until it absorbs almost all the light and appears black. A surface that reflects perhaps 70% or 80% or more, but does so specularly, will appear the familiar shiny gray of a typical metal. Metals possess tremendous numbers of free electrons (p. 127) that scatter light very effectively, independent of frequency: they are not bound to the atoms and have no associated resonances. Moreover, the amplitudes of the vibrations are an order of magnitude larger than they were for the bound electrons. The incident light cannot penetrate into the metal any more than a fraction of a wavelength or so before it's canceled completely. There is little or no refracted light; most of the energy is reflected out, and only the small remainder is absorbed. Note that the primary difference between a gray surface and a mirrored surface is one of diffuse versus specular reflection. An artist paints a picture of a polished "white" metal, such as silver or aluminum, by "reflecting" images of things in the room on top of a gray surface.

When the distribution of energy in a beam of light is not effectively uniform across the spectrum, the light appears colored. Figure 4.60 depicts typical frequency distributions for what would be perceived as red, green, and blue light. These curves show the predominant frequency regions, but there can be a great deal of variation in the distributions, and they will still provoke the responses of red, green, and blue. In the early 1800s Thomas Young showed that a broad range of colors could be generated by mixing three beams of light, provided their frequencies were widely separated. When three such beams combine to produce white light, they are called **primary colors**. There is no single unique set of these primaries, nor do they have to be quasimonochromatic. Since a wide range of colors can be created by mixing red (R), green (G), and blue (B), these tend to be used most frequently. They are the three components (emitted by three phosphors) that generate the whole gamut of hues seen on a color television set.

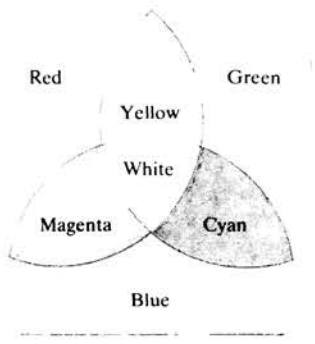


Figure 4.61 Three overlapping beams of colored light. A color television set uses these same three primary light sources—red, green, and blue.

Figure 4.61 summarizes the results when beams of these three primaries are overlapped in a number of different combinations: Red light plus blue light is seen as *magenta* (M), a reddish purple; blue light plus green light is seen as *cyan* (C), a bluish green or turquoise; and perhaps most surprising, red light plus green light is seen as *yellow* (Y). The sum of all three primaries is white:

$$R + B + G = W$$

$$M + G = W, \text{ since } R + B = M$$

$$C + R = W, \text{ since } B + G = C$$

$$Y + B = W, \text{ since } R + G = Y$$

Any two colored light beams that together produce white are said to be **complementary**, and the last three symbolic statements exemplify that situation. Thus

$$R + B + G = W$$

$$R + B = W - G = M$$

$$B + G = W - R = C$$

$$R + G = W - B = Y$$

which means, for example, that a filter that absorbs blue out of white light passes yellow.

Suppose we overlap beams of magenta and yellow light:

$$M + Y = (R + B) + (R + G) = W + R$$

The result is a combination of red and white, or pink. That raises another point: we say a color is **saturated**, that it is deep and intense, when it does not contain any white light. As Fig. 4.62 shows, pink is unsaturated red—red superimposed on a background of white.

The mechanism responsible for the yellowish red hue of gold and copper is, in some respects, similar to the process that causes the sky to appear blue. Putting it rather succinctly, the molecules of air have resonances in the ultraviolet and will be driven into larger-amplitude oscillations as the frequency of the incident light increases toward the ultraviolet. They effectively take energy from and re-emit the blue component of sunlight in all directions, transmitting the complementary red

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

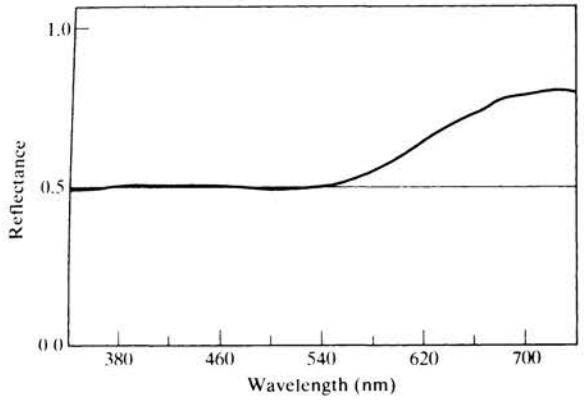


Figure 4.62 Spectral reflection of a pink pigment.

end of the spectrum with little alteration. This is analogous to the selective reflection or scattering of yellow-red light that takes place at the surface of a gold film and the concomitant transmission of blue-green light.

The characteristic colors of most substances have their origin in the phenomenon of **selective** or **preferential absorption**. For example, water has a very faint green-blue tint because of its absorption of red light. That is, the H_2O molecules have a broad resonance in the infrared, which extends somewhat into the visible. The absorption isn't very strong, so there is no accentuated reflection of red light at the surface. Instead it is transmitted and gradually absorbed out until at a depth of about 30 m of seawater, red is almost completely removed from the sunlight. This same process of selective absorption is responsible for the colors of brown eyes and butterflies, of birds and bees and cabbages and kings. Indeed, the great majority of objects in nature appear to have characteristic colors as the result of preferential absorption by pigment molecules. In contrast with most atoms and molecules, which have resonances in the ultraviolet and infrared, the pigment molecules must obviously have resonances in the visible. Yet visible photons have energies of roughly 1.6 eV to 3.2 eV, which, as you might expect, are on the low side for ordinary electron excitation and on the high side for excitation via molecular vibration. Despite this, *there are atoms where the bound electrons form incomplete shells* (gold, for example) and variations in the configuration of these shells provide a mode for low-energy excitation. In addition, there is the large group of organic dye molecules, which evidently also have resonances in the visible. All such substances, whether natural or synthetic, consist of long-chain molecules made up of regularly alternating single and double bonds in what is called a conjugated system. This structure is typified by the carotene molecule $C_{40}H_{56}$ (Fig. 4.63). The carotenoids range in color from yellow to red and are found in carrots, tomatoes, daffodils, dandelions, autumn leaves, and people. The chlorophylls are another group of familiar natural pigments, but here a portion of the long chain is turned around on itself to form a ring. In any event, conjugated systems of this sort contain a number of particularly mobile electrons known as *pi electrons*. They are not bound to specific atomic sites but instead can range over the relatively large dimensions of the molecular chain or ring. In the phraseology of Quantum Mechanics, we say that these are long-wavelength, low-frequency, and therefore low-energy, electron states. The energy required to raise a pi electron to an excited state is comparatively low, corresponding to that of visible photons. In effect, the molecule can be imagined as an oscillator having a resonance frequency in the visible.

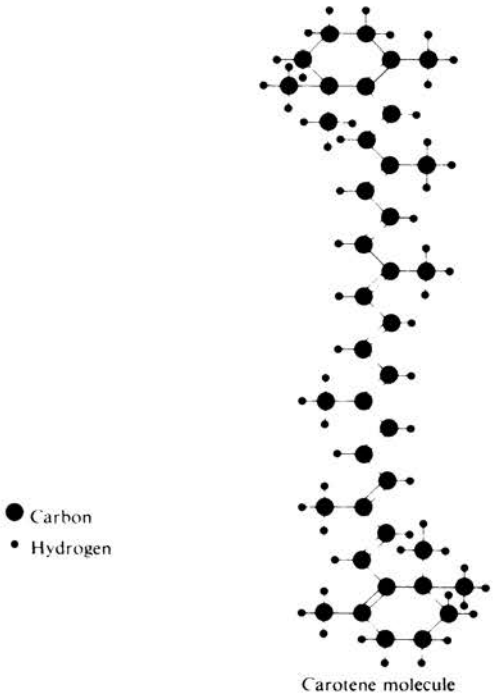


Figure 4.63 The carotene molecule.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

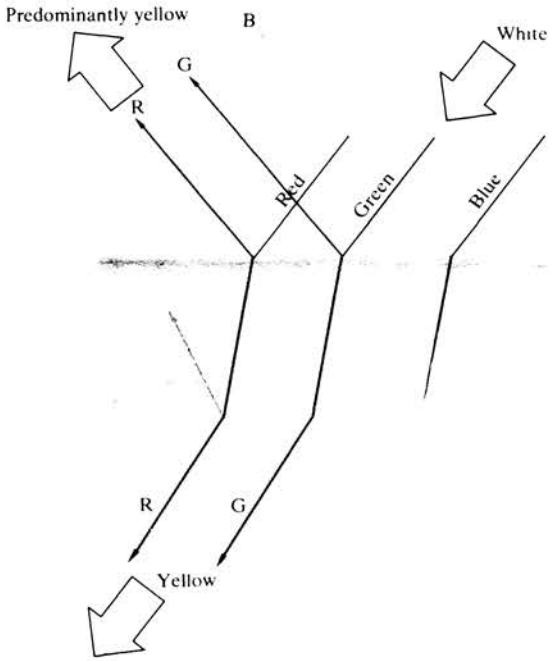


Figure 4.64 Yellow stained glass

The energy levels of an individual atom are precisely defined; that is, the resonances are very sharp. With solids and liquids, however, the proximity of the atoms results in a broadening of the energy levels into wide bands. The resonances spread over a broad range of frequencies. Consequently, we can expect that a dye will not absorb just a narrow portion of the spectrum; indeed if it did, it would reflect most frequencies and appear nearly white.

Imagine a piece of stained glass with a resonance in the blue where it strongly absorbs. If you look through it at a white-light source composed of red, green, and blue, the glass will absorb blue, passing red and green, which is yellow (Fig. 4.64). The glass looks yellow: yellow cloth, paper, dye, paint, and ink all selectively absorb blue. If you peer at something that is a pure blue through a yellow filter, one that passes yellow and absorbs blue, the object will appear black. Here the filter colors the light yellow by removing blue, and we speak of the process as **subtractive coloration**, as opposed to **additive coloration**, which results from overlapping beams of light.

In the same way, fibers of a sample of white cloth or paper are essentially transparent, but when dyed each fiber behaves as if it were a chip of colored glass. The incident light penetrates the paper, emerging for the most part as a reflected beam only after undergoing numerous reflections and refractions within the dyed fibers. The exiting light will be colored to the extent that it lacks the frequency component absorbed by the dye. This is precisely why a leaf appears green, or a banana yellow.

A bottle of ordinary blue ink looks blue in either reflected or transmitted light. But if the ink is painted on a glass slide and the solvent evaporates, something rather interesting happens. The concentrated pigment absorbs so effectively that it preferentially reflects at the resonant frequency, and we are back to the idea that a strong absorber (larger n_I) is a strong reflector. Thus, concentrated blue-green ink reflects red, whereas red-blue ink reflects green. Try it with a felt marker (overhead projector pens are best), but you must use reflected light, being careful not to inundate the sample with unwanted light from below. The most convenient way to accomplish that is to put colored ink onto a black surface that isn't very absorbant. For example, smear red ink over a black area on a glossy printed page (or better yet, on a black piece of plastic) and it will glow green in reflected light. Gentian violet, which you can buy in any drugstore, works beautifully. Put some on a glass slide and let it dry in a thick coat. Examine both the reflected and transmitted light—they will be complementary.

The whole range of colors (including red, green, and blue) can be produced by passing white light through various combinations of magenta, cyan, and yellow filters (Fig. 4.65). These are the primary colors of subtractive mixing, the primaries of the paint box, although they are often mistakenly spoken of as red, blue, and yellow. They are the basic colors of the dyes used to make photographs and the inks used to print them. Ideally, if you mix all the subtractive primaries together (either by combining paints or by stacking filters), you get no color, no light—black. Each removes a region of the spectrum, and together they absorb it all.

If the range of frequencies being absorbed spreads across the visible, the object will appear black. That is not to say that there is no reflection at all—you obviously can see a reflected image in a piece of black patent leather, and a rough black surface reflects also, only diffusely. If you still have those red and blue inks, mix them, add some green, and you'll get black.

In addition to the above processes specifically related to reflection, refraction, and absorption, there are other color-generating mechanisms, which we shall explore later on. For

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

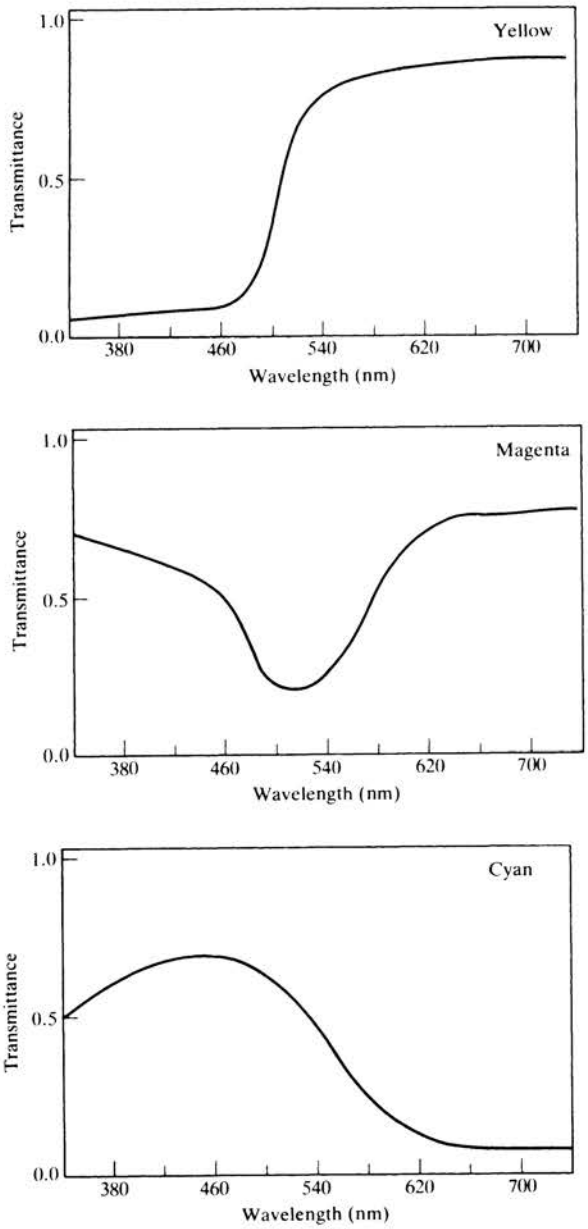


Figure 4.65 Transmission curves for colored filters.

example, scarabaeid beetles mantle themselves in the brilliant colors produced by diffraction gratings on their wing cases, and wavelength-dependent interference effects contribute to the color patterns seen on oil slicks, mother-of-pearl, soap bubbles, peacocks, and hummingbirds.

4.10 The Stokes Treatment of Reflection and Refraction

A rather elegant and novel way of looking at reflection and transmission at a boundary was developed by the British physicist Sir George Gabriel Stokes (1819–1903). Suppose that we have an incident wave of amplitude E_{0i} impinging on the planar interface separating two dielectric media, as in Fig. 4.66*a*. As we saw earlier in this chapter, because r and t are the fractional amplitudes reflected and transmitted, respectively (where $n_i = n_1$ and $n_t = n_2$), then $E_{0r} = rE_{0i}$ and $E_{0t} = tE_{0i}$. Again we are reminded that Fermat's Principle led to the Principle of Reversibility, which implies that the situation depicted in Fig. 4.66*b*, where all the ray directions are reversed, must also be physically possible. With the one proviso that there be no energy dissipation (no absorption), a wave's meanderings must be reversible. Equivalently, in the idiom of modern physics one speaks of *time-reversal invariance*, that is, if a process occurs, the reverse process can also occur. Thus if we take a hypothetical motion picture of the wave incident on, reflecting from, and transmitting through the interface, the behavior depicted when the film is run backward must also be physically realizable. Accordingly, examine Fig. 4.66*c*, where there are now two incident waves of amplitudes $E_{0i}r$ and $E_{0i}t$. A portion of the wave whose amplitude is $E_{0i}t$ is both reflected and transmitted at the interface. Without making any assumptions, let r' and t' be the amplitude reflection and transmission coefficients, respectively, for a wave incident from below (i.e., $n_i = n_2$, $n_t = n_1$). Consequently, the reflected portion is $E_{0i}tr'$, and the transmitted portion is $E_{0i}tt'$. Similarly, the incoming wave whose amplitude is $E_{0i}r$ splits into segments of amplitude $E_{0i}rr$ and $E_{0i}rt$. If the configuration in Fig. 4.66*c* is to be identical with that in Fig. 4.66*b*, then obviously

$$E_{0i}tt' + E_{0i}rr = E_{0i} \tag{4.84}$$

and

$$E_{0i}rt + E_{0i}tr' = 0 \tag{4.85}$$

Hence

$$tt' = 1 - r^2 \tag{4.86}$$

and

$$r' = -r \tag{4.87}$$

the latter two equations being known as the Stokes Relations. This discussion calls for a bit more caution than is usually granted it. It must be pointed out that *the amplitude coefficients are functions of the incident angles*, and therefore the

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

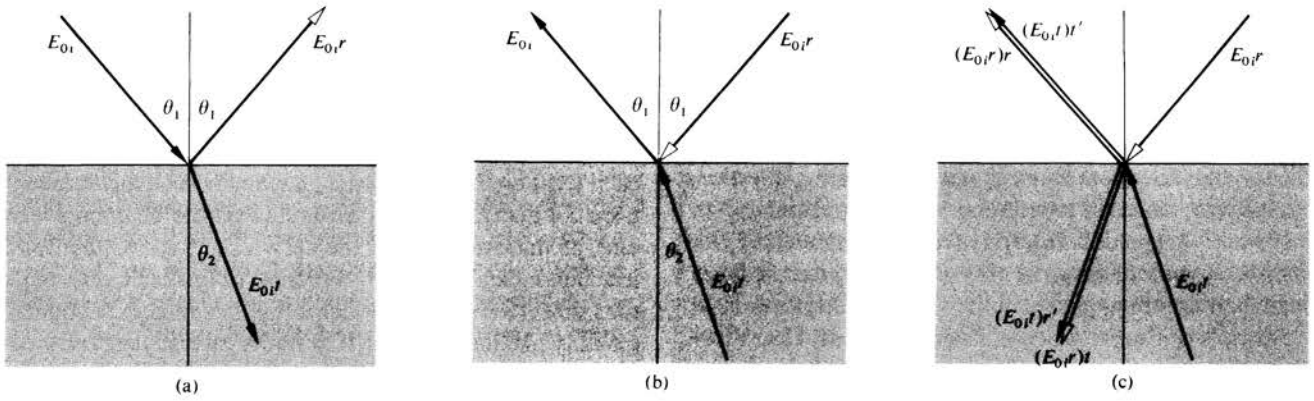


Figure 4.66 Reflection and refraction via the Stokes treatment.

Stokes Relations might better be written as

$$t(\theta_I)t'(\theta_2) = 1 - r^2(\theta_I) \tag{4.88}$$

and

$$r'(\theta_2) = -r(\theta_I) \tag{4.89}$$

where $n_1 \sin \theta_1 = n_2 \sin \theta_2$. The second equation indicates, by virtue of the minus sign, that *there is a 180° phase difference between the waves internally and externally reflected.* It is most important to keep in mind that here θ_1 and θ_2 are pairs of angles that are related by way of Snell's Law. Note as well that we never did say whether n_1 was greater or less than n_2, so Eqs. (4.88) and (4.89) apply in either case. Let's return for a moment to one of the Fresnel Equations:

$$r_\perp = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \tag{4.42}$$

If a ray enters from above, as in Fig. 4.66*a*, and we assume $n_2 > n_1$, $r_\perp$ is computed by setting $\theta_i = \theta_1$ and $\theta_t = \theta_2$ (external reflection), the latter being derived from Snell's Law. If, on the other hand, the wave is incident at that same angle from below (in this instance internal reflection), $\theta_i = \theta_1$ and we again substitute in Eq. (4.42), but here θ_t is not θ_2, as before. The values of $r_\perp$ for internal and external reflection *at the same incident angle* are obviously different. Now suppose, in this case of internal reflection, that $\theta_i = \theta_2$. Then $\theta_t = \theta_1$, the ray directions are the reverse of those in the first situation, and Eq. (4.42) yields

$$r'_\perp(\theta_2) = -\frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)}$$

Although it may be unnecessary we once again point out that this is just the negative of what was determined for $\theta_i = \theta_1$ and external reflection, that is,

$$r'_\perp(\theta_2) = -r_\perp(\theta_I) \tag{4.90}$$

The use of primed and unprimed symbols to denote the amplitude coefficients should serve as a reminder that we are once more dealing with angles related by Snell's Law. In the same way, interchanging θ_i and θ_t in Eq. (4.43) leads to

$$r'_{\parallel}(\theta_2) = -r_{\parallel}(\theta_I) \tag{4.91}$$

The 180° phase difference between each pair of components is evident in Fig. 4.44, but keep in mind that when $\theta_i = \theta_p$, $\theta_t = \theta'_p$ and vice versa (Problem 4.78). Beyond $\theta_i = \theta_c$ there is no transmitted wave, Eq. (4.89) is not applicable, and as we have seen, the phase difference is no longer 180°.

It is common to conclude that both the parallel and perpendicular components of the externally reflected beam change phase by π radians, while the internally reflected beam undergoes no phase shift at all. This is quite incorrect (compare Figs. 4.45*a* and 4.46*a*).

4.11 Photons, Waves, and Probability

Much of the theoretical grounding of Optics is predicated on wave theory. We take for granted both that we understand the phenomenon and that it's "real." As one example out of the many that will be encountered, the process of scattering seems

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

to be understandable only in terms of interference; classical particles simply do not interfere. When a beam propagates through a dense medium, interference in the forward direction is constructive, whereas in all other directions it's almost completely destructive. Thus nearly all the light energy advances in the forward direction. But this raises interesting questions about the basic nature of interference and the usual interpretation of what's happening. ***Interference is a nonlocalized phenomenon; it cannot happen at only one single point in space***, even though we often talk about the interference at a point P. The principle of Conservation of Energy makes it clear that if there is constructive interference at one point, the "extra" energy at that location must have come from elsewhere. There must therefore be destructive interference somewhere else. ***Interference takes place over an extended region of space in a coordinated fashion that leaves the total amount of radiant energy unchanged***.

Now imagine a light beam traversing a dense medium, as in Fig. 4.6. Do real energy-carrying electromagnetic wavelets (which are never actually measured) propagate out laterally, only to interfere destructively everywhere beyond the beam? If so, these wavelets cancel and the energy they transport outward is inexplicably returned to the beam since, in the end, there is no net lateral scattering. That's true no matter how far away P is. Moreover, this applies to *all* interference effects (Chapter 9). If two or more electromagnetic waves arrive at point P out-of-phase and cancel, "What does that mean as far as their energy is concerned?" Energy can be redistributed, but *it* doesn't cancel out. We've learned from Quantum Mechanics that at base interference is one of the most fundamental mysteries in physics.

Remembering Einstein's admonition that there are no spherical wavelets emitted by atoms, perhaps we're being too literal in our interpretation of the classical wave field. After all, strictly speaking, the classical electromagnetic wave with its continuous distribution of energy does not actually exist. Perhaps we should think of the wavelets and the overall pattern they produce (rather than being a *real* wave field) as a theoretical device that, wonderfully enough, tells us where the light will end up. In any event, Maxwell's Equations provide a means of calculating the macroscopic distribution of electromagnetic energy in space.

Moving ahead in a semiclassical way, imagine a distribution of light given by some function of the off-axis angle θ. For example, consider the irradiance on a screen placed far beyond a slit-shaped aperture (p. 453) such that $I(\theta) = I(0)\ \mathrm{sinc}^2\beta(\theta)$. Suppose that instead of observing the pattern by eye a detector composed of a diaphragm followed by a photomultiplier tube is used. Such a device could be moved around from one point to another, and over a constant time interval, it could measure the number of photons arriving at each location, $N(\theta)$. Taking a great many such measurements, a spatial distribution of the number of photon counts would emerge that would be of the very same form as that for the irradiance, namely, $N(\theta) = N(0)\mathrm{sinc}^2\beta(\theta)$: the number of photons detected is proportional to the irradiance. A countable quantity like this lends itself to statistical analysis, and we can talk about the probability of detecting a photon at any point on the screen. That is, a probability distribution can be constructed, reminiscent of Fig. 3.23. Because the space variables (θ, x, y, or z) are continuous, it's necessary to introduce a **probability density**, let it be $\wp(\theta)$. Then $\wp(\theta)\,d\theta$ is the probability that a photon will be found in the infinitesimal range from θ to $\theta + d\theta$. In this case $\wp(\theta) = \wp(0)\ \mathrm{sinc}^2\beta(\theta)$.

The square of the net electric field amplitude at every point in space corresponds to the irradiance (which can be measured directly), and that's equivalent to the likelihood of finding photons at any point. Accordingly, let's tentatively define the **probability amplitude** as that quantity whose absolute value squared equals the probability density. Thus the net E_0 at P can be interpreted as being proportional to a **semiclassical** probability amplitude inasmuch as ***the probability of detecting a photon at some point in space depends on the irradiance at that location and*** $I \propto E_0^2$. This is in accord with Einstein's conception of the light field, which Max Born (who initiated the statistical interpretation of Quantum Mechanics) described as a *Gespensterfeld*, or phantom field. In that view the waves of that field reveal how the photons distribute in space in the sense that the square of the absolute value of the wave amplitude somehow relates to the probability density of arriving photons. In the formal treatment of Quantum Mechanics, *the probability amplitude is generally a complex quantity* whose absolute value squared corresponds to the probability density (e.g., the Schrödinger wavefunction is a probability amplitude). Thus, however reasonable it was to consider E_0 as equivalent to a semiclassical probability amplitude, that usage cannot be carried over, as is, into quantum theory.

Still, all of this suggests that we might take the scattering process, considered in terms of probabilities, as the basis for a computational scheme. Each scattered wavelet is then a measure of the probability amplitude for light taking a particular route from one point to another, and the net electric field at P

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

is the sum of all the scattered fields arriving via all possible routes. A quantum-mechanical methodology analogous to this was devised by Feynman, Schwinger, Tomonaga, and Dyson in the course of their development of Quantum Electrodynamics. In brief, the final observable outcome of an event is determined by the superposition of all the various probability amplitudes associated with each and every possible way that the event can occur. In other words, each "route" along which an event can take place, each way it can happen, is given an abstract mathematical representation, a complex probability amplitude. All of these then combine—and interfere, as complex quantities are wont to do—to produce a net probability amplitude for the event to take place.

What follows is a greatly simplified version of that analysis.

4.11.1 QED

Feynman was rather unequivocal in his stance regarding the nature of light:

> I want to emphasize that light comes in this form—particles. It is very important to know that light behaves like particles, especially for those of you who have gone to school, where you were probably told something about light behaving like waves. I'm telling you the way it *does* behave—like particles.

For him "light is made of particles (as Newton originally thought)"; it's a stream of photons whose behavior *en masse* can be determined statistically. For example, if 100 photons are incident perpendicularly on a piece of glass in air, on average 4 will be reflected backward from the first surface they encounter. Which 4 cannot be known, and in fact how those particular 4 photons are selected is a mystery. What can be deduced and confirmed experimentally is that 4% of the incident light will be reflected (p. 121).

Feynman's analysis proceeds from a few general computational rules, with the ultimate justification being that it works; the scheme makes accurate predictions. (1) ***The probability amplitude associated with the occurrence of an event is the "sum" of the constituent probability amplitudes corresponding to each and every possible way the event can occur***. (2) ***Each such constituent probability amplitude is generally expressible as a complex quantity***. Rather than analytically combining these constituent probability amplitudes, we can use the phasor representation (p. 23) to approximate the summation and thereby arrive at a resultant probability amplitude. (3) ***The probability of occurrence of the event as a whole is proportional to the absolute square of the resultant probability amplitude***.

We can appreciate how all of this comes together by treating the reflection pictured in Fig. 4.67: a point source S illuminates a mirror, and light is subsequently scattered upward in every direction from every point on the mirror. We wish to determine the probability of a detector at P, recording the arrival of a photon. Here the classical perspective, with its familiar wavelet model, can be used as an analogue to provide guidance (and perhaps a little intellectual comfort, if you still believe in classical EM waves).

For simplicity, take the mirror to be a narrow strip (which is essentially one-dimensional); that doesn't change things conceptually. Divide it into a number of equal-sized lengths

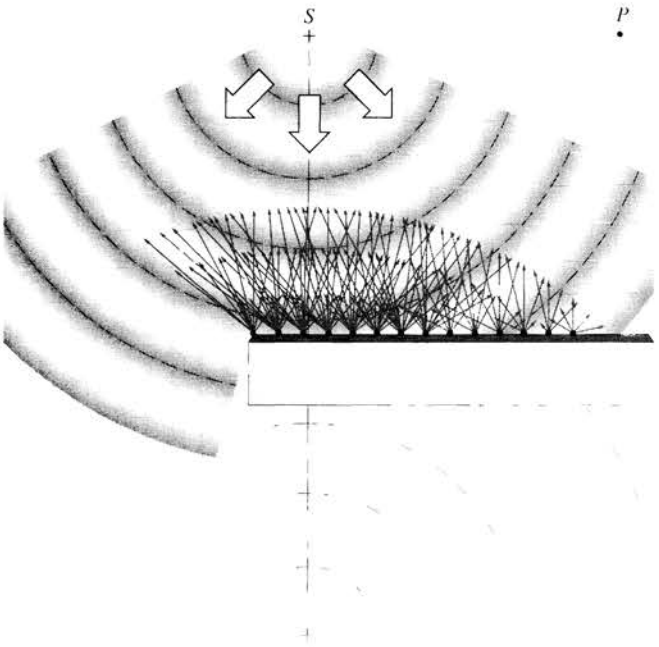


Figure 4.67 A schematic representation of reflection. A wave from S sweeps down and spreads across the surface of the mirror. Every atom on the interface subsequently scatters light back in all upward directions. And some of it ultimately arrives at P having come from every scatterer on the surface.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

(Fig. 4.68*a*), each of which establishes a possible path to *P*. (Of course, every atom on the surface is a scatterer, and so there are a multitude of paths, but the several we have drawn will do.) Classically, we know that every route from *S* to the mirror to *P* corresponds to the path of a scattered wavelet, and that the amplitude (E_{0j}) and phase of each such wavelet at *P* will determine the net resultant amplitude, E_0. As we saw with Fermat's Principle (p. 106), the optical path length from *S* to the mirror to *P* establishes the phase of each wavelet arriving at *P*. Moreover, the greater the *path length* is, the more the light spreads out (via the Inverse Square Law) and the smaller is the amplitude of the wavelet arriving at *P*.

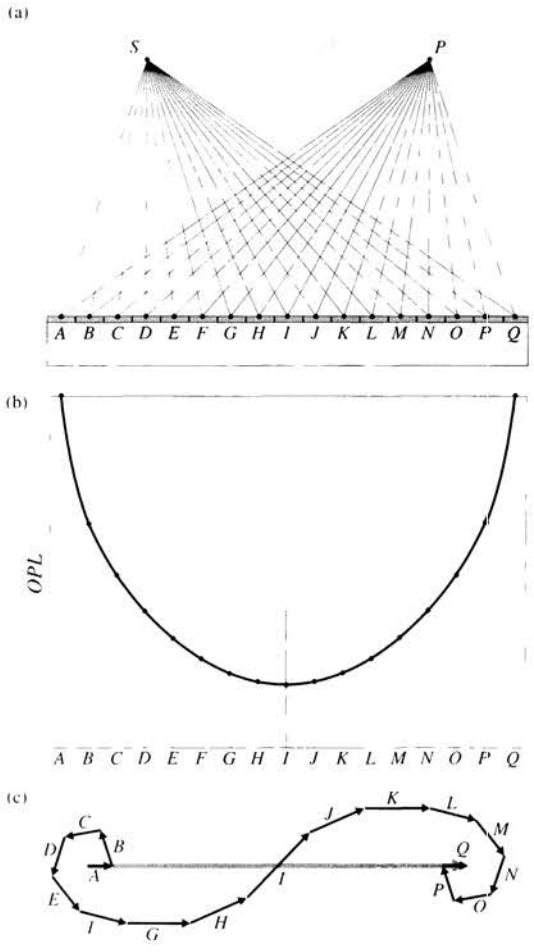


Figure 4.68 (*a*) Feynman's analysis of the problem of reflection via QED. (*a*) A number of paths from *S* to the mirror to *P*. (*b*) The *OPL* for light going from *S* to *P* along the paths depicted in (*a*). Each path has a probability amplitude associated with it. These add to produce a net amplitude.

Figure 4.68*b* is a plot of the *OPL* with its minimum at the observed path (*S-I-P*), for which $\theta_i = \theta_r$. A large change in *OPL*, as between (*S-A-P*) and (*S-B-P*) is accompanied by a large phase difference and a correspondingly large rotation of the phasors drawn in Fig. 4.68*c*. Going from *A* to *B* to *C* and so on to *I*, the optical path lengths decrease less and less rapidly, and each phasor leads the previous one by a smaller angle (set by the slope of the curve). In effect, the phasors to the left of *I* rotate counterclockwise from *A* to *I*. Since the *OPL* is a minimum at *I*, the phasors from that region are large and differ very little in phase angle. Going from *I* to *J* to *K* and so on to *Q*, the optical path lengths increase more and more rapidly, and each phasor lags the previous one by a larger angle. In effect, the phasors to the right of *I* rotate clockwise from *I* to *Q*.

In Fig. 4.68*c* the resultant amplitude is drawn from the starting tail to the ending tip, and classically it corresponds to the net electric field amplitude at *P*. The irradiance, *I*, is proportional to the square of the net field amplitude, and that, in turn, should be a measure of the likelihood of finding a photon when a detector is placed at *P*.

Let's move beyond the classical ideas of scattered wavelets and electric fields (nonetheless being guided by them) and construct a quantum-mechanical treatment. Photons can go from *S* to the mirror to *P* along each of an innumerable number of distinct paths. It's reasonable to assume that each such path makes a specific contribution to the end result; an exceedingly long route out to the very edge of the mirror and back to *P* should contribute differently than a more direct route. Following Feynman, we associate some (as yet unspecified) complex quantity, a constituent **quantum-mechanical (QM) probability amplitude**, with each possible path. Each such constituent QM probability amplitude can be represented as a phasor whose angle is determined by the total time of flight from *S* to the mirror to *P*, and whose size is determined by the path length traversed. (Of course, this is just what obtained with each phasor in Fig. 4.68*c*. Still there are convincing reasons why the classical *E*-field cannot be the QM probability amplitude.) The total QM probability amplitude is the sum of all such phasors corresponding to all possible paths, and that is analogous to the resultant phasor in Fig. 4.68*c*.

Now re-label Fig. 4.68*c* so that it represents the quantum-mechanical formulation. Clearly, ***most of the length of the resultant QM probability amplitude arises from contributions in the immediate vicinity of path S-I-P, where the constituent phasors are large and nearly in-phase***. Most of the

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

accumulated probability for light to go from S to P via reflection arises along, and immediately adjacent to, path S-I-P. The regions at the ends of the mirror contribute very little because the phasors from those areas form tight spirals at both extremes (Fig. 4.68c). Covering the ends of the mirror will have little effect on the length of the resultant amplitude and therefore little effect on the amount of light reaching P. Keep in mind that this diagram is rather crude; instead of 17 routes from S to P, there are billions of possible paths, and the phasors on both ends of the spiral wind around countless times.

QED predicts that light emitted by a point source S reflects out to P from all across the mirror, but that the most likely route is S-I-P, in which case $\theta_i = \theta_r$. With your eye at P looking into the mirror, you see one sharp image of S.

4.11.2 Photons and the Laws of Reflection and Refraction

Suppose that light consists of a stream of photons and consider one such quantum that strikes the interface between two dielectric media (e.g., air and glass) at an angle θ_i. That photon is absorbed by an atom (e.g., in the glass), and an identical photon is subsequently transmitted at an angle θ_t. We know that if this were just one out of billions of such quanta in a narrow laserbeam, it would conform to Snell's Law. To explore this behavior, let's examine the dynamics associated with the odyssey of our single photon. Recall Eq. (3.54), namely, $p = h/\lambda$ and so its vector momentum would be

$$\vec{\mathbf{p}} = \hbar\vec{\mathbf{k}}$$

where $\vec{\mathbf{k}}$ is the propagation vector and $\hbar \equiv h/2\pi$. Consequently, the incident and transmitted momenta are $\vec{\mathbf{p}}_i = \hbar\vec{\mathbf{k}}_i$ and $\vec{\mathbf{p}}_t = \hbar\vec{\mathbf{k}}_t$ respectively. We assume (without much justification) that while the material in the vicinity of the interface affects the component of momentum perpendicular to the interface it leaves the parallel component unchanged. Indeed, we know experimentally that linear momentum perpendicular to the interface can be transferred to a medium from a light beam (Section 3.3.4). The statement of conservation of the component of momentum parallel to the interface for a single photon takes the form

$$p_i \sin\theta_i = p_t \sin\theta_t$$

We are at an important juncture here; classically the momentum of a material particle depends on its speed. When $n_t > n_i$ it follows (from the Snell's Law and the above equation) that $p_t > p_i$, and the particles of light must presumably speed up. Indeed, the first derivation of the Law of Refraction, published by René Descartes (1637), wrongly treated light as a stream of particles that sped up as it entered the optically more dense medium (see Problem 4.10). By contrast, the first person to measure the shortening of the wavelength of light as it entered a more dense medium was probably Thomas Young ($\approx$1802).* He rightly inferred that the speed of a light beam was thereby actually reduced: $v < c$.

We now know from Quantum Mechanics that the speed of a photon is always c and that its momentum depends on its wavelength and not its speed. Thus

$$\frac{h}{\lambda_i}\sin\theta_i = \frac{h}{\lambda_t}\sin\theta_t$$

Multiplying both sides by c/ν, we get Snell's Law.

Do keep in mind that the above analysis is a bit simplistic, but it is appealing pedagogically.

*Foucault's definitive experiments proving the point were done in 1850.

PROBLEMS

Complete solutions to all problems—except those with an asterisk—can be found in the back of the book .

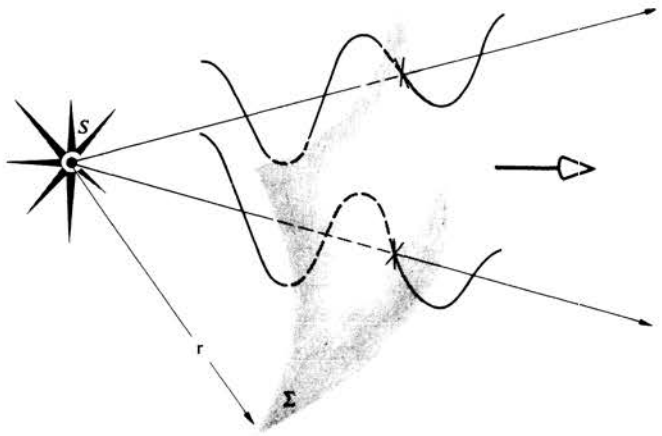
4.1 Work your way through an argument using dimensional analysis to establish the λ^{-4} dependence of the percentage of light scattered in Rayleigh Scattering. Let E_{0i} and E_{0s} be the incident and scattered amplitudes, the latter at a distance r from the scatterer. Assume $E_{0s} \propto E_{0i}$ and $E_{0s} \propto 1/r$. Furthermore, plausibly assume that the scattered amplitude is proportional to the volume, V, of the scatterer; within limits this is reasonable. Determine the units of the constant of proportionality.

4.2* A white floodlight beam crosses a large volume containing a tenuous molecular gas mixture of mostly oxygen and nitrogen. Compare the relative amount of scattering occurring for the yellow (580 nm) component with that of the violet (400 nm) component.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

4.3* Figure P.4.3 depicts light emerging from a point source. It shows three different representations of radiant energy streaming outward. Identify each one and discuss its relationship to the others.

Figure P.4.3 A segment of a spherical wave.



4.4 The equation for a driven damped oscillator is

$$m_e\ddot{x} + m_e\gamma\dot{x} + m_e\omega_0^2 x = q_e E(t)$$

(a) Explain the significance of each term.

(b) Let $E = E_0 e^{i\omega t}$ and $x = x_0 e^{i(\omega t - \alpha)}$, where E_0 and x_0 are real quantities. Substitute into the above expression and show that

$$x_0 = \frac{q_e E_0}{m_e} \frac{1}{[(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2]^{1/2}}$$

(c) Derive an expression for the phase lag, α, and discuss how α varies as ω goes from $\omega \ll \omega_0$ to $\omega = \omega_0$ to $\omega \gg \omega_0$.

4.5 Imagine that we have a nonabsorbing glass plate of index n and thickness Δy, which stands between a source S and an observer P.

(a) If the unobstructed wave (without the plate present) is $E_u = E_0 \exp i\omega(t - y/c)$, show that with the plate in place the observer sees a wave

$$E_p = E_0 \exp i\omega[t - (n - 1)\,\Delta y/c - y/c]$$

(b) Show that if either $n \approx 1$ or Δy is very small, then

$$E_p = E_u + \frac{\omega(n - 1)\,\Delta y}{c} E_u e^{-i\pi/2}$$

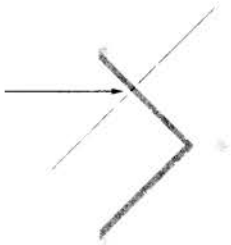
The second term on the right may be envisioned as the field arising from the oscillators in the plate.

4.6* A very narrow laserbeam is incident at an angle of 58° on a horizontal mirror. The reflected beam strikes a wall at a spot 5.0 m away from the point of incidence where the beam hit the mirror. How far horizontally is the wall from that point of incidence?

4.7* On entering the tomb of FRED the Hero of Nod, you find yourself in a dark closed chamber with a small hole in a wall 3.0 m up from the floor. Once a year, on FRED's birthday, a beam of sunlight enters via the hole, strikes a small polished gold disk on the floor 4.0 m from the wall and reflects off it, lighting up a great diamond embedded in the forehead of a glorious statue of FRED, 20 m from the wall. Roughly how tall is the statue?

4.8* Figure P.4.8 shows what's called a corner mirror. Determine the direction of the exiting ray with respect to the incident ray.

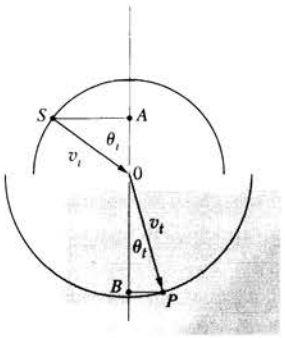
Figure P.4.8



4.9 Calculate the transmission angle for a ray incident in air at 30° on a block of crown glass ($n_g = 1.52$).

4.10* The construction in Fig. P.4.10 corresponds to Descartes's derivation of the Law of Refraction. Light moves from S to O in the same time it travels from O to P. Moreover, its transverse momentum is unchanged on traversing the interface. Use all of this to "derive" Snell's Law.

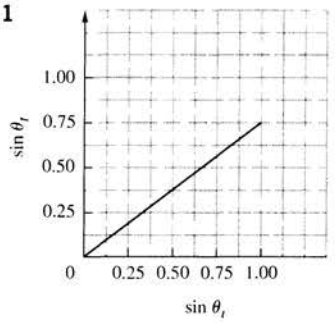
Figure P.4.10



Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

4.11* Figure P.4.11 is a plot of the sine of the angle-of-incidence versus the sine of the transmission angle measured as light passed from air into a more optically dense medium. Discuss the curve. What is the significance of the slope of the line? Guess at what the dense medium might be.

Figure P.4.11



4.12* A ray of yellow light from a sodium discharge lamp falls on the surface of a diamond in air at 45°. If at that frequency $n_d = 2.42$, compute the angular deviation suffered upon transmission.

4.13* Given an interface between water ($n_w = \frac{4}{3}$) and glass ($n_g = \frac{3}{2}$), compute the transmission angle for a beam incident in the water at 45°. If the transmitted beam is reversed so that it impinges on the interface, show that $\theta_t = 45°$.

4.14 A beam of 12-cm planar microwaves strikes the surface of a dielectric at 45°. If $n_{ti} = \frac{4}{3}$, compute (a) the wavelength in the transmitting medium, and (b) the angle θ_t.

4.15* Light of wavelength 600 nm in vacuum enters a block of glass where $n_g = 1.5$. Compute its wavelength in the glass. What color would it appear to someone embedded in the glass (see Table 3.4)?

4.16* A laserbeam impinges on an air–liquid interface at an angle of 55°. The refracted ray is observed to be transmitted at 40°. What is the refractive index of the liquid?

4.17* An underwater swimmer shines a beam of light up toward the surface. It strikes the air–water interface at 35°. At what angle will it emerge into the air?

4.18 Make a plot of θ_i versus θ_t for an air–glass boundary where $n_{ga} = 1.5$. Discuss the shape of the curve.

4.19* A laserbeam having a diameter D in air strikes a piece of glass (n_g) at an angle θ_i. What is the diameter of the beam in the glass?

4.20* An exceedingly narrow beam of white light is incident at 60.0° on a sheet of glass 10.0 cm thick in air. The index of refraction for red light is 1.505 and for violet light it's 1.545. Determine the approximate diameter of the emerging beam.

4.21* Prove that to someone looking straight down into a swimming pool, any object in the water will appear to be at $\frac{3}{4}$ of its true depth.

4.22* A laserbeam impinges on the top surface of a 2.00-cm-thick parallel glass ($n = 1.50$) plate at an angle of 35°. How long is the actual path through the glass?

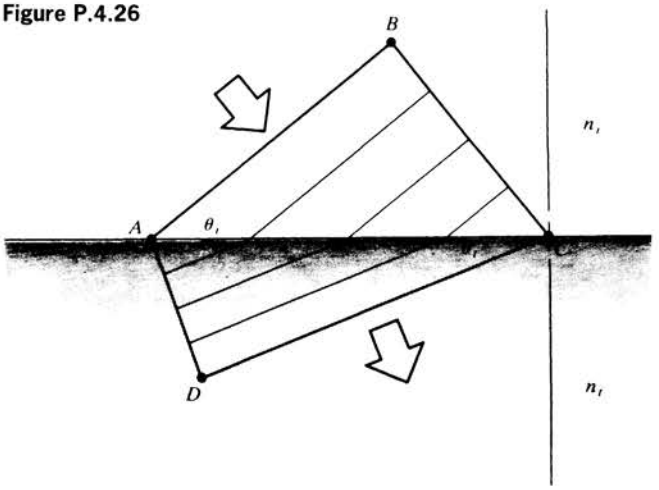
4.23* Light is incident in the air on an air–glass interface. If the index of refraction of the glass is 1.70, find the incident angle such that the transmission angle is to equal $\frac{1}{2}\theta_i$.

4.24* Suppose that you focus a camera with a close-up bellows attachment directly down on a letter printed on this page. The letter is then covered with a 1.00-mm-thick microscope slide ($n = 1.55$). How high must the camera be raised in order to keep the letter in focus?

4.25* A coin is resting on the bottom of a tank of water ($n_W = 1.33$) 1.00 m deep. On top of the water floats a layer of benzene ($n_b = 1.50$), which is 20.0 cm thick. Looking down nearly perpendicularly, how far beneath the topmost surface does the coin appear? Draw a ray diagram.

4.26 In Fig. P.4.26 the wavefronts in the incident medium match the fronts in the transmitting medium everywhere on the interface—a concept known as *wavefront continuity*. Write expressions for the number of waves per unit length along the interface in terms of θ_i and λ_i in one case and θ_t and λ_t in the other. Use these to derive Snell's Law. Do you think Snell's Law applies to sound waves? Explain.

Figure P.4.26

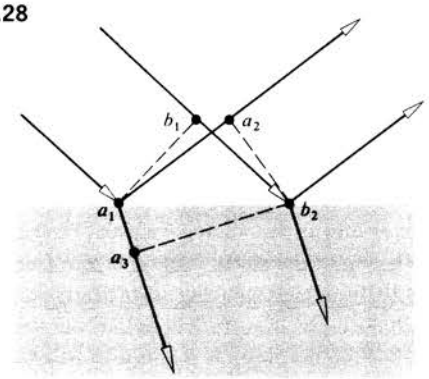


Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

4.27* With the previous problem in mind, return to Eq. (4.19) and take the origin of the coordinate system in the plane-of-incidence and on the interface (Fig. 4.39). Show that that equation is then equivalent to equating the x-components of the various propagation vectors. Show that it is also equivalent to the notion of wavefront continuity.

4.28 Making use of the ideas of equal transit times between corresponding points and the orthogonality of rays and wavefronts, derive the law of reflection and Snell's Law. The ray diagram of Fig. P.4.28 should be helpful.

Figure P.4.28



4.29 Starting with Snell's Law, prove that the vector refraction equation has the form

$$n_t\hat{\mathbf{k}}_t - n_i\hat{\mathbf{k}} = (n_t \cos\theta_t - n_i \cos\theta_i)\hat{\mathbf{u}}_n \qquad [4.7]$$

4.30 Derive a vector expression equivalent to the Law of Reflection. As before, let the normal go from the incident to the transmitting medium, even though it obviously doesn't really matter.

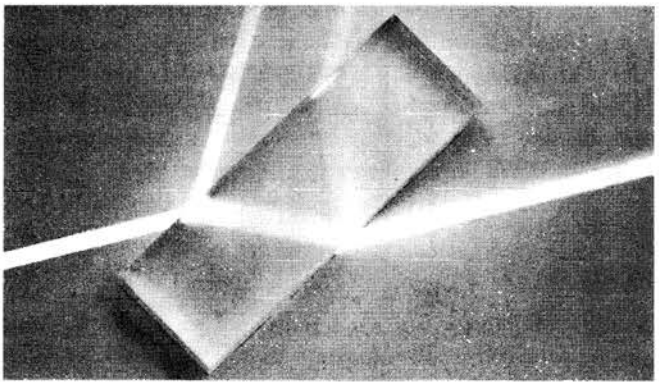
4.31 In the case of reflection from a planar surface, use Fermat's Principle to prove that the incident and reflected rays share a common plane with the normal $\hat{\mathbf{u}}_n$, namely, the plane-of-incidence.

4.32* Derive the Law of Reflection, $\theta_i = \theta_r$, by using the calculus to minimize the transit time, as required by Fermat's Principle.

4.33* According to the mathematician Hermann Schwarz, there is one triangle that can be inscribed within an acute triangle such that it has a minimal perimeter. Using two planar mirrors, a laserbeam, and Fermat's Principle, explain how you can show that this inscribed triangle has its vertices at the points where the altitudes of the acute triangle intersect its corresponding sides.

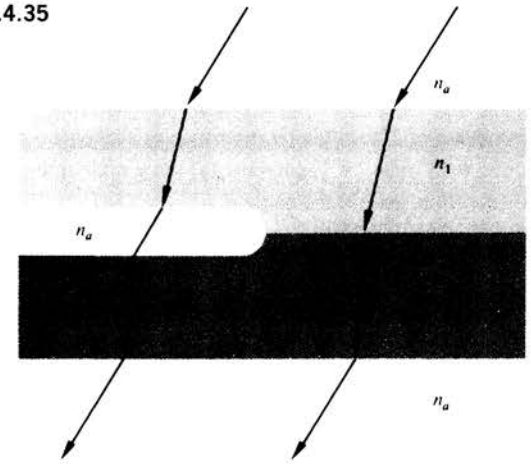
4.34 Show analytically that a beam entering a planar transparent plate, as in Fig. P.4.34, emerges parallel to its initial direction. Derive an expression for the lateral displacement of the beam. Incidentally, the incoming and outgoing rays would be parallel even for a stack of plates of different material.

Figure P.4.34



4.35* Show that the two rays that enter the system in Fig. P.4.35 parallel to each other emerge from it being parallel.

Figure P.4.35



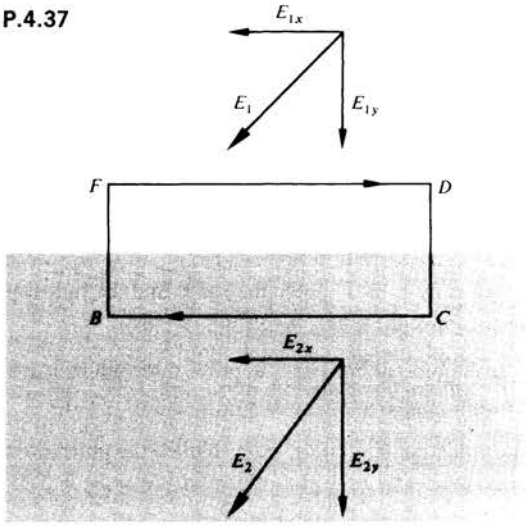
4.36 Discuss the results of Problem 4.34 in the light of Fermat's Principle: that is, how does the relative index n_{21} affect things? To see the lateral displacement, look at a broad source through a thick piece of glass ($\approx\frac{1}{4}$ inch) or a stack (four will do) of microscope slides *held at an angle*. There will be an obvious shift between the region of the source seen directly and the region viewed through the glass.

4.37 Show that even in the nonstatic case the tangential component of the electric field intensity $\mathbf{E}$ is continuous across an interface.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

[*Hint:* Using Fig. P.4.37 and Eq. (3.5), shrink sides *FB* and *CD*, thereby letting the area bounded go to zero.]

Figure P.4.37



4.38 Suppose a lightwave that is linearly polarized in the plane-of-incidence impinges at 30° on a crown-glass ($n_g = 1.52$) plate in air. Compute the appropriate amplitude reflection and transmission coefficients at the interface. Compare your results with Fig. 4.39.

4.39 Derive Eqs. (4.42) through (4.45) for $r_\perp$, $r_\parallel$, $t_\perp$, and $t_\parallel$.

4.40* A beam of light in air strikes the surface of a smooth piece of plastic having an index of refraction of 1.55 at an angle with the normal of 20.0°. The incident light has component *E*-field amplitudes parallel and perpendicular to the plane-of-incidence of 10.0 V/m and 20.0 V/m, respectively. Determine the corresponding reflected field amplitudes.

4.41* A laserbeam is incident on the interface between air and some dielectric of index n. For small values of θ_i show that $\theta_t = \theta_i/n$. Use this and Eq. (4.42) to establish that at near-normal incidence $[-r_\perp]_{\theta_i \approx 0} = (n-1)/(n+1)$.

4.42* Light is incident in air perpendicularly on a sheet of crown glass having an index of refraction of 1.522. Determine both the reflectance and the transmittance.

4.43* A beam of quasimonochromatic light having an irradiance of 500 W/m^2 is incident in air perpendicularly on the surface of a tank of water ($n_w = 1.333$). Determine the transmitted irradiance.

4.44* Quasimonochromatic light having an irradiance of 400 W/m^2 is incident normally on the cornea ($n_c = 1.376$) of the human eye. If the person is swimming under the water ($n_w = 1.33$), determine the transmitted irradiance into the cornea.

4.45* Compare the amplitude reflection coefficients for an air–water ($n_w = 4/3$) interface with that of an air–crown glass ($n_g = 3/2$) interface, both at near-normal incidence. What are the corresponding ratios of the reflected to the incident irradiances?

4.46* Use Eq. (4.42) and the power series expansion of the sine function to establish that at near-normal incidence we can obtain a better approximation than the one in Problem 4.41, which is $[-r_\perp]_{\theta_i \approx 0} = (n-1)/(n+1)$, namely

$$[-r_\perp]_{\theta_i \approx 0} = \left(\frac{n-1}{n+1}\right)\left(1 + \frac{\theta_i^2}{n}\right)$$

4.47* Establish that at near-normal incidence the equation

$$[r_\parallel]_{\theta_i \approx 0} = \left(\frac{n-1}{n+1}\right)\left(1 - \frac{\theta_i^2}{n}\right)$$

is a good approximation. [*Hint:* Use the results of the previous problem, Eq. (4.43), and the power series expansions of the sine and cosine functions.]

4.48* Prove that for a vacuum-dielectric interface at glancing incidence $r_\perp \to -1$, as in Fig. 4.41.

4.49* In Fig. 4.41 the curve of $r_\perp$ approaches -1.0 as the angle-of-incidence approaches 90°. Prove that if $\alpha_\perp$ is the angle the curve makes with the vertical at $\theta_i = 90°$, then

$$\tan \alpha_\perp = \frac{\sqrt{n^2 - 1}}{2}$$

[*Hint:* First show that $d\theta_t/d\theta_i = 0$.]

4.50 Prove that

$$t_\perp + (-r_\perp) = 1 \qquad [4.49]$$

for all θ_i, first from the boundary conditions and then from the Fresnel Equations.

4.51* Verify that

$$t_\perp + (-r_\perp) = 1 \qquad [4.49]$$

for $\theta_i = 30°$ at a crown–glass and air interface ($n_u = 1.52$).

4.52* Calculate the critical angle beyond which there is total internal reflection at an air–glass ($n_g = 1.5$) interface. Compare this result with that of Problem 4.15.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

4.53* Referring back to Problem 4.18, note that as θ_i increases θ_t increases. Prove that the maximum value θ_t may have is θ_c.

4.54* What is the critical angle for total internal reflection for diamond? What, if anything, does the critical angle have to do with the luster of a well-cut diamond?

4.55* Using a block of a transparent, unknown material, it is found that a beam of light inside the material is totally internally reflected at the air–block interface at an angle of 48.0°. What is its index of refraction?

4.56* A prism, ABC, is configured such that angle $BCA = 90°$ and angle $CBA = 45°$. What is the minimum value of its index of refraction if, while immersed in air, a beam traversing face AC is to be totally internally reflected from face BC?

4.57* A fish looking straight up toward the smooth surface of a pond receives a cone of rays and sees a circle of light filled with the images of sky and birds and whatever else is up there. This bright circular field is surrounded by darkness. Explain what is happening and compute the cone-angle.

4.58* A glass block having an index of 1.55 is covered with a layer of water of index 1.33. For light traveling in the glass, what is the critical angle at the interface?

4.59 Derive an expression for the speed of the evanescent wave in the case of internal reflection. Write it in terms of c, n_i, and θ_i.

4.60 Light having a vacuum wavelength of 600 nm, traveling in a glass ($n_g = 1.50$) block, is incident at 45° on a glass–air interface. It is then totally internally reflected. Determine the distance into the air at which the amplitude of the evanescent wave has dropped to a value of $1/e$ of its maximum value at the interface.

4.61 Figure P.4.61 shows a laserbeam incident on a wet piece of filter paper atop a sheet of glass whose index of refraction is to be measured—the photograph shows the resulting light pattern. Explain what is happening and derive an expression for n_i in terms of R and d.

4.62 Consider the common mirage associated with an inhomogeneous distribution of air situated above a warm roadway. Envision the bending of the rays as if it were instead a problem in total internal reflection. If an observer, at whose head $n_a = 1.000\,29$, sees an apparent wet spot at $\theta_i \geq 88.7°$ down the road, find the index of the air immediately above the road.

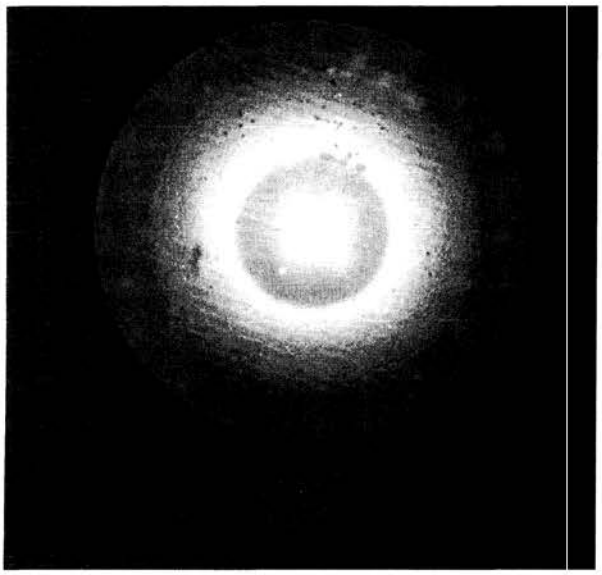
4.63* Use the Fresnel Equations to prove that light incident at $\theta_p = \frac{1}{2}\pi - \theta_i$ results in a reflected beam that is indeed polarized.

4.64 Show that $\tan\theta_p = n_t/n_i$ and calculate the polarization angle for external incidence on a plate of crown glass ($n_g = 1.52$) in air.

4.65* Beginning with Eq. (4.38), show that for two dielectric media, in general $\tan\theta_p = [\epsilon_i(\epsilon_t\mu_t - \epsilon_t\mu_t)/\epsilon_i(\epsilon_t\mu_t - \epsilon_i\mu_i)]^{1/2}$.

4.66 Show that the polarization angles for internal and external reflection at a given interface are complementary, that is, $\theta_p + \theta'_p = 90°$ (see Problem 4.64).

4.67 It is often useful to work with the *azimuthal angle* γ, which is



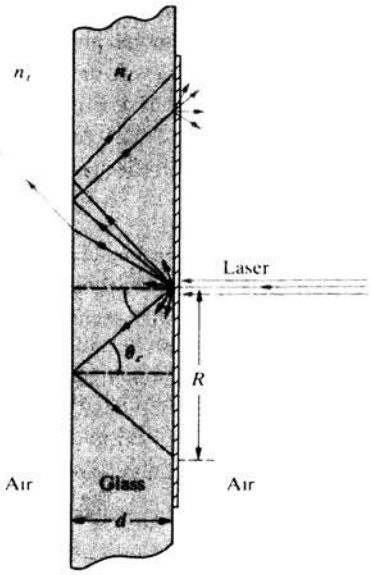


Figure P.4.61 (Photo and diagram courtesy S. Reich, The Weizmann Institute of Science, Israel.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

defined as the angle between the plane-of-vibration and the plane-of-incidence. Thus for linearly polarized light,

$$\tan \gamma_i = [E_{0i}]_\perp / [E_{0i}]_\parallel \tag{4.92}$$

$$\tan \gamma_t = [E_{0t}]_\perp / [E_{0t}]_\parallel \tag{4.93}$$

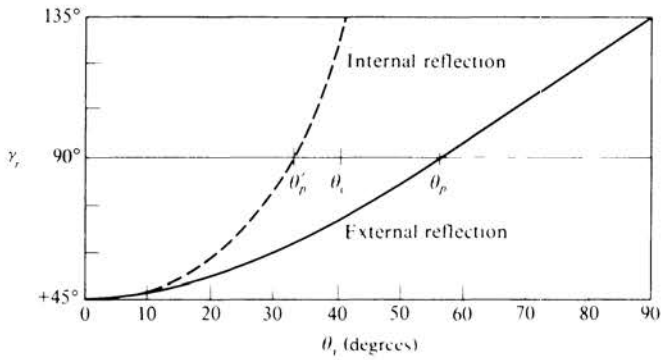
and

$$\tan \gamma_r = [E_{0r}]_\perp / [E_{0r}]_\parallel \tag{4.94}$$

Figure P.4.67 is a plot of γ_r versus θ_i for internal and external reflection at an air–glass interface ($n_{ga} = 1.51$), where $\gamma_i = 45°$. Verify a few of the points on the curves and in addition show that

$$\tan \gamma_r = -\frac{\cos(\theta_i - \theta_t)}{\cos(\theta_i + \theta_t)} \tan \gamma_i \tag{4.95}$$

Figure P.4.67



4.68* Making use of the definitions of the azimuthal angles in Problem 4.67, show that

$$R = R_\parallel \cos^2 \gamma_i + R_\perp \sin^2 \gamma_i \tag{4.96}$$

and

$$T = T_\parallel \cos^2 \gamma_i + T_\perp \sin^2 \gamma_i \tag{4.97}$$

4.69 Make a sketch of $R_\perp$ and $R_\parallel$ for $n_i = 1.5$ and $n_t = 1$ (i.e., internal reflection).

4.70 Show that

$$T_\parallel = \frac{\sin 2\theta_i \sin 2\theta_t}{\sin^2(\theta_i + \theta_t)\cos^2(\theta_i - \theta_t)} \tag{4.98}$$

and

$$T_\perp = \frac{\sin 2\theta_i \sin 2\theta_t}{\sin^2(\theta_i + \theta_t)} \tag{4.99}$$

4.71* Using the results of Problem 4.70, that is, Eqs. (4.98) and (4.99), show that

$$R_\parallel + T_\parallel = 1 \tag{4.65}$$

and

$$R_\perp + T_\perp = 1 \tag{4.66}$$

4.72 Suppose that we look at a source perpendicularly through a stack of N microscope slides. The source seen through even a dozen slides will be noticeably darker. Assuming negligible absorption, show that the total transmittance of the stack is given by

$$T_t = (1 - R)^{2N}$$

and evaluate T_t for three slides in air.

4.73 Making use of the expression

$$I(y) = I_0 e^{-\alpha y} \tag{4.78}$$

for an absorbing medium, we define a quantity called the *unit transmittance* T_1. At normal incidence, Eq. (4.55), $T = I_t/I_i$, and thus when $y = 1$, $T_1 \equiv I(1)/I_0$. If the total thickness of the slides in the previous problem is d and if they now have a transmittance per unit length T_1, show that

$$T_t = (1 - R)^{2N} (T_1)^d$$

4.74 Show that at normal incidence on the boundary between two dielectrics, as $n_{ti} \to 1$, $R \to 0$, and $T \to 1$. Moreover, prove that as $n_{ti} \to 1$, $R_\parallel \to 0$, $R_\perp \to 0$, $T_\parallel \to 1$, and $T_\perp \to 1$ for all θ_i. Thus as the two media take on more similar indices of refraction, less and less energy is carried off in the reflected wave. It should be obvious that when $n_{ti} = 1$ there will be no interface and no reflection.

4.75* Derive the expressions for $r_\perp$ and $r_\parallel$ given by Eqs. (4.70) and (4.71).

4.76 Show that when $\theta_i > \theta_c$ at a dielectric interface, $r_\parallel$ and $r_\perp$ are complex and $r_\perp r_\perp^* = r_\parallel r_\parallel^* = 1$.

4.77 Figure P.4.77 depicts a ray being multiply reflected by a transparent dielectric plate (the amplitudes of the resulting fragments are indicated). As in Section 4.5, we use the primed coefficient notation because the angles are related by Snell's Law.

(a) Finish labeling the amplitudes of the last four rays.

(b) Show, using the Fresnel Equations, that

$$t_\parallel t_\parallel' = T_\parallel \tag{4.100}$$

$$t_\perp t_\perp' = T_\perp \tag{4.101}$$

$$r_\parallel^2 = r_\parallel'^2 = R_\parallel \tag{4.102}$$

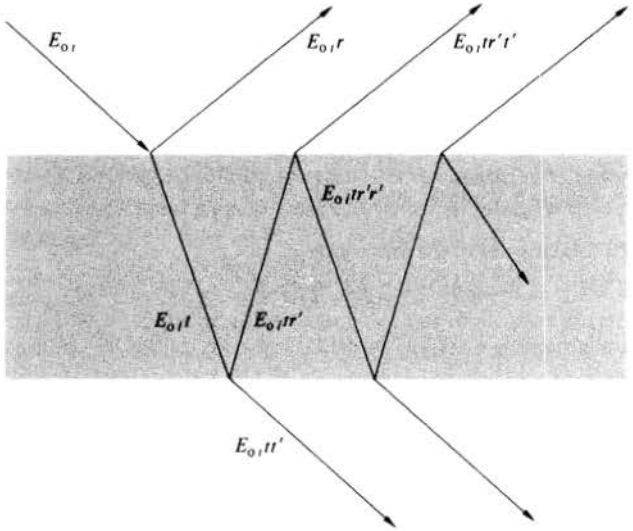
and

$$r_\perp^2 = r_\perp'^2 = R_\perp \tag{4.103}$$

4.78* A wave, linearly polarized in the plane-of-incidence, impinges on the interface between two dielectric media. If $n_i > n_t$ and $\theta_i = \theta_p'$, there is no reflected wave, that is, $r_\parallel'(\theta_p') = 0$. Using Stokes's tech-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Figure P.4.77

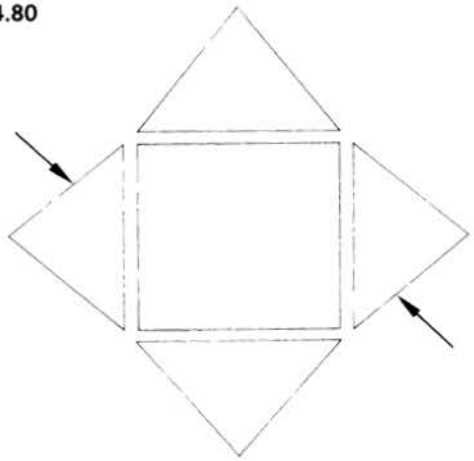


nique, start from scratch to show that $t_{\parallel}(\theta_p)t'_{\parallel}(\theta'_p) = 1$, $r_{\parallel}(\theta_p) = 0$, and $\theta_t = \theta_p$ (Problem 4.66). How does this compare with Eq. (4.100)?

4.79 Making use of the Fresnel Equations, show that $t_{\parallel}(\theta_p)t'_{\parallel}(\theta'_p) = 1$, as in the previous problem.

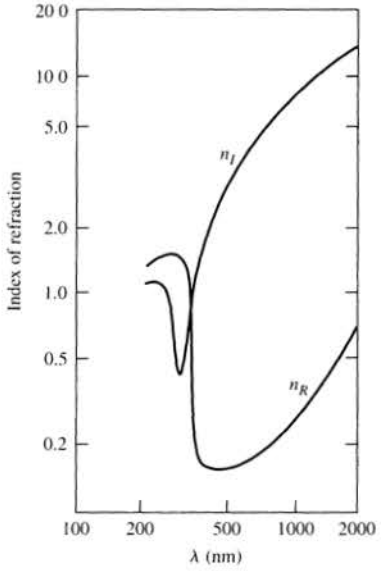
4.80 Figure P.4.80 depicts a glass cube surrounded by four glass prisms in very close proximity to its sides. Sketch in the paths that will be taken by the two rays shown and discuss a possible application for the device.

Figure P.4.80



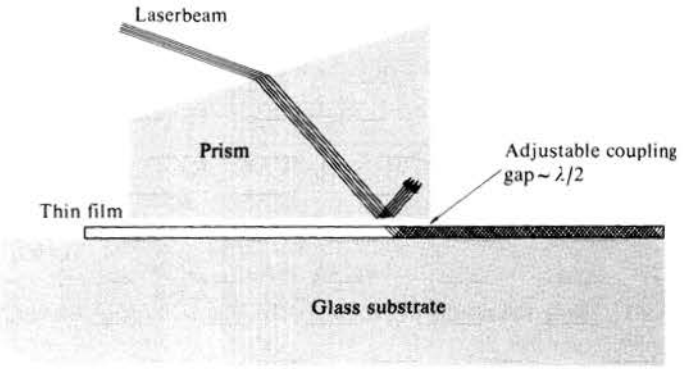
4.81 Figure P.4.81 is a plot of n_I and n_R versus λ for a common metal. Identify the metal by comparing its characteristics with those considered in the chapter and discuss its optical properties.

Figure P.4.81



4.82 Figure P.4.82 shows a prism-coupler arrangement developed at the Bell Telephone Laboratories. Its function is to feed a laserbeam into a thin (0.000 01-inch) transparent film, which then serves as a sort of waveguide. One application is that of thin-film laserbeam circuitry—a kind of integrated optics. How do you think it works?

Figure P.4.82



Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

5 Geometrical Optics

5.1 Introductory Remarks

The surface of an object that is either self-luminous or externally illuminated behaves as if it consisted of a very large number of radiating point sources. Each of these emits spherical waves; rays emanate radially in the direction of energy flow, that is, in the direction of the Poynting vector. In this case, the rays *diverge* from a given point source S, whereas if the spherical wave were collapsing to a point, the rays would of course be *converging*. Generally, one deals only with a small portion of a wavefront. ***A point from which a portion of a spherical wave diverges, or one toward which the wave segment converges, is known as a focus of the bundle of rays.***

Figure 5.1 depicts a point source in the vicinity of some arrangement of reflecting and refracting surfaces representing an *optical system*. Of the infinity of rays emanating from S, generally speaking, only one will pass through an arbitrary point in space. Even so, it is possible to arrange for an infinite number of rays to arrive at a certain point P, as in Fig. 5.1. If for a cone of rays coming from S there is a corresponding cone of rays passing through P, the system is said to be **stigmatic** for these two points. The energy in the cone (apart from some inadvertent losses due to reflection, scattering, and absorption) reaches P, which is then referred to as a **perfect image** of S. The wave could conceivably arrive to form a finite patch of light, or **blur spot**, about P; it would still be an image of S but no longer a perfect one.

It follows from the Principle of Reversibility (p. 110) that a point source placed at P would be equally well imaged at S, and accordingly the two are spoken of as **conjugate points**. In an *ideal optical system*, every point of a three-dimensional region will be perfectly (or stigmatically) imaged in another region, the former being the **object space**, the latter the **image space**.

Most commonly, the function of an optical device is to collect and reshape a portion of the incident wavefront, often with the ultimate purpose of forming an image of an object. Notice that inherent in realizable systems is the limitation of being unable to collect all the emitted light; a system generally accepts only a segment of the wavefront. As a result, there will always be an apparent deviation from rectilinear propagation even in homogeneous media—the waves will be *diffracted*. The attainable degree of perfection of a real imaging optical system will be **diffraction-limited** (there will always be a blur spot, p. 467). As the wavelength of the radiant energy decreases in comparison to the physical dimensions of the optical system, the effects of diffraction become less significant. In the conceptual limit as $\lambda_0 \to 0$, rectilinear propagation obtains in homogeneous media, and we have the idealized domain of **Geometrical Optics**.* Behavior that is specifically attributable to the wave nature of light (e.g., interference and diffraction) would no longer be observable. In many situations, the great simplicity arising from the approximation of Geometrical Optics more than compensates for its inaccuracies. In short, *the subject treats the controlled manipulation of wavefronts (or rays) by means of the interpositioning of reflecting and/or refracting bodies, neglecting any diffraction effects.*

Physical Optics deals with situations in which the nonzero wavelength of light must be reckoned with. Analogously, when the de Broglie wavelength of a material object is negligible, we have *Classical Mechanics*; when it is not, we have the domain of *Quantum Mechanics*.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

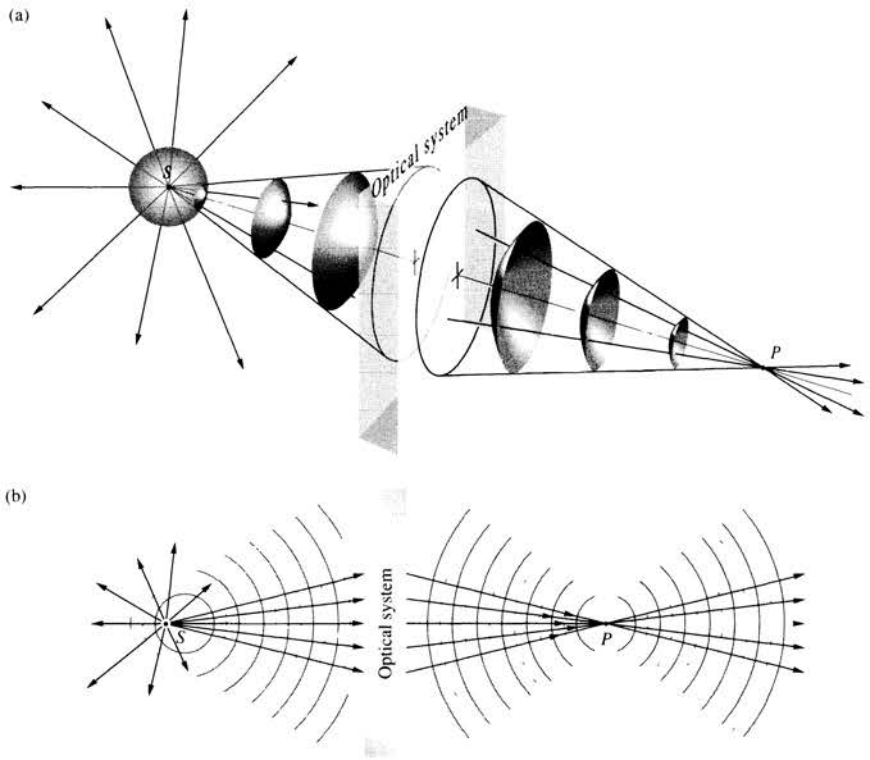


Figure 5.1 Conjugate foci. (*a*) A point source *S* sends out spherical waves. A cone of rays enters an optical system that inverts the wavefronts, causing them to converge on point *P*. (*b*) In cross section rays diverge from *S*, and a portion of them converge to *P*. If nothing stops the light at *P*, it continues on.

5.2 Lenses

The lens is no doubt the most widely used optical device, and that's not even considering the fact that we see the world through a pair of them. Human-made lenses date back at least to the burning-glasses of antiquity, which, as the name implies, were used to start fires long before the advent of matches. In the most general terms, **a lens is a refracting device (i.e., a discontinuity in the prevailing medium) that reconfigures a transmitted energy distribution**. That much is true whether we are dealing with UV, lightwaves, IR, microwaves, radiowaves, or even sound waves.

The configuration of a lens is determined by the required reshaping of the wavefront it is designed to perform. Point sources are basic, and so it is often desirable to convert diverging spherical waves into a beam of plane waves. Flashlights, projectors, and searchlights all do this in order to keep the beam from spreading out and weakening as it progresses. In just the reverse, it's frequently necessary to collect incoming parallel rays and bring them together at a point, thereby focusing the energy, as is done with a burning-glass or a telescope lens. Moreover, since the light reflected from someone's face scatters out from billions of point sources, a lens that causes each diverging wavelet to converge could form an image of that face (Fig. 5.2).

5.2.1 Aspherical Surfaces

To see how a lens works, imagine that we interpose in the path of a wave a transparent substance in which the wave's speed is different than it was initially. Figure 5.3*a* presents a cross-sectional view of a diverging spherical wave traveling in an inci-



Figure 5.2 A person's face, like everything else we ordinarily see in reflected light, is covered with countless atomic scatterers.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

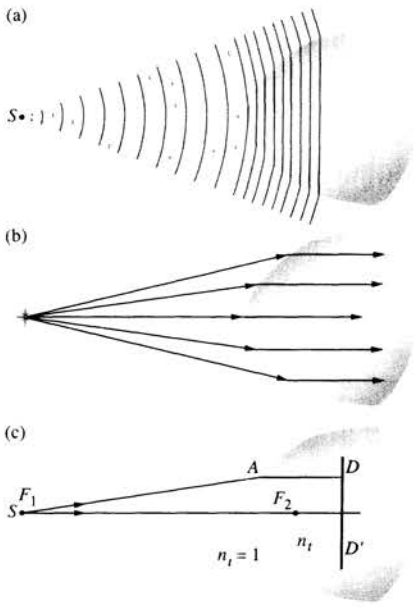


Figure 5.3 A hyperbolic interface between air and glass. (*a*) The wavefronts bend and straighten out. (*b*) The rays become parallel. (*c*) The hyperbola is such that the optical path from *S* to *A* to *D* is the same no matter where *A* is.

dent medium of index n_i impinging on the curved interface of a transmitting medium of index n_t. When n_t is greater than n_i, the wave slows upon entering the new substance. The central area of the wavefront travels more slowly than its outer extremities, which are still moving quickly through the incident medium. These extremities overtake the midregion, continuously flattening the wavefront. If the interface is properly configured, the spherical wavefront bends into a plane wave. The alternative ray representation is shown in Fig. 5.3*b*; the rays simply bend toward the local normal upon entering the more dense medium, and if the surface configuration is just right, the rays emerge parallel.

To find the required shape of the interface, refer to Fig. 5.3*c*, wherein point *A* can lie anywhere on the boundary. One wavefront is transformed into another, provided the paths along which the energy propagates are all "equal," thereby maintaining the phase of the wavefront (p. 26). A little spherical surface of constant phase emitted from *S* must evolve into a flat surface of constant phase at $\overline{DD'}$. Whatever path the light takes from *S* to $\overline{DD'}$, it must always be the same number of wavelengths long, so that the disturbance begins and ends in-phase. Radiant energy leaving *S* as a single wavefront must arrive at the plane $\overline{DD'}$, having traveled for the same amount of time to get there, no matter what the actual route taken by any particular ray. In other words, $\overline{F_1A}/\lambda_t$ (the number of wavelengths along the arbitrary ray from F_1 to *A*) plus $\overline{AD}/\lambda_t$ (the number of wavelengths along the ray from *A* to *D*) must be constant regardless of where on the interface *A* happens to be. Now, adding these and multiplying by λ_0, yields

$$n_i(\overline{F_1A}) + n_t(\overline{AD}) = \text{constant} \tag{5.1}$$

Each term on the left is the length traveled in a medium multiplied by the index of that medium, and, of course, each represents the optical path length, *OPL*, traversed. The optical path lengths from *S* to $\overline{DD'}$ are all equal. If Eq. (5.1) is divided by *c*, the first term becomes the time it takes light to travel from *S* to *A* and the second term, the time from *A* to *D*; the right side remains constant (not the same constant, but constant). Equation (5.1) is equivalent to saying that all paths from *S* to $\overline{DD'}$ must take the same amount of time to traverse.

Let's return to finding the shape of the interface. Divide Eq. (5.1) by n_i, and it becomes

$$\overline{F_1A} + \left(\frac{n_t}{n_i}\right)\overline{AD} = \text{constant} \tag{5.2}$$

This is the equation of a hyperbola in which the eccentricity (e), which measures the bending of the curve, is given by $(n_t/n_i) > 1$; that is, $e = n_{ti} > 1$. The greater the eccentricity, the flatter the hyperbola (the larger the difference in the indices, the less the surface need be curved). When a point source is located at the focus F_1 and the interface between the two media is hyperbolic, plane waves are transmitted into the higher index material. It's left for Problem 5.3 to establish that when $(n_t/n_i) < 1$, the interface must be ellipsoidal. In each case pictured in Fig. 5.4, the rays either diverge from or converge toward a focal point, *F*. Furthermore, the rays can be reversed so that they travel either way; if a plane wave is inci-

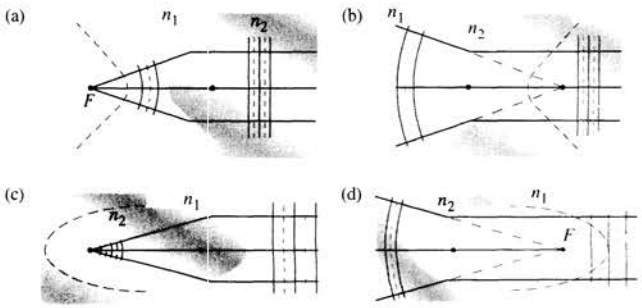


Figure 5.4 (*a*) and (*b*) Hyperboloidal and (*c*) and (*d*) ellipsoidal refracting surfaces ($n_2 > n_1$) in cross section.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

dent (from the right) on the interface in Fig. 5.4*c*, it will converge (off to the left) at the farthest focus of the ellipsoid.

The first person to suggest using conic sections as surfaces for lenses and mirrors was Johann Kepler (1611), but he wasn't able to go very far with the idea without Snell's Law. Once that relationship was discovered, Descartes (1637) using his Analytic Geometry could develop the theoretical foundations of the optics of aspherical surfaces. The analysis presented here is in essence a gift from Descartes.

It's an easy matter now to construct lenses such that both the object and image points (or the incident and emerging light) will be outside of the medium of the lens. In Fig. 5.5*a* diverging incident spherical waves are made into plane waves at the first interface via the mechanism of Fig. 5.4*a*. These plane waves within the lens strike the back face perpendicularly and emerge unaltered: $\theta_i = 0$ and $\theta_t = 0$. Because the rays are reversible, plane waves incoming from the right will converge to point F_1, which is known as the focal point of the lens. Exposed on its flat face to the parallel rays from the Sun, our rather sophisticated lens would serve nicely as a burning-glass.

In Fig. 5.5*b*, the plane waves within the lens are made to converge toward the axis by bending at the second interface. Both of these lenses are thicker at their midpoints than at their edges and are therefore said to be **convex** (from the Latin *convexus*, meaning arched). Each lens causes the incoming beam to converge somewhat, to bend a bit more toward the central axis; therefore, they are referred to as **converging lenses**.

In contrast, a **concave** lens (from the Latin *concavus*, meaning hollow—and most easily remembered because it contains the word cave) is thinner in the middle than at the edges, as is evident in Fig. 5.5*c*. It causes the rays that enter as a parallel bundle to diverge. All such devices that turn rays outward away from the central axis (and in so doing add divergence to the beam) are called **diverging lenses**. In Fig. 5.5*c*, parallel rays enter from the left and, on emerging, seem to diverge from F_2; still, that point is taken as a focal point. **When a parallel bundle of rays passes through a converging lens, the point to which it converges (or when passing through a diverging lens, the point from which it diverges) is a focal point of the lens.**

If a point source is positioned on the central or optical axis at the point F_1 in front of the lens in Fig. 5.5*b*, rays will *converge* to the conjugate point F_2. A luminous image of the source would appear on a screen placed at F_2, an image that is therefore said to be **real**. On the other hand, in Fig. 5.5*c* the point source is at infinity, and the rays emerging from the system this time are *diverging*. They appear to come from a point F_2, but no actual luminous image would appear on a screen at that location. The image here is spoken of as **virtual**, as is the familiar image generated by a plane mirror.

Optical elements (lenses and mirrors) of the sort we have talked about, with one or both surfaces neither planar nor spherical, are referred to as *aspherics*. Although their operation is easy to understand and they perform certain tasks exceedingly well, they are still difficult to manufacture with great accuracy. Nonetheless, where the costs are justifiable or the required precision is not restrictive or the volume produced is large enough, aspherics are being used and will surely have an increasingly important role. The first quality glass aspheric

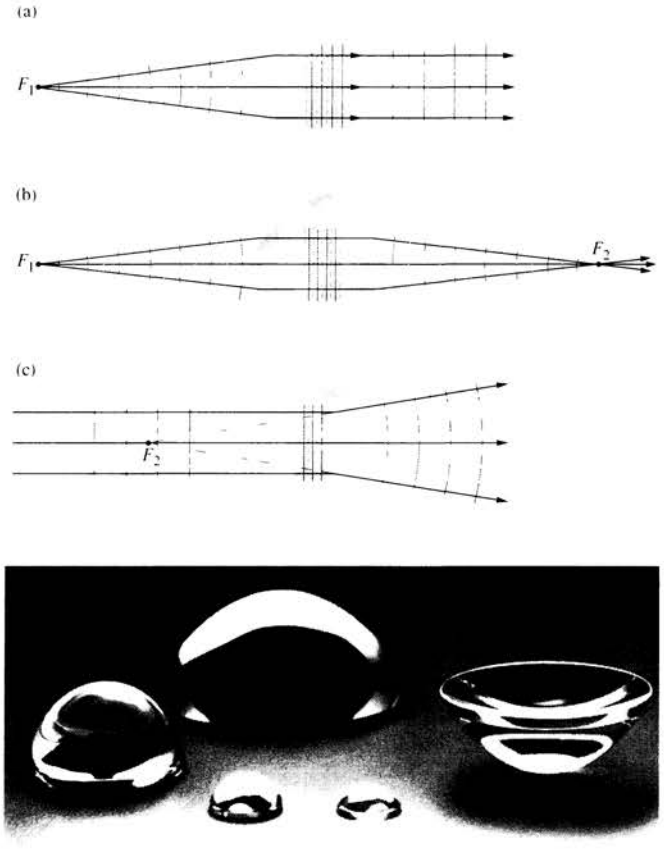


Figure 5.5 (*a*), (*b*), and (*c*) Several hyperbolic lenses seen in cross section. To explore an ellipsoidal lens, see Problem 5.4. (*d*) A selection of aspherical lenses. (Photo courtesy Melles Griot.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

to be manufactured in great quantities (tens of millions) was a lens for the Kodak disk camera (1982). Today aspherical lenses are frequently used as an elegant means of correcting imaging errors in complicated optical systems. Aspherical eyeglass lenses are flatter and lighter than regular spherical ones. As such, they're well suited for strong prescriptions. Furthermore, they minimize the magnification of the wearer's eyes as seen by other people.

A new generation of computer-controlled machines, aspheric generators, is producing elements with tolerances (i.e., departures from the desired surface) of better than 0.5 μm (0.000 020 inch). This is still about a factor of 10 away from the generally required tolerance of $\lambda/4$ for quality optics. After grinding, aspheres can be polished magnetorheologically. This technique, used to figure and finish the surface, magnetically controls the direction and pressure applied to the workpiece by the abrasive particles during polishing.

Nowadays aspherics made in plastic and glass can be found in all kinds of instruments across the whole range of quality, including telescopes, projectors, cameras, and reconnaissance devices.

5.2.2 Refraction at Spherical Surfaces

Consider two pieces of material, one with a concave and the other a convex spherical surface, both having the same radius. It is a unique property of the sphere that such pieces will fit together in intimate contact regardless of their mutual orientation. If we take two roughly spherical objects of suitable curvature, one a grinding tool and the other a disk of glass, separate them with some abrasive, and then randomly move them with respect to each other, we can anticipate that any high spots on either object will wear away. As they wear, both pieces will gradually become more spherical (see photo). Such surfaces are commonly generated in batches by automatic grinding and polishing machines.

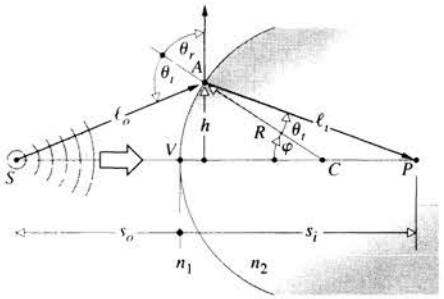


Figure 5.6 Refraction at a spherical interface. Conjugate foci.

Not surprisingly, the vast majority of quality lenses in use today have surfaces that are segments of spheres. Our intent here is to establish techniques for using such surfaces to simultaneously image a great many object points in light composed of a broad range of frequencies. Image errors, known as **aberrations**, will occur, but it is possible with the present technology to construct high-quality spherical lens systems whose aberrations are so well controlled that image fidelity is limited only by diffraction.

Figure 5.6 depicts a wave from the point source S impinging on a spherical interface of radius R centered at C. The point V is called the **vertex** of the surface. The length $s_o = \overline{SV}$ is known as the **object distance**. The ray $\overline{SA}$ will be refracted at the interface toward the local normal ($n_2 > n_1$) and therefore toward the central or **optical axis**. Assume that at some



Polishing a spherical lens. (Photo courtesy Optical Society of America.)

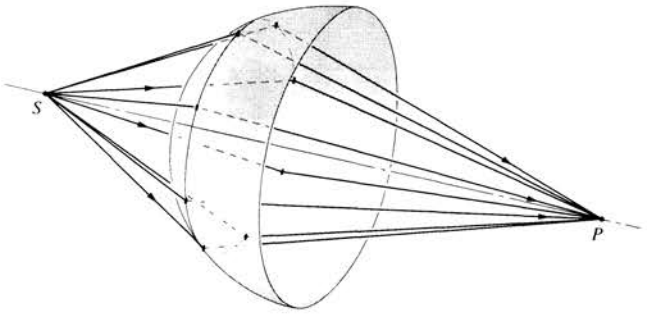


Figure 5.7 Rays incident at the same angle.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

point P the ray will cross the axis, as will all other rays incident at the same angle θ_i (Fig. 5.7). The length $s_i = \overline{VP}$ is the **image distance**. Fermat's Principle maintains that the optical path length OPL will be stationary; that is, its derivative with respect to the position variable will be zero. For the ray in question,

$$OPL = n_1\ell_o + n_2\ell_i \tag{5.3}$$

Using the law of cosines in triangles SAC and ACP along with the fact that $\cos\varphi = -\cos(180° - \varphi)$, we get

$$\ell_o = [R^2 + (s_o + R)^2 - 2R(s_o + R)\cos\varphi]^{1/2}$$

and $$\ell_i = [R^2 + (s_i - R)^2 + 2R(s_i - R)\cos\varphi]^{1/2}$$

The OPL can be rewritten as

$$OPL = n_1[R^2 + (s_o + R)^2 - 2R(s_o + R)\cos\varphi]^{1/2}$$

$$+ n_2[R^2 + (s_i - R)^2 + 2R(s_i - R)\cos\varphi]^{1/2}$$

All the quantities in the diagram (s_i, s_o, R, etc.) are positive numbers, and these form the basis of a *sign convention* which is gradually unfolding and to which we shall return time and again (see Table 5.1). Inasmuch as the point A moves at the end of a fixed radius (i.e., R = constant), φ is the position variable, and thus setting $d(OPL)/d\varphi = 0$, via Fermat's Principle we have

$$\frac{n_1R(s_o + R)\sin\varphi}{2\ell_o} - \frac{n_2R(s_i - R)\sin\varphi}{2\ell_i} = 0 \tag{5.4}$$

from which it follows that

$$\frac{n_1}{\ell_o} + \frac{n_2}{\ell_i} = \frac{1}{R}\left(\frac{n_2s_i}{\ell_i} - \frac{n_1s_o}{\ell_o}\right) \tag{5.5}$$

This is the relationship that must hold among the parameters for a ray going from S to P by way of refraction at the spherical interface. Although this expression is exact, it is rather complicated. If A is moved to a new location by changing φ, the new ray will not intercept the optical axis at P. (See Problem 5.1 concerning the Cartesian oval which is the interface configuration that would bring any ray, regardless of φ, to P.) The approximations that are used to represent ℓ_o and ℓ_i, and thereby simplify Eq. (5.5), are crucial in all that is to follow. Recall that

$$\cos\varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \frac{\varphi^6}{6!} + \cdots \tag{5.6}$$

and $$\sin\varphi = \varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \frac{\varphi^7}{7!} + \cdots \tag{5.7}$$

If we assume small values of φ (i.e., A close to V), $\cos\varphi \approx 1$. Consequently, the expressions for ℓ_o and ℓ_i yield $\ell_o \approx s_o$, $\ell_i \approx s_i$, and to that approximation

$$\frac{n_1}{s_o} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R} \tag{5.8}$$

We could have begun this derivation with Snell's Law rather than Fermat's Principle (Problem 5.5), in which case small values of φ would have led to $\sin\varphi \approx \varphi$ and Eq. (5.8) once again. This approximation delineates the domain of what is called *first-order theory*; we'll examine *third-order theory* (sin $\varphi \approx \varphi - \varphi^3/3!$) in the next chapter. Rays that arrive at shallow angles with respect to the optical axis (such that φ and h are appropriately small) are known as **paraxial rays**. The *emerging wavefront segment corresponding to these paraxial rays is essentially spherical and will form a "perfect" image at its center P located at* s_i. Notice that Eq. (5.8) is independent of the location of A over a small area about the symmetry axis, namely, the *paraxial region*. Gauss, in 1841, was the first to give a systematic exposition of the formation of images

TABLE 5.1 Sign Convention for Spherical Refracting Surfaces and Thin Lenses* (Light Entering from the Left)

s_o, f_o	+ left of V
x_o	+ left of F_o
s_i, f_i	+ right of V
x_i	+ right of F_i
R	+ if C is right of V
y_o, y_i	+ above optical axis

*This table anticipates the imminent introduction of a few quantities not yet spoken of.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

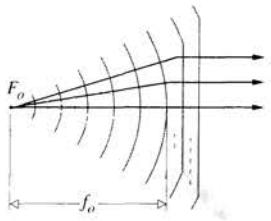


Figure 5.8 Plane waves propagating beyond a spherical interface—the object focus.

under the above approximation, and the result is variously known as *first-order*, *paraxial*, or **Gaussian Optics**. It soon became the basic theoretical tool by which lenses would be designed for several decades to come. If the optical system is well corrected, an incident spherical wave will emerge in a form very closely resembling a spherical wave. Consequently, as the perfection of the system increases, it more closely approaches first-order theory. Deviations from that of paraxial analysis will provide a convenient measure of the quality of an actual optical device.

If the point F_o in Fig. 5.8 is imaged at infinity ($s_i = \infty$), we have

$$\frac{n_1}{s_o} + \frac{n_2}{\infty} = \frac{n_2 - n_1}{R}$$

That special object distance is defined as the **first focal length** or the **object focal length**, $s_o \equiv f_o$, so that

$$f_o = \frac{n_1}{n_2 - n_1} R \tag{5.9}$$

The point F_o is known as the **first** or **object focus**. Similarly,

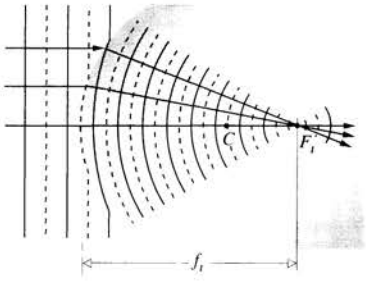


Figure 5.9 The reshaping of plane into spherical waves at a spherical interface—the image focus.

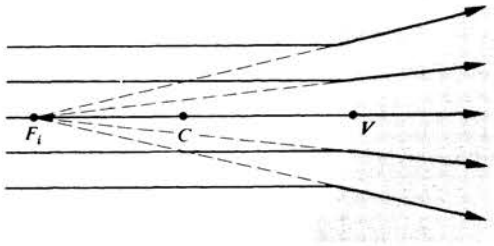


Figure 5.10 A virtual image point.

the **second** or **image focus** is the axial point F_i, where the image is formed when $s_o = \infty$; that is,

$$\frac{n_1}{\infty} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R}$$

Defining the **second** or **image focal length** f_i as equal to s_i in this special case (Fig. 5.9), we have

$$f_i = \frac{n_2}{n_2 - n_1} R \tag{5.10}$$

Recall that an image is virtual when the rays diverge from it (Fig. 5.10). Analogously, ***an object is virtual when the rays converge toward it*** (Fig. 5.11). Observe that the virtual object is now on the right-hand side of the vertex, and therefore s_o will be a negative quantity. Moreover, the surface is concave, and its radius will also be negative, as required by Eq. (5.9), since f_o would be negative. In the same way, the virtual image distance appearing to the left of V is negative.

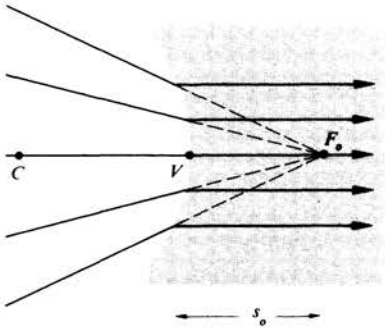


Figure 5.11 A virtual object point.

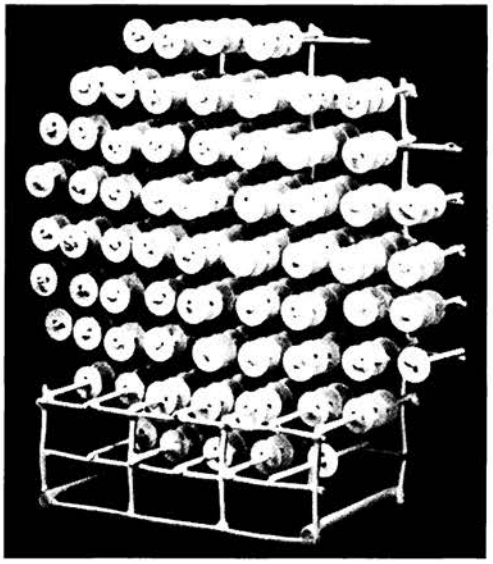
Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

5.2.3 Thin Lenses

Lenses are made in a wide range of forms; for example, there are acoustic and microwave lenses. Some of the latter are made of glass or wax in easily recognizable shapes, whereas others are far more subtle in appearance (see photo). Most often a lens has two or more refracting interfaces, and at least one of these is curved. Generally, the nonplanar surfaces are centered on a common axis. These surfaces are most frequently spherical segments and are often coated with thin dielectric films to control their transmission properties (see Section 9.9).

A lens that consists of one element (i.e., it has only two refracting surfaces) is a *simple lens*. The presence of more than one element makes it a *compound lens*. A lens is also classified as to whether it is *thin* or *thick*—that is, whether or not its thickness is effectively negligible. We will limit ourselves, for the most part, to *centered systems* (for which all surfaces are rotationally symmetric about a common axis) of spherical surfaces. Under these restrictions, the simple lens can take the forms shown in Fig. 5.12.

Lenses that are variously known as ***convex***, ***converging***, or ***positive*** are thicker at the center and so tend to decrease the radius of curvature of the wavefronts. In other words, the wave converges more as it traverses the lens, assuming, of course, that the index of the lens is greater than that of the media in which it is immersed. ***Concave***, ***diverging***, or ***negative*** lenses, on the other hand, are thinner at the center and tend to advance that portion of the wavefront, causing it to diverge more than it did upon entry.



A lens for short-wavelength radiowaves. The disks serve to refract these waves much as rows of atoms refract light. (Photo courtesy Optical Society of America.)

Thin-Lens Equations

Return to the discussion of refraction at a single spherical interface, where the location of the conjugate points S and P is given by

$$\frac{n_1}{s_o} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R} \qquad [5.8]$$

When s_o is large for a fixed $(n_2 - n_1)/R$, s_i is relatively small. As s_o decreases, s_i moves away from the vertex: that is, both θ_i and θ_t increase until finally $s_o = f_o$ and $s_i = \infty$. At that point, $n_1/s_o = (n_2 - n_1)/R$, so that if s_o gets any smaller, s_i will have to be negative, if Eq. (5.8) is to hold. In other words, the image becomes virtual (Fig. 5.13).

CONVEX	CONCAVE
$R_1 > 0$ $R_2 < 0$ Bi-convex	$R_1 < 0$ $R_2 > 0$ Bi-concave
$R_1 = \infty$ $R_2 < 0$ Planar convex	$R_1 = \infty$ $R_2 > 0$ Planar concave
$R_1 > 0$ $R_2 > 0$ Meniscus convex	$R_1 > 0$ $R_2 > 0$ Meniscus concave



Figure 5.12 Cross sections of various centered spherical simple lenses. The surface on the left is #1 since it is encountered first. Its radius is R_1. (Photo courtesy of Melles Griot.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

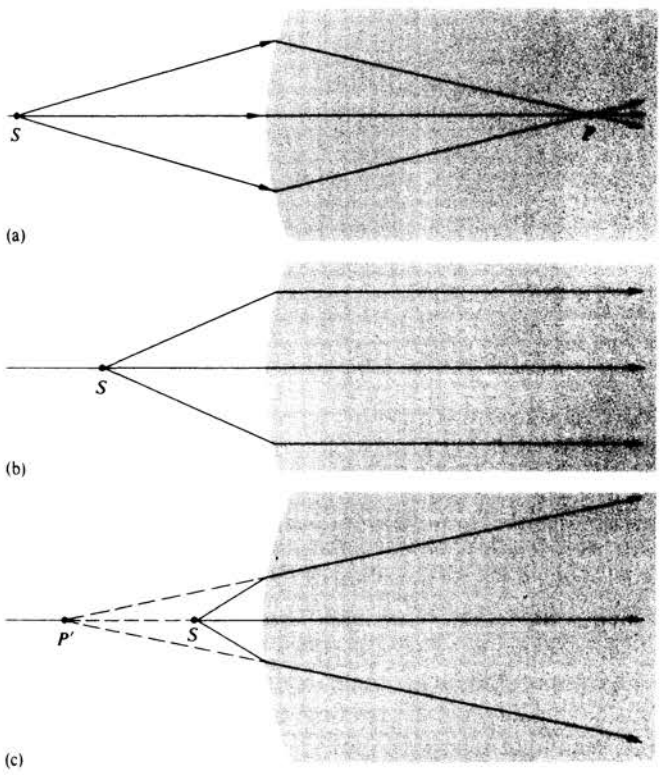


Figure 5.13 Refraction at a spherical interface between two transparent media shown in cross section.

Let's now locate the conjugate points for a lens of index n_l surrounded by a medium of index n_m, as in Fig. 5.14, where another end has simply been ground onto the piece in Fig. 5.13*c*. This certainly isn't the most general set of circumstances, but it is the most common, and even more cogently, it is the simplest.* We know from Eq. (5.8) that the paraxial rays issuing from S at s_{o1} will meet at P', a distance, which we now call s_{i1}, from V_1, given by

$$\frac{n_m}{s_{o1}} + \frac{n_l}{s_{i1}} = \frac{n_l - n_m}{R_1} \tag{5.11}$$

Thus, as far as the second surface is concerned, it "sees" rays coming toward it from P', which serves as its object point a distance s_{o2} away. Furthermore, the rays arriving at that second surface are in the medium of index n_l. Thus the object space for the second interface that contains P' has an index n_l.

*See Jenkins and White, *Fundamentals of Optics*, p. 57, for a derivation containing three different indices.

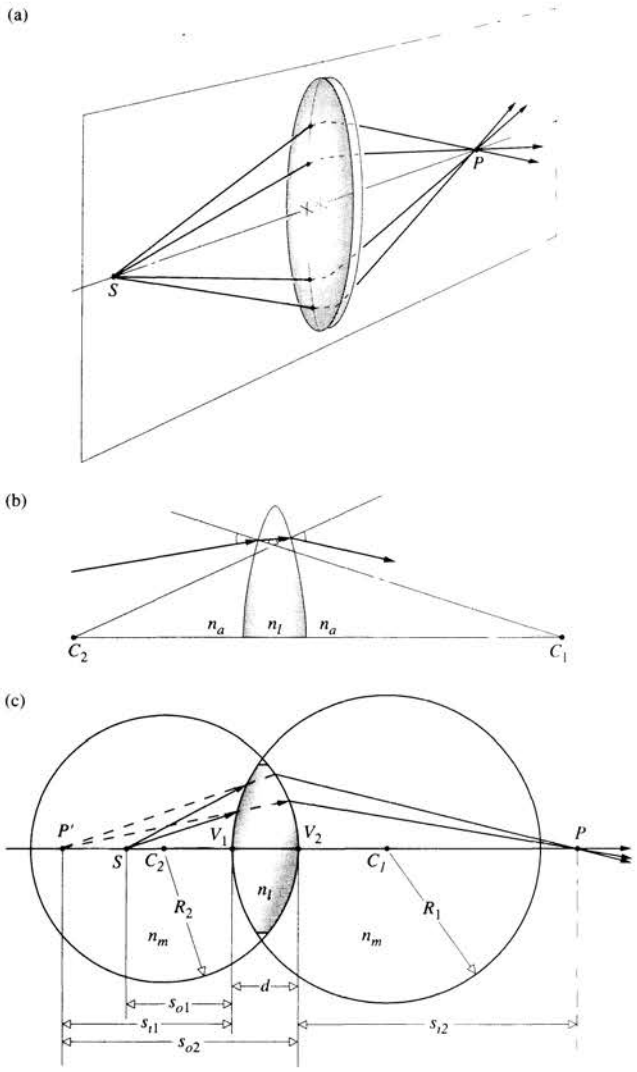


Figure 5.14 A spherical lens. (*a*) Rays in a vertical plane passing through a lens. Conjugate foci. (*b*) Refraction at the interfaces. The radius drawn from C_1 is normal to the first surface, and as the ray enters the lens it bends down *toward* that normal. The radius from C_2 is normal to the second surface; and as the ray emerges, since $n_l > n_a$, the ray bends down *away* from that normal. (c) The geometry.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Note that the rays from P' to that surface are indeed straight lines. Considering the fact that

$$|s_{o2}| = |s_{i1}| + d$$

since s_{o2} is on the left and therefore positive, $s_{o2} = |s_{o2}|$, and s_{i1} is also on the left and therefore negative, $-s_{i1} = |s_{i1}|$, we have

$$s_{o2} = -s_{i1} + d \tag{5.12}$$

Thus at the second surface Eq. (5.8) yields

$$\frac{n_l}{(-s_{i1} + d)} + \frac{n_m}{s_{i2}} = \frac{n_m - n_l}{R_2} \tag{5.13}$$

Here $n_l > n_m$ and $R_2 < 0$, so that the right-hand side is positive. Adding Eqs. (5.11) and (5.13), we have

$$\frac{n_m}{s_{o1}} + \frac{n_m}{s_{i2}} = (n_l - n_m)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + \frac{n_l d}{(s_{i1} - d)s_{i1}} \tag{5.14}$$

If the lens is thin enough ($d \to 0$), the last term on the right is effectively zero. As a further simplification, assume the surrounding medium to be air (i.e., $n_m \approx 1$). Accordingly, we have the very useful **Thin-Lens Equation**, often referred to as the **Lensmaker's Formula**:

$$\frac{1}{s_o} + \frac{1}{s_i} = (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{5.15}$$

where we let $s_{o1} = s_o$ and $s_{i2} = s_i$. The points V_1 and V_2 tend to coalesce as $d \to 0$, so that s_o and s_i can be measured from either the vertices or the lens center.

Just as in the case of the single spherical surface, if s_o is moved out to infinity, the image distance becomes the focal length f_i, or symbolically,

$$\lim_{s_o \to \infty} s_i = f_i$$

Similarly

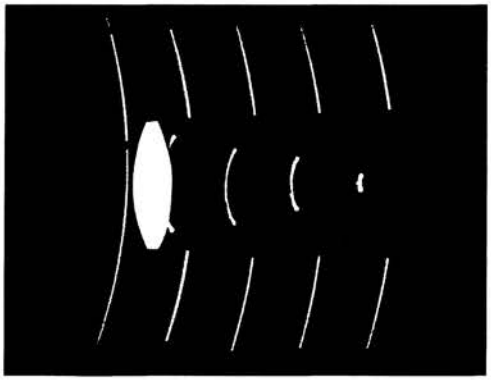
$$\lim_{s_i \to \infty} s_o = f_o$$

It is evident from Eq. (5.15) that for a thin lens $f_i = f_o$, and consequently we drop the subscripts altogether. Thus

$$\frac{1}{f} = (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{5.16}$$

and

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f} \tag{5.17}$$



The actual wavefronts of a diverging lightwave partially focused by a lens. The photo shows five exposures, each separated by about 100 ps (i.e., 100×10^{-12} s), of a spherical pulse 10 ps long as it swept by and through a converging lens. The picture was made by using a holographic technique. (Courtesy of N.H. Abramson.)

which is the famous **Gaussian Lens Formula** (see photo).

As an example of how these expressions might be used, let's compute the focal length in air of a thin planar-convex lens having a radius of curvature of 50 mm and an index of 1.5. With light entering on the planar surface ($R_1 = \infty$, $R_2 = -50$),

$$\frac{1}{f} = (1.5 - 1)\left(\frac{1}{\infty} - \frac{1}{-50}\right)$$

whereas if instead it arrives at the curved surface ($R_1 = +50$, $R_2 = \infty$),

$$\frac{1}{f} = (1.5 - 1)\left(\frac{1}{+50} - \frac{1}{\infty}\right)$$

and in either case $f = 100$ mm. If an object is alternately placed at distances 600 mm, 200 mm, 150 mm, 100 mm, and 50 mm from the lens on either side, we can find the image points from Eq. (5.17):

$$s_i = \frac{s_o f}{s_o - f} = \frac{(600)(100)}{600 - 100}$$

and $s_i = 120$ mm. Similarly, the other image distances are 200 mm, 300 mm, ∞, and -100 mm, respectively.

Interestingly enough, when $s_o = \infty$, $s_i = f$; as s_o decreases, s_i increases positively until $s_o = f$ and s_i is negative thereafter. You can qualitatively check this out with a simple convex lens and a small electric light—the high-intensity variety is probably the most convenient. Standing as far as you can from the source, project a clear image of it onto a white sheet of paper. You should be able to see the lamp quite clearly and not just as

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

a blur. That image distance approximates f. Now move the lens in toward S, adjusting s_i to produce a clear image. It will surely increase. As $s_o \to f$, a clear image of the lamp can be projected, but only on an increasingly distant screen. For $s_o < f$, there will just be a blur where the farthest wall intersects the diverging cone of rays—the image is virtual.

Focal Points and Planes

Figure 5.15 summarizes pictorially some of the situations described analytically by Eq. 5.16. Observe that if a lens of index n_l is immersed in a medium of index n_m,

$$\frac{1}{f} = (n_{lm} - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{5.18}$$

The focal lengths in (*a*) and (*b*) of Fig. 5.15 are equal, because the same medium exists on either side of the lens. Since $n_l > n_m$, it follows that $n_{lm} > 1$. In both cases $R_1 > 0$ and $R_2 < 0$, so that each focal length is positive. We have a real object in (*a*) and a real image in (*b*). In (*c*), $n_l < n_m$, and consequently f is negative. In (*d*) and (*e*), $n_{lm} > 1$ but $R_1 < 0$, whereas $R_2 > 0$, so f is again negative, and the object in one case and the image in the other are virtual. In (*f*) $n_{lm} < 1$, yielding an $f > 0$.

Notice that in each instance it is particularly convenient to draw a ray through the center of the lens, which, because it is perpendicular to both surfaces, is undeviated. Suppose, instead, that an off-axis paraxial ray emerges from the lens parallel to its incident direction, as in Fig. 5.16. We maintain that all such rays will pass through the point defined as the **optical center** O of the lens. To see this, draw two parallel planes, one on each side tangent to the lens at any pair of points A and B. This can easily be done by selecting A and B such that the radii $\overline{AC}_1$ and $\overline{BC}_2$ are themselves parallel. It is to be shown that the paraxial ray traversing $\overline{AB}$ enters and leaves the lens in the same direction. It's evident from the diagram that triangles AOC_1 and BOC_2 are similar, in the geometric sense, and therefore their sides are proportional. Hence, $|R_1|(\overline{OC}_2) = |R_2|(\overline{OC}_1)$, and since the radii are constant, the location of O is constant, independent of A and B. As we saw earlier (Problem 4.34 and Fig. P.4.34), a ray traversing a medi-

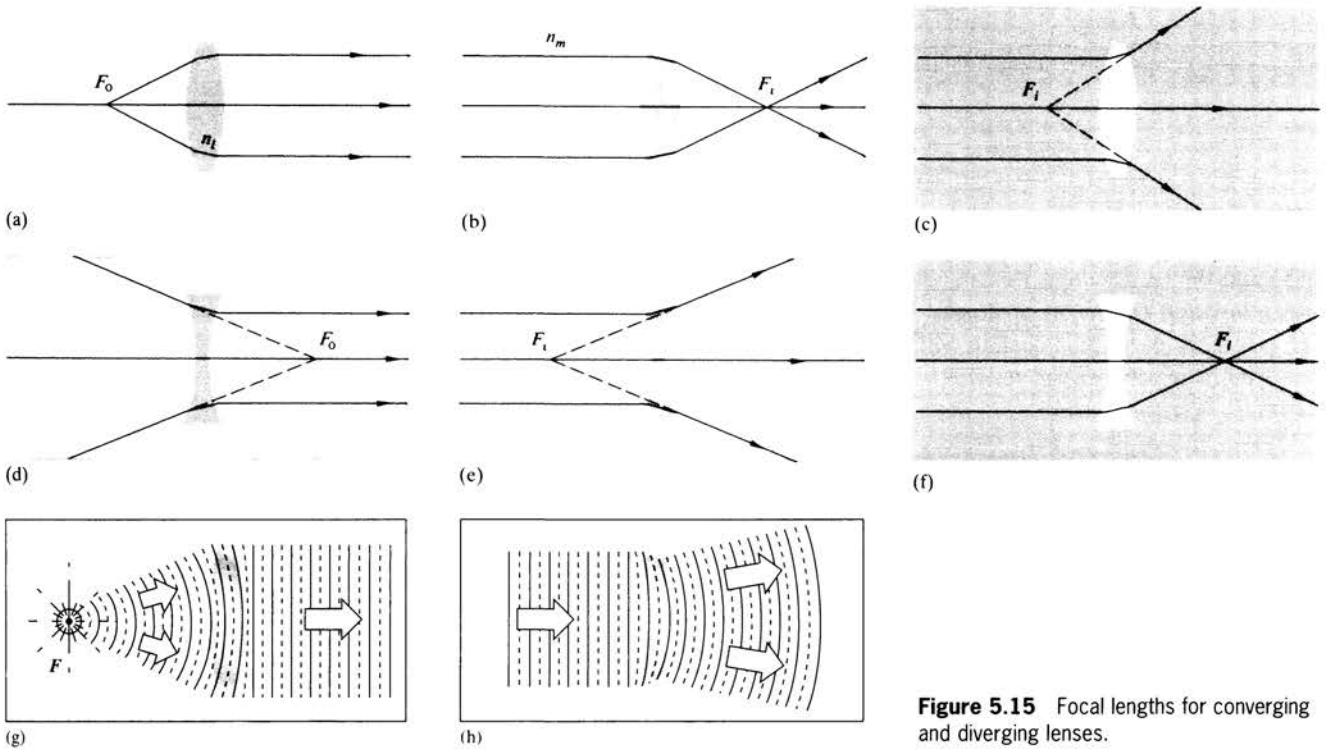


Figure 5.15 Focal lengths for converging and diverging lenses.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

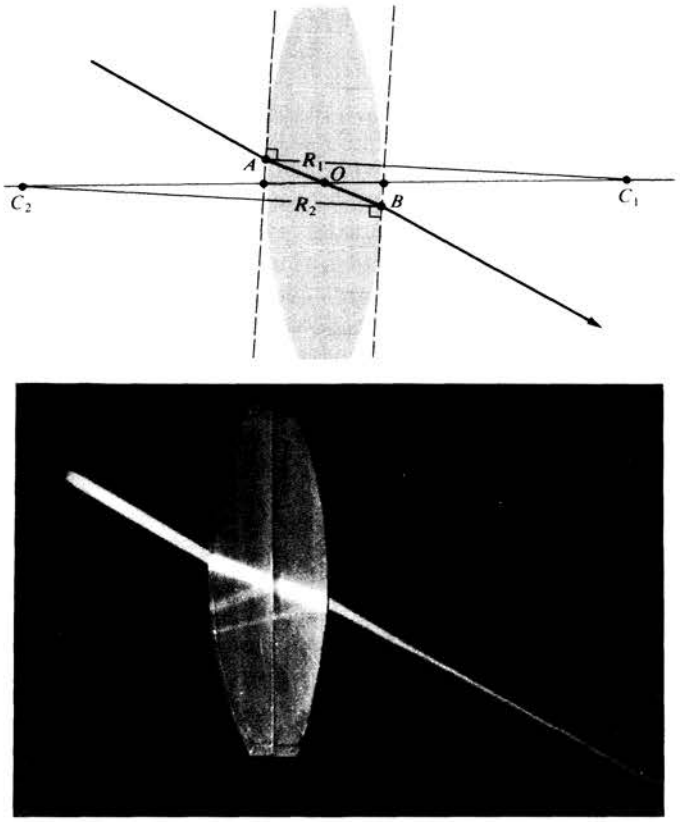


Figure 5.16 The optical center of a lens. (Photo by E.H.)

um bounded by parallel planes will be displaced laterally but will suffer no angular deviation. This displacement is proportional to the thickness, which for a thin lens is negligible. ***Rays passing through O may, accordingly, be drawn as straight lines***. It is customary when dealing with thin lenses simply to place O midway between the vertices.

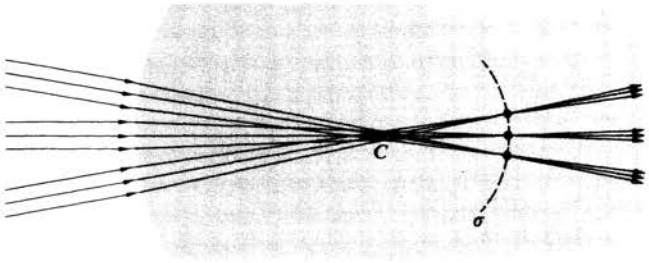


Figure 5.17 Focusing of several ray bundles.

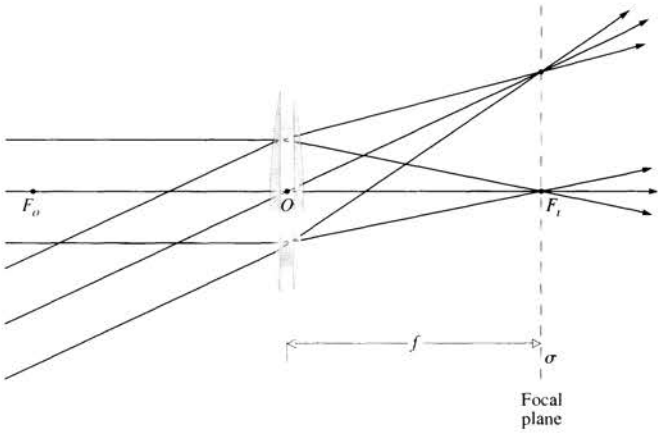


Figure 5.18 The focal plane of a lens.

Recall that a bundle of parallel paraxial rays incident on a spherical refracting surface comes to a focus at a point on the optical axis (Fig. 5.10). As shown in Fig. 5.17, this implies that several such bundles entering in a narrow cone will be focused on a spherical segment σ, also centered on C. The undeviated rays normal to the surface, and therefore passing through C, locate the foci on σ. Since the ray cone must indeed be narrow, σ can satisfactorily be represented as a plane normal to the symmetry axis and passing through the image focus. It is known as a **focal plane**. In the same way, limiting ourselves to paraxial theory, a lens will focus all incident parallel bundles of rays* onto a surface called the **second** or **back focal plane**, as in Fig. 5.18. Here each point on σ is located by the undeviated ray through O. Similarly, the **first** or **front focal plane** contains the object focus F_o.

Finite Imagery

Thus far we've treated the mathematical abstraction of a single-point source. Now let's deal with the fact that a great many such points combine to form a continuous finite object (Fig. 5.2). For the moment, imagine the object to be a segment of a sphere, σ_o, centered on C, as in Fig. 5.19. If σ_o is close to the

*Perhaps the earliest literary reference to the focal properties of a lens appears in Aristophanes' play, *The Clouds*, which dates back to 423 B.C.E. In it Strepsiades plots to use a burning-glass to focus the Sun's rays onto a wax tablet and thereby melt out the record of a gambling debt.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

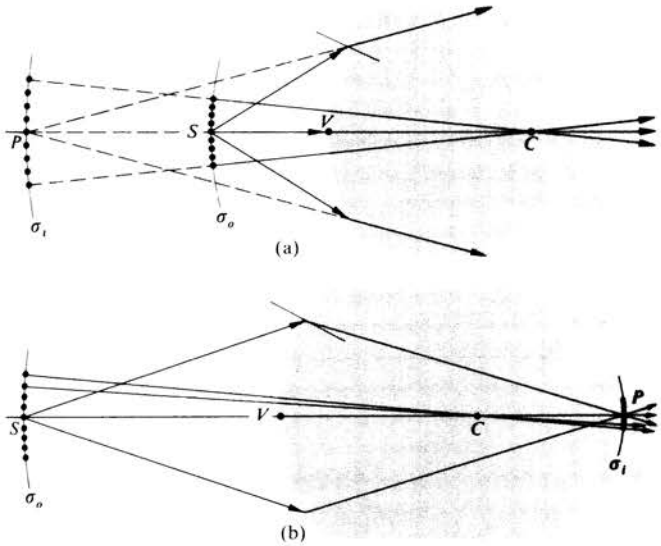


Figure 5.19 Finite imagery.

spherical interface, point S will have a virtual image P ($s_i < 0$ and therefore on the left of V). With S farther away, its image will be real ($s_i > 0$ and therefore on the right-hand side). In either case, each point on σ_o has a conjugate point on σ_i lying on a straight line through C. Within the restrictions of paraxial theory, these surfaces can be considered planar. Thus a small planar object normal to the optical axis will be imaged into a small planar region also normal to that axis. Note that if σ_o is moved out to infinity, the cone of rays from each source point will become **collimated** (i.e., parallel), and the image points will lie on the focal plane (Fig. 5.18).

By cutting and polishing the right side of the piece depicted in Fig. 5.19, we can construct a thin lens. Once again, the image (σ_i in Fig. 5.19) formed by the first surface of the lens will serve as the object for the second surface, which in turn will generate a final image. Suppose then that σ_i in Fig. 5.19*a* is the object for the second surface, which is assumed to have a negative radius. We already know what will happen—the situation is identical to Fig. 5.19*b* with the ray directions reversed. The *final image formed by a lens of a small planar object normal to the optical axis will itself be a small plane normal to that axis.*

The location, size, and orientation of an image produced by a lens can be determined, particularly simply, with ray diagrams. To find the image of the object in Fig. 5.20, we must locate the image point corresponding to each object point. Since all rays issuing from a source point in a paraxial cone will arrive at the image point, any two such rays will suffice to fix that point. Because we know the positions of the focal points, there are three rays that are especially easy to apply. The first (ray-1) is the undeviated ray through the center of the lens O. The other two (ray-2 and ray-3) make use of the fact that a ray passing through the focal point will emerge from the lens parallel to the central axis and vice versa. As a rule-of-thumb when sketching ray diagrams, draw the lens diameter (the vertical extent) roughly the size of the focal length. Then put in points on the central optical axis at one and two focal lengths, both in front of and behind the lens. You can usually locate the image by just tracing ray-1 and ray-2 from either the upper or lowermost points on the object.

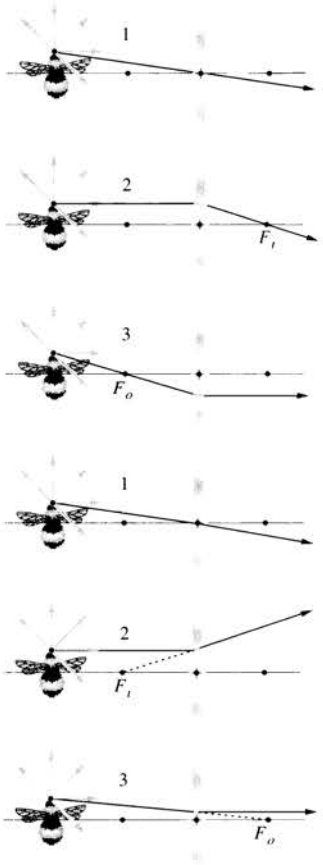


Figure 5.20 Tracing a few key rays through a positive and negative lens.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

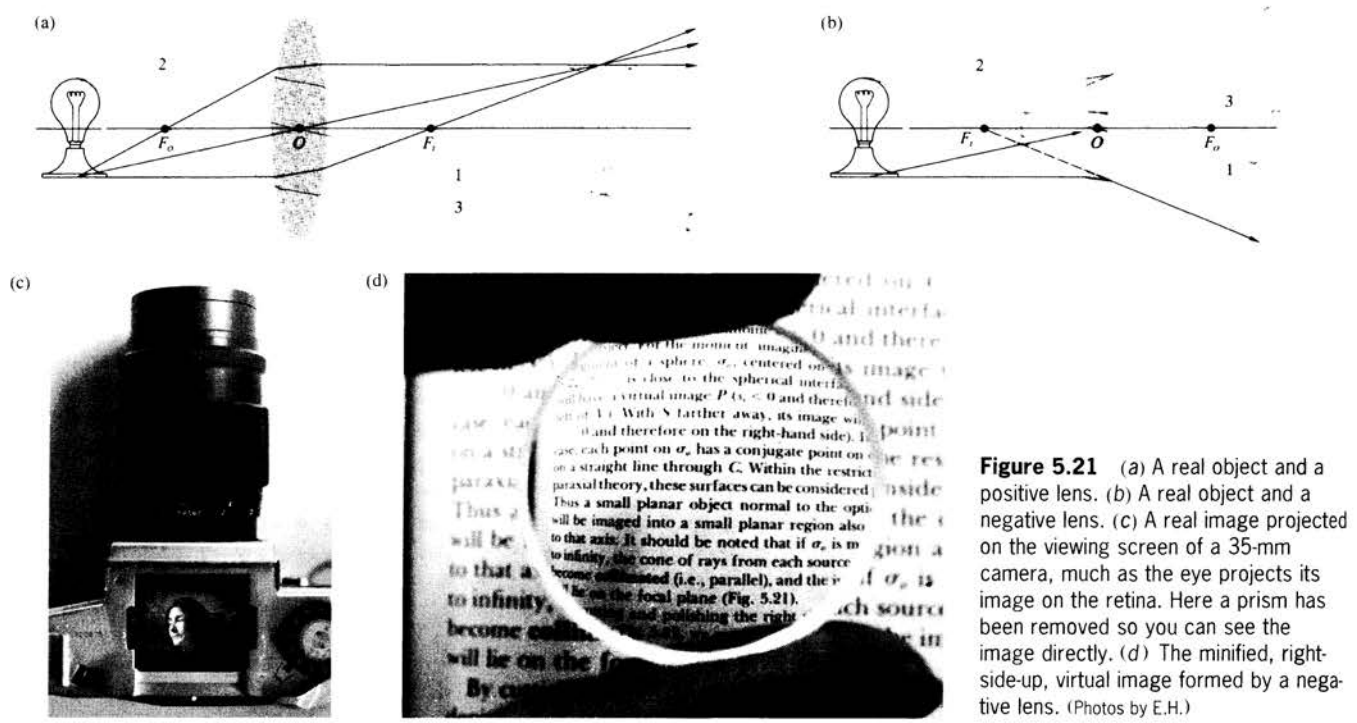


Figure 5.21 (*a*) A real object and a positive lens. (*b*) A real object and a negative lens. (*c*) A real image projected on the viewing screen of a 35-mm camera, much as the eye projects its image on the retina. Here a prism has been removed so you can see the image directly. (*d*) The minified, right-side-up, virtual image formed by a negative lens. (Photos by E.H.)

Figure 5.21 shows how any *two* of these three rays locate the image of a point on the object. Incidentally, this technique dates back to the work of Robert Smith as long ago as 1738. This graphical procedure can be made even simpler by replacing the thin lens with a plane passing through its center (Fig. 5.22). Presumably, if we were to extend every incoming ray forward a little and every outgoing ray backward a bit, each pair would meet on this plane. The total deviation of any ray can be envisaged as occurring all at once on that plane. This is equivalent to the actual process consisting of two separate angular shifts, one at each interface. (As we'll see later, this is tantamount to saying that the two principal planes of a thin lens coincide.)

In accord with convention, transverse distances above the optical axis are taken as positive quantities, and those below the axis are given negative numerical values. Therefore in Fig. 5.22 $y_o > 0$ and $y_i < 0$. Here the image is said to be **inverted**, whereas if $y_i > 0$ when $y_o > 0$, it is **right-side-up** or **erect**. Observe that triangles AOF_i and $P_2P_1F_i$ are similar. Ergo

$$\frac{y_o}{|y_i|} = \frac{f}{(s_i - f)} \tag{5.19}$$

Similarly, triangles S_2S_1O and P_2P_1O are similar, and

$$\frac{y_o}{|y_i|} = \frac{s_o}{s_i} \tag{5.20}$$

where all quantities other than y_i are positive. Hence

$$\frac{s_o}{s_i} = \frac{f}{(s_i - f)} \tag{5.21}$$

and

$$\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i}$$

which is, of course, the Gaussian Lens Equation [Eq. (5.17)]. Furthermore, triangles $S_2S_1F_o$ and BOF_o are similar and

$$\frac{f}{(s_o - f)} = \frac{|y_i|}{y_o} \tag{5.22}$$

Using the distances measured from the focal points and combining this information with Eq. (5.19) leads to

$$x_o x_i = f^2 \tag{5.23}$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

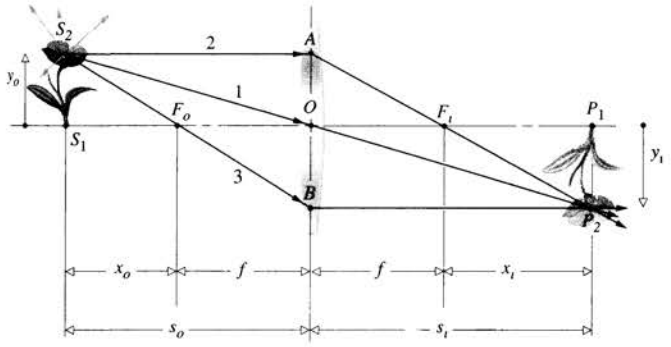


Figure 5.22 Object and image location for a thin lens.

TABLE 5.2 Meanings Associated with the Signs of Various Thin Lens and Spherical Interface Parameters

Quantity	Sign	
	+	−
s_o	Real object	Virtual object
s_i	Real image	Virtual image
f	Converging lens	Diverging lens
y_o	Erect object	Inverted object
y_i	Erect image	Inverted image
M_T	Erect image	Inverted image

This is the **Newtonian form** of the lens equation, the first statement of which appeared in Newton's *Opticks* in 1704. The signs of x_o and x_i are reckoned with respect to their concomitant foci. By convention, x_o is taken to be positive left of F_o, whereas x_i is positive on the right of F_i. It is evident from Eq. (5.23) that x_o and x_i have like signs, which means that ***the object and image must be on opposite sides of their respective focal points***. This is a good thing for the neophyte to remember when making those hasty freehand ray diagrams for which he or she is already infamous.

The ratio of the transverse dimensions of the final image formed by any optical system to the corresponding dimension of the object is defined as the *lateral* or **transverse magnification**, M_T, that is,

$$M_T \equiv \frac{y_i}{y_o} \tag{5.24}$$

Or from Eq. (5.20)

$$M_T = -\frac{s_i}{s_o} \tag{5.25}$$

Thus ***a positive M_T connotes an erect image, while a negative value means the image is inverted*** (see Table 5.2). Bear in mind that s_i and s_o are both positive for real objects and images. Clearly, then, ***all real images formed by a single thin lens will be inverted***. The Newtonian expression for the magnification follows from Eqs. (5.19) and (5.22) and Fig. 5.22:

$$M_T = -\frac{x_i}{f} = -\frac{f}{x_o} \tag{5.26}$$

The term *magnification* is a bit of a misnomer, since the magnitude of M_T can certainly be less than 1, in which case the image is smaller than the object. We have $M_T = -1$ when the object and image distances are positive and equal, and that happens [Eq. (5.17)] only when $s_o = s_i = 2f$. This turns out to be the configuration in which the object and image are as close together as they can possibly get (i.e., a distance $4f$ apart; see Problem 5.15). Table 5.3 summarizes a number of image configurations resulting from the juxtaposition of a thin lens and a real object.

TABLE 5.3 Images of Real Objects Formed by Thin Lenses

Convex

Object	Image			
Location	Type	Location	Orientation	Relative Size
$\infty > s_o > 2f$	Real	$f < s_i < 2f$	Inverted	Minified
$s_o = 2f$	Real	$s_i = 2f$	Inverted	Same size
$f < s_o < 2f$	Real	$\infty > s_i > 2f$	Inverted	Magnified
$s_o = f$		$\pm\infty$		
$s_o < f$	Virtual	$\lvert s_i\rvert > s_o$	Erect	Magnified

Concave

Object	Image			
Location	Type	Location	Orientation	Relative Size
Anywhere	Virtual	$\lvert s_i\rvert < \lvert f\rvert$, $s_o > \lvert s_i\rvert$	Erect	Minified

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

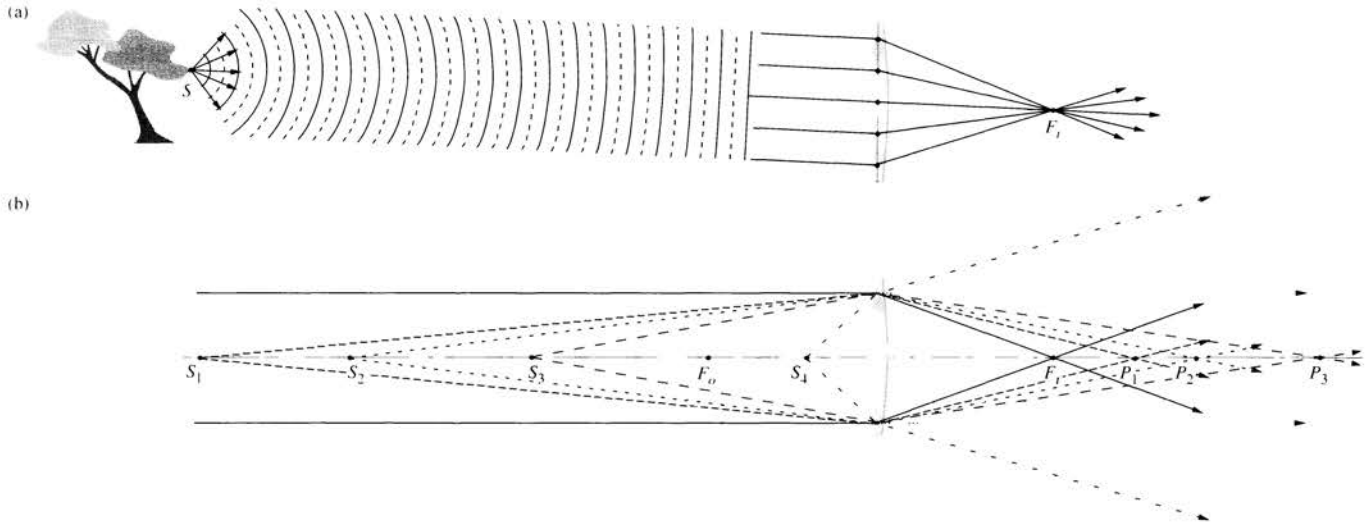


Figure 5.23 (*a*) The waves from a distant object flatten out as they expand, and the radii get larger and larger. Viewed from far away the rays from any point are essentially parallel, and the lens causes them to converge at F_i. (*b*) As a point source moves closer, the rays diverge more and the image point moves out away from the lens. The emerging rays no longer converge once the object reaches the focal point; nearer in still, they diverge.

We are now in a position to understand the entire range of behavior of a single convex or concave lens. To that end, suppose that a distant point source sends out a cone of light that is intercepted by a positive lens (Fig. 5.23). If the source is at infinity (i.e., so far away that it might just as well be infinity), rays coming from it entering the lens are essentially parallel (Fig. 5.23*a*) and will be brought together at the focal point F_i. If the source point S_1 is closer (Fig. 5.23*b*), but still fairly far away, the cone of rays entering the lens is narrow, and the rays come in at shallow angles to the surface of the lens. Because the rays do not diverge greatly, the lens bends each one into convergence, and they arrive at point P_1. As the source moves closer, the entering rays diverge more, and the resulting image point moves farther to the right. Finally, when the source point is at F_o, the rays are diverging so strongly that the lens can no longer bring them into convergence, and they emerge parallel to the central axis. Moving the source point closer results in rays that diverge so much on entering the lens that they still diverge on leaving. The image point is now virtual—*there are no real images of objects at or closer in than f.*

Figure 5.24 illustrates the behavior pictorially. **As the object approaches the lens, the real image moves away from it**. When the object is very far away, the image (real, inverted, and minified $M_T < 1$) is just to the right of the focal plane. As the object approaches the lens, the image (still real, inverted, and minified $M_T < 1$) moves away from the focal plane, to the right, getting larger and larger. With the object between infinity and $2f$ we have the arrangement for cameras and eyeballs, both of which require a minified, real image. By the way, it's the brain that flips the image so that you see things right-side-up.

When the object is at two focal lengths, the image (real and inverted) is now life size, that is, $M_T = 1$. This is the usual configuration of the photocopy machine.

As the object comes closer to the lens (between $2f$ and f), the image (real, inverted, and enlarged $M_T > 1$) rapidly moves to the right and continues to increase in size. This configuration corresponds to the film projector where the crucial feature is that the image is real and enlarged. To compensate for the image being inverted, the film is simply put in upside-down.

When the object arrives at a distance from the lens of precisely one focal length, the image has, in effect, moved off to infinity. (There is no image, the emerging rays are parallel.)

With the object closer in than one focal length, the image

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

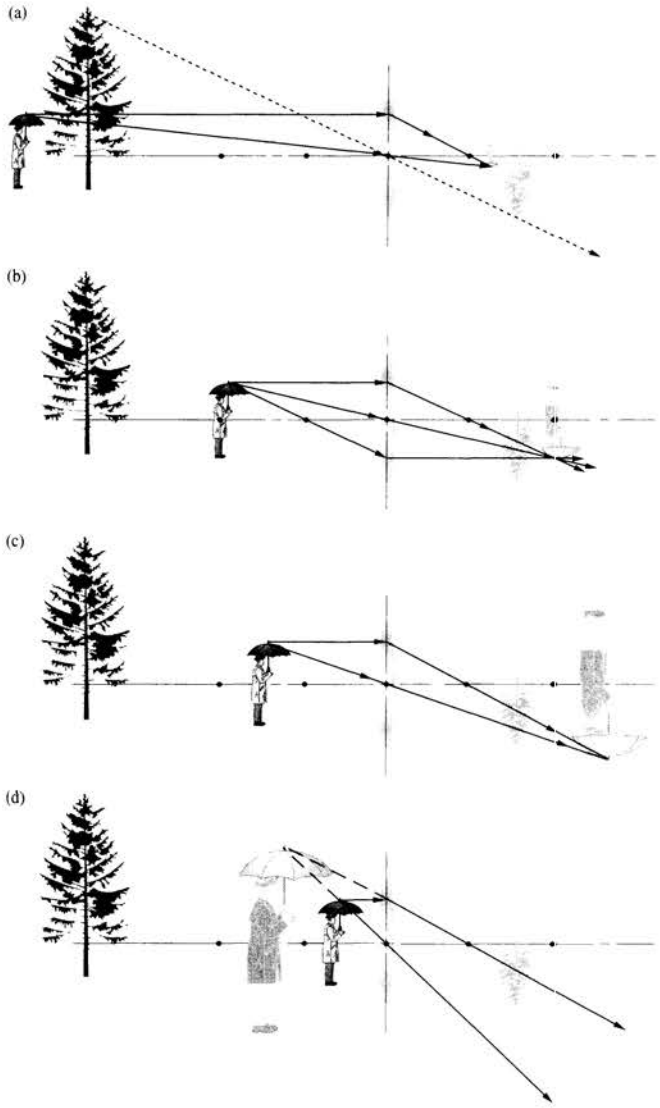


Figure 5.24 The image-forming behavior of a thin positive lens.

(virtual, right-side-up, and enlarged $M_T > 1$) reappears. This is the configuration of the magnifying glass.

It's useful to remember that *the ray entering the lens parallel to the central axis fixes the height of the real image* (Fig. 5.25). Because that ray diverges from the central axis, the size of the image increases rapidly as the object approaches F. Note, too, that the transformation from object to image space is not linear; all of the object space from $2f$ out to infinity, on the left of the lens, is compressed in the image space between f and $2f$, on the right of the lens. Figure 5.25 suggests that the image space is distorted, in the sense that advancing the object uniformly toward the lens has the effect of changing the image differently along and transverse to the central axis. The axial image intervals increase much more rapidly than the corresponding successive changes in the height of the image. This relative "flattening" of distant-object space is easily observable using a telescope (i.e., a long focal-length lens). You've probably seen the effect in a motion picture shot through a telephoto lens. Always staying far away, the hero vigorously runs a great distance toward the camera, but psychologically he seems to make no progress because his perceived size increases very little despite all his effort.

Presumably, the image of a three-dimensional object will itself occupy a three-dimensional region of space. The optical system can apparently affect both the transverse and longitudinal dimensions of the image. The **longitudinal magnification**, M_L, which relates to the axial direction, is defined as

$$M_L \equiv \frac{dx_i}{dx_o} \tag{5.27}$$

This is the ratio of an infinitesimal axial length in the region of the image to the corresponding length in the region of the object. Differentiating Eq. (5.23) leads to

$$M_L = -\frac{f^2}{x_o^2} = -M_T^2 \tag{5.28}$$

for a thin lens in a single medium (Fig. 5.26). Evidently, $M_L < 0$, which implies that a positive dx_o corresponds to a negative dx_i and vice versa. In other words, a finger pointing toward the lens is imaged pointing away from it (Fig. 5.27).

Form the image of a window on a sheet of paper, using a simple convex lens. Assuming a lovely arboreal scene, image the distant trees on the screen. Now move the paper *away* from the lens, so that it intersects a different region of the image space. The trees will fade while the nearby window itself comes into view.

Thin-Lens Combinations

Our purpose here is not to become proficient in the intricacies of modern lens design, but rather to gain the familiarity necessary to utilize, and adapt, those lens systems already available commercially.

In constructing a new optical system, one generally begins by sketching out a rough arrangement using the quickest

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

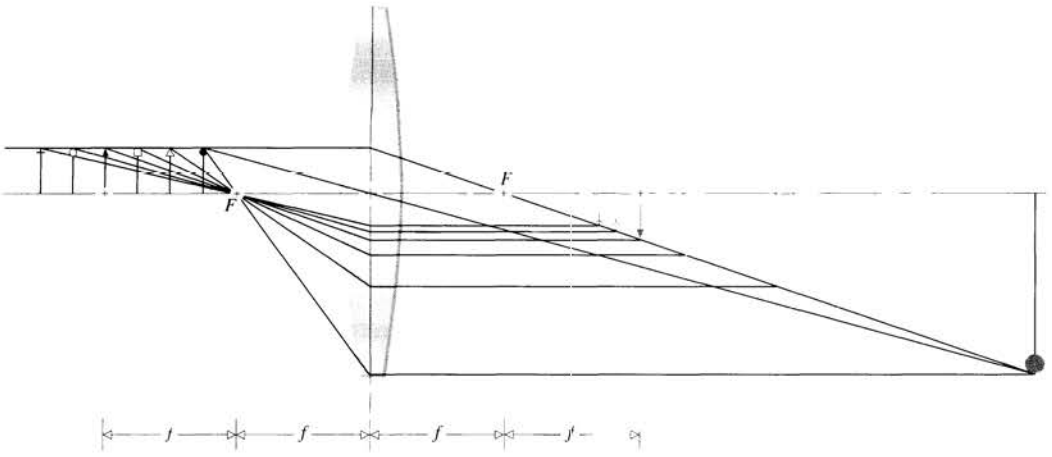


Figure 5.25 The number-2 ray entering the lens parallel to the central axis limits the image height.

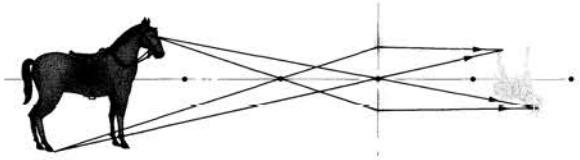


Figure 5.26 The transverse magnification is different from the longitudinal magnification.

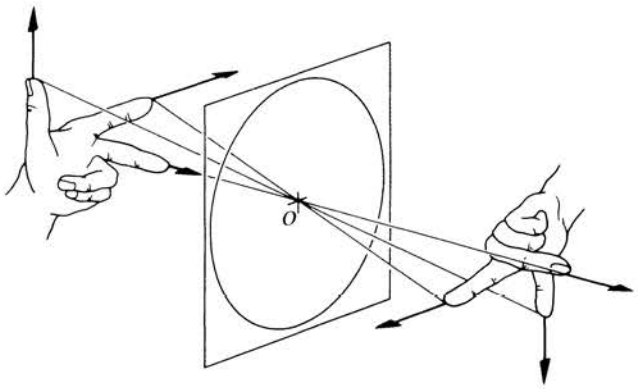


Figure 5.27 Image orientation for a thin lens.

approximate calculations. Refinements are then added as the designer goes on to the prodigious and more exact ray-tracing techniques. Nowadays these computations are carried out by computers. Even so, the simple thin-lens concept provides a highly useful basis for preliminary calculations in a broad range of situations.

No lens is actually a thin lens in the strict sense of having a thickness that approaches zero. Yet many simple lenses, for all practical purposes, function in a fashion equivalent to that of a thin lens (i.e., one that is thin in comparison to its diameter). Almost all spectacle lenses (which, by the way, have been used at least since the thirteenth century) are in this category. When the radii of curvature are large and the lens diameter is small, the thickness will usually be small as well. A lens of this sort would generally have a large focal length, compared with which the thickness would be quite small; many early telescope objectives fit that description perfectly.

We'll now derive expressions for parameters associated with thin-lens combinations. The approach will be fairly simple, leaving the more elaborate traditional treatment for those tenacious enough to pursue the matter into the next chapter.

Consider two thin positive lenses L_1 and L_2 separated by a distance d, which is smaller than either focal length, as in Fig. 5.28. The resulting image can be located graphically as follows. Overlooking L_2 for a moment, construct the image formed exclusively by L_1 using rays-2 and 3. As usual, these pass through the lens object and image foci, F_{o1} and F_{i1}, respectively. The object is in a normal plane, so that two rays determine the top of the image, and a perpendicular to the optical axis finds its bottom. Ray-4 is then constructed running

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

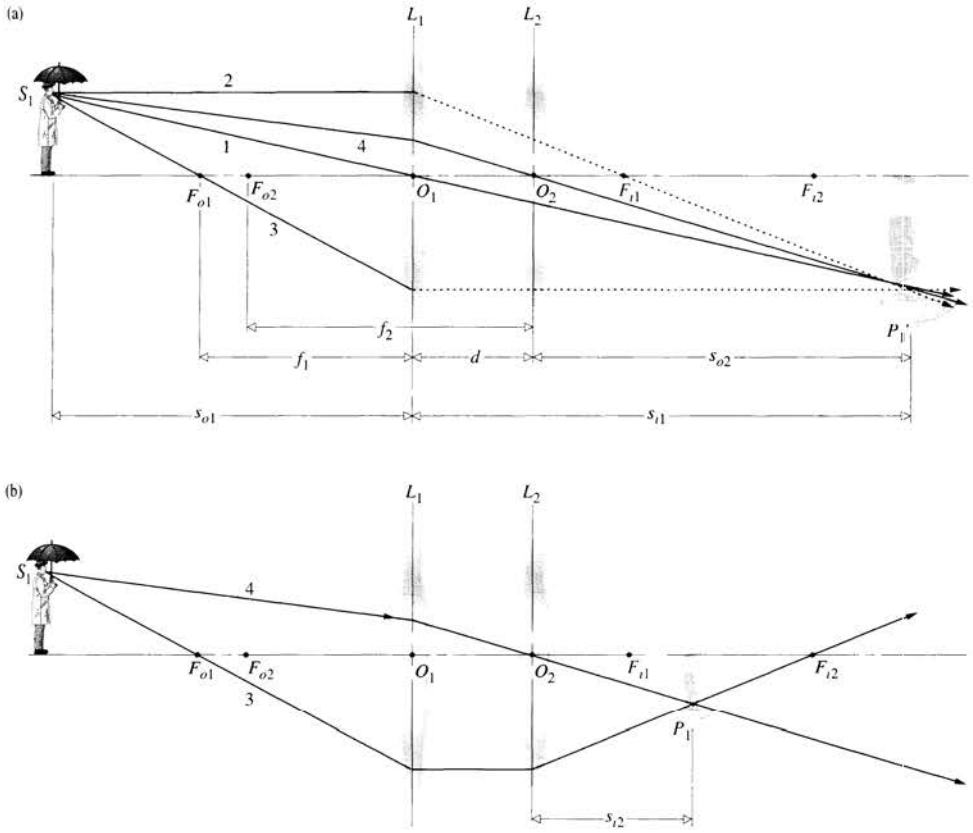


Figure 5.28 Two thin lenses separated by a distance smaller than either focal length.

backward from P_1' through O_2. Insertion of L_2 has no effect on ray-4, whereas ray-3 is refracted through the image focus F_{i2} of L_2. The intersection of rays-4 and 3 fixes the image, which in this particular case is real, minified, and inverted. When the two lenses are close together, as they are here, the presence of L_2 essentially adds convergence ($f_2 > 0$) or divergence ($f_2 < 0$) to the bundle of rays emerging from L_1; see Fig. 5.29.

A similar pair of lenses is illustrated in Fig. 5.30, in which the separation has been increased. Once again rays-2 and 3 through F_{i1} and F_{o1} fix the position of the intermediate image generated by L_1 alone. As before, ray-4 is drawn backward from O_2 to P_1' to S_1. The intersection of rays-3 and 4, as the former is refracted through F_{i2}, locates the final image. This time it is real and erect. Notice that if the focal length of L_2 is increased with all else constant, the size of the image increases as well.

Analytically, for L_1

$$\frac{1}{s_{i1}} = \frac{1}{f_1} - \frac{1}{s_{o1}} \tag{5.29}$$

or

$$s_{i1} = \frac{s_{o1} f_1}{s_{o1} - f_1} \tag{5.30}$$

This is positive, and the intermediate image is to the right of L_1, when $s_{o1} > f_1$ and $f_1 > 0$. For L_2

$$s_{o2} = d - s_{i1} \tag{5.31}$$

and if $d > s_{i1}$, the object for L_2 is real (as in Fig. 5.30), whereas if $d < s_{i1}$, it is virtual ($s_{o2} < 0$, as in Fig. 5.28). In the former instance the rays approaching L_2 are diverging from P_1',

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

whereas in the latter they are converging toward it. Furthermore,

$$\frac{1}{s_{i2}} = \frac{1}{f_2} - \frac{1}{s_{o2}}$$

or

$$s_{i2} = \frac{s_{o2} f_2}{s_{o2} - f_2}$$

Using Eq. (5.31), we obtain

$$s_{i2} = \frac{(d - s_{i1}) f_2}{(d - s_{i1} - f_2)} \tag{5.32}$$

In this same way we could compute the response of any number of thin lenses. It will often be convenient to have a single expression, at least when dealing with only two lenses, so substituting for s_{i1} from Eq. (5.29),

$$s_{i2} = \frac{f_2 d - f_2 s_{o1} f_1/(s_{o1} - f_1)}{d - f_2 - s_{o1} f_1/(s_{o1} - f_1)} \tag{5.33}$$

Here s_{o1} and s_{i2} are the object and image distances, respectively, of the compound lens. As an example, let's compute the image distance associated with an object placed 50.0 cm from the first of two positive lenses. These in turn are separated by 20.0 cm and have focal lengths of 30.0 cm and 50.0 cm, respectively. By direct substitution

$$s_{i2} = \frac{50(20) - 50(50)(30)/(50 - 30)}{20 - 50 - 50(30)/(50 - 30)} = 26.2 \text{ cm}$$

and the image is real. Inasmuch as L_2 "magnifies" the intermediate image formed by L_1, the total transverse magnification of the compound lens is the product of the individual magnifications, that is,

$$M_T = M_{T1} M_{T2}$$

It is left as Problem 5.35 to show that

$$M_T = \frac{f_1 s_{i2}}{d(s_{o1} - f_1) - s_{o1} f_1} \tag{5.34}$$

In the above example

$$M_T = \frac{30(26.2)}{20(50 - 30) - 50(30)} = -0.72$$

and just as we should have guessed from Fig. 5.28, the image is minified and inverted.

The distance from the last surface of an optical system to the second focal point of that system as a whole is known as the **back focal length**, or b.f.l. Similarly, the distance from the vertex of the first surface to the first or object focus is the **front local length**, or f.f.l. Consequently, if we let $s_{i2} \to \infty$, s_{o2} approaches f_2, which combined with Eq. (5.31) tells us that $s_{i1} \to d - f_2$. Hence from Eq. (5.29)

$$\left.\frac{1}{s_{o1}}\right|_{s_{i2}=\infty} = \frac{1}{f_1} - \frac{1}{(d - f_2)} = \frac{d - (f_1 + f_2)}{f_1(d - f_2)}$$

But this special value of s_{o1} is the f.f.l.:

$$\text{f.f.l.} = \frac{f_1(d - f_2)}{d - (f_1 + f_2)} \tag{5.35}$$

In the same way, letting $s_{o1} \to \infty$ in Eq. (5.33), $(s_{o1} - f_1) \to s_{o1}$, and since s_{i2} is then the b.f.l., we have

$$\text{b.f.l.} = \frac{f_2(d - f_1)}{d - (f_1 + f_2)} \tag{5.36}$$

To see how this works numerically, let's find both the b.f.l.

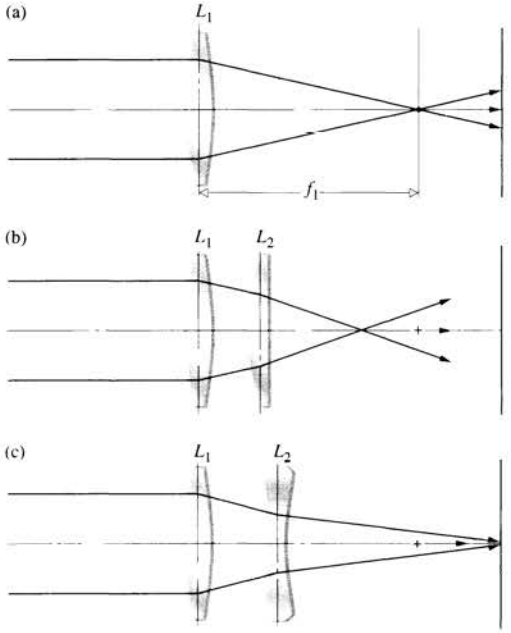


Figure 5.29 (*a*) The effect of placing a second lens, L_2, within the focal length of a positive lens, L_1. (*b*) When L_2 is positive, its presence adds convergence to the ray bundle. (*c*) When L_2 is negative, it adds divergence to the ray bundle.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

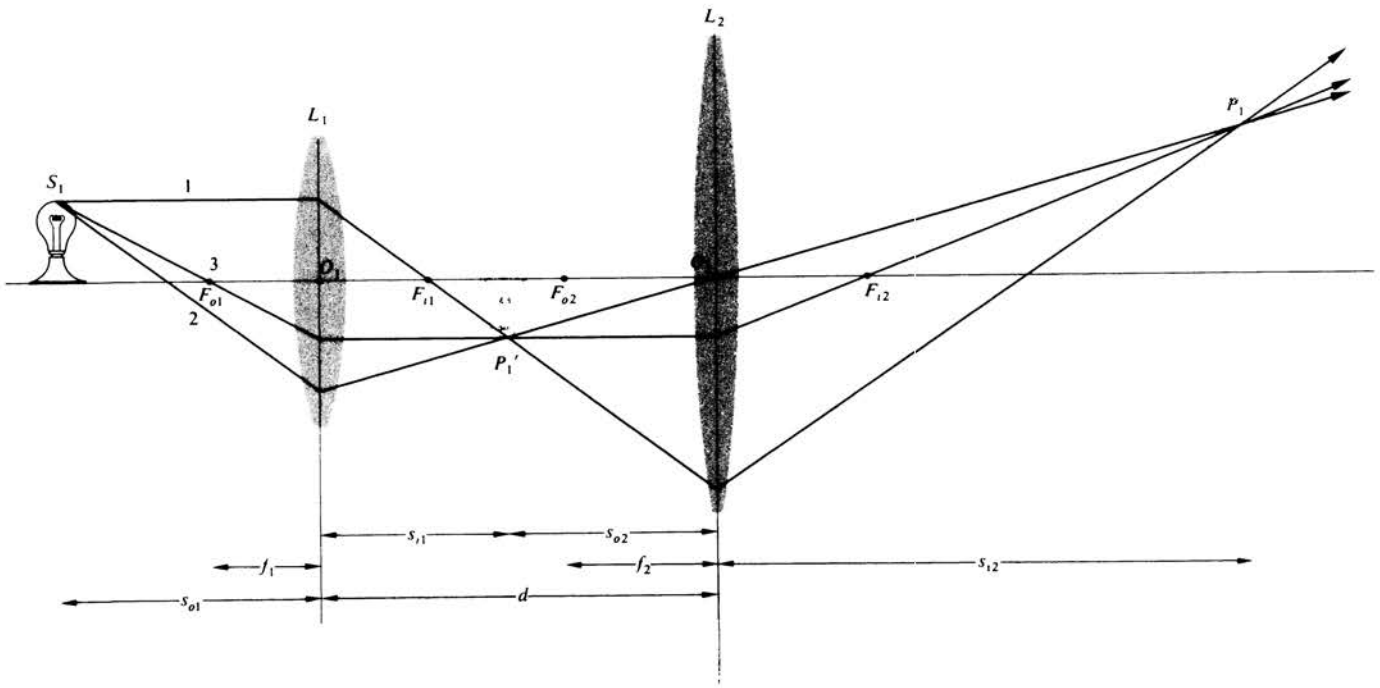


Figure 5.30 Two thin lenses separated by a distance greater than the sum of their focal lengths. Because the intermediate image is real, you could start with point P_1' and treat it as if it were a real object point for L_2. Thus a ray from P_1' through F_{o2} would arrive at P_1.

and f.f.l. for the thin-lens system in Fig. 5.31*a*, where $f_1 = -30$ cm and $f_2 = +20$ cm. Then

$$\text{b.f.l.} = \frac{20[10 - (-30)]}{10 - (-30 + 20)} = 40 \text{ cm}$$

and similarly f.f.l. = 15 cm. Incidentally, notice that if $d = f_1 + f_2$, plane waves entering the compound lens from either side will emerge as plane waves (Problem 5.39), as in telescopic systems.

Observe that if $d \to 0$, that is, the lenses are brought into contact, as in the case of some achromatic doublets,

$$\text{b.f.l.} = \text{f.f.l.} = \frac{f_2 f_1}{f_2 + f_1} \tag{5.37}$$

The resultant thin lens has an *effective focal length*, f, such that

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \tag{5.38}$$

This implies that if there are N such lenses in contact,

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \cdots + \frac{1}{f_N} \tag{5.39}$$

Many of these conclusions can be verified, at least qualitatively, with a few simple lenses. Figure 5.28 is easy to duplicate, and the procedure should be self-evident, whereas Fig. 5.30 requires a bit more care. First, determine the focal lengths of the two lenses by imaging a distant source. Then hold one of the lenses (L_2) at a fixed distance *slightly greater than its focal length* from the plane of observation (i.e., a piece of white paper). Now comes the maneuver that requires some effort if you don't have an optical bench. Move the second lens (L_1) toward the source, keeping it reasonably centered. Without any attempts to block out light entering L_2 directly, you will probably see a blurred image of your hand holding L_1. Position the lenses so that the region on the screen corresponding to L_1 is as bright as possible. The scene spread across L_1 (i.e., its image within the image) will become clear and erect, as in Fig. 5.30.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

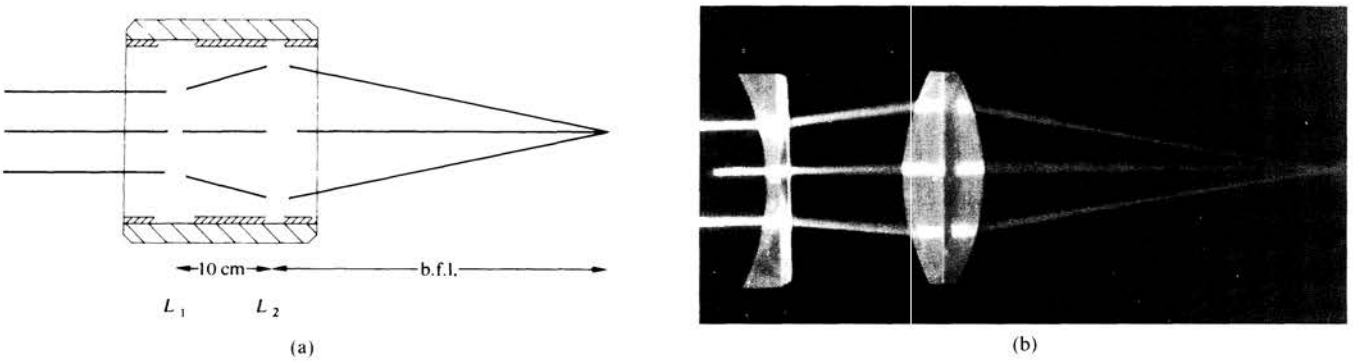


Figure 5.31 (*a*) A positive and negative thin-lens combination, (*b*) photo (Photo by E.H.)

QED and the Lens

One excellent reason for deriving the basic equations of this chapter from Fermat's Principle is that it keeps us thinking in terms of optical path lengths, and that naturally leads to the Feynman treatment of Quantum Electrodynamics. Keep in mind that many physicists consider their theories to be nothing more than the conceptual machinery for calculating the results of observations. And no matter how sophisticated a theory is, it must be in agreement with even the most "ordinary" observation. Thus to see how the operation of a lens fits into the QED worldview, return to Fig. 4.68 and the mirror for a brief review.

Light goes from point S to the mirror to point P along a tremendous number of possible routes. Classically, we note that the *OPL*s are different as, therefore, are the traversal times. In QED, each path has an associated probability amplitude (which has a phase angle proportional to the traversal time). When these are all summed, the most effective contribution to the overall probability of light arriving at P is seen to come from the paths immediately adjacent to the one that has the minimum *OPL*.

For a lens (Fig. 5.32) the situation is very different. We can again approximate things by dividing the device into a manageable number of segments with a possible light path, and therefore a tiny probability amplitude, corresponding to each one. Of course, there should be a lot more than 17 paths, so think of each of these as representing a cluster of billions of neighboring trajectories—the logic doesn't change. Each path has a little probability-amplitude phasor associated with it.

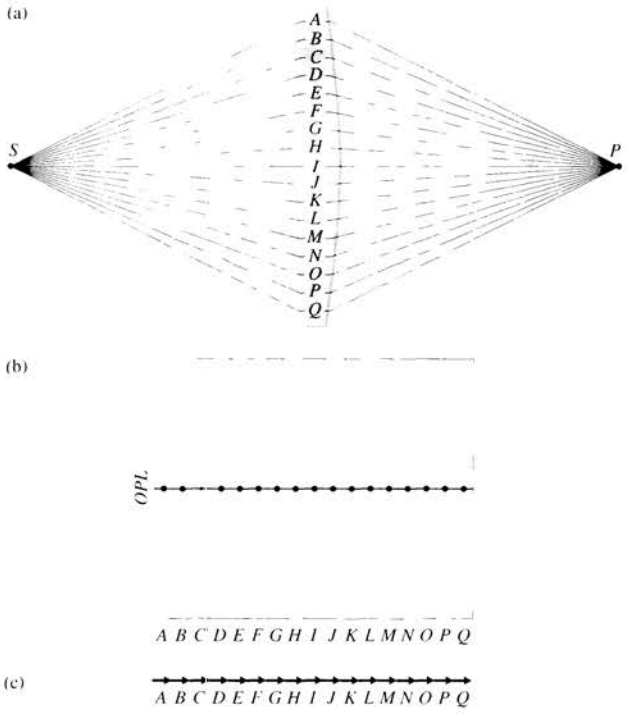


Figure 5.32 Feynman's analysis of the thin lens via QED. (*a*) A number of possible paths from S to P. (*b*) The *OPL* for light along each path. (*c*) The corresponding probability-amplitude phasors all adding in-phase.

Because the lens was designed specifically to make all the *OPL*s equal, a plot of *OPL* (or equivalently the transit times) against distance across the breadth of the lens is a straight line.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Consequently, a photon takes the same time to traverse any one path; all the phasors (each assumed to be the same size) have the same phase angle. Thus, they all contribute equally to the likelihood of a photon arriving at P. Putting the phasors tip-to-tail results in a very large net amplitude, which when squared yields a very high probability of light reaching P via the lens. In the language of QED, ***a lens focuses light, by causing all the constituent probability amplitudes to have the same phase angle***.

For other points in the plane containing P that are close to the optical axis, the phase angles will differ proportionately. The phasors placed tip-to-tail will gradually spiral, and the net probability amplitude will initially diminish quickly, but not discontinuously so. Notice that the probability distribution is not a single infinitesimally narrow spike; the light cannot be focused to a point. The phasors for off-axis points cannot all at once add to zero; what happens, happens gradually and continuously. The resulting circularly symmetric probability distribution, $I(r)$, is known as the Airy pattern (p. 469).

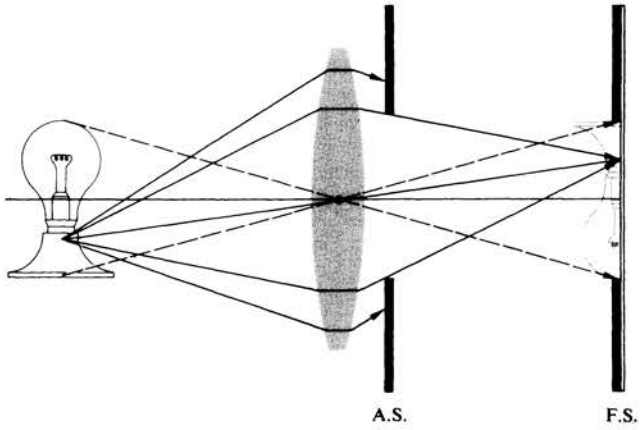


Figure 5.33 Aperture stop and field stop.

5.3 Stops

5.3.1 Aperture and Field Stops

The intrinsically finite nature of all lenses demands that they collect only a fraction of the energy emitted by a point source. The physical limitation presented by the periphery of a simple lens therefore determines which rays shall enter the system to form an image. In that respect, the unobstructed or *clear diameter* of the lens functions as an aperture into which energy flows. Any element, be it the rim of a lens or a separate diaphragm, that determines the amount of light reaching the image is known as the **aperture stop** (abbreviated A.S.). The adjustable leaf diaphragm that is usually located behind the first few elements of a compound camera lens is just such an aperture stop. Evidently, it determines the light-gathering capability of the lens as a whole. As shown in Fig. 5.33, highly oblique rays can still enter a system of this sort. Usually, however, they are deliberately restricted in order to control the quality of the image. The element limiting the size or angular breadth of the object that can be imaged by the system is called the **field stop** or F.S.—it determines the field of view of the instrument. In a camera, the edge of the film itself bounds the image plane and serves as the field stop. Thus, while the aperture stop controls the number of rays from an object point reaching the conjugate image point (Fig. 5.33), it is the field stop that will or will not obstruct those rays *in toto*. Neither the region above the top nor the region below the bottom of the object in Fig. 5.33 passes the field stop. Opening the circular aperture stop would cause the system to accept a larger energy cone and in so doing increase the irradiance at each image point. In contrast, opening the field stop would allow the regions beyond the extremities of the object, which were previously blocked, to be imaged.

5.3.2 Entrance and Exit Pupils

Another concept, useful in determining whether or not a given ray will traverse the entire optical system, is the *pupil*. This is simply an *image of the aperture stop*. The **entrance pupil** of a system is the ***image of the aperture stop as seen from an axial point on the object through those elements preceding the stop***. If there are no lenses between the object and the A.S., the latter itself serves as the entrance pupil. To illustrate the point, examine Fig. 5.34, which is a lens with a *rear aperture stop*. The image of the aperture stop in L is virtual (see Table 5.3) and magnified. It can be located by sending a few rays out from the edges of the A.S. in the usual way. In contrast, the

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

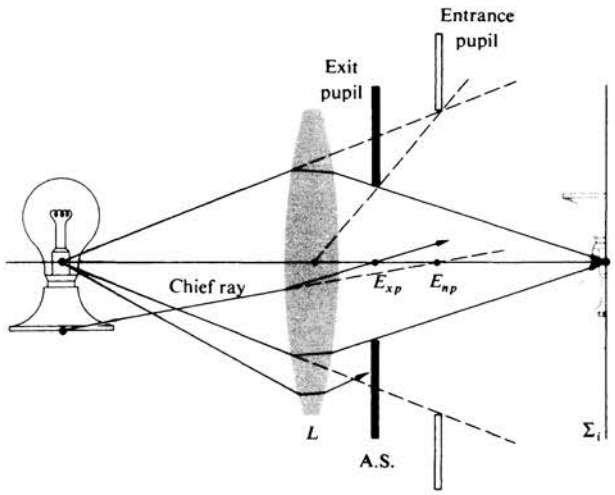


Figure 5.34 Entrance pupil and exit pupil.

exit pupil is the ***image of the A.S. as seen from an axial point on the image plane through the interposed lenses, if there are any***. In Fig. 5.34 there are no such lenses, so the aperture stop itself serves as the exit pupil. Notice that all of this just means that the cone of light actually entering the optical system is determined by the entrance pupil, whereas the cone leaving it is controlled by the exit pupil. No rays from the source point proceeding outside of either cone will make it to the image plane.

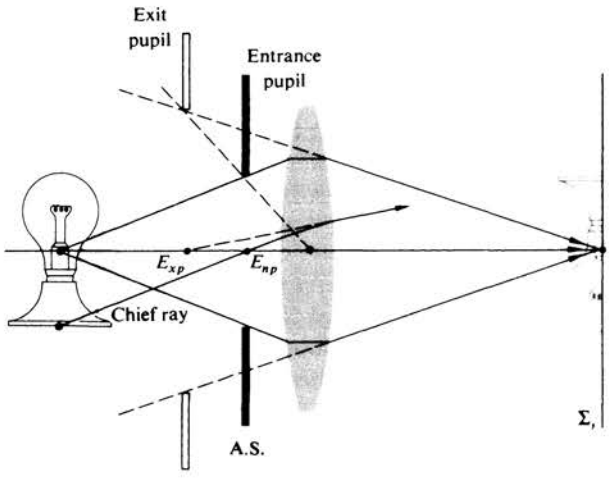


Figure 5.35 A front aperture stop.

To use a telescope or a monocular as a camera lens, you might attach an external *front aperture stop* to control the amount of incoming light for exposure purposes. Figure 5.35 represents a similar arrangement in which the entrance and exit pupil locations should be self-evident. The last two diagrams include a ray labeled the **chief ray**. It is defined to be *any ray from an off-axis object point that passes through the center of the aperture stop. The chief ray enters the optical system along a line directed toward the midpoint of the entrance pupil, E_{np}, and leaves the system along a line passing through the center of the exit pupil, E_{xp}.* The chief ray, associated with a conical bundle of rays from a point on the object, effectively behaves as the central ray of the bundle and is representative of it. Chief rays are of particular importance when the aberrations of a lens design are being corrected.

Figure 5.36 depicts a somewhat more involved arrangement. The two rays shown are those that are usually traced through an optical system. One is the chief ray from a point on the periphery of the object that is to be accommodated by the system. The other is called a **marginal ray**, since it goes from the axial object point to the rim or margin of the entrance pupil (or aperture stop).

In a situation where it is not clear which element is the actual aperture stop, each component of the system must be imaged by the remaining elements to its left. *The image that subtends the smallest angle at the axial object point is the entrance pupil.* The element whose image is the entrance pupil is then the aperture stop of the system for that object point. Problem 5.44 deals with just this kind of calculation.

Notice how the cone of rays, in Fig. 5.37, that can reach the image plane becomes narrower as the object point moves off-axis. The effective aperture stop, which for the axial bundle of rays was the rim of L_1, has been markedly reduced for the off-axis bundle. The result is a gradual fading out of the image at points near its periphery, a process known as **vignetting**.

The locations and sizes of the pupils of an optical system are of considerable practical importance. In visual instruments, the observer's eye is positioned at the center of the exit pupil. The pupil of the eye itself will vary from 2 mm to about 8 mm, depending on the general illumination level. Thus a telescope or binocular designed primarily for evening use might have an exit pupil of at least 8 mm. (You may have

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

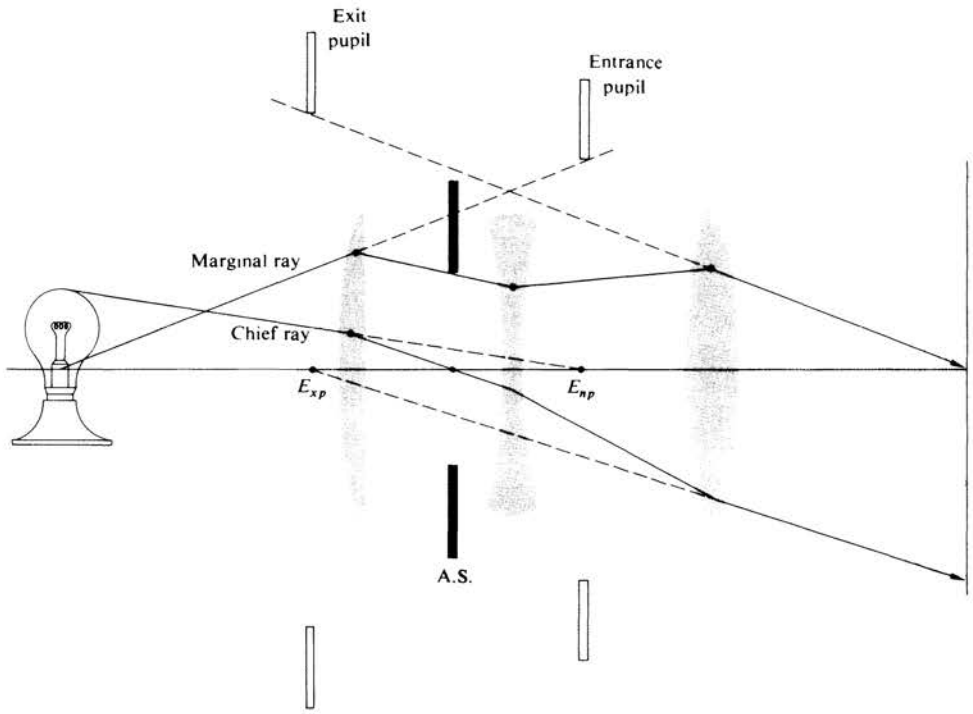


Figure 5.36 Pupils and stops for a three-lens system.

heard the term *night glasses*—they were quite popular on roofs during the Second World War.) In contrast, a daylight version will suffice with an exit pupil of 3 or 4 mm. The larger the exit pupil, the easier it is to align your eye properly with the instrument. Obviously, a telescopic sight for a high-powered rifle should have a large exit pupil located far enough behind the scope so as to avoid injury from recoil.

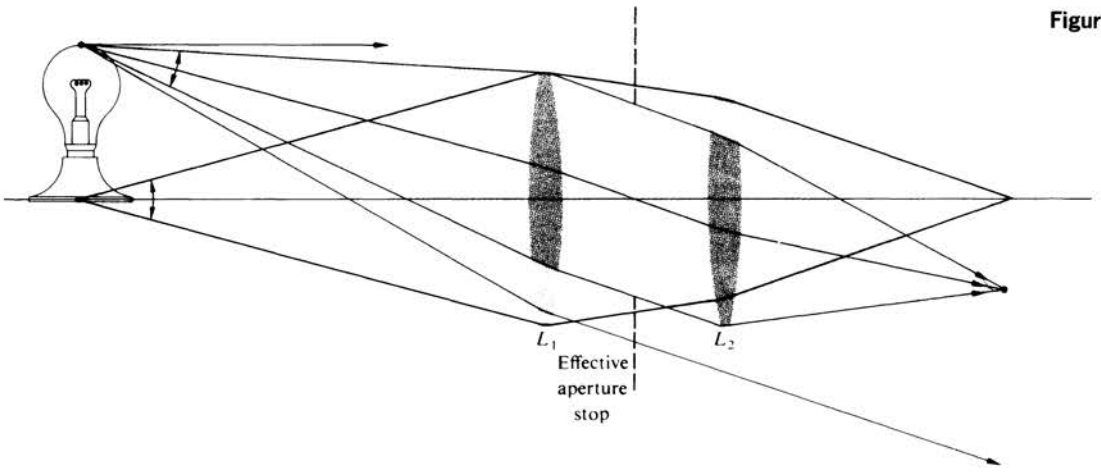


Figure 5.37 Vignetting.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

5.3.3 Relative Aperture and *f*-Number

Suppose we collect the light from an extended source and form an image of it using a lens (or mirror). The amount of energy gathered by the lens (or mirror) from some small region of a distant source will be directly proportional to the area of the lens or, more generally, to the area of the entrance pupil. A large *clear aperture* will intersect a large cone of rays. Obviously, if the source were a laser with a very narrow beam, this would not necessarily be true. If we neglect losses due to reflection, absorption, and so forth, the incoming energy will be spread across a corresponding region of the image (Fig. 5.38). The energy per unit area per unit time (i.e., the flux density or irradiance) will be inversely proportional to the image area.

The entrance pupil area, if circular, varies as the square of its radius and is therefore proportional to the square of its diameter D. Furthermore, the image area will vary as the square of its lateral dimension, which in turn [Eqs. (5.24) and (5.26)] is proportional to f^2. (Keep in mind that we are talking about an extended object rather than a point source. In the latter case, the image would be confined to a very small area independent of f.) Thus the flux density at the image plane varies as $(D/f)^2$. The ratio D/f is known as the *relative aperture*, and its inverse is the **focal ratio** or ***f*-number**, often written $f/\#$, that is,

$$f/\# \equiv \frac{f}{D} \tag{5.40}$$

where $f/\#$ should be understood as a single symbol. For example, a lens with a 25-mm aperture and a 50-mm focal length has an *f-number* of 2, which is usually designated $f/2$. Figure 5.39 illustrates the point by showing a thin lens behind a variable iris diaphragm operating at either $f/2$ or $f/4$. A smaller *f-number* clearly permits more light to reach the image plane.

Camera lenses are usually specified by their focal lengths and largest possible apertures; for example, you might see "50 mm, $f/1.4$" on the barrel of a lens. Since the photographic exposure time is proportional to the square of the *f-number*, the latter is sometimes spoken of as the **speed** of the lens. An $f/1.4$ lens is said to be twice as fast as an $f/2$ lens. Usually, lens diaphragms have *f-number* markings of 1, 1.4, 2, 2.8, 4, 5.6, 8, 11, 16, 22, and so on. The largest relative aperture in this case corresponds to $f/1$, and that's a fast lens—$f/2$ is more typical. Each consecutive diaphragm setting increases the *f-number* by a multiplicative factor of $\sqrt{2}$ (numerically

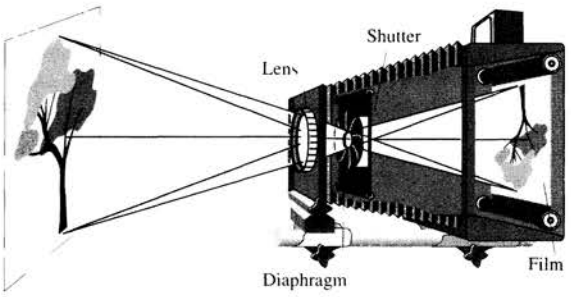


Figure 5.38 A large-format camera usually consists of a lens, followed by an adjustable diaphragm, a shutter that can rapidly open and close, and a sheet of film on which the image is formed.

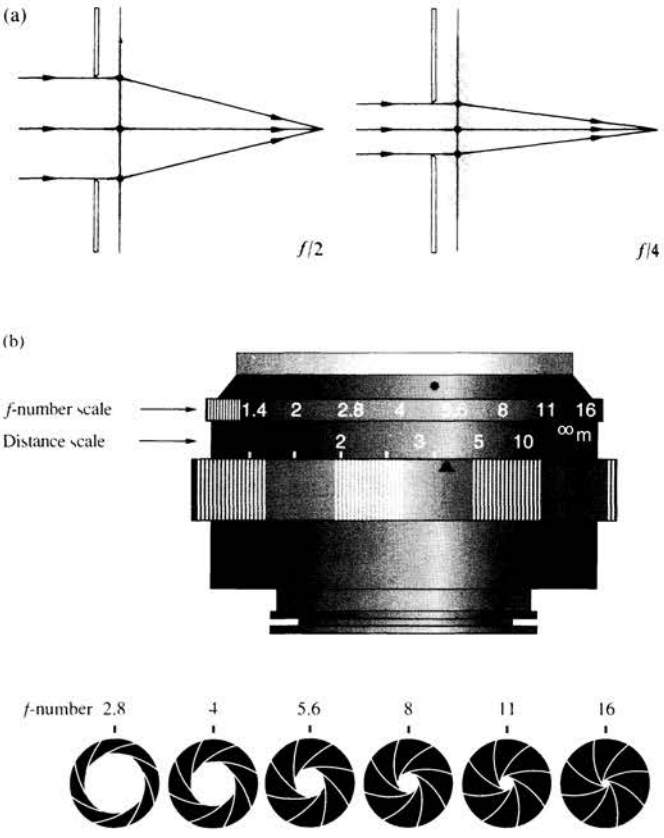


Figure 5.39 (*a*) Stopping down a lens to change the *f-number*. (*b*) A camera lens showing possible settings of the variable diaphragm usually located within the lens.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

rounded off). This corresponds to a decrease in relative aperture by a multiplicative factor of $1/\sqrt{2}$ and therefore a decrease in flux density by one half. Thus, the same amount of light will reach the film whether the camera is set for $f/1.4$ at 1/500th of a second, $f/2$ at 1/250th of a second, or $f/2.8$ at 1/125th of a second.

The largest refracting telescope in the world, located at the Yerkes Observatory of the University of Chicago, has a 40-inch diameter lens with a focal length of 63 feet and therefore an *f-number* of 18.9. The entrance pupil and focal length of a mirror will, in exactly the same way, determine its *f-number*. Accordingly, the 200-inch diameter mirror of the Mount Palomar telescope, with a prime focal length of 666 inches, has an *f-number* of 3.33.

5.4 Mirrors

Mirror systems are increasingly being used, particularly in the X-ray, ultraviolet, and infrared regions of the spectrum. Although it is relatively simple to construct a reflecting device that will perform satisfactorily across a broad-frequency band, the same cannot be said of refracting systems. For example, a silicon or germanium lens designed for the infrared will be completely opaque in the visible (see photo on p. 73). As we will see later (p. 256), mirrors have other attributes that also contribute to their usefulness.

A mirror might simply be a piece of black glass or a finely polished metal surface. In the past, mirrors were usually made by coating glass with silver, which was chosen because of its high efficiency in the UV and IR (see Fig. 4.59). Vacuum-evaporated coatings of aluminum on highly polished substrates have become the accepted standard for quality mirrors. Protective coatings of silicon monoxide or magnesium fluoride are often layered over the aluminum as well. In special applications (e.g., in lasers), where even the small losses due to metal surfaces cannot be tolerated, mirrors formed of multilayered dielectric films (see Section 9.9) are indispensable.

A new generation of lightweight precision mirrors continues to be developed for use in large-scale orbiting telescopes; the technology is by no means static.

5.4.1 Planar Mirrors

As with all mirror configurations, those that are planar can be either front- or back-surfaced. The latter type are most commonly found in everyday use because it allows the metallic reflecting layer to be completely protected behind glass. In contrast, the majority of mirrors designed for more critical technical usage are front-surfaced (Fig. 5.40).

From Section 4.3.1, it's an easy matter to determine the image characteristics of a planar mirror. Examining the point source and mirror arrangement of Fig. 5.40, we can quickly show that $|s_o| = |s_i|$; that is, the image P and object S are

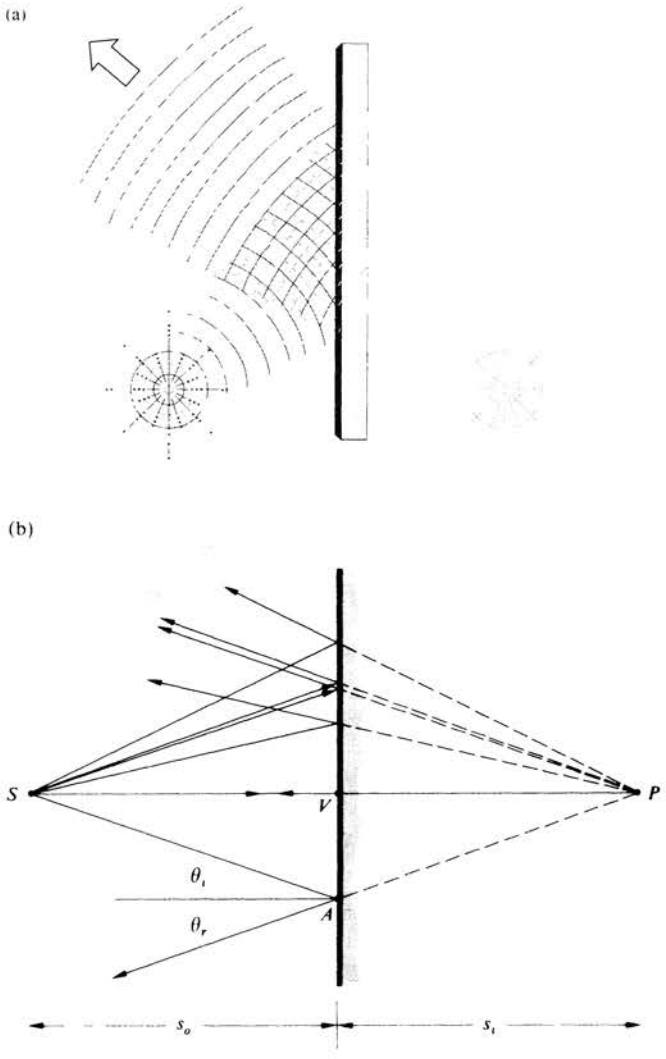


Figure 5.40 A planar mirror. (*a*) Reflection of waves. (*b*) Reflection of rays.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

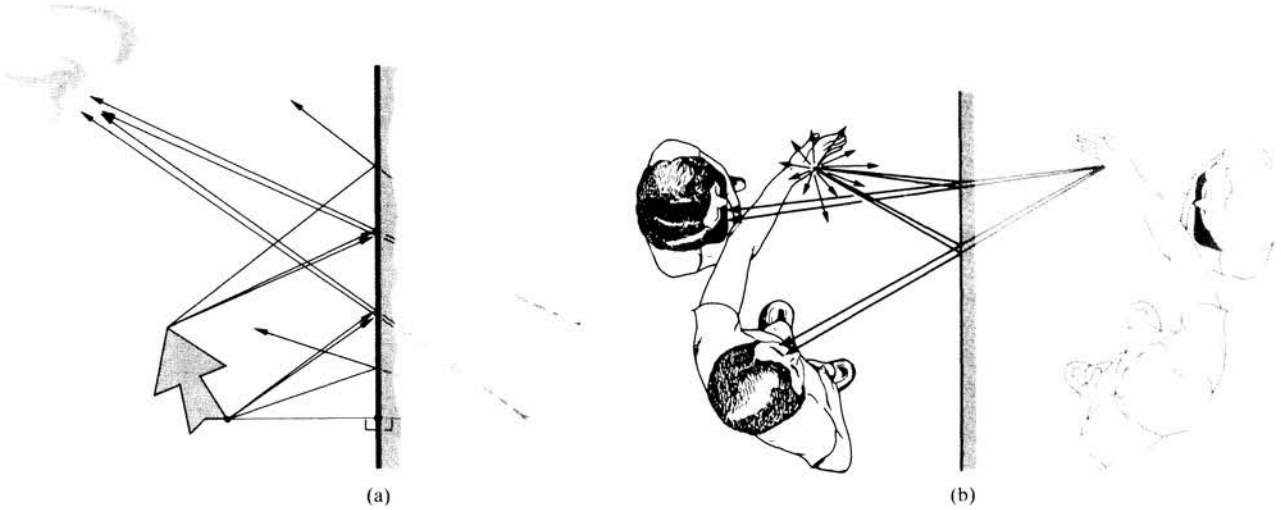


Figure 5.41 (*a*) The image of an extended object in a planar mirror. (*b*) Images in a planar mirror.

equidistant from the surface. To wit, $\theta_i = \theta_r$, from the Law of Reflection; $\theta_i + \theta_r$ is the exterior angle of triangle *SPA* and is therefore equal to the sum of the alternate interior angles, $\measuredangle VSA + \measuredangle VPA$. But $\measuredangle VSA = \theta_i$, and therefore $\measuredangle VSA = \measuredangle VPA$. This makes triangles *VAS* and *VPA* congruent, in which case $|s_o| = |s_i|$.

We are now faced with the problem of determining a sign convention for mirrors. Whatever we choose, and you should certainly realize that there is a choice, we need only be faithful unto it for all to be well. One obvious dilemma with respect to the convention for lenses is that now the virtual image is to the right of the interface. The observer sees *P* to be positioned behind the mirror because the eye (or camera) cannot perceive the actual reflection; it merely interpolates the rays backward along straight lines. The rays from *P* in Fig. 5.41 are diverging, and no light can be cast on a screen located at *P*—the image is certainly virtual. Clearly, it is a matter of taste whether s_i should be defined as positive or negative in this instance. Since we rather like the idea of virtual object and image distances being negative, we define *s_o and s_i as negative when they lie to the right of the vertex V*. This will have the added benefit of yielding a mirror formula identical to the Gaussian Lens Equation [Eq. (5.17)]. Evidently, the same definition of the transverse magnification [Eq. (5.24)] holds, where now, as before, $M_T = +1$ indicates a *life-size*, virtual, erect image.

Each point of the extended object in Fig. 5.41, a perpendicular distance s_i from the mirror, is imaged that same distance behind the mirror. In this way, the entire image is built up point by point. This is considerably different from the way a lens locates an image. The object in Fig. 5.27 was a left

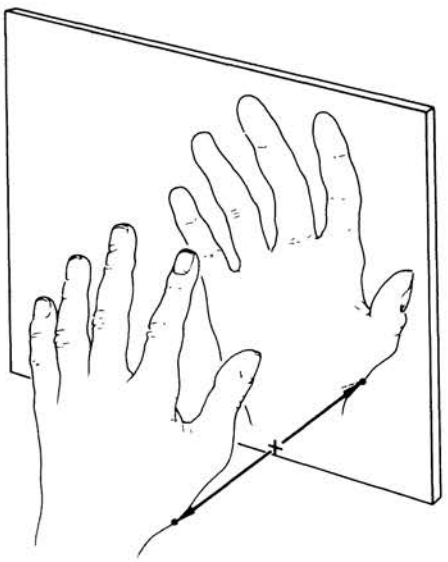


Figure 5.42 Mirror images—inversion.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

hand, and the image formed by the lens was also a left hand. To be sure, it might have been distorted ($M_L \neq M_T$), but it was still a left hand. The only evident change was a 180° rotation about the optical axis—an effect known as *reversion*. Contrarily, the mirror image of the left hand, determined by dropping perpendiculars from each point, is a right hand (Fig. 5.42). Such an image is sometimes said to be *perverted*. In deference to the more usual lay connotation of the word, its use in optics is happily waning. The process that converts a right-handed coordinate system in the object space into a left-handed one in the image space is known as *inversion*. Systems with more than one planar mirror can be used to produce either an odd or even number of inversions. In the latter case a right-handed (r-h) object will generate a right-handed image (Fig. 5.43), whereas in the former instance, the image will be left-handed (l-h).

A number of practical devices utilize rotating planar mirror systems—for example, choppers, beam deflectors, image rotators, and scanners. Mirrors are frequently used to amplify and measure the slight rotations of certain laboratory apparatus (galvanometers, torsion pendulums, current balances, etc.). As Fig. 5.44 shows, if the mirror rotates through an angle α, the reflected beam or image will move through an angle of 2α.

The ability to rapidly redirect a beam of light is an inherent virtue of planar mirrors that has been utilized for centuries; the single-lens reflex camera (see photo on p. 178) is just one

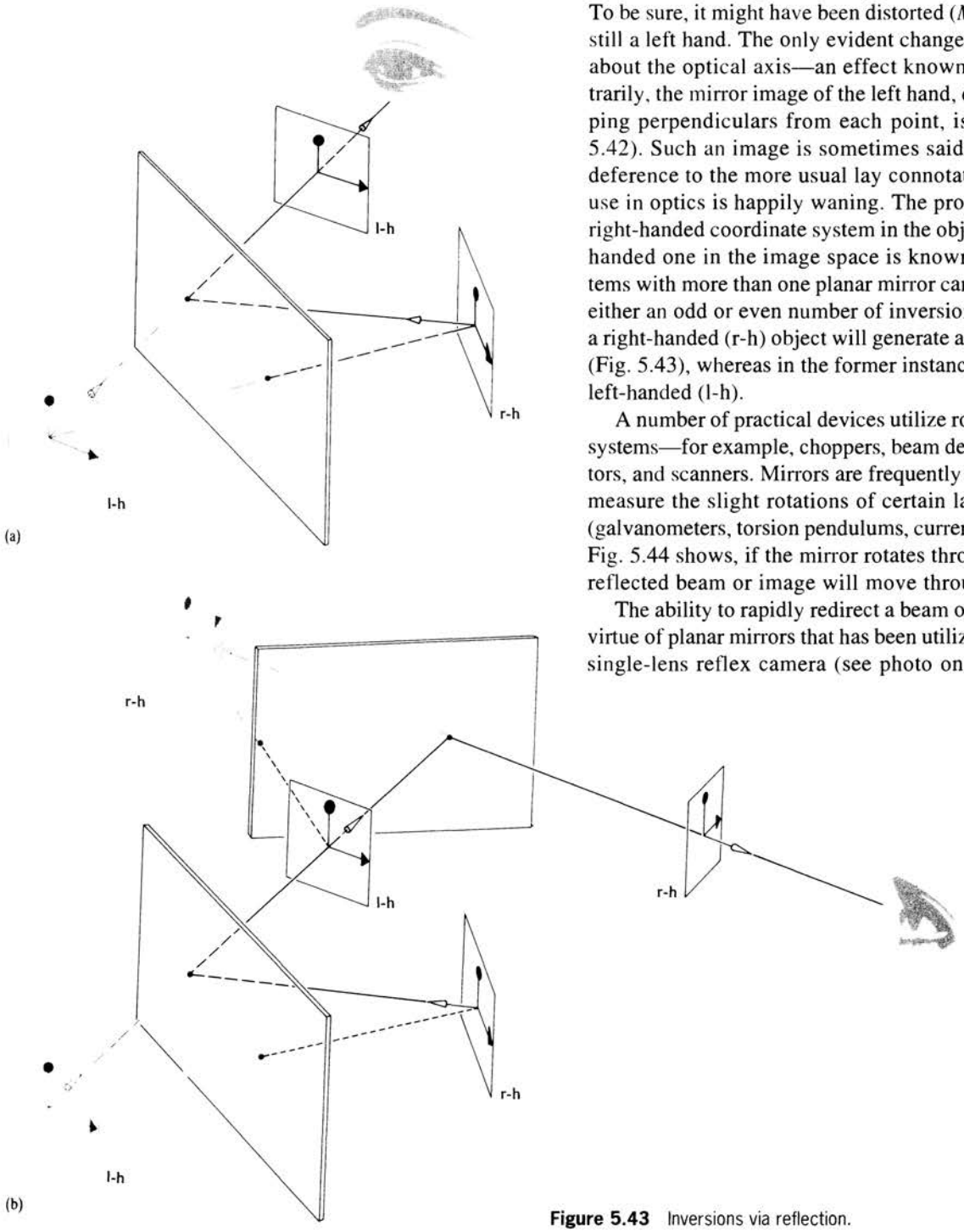


Figure 5.43 Inversions via reflection.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

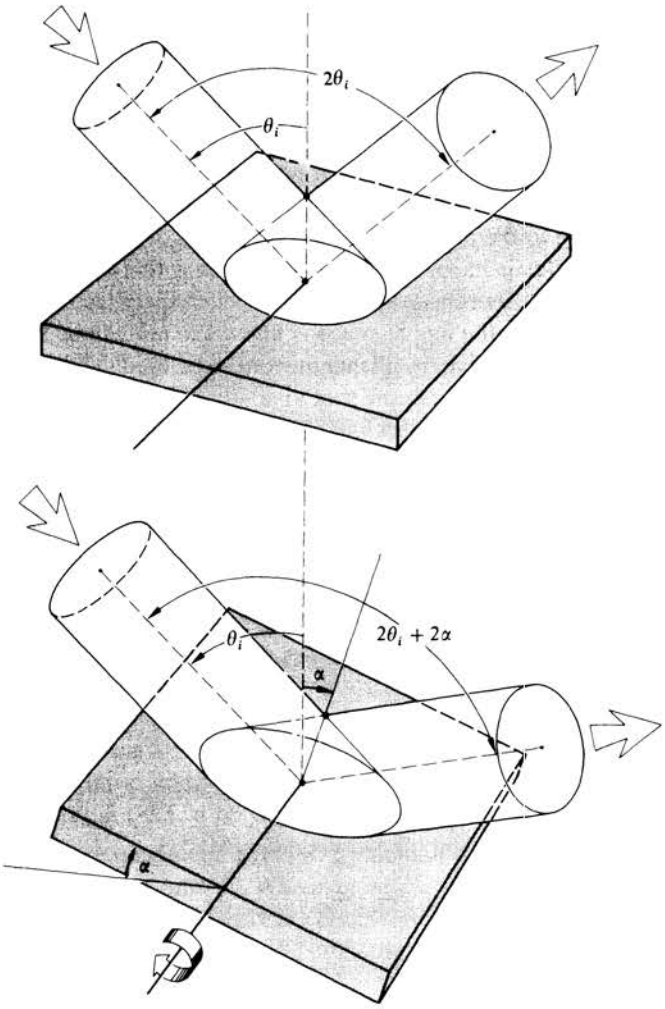
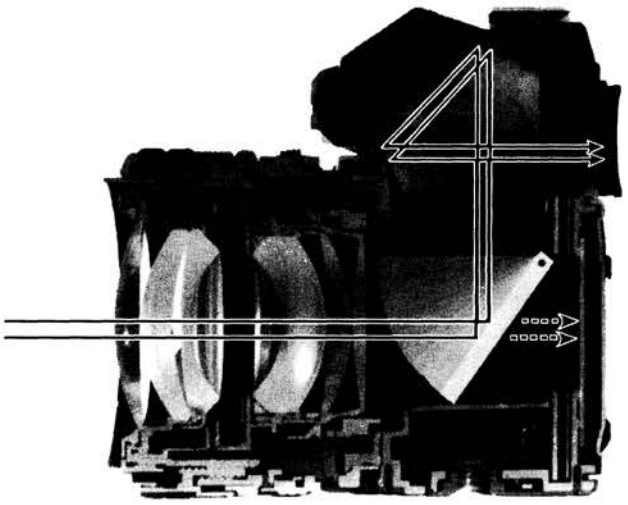
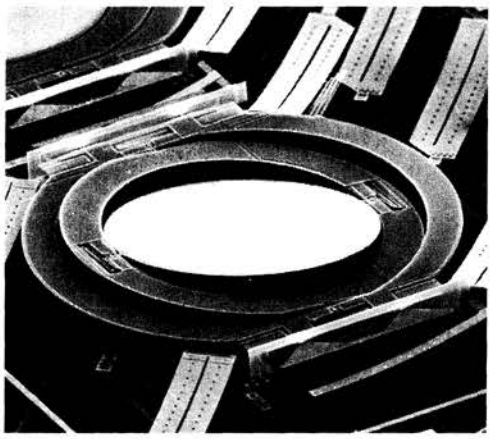


Figure 5.44 Rotation of a mirror and the concomitant angular displacement of a beam.

application that comes to mind. Today micromirrors (see photo), small enough to pass through the eye of a needle, have become part of the flourishing **MOEMS** (Micro-OptoElectro-Mechanical Systems) technology. The telecommunications network that delivers worldwide telephone, fax, and Internet services is undergoing a quiet microphotonic revolution as it sheds its electronic elements and moves toward becoming entirely optical. Electronic switches are expensive, bulky, and, by optic standards, unacceptably slow. Accordingly, the crucial component needed for that transition is the optical switch. Micromirrors that can tilt side-to-side and top-to-bottom in a matter of milliseconds are presently one of the most promising approaches (see p.200).



A single-lens reflex camera. Light from the lens hits the mirror and goes up to the prism and out to the eye. When the shutter is released the mirror pops up, the light goes directly to the film, and then the mirror pops back down. (Photo by E.H.)



This tiny tiltable mirror (which is so small it can fit through the eye of a needle) is used to steer light beams in one of today's most important telecommunications devices. (Photo courtesy Lucent Technologies' Bell Laboratories.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

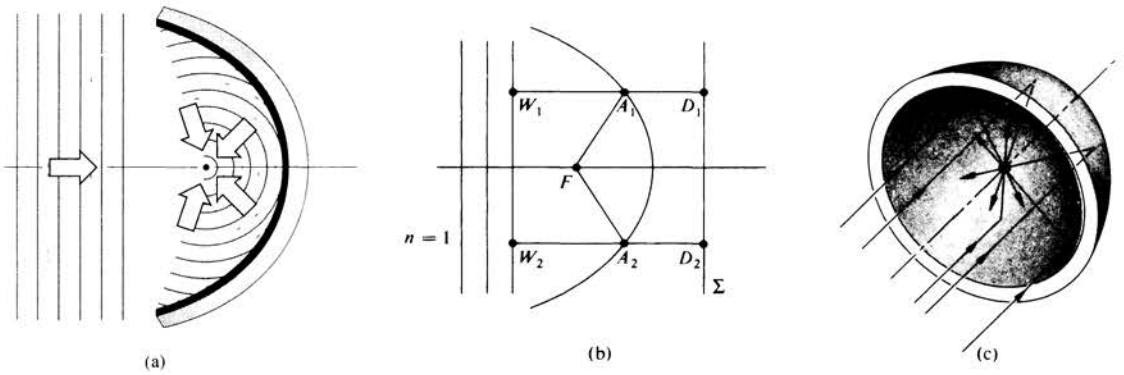


Figure 5.45
A paraboloidal mirror.

5.4.2 Aspherical Mirrors

Curved mirrors that form images very much like those of lenses or curved refracting surfaces have been known since the time of the ancient Greeks. Euclid, who is presumed to have authored the book entitled *Catoptrics*, discusses in it both concave and convex mirrors.* Fortunately, the conceptual basis for designing such mirrors was developed earlier when we studied Fermat's Principle as applied to imagery in refracting systems. Accordingly, let's determine the configuration a mirror must have if an incident plane wave is to be reformed upon reflection into a converging spherical wave (Fig. 5.45). Because the plane wave is ultimately to converge on point F, the optical path lengths for all rays must be equal; accordingly, for arbitrary points A_1 and A_2

$$OPL = \overline{W_1A_1} + \overline{A_1F} = \overline{W_2A_2} + \overline{A_2F} \qquad (5.41)$$

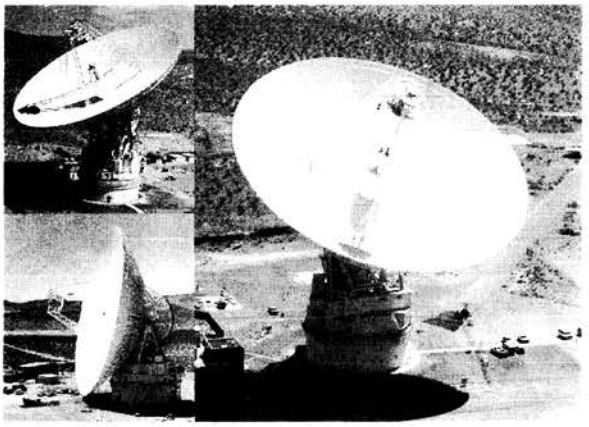
Since the plane Σ is parallel to the incident wavefronts,

$$\overline{W_1A_1} + \overline{A_1D_1} = \overline{W_2A_2} + \overline{A_2D_2} \qquad (5.42)$$

Equation (5.41) will therefore be satisfied for a surface for which $\overline{A_1F} = \overline{A_1D_1}$ and $\overline{A_2F} = \overline{A_2D_2}$ or, more generally, one for which $\overline{AF} = \overline{AD}$ for any point A on the mirror. In general, $\overline{AF} = e(\overline{AD})$, where e is the eccentricity of a **conic section**.

Dioptrics denotes the optics of refracting elements, whereas *catoptrics* denotes the optics of reflecting surfaces.

Earlier (Section 5.2.1) the figure studied was a hyperbola for which $e = n_{ti} > 1$. In Problem 5.3 the figure is an ellipse and $e = n_{ti} < 1$. Here the second medium is identical to the first, $n_t = n_i$, and $e = n_{ti} = 1$; in other words, the surface is a paraboloid with F as its focus and Σ as its directrix. The rays could equally well be reversed (i.e., a point source at the focus of a paraboloid would result in the emission of plane waves from the system). Paraboloidals are used in a great variety of applications from flashlight and automobile headlight reflectors to giant radiotelescope antennas (see photo), from microwave horns and acoustical dishes to optical telescope mirrors and Moon-based communications antennas. The convex paraboloidal mirror is also possible but is far less widely in use. Applying what we already know, it should be evident from



A large paraboloidal radio antenna at the Goldstone Deep Space Communications Complex. (Photo courtesy of NASA.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

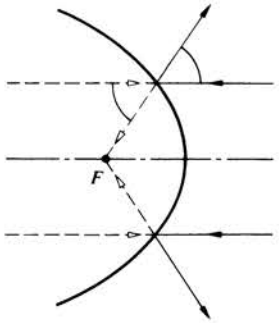


Figure 5.46 Real and virtual images for a paraboloidal mirror.

Fig. 5.46 that an incident parallel bundle of rays will form a virtual image at F when the mirror is convex and a real image when it's concave.

There are other aspherical mirrors of interest, namely, the ellipsoid ($e < 1$) and hyperboloid ($e > 1$). Both produce perfect imagery between a pair of conjugate axial points corresponding to their two foci (Fig. 5.47). As we'll see presently, the Cassegrain and Gregorian telescope configurations utilize convex secondary mirrors that are hyperboloidal and ellipsoidal, respectively. Like many new instruments, the primary mirror of the Hubble Space Telescope is hyperboloidal (see photo).

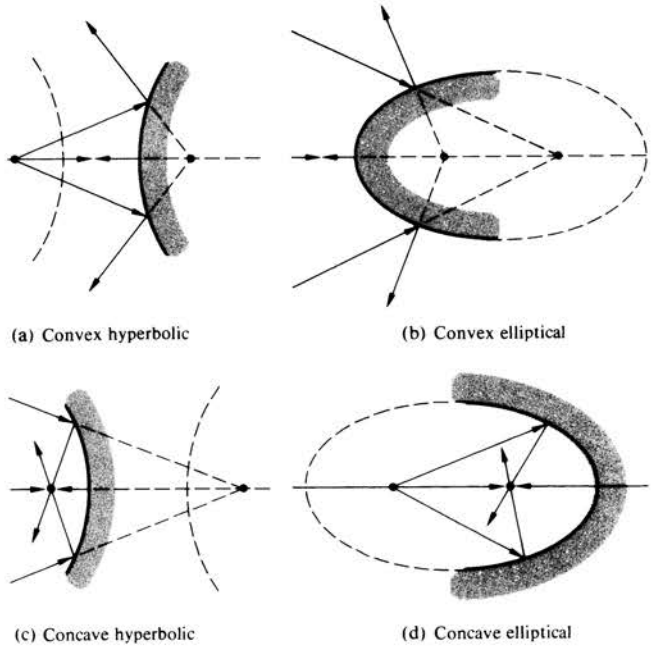
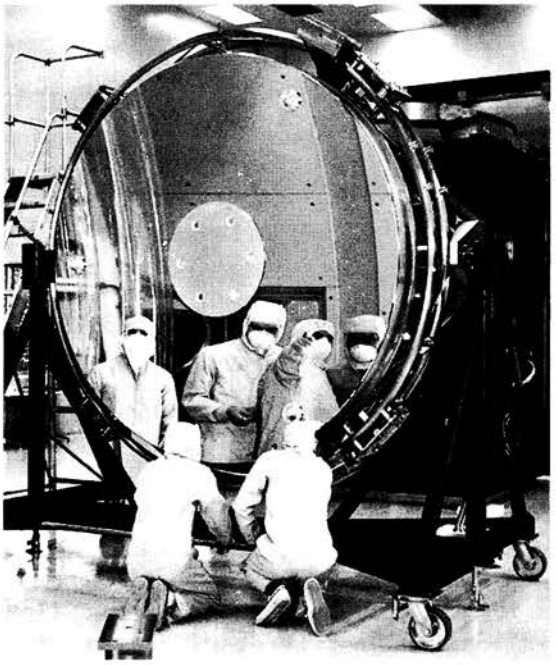


Figure 5.47 Hyperbolic and elliptical mirrors.



The 2.4-m-diameter hyperboloidal primary mirror of the Hubble Space Telescope. (Photo courtesy of NASA.)

A variety of aspherical mirrors are readily available commercially. In fact, one can purchase *off-axis elements*, in addition to the more common centered systems. Thus in Fig. 5.48 the focused beam can be further processed without obstructing the mirror. Incidentally, this geometry also obtains in large microwave horn antennas.

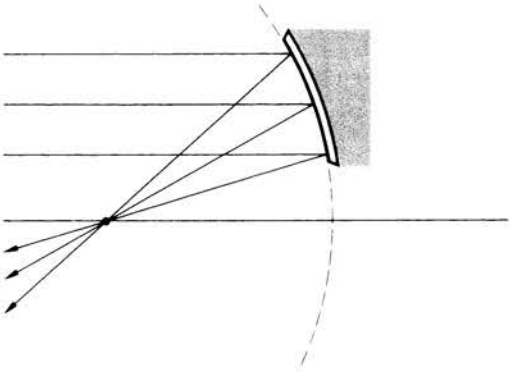


Figure 5.48 An off-axis parabolic mirror element.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

5.4.3 Spherical Mirrors

Precise aspheric surfaces are considerably more difficult to fabricate than are spherical ones, and, not surprisingly, they're considerably more expensive. Accordingly, we again turn to the spherical configuration to determine the circumstances under which it might perform adequately.

The Paraxial Region

The well-known equation for the circular cross section of a sphere (Fig. 5.49*a*) is

$$y^2 + (x - R)^2 = R^2 \tag{5.43}$$

where the center C is shifted from the origin O by one radius R. After writing this as

$$y^2 - 2Rx + x^2 = 0$$

we can solve for x:

$$x = R \pm (R^2 - y^2)^{1/2} \tag{5.44}$$

Let's just concern ourselves with values of x less than R; that is, we'll study a hemisphere, open on the right, corresponding to the minus sign in Eq. (5.44). After expansion in a binomial series, x takes the form

$$x = \frac{y^2}{2R} + \frac{1y^4}{2^2 2! R^3} + \frac{1 \cdot 3y^6}{2^3 3! R^5} + \cdots \tag{5.45}$$

This expression becomes quite meaningful as soon as we realize that the standard equation for a parabola with its vertex at the origin and its focus a distance f to the right (Fig. 5.49*b*) is simply

$$y^2 = 4fx \tag{5.46}$$

By comparing these two formulas, we see that if $4f = 2R$ (i.e., if $f = R/2$), the first contribution in the series can be thought of as parabolic, and the remaining terms represent the deviation. If that deviation is Δx, then

$$\Delta x = \frac{y^4}{8R^3} + \frac{y^6}{16R^5} + \cdots$$

Evidently, this difference will be appreciable only when y is relatively large (Fig. 5.49*c*) in comparison to R. *In the paraxial region, that is, in the immediate vicinity of the central axis, these two configurations will be essentially indistinguishable.* If we stay within the paraxial theory of spherical mirrors as a first approximation, the conclusions drawn from our study of the stigmatic imagery of paraboloids are again applicable. In actual use, however, y will not be so limited, and aberrations will appear. Moreover, aspherical surfaces produce perfect images only for pairs of axial points—they too will suffer aberrations.

The Mirror Formula

The paraxial equation that relates conjugate object and image points to the physical parameters of a spherical mirror can be derived with the help of Fig. 5.50. To that end, observe that since $\theta_i = \theta_r$, the $\measuredangle SAP$ is bisected by $\overline{CA}$, which therefore divides the side $\overline{SP}$ of triangle SAP into segments proportion-

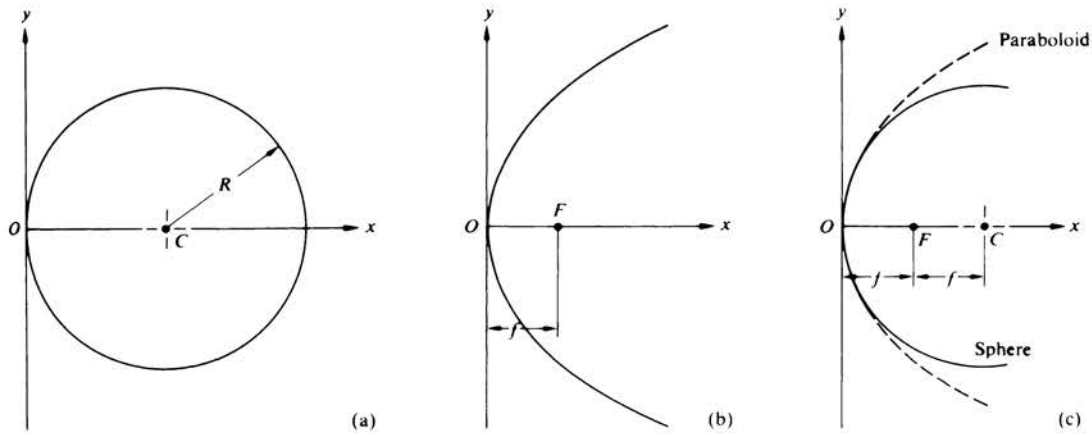


Figure 5.49
Comparison of spherical and paraboloidal mirrors.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

al to the remaining two sides; that is,

$$\frac{\overline{SC}}{\overline{SA}} = \frac{\overline{CP}}{\overline{PA}} \tag{5.47}$$

Furthermore,

$$\overline{SC} = s_o - |R| \quad \text{and} \quad \overline{CP} = |R| - s_i$$

where s_o and s_i are on the left and therefore positive. Using the same sign convention as we did with refraction, R will be negative because C is to the left of V (i.e., the surface is concave).

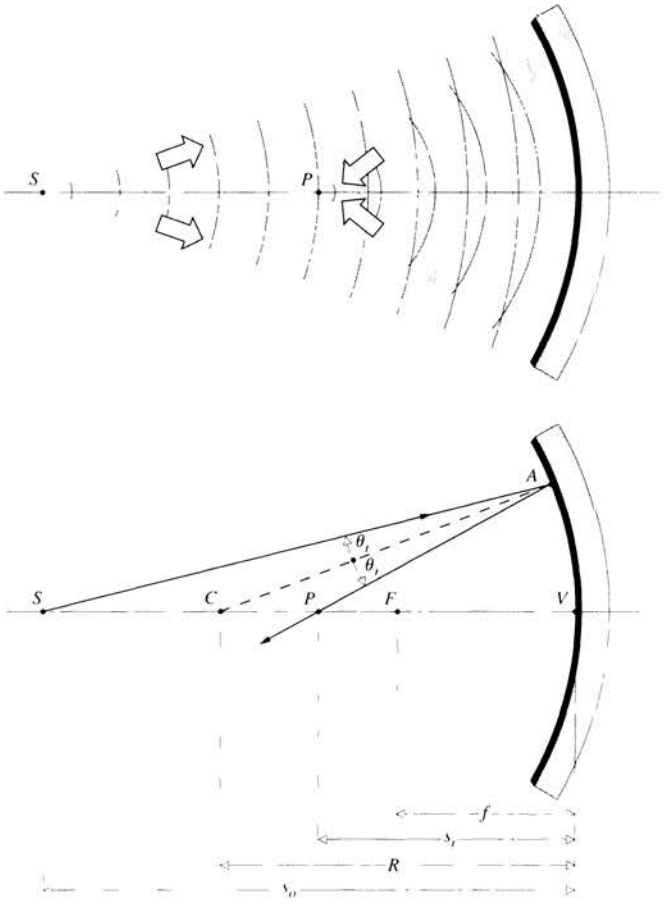


Figure 5.50 A concave spherical mirror. Conjugate foci.

Thus $|R| = -R$ and

$$\overline{SC} = s_o + R \quad \text{and} \quad \overline{CP} = -(s_i + R)$$

In the paraxial region $\overline{SA} \approx s_o$, $\overline{PA} \approx s_i$, and Eq. (5.47) becomes

$$\frac{s_o + R}{s_o} = -\frac{s_i + R}{s_i}$$

or

$$\frac{1}{s_o} + \frac{1}{s_i} = -\frac{2}{R} \tag{5.48}$$

which is the **Mirror Formula**. It's equally applicable to concave ($R < 0$) and convex ($R > 0$) mirrors. The *primary* or *object focus* is again defined by

$$\lim_{s_i \to \infty} s_o = f_o$$

and the *secondary* or *image focus* corresponds to

$$\lim_{s_o \to \infty} s_i = f_i$$

Consequently, from Eq. (5.48)

$$\frac{1}{f_o} + \frac{1}{\infty} = \frac{1}{\infty} + \frac{1}{f_i} = -\frac{2}{R}$$



A convex spherical mirror forming a virtual, right-side-up, minified image. See if you can locate the image of the author. (Photo by E.H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

to wit,

$$f_o = f_i = -\frac{R}{2} \tag{5.49}$$

as can be seen in Fig. 5.49*c*. Dropping the subscripts on the focal lengths yields

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f} \tag{5.50}$$

Observe that f will be positive for concave mirrors ($R < 0$) and negative for convex mirrors ($R > 0$). In the latter instance, the image is formed behind the mirror and is virtual (Fig. 5.51).

Finite Imagery

The remaining mirror properties are so similar to those of lenses and spherical refracting surfaces that we need only mention them briefly, without repeating the entire logical development of each item. Within the restrictions of paraxial theory, any parallel off-axis bundle of rays will be focused to a point on the *focal plane* passing through F normal to the optical axis. Likewise, a finite planar object perpendicular to the optical axis will be imaged (to a first approximation) in a plane similarly oriented; each object point will have a corresponding image point in that plane. This is certainly true for a plane mirror, but it only approximates the case for other configurations.

If a spherical mirror is used in a restricted fashion, the reflected waves arising from each object point will closely

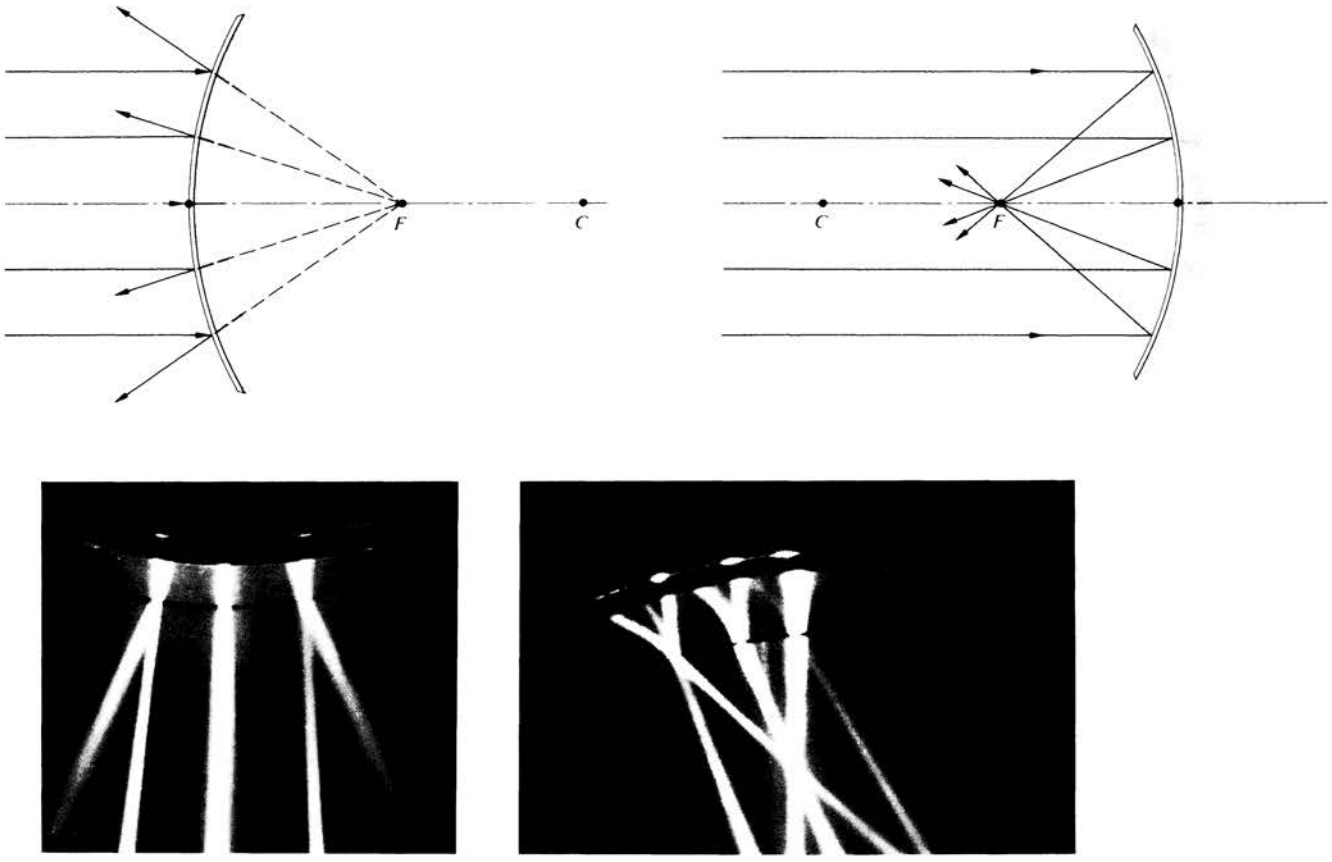


Figure 5.51 Focusing of rays via a spherical mirror. (Photo by E.H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

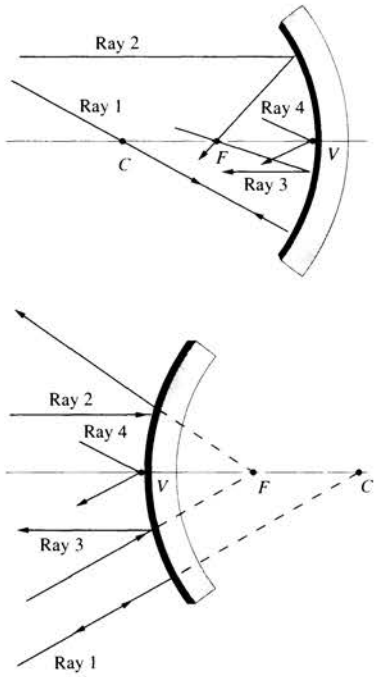


Figure 5.52 Four easy rays to draw. Ray 1 heads toward C and reflects back along itself. Ray 2 comes in parallel to the central axis and reflects toward (or away from) F. Ray 3 passes through (or heads toward) F and reflects off parallel to the axis. Ray 4 strikes point V and reflects such that $\theta_i = \theta_r$.

TABLE 5.4 Sign Convention for Spherical Mirrors

Quantity	Sign	
	+	−
s_o	Left of V, real object	Right of V, virtual object
s_i	Left of V, real image	Right of V, virtual image
f	Concave mirror	Convex mirror
R	C right of V, convex	C left of V, concave
y_o	Above axis, erect object	Below axis, inverted object
y_i	Above axis, erect image	Below axis, inverted image

approximate spherical waves. Under such circumstances, good finite images of extended objects can be formed.

Just as each image point produced by a thin lens lies along a straight line through the optical center O, each image point for a spherical mirror will lie on a ray passing through both the center of curvature C and the object point (Fig. 5.52). As with the thin lens (Fig. 5.21), the process for graphically locating the image is straightforward (Fig. 5.53). The top of the image is fixed at the intersection of two rays, one initially parallel to the axis and passing through F after reflection, and the other going straight through C (Fig. 5.54). The ray from any off-axis object point to the vertex forms equal angles with the central axis on reflection and is therefore particularly convenient to construct. So too is the ray that first passes through the focus and after reflection emerges parallel to the axis.

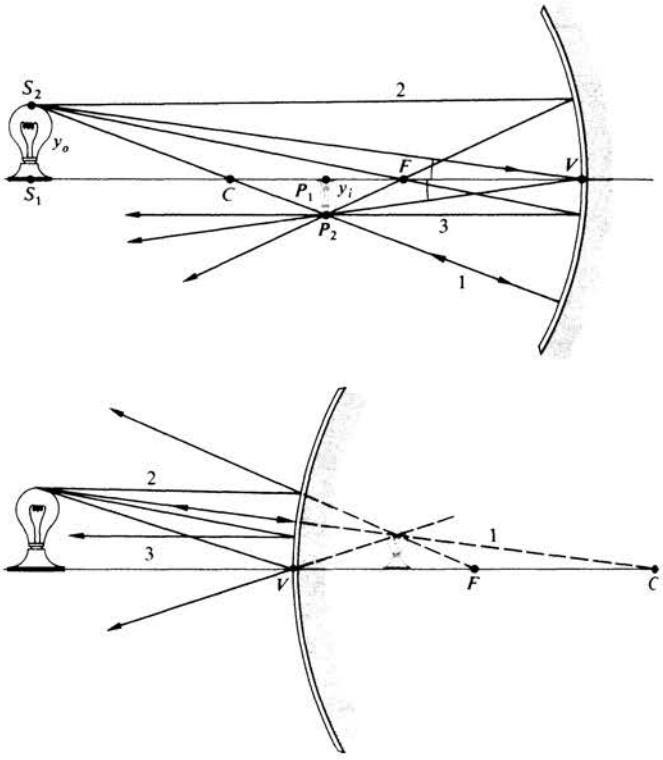


Figure 5.53 Finite imagery with spherical mirrors.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

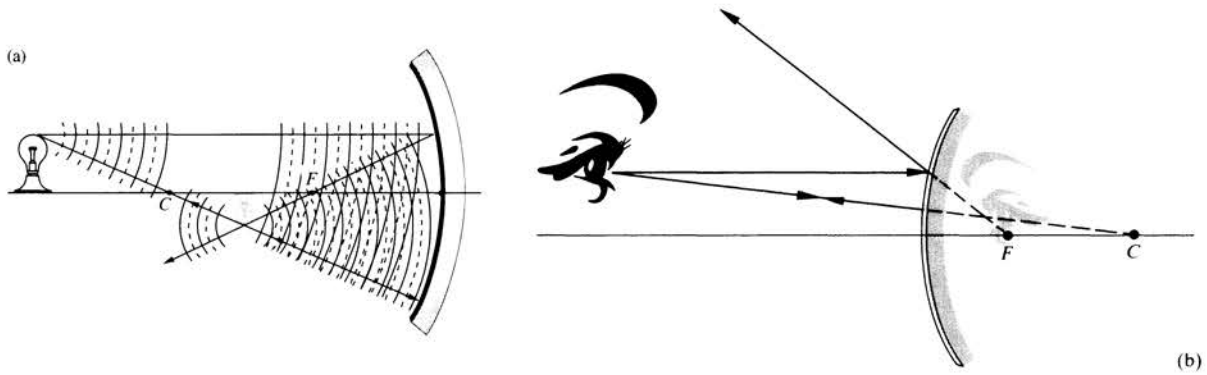


Figure 5.54 (*a*) Reflection from a concave mirror. (*b*) Reflection from a convex mirror.

Notice that triangles S_1S_2V and P_1P_2V in Fig. 5.53*a* are similar, and hence their sides are proportional. Taking y_i to be negative, as we did before, since it's below the axis, $y_i/y_o = -s_i/s_o$, which is equal to M_T. This is the *transverse magnification*, just as it was for the lens [Eq. (5.25)].

The only equation that contains information about the structure of the optical element (n, R, etc.) is that for f, and so, understandably, it differs for the thin lens [Eq. (5.16)] and spherical mirror [Eq. (5.49)]. The other functional expressions that relate s_o, s_i, and f or y_o, y_i, and M_T are, however, precisely the same. The only alteration in the previous sign convention appears in Table 5.4, where s_i on the left of V is now taken as positive. The striking similarity between the properties of a concave mirror and a convex lens on one hand and a convex mirror and a concave lens on the other are quite evident from a comparison of Tables 5.3 and 5.5, which are identical in all respects.

Table 5.5 Images of Real Objects Formed by Spherical Mirrors

Concave

Object	Image			
Location	Type	Location	Orientation	Relative Size
$\infty > s_o > 2f$	Real	$f < s_i < 2f$	Inverted	Minified
$s_o = 2f$	Real	$s_i = 2f$	Inverted	Same size
$f < s_o < 2f$	Real	$\infty > s_i > 2f$	Inverted	Magnified
$s_o = f$		$\pm\infty$		
$s_o < f$	Virtual	$\lvert s_i\rvert > s_o$	Erect	Magnified

Convex

Object	Image			
Location	Type	Location	Orientation	Relative Size
Anywhere	Virtual	$\lvert s_i\rvert < \lvert f\rvert$, $s_o > \lvert s_i\rvert$	Erect	Minified

The properties summarized in Table 5.5 and depicted in Fig. 5.55 can easily be verified empirically. If you don't have a spherical mirror at hand, a fairly crude but functional one can be made by carefully shaping aluminum foil over a spherical form, such as the end of a lightbulb (in that particular case R and therefore f will be small). A rather nice qualitative experiment involves examining the image of some small object formed by a short focal-length concave mirror. As you move it toward the mirror from beyond a distance of $2f = R$, the image will gradually increase, until at $s_o = 2f$ it will appear inverted and life-size. Bringing it closer will cause the image to increase even more, until it fills the entire mirror with an unrecognizable blur. As s_o becomes smaller, the now erect, magnified image will continue to decrease until the object finally rests on the mirror, where the image is again life-size. If you are not moved by all of this to jump up and make a mirror, you might try examining the image formed by a shiny spoon—either side will be interesting.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

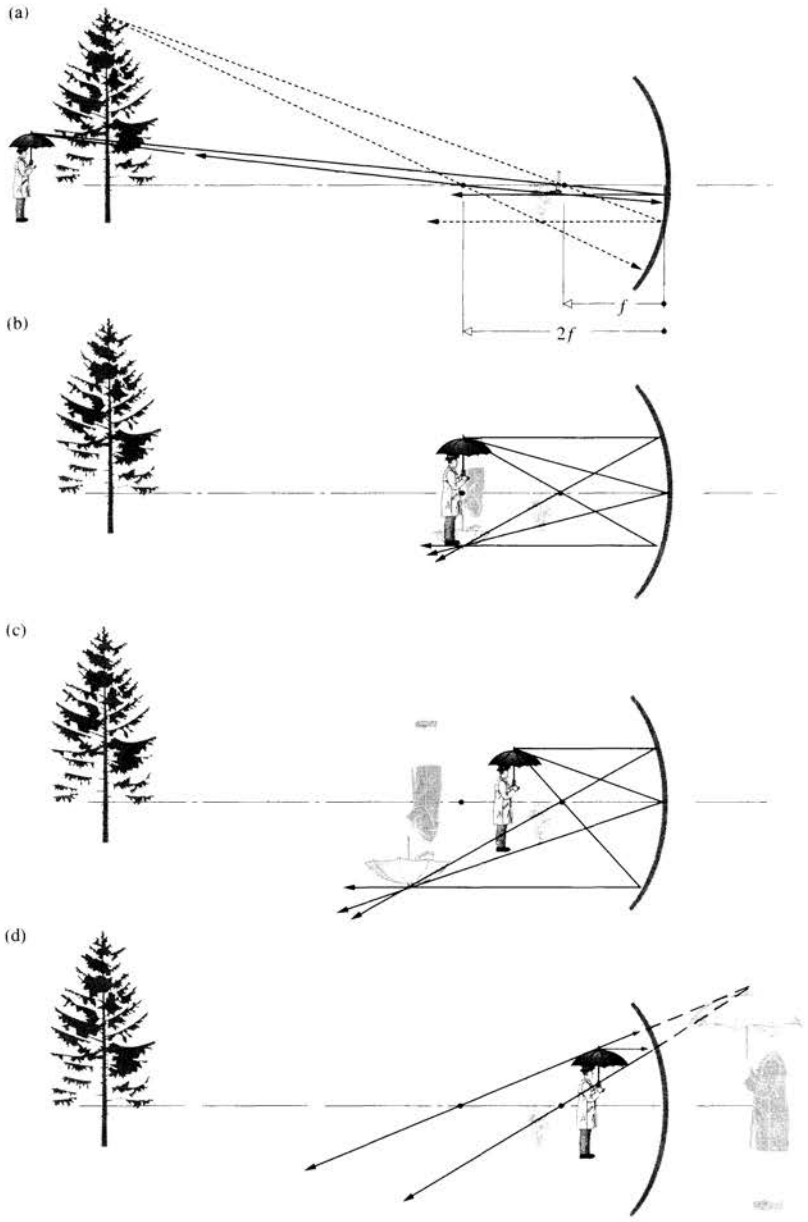


Figure 5.55 The image-forming behavior of a concave spherical mirror.

5.5 Prisms

Prisms play many different roles in Optics; there are prism combinations that serve as beamsplitters (p. 127), polarizing devices (see Section 8.4.3), and interferometers. Despite this diversity, the vast majority of applications make use of only one of two main prism functions. First, a prism can serve as a dispersive device, as it does in a variety of spectrum analyzers (p. 187). As such it is capable of separating, to some extent,

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

the constituent frequency components in a polychromatic light beam. Recall that the term *dispersion* was introduced earlier (p. 188) in connection with the frequency dependence of the index of refraction, $n(\omega)$, for dielectrics. In fact, the prism provides a highly useful means of measuring $n(\omega)$ over a wide range of frequencies and for a variety of materials (including gases and liquids).

Its second and more common function is to effect a change in the orientation of an image or in the direction of propagation of a beam. Prisms are incorporated in many optical instruments, often simply to fold the system into a confined space. There are inversion prisms, reversion prisms, and prisms that deviate a beam without inversion or reversion—and all of this without dispersion.

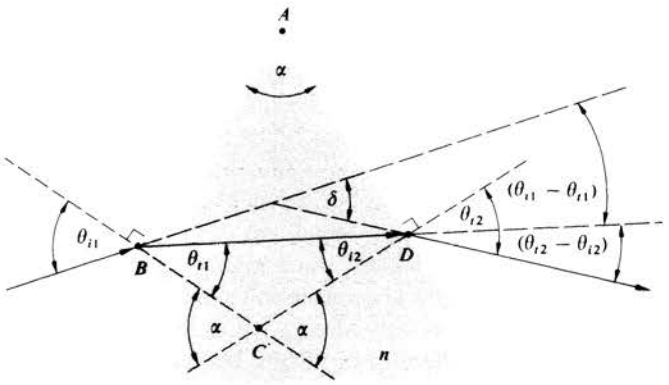


Figure 5.56 Geometry of a dispersing prism.

5.5.1 Dispersing Prisms

Prisms come in many sizes and shapes and perform a variety of functions (see photo). Let's first consider the group known as **dispersing prisms**. Typically, a ray entering a dispersing prism, as in Fig. 5.56, will emerge having been deflected from its original direction by an angle δ known as the **angular deviation**. At the first refraction the ray is deviated through an angle $(\theta_{i1} - \theta_{t1})$, and at the second refraction it is further deflected through $(\theta_{t2} - \theta_{i2})$. The total deviation is then

$$\delta = (\theta_{i1} - \theta_{t1}) + (\theta_{t2} - \theta_{i2})$$

Since the polygon $ABCD$ contains two right angles, $\measuredangle BCD$ must be the supplement of the **apex angle** α. As the exterior angle to triangle BCD, α is also the sum of the alternate interior angles, that is,

$$\alpha = \theta_{t1} + \theta_{i2} \tag{5.51}$$

Thus

$$\delta = \theta_{i1} + \theta_{t2} - \alpha \tag{5.52}$$

We would like to write δ as a function of both the angle-of-incidence for the ray (i.e., θ_{i1}) and the prism angle α; these presumably would be known. If the prism index is n and it's immersed in air ($n_a \approx 1$), it follows from Snell's Law that

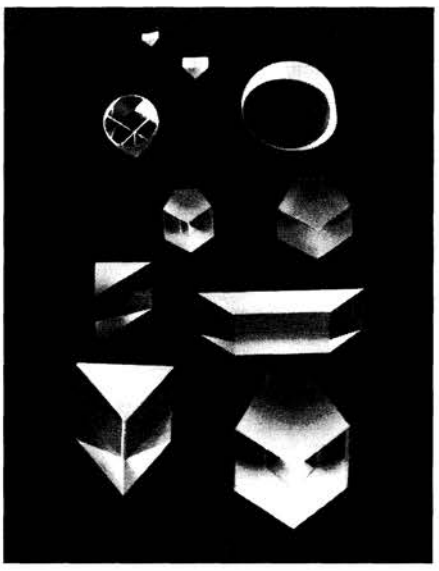
$$\theta_{t2} = \sin^{-1}(n \sin \theta_{i2}) = \sin^{-1}[n \sin(\alpha - \theta_{t1})]$$

Upon expanding this expression, replacing $\cos \theta_{t1}$ by $(1 - \sin^2 \theta_{t1})^{1/2}$, and using Snell's Law we have

$$\theta_{t2} = \sin^{-1}[(\sin \alpha)(n^2 - \sin^2 \theta_{i1})^{1/2} - \sin \theta_{i1} \cos \alpha]$$

The deviation is then

$$\delta = \theta_{i1} + \sin^{-1}[(\sin \alpha)(n^2 - \sin^2 \theta_{i1})^{1/2} - \sin \theta_{i1} \cos \alpha] - \alpha \tag{5.53}$$



Prisms. (Photo courtesy Melles Griot.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Apparently, δ increases with n, which is itself a function of frequency, so we might designate the deviation as $\delta(\nu)$ or $\delta(\lambda)$. For most transparent dielectrics of practical concern, $n(\lambda)$ decreases as the wavelength increases across the visible [refer back to Fig. 3.41 for a plot of $n(\lambda)$ versus λ for various glasses]. Clearly, then, $\delta(\lambda)$ will be less for red light than it is for blue.

Missionary reports from Asia in the early 1600s indicated that prisms were well known and highly valued in China because of their ability to generate color. A number of scientists of the era, particularly Marci, Grimaldi, and Boyle, had made some observations using prisms, but it remained for the great Sir Isaac Newton to perform the first definitive studies of dispersion. On February 6, 1672, Newton presented a classic paper to the Royal Society entitled "A New Theory about Light and Colours." He had concluded that white light consisted of a mixture of various colors and that the process of refraction was color-dependent.

Returning to Eq. (5.53), it's evident that the deviation suffered by a monochromatic beam on traversing a given prism (i.e., n and α are fixed) is a function only of the incident angle at the first face, θ_{i1}. A plot of the results of Eq. (5.53) as applied to a typical glass prism is shown in Fig. 5.57. The smallest value of δ is known as the **minimum deviation**, δ_m, and it is of particular interest for practical reasons. The value of δ_m can be determined analytically by differentiating Eq. (5.53) and then setting $d\delta/d\theta_{i1} = 0$, but a more indirect route will certainly be simpler. Differentiating Eq. (5.52) and setting it equal to zero yields

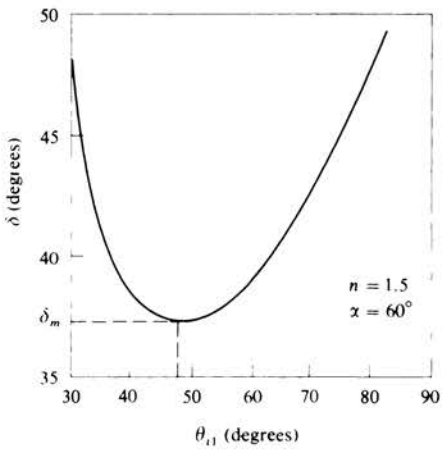


Figure 5.57 Deviation versus incident angle.

$$\frac{d\delta}{d\theta_{i1}} = 1 + \frac{d\theta_{t2}}{d\theta_{i1}} = 0$$

or $d\theta_{t2}/d\theta_{i1} = -1$. Taking the derivative of Snell's Law at each interface, we get

$$\cos\theta_{i1}\, d\theta_{i1} = n \cos\theta_{t1}\, d\theta_{t1}$$

and

$$\cos\theta_{t2}\, d\theta_{t2} = n \cos\theta_{i2}\, d\theta_{i2}$$

Note as well, on differentiating Eq. (5.51), that $d\theta_{t1} = -d\theta_{i2}$, since $d\alpha = 0$. Dividing the last two equations and substituting for the derivatives leads to

$$\frac{\cos\theta_{i1}}{\cos\theta_{t2}} = \frac{\cos\theta_{t1}}{\cos\theta_{i2}}$$

Making use of Snell's Law once again, we can rewrite this as

$$\frac{1 - \sin^2\theta_{t1}}{1 - \sin^2\theta_{t2}} = \frac{n^2 - \sin^2\theta_{i1}}{n^2 - \sin^2\theta_{t2}}$$

The value of θ_{i1} for which this is true is the one for which $d\delta/d\theta_{i1} = 0$. Inasmuch as $n \neq 1$, it follows that

$$\theta_{i1} = \theta_{t2}$$

and therefore

$$\theta_{t1} = \theta_{i2}$$

This means that *the ray for which the deviation is a minimum traverses the prism symmetrically, that is, parallel to its base.* Incidentally, there is a lovely argument for why θ_{i1} must equal θ_{t2}, which is neither as mathematical nor as tedious as the one we have evolved. In brief, suppose a ray undergoes a minimum deviation and $\theta_{i1} \neq \theta_{t2}$. Then if we reverse the ray, it will retrace the same path, so δ must be unchanged (i.e., $\delta = \delta_m$). But this implies that there are two different incident angles for which the deviation is a minimum, and this we know is not true—ergo $\theta_{i1} = \theta_{t2}$.

In the case when $\delta = \delta_m$, it follows from Eqs. (5.51) and (5.52) that $\theta_{i1} = (\delta_m + \alpha)/2$ and $\theta_{t1} = \alpha/2$, whereupon Snell's Law at the first interface leads to

$$n = \frac{\sin[(\delta_m + \alpha)/2]}{\sin\alpha/2} \tag{5.54}$$

This equation forms the basis of one of the most accurate techniques for determining the refractive index of a transparent substance. Effectively, one fashions a prism out of the materi-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

al in question, and then, measuring α and $\delta_m(\lambda)$, $n(\lambda)$ is computed employing Eq. (5.54) at each wavelength of interest. Hollow prisms whose sides are fabricated of plane-parallel glass can be filled with liquids or gases under high pressure; the glass plates will not result in any deviation of their own.

Figures 5.58 and 5.59 show two examples of **constant-deviation dispersing prisms**, which are important primarily in spectroscopy. The **Pellin–Broca** prism is probably the most common of the group. Albeit a single block of glass, it can be envisaged as consisting of two 30°–60°–90° prisms and one 45°–45°–90° prism. Suppose that in the position shown a single monochromatic ray of wavelength λ traverses the component prism *DAE* symmetrically, thereafter to be reflected at 45° from face *AB*. The ray will then traverse prism *CDB* symmetrically, having experienced a total deviation of 90°. The ray can be thought of as having passed through an ordinary 60° prism (*DAE* combined with *CDB*) at minimum deviation. All other wavelengths present in the beam will emerge at other angles. If the prism is now rotated slightly about an axis normal to the paper, the incoming beam will have a new incident angle. A different wavelength component, say λ_2, will now undergo a minimum deviation, which is again 90°—hence the name *constant deviation*. With a prism of this sort, one can conveniently set up the light source and viewing system at a fixed angle (here 90°), and then simply rotate the prism to look at a particular wavelength. The device can be calibrated so that the prism-rotating dial reads directly in wavelength.

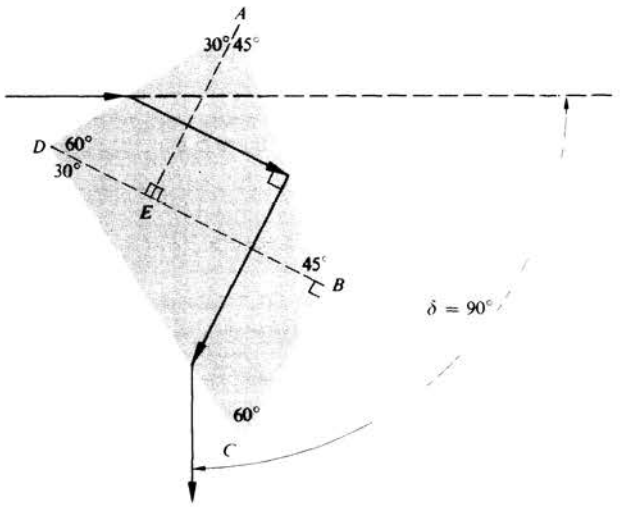


Figure 5.58 The Pellin–Broca prism.

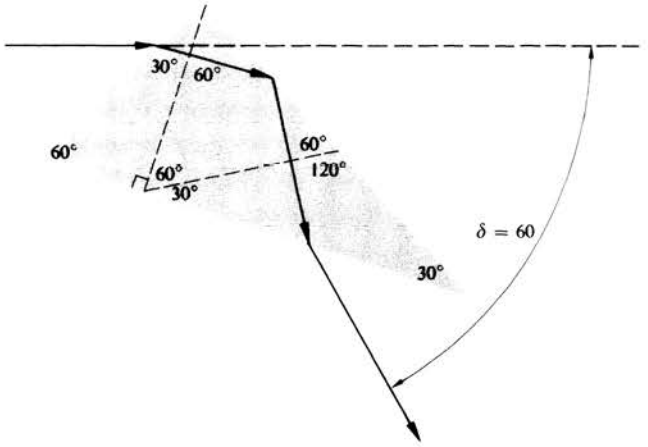


Figure 5.59 The Abbe prism.

5.5.2 Reflecting Prisms

We now examine **reflecting prisms**, in which dispersion is not desirable. In this case, the beam is introduced in such a way that at least one internal reflection takes place, for the specific purpose of changing either the direction of propagation or the orientation of the image, or both.

Let's first establish that it is actually possible to have such an internal reflection without dispersion. In other words, is δ independent of λ? The prism in Fig. 5.60 is assumed to have as its profile an isosceles triangle—this happens to be a rather common configuration in any event. The ray refracted at the first interface is later reflected from face *FG*. As we saw earlier (Section 4.7), this will occur when the internal incident angle is greater than the critical angle θ_c, defined by

$$\sin \theta_c = n_{ti} \qquad [4.69]$$

For a glass–air interface, this requires that θ_i be greater than roughly 42°. To avoid any difficulties at smaller angles, let's further suppose that the base of our hypothetical prism is silvered as well—certain prisms do in fact require silvered faces. The angle of deviation between the incoming and outgoing rays is

$$\delta = 180° - \measuredangle BED \qquad (5.55)$$

From the polygon *ABED* it follows that

$$\alpha + \measuredangle ADE + \measuredangle BED + \measuredangle ABE = 360°$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

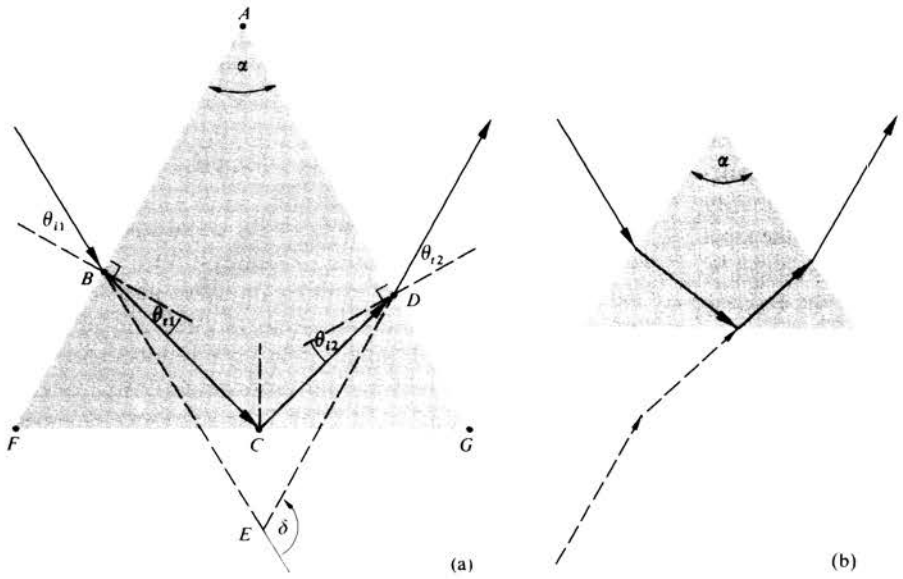


Figure 5.60 Geometry of a reflecting prism.

Moreover, at the two refracting surfaces

$$\measuredangle ABE = 90° + \theta_{i1}$$

and

$$\measuredangle ADE = 90° + \theta_{t2}$$

Substituting for $\measuredangle BED$ in Eq. (5.55) leads to

$$\delta = \theta_{i1} + \theta_{t2} + \alpha \tag{5.56}$$

Since the ray at point C has equal angles-of-incidence and reflection, $\measuredangle BCF = \measuredangle DCG$. Thus, because the prism is isosceles, $\measuredangle BFC = \measuredangle DGC$, and triangles FBC and DGC are similar. It follows that $\measuredangle FBC = \measuredangle CDG$, and therefore $\theta_{t1} = \theta_{i2}$. From Snell's Law we know that this is equivalent to $\theta_{i1} = \theta_{t2}$, whereupon the deviation becomes

$$\delta = 2\theta_{i1} + \alpha \tag{5.57}$$

which is certainly independent of both λ and n. The reflection will occur without any color preferences, and the prism is said to be **achromatic**. Unfolding the prism, that is, drawing its image in the reflecting surface FG, as in Fig. 5.60*b*, we see that it is equivalent in a sense to a parallelepiped or thick planar plate. The image of the incident ray emerges parallel to itself, regardless of wavelength.

A few of the many widely used reflecting prisms are shown in the next several figures. These are often made from BSC-2 or C-1 glass (see Table 6.2). For the most part, the illustrations are self-explanatory, so the descriptive commentary will be brief.

The **right-angle prism** (Fig. 5.61) deviates rays normal to the incident face by 90°. Notice that the top and bottom of the image have been interchanged; that is, the arrow has been flipped over, but the right and left sides have not. It is therefore an inversion system with the top face acting like a plane mirror. (To see this, imagine that the arrow and lollypop are vec-

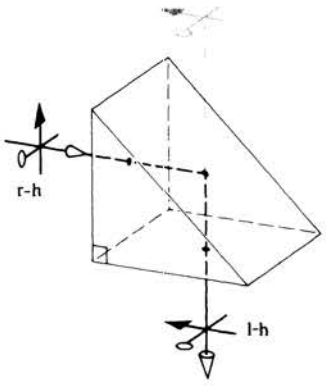


Figure 5.61 The right-angle prism.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

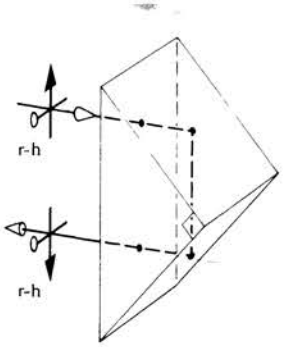


Figure 5.62 The Porro prism.

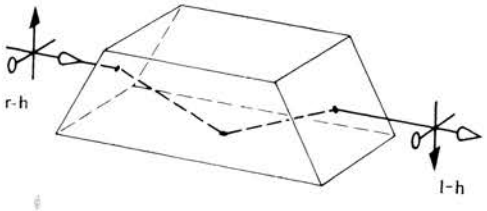


Figure 5.63 The Dove prism.

tors and take their cross-product. The resultant, arrow × lollypop, was initially in the propagation direction but is reversed by the prism.)

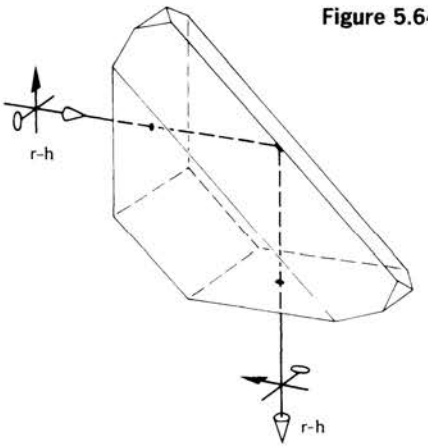
The **Porro prism** (Fig. 5.62) is physically the same as the right-angle prism but is used in a different orientation. After two reflections, the beam is deviated by 180°. Thus, if it enters right-handed, it leaves right-handed.

The **Dove prism** (Fig. 5.63) is a truncated version (to reduce size and weight) of the right-angle prism, used almost exclusively in collimated light. It has the interesting property (Problem 5.70) of rotating the image twice as fast as it is itself rotated about the longitudinal axis.

The **Amici prism** (Fig. 5.64) is essentially a truncated right-angle prism with a roof section added on to the hypotenuse face. In its most common use, it has the effect of splitting the image down the middle and interchanging the right and left portions.* These prisms are expensive, because the 90° roof angle must be held to roughly 3 or 4 seconds of arc, or a troublesome double image will result. They are often used in simple telescope systems to correct for the reversion introduced by the lenses.

The **penta prism** (Fig. 5.65) will deviate the beam by 90° without affecting the orientation of the image. Note that two of its surfaces must be silvered. These prisms are often used as end reflectors in small range finders.

Figure 5.64 The Amici prism.



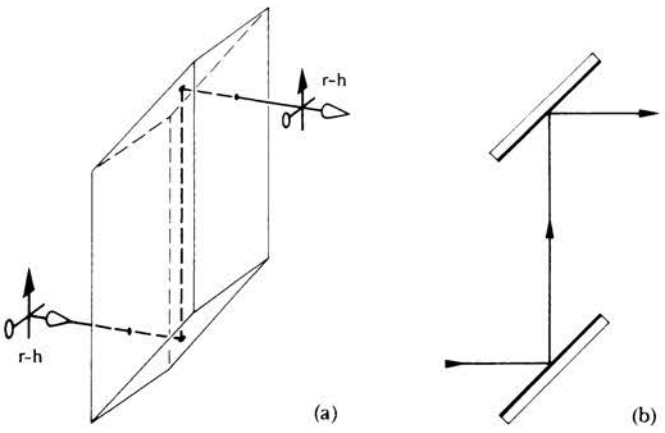


Figure 5.65 The rhomboid prism and its mirror equivalent.

*You can see how it actually works by placing two plane mirrors at right angles and looking directly into the combination. If you wink your *right* eye, the image will wink its *right* eye. Incidentally, if your eyes are equally strong, you will see two seams (images of the line where the mirrors meet), one running down the middle of each eye, with your nose presumably between them. If one eye is stronger, there will be only one seam, down the middle of that eye. If you close it, the seam will jump over to the other eye. This must be tried to be appreciated.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

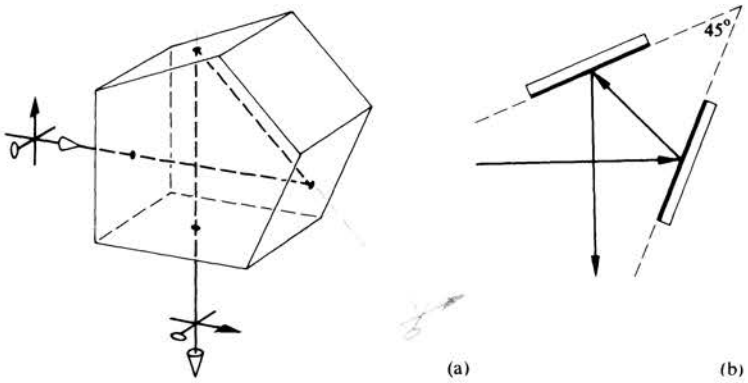


Figure 5.66 The penta prism and its mirror equivalent.

The **rhomboid prism** (Fig. 5.66) displaces the line-of-sight without producing any angular deviation or changes in the orientation of the image.

The **Leman–Springer prism** (Fig. 5.67) also has a 90° roof. Here the line-of-sight is displaced without being deviated, but the emerging image is right-handed and rotated through 180°. The prism can therefore serve to erect images in telescope systems, such as gun sights and the like.

Many more reflecting prisms perform specific functions. For example, if one cuts a cube so that the piece removed has three mutually perpendicular faces, it is called a **corner-cube prism**. It has the property of being retrodirective; that is, it

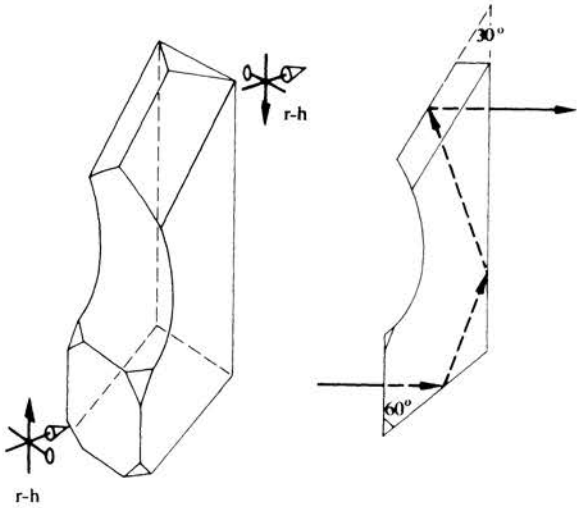


Figure 5.67 The Leman–Springer prism.

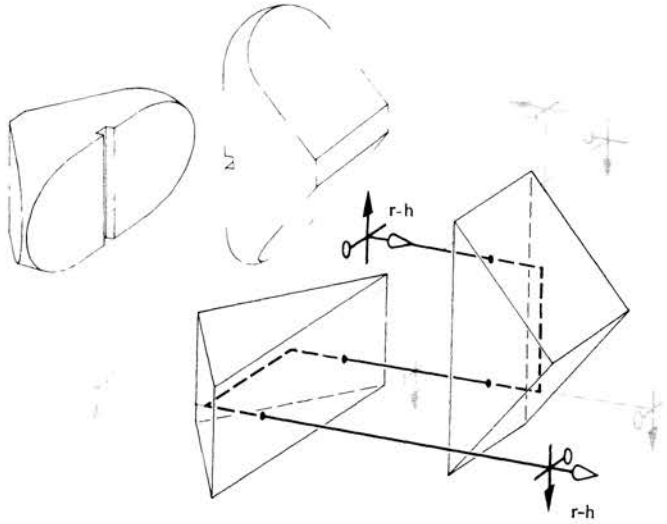


Figure 5.68 The double Porro prism.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

will reflect all incoming rays back along their original directions. One hundred of these prisms are sitting in an 18-inch square array 240 000 miles from here, having been placed on the Moon during the Apollo 11 flight.*

The most common erecting system consists of two Porro prisms, as illustrated in Fig. 5.68. These are relatively easy to manufacture and are shown here with rounded corners to reduce weight and size. Since there are four reflections, the exiting image will be right-handed. A small slot is often cut in the hypotenuse face to obstruct rays that are internally reflected at glancing angles. Finding these slots after dismantling the family's binoculars is often an inexplicable surprise.

5.6 Fiberoptics

The concept of channeling light within a long narrow dielectric (via total internal reflection) has been around for quite a while. John Tyndall (1870) showed that light could be contained within and guided along a thin stream of water. Soon after that, glass "light pipes" and, later, threads of fused quartz were used to further demonstrate the effect. But it wasn't until the early 1950s that serious work was done to transport images along bundles of short glass fibers.

After the advent of the laser (1960), there was an immediate appreciation of the potential benefits of sending information from one place to another using light, as opposed to electric currents or even microwaves. At those high optical frequencies (of the order of 10^{15} Hz), one hundred thousand times more information can be carried than with microwaves. Theoretically, that's the equivalent of sending tens of millions of television programs all at once on a beam of light. It wasn't long (1966) before the possibility of coupling lasers with fiberoptics for long-distance communications was pointed out. Thus began a tremendous technological transformation that's still roaring along today.

In 1970 researchers at the Corning Glass Works produced a silica fiber with a signal-power transmission of better than 1% over a distance of 1 km (i.e., an attenuation of 20 dB/km), which was comparable to existing copper electrical systems. During the next two decades, the transmission rose to about 96% over 1 km (i.e., an attenuation of only 0.16 dB/km).

Because of its low-loss transmission, high-information-carrying capacity, small size and weight, immunity to electromagnetic interference, unparalleled signal security, and the abundant availability of the required raw materials (i.e., ordinary sand), ultrapure glass fibers have become the premier communications medium.

As long as the diameter of these fibers is large compared with the wavelength of the radiant energy, the inherent wave nature of the propagation is of little importance, and the process obeys the familiar laws of Geometrical Optics. On the other hand, if the diameter is of the order of λ, the transmission closely resembles the manner in which microwaves advance along waveguides. Some of the propagation modes are evident in the photomicrographic end views of fibers shown in Fig. 5.69. Here the wave nature of light must be reckoned with, and this behavior resides in the domain of Physical Optics. Although optical waveguides, particularly of the thin-film variety, are of increasing interest, this discussion will be limited to the case of relatively large diameter fibers, those about the thickness of a human hair.

Consider the straight glass cylinder of Fig. 5.70 surrounded by an incident medium of index n_i—let it be air, $n_i = n_a$. Light striking its walls from within will be totally internally reflected, provided that the incident angle at each reflection is greater than $\theta_c = \sin^{-1} n_a/n_f$, where n_f is the index of the cylinder or fiber. As we will show, a *meridional ray* (i.e., one that is coplanar with the central or optical axis) might undergo several thousand reflections per foot as it bounces back and forth

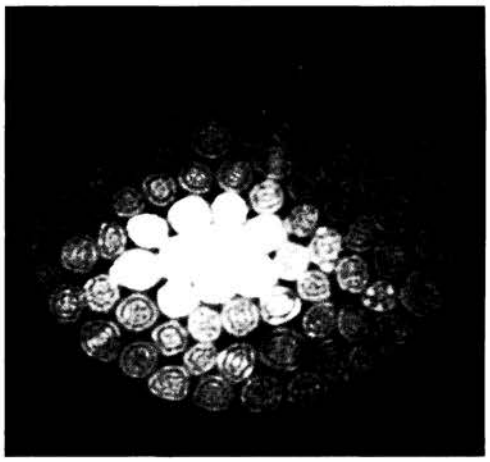


Figure 5.69 Optical waveguide mode patterns seen in the end faces of small-diameter fibers. (Photo courtesy of Narinder S. Kapany, AMP Fellow)

*J. E. Foller and E. J. Wampler, "The Lunar Laser Reflector." *Sci. Am.*, March 1970, p. 38.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

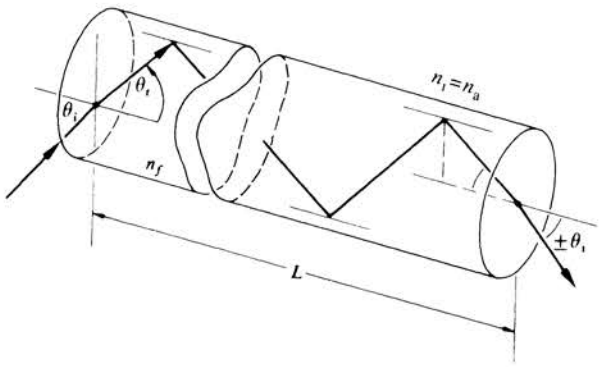


Figure 5.70 Rays reflected within a dielectric cylinder.

along a fiber, until it emerges at the far end (see photo). If the fiber has a diameter D and a length L, the path length ℓ traversed by the ray will be

$$\ell = L/\cos\,\theta_t \tag{5.58}$$

or from Snell's Law

$$\ell = n_f L(n_f^2 - \sin^2\,\theta_i)^{-1/2} \tag{5.59}$$

The number of reflections N is then given by

$$N = \frac{\ell}{D/\sin\,\theta_t} \pm 1$$

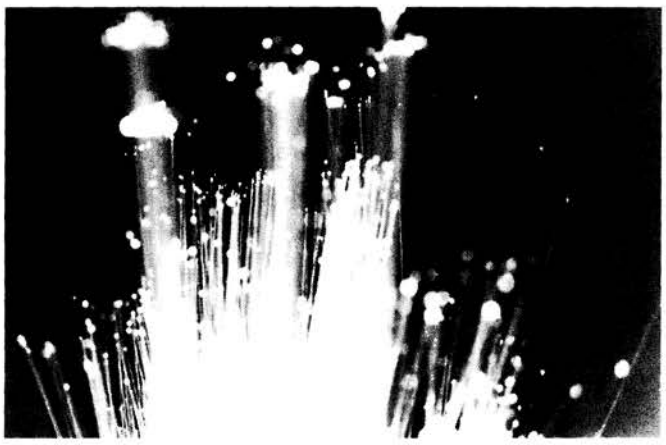
or

$$N = \frac{L\sin\,\theta_i}{D(n_f^2 - \sin^2\,\theta_i)^{1/2}} \pm 1 \tag{5.60}$$

rounded off to the nearest whole number. The ± 1, which depends on where the ray strikes the end face, is of no significance when N is large, as it is in practice. Thus, if D is 50 μm (i.e., 50 microns where 1 μm = 10^{-6} m = 39.37×10^{-6} in), which is about 2×10^{-3} in (a hair from the head of a human is roughly 50 μm in diameter), and if $n_f = 1.6$ and $\theta_i = 30°$, N turns out to be approximately 2000 reflections per foot. Fibers are available in diameters as small as 2 μm or so but are seldom used in sizes much less than about 10 μm. Extremely thin glass (or plastic) filaments are quite flexible and can even be woven into fabric.

The smooth surface of a single fiber must be kept clean (of moisture, dust, oil, etc.), if there is to be no leakage of light (via frustrated total internal reflection). Similarly, if large numbers of fibers are packed in close proximity, light may leak from one fiber to another in what is known as *cross-talk*. For these reasons, it is customary to enshroud each fiber in a transparent sheath of lower index called a **cladding**. This layer need only be thick enough to provide the desired isolation, but for other reasons it generally occupies about one-tenth of the cross-sectional area. Although references in the literature to simple light pipes go back 100 years, the modern era of fiberoptics began with the introduction of clad fibers in 1953.

Typically, a fiber core might have an index (n_f) of 1.62, and the cladding an index (n_c) of 1.52, although a range of values is available. A clad fiber is shown in Fig. 5.71. Notice that there is a maximum value θ_{max} of θ_i, for which the internal ray will impinge at the critical angle, θ_c. Rays incident on the face at angles greater than θ_{max} will strike the interior wall at angles less than θ_c. They will be only partially reflected at each such



Light emerging from the ends of a loose bundle of glass fibers.
(Photo by E.H.)

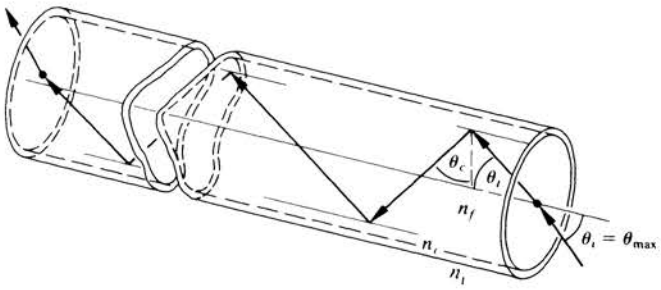


Figure 5.71 Rays in a clad optical fiber.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

encounter with the core–cladding interface and will quickly leak out of the fiber. Accordingly, θ_{max}, which is known as the acceptance angle, defines the half-angle of the acceptance cone of the fiber. To determine it, start with

$$\sin \theta_c = n_c/n_f = \sin (90° - \theta_t)$$

Thus
$$n_c/n_f = \cos \theta_t \tag{5.61}$$

or
$$n_c/n_f = (1 - \sin^2 \theta_t)^{1/2}$$

Making use of Snell's Law and rearranging terms, we have

$$\sin \theta_{max} = \frac{1}{n_i}(n_f^2 - n_c^2)^{1/2} \tag{5.62}$$

The quantity $n_i \sin \theta_{max}$ is defined as the **numerical aperture**, or NA. Its square is a measure of the light-gathering power of the system. The term originates in microscopy, where the equivalent expression describes the corresponding capabilities of the objective lens. It should clearly relate to the *speed* of the system, and, in fact,

$$f/\# = \frac{1}{2(\text{NA})} \tag{5.63}$$

Thus for a fiber

$$\text{NA} = (n_f^2 - n_c^2)^{1/2} \tag{5.64}$$

The left-hand side of Eq. (5.62) cannot exceed 1, and in air ($n_a = 1.000\,28 \approx 1$) that means that the largest value of NA is 1. In this case, the half-angle θ_{max} equals 90°, and the fiber totally internally reflects all light entering its face (Problem 5.71). Fibers with a wide variety of numerical apertures, from about 0.2 up to and including 1.0, are commercially obtainable.

Bundles of free fibers whose ends are bound together (e.g., with epoxy), ground, and polished form flexible light guides. If no attempt is made to align the fibers in an ordered array, they form an *incoherent bundle*. This unfortunate use of the term *incoherent* (which should not be confused with coherence theory) just means, for example, that the first fiber in the top row at the entrance face may have its terminus anywhere in the bundle at the exit face. These *flexible light carriers* are, for that reason, relatively easy to make and inexpensive. Their primary function is simply to conduct light from one region to another. Conversely, when the fibers are carefully arranged so that their terminations occupy the same relative positions in both of the bound ends of the bundle, it is said to be *coherent*. Such an arrangement is capable of transmitting images and is consequently known as a *flexible image carrier*.

Coherent bundles are frequently fashioned by winding fibers on a drum to make ribbons, which are then carefully layered. When one end of such a device is placed face down flat on an illuminated surface, a point-by-point image of whatever is beneath it will appear at the other end (see photo). These bundles can be tipped off with a small lens, so that they need not be in contact with the object under examina-tion. Nowadays it is common to use fiberoptic instruments to poke into all sorts of unlikely places, from nuclear reactor cores and jet engines to stomachs and reproductive organs. When a device is used to examine internal body cavities, it's called an *endoscope*. This category includes bronchoscopes, colonoscopes, gastroscopes, and so forth, all of which are generally less than about 200 cm in length. Similar industrial instruments are usually two or three times as long and often contain from 5000 to 50 000 fibers, depending on the required image resolution and the overall diameter that can be accommodated. An additional incoherent bundle incorporated into the device usually supplies the illumination.



A coherent bundle of 10 μm glass fibers transmitting an image even though knotted and sharply bent. (Photo courtesy of American ACMI Div., American Hospital Supply Corp.)

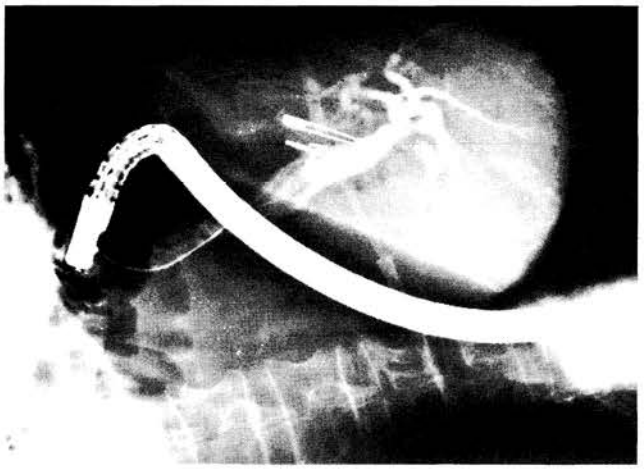
Not all fiberoptic arrays are made flexible; for example, fused, rigid, coherent fiber faceplates, or mosaics, are used to replace homogeneous low-resolution sheet glass on cathode-ray tubes, vidicons, image intensifiers, and other devices. Mosaics consisting of literally millions of fibers with their claddings fused together have mechanical properties almost identical to homo-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

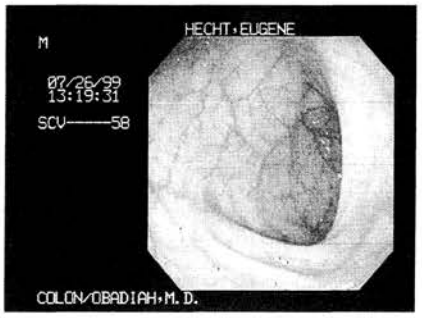
geneous glass. Similarly, a sheet of fused tapered fibers can either magnify or minify an image, depending on whether the light enters the smaller or larger end of the fiber. The compound eye of an insect such as the housefly is effectively a bundle of tapered fiberoptical filaments. The rods and cones that make up the human retina may also channel light through total internal reflection. Another common application of mosaics involving imaging is the field flattener. If the image formed by a lens system resides on a curved surface, it is often desirable to reshape it into a plane, for example, to match a film plate. A mosaic can be ground and polished on one of its end surfaces to correspond to the contour of the image and on the other to match the detector. Incidentally, a naturally occurring fibrous crystal known as ulexite, when polished, responds surprisingly like a fiberoptic mosaic. (Hobby shops often sell it for use in making jewelry.)

If you have never seen the kind of light conduction we've been talking about, try looking down the edges of a stack of microscope slides. Even better are the much thinner (0.18-mm) cover-glass slides (see photo).

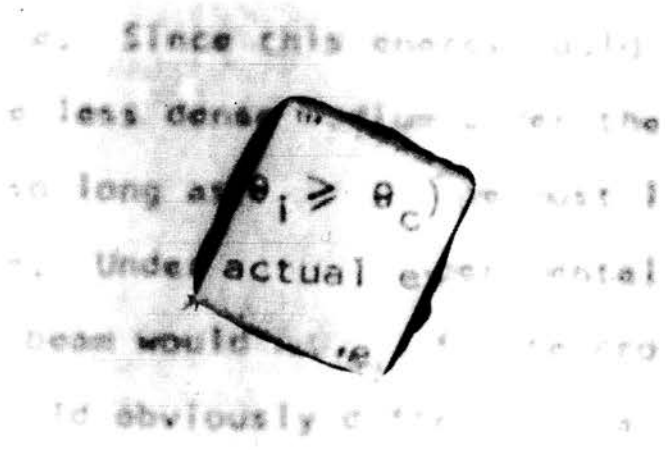
Today fiberoptics has three very different applications: it is used for the direct (short-distance) transmission of images and illumination, it provides a variety of remarkable waveguides used in telecommunications, and it serves as the core of a new family of sensors. Transmitting images over distances of a few meters with coherent bundles, however beautiful and however useful, is a rather unsophisticated business that doesn't start to utilize the full potential inherent in fiberoptics. The application of lightguides to telecommunications is rapidly replacing copper wires and electricity as the primary information pathway.



An X-ray showing a colonoscope being used to examine a patient for cancer of the colon.



A remarkably detailed view as seen through a fiberoptic colonoscope.



A stack of cover-glass slides held together by a rubber band serves as a coherent light guide. (Photo by E.H.)

Worldwide, in the first few decades after 1970 well over 100 million kilometers of fibers were installed. It's been estimated that today, every day, enough fiberoptic cable is installed to circle the Earth several times. And it's not unrealistic to anticipate that within a decade the data-carrying capacity of a single fiber will exceed one hundred trillion (i.e., in America, 100×10^{12}) bits per second (i.e., 100 Tb/s). In a different vein, fiberoptic sensors—devices that measure pressure, sound, temperature, voltage, current, liquid levels, electric and magnetic fields, rotations, and so forth—have become the latest manifestation of the versatility of fibers.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

5.6.1 Fiberoptic Communications Technology

The high frequencies of light allow for an incredible data-handling capacity. For example, with sophisticated transmitting techniques, a pair of copper telephone wires can be made to carry about two dozen simultaneous conversations. That should be compared with a single, ongoing, simple television transmission, which is equivalent to about 1300 simultaneous telephone conversations, and that, in turn, is roughly the equal of sending some 2500 typewritten pages each second. Clearly, at present it's quite impractical to attempt to send television over copper telephone lines. Yet by the mid-1980s it was already possible to transmit in excess of 12 000 simultaneous conversations over a *single pair* of fibers—that's more than nine television channels. Each such fiber has a line rate of about 400 million bits of information per second (400 Mb/s), or 6000 voice circuits. Fibers of this sort (with repeaters spaced every 40 km or so) formed the world's new intercity long-haul telecommunications grids. In the early 1990s researchers used **solitons**—carefully shaped pulses that travel without changing—to attain transmission rates of around 4 Gb/s. This is the equivalent of 70 simultaneous color TV channels sent more than a million kilometers.

The first fiberoptic transatlantic cable TAT-8 was designed, using some clever data-handling techniques, to carry 40 000 conversations at once over just two pairs of glass fibers. TAT-1, a copper cable installed in 1956, could carry a mere 51 conversations, and the last of the bulky copper versions, TAT-7 (1983), can handle only about 8000. The TAT-8, which began operations in 1988, functions at 296 Mb/s (using single-mode 1300-nm fibers—see p. 198). It has regenerators or repeaters (to boost the signal strength) every 50 km (30 mi) or more. This feature is tremendously important in long-distance communications. Ordinary wire systems require repeaters roughly every kilometer; electrical coaxial networks extend that range to about 2 to 6 km; even radio transmissions through the atmosphere need regeneration every 30 to 50 km. The repeaters used until the mid-1990s were electro-optical hybrids that converted the weakened optical signal into an electrical one, amplified it, and then, using semiconductor lasers, reintroduced it into the fiber.

A major determining factor in the spacing of repeaters is the power loss due to attenuation of the signal as it propagates down the line. The decibel (dB) is the customary unit used to designate the ratio of two power levels, and as such it can provide a convenient indication of the power-out (P_o) with respect to the power-in (P_i). The number of dB $= -10 \log (P_o/P_i)$, and hence a ratio of 1:10 is 10 dB, 1:100 is 20 dB, 1:1000 is 30 dB, and so on. The attenuation (α) is usually specified in decibels per kilometer (dB/km) of fiber length (L). Thus $-\alpha L/10 = \log (P_o/P_i)$, and if we raise 10 to the power of both sides,

$$P_o/P_i = 10^{-\alpha L/10} \tag{5.65}$$

As a rule, reamplification of the signal is necessary when the power has dropped by a factor of about 10^{-5}. Commercial optical glass, the kind of material available for fibers in the mid-1960s, has an attenuation of about 1000 dB/km. Light, after being transmitted 1 km through the stuff, would drop in power by a factor of 10^{-100}, and regenerators would be needed every 50 m (which is little better than communicating with a string and two tin cans). By 1970 α was down to about 20 dB/km for fused silica (quartz, SiO_2), and it was reduced to as little as 0.16 dB/km in 1982. This tremendous decrease in attenuation was achieved mostly by removing impurities (especially the ions of iron, nickel, and copper) and reducing contamination by OH groups, largely accomplished by scrupulously eliminating any traces of water in the glass (p. 72). Today the purest fibers can carry signals up to 80 km before needing reamplification.

By the beginning of the twenty-first century, two major advances had already begun to dramatically increase the data handling capacity of long-distance fiberoptic cables. The first innovation was the introduction of **erbium-doped fiber amplifiers** (EDFAs). These are single-mode fibers, which have ions of the rare-earth element erbium infused into their cores at levels of 100 to 1000 ppm. Having a good conversion efficiency, they're typically pumped at 980 nm (for the highest level of inversion) or 1480 nm (for the highest quantum efficiency) by diode lasers putting out around 200 mW. The resulting excited erbium atoms reradiate, via stimulated emission induced by photons from the faded signal, and thereby re-energize the flow of data. This happens along the entire length of the amplifier, and it can boost the power (usually held to milliwatt levels) of a wide range of frequencies simultaneously. Fiberoptic amplifiers eliminated the bottleneck caused by the previous generation of electronic hybrid repeaters.

The second innovation was the application of a new data-handling technique called **dense wavelength division multiplexing** (DWDM). The word "multiplexing" means the use of a single pathway to simultaneously transmit several signals which nonetheless retain their individuality. At the pres-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

ent it's not hard to send upwards of 160 optical channels carrying different signals, all transmitted at the same time over the same fiber at different frequencies. And it won't be long before 1000 channels per fiber is commonplace. Typically, each channel has a data rate of 10 Gb/s or more, and these are each spaced by 50 to 100 GHz. Every major telecommunications carrier is already using DWDM. The latest transatlantic cables contain four fiber pairs, each capable of carrying 48 DWDM channels and each of these flows data at a rate of 10 Gb/s. That's a net capacity of $4 \times 48 \times 10$ Gb/s or 1.9 Tb/s. Commercial links operating at 40 Gb/s per channel are already in service.

Figure 5.72 depicts the three major fiber configurations used in communications today. In (*a*) the core is relatively wide, and the indices of core and cladding are both constant throughout. This is the so-called **stepped-index fiber**, with a homogeneous core of roughly 50 to 200 μm and cladding typically 20 μm thick. The oldest of the three types, the stepped-index fiber was widely used in first-generation systems (1975–1980). The comparatively large central core makes it rugged and easily infused with light, as well as easily terminated and coupled. It's the least expensive but also the least effective of the lot, and for long-range applications, it has some serious drawbacks.

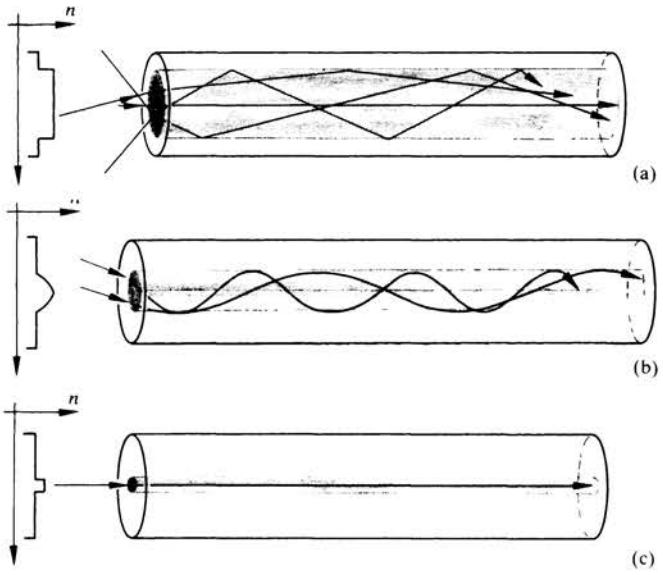


Figure 5.72 The three major fiberoptic configurations and their index profiles. (*a*) Multimode step-index fiber. (*b*) Multimode graded-index fiber. (*c*) Single-mode step-index fiber.

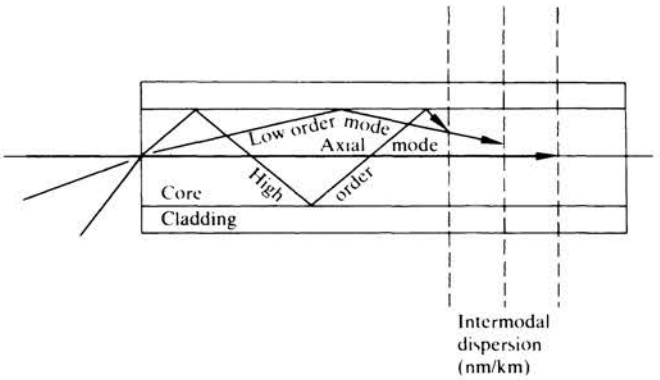


Figure 5.73 Intermodal dispersion in a stepped-index multimode fiber.

Depending on the launch angle into the fiber, there can be hundreds, even thousands, of different ray paths or modes by which energy can propagate down the core (Fig. 5.73). This then is a **multimode fiber**, wherein each mode corresponds to a slightly different transit time. Higher-angle rays travel longer paths; reflecting from side to side, they take longer to get to the end of the fiber than do rays moving along the axis. This is loosely spoken of as **intermodal dispersion** (or often just *modal dispersion*), even though it has nothing to do with a frequency-dependent index of refraction. Information to be transmitted is usually digitized in some coded fashion and then sent along the fibers as a flood of millions of pulses or bits per second. The different transit times have the undesirable effect of changing the shape of the pulses of light that represent the signal. What started as a sharp rectangular pulse can smear out, after traveling a few kilometers within the fiber, into an unrecognizable blur (Fig. 5.74).

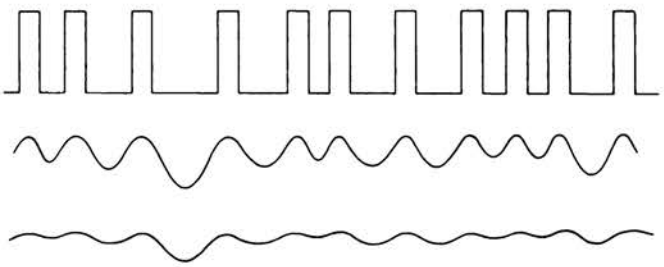


Figure 5.74 Rectangular pulses of light smeared out by increasing amounts of dispersion. Note how the closely spaced pulses degrade more quickly.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

The total time delay between the arrival of the axial ray and the slowest ray, the one traveling the longest distance, is $\Delta t = t_{max} - t_{min}$. Here, referring back to Fig. 5.71, the minimum time of travel is just the axial length L divided by the speed of light in the fiber:

$$t_{min} = \frac{L}{v_f} = \frac{L}{c/n_f} = \frac{Ln_f}{c} \tag{5.66}$$

The nonaxial route (ℓ), given by Eq. (5.58), is longest when the ray is incident at the critical angle, whereupon Eq. (5.61) holds. Combining these two, we get $\ell = Ln_f/n_c$, and so

$$t_{max} = \frac{\ell}{v_f} = \frac{Ln_f/n_c}{c/n_f} = \frac{Ln_f^2}{cn_c} \tag{5.67}$$

Thus it follows that, subtracting Eq. (5.66) from Eq. (5.67), we get

$$\Delta t = \frac{Ln_f}{c}\left(\frac{n_f}{n_c} - 1\right) \tag{5.68}$$

As an example, suppose $n_f = 1.500$ and $n_c = 1.489$. The delay, $\Delta t/L$, then turns out to be 37 ns/km. In other words, a sharp pulse of light entering the system will be spread out in time some 37 ns for each kilometer of fiber traversed. Moreover, traveling at a speed $v_f = c/n_f = 2.0 \times 10^8$ m/s, it will spread in space over a length of 7.4 m/km. To make sure that the transmitted signal will still be easily readable, we might require that the spatial (or temporal) separation be at least twice the spread-out width (Fig. 5.75). Now imagine the line to be 1.0 km long. In that case, the output pulses are 7.4 m wide on emerging from the fiber and so must be separated by 14.8 m. This means that the input pulses must be at least 14.8 m apart; they must be separated in time by 74 ns and so cannot come any faster than one every 74 ns, which is a rate of 13.5 million pulses per second. In this way the intermodal dispersion (which is typically 15 to 30 ns/km) limits the frequency of the input signal, thereby dictating the rate at which information can be fed through the system. Step-index multimode fibers are used for low-speed, short-distance lines.

These large-core fibers are used mainly in image transmission and illumination bundles. They're also useful for carrying high-power laserbeams where the energy is distributed over a larger volume, thereby avoiding damage to the fiber.

The problem of delay differences can be reduced as much as a hundredfold by gradually varying the refractive index of the core, decreasing it radially outward to the cladding (Fig. 5.72*b*). Instead of following sharp zigzag paths, the rays then smoothly spiral around the central axis. Because the index is higher along the center, rays taking shorter paths are slowed down by proportionately greater amounts, and rays spiralling around near the cladding move more swiftly over longer paths. The result is that all the rays tend to stay more or less together in these multimode **graded-index fibers**. Typically, a graded-index fiber has a core diameter of about 20 μm to 90 μm and an intermodal dispersion of only around 2 ns/km. They are intermediate in price and have been widely used in medium-distance intercity applications.

Multimode fibers with core diameters of 50 μm or more are often fed by *light-emitting diodes* (LEDs). These are comparatively inexpensive and are commonly used over relatively short spans at low transmission rates. The problem with them is that they emit a fairly broad range of frequencies. As a result, ordinary *material* or *spectral dispersion*, the fact that the fiber index is a function of frequency, becomes a limiting factor. That difficulty is essentially avoided by using spectrally pure laserbeams. Alternatively, the fibers can be operated at wavelengths near 1.3 μm, where silica glass (see Figs. 3.40 and 3.41) has little dispersion.

The last, and best, solution to the problem of intermodal dispersion is to make the core so narrow (less than 10 μm) that it will provide only one mode wherein the rays travel parallel to the central axis (Fig. 5.72*c*). Such **single-mode fibers** of

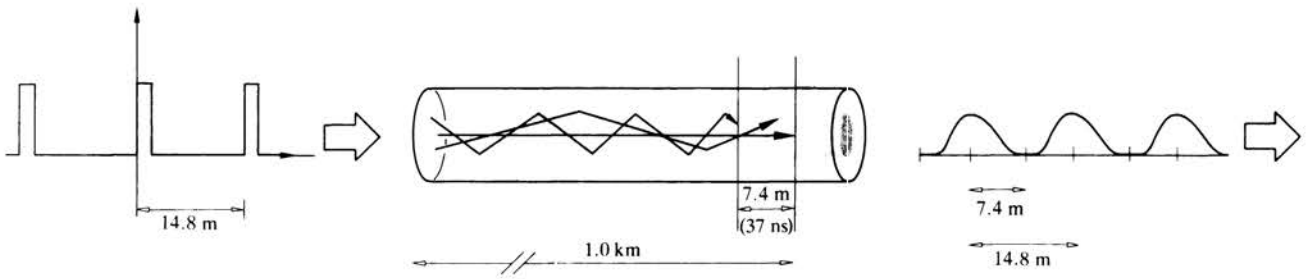
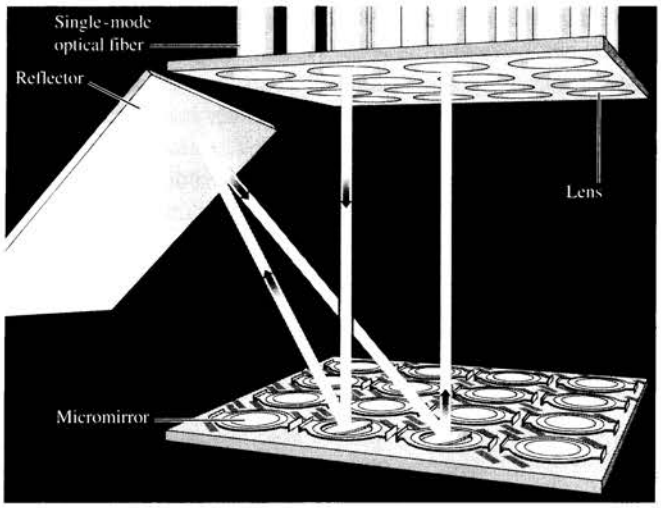
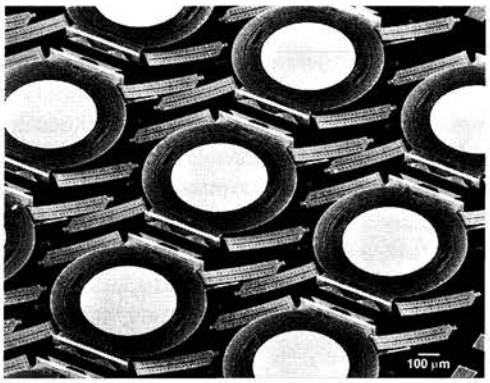


Figure 5.75 The spreading of an input signal due to intermodal dispersion.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(a)



(b)

Figure 5.76 (*a*) An optical switch that uses tiny steerable mirrors to redirect pulses of light. (*b*) The array of tiltable mirrors. (Photo courtesy Lucent Technologies' Bell Laboratories.)

ultrapure glass (both stepped-index and the newer graded-index) provide the best performance. Typically having core diameters of only 2 μm to 9 μm (around 10 wavelengths), they essentially eliminate intermodal dispersion. Although they are relatively expensive and require laser sources, these single-mode fibers operated at 1.55 μm (where the attenuation is about 0.2 dB/km, not far from the ideal silica value of 0.1 dB/km) are today's premier long-haul lightguides. A pair of such fibers may someday connect your home to a vast network of communications and computer facilities, making the era of the copper wire seem charmingly primitive.

Wandering through the Internet requires rapidly channeling vast amounts of data from one fiberoptic route to another. At the end of the twentieth century, this was accomplished at network hubs where pulses of light were converted into electrical signals that could subsequently be switched around electronically. Only then were the packets of data converted back into pulses of light to continue the journey. Unfortunately, electronic switches are bulky, expensive, and relatively slow—not up to the task of meeting future demands. Until very recently, there was little hope that this so-called electronic bottleneck would soon be alleviated. But things changed dramatically at the turn of the new millennium with the introduction of several photonic switching systems.

Figure 5.76 depicts an all-optical switch utilizing **MOEMS** (Micro-OptoElectroMechanical Systems) technology (p. 178). The end faces of hundreds of incoming and outgoing fibers are capped with tiny lenses at the top of the assembly. A downward pulse of photons enters, hits a micromirror (only 0.5 mm in diameter) whose orientation is electronically controlled, "bounces off" a large reflector, strikes another controllable micromirror, and emerges into a designated output fiber, all readjustable in a matter of milliseconds. **MOEMS** switches have already been deployed into the network to control data traffic. Eventually, optical switches will support the petabit-per-second, Pb/s (that's American quadrillion, 10^{15}) telecommunications system that's not far off in the future. Beyond that is the all-optical worldwide Telephone-TV-Internet purring along at rates as yet unimagined.

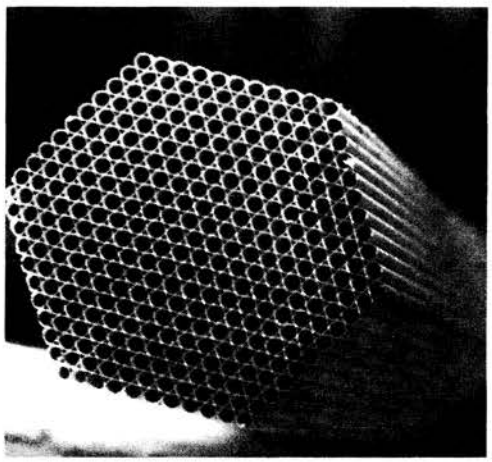
Capillary Optics

Fiberoptics works by having radiant energy (of a relatively low frequency, namely, light or IR) totally internally reflect off a high-index/low-index interface within a narrow solid waveguide. Similarly, high-frequency EM-radiation (especially X-rays) can also be totally internally reflected (p. 122) off an air–glass interface (rather than a glass–air interface). The critical angle, *measured up from the surface*, is typically only about 0.2° for 10 keV (≈0.12 nm) X-rays. Figure 5.77 shows



Figure 5.77 Multiple grazing-incidence reflections of X-rays within a hollow glass fiber.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



A scanning electron micrograph of a single multichannel thread containing hundreds of hollow capillaries. (Photo courtesy X-Ray Optical Systems, Inc. Albany, NY, USA)

how a beam follows the curve of a hollow capillary tube via multiple grazing-incidence reflection at the internal air–glass interface. Bending the path of X-rays is otherwise a daunting business.

A single glass thread with a diameter of 300 to 600 μm can be fabricated so that it contains thousands of fine capillary channels each from 3 to 50 μm in diameter (see photo). Thousands of such multichannel threads (Fig. 5.78) are then used together to conveniently focus or collimate X-ray beams in a way never before possible.

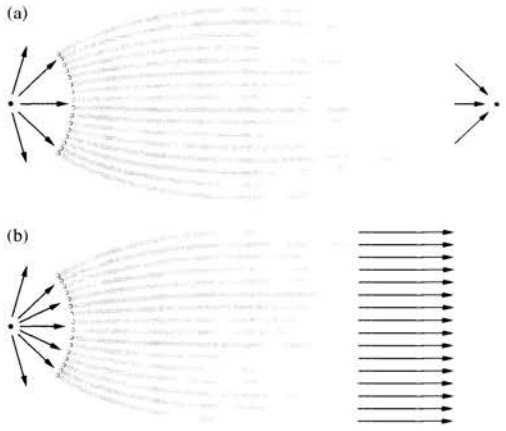


Figure 5.78 A bundle of multicapillary threads used to (*a*) focus or (*b*) collimate the X-rays from a point source.

5.7 Optical Systems

We have developed paraxial theory to a point where it is possible to appreciate the principles underlying the majority of practical optical systems. To be sure, the subtleties involved in controlling aberrations are extremely important and still beyond this discussion. Even so, one could build, for example, a telescope (admittedly not a very good one, but a telescope nonetheless) using the conclusions already drawn from first-order theory.

What better starting point for a discussion of optical instruments than the most common of all—the eye?

5.7.1 Eyes

For our purposes, three main groupings of eyes can readily be distinguished: those that gather radiant energy and form images via a single-centered lens system, those that utilize a multifaceted arrangement of tiny lenses (feeding into channels resembling optical fibers), and the most rudimentary, those that simply function with a small lensless hole (p. 216). In addition to light-sensing eyes, the rattlesnake has infrared pinhole "eyes" called pits, which might be included in this last group.

Visual lens systems of the first type have evolved independently and remarkably similarly in at least three distinct kinds of organisms. Some of the more advanced mollusks (e.g., the octopus), certain spiders (e.g., the avicularia), and the vertebrates, ourselves included, possess eyes that each form a single continuous real image on a light-sensitive screen or retina. By comparison, the multifaceted compound eye (Fig. 5.79) developed independently among arthropods, the creatures with articulated bodies and limbs (e.g., insects and crayfish). It produces a mosaic sensory image composed of many small-field-of-view spot contributions, one from each tiny segment of the eye (as if one were looking at the world through a tightly packed bundle of exceedingly fine tubes). Like a television picture made up of different-intensity dots, the compound eye divides and digitizes the scene being viewed. There is no real image formed on a retinal screen; the synthesis takes place electrically in the nervous system. The horsefly has about 7000 such segments, and the predatory dragonfly, an especially fast flyer, gets a better view with 30 000, as compared with some ants that manage with only about 50. The more facets, the more image dots, and the better the resolution, the sharper the composite picture. This may well be the oldest of eye

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

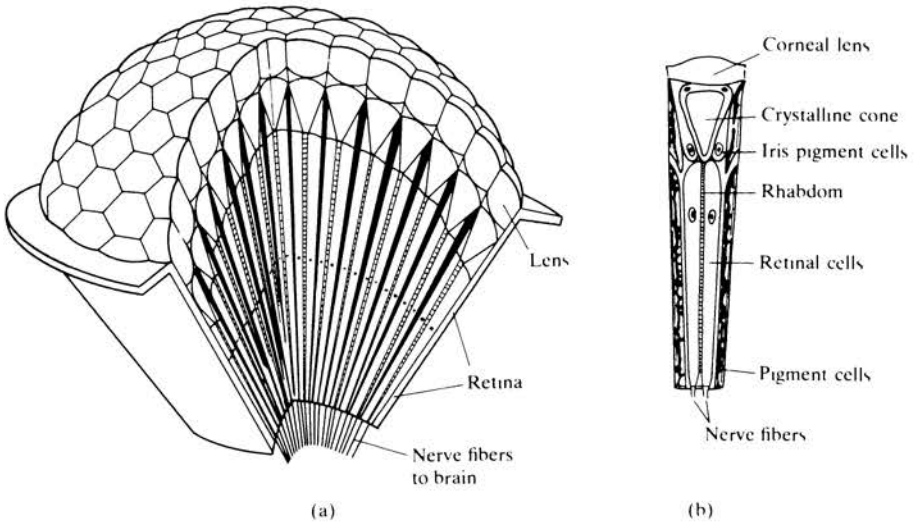


Figure 5.79 (*a*) The compound eye made up of many ommatidia. (*b*) An ommatidium, the little individual eye that each "sees" a small region in a particular direction. The corneal lens and crystalline cone channel the light into the sensing structure, the clear, rod-shaped rhabdom. Each of these is surrounded by retinal cells that lead via nerve fibers to the brain. (From Ackerman et al., *Biophysical Science*, © 1962, 1979. Englewood Cliffs, NJ: Prentice-Hall, Inc., p. 31. After R. Bushman, *Animals Without Backbones.*)

types: trilobites, the little sea creatures of 500 million years ago had well-developed compound eyes. Remarkably, however different the optics, the chemistry of the image-sensing mechanisms in all Earth animals is quite similar.

Structure of the Human Eye

The human eye can be thought of as a positive double-lens arrangement that casts a real image on a light-sensitive surface. That notion, in a rudimentary form, was apparently proposed by Kepler (1604), who wrote "Vision, I say, occurs when the image of the ... external world ... is projected onto the ... concave retina." This insight gained wide acceptance only after a lovely experiment was performed in 1625 by the German Jesuit Christopher Scheiner (and independently, about five years later, by Descartes). Scheiner removed the coating on the back of an animal's eyeball and, peering through the nearly transparent retina from behind, was able to see a minified, inverted image of the scene beyond the eye. Although it resembles a simple camera (p. 174), the seeing system (eye, optic nerve, and visual cortex) functions much more like a closed-circuit computerized television unit.

The eye (Fig. 5.80) is an almost spherical (24 mm long by about 22 mm across) jellylike mass contained within a tough flexible shell, the **sclera**. Except for the front portion, or **cornea**, which is transparent, the sclera is white and opaque. Bulging out from the body of the sphere, the cornea's curved surface (which is slightly flattened, thereby cutting down on spherical aberration) serves as the first and strongest convex element of the lens system. Indeed, most of the bending imparted to a bundle of rays takes place at the air–cornea interface. Incidentally, one of the reasons you can't see very well under water ($n_W \approx 1.33$) is that its index is too close to that of the cornea ($n_C \approx 1.376$) to allow for adequate refraction.

Light emerging from the cornea passes through a chamber filled with a clear watery fluid called the **aqueous humor** ($n_{ah} \approx 1.336$). It nourishes the anterior portion of the eye. A ray that is strongly refracted toward the optical axis at the air–cornea interface will be only slightly redirected at the cornea–aqueous humor interface because of the similarity of their indices. Immersed in the aqueous is a diaphragm known as the **iris**, which serves as the aperture stop controlling the amount of light entering the eye through the hole, or **pupil**. It is the iris (from the Greek word for rainbow) that gives the eye its characteristic blue, brown, gray, green, or hazel color. Made up of circular and radial muscles, the iris can expand or contract the pupil over a range from about 2 mm in bright light to roughly 8 mm in darkness. In addition to this function, it is also linked to the focusing response and will contract to increase image sharpness when doing close work.

Immediately behind the iris is the **crystalline lens**. The name, which is somewhat misleading, dates back to about A.D. 1000 and the work of Abû 'Alî al Hasan ibn al Hasan ibn al Haitham, alias Alhazen of Cairo, who described the eye as partitioned into three regions that were watery, crystalline, and glassy, respectively. The lens, which has both the size and shape of a small bean (9 mm in diameter and 4 mm thick), is a complex layered fibrous mass surrounded by an elastic mem-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

brane. In structure it is somewhat like a transparent onion, formed of roughly 22 000 very fine layers. It has some remarkable characteristics that distinguish it from man-made lenses, in addition to the fact that it continues to grow in size. Because of its laminar structure, rays traversing it will follow paths made up of minute, discontinuous segments. The lens as a whole is quite pliable, albeit less so with age. Moreover, its index of refraction ranges from about 1.406 at the inner core to approximately 1.386 at the less dense cortex, and as such it represents a gradient-index or GRIN system (p. 273). The crystalline lens provides the needed fine-focusing mechanism through changes in its shape; that is, it has a variable focal length—a feature we'll come back to presently.

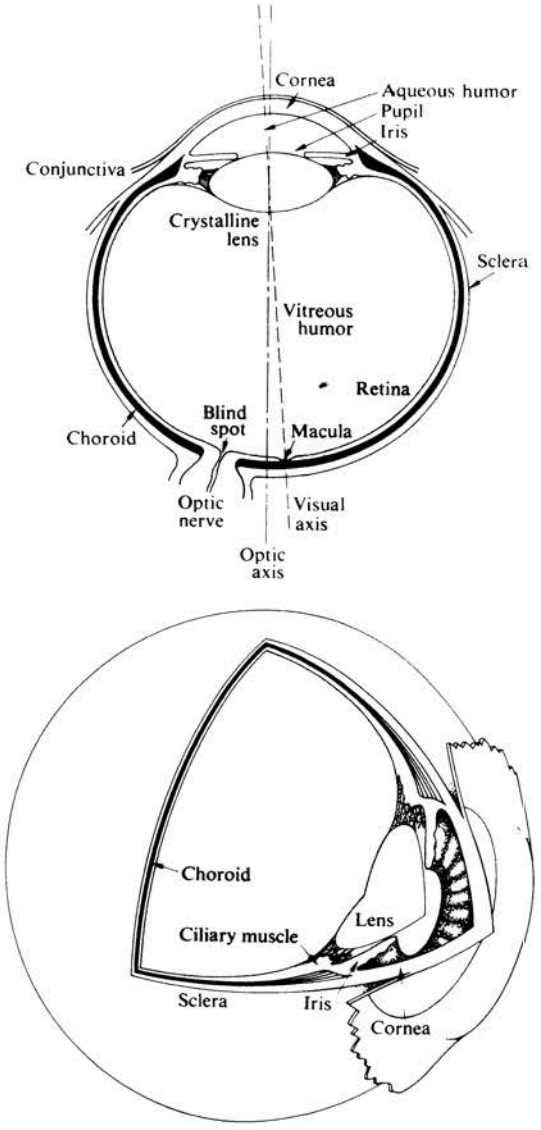


Figure 5.80 The human eye.

The refracting components of the eye, the cornea and crystalline lens, can be treated as forming an effective double-element lens with an object focus of about 15.6 mm in front of the anterior surface of the cornea and an image focus of about 24.3 mm behind it on the retina. To simplify things a little, we can take the combined lens to have an optical center 17.1 mm in front of the retina, which falls just at the rear edge of the crystalline lens.

Behind the lens is another chamber filled with a transparent gelatinous substance made of collagen (a protein polymer) and hyaluronic acid (a protein concentrate). Known as the **vitreous humor** ($n_{vh} \approx 1.337$), this thick gel gives support to the eyeball. As an aside, it should be noted that the vitreous humor contains microscopic particles of cellular debris floating freely about. You can easily see their shadows, outlined with diffraction fringes, within your own eye by squinting at a light source or looking at the sky through a pinhole—strange little amoebalike objects (*muscae volitantes*) will float across the field of view. Incidentally, a marked increase in one's perception of these floaters may be indicative of retinal detachment. While you're at it, squint at the source again (a broad diffuse fluorescent light works well). Closing your lids almost completely, you'll actually be able to see the near circular periphery of your own pupil, beyond which the glare of light will disappear into blackness. If you don't believe it, block and then unblock some of the light; the glare circle will visibly expand and contract, respectively. You are seeing the shadow cast by the iris from the inside! Seeing internal objects like this is known as entoptic perception.

Within the tough sclerotic wall is an inner shell, the choroid. It is a dark layer, well supplied with blood vessels and richly pigmented with melanin. The choroid is the absorber of stray light, as is the coat of black paint on the inside of a camera. A thin layer (about 0.5 mm to 0.1 mm thick) of light receptor cells covers much of the inner surface of the choroid—this is the **retina** (from the Latin *rete*, meaning net). The focused beam of light is absorbed via electrochemical reactions in this pinkish multilayered structure.

The human eye contains two kinds of photoreceptor cells: **rods** and **cones** (see photo). Roughly 125 million of them are intermingled nonuniformly over most of the retina. The ensemble of rods (each about 0.002 mm in diameter) in some respects has the characteristics of a high-speed, black-and-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



An electron micrograph of the retina of a salamander (*Necturus Maculosus*). Two visual cones appear in the foreground and several rods behind them. (Photo from E. R. Lewis, Y. Y. Zeevi, and F. S. Werblin, *Brain Research* **15**, 559 (1969)).

white film (such as Tri-X). It is exceedingly sensitive, performing in light too dim for the cones to respond to; yet it is unable to distinguish color, and the images it relays are not well defined. In contrast, the ensemble of 6 or 7 million cones (each about 0.006 mm in diameter) can be imagined as a separate, but overlapping, low-speed color film. It performs in bright light, giving detailed colored views, but it is fairly insensitive at low light levels.

The normal wavelength range of human vision is roughly 390 nm to 780 nm (Table 3.4). However, studies have extended these limits down to about 310 nm in the ultraviolet and up to roughly 1050 nm in the infrared. Indeed, people have reported "seeing" X-radiation. The limitation on ultraviolet transmission in the eye is set by the crystalline lens, which absorbs in the UV. People who have had a lens removed surgically have greatly improved UV sensitivity.

The area of exit of the optic nerve from the eye contains no receptors and is insensitive to light; accordingly, it is known as the **blind spot** (see Fig. 5.81). The optic nerve spreads out over the back of the interior of the eye in the form of the retina.

× **1** **2**

Figure 5.81 To verify the existence of the blind spot, close one eye and, at a distance of about 10 inches, look directly at the X—the 2 will disappear. Moving closer will cause the 2 to reappear while the 1 vanishes.

Just about at the center of the retina is a small depression from 2.5 to 3 mm in diameter known as the yellow spot, or **macula**. It is composed of more than twice as many cones as rods. There is a tiny rod-free region about 0.3 mm in diameter at the center of the macula called the **fovea centralis**. (In comparison, the image of the full Moon on the retina is about 0.2 mm in diameter—Problem 5.75.) Here the cones are thinner (with diameters of 0.0030 mm to 0.0015 mm) and more densely packed than anywhere else in the retina. Since the fovea provides the sharpest and most detailed information, the eyeball is continuously moving, so that light coming from the area on the object of primary interest falls on this region. An image is constantly shifted across different receptor cells by these normal eye movements. If such movements did not occur and the image was kept stationary on a given set of photoreceptors, it would actually tend to fade out. Without the fovea the eye would lose 90 to 95% of its capability, retaining only peripheral vision.

Another fact that indicates the complexity of the sensing system is that the rods are multiply connected to nerve fibers, and a single such fiber can be activated by any one of about a hundred rods. By contrast, cones in the fovea are individually connected to nerve fibers. The actual perception of a scene is constructed by the eye–brain system in a continuous analysis of the time-varying retinal image. Just think how little trouble the blind spot causes, even with one eye closed.

Between the nerve-fiber layer of the retina and the humor is a network of large retinal blood vessels, which can be observed entoptically. One way is to close your eye and place a bright small source against the lid. You'll "see" a pattern of shadows (*Purkinje figures*) cast by the blood vessels on the sensitive retinal layer.

Accommodation

The fine focusing, or **accommodation**, of the human eye is a function performed by the crystalline lens. The lens is suspended in position behind the iris by ligaments that are connected to a circular yoke composed of the **ciliary muscles**. Ordinarily, these muscles are relaxed, and in that state they pull outward radially on the network of fine fibers holding the rim of the lens. This draws the pliable lens into a fairly flat configuration, increasing its radii, which in turn increases its focal length [Eq. (5.16)]. With the muscles completely relaxed, the light from an object at infinity will be focused on the retina (Fig. 5.82). As the object moves closer to the eye, the ciliary muscles contract, relieving the external tension on

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

the periphery of the lens, which then bulges slightly under its own elastic forces. In so doing, the focal length decreases such that s_i is kept constant. As the object comes still closer, the yoke of ciliary muscles becomes more tensely contracted, the circular region they encompass gets still smaller, and the lens surfaces take on even smaller radii. The closest point on which the eye can focus is known as the **near point**. In a normal eye it might be about 7 cm for a teenager, 12 cm or so for a young adult, roughly 28 to 40 cm in the middle-aged, and about 100 cm by 60 years of age. Visual instruments are designed with this in mind, so that the eye need not strain unnecessarily. Clearly, the eye cannot focus on two different objects at once. This will be made obvious if, while looking through a piece of glass, you try to focus on it and the scene beyond at the same time.

Mammals generally accommodate by varying the lens curvature, but there are other means. Fish move only the lens itself toward or away from the retina, just as the camera lens is moved to focus. Some mollusks accomplish the same thing by contracting or expanding the whole eye, thus altering the relative distance between lens and retina. For birds of prey, which must keep a rapidly moving object in constant focus over a wide range of distances as a matter of survival, the accommodation mechanism is quite different. They accommodate by greatly changing the curvature of the cornea.

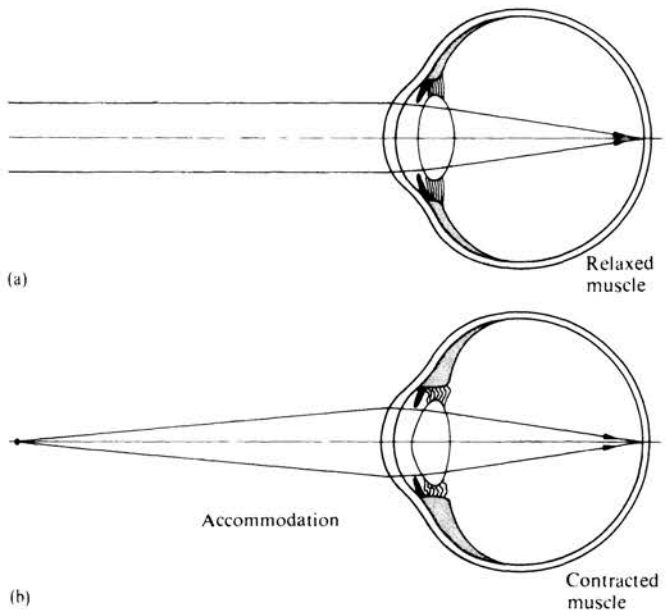


Figure 5.82 Accommodation—changes in the lens configuration.

5.7.2 Eyeglasses

Spectacles were probably invented some time in the late thirteenth century, possibly in Italy. A Florentine manuscript of the period (1299), which no longer exists, spoke of "spectacles recently invented for the convenience of old men whose sight has begun to fail." These were biconvex lenses, little more than variations on the handheld magnifying or reading glasses, and polished gemstones were no doubt employed as lorgnettes long before that. Roger Bacon (ca. 1267) wrote about negative lenses rather early on, but it was almost another two hundred years before Nicholas Cusa first discussed their use in eyeglasses and a hundred years more before such glasses ceased to be a novelty, in the late 1500s. Amusingly, it was considered improper to wear spectacles in public even as late as the eighteenth century, and we see few users in the paintings up until that time. In 1804 Wollaston, recognizing that traditional (fairly flat, biconvex, and concave) eyeglasses provided good vision only while one looked through their centers, patented a new, deeply curved lens. This was the forerunner of modern-day meniscus (from the Greek *meniskos*, the diminutive for moon, i.e., crescent) lenses, which allow the turning eyeball to see through them from center to margin without significant distortion.

It is customary and quite convenient in physiological optics to speak about the **dioptric power**, $\mathcal{D}$, of a lens, which is simply the reciprocal of the focal length. When f is in meters, the unit of power is the inverse meter, or *diopter*, symbolized by D: $1\ \text{m}^{-1} = 1$ D. For example, if a converging lens has a focal length of $+1$ m, its power is $+1$ D; with a focal length of -2 m (a diverging lens), $\mathcal{D} = -\frac{1}{2}$ D; for $f = +10$ cm, $\mathcal{D} = 10$ D. Since a thin lens of index n_l in air has a focal length given by

$$\frac{1}{f} = (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad [5.16]$$

its power is

$$\mathcal{D} = (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad (5.69)$$

You can get a sense of the direction in which we are moving by considering, in rather loose terms, that each surface of a lens bends the incoming rays—the more bending, the stronger the surface. A convex lens that strongly bends the rays at both surfaces has a short focal length and a large dioptric power. We already know that the focal length for two thin lenses in contact is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \qquad [5.38]$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

This means that the combined power is the sum of the individual powers, that is,

$$\mathscr{D} = \mathscr{D}_1 + \mathscr{D}_2$$

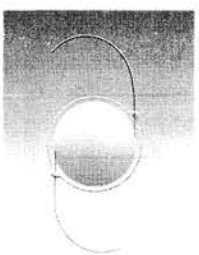
Thus a convex lens with $\mathscr{D}_1 = +10$ D in contact with a negative lens of $\mathscr{D}_2 = -10$ D results in $\mathscr{D} = 0$; the combination behaves like a parallel sheet of glass. Furthermore, we can imagine a lens, for example, a double convex lens, as being composed of two planar-convex lenses in intimate contact, back to back. The power of each of these follows from Eq. (5.69); thus for the first planar-convex lens ($R_2 = \infty$),

$$\mathscr{D}_1 = \frac{(n_l - 1)}{R_1} \tag{5.70}$$

and for the second,

$$\mathscr{D}_2 = \frac{(n_l - 1)}{-R_2} \tag{5.71}$$

These expressions may be equally well defined as giving the *powers of the respective surfaces* of the initial double convex lens. In other words, *the power of any thin lens is equal to the sum of the powers of its surfaces*. Because R_2 for a convex lens is a negative number, both $\mathscr{D}_1$ and $\mathscr{D}_2$ will be positive in that case. The power of a surface, defined in this way, is not generally the reciprocal of its focal length, although it is when immersed in air. Relating this terminology to the generally used model for the human eye, we note that the power of the crystalline lens *surrounded by air* is about +19 D. The cornea provides roughly +43 of the total +58.6 D of the intact unaccommodated eye.



When the normally clear lens in the eye becomes cloudy, the condition is referred to as a **cataract**. The resulting haziness can have a devastating effect on vision. In extreme cases the crystalline lens is usually surgically removed. A small convex plastic lens (an **intraocular lens implant**) is then inserted in the eye to enhance its convergence. (The photo shows an enlarged image of this type of converging spherical lens, it's actually only about 6 mm in diameter.) Its use has all but eliminated the need for the thick "cataract eye glasses" that were once required after surgery.

A normal eye, despite the connotation of the word, is not as common as one might expect. By the term *normal*, or its synonym *emmetropic*, we mean an eye that is capable of focusing parallel rays on the retina while in a relaxed condition—that is, one whose second focal point lies on the retina. For the unaccommodated eye, we define the object point whose image lies on the retina to be the **far point**. Thus for the normal eye the most distant point that can be brought to a focus on the retina, the far point, is located at infinity (which for all practical purposes is anywhere beyond about 5 m). In contrast, when the focal point does not lie on the retina, the eye is *ametropic* (e.g., it suffers hyperopia, myopia, or astigmatism). This can arise either because of abnormal changes in the refracting mechanism (cornea, lens, etc.) or because of alterations in the length of the eyeball that change the distance between the lens and the retina. The latter is by far the more common cause. Just to put things in proper perspective, note that about 25% of young adults require ±0.5 D or less of eyeglass correction, and perhaps as many as 65% need only ±1.0 D or less.

Nearsightedness—Negative Lenses

Myopia is the condition in which parallel rays are brought to focus in front of the retina; the power of the lens system as configured is too large for the anterior-posterior axial length of the eye. Images of distant objects fall in front of the retina, the far point is closer in than infinity, and all points beyond it will appear blurred. This is why myopia is often called **nearsightedness**; an eye with this defect sees nearby objects clearly (Fig. 5.83). To correct the condition, or at least its symptoms, we place an additional lens in front of the eye such that the combined spectacle–eye lens system has its focal point on the retina. Since the myopic eye can clearly see objects closer than the far point, the spectacle lens must cast relatively nearby images of distant objects. Hence we introduce a negative lens that will diverge the rays a bit. Resist the temptation to suppose that we are merely reducing the power of the system. In point of fact, the power of the lens–eye combination is most often made to equal that of the unaided eye. If you are wearing glasses to correct myopia, take them off; the world gets blurry, but it doesn't change size. Try casting a real image on a piece of paper using your glasses—it can't be done.

Suppose an eye has a far point of 2 m. All would be well if the spectacle lens appeared to bring more distant objects in closer than 2 m. If the virtual image of an object at infinity is formed by a concave lens at 2 m, the eye will see the object

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

clearly with an unaccommodated lens. Thus using the thin-lens approximation (eyeglasses are generally thin to reduce weight and bulk), we have

$$\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{\infty} + \frac{1}{-2} \qquad [5.17]$$

and $f = -2$ m while $\mathcal{D} = -\frac{1}{2}$ D. Notice that the far-point distance, measured from the correction lens, equals its focal length (Fig. 5.84). The eye views the right-side-up virtual images of all objects formed by the correction lens, and those images are located between its far and near points. Incidentally, the near point also moves away a little, which is why myopes often prefer to remove their spectacles when threading needles or reading small print; they can then bring the material closer to the eye, thereby increasing the magnification.

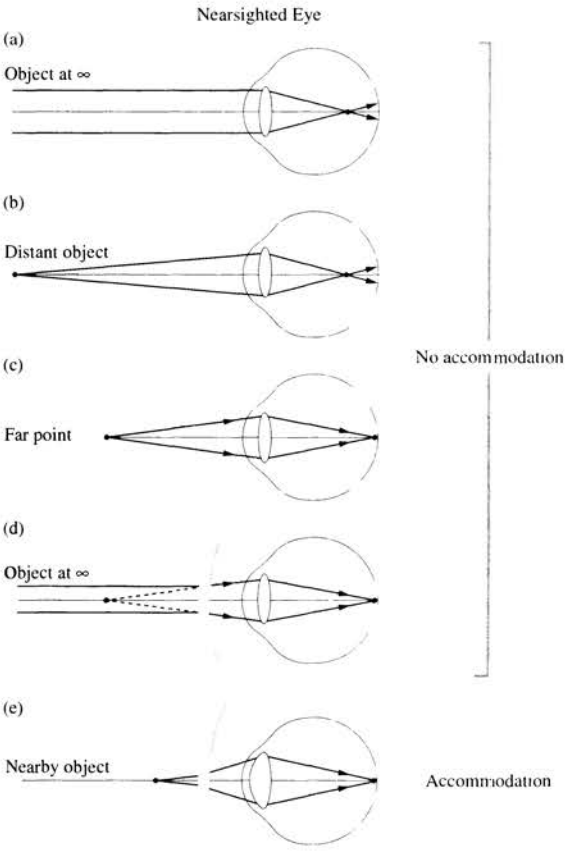


Figure 5.83 Correction of the nearsighted eye

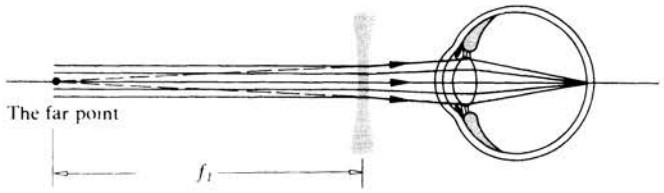


Figure 5.84 The far-point distance equals the focal length of the correction lens.

The calculation we have just performed overlooks the separation between the correction lens and the eye—in effect, it applies to contact lenses more than to spectacles. The separation is usually made equal to the distance of the first focal point of the eye (≈16 mm) from the cornea, so that no magnification of the image over that of the unaided eye occurs. Many people have unequal eyes, yet both yield the same magnification. A change in M_T for one and not the other would be a disaster. Placing the correcting lens at the eye's first focal point avoids the problem completely, regardless of the power of that lens [take a look at Eq. (6.8)]. To see this, just draw a ray from the top of some object through that focal point. The ray will enter the eye and traverse it parallel to the optic axis, thus establishing the height of the image. Yet, since this ray is unaffected by the presence of the spectacle lens, whose center is at the focal point, the image's location may change on insertion of such a lens, but its height and therefore M_T will not [see Eq. (5.24)].

The question now becomes: What is the equivalent power of a spectacle lens at some distance d from the eye (i.e., equivalent to that of a contact lens with a focal length f_c that equals the far-point distance)? It will do for our purposes to approximate the eye by a single lens and take d from that lens to the spectacle as roughly equal to the cornea–eyeglass distance, around 16 mm. Given that the focal length of the correction lens is f_l and the focal length of the eye is f_e, the combination has a focal length provided by Eq. (5.36), that is,

$$\text{b.f.l.} = \frac{f_e(d - f_l)}{d - (f_l + f_e)} \qquad (5.72)$$

This is the distance from the eye-lens to the retina. Similarly, the equivalent contact lens combined with the eye-lens has a focal length given by Eq. (5.38):

$$\frac{1}{f} = \frac{1}{f_c} + \frac{1}{f_e} \qquad (5.73)$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

where f = b.f.l. Inverting Eq. (5.72), setting it equal to Eq. (5.73), and simplifying, we obtain the result $1/f_c = 1/(f_l - d)$, independent of the eye itself. In terms of power,

$$\mathcal{D}_c = \frac{\mathcal{D}_l}{1 - \mathcal{D}_l d} \tag{5.74}$$

A spectacle lens of power $\mathcal{D}_l$ a distance d from the eye-lens has an effective power the same as that of a contact lens of power $\mathcal{D}_c$. Notice that since d is measured in meters and thus is quite small, unless $\mathcal{D}_l$ is large, as it often is, $\mathcal{D}_c \approx \mathcal{D}_l$. Usually, the point on your nose where you choose to rest your eyeglasses has little effect, but that's certainly not always the case; an improper value of d has resulted in many a headache.

Farsightedness—Positive Lenses

Hyperopia (or *hypermetropia*) is the defect that causes the second focal point of the unaccommodated eye to lie behind the retina (Fig. 5.85). **Farsightedness**, as you might have guessed it would be called, is often due to a shortening of the anteroposterior axis of the eye—the lens is too close to the retina. To increase the bending of the rays, a positive spectacle lens is placed in front of the eye. The hyperopic eye can and must accommodate to see distant objects distinctly, but it will be at its limit to do so for a near point, which is much farther away than it would be normally (this we take as 254 mm or just 25 cm). It will consequently be unable to see nearby objects clearly. A converging corrective lens with positive power will effectively move a close object out beyond the near point where the eye has adequate acuity; that is, it will form a distant virtual image, which the eye can then see clearly. Suppose that a hyperopic eye has a near point of 125 cm. For an object at +25 cm to have its image at $s_i = -125$ cm so that it can be seen as if through a normal eye, the focal length must be

$$\frac{1}{f} = \frac{1}{(-1.25)} + \frac{1}{0.25} = \frac{1}{0.31}$$

or $f = 0.31$ m and $\mathcal{D} = +3.2$ D. This is in accord with Table 5.3, where $s_o < f$. These spectacles will cast real images—try it if you're hyperopic.

As shown in Fig. 5.86, the correcting lens allows the relaxed eye to view objects at infinity. In effect, it creates an image on its focal "plane" (passing through F), which then serves as a virtual object for the eye. The point (whose image lies on the retina) is once again the *far point*, and it's a distance f_l *behind* the lens. The hyperope can comfortably "see" the far point, and any lens located anywhere in front of the eye that has an appropriate focal length will serve that purpose.

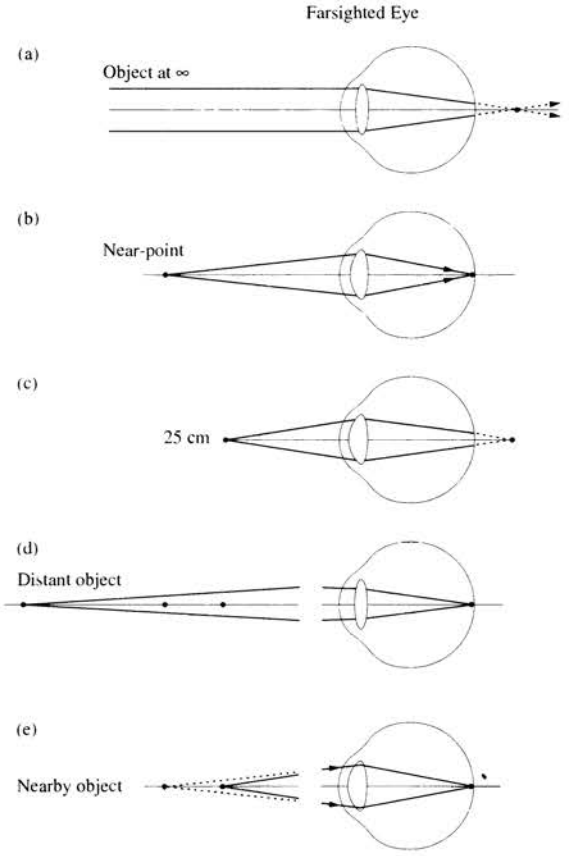


Figure 5.85 Correction of the farsighted eye.

Very gentle finger pressure on the lids above and below the cornea will temporarily distort it, changing your vision from blurred to clear and vice versa.

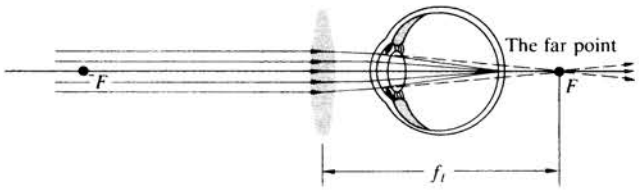


Figure 5.86 Again the far-point distance equals the focal length of the correction lens.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Astigmatism—Anamorphic Lenses

Perhaps the most common eye defect is **astigmatism**. It arises from an uneven curvature of the cornea. In other words, the cornea is asymmetric. Suppose we pass two meridional planes (one containing the optical axis) through the eye such that the (curvature or) power is maximal on one and minimal on the other. If these planes are perpendicular, the *astigmatism* is *regular* and correctable; if not, it is *irregular* and not easily corrected. Regular astigmatism can take different forms; the eye can be emmetropic, myopic, or hyperopic in various combinations and degrees on the two perpendicular meridional planes. Thus, as a simple example, the columns of a checkerboard might be well focused while the rows are blurred due to myopia or hyperopia. Obviously, these meridional planes need not be horizontal and vertical (Fig. 5.87).

The great astronomer Sir George B. Airy used a concave sphero-cylindrical lens to ameliorate his own myopic astigmatism in 1825. This was probably the first time astigmatism had been corrected. But it was not until the publication in 1862 of a treatise on cylindrical lenses and astigmatism by the Dutchman Franciscus Cornelius Donders (1818–1889) that ophthalmologists were moved to adopt the method on a large scale.

Any optical system that has a different value of M_T or $\mathscr{D}$ in two principal meridians is said to be *anamorphic*. Thus, for example, if we rebuilt the system depicted in Fig. 5.31, this time using cylindrical lenses (Fig. 5.88), the image would be distorted, having been magnified in only one plane. This is just the sort of distortion needed to correct for astigmatism when a defect exists in only one meridian. An appropriate planar cylindrical spectacle lens, either positive or negative, would restore essentially normal vision. When both perpendicular meridians require correction, the lens may be *sphero-cylindrical* or even *toric* as in Fig. 5.89.

Just as an aside, we note that anamorphic lenses are used in other areas, as for example, in the making of wide-screen

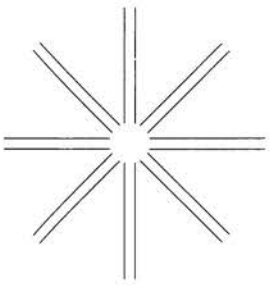
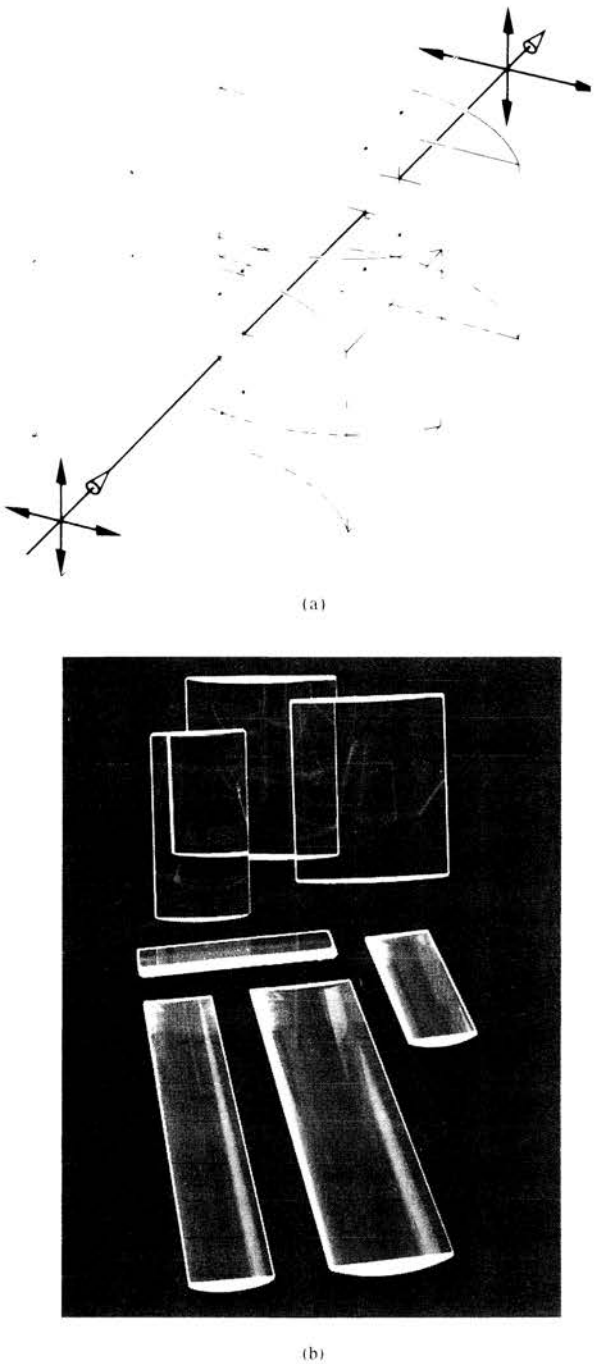


Figure 5.87 A test for astigmatism of the eye. View this figure through one unaided eye. If one set of lines appears bolder than the others, you have astigmatism. Hold the figure close to your eye; move it away slowly and note which set of lines comes into focus first. If two sets seem to be equally clear, rotate the figure until only one set is in focus. If all sets are clear you don't have astigmatism.



(a)

(b)

Figure 5.88 (*a*) An anamorphic system. (*b*) Cylindrical lenses (Photo courtesy Melles Griot.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027