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. 109) 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. 459). 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*.

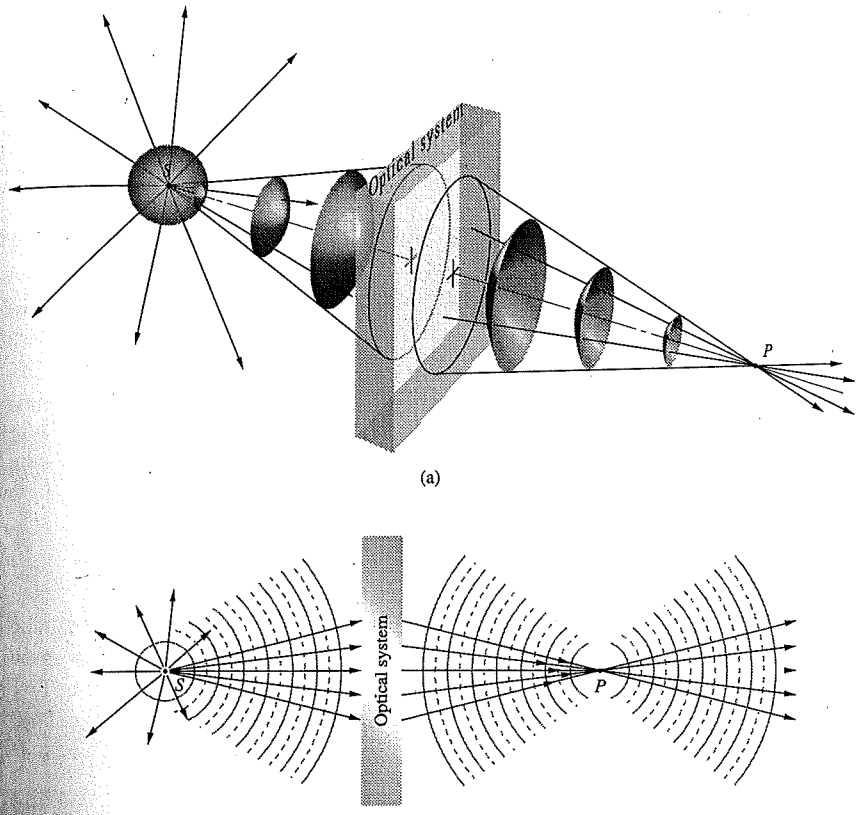


FIGURE 5.1 Conjugate foci. (*a*) A point source *S* sends out spherical waves. A cone of rays enters an optical system that inverts the wavefronts, causing them to converge on point *P*. (*b*) In cross section rays diverge from *S*, and a portion of them converge to *P*. If nothing stops the light at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

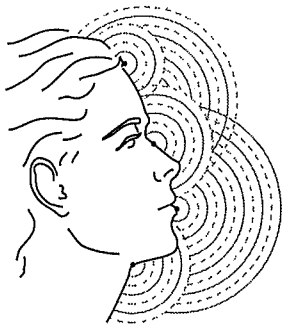


FIGURE 5.2 A person's face, like everything else we ordinarily see in reflected light, is covered with countless atomic scatterers.

tional view of a diverging spherical wave traveling in an incident 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. 28). 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, $\overline{F_1A}/\lambda_i$ (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

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

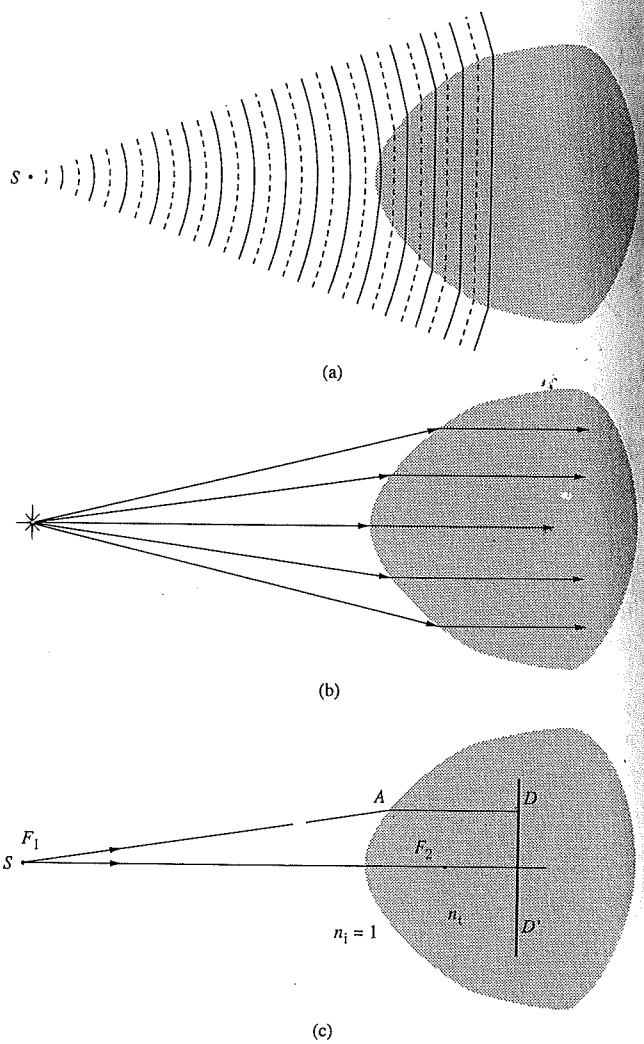


FIGURE 5.3 A hyperbolic interface between air and glass. (*a*) The wavefronts bend and straighten out. (*b*) The rays become parallel. (*c*) The hyperbola is such that the optical path from S to A to D is the same no matter where A is.

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

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

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 incident (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**.

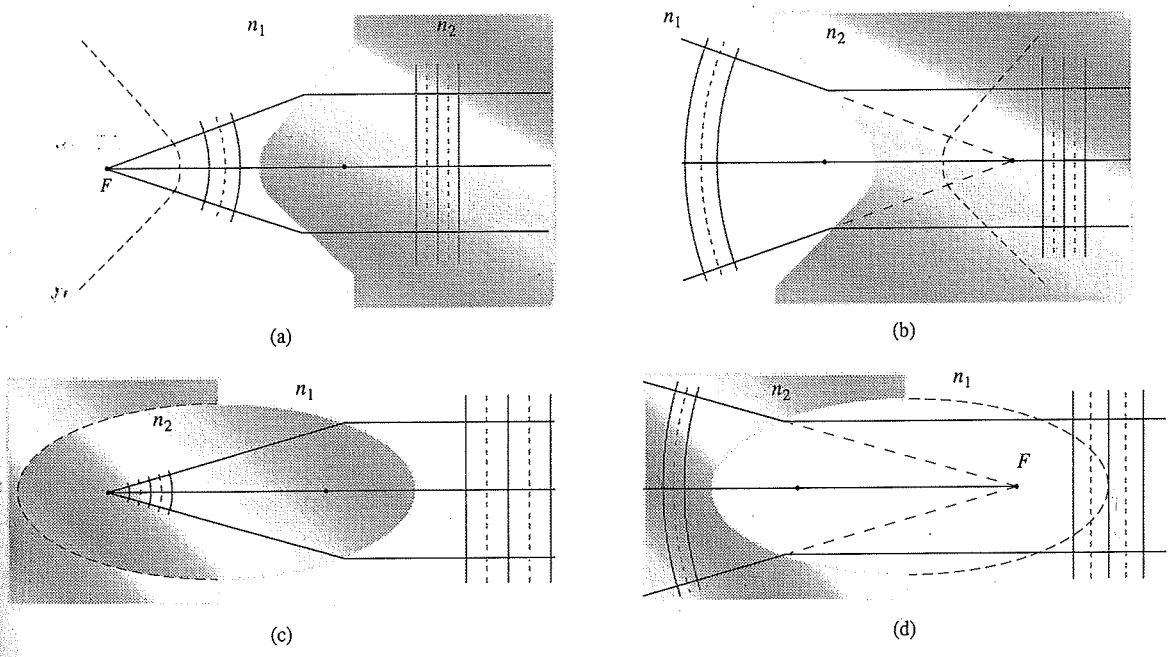
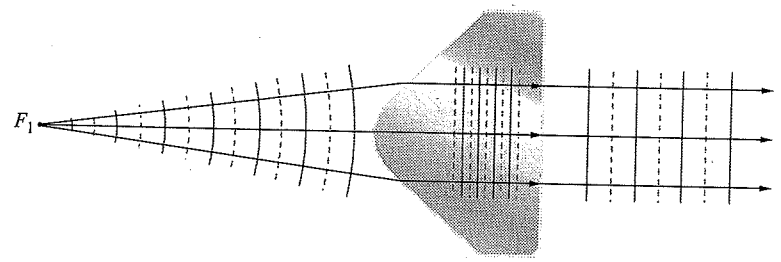
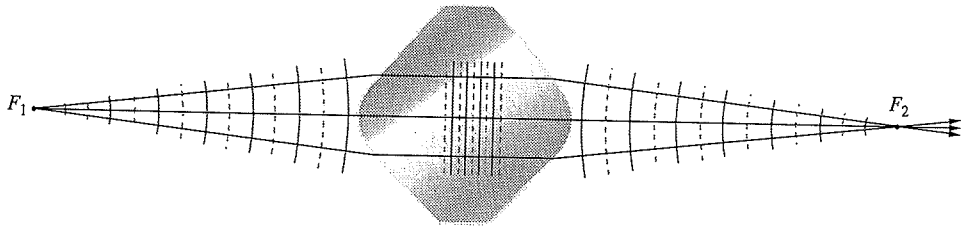


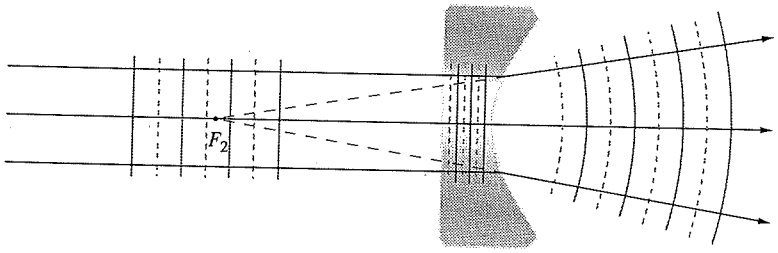
FIGURE 5.4 (*a*) and (*b*) Hyperboloidal and (*c*) and (*d*) ellipsoidal refracting surfaces ($n_2 > n_1$) in cross section.



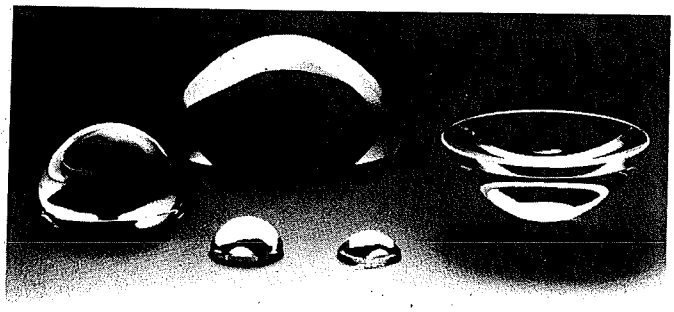
(a)



(b)



(c)



(d)

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.)

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 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.

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, but that will surely come in time. 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.

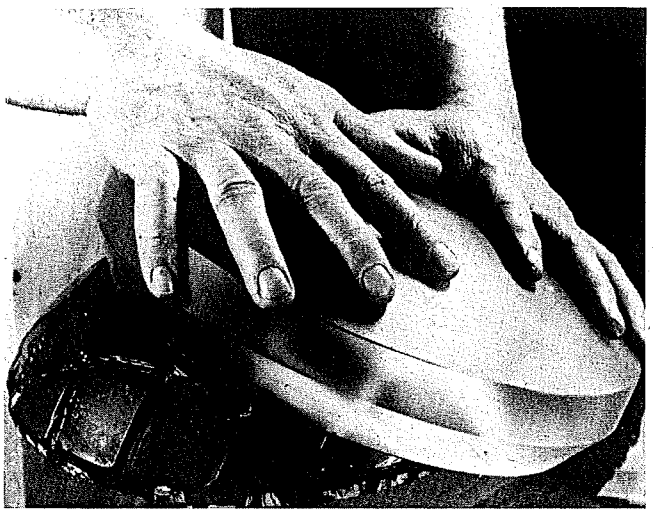


FIGURE 5.6 Polishing a spherical lens. (Photo courtesy Optical Society of America.)

5.2.2 Refraction at Spherical Surfaces

Imagine 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. Thus, 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 (Fig. 5.6). Such surfaces are commonly generated in batches by automatic grinding and polishing machines.

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 5.7 depicts a wave from the point source S impinging on a spherical interface of radius R centered at C. The

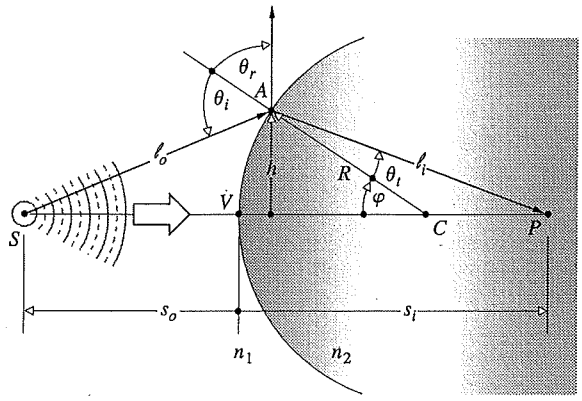


FIGURE 5.7 Refraction at a spherical interface. Conjugate foci.

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 point P the ray will cross the axis, as will all other rays incident at the same angle θ_i (Fig. 5.8). 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), $\cos\varphi \approx 1$.

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.

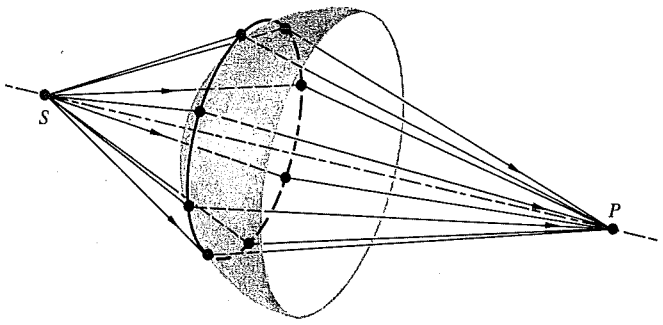


FIGURE 5.8 Rays incident at the same angle

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

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

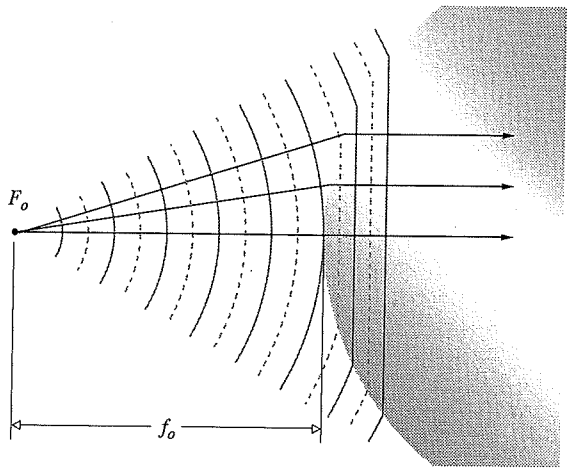


FIGURE 5.9 Plane waves propagating beyond a spherical interface—the object focus.

Defining the **second** or **image focal length** f_i as equal to s_i in this special case (Fig. 5.10), we have

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

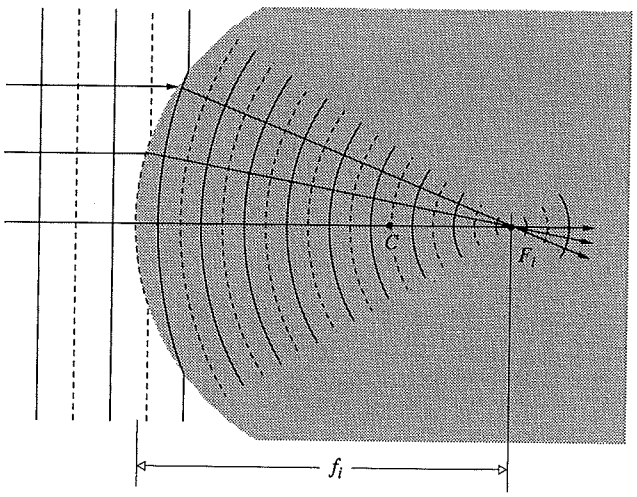


FIGURE 5.10 The reshaping of plane into spherical waves at a spherical interface—the image focus.

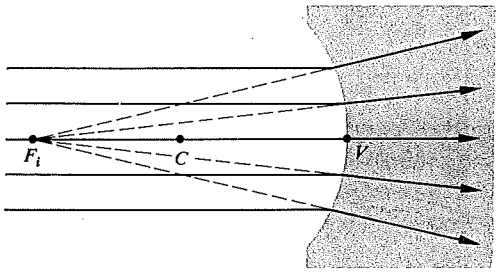


FIGURE 5.11 A virtual image point.

Recall that an image is virtual when the rays diverge from it (Fig. 5.11). Analogously, ***an object is virtual when the rays converge toward it*** (Fig. 5.12). 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.

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 (Fig. 5.13). 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).

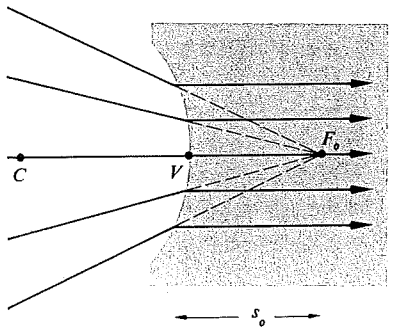


FIGURE 5.12 A virtual object point.

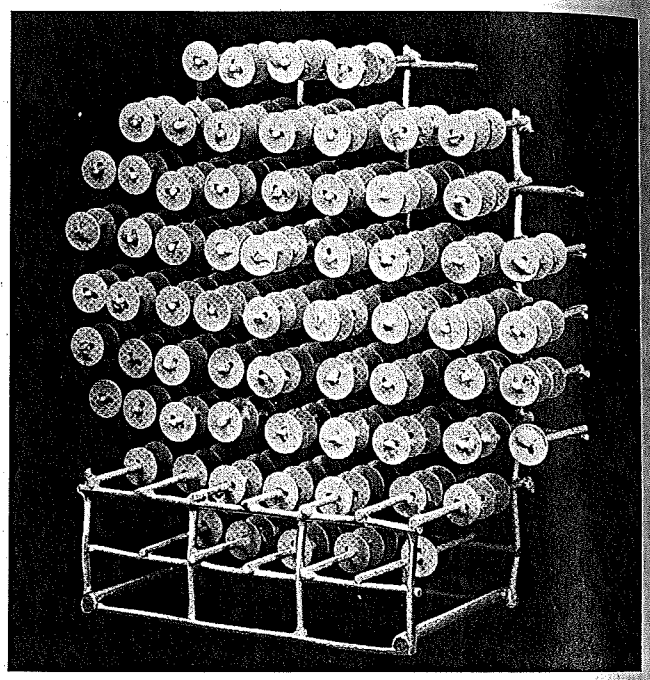
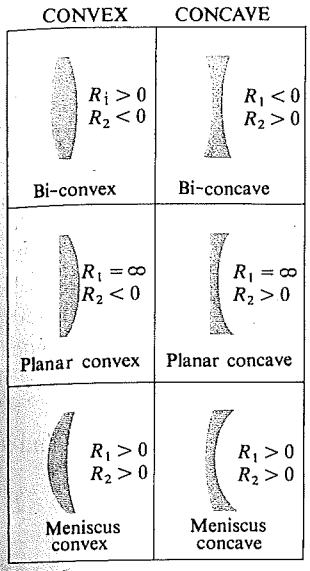


FIGURE 5.13 A lens for short-wavelength radiowaves. The disks serve to refract these waves much as rows of atoms refract light. (Photo courtesy Optical Society of America.)

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. 5.14.

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.



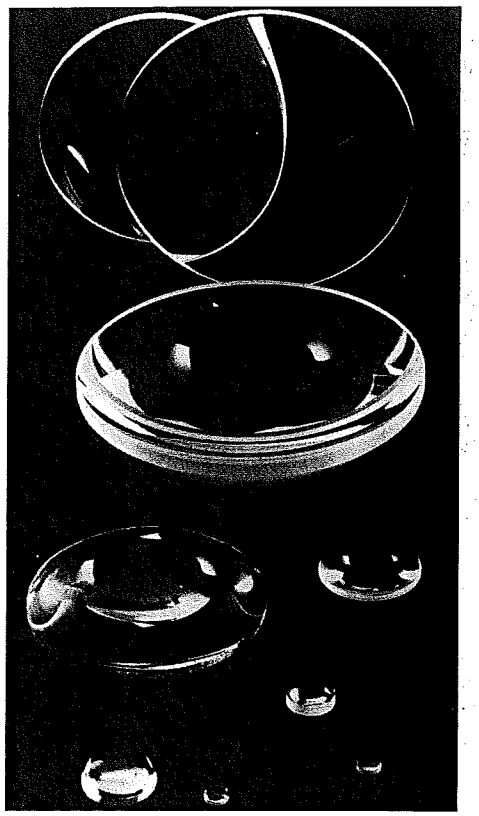


FIGURE 5.14 Cross sections of various centered spherical simple lenses. The surface on the left is #1 since it is encountered first. Its radius is R_1. (Photo courtesy of Melles Griot.)

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. 5.15).

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. 5.16, where another end has simply been ground onto the piece in Fig. 5.15*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

$$\frac{n_m}{s_{o1}} + \frac{n_l}{s_{i1}} = \frac{n_l - n_m}{R_1} \qquad (5.11)$$

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

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

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

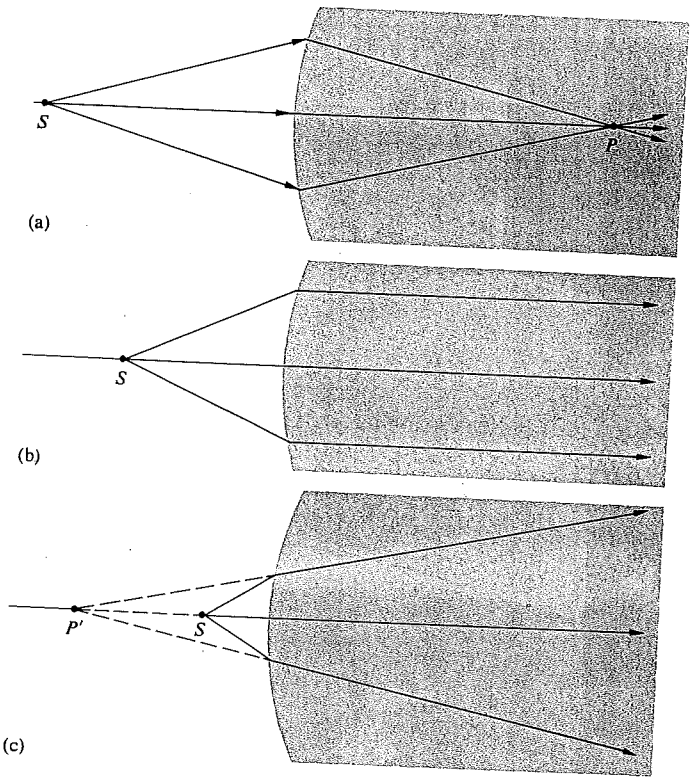


FIGURE 5.15 Refraction at a spherical interface between two transparent media shown in cross section.

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

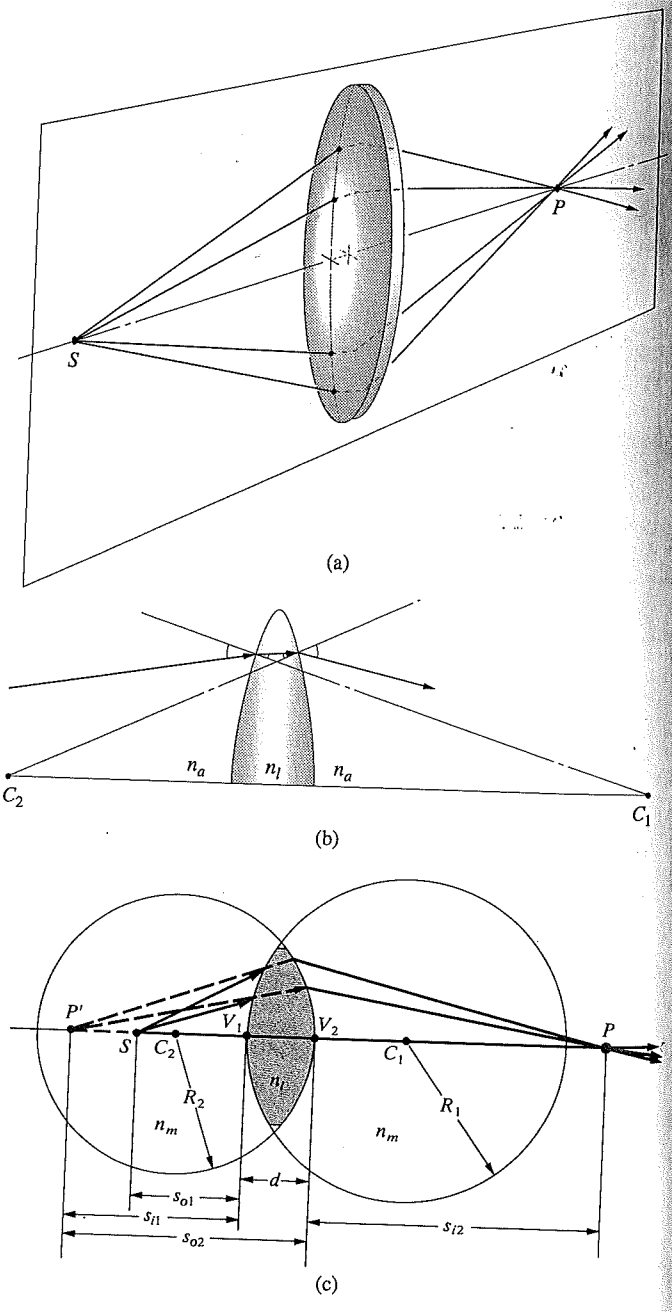


FIGURE 5.16 A spherical lens. (*a*) Rays in a vertical plane passing through a lens. Conjugate foci. (*b*) Refraction at the interfaces. The radius drawn from C_1 is normal to the first surface, and as the ray enters the lens it bends down *toward* that normal. The radius from C_2 is normal to the second surface; and as the ray emerges, since $n_l > n_a$, the ray bends down *away* from that normal. (*c*) The geometry.

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

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

which is the famous **Gaussian Lens Formula** (Fig. 5.17).

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 proba-

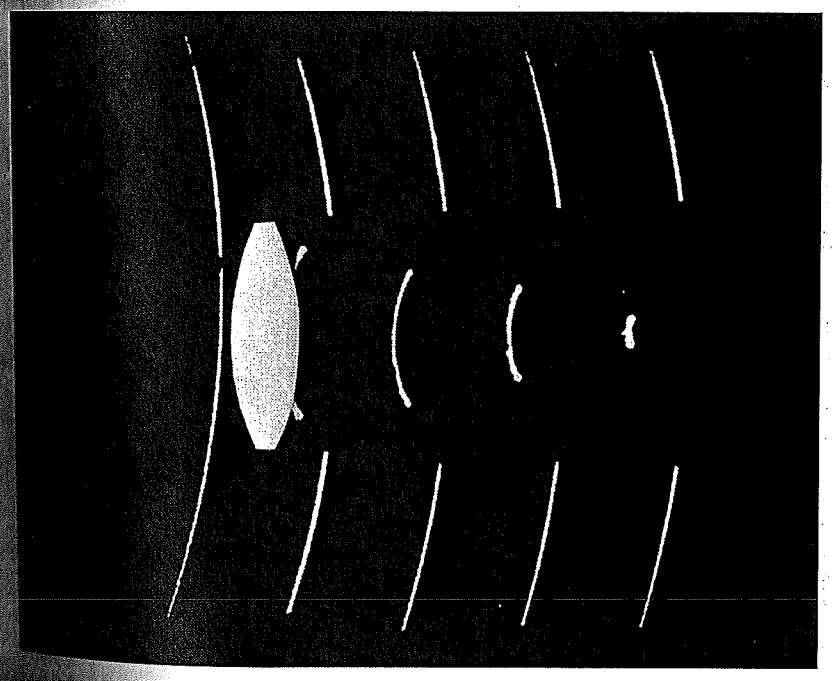


FIGURE 5.17 The actual wavefronts of a diverging lightwave partially focused by a lens. The photo shows five exposures, each separated by about 100 ps (i.e., 100×10^{-12} s), of a spherical pulse 10 ps long as it swept by and through a converging lens. The picture was made by Nils Abramson using a holographic technique.(Courtesy of N.H. Abramson)

bly 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 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 5.18 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. 5.18 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. 5.19. 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

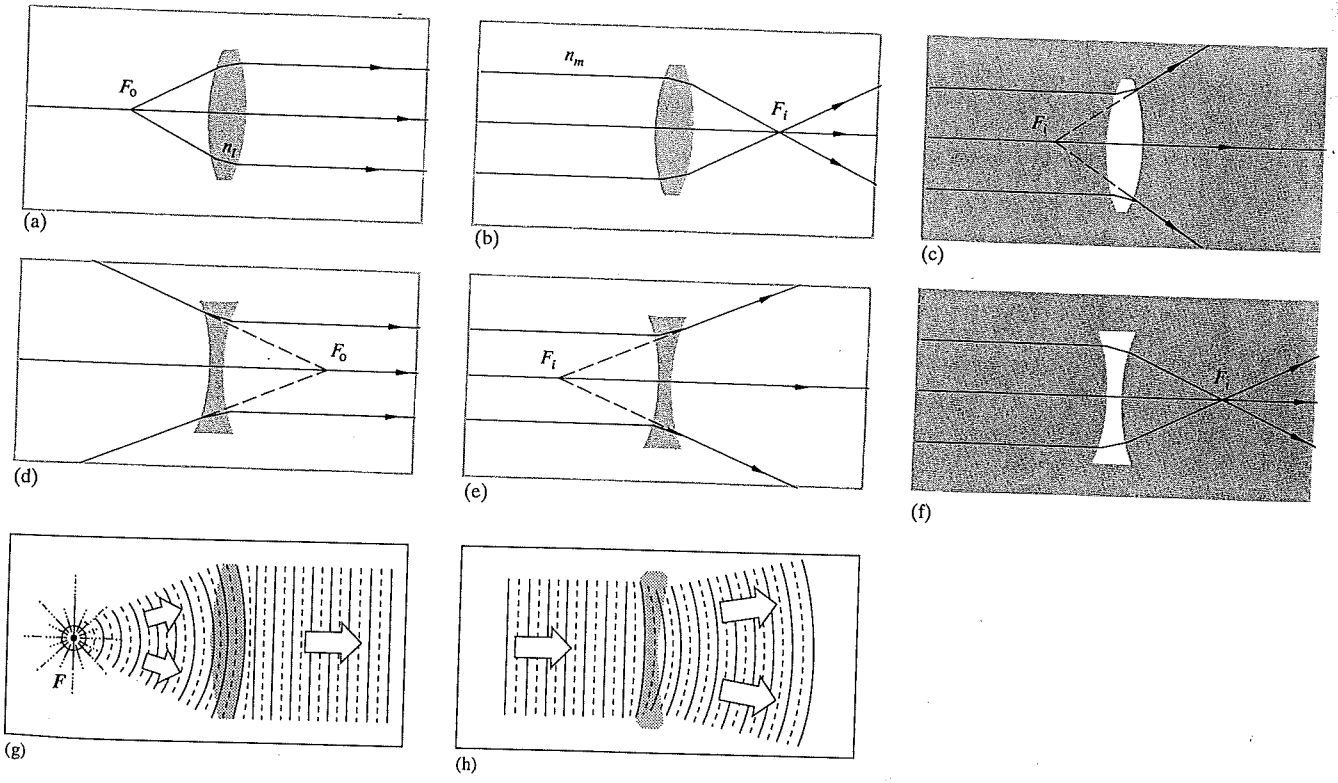


FIGURE 5.18 Focal lengths for converging and diverging lenses.

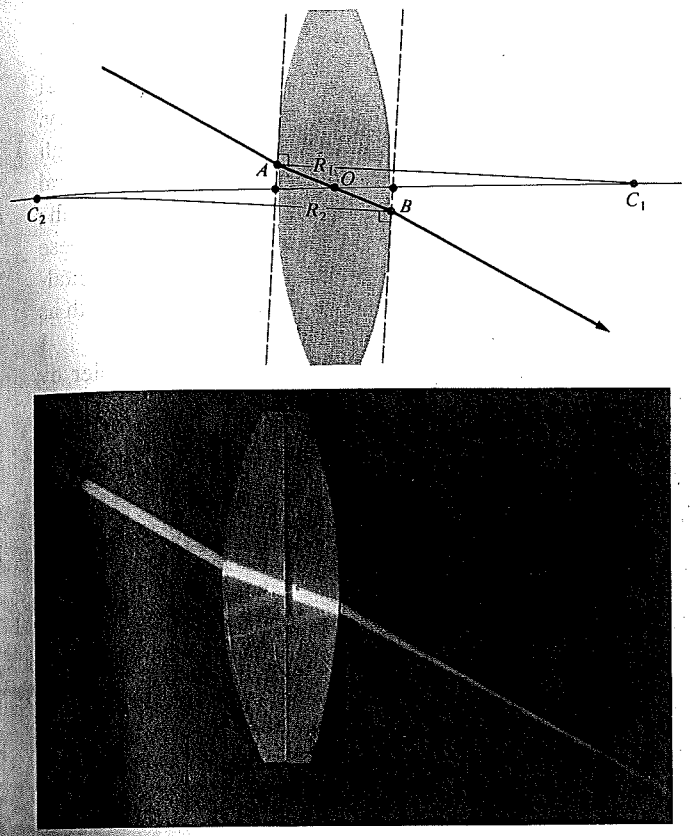


FIGURE 5.19 The optical center of a lens. (E.H.)

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 4.30 and Fig. P.4.30), a ray traversing a medium 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.

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. 5.11). As shown in Fig. 5.20, 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. 5.21. 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

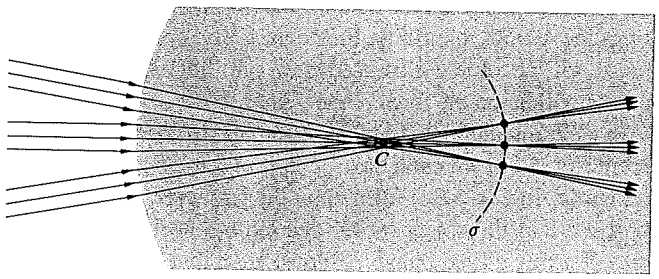


FIGURE 5.20 Focusing of several ray bundles.

*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.

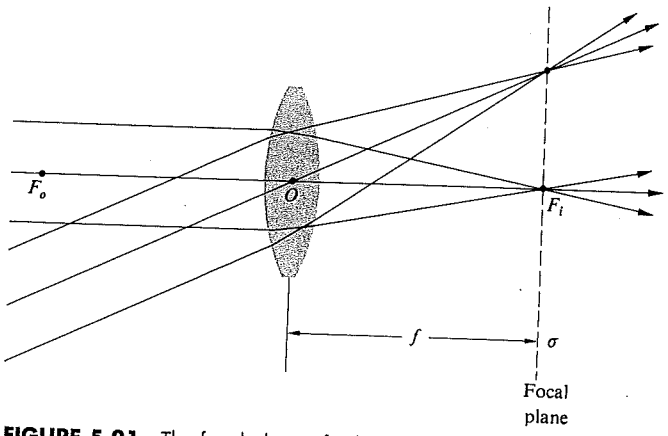


FIGURE 5.21 The focal plane of a lens.

sphere, σ_o, centered on C, as in Fig. 5.22. If σ_o is close to the 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. 5.21).

By cutting and polishing the right side of the piece depicted in Fig. 5.22, we can construct a thin lens. Once again, the image (σ_i in Fig. 5.22) 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. 5.22*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. 5.22*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. 5.23, 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. Two of these 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; the third is the undeviated ray through the center of the lens O. Figure 5.24 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. 5.25). 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,

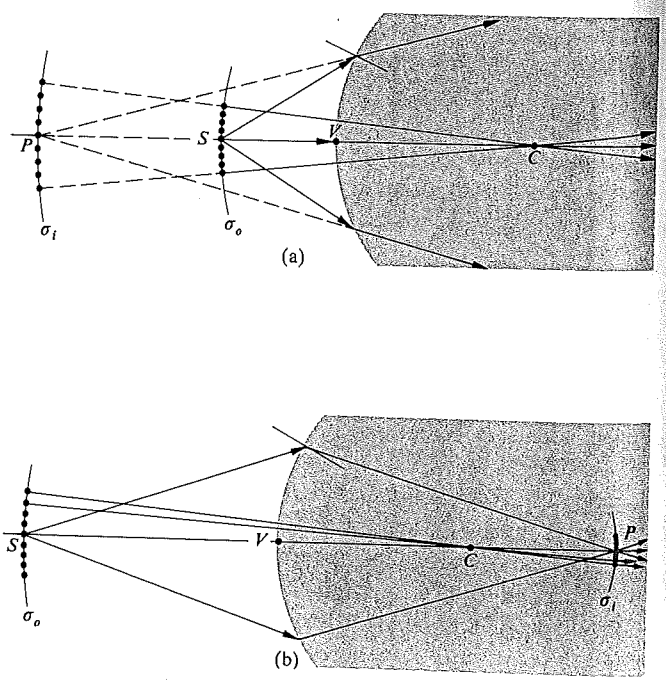


FIGURE 5.22 Finite imagery.

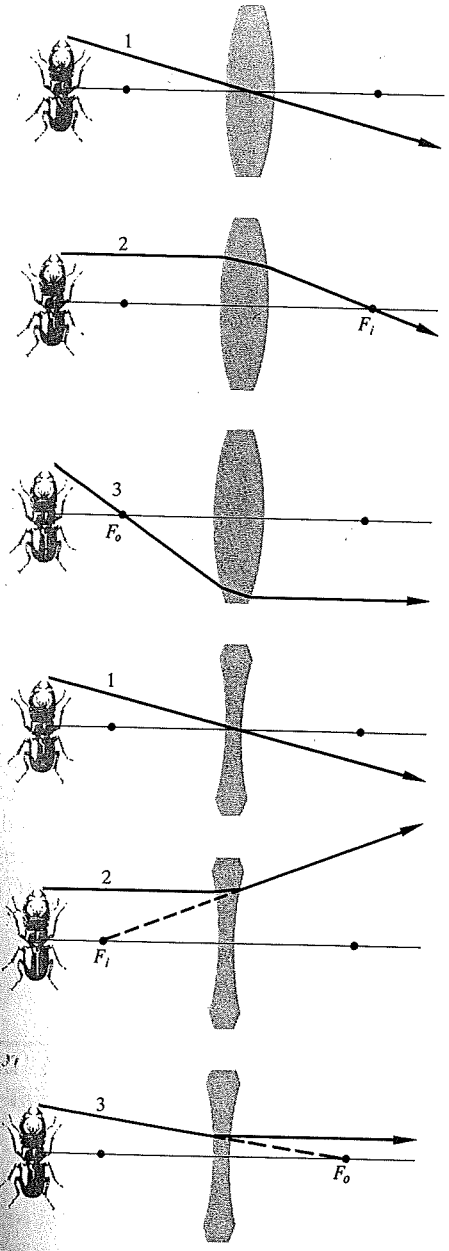


FIGURE 5.23 Tracing a few key rays through a positive and negative lens.

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. 5.25 $y_o > 0$ 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

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

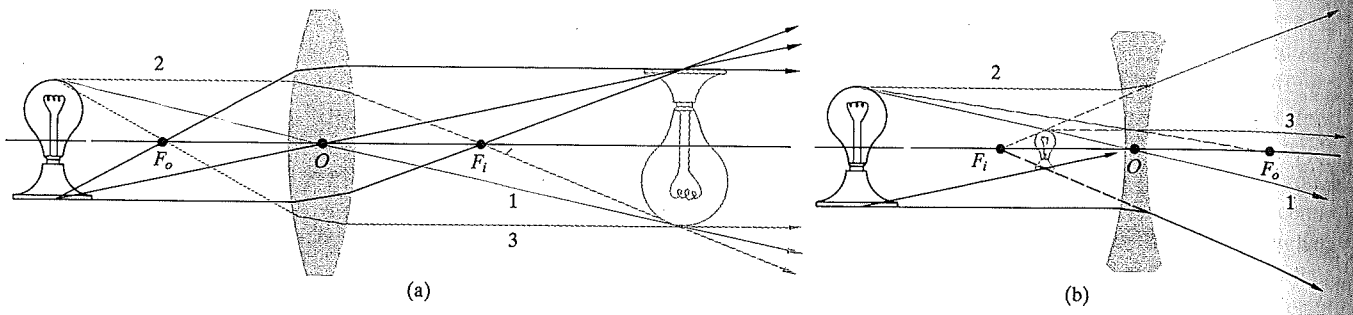
which is, of course, the Gaussian Lens Equation [Eq. (5.17)]. Furthermore, triangles $S_2S_1F_o$ and BOF_o are similar and

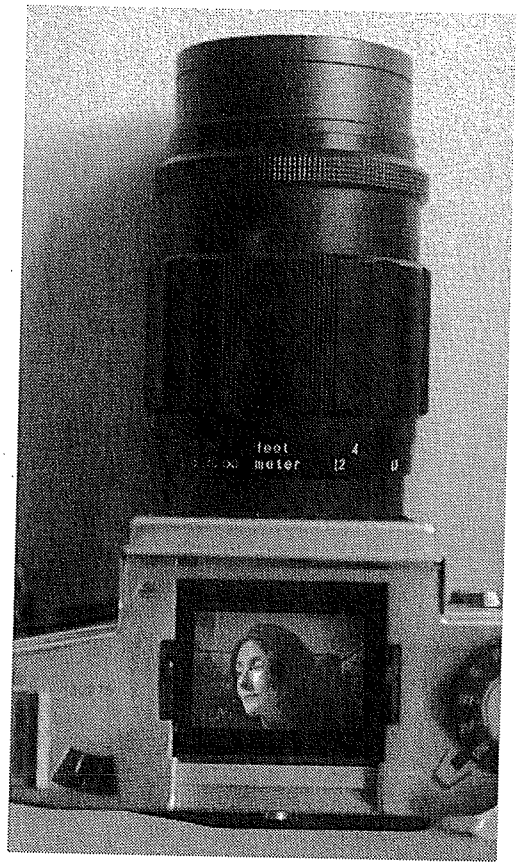
$$\frac{f}{(s_o - f)} = \frac{|y_i|}{y_o} \tag{5.22}$$

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

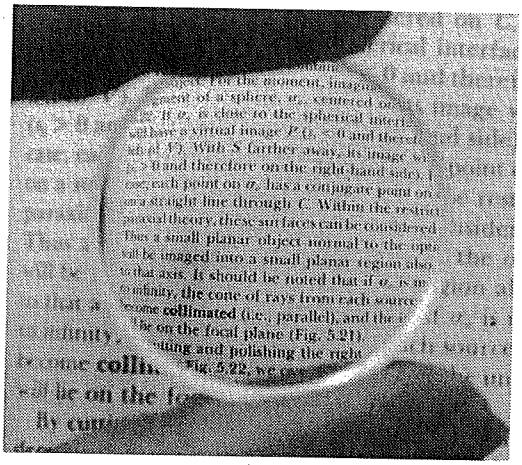
$$x_o x_i = f^2 \tag{5.23}$$

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.





(c)



(d)

FIGURE 5.24 (*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. (E.H.) (*d*) The minified, right-side-up, virtual image formed by a negative lens.

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

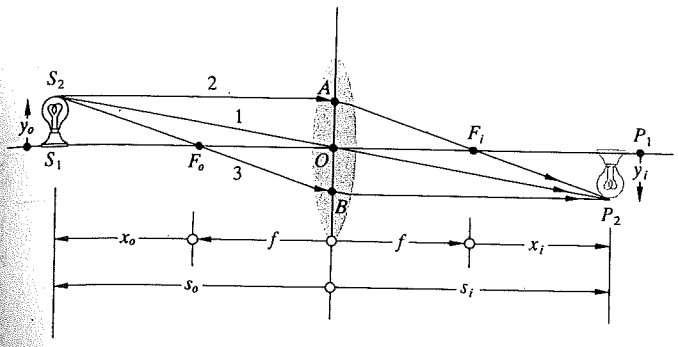


FIGURE 5.25 Object and image location for a thin lens.

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. 5.25:

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

The term *magnification* is 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 5.11). Table 5.3 summarizes a number of image configurations resulting from the juxtaposition of a thin lens and a real object.

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

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

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. 5.26). 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. 5.26*a*) and will be brought together at the focal point F_i. If the source point S_1 is closer (Fig. 5.26*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

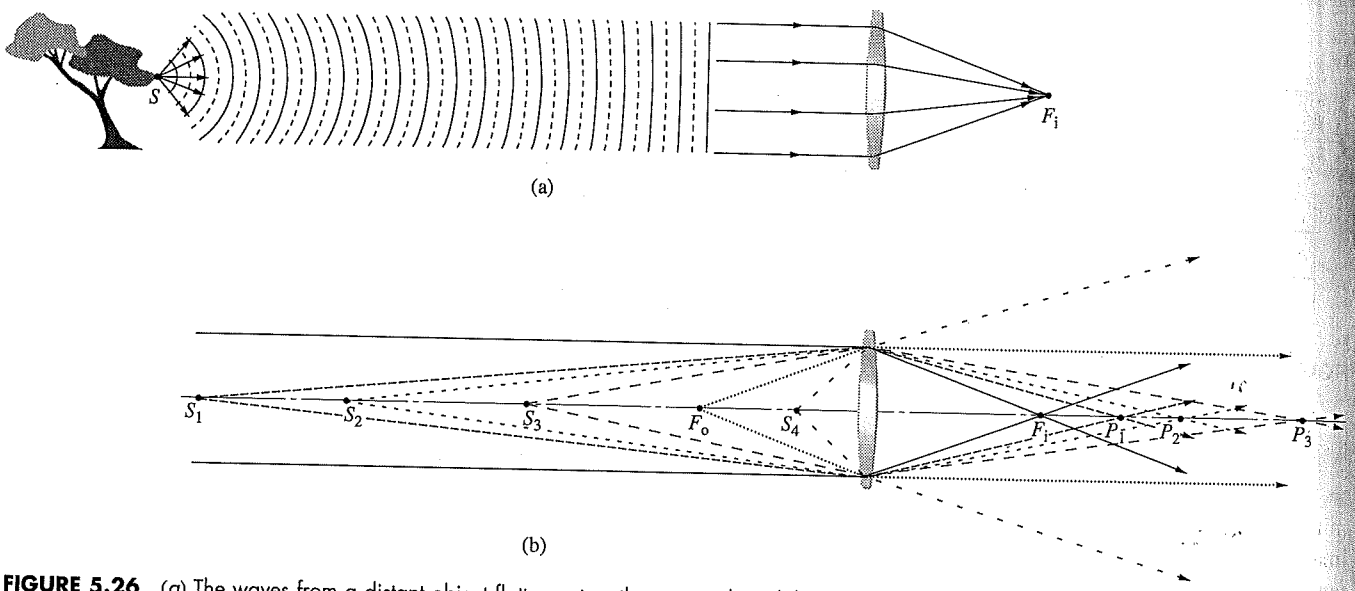


FIGURE 5.26 (*a*) The waves from a distant object flatten out as they expand, and the radii get larger and larger. Viewed from far away the rays from any point are essentially parallel, and the lens causes them to converge at F_i. (*b*) As a point source moves closer, the rays diverge more and the image 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.

diverge on leaving. The image point is now virtual—*there are no real images of objects at or closer in than f.* Figure 5.27 illustrates the behavior pictorially. **As the object approaches the lens, the real image moves away from it.**

It's useful to remember that *the ray entering the lens parallel to the central axis fixes the height of the real image* (Fig. 5.28). 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 5.28 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

FIGU

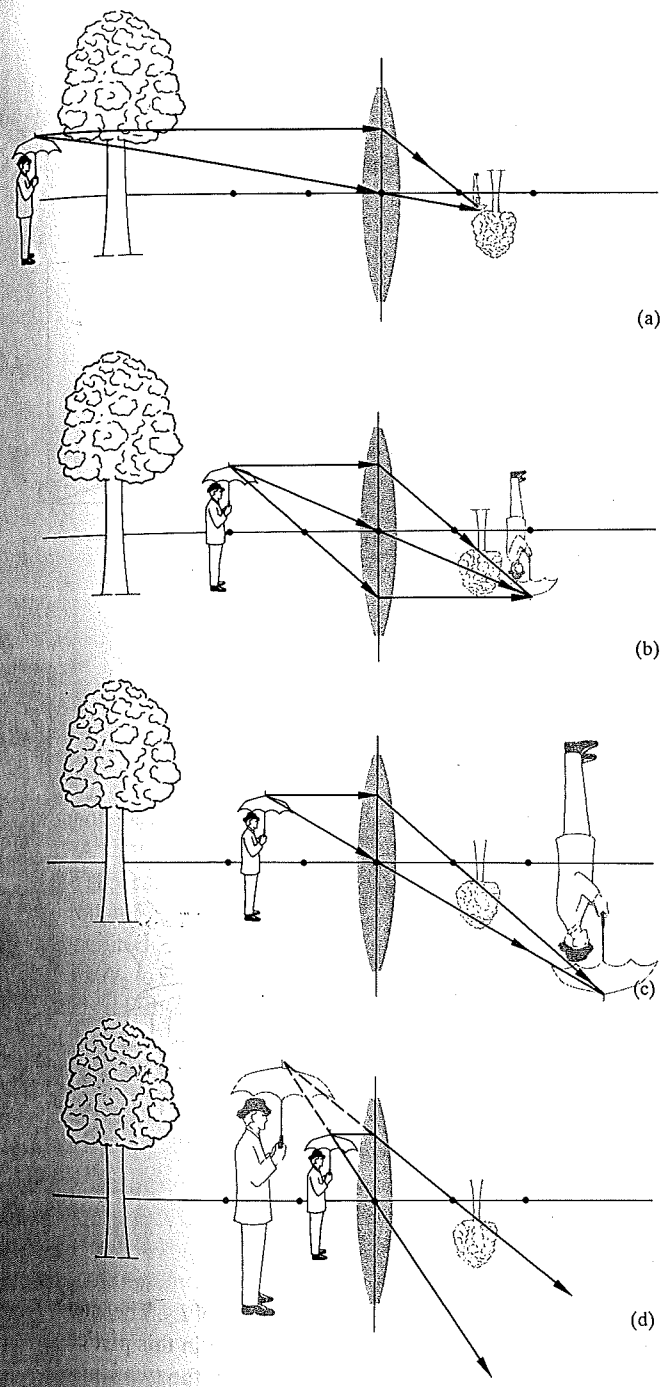


FIGURE 5.27 The image-forming behavior of a thin positive lens.

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. 5.29). 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. 5.30).

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 approximate calculations. Refinements are then added as the designer goes on to the prodigious and more exact ray-tracing techniques. Nowadays these computations are most often 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 will 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.

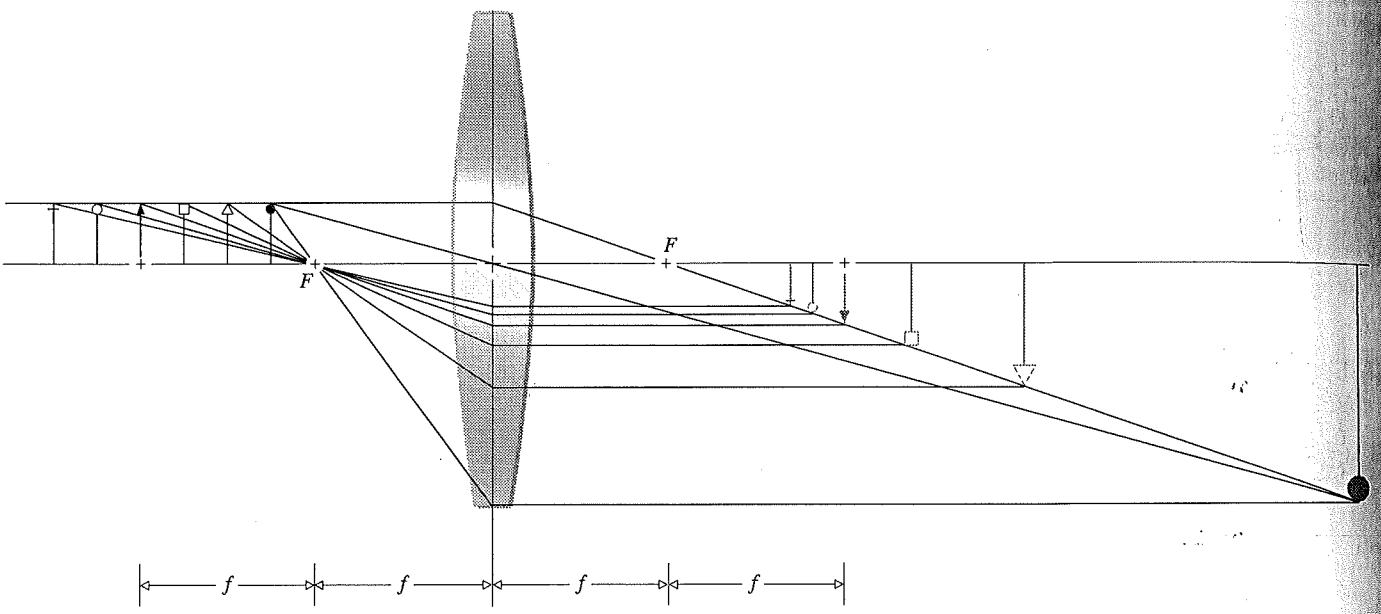


FIGURE 5.28 The number-2 ray entering the lens parallel to the central axis limits the image height.

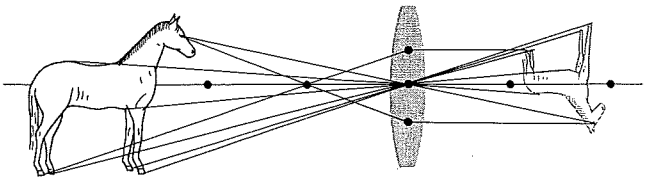


FIGURE 5.29 The transverse magnification is different from the longitudinal magnification.

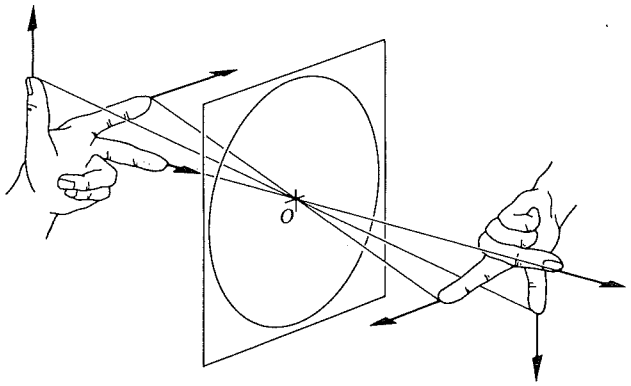


FIGURE 5.30 Image orientation for a thin lens.

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. 5.31. 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 1 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 its top, and a perpendicular to the optical axis finds its bottom. Ray 2 is then constructed running backward from P_1' through O_2. Insertion of L_2 has no effect on ray 2, whereas ray 3 is refracted through the image focus F_{i2} of L_2. The intersection of rays 2 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 essen-

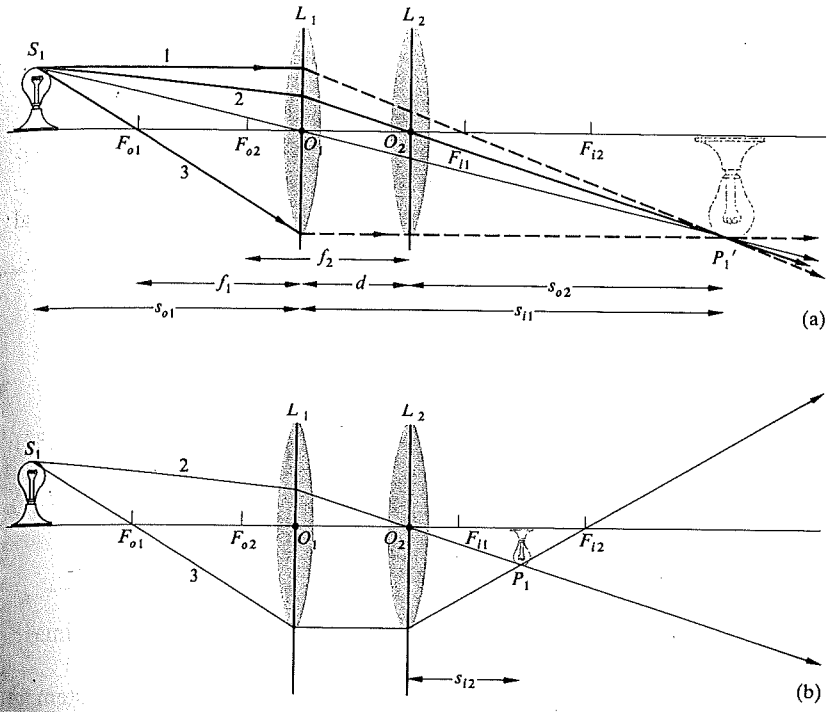


FIGURE 5.31 Two thin lenses separated by a distance smaller than either focal length.

tially adds convergence ($f_2 > 0$) or divergence ($f_2 < 0$) to the bundle of rays emerging from L_1; see Fig. 5.32.

A similar pair of lenses is illustrated in Fig. 5.33, in which the separation has been increased. Once again rays 1 and 3 through F_{i1} and F_{o1} fix the position of the intermediate image generated by L_1 alone. As before, ray 2 is drawn backward from O_2 to P_1' to S_1. The intersection of rays 2 and 3, as the latter 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. 5.33), whereas if $d < s_{i1}$, it is virtual ($s_{o2} < 0$, as in Fig. 5.31). In the former instance the rays approaching L_2 are diverging from P_1', 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),

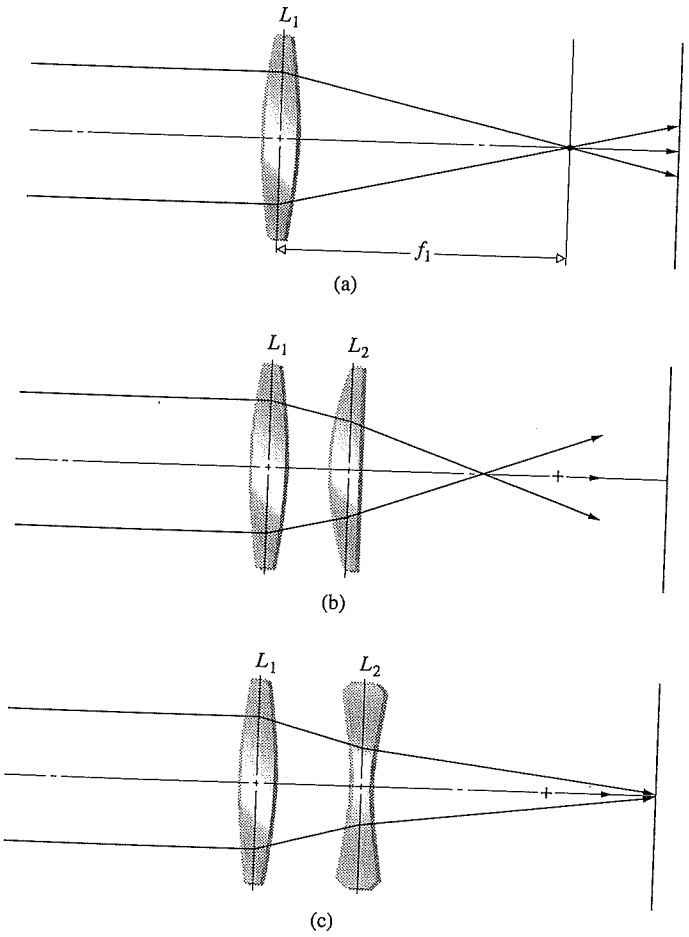


FIGURE 5.32 (*a*) The effect of placing a second lens, L_2, within the focal length of a positive lens, L_1. (*b*) When L_2 is positive, its presence adds convergence to the ray bundle. (*c*) When L_2 is negative, it adds divergence to the ray bundle.

$$s_{i2} = \frac{f_2 d - f_2 s_{o1} f_1/(s_{o1} - f_1)}{d - f_2 - s_{o1} f_1/(s_{o1} - f_1)} \tag{5.33}$$

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 5.31 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. 5.31, 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. and f.f.l. for the thin-lens system in Fig. 5.34*a*, where $f_1 = -30$ cm and $f_2 = +20$ cm. Then

$$\text{b.f.l.} = \frac{20[10 - (-30)]}{10 - (-30 + 20)} = 40 \text{ cm}$$

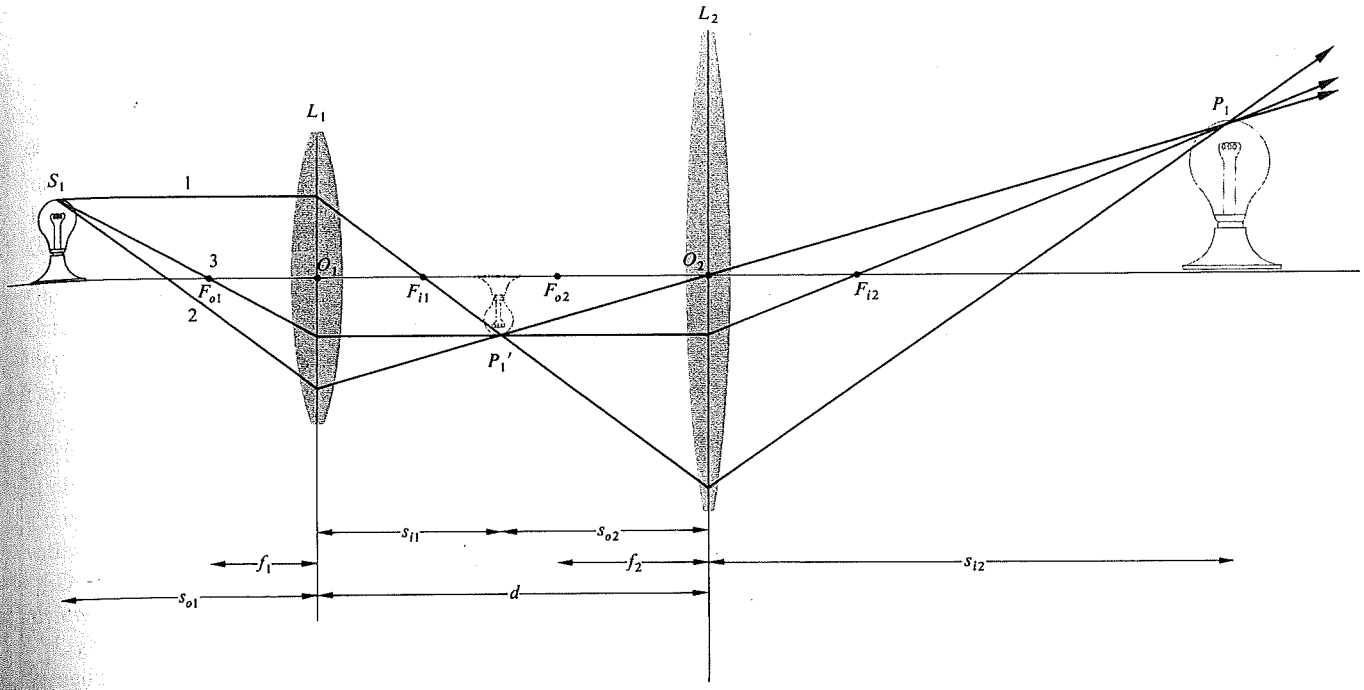
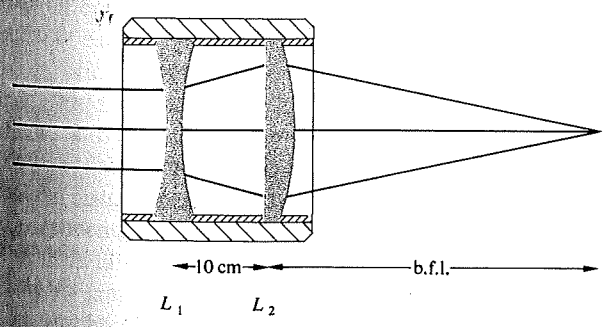


FIGURE 5.33 Two thin lenses separated by a distance greater than the sum of their focal lengths. Because the intermediate image is real, you could start with point P_i' and treat it as if it were a real object point for L_2. Thus a ray from P_i' through F_{o2} would arrive at P_1.

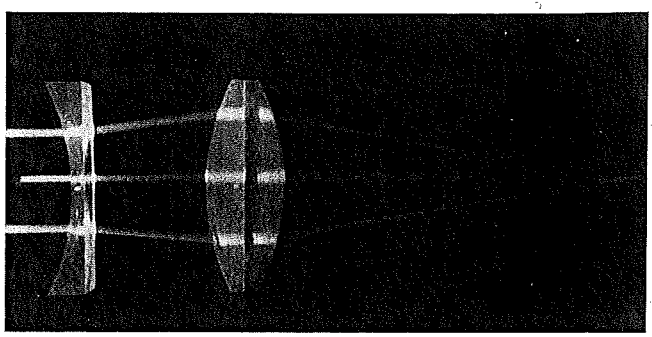
and similarly f.f.l. = 15 cm. Incidentally, notice that if $d = f_1 + f_2$, plane waves entering the compound lens from either side will emerge as plane waves (Problem 5.34), as in telescopic systems.

Observe that if $d \to 0$, that is, the lenses are brought into contact, as in the case of some achromatic doublets,

$$\text{b.f.l.} = \text{f.f.l.} = \frac{f_2 f_1}{f_2 + f_1} \tag{5.37}$$



(a)



(b)

FIGURE 5.34 (a) A positive and negative thin-lens combination, (b) photo (E.H.)

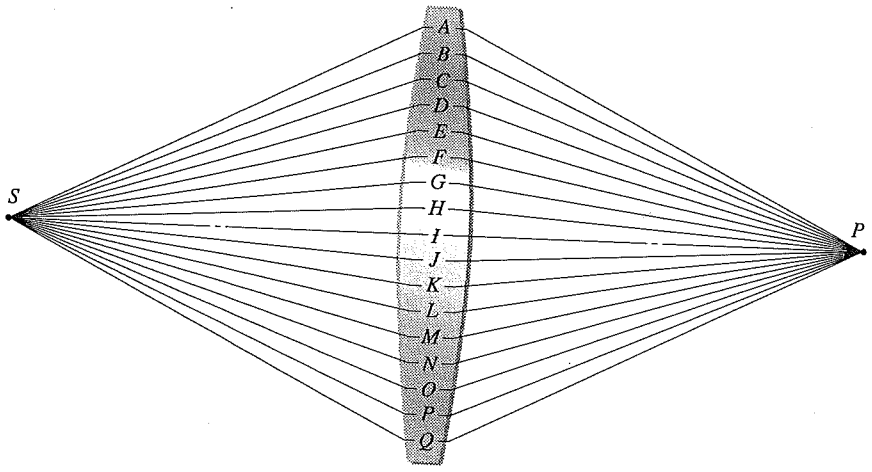
The resultant thin lens has an *effective focal length*, f, such that

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \tag{5.38}$$

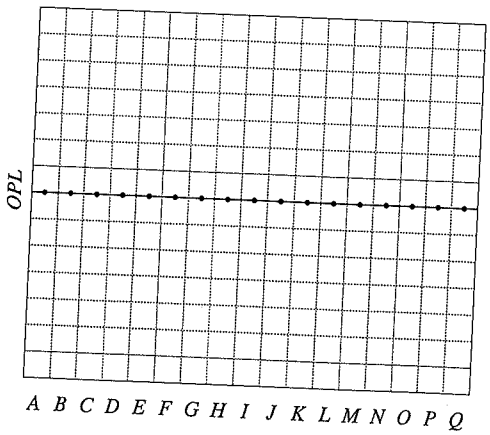
This implies that if there are N such lenses in contact,

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \cdots + \frac{1}{f_N} \tag{5.39}$$

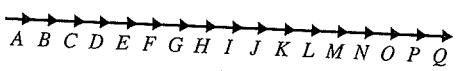
Many of these conclusions can be verified, at least qualitatively, with a few simple lenses. Figure 5.31 is easy to duplicate, and the procedure should be self-evident, whereas Fig. 5.33 requires a bit more care. First, determine the focal lengths of the two lenses by imaging a distant source. Then hold 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,



(a)



(b)



(c)

FIGURE 5.35 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.

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. 5.33.

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. 4.71 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. 5.35) 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. 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. 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. 459).

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. 5.36, 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 (Fig. 5.36) the aperture stop controls the number of rays from an object point reaching the conjugate image point, 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. 5.36 passes the field stop. Opening the circular

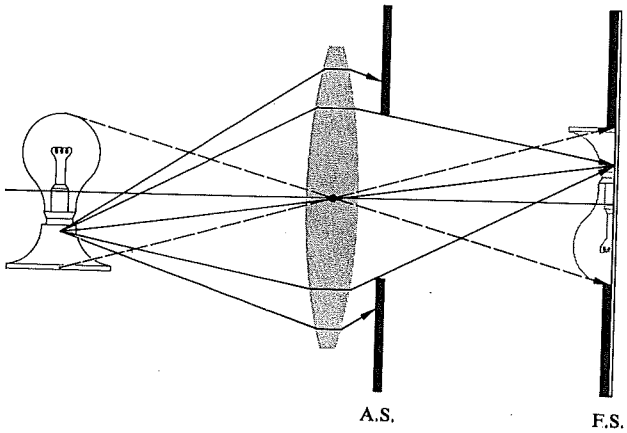


FIGURE 5.36 Aperture stop and field stop.

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. 5.37, 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 **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. 5.37 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.

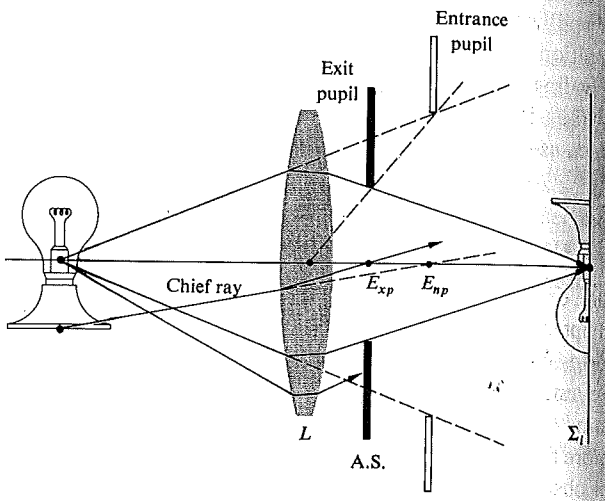


FIGURE 5.37 Entrance pupil and exit pupil.

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 5.38 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*

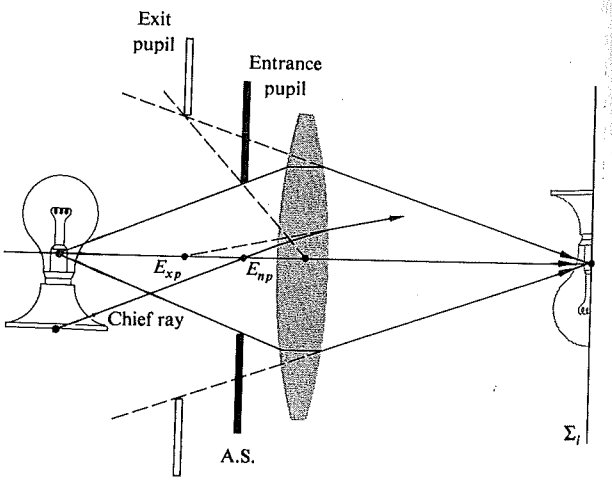


FIGURE 5.38 A front aperture stop.

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 5.39 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 5.38 deals with just this kind of calculation.

Notice how the cone of rays, in Fig. 5.40, 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 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.

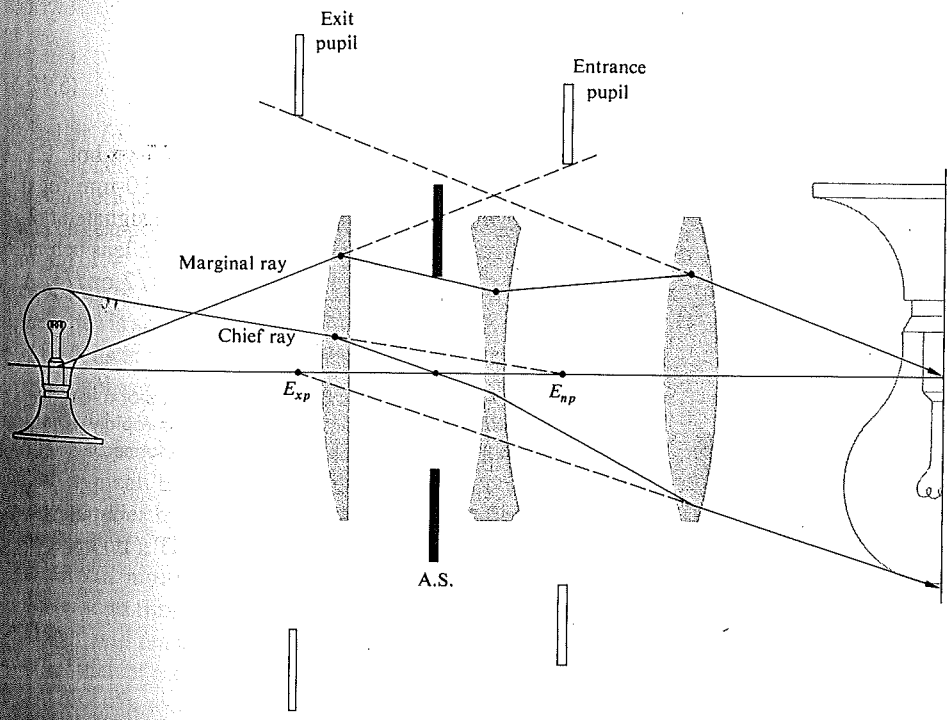


FIGURE 5.39 Pupils and stops for a three-lens system.

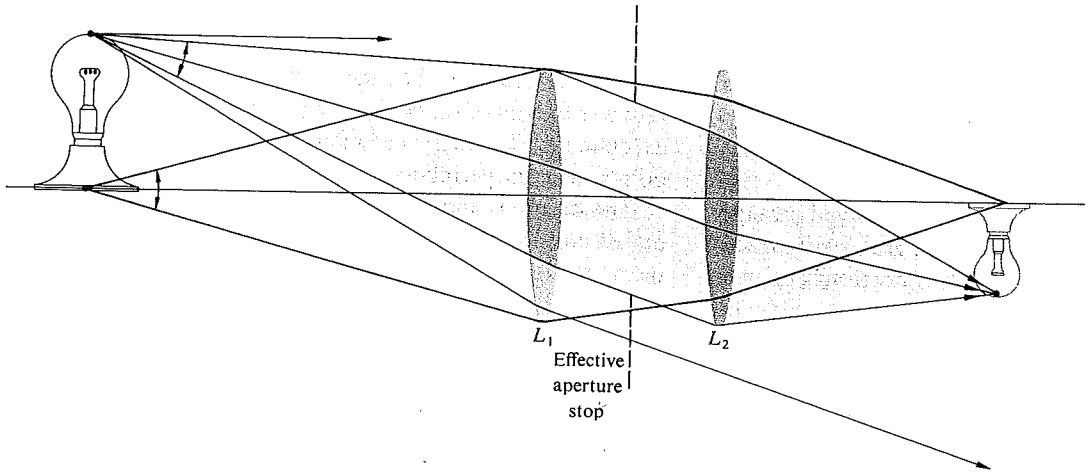


FIGURE 5.40 Vignetting.

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. 5.41). 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 5.42 illustrates the point by showing a thin lens behind a variable iris diaphragm operating at either $f/2$ or $f/4$. A

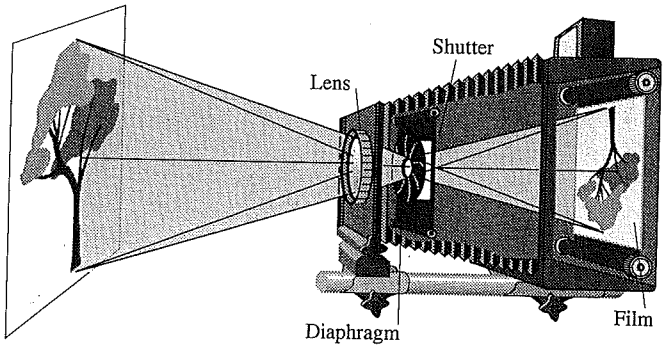


FIGURE 5.41 A large-format camera usually consists of a lens, followed by an adjustable diaphragm, a shutter that can rapidly open and close, and a sheet of film on which the image is formed.

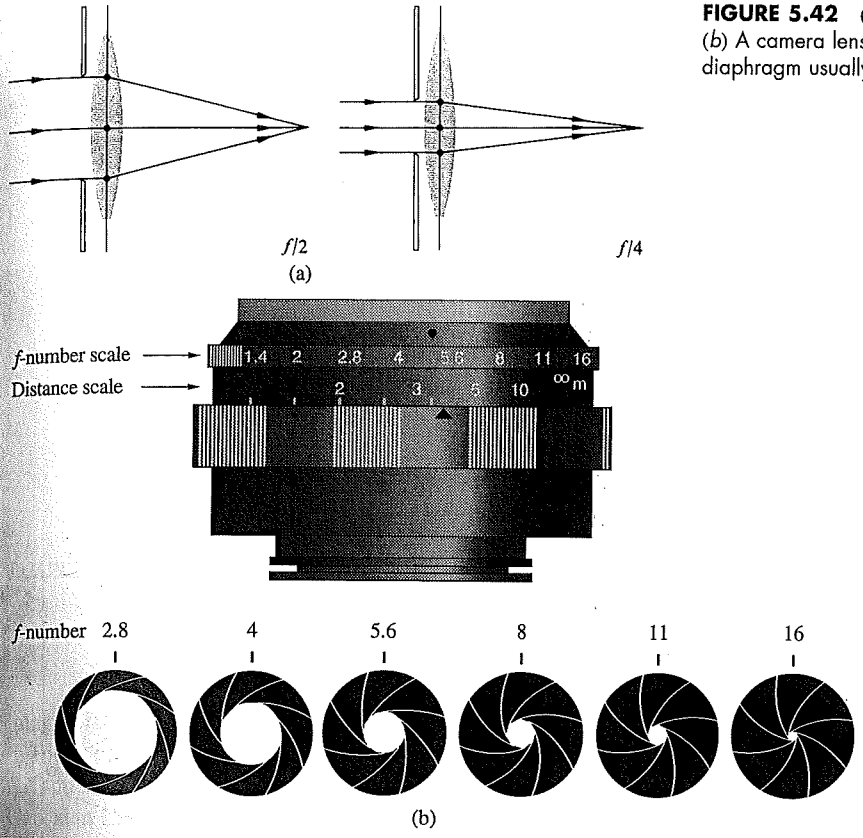


FIGURE 5.42 (*a*) Stopping down a lens to change the *f-number*. (*b*) A camera lens showing possible settings of the variable diaphragm usually located within the lens.

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 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 (Fig. 3.40). As we will see later (p. 260), 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. 4.62). 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. 5.43).

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. 5.43, we can quickly show that $|s_o| = |s_i|$; that is, the image P and object S are 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. 5.43 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

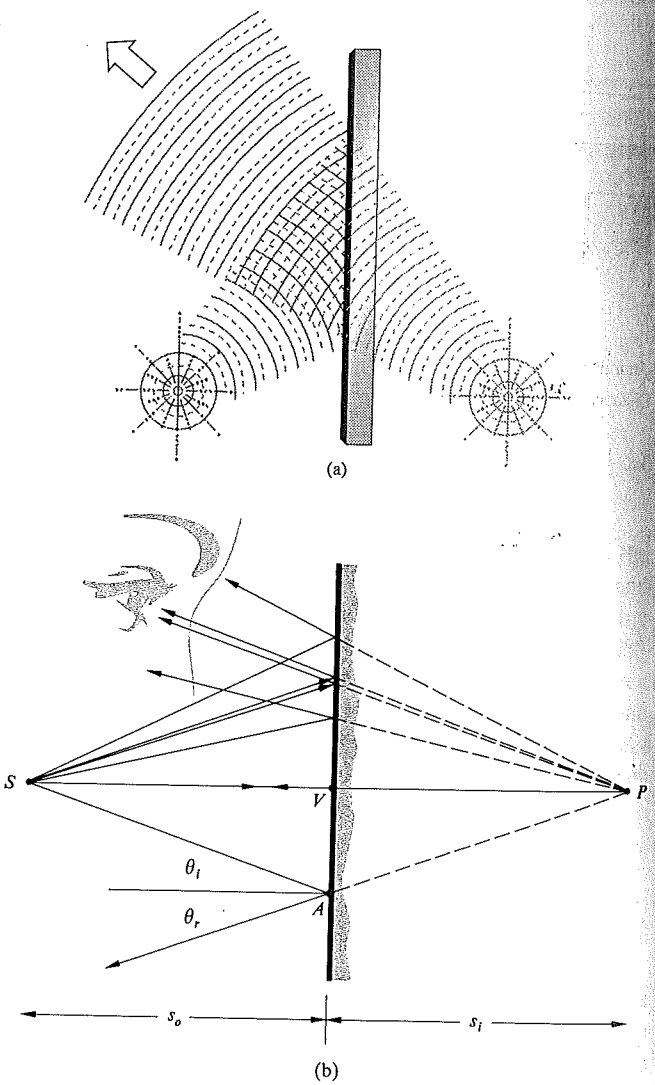


FIGURE 5.43 A planar mirror. (*a*) Reflection of waves. (*b*) Reflection of rays.

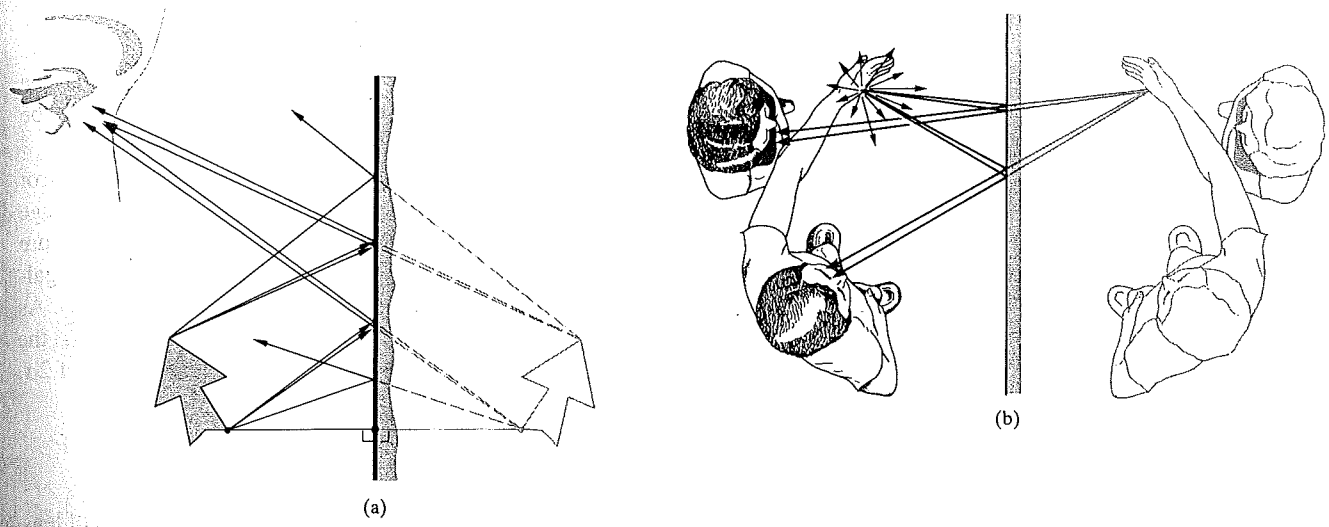


FIGURE 5.44 (*a*) The image of an extended object in a planar mirror. (*b*) Images in a planar mirror.

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. 5.44, 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. 5.30 was a left 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. 5.45). 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*. Sys-

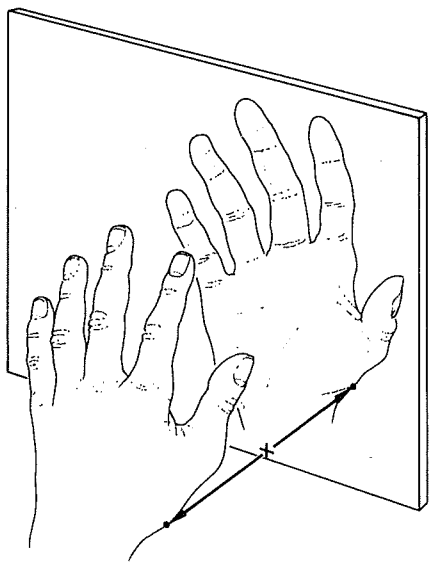
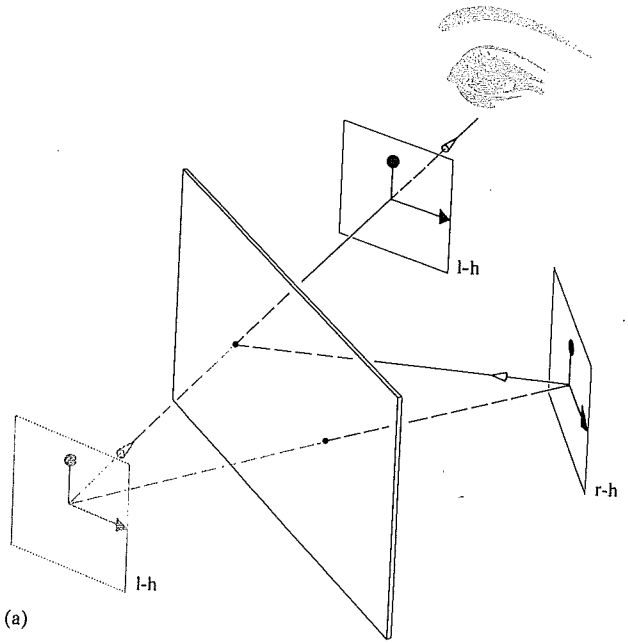
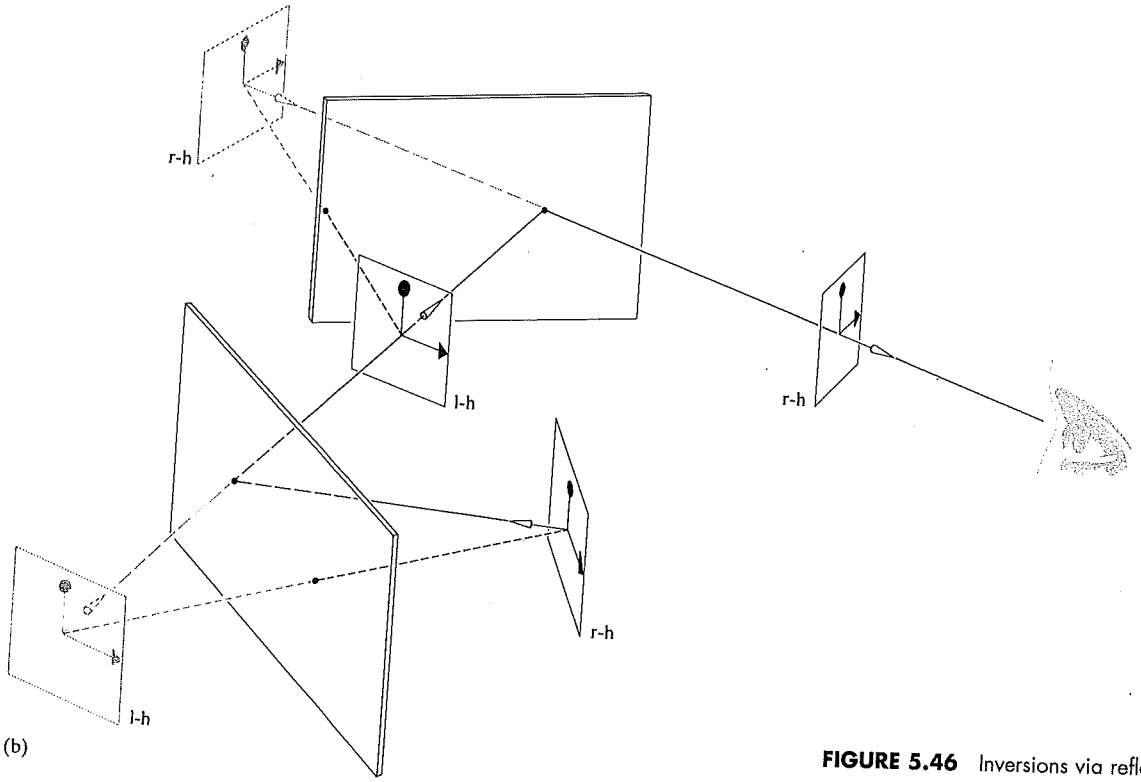


FIGURE 5.45 Mirror images—inversion.



(a)



(b)

FIGURE 5.46 Inversions via reflection.

tems 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. 5.46), 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, and image rotators. Mirrors are frequently used to amplify and measure the slight rotations of certain laboratory apparatus (galvanometers, torsion pendulums, current balances, etc.). As Fig. 5.47 shows, if the mirror rotates through an angle α, the reflected beam or image will move through an angle of 2α.

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

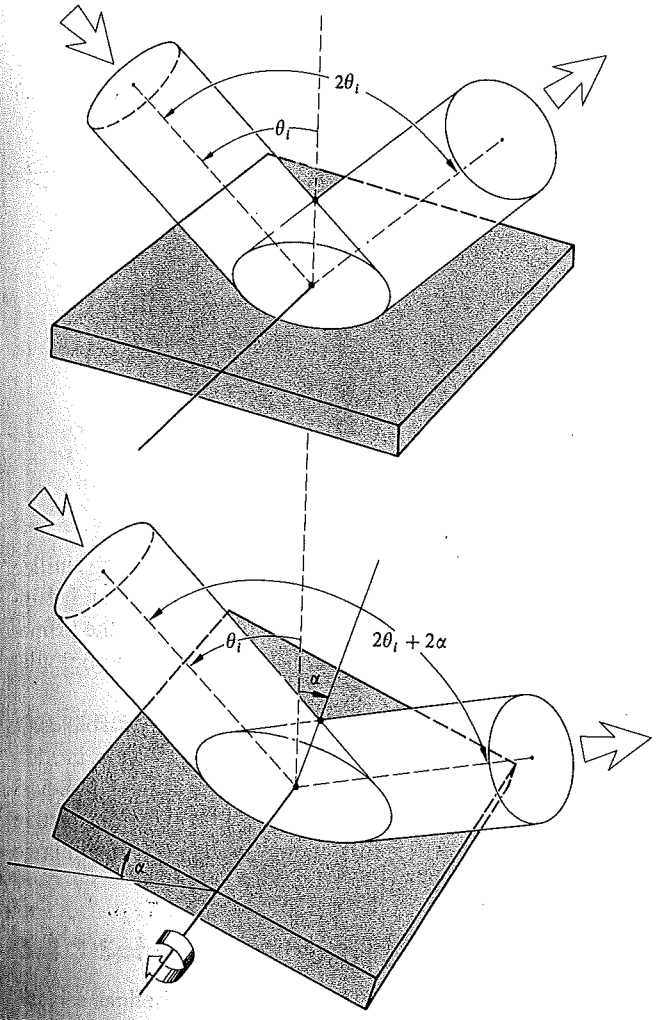


FIGURE 5.47 Rotation of a mirror and the concomitant angular displacement of a beam.

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. 5.48). 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**. 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 (Fig. 5.49), 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 Fig. 5.50 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. 5.51). As we'll see presently,

Dioptrics denotes the optics of refracting elements, whereas *catoptrics* denotes the optics of reflecting surfaces.

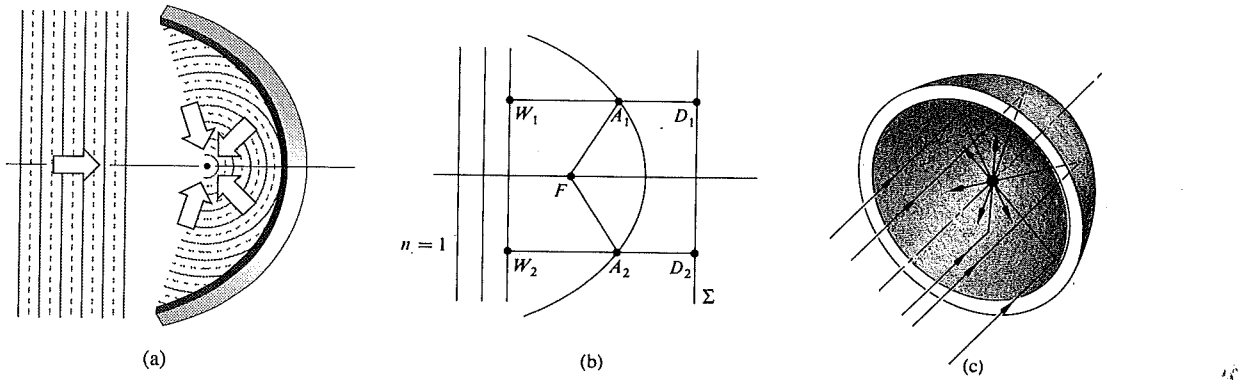


FIGURE 5.48 A paraboloidal mirror.

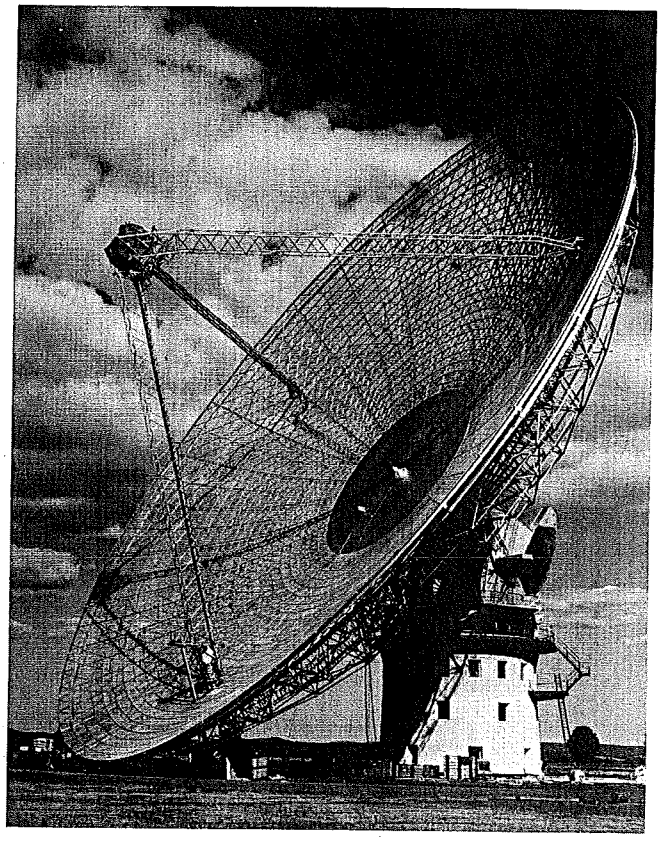


FIGURE 5.49 A classic paraboloidal radio antenna. (Photo courtesy of the Australian News and Information Bureau.)

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 (Fig. 5.52).

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. 5.53 the focused beam can be further processed without obstructing the mirror. Incidentally, this geometry also obtains in large microwave horn antennas.

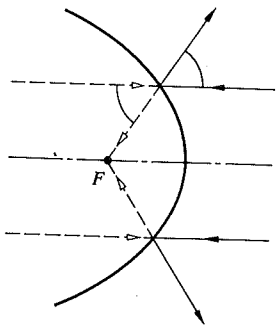


FIGURE 5.50 Real and virtual images for a paraboloidal mirror.

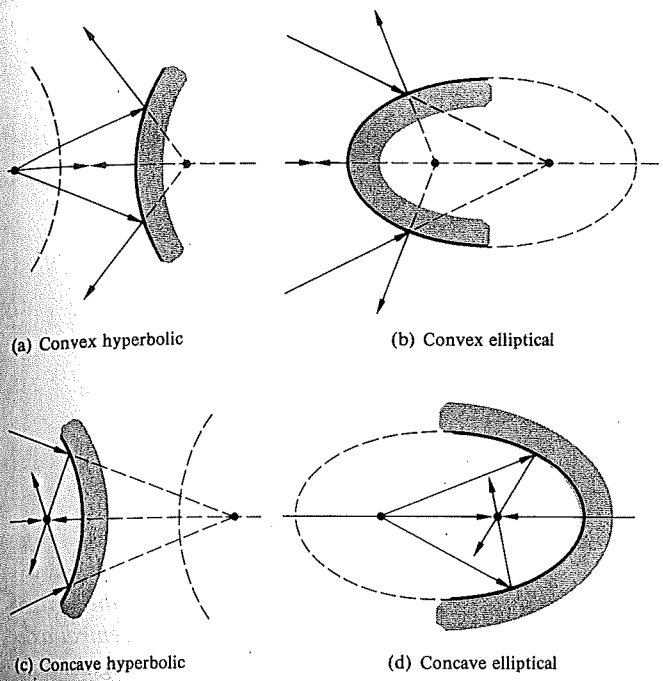


FIGURE 5.51 Hyperbolic and elliptical mirrors.

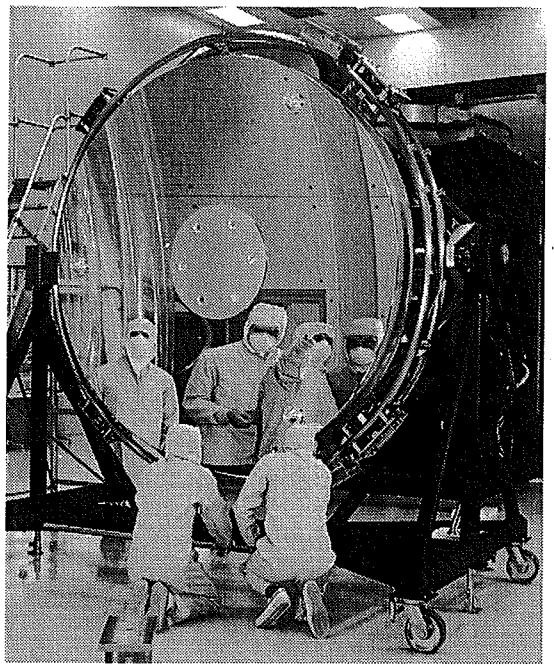


FIGURE 5.52 The 2.4-m-diameter hyperboloidal primary mirror of the Hubble Space Telescope. (Courtesy of NASA)

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. 5.54*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

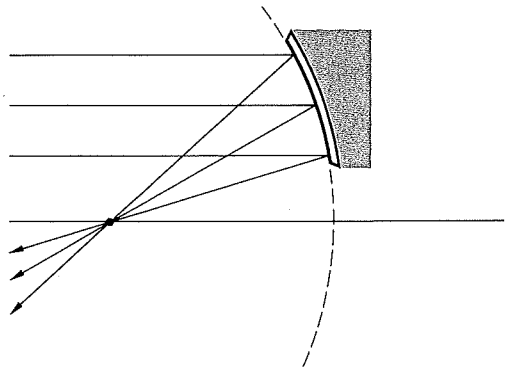


FIGURE 5.53 An off-axis parabolic mirror element.

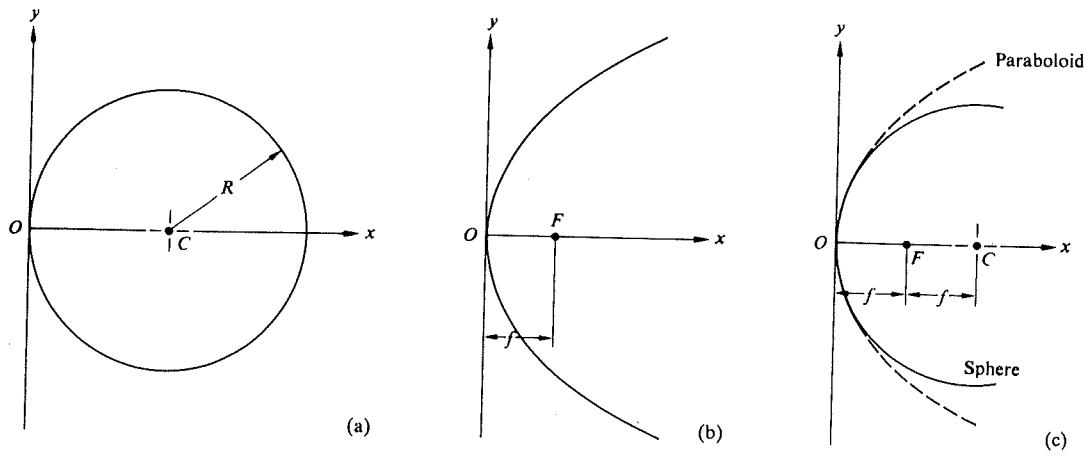


FIGURE 5.54 Comparison of spherical and paraboloidal mirrors.

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. 5.54*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. 5.54*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. 5.55. 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 proportional to the remaining two sides; that is,

$$\frac{\overline{SC}}{\overline{SA}} = \frac{\overline{CP}}{\overline{PA}} \tag{5.47}$$

Furthermore,

$$\overline{SC} = s_o - |R| \quad \text{and} \quad \overline{CP} = |R| - s_i$$

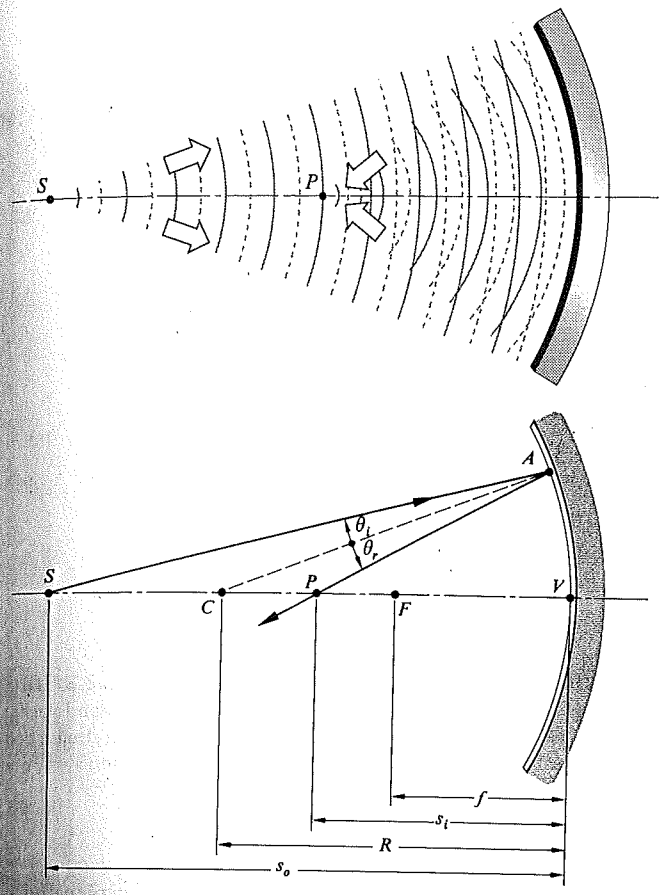


FIGURE 5.55 A concave spherical mirror. Conjugate foci.

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). 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)

$$\frac{1}{f_o} + \frac{1}{\infty} = \frac{1}{\infty} + \frac{1}{f_i} = -\frac{2}{R}$$

to wit,

$$f_o = f_i = -\frac{R}{2} \tag{5.49}$$

as can be seen in Fig. 5.54*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 (Figs. 5.56 and 5.57).

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

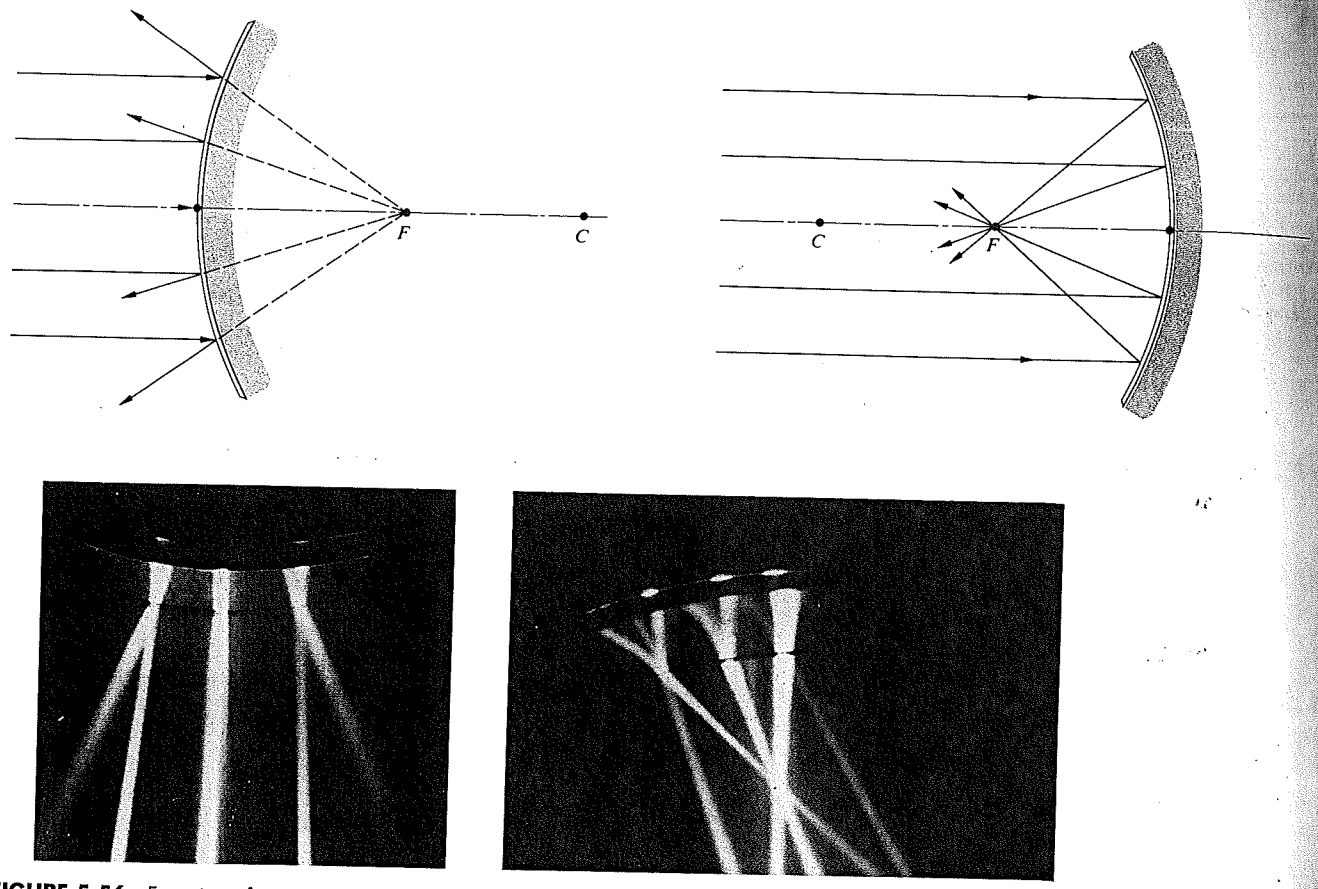


FIGURE 5.56 Focusing of rays via a spherical mirror. (E.H.)

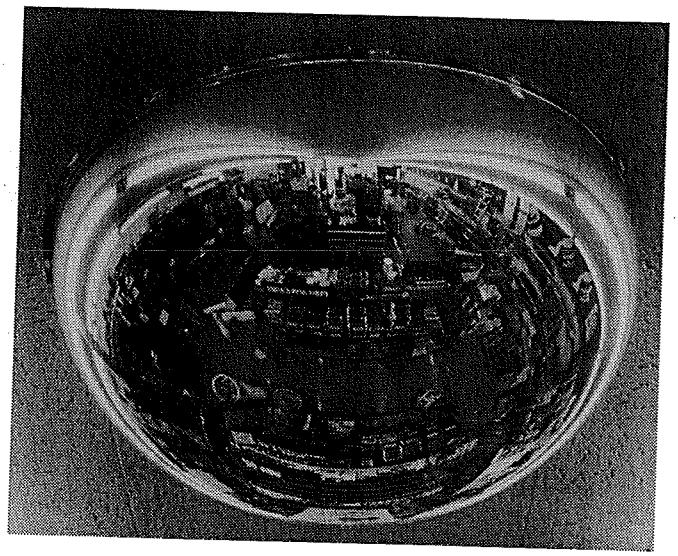


FIGURE 5.57 A convex spherical mirror forming a virtual, right-side-up, minified image. See if you can locate the image of the author.

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 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. 5.58). As with the thin lens (Fig. 5.24), the process for graphically locating the image is straightforward (Fig. 5.59). 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. 5.60). 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.

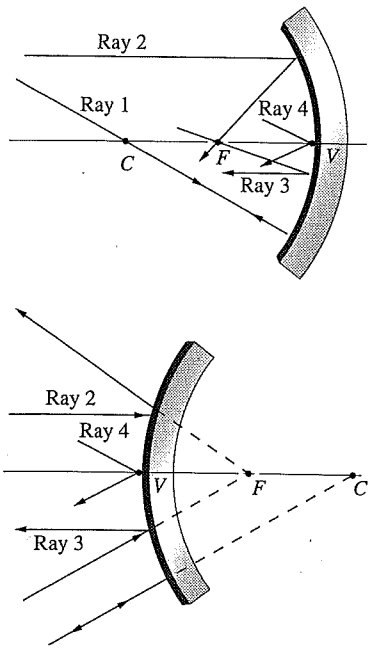


FIGURE 5.58 Four easy rays to draw. Ray 1 heads toward C and reflects back along itself. Ray 2 comes in parallel to the central axis and reflects toward (or away from) F. Ray 3 passes through (or heads toward) F and reflects off parallel to the axis. Ray 4 strikes point V and reflects such that $\theta_i = \theta_r$.

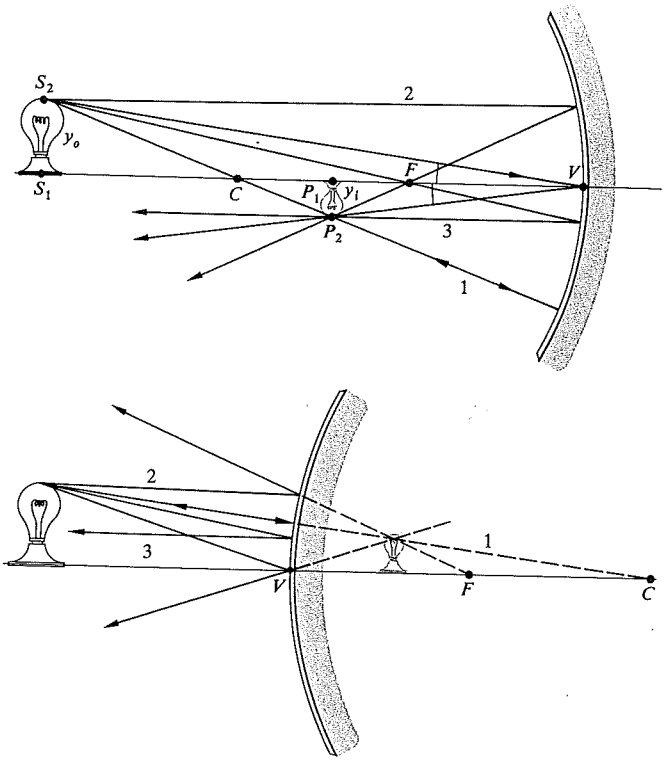


FIGURE 5.59 Finite imagery with spherical mirrors.

Notice that triangles S_1S_2V and P_1P_2V in Fig. 5.59*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

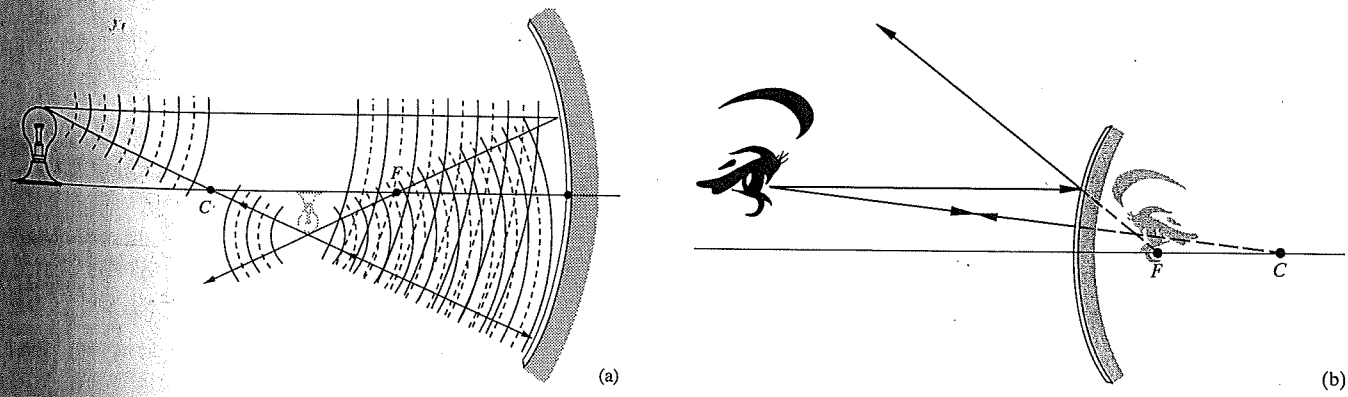


FIGURE 5.60 (*a*) Reflection from a concave mirror. (*b*) Reflection from a convex mirror.

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

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. 5.61 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 exper-

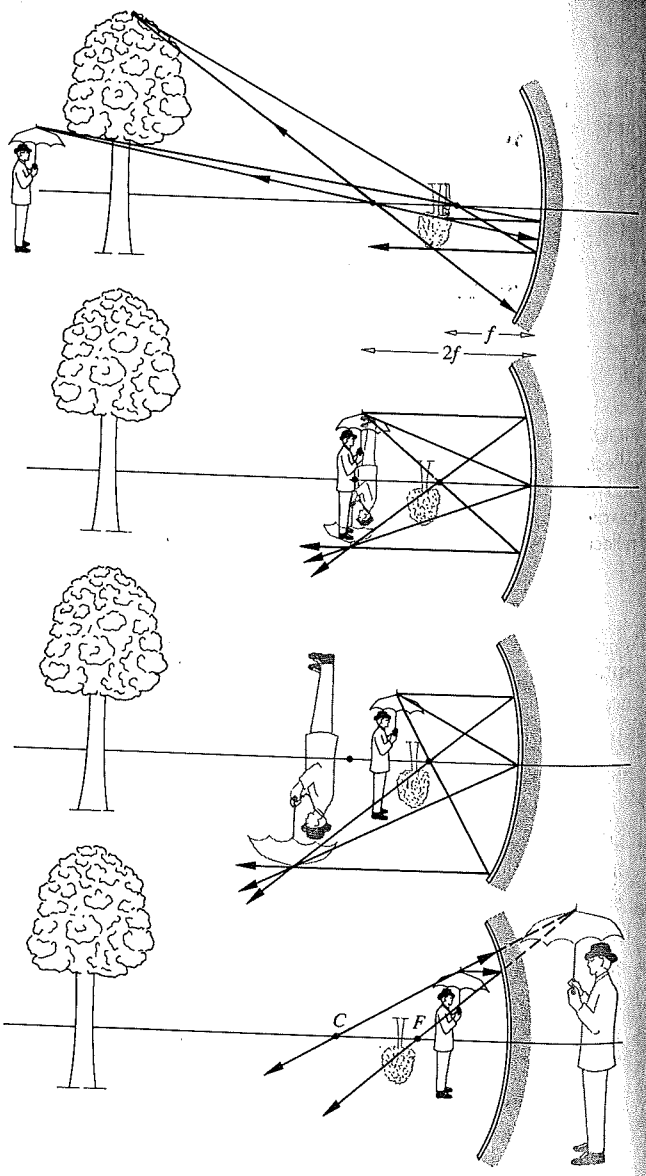


FIGURE 5.61 The image-forming behavior of a concave spherical mirror.

iment 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.

5.5 Prisms

Prisms play many different roles in Optics; there are prism combinations that serve as beamsplitters (p. 000), 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. 191). As such it is capable of separating, to some extent, the constituent frequency components in a polychromatic light beam. Recall that the term *dispersion* was introduced earlier (p. 67) 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.

5.5.1 Dispersing Prisms

Prisms come in many sizes and shapes and perform a variety of functions (Fig. 5.62). Let's first consider the group known as **dispersing prisms**. Typically, a ray entering a dispersing prism, as in Fig. 5.63, will emerge having been deflected from its original direction by an angle δ known as the **angular devi-**

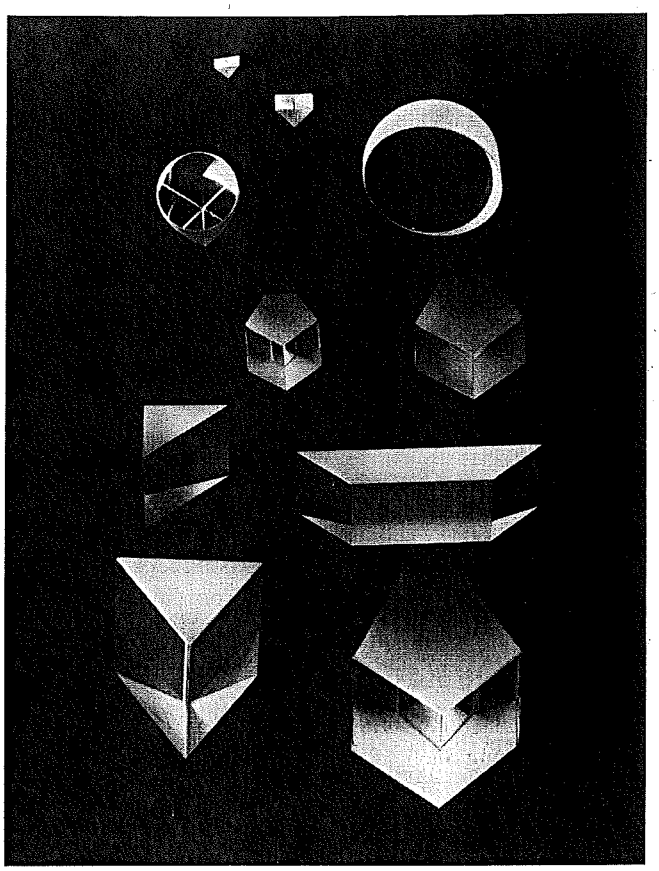


FIGURE 5.62 Prisms. (Photo courtesy Melles Griot.)

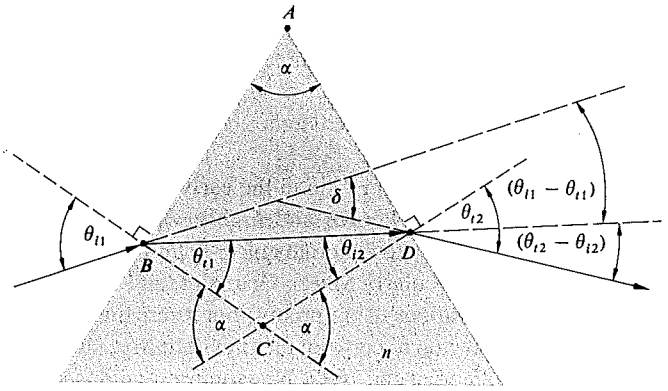


FIGURE 5.63 Geometry of a dispersing prism.