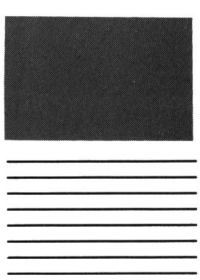


OPTICS

SECOND EDITION

EUGENE HECHT
Adelphi University

With Contributions by Alfred Zajac

ADDISON-WESLEY PUBLISHING COMPANY
Reading, Massachusetts ▪ Menlo Park, California ▪ Don Mills, Ontario
Wokingham, England ▪ Amsterdam ▪ Sydney ▪ Singapore
Tokyo ▪ Madrid ▪ Bogotá ▪ Santiago ▪ San Juan

Sponsoring editor: Bruce Spatz
Production supervisors: Margaret Pinette and Lorraine Ferrier
Text designer: Joyce Weston
Illustrators: Oxford Illustrators
Art consultant: Loretta Bailey
Manufacturing supervisor: Ann DeLacey

Library of Congress Cataloging-in-Publication Data

Hecht, Eugene.
Optics.

Bibliography: p.
Includes indexes.
1. Optics. I. Zajac, Alfred. II. Title.
QC355.2.H42 1987 535 86-14067
ISBN 0-201-11609-X

Reprinted with corrections, February 1989.

DEFGHIJK-HA-89

5 GEOMETRICAL OPTICS—PARAXIAL THEORY

5.1 INTRODUCTORY REMARKS

Suppose we have an object that is either self-luminous or externally illuminated, and imagine its surface as consisting of a large number of point sources. Each of these emits spherical waves, that is, rays emanate radially in the direction of energy flow or, if you like, in the direction of the Poynting vector (Fig. 4.1). 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 focal point of the bundle of rays.*

Now envision the situation in which we have 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. Thus, 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 (see Section 4.2.4) 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; the system accepts only a segment of the wavefront. As a result, there will always be an

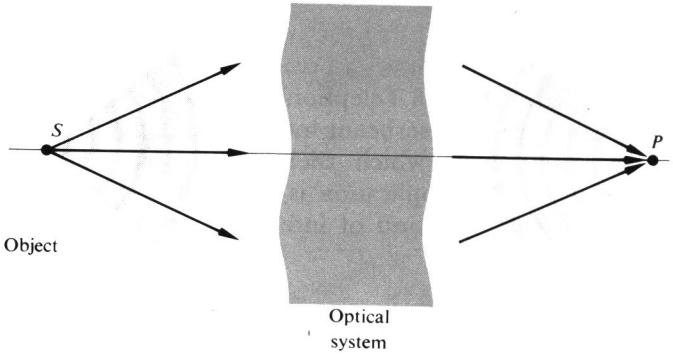


Figure 5.1 Converging and diverging waves.

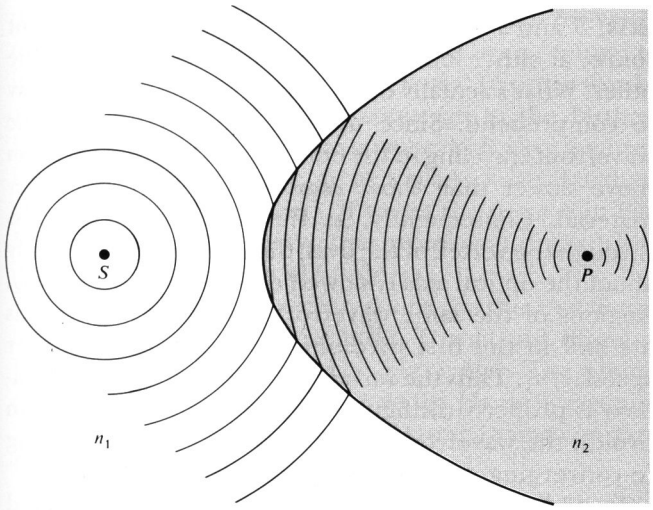


Figure 5.2 Reshaping a spherical wave at a refracting interface ($n_1 < n_2$).

apparent deviation from rectilinear propagation even in homogeneous media—the waves will be *diffracted.* The attainable degree of perfection in the imaging capability of a real optical system will therefore be **diffraction-limited** (there will always be a blur spot). 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. There are many situations in which 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* (see Chapter 13).

5.2 LENSES

No doubt the most widely used optical device is the lens, and that notwithstanding the fact that we see the world through a pair of them. Lenses date back to the burning glasses of antiquity, and indeed who can say when people first peered through the liquid lens formed by a droplet of water?

As an initial step toward an understanding of what lenses do and how they manage to do it, let's examine what happens when light impinges on the curved surface of a transparent dielectric medium.

5.2.1 Refraction at Aspherical Surfaces

Imagine that we have a point source S whose spherical waves arrive at a boundary between two transparent media, as shown in Fig. 5.2. We would like to determine the shape that the interface must have for the wave traveling within the second medium to converge at a point P, there forming a perfect image of S. Practical reasons for wanting to focus a diverging wave to a point will become evident as we proceed.

The time it takes for each and every portion of a wavefront leaving S to converge at P must be identical, if a perfect image is to be formed—that much was implied by Huygens in 1678. Or as we saw in Section

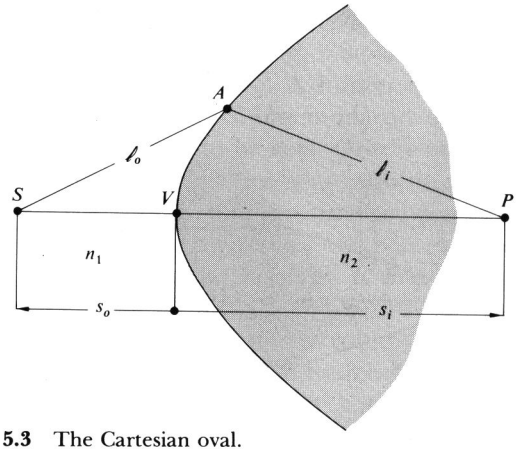


Figure 5.3 The Cartesian oval.

4.2.3, the distance between corresponding points on any and all rays will be traversed in that same time. Another way to say essentially the same thing from the perspective of Fermat's principle is that if a great many different rays are to go from S to P (i.e., if point A in Fig. 5.3 can be anywhere on the interface), each ray must traverse the same optical path length. Thus, for example, if S is in a medium of index n_1 and P is in an optically more dense medium of index n_2,

$$\ell_o n_1 + \ell_i n_2 = s_o n_1 + s_i n_2, \tag{5.1}$$

where s_o and s_i are the **object** and **image distances** measured from the *vertex* or *pole* V, respectively. Once we choose s_o and s_i, the right-hand side of this equation becomes fixed, and so

$$\ell_o n_1 + \ell_i n_2 = \text{constant}. \tag{5.2}$$

This is the equation of a *Cartesian oval* whose significance in optics was studied extensively by René Descartes in the early 1600s (Problem 5.1). Hence, when the boundary between two media has the shape of a Cartesian oval of revolution about the $\overline{SP}$, or **optical axis**, S and P will be conjugate points, that is, a point source at either location will be perfectly imaged at the other. What's actually occurring physically is rather easy to comprehend. Since $n_2 > n_1$, those regions of the wavefront traveling in the optically more dense medium move slower than those regions traversing the rarer material. Consequently, as the wave begins to pass through the vertex of the oval, the segment immediately about the optical axis is slowed down from c/n_1 to c/n_2. Regions of the same wavefront remote from the axis are still in the first medium traveling with a greater speed, c/n_1. Thus the wavefronts bend, and if the boundary is properly configured (in the form of a Cartesian ovoid), the wavefronts will be inverted from diverging to converging spherical segments.

In addition to focusing a spherical wave, we would like to be able to perform a few other reshaping operations using refracting interfaces; some of these are illustrated in Fig. 5.4. We shall consider them only briefly and more for pedagogical than practical reasons. The surfaces in Fig. 5.4(a) and (b) are ellipsoidal, whereas those in (c) and (d) are hyperboloidal. Notice

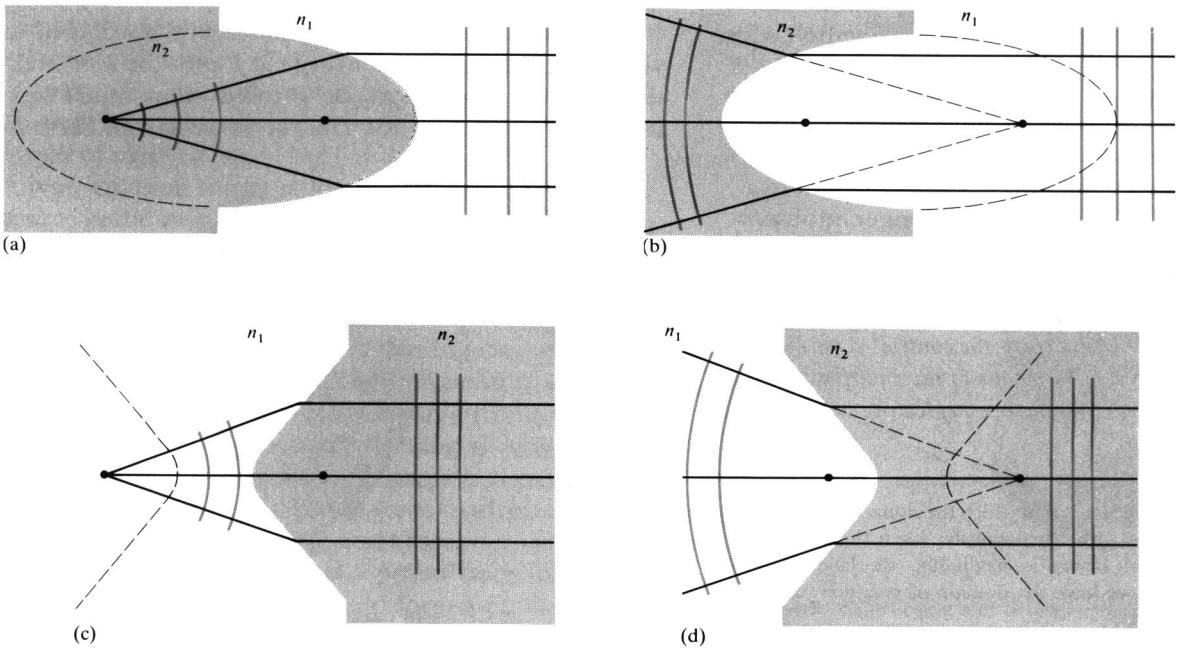


Figure 5.4 Ellipsoidal and hyperboloidal refracting surfaces ($n_2 > n_1$).

that in all cases, the rays either diverge from or converge toward the foci. The arrowheads have been omitted to indicate that the rays can go either way. In other words, an incident plane wave will converge to the farthest focus of an ellipsoid just as a spherical wave emitted from that focus will emerge as a plane wave. Furthermore, as you might expect, if we let the point S in Fig. 5.2 move out to infinity, the ovoid would gradually metamorphose into an ellipsoid.

Rather than deriving expressions for these surfaces, let's just justify the above remarks. To that end, examine Fig. 5.5, which relates back to Fig. 5.4(a). The optical path lengths from any point D on the planar wavefront Σ to the focus F_1 must all be equal to the same constant C, that is,

$$(\overline{F_1A})n_2 + (\overline{AD})n_1 = C$$

or

$$(\overline{F_1A}) + (\overline{AD})n_{12} = C/n_2. \tag{5.3}$$

To see that this relationship is indeed satisfied by an ellipsoid of revolution, recall that if Σ corresponds to the directrix of the ellipse, $(\overline{F_2A}) = e(\overline{AD})$, where e is the eccentricity. Thus if $e = n_{12}$, the left-hand side of Eq. (5.3) becomes $(\overline{F_1A}) + (\overline{F_2A})$, which is certainly constant for an ellipse. Here the eccentricity is less than 1 ($e = n_1/n_2$) and it is left for Problem 5.2 to show that had it been greater than 1 (i.e., $n_1 > n_2$), the curve would have been a hyperbola instead [compare (a) with (c) and (b) with (d) in Fig. 5.4]. If all this brings back memories of analytic geometry, you might keep in mind that that subject was originated by Descartes. Interestingly, it was Kepler who first (1611) suggested using conic sections for mirrors and lenses.

The knowledge we have at hand now may be used to construct lenses such that both the object and image points can be in the same medium, which is usually air. The first such device to be considered [Fig. 5.6(a)] is a *double convex hyperbolic* lens, which utilizes the response characterized in Fig. 5.4(c). A diverging spherical wave becomes planar after traversing the first hyperbolic surface and then spherically converging on leaving the lens. Alternatively, if the second surface is made planar so that we have a *hyperbolic planar convex* lens, as in Fig. 5.6(b), the plane waves within the lens will strike the back surface perpendicularly and emerge unaltered.

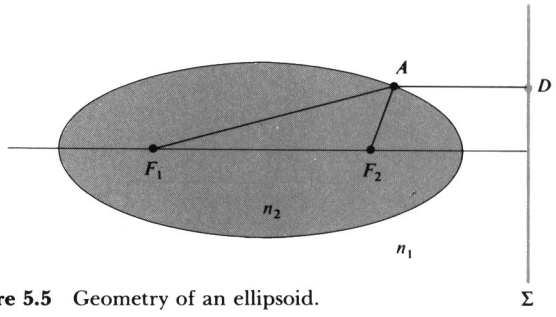
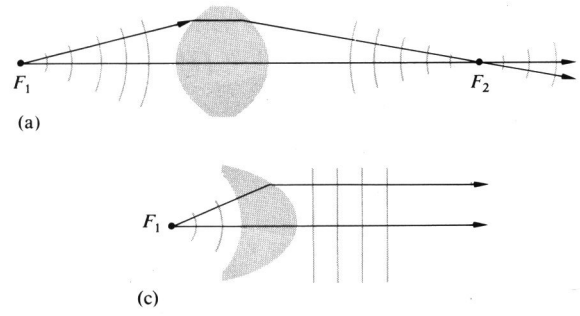


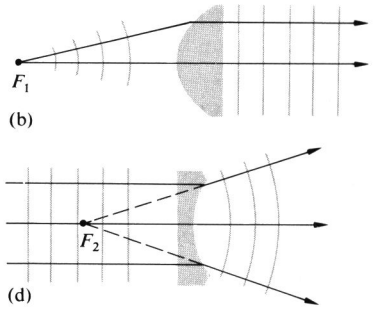
Figure 5.5 Geometry of an ellipsoid.

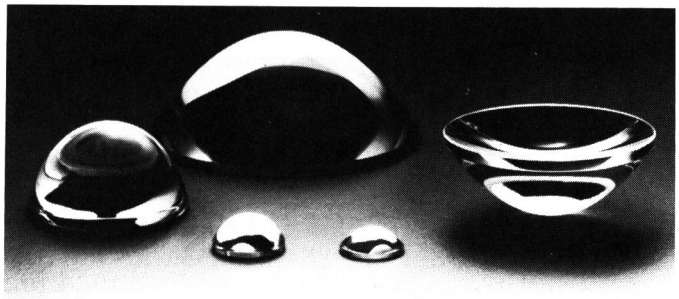
Another arrangement that will convert diverging spherical waves into plane waves is illustrated in Fig. 5.6(c). This is a *sphero-elliptic convex* lens, where F_1 is simultaneously at the center of the spherical surface and at the focus of the ellipsoid. Rays from F_1 strike the first surface perpendicularly and are therefore undeviated by it. As in Fig. 5.4(a), the exiting wavefronts are planar. All the elements thus far examined have been thicker at their midpoints than at their edges and are for that reason said to be *convex* (from the Latin *convexus*, meaning arched). In contrast, the *planar hyperbolic concave* lens (from the Latin *concavus*, meaning hollow, and easily remembered because it contains the word cave) is thinner at the middle than at the edges, as is evident in Fig. 5.6(d). A number of other arrangements are possible, and a few will be considered in the problems (5.3). Note that each of these lenses will work just as well in reverse: the waves shown emerging can instead be thought of as entering from the right.

If a point source is positioned on the optical axis at the point F_1 of the lens in Fig. 5.6(a), 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. 5.6(d) 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







(e)

Figure 5.6 (a) A double hyperbolic lens. (b) A hyperbolic planar convex lens. (c) A sphero-elliptic lens. (d) A planar hyperbolic lens. (e) Photo courtesy Melles Griot.

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 extensively 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). And the small-scale production of diffraction-limited molded-glass aspheric lenses has been reported in recent times. 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 (0.000020 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.

5.2.2 Refraction at Spherical Surfaces

Imagine that we have 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 cur-

Figure 5.7 Polishing a spherical lens. (Photo courtesy Optical Society of America.)

vature, 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. 5.7). Such surfaces are now commonly generated in batches by automatic grinding and polishing machines. In contrast, high-quality aspherical shapes require considerably more effort to produce.

It should therefore come as no surprise that the vast majority of quality lenses in use today have spherical surfaces. Our intent here is to establish techniques for using such surfaces whereby a great many object points can be satisfactorily imaged simultaneously in light composed of a broad frequency range. 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.

Now that we know why and where we are going, let's move on. Figure 5.8 depicts a wave from the point source S impinging on a spherical interface of radius R centered at C. The ray $(\overline{SA})$ will be refracted at the interface toward the local normal ($n_2 > n_1$) and therefore toward the optical axis. Assume that at some point P it will cross the axis, as will all other rays incident at the same angle θ_i (Fig. 5.9). 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,

$$(\text{OPL}) = n_1\ell_o + + n_2\ell_i. \tag{5.4}$$

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

$$(\text{OPL}) = n_1[R^2 + (s_o + R)^2 - 2R(s_o + R)\cos\varphi]^{1/2} + n_2[R^2 + (s_i - R)^2 + 2R(s_i - R)\cos\varphi]^{1/2}.$$

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

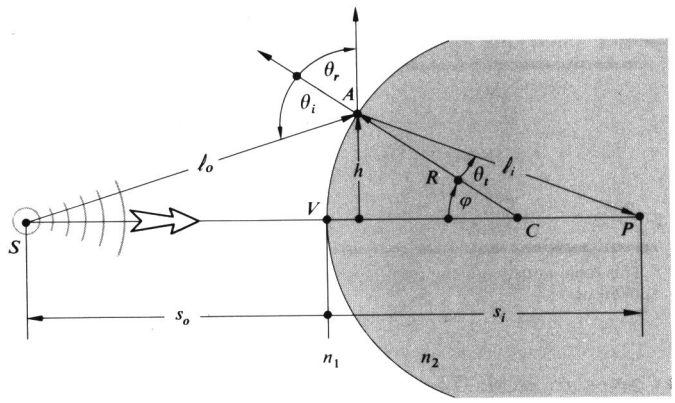


Figure 5.8 Refraction at a spherical interface.

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 $d(\text{OPL})/d\varphi = 0$, via Fermat's principle we have

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

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. We already know that if A is moved to a new location by changing φ, the new ray will not intercept the optical axis at P—this is not a

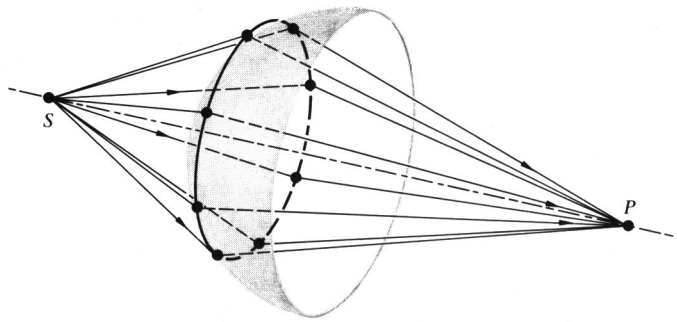


Figure 5.9 Rays incident at the same angle.

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.

Cartesian oval. The approximations that are used to represent ℓ_o and ℓ_i, and thereby simplify Eq. (5.5), are crucial in all that is to follow. Recall that

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

and

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

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

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

We could have begun this derivation with Snell's law rather than Fermat's principle (Problem 5.4), 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. 5.10 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,

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

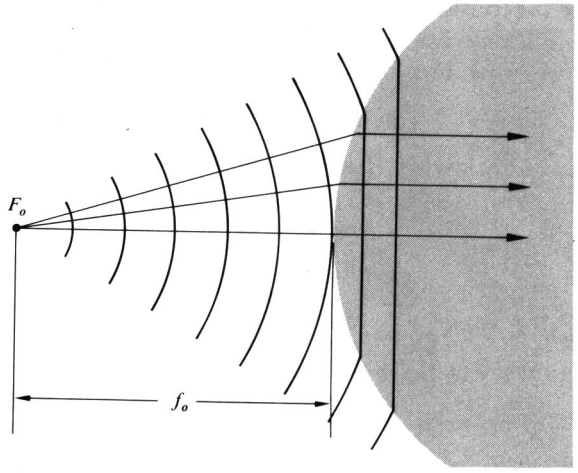


Figure 5.10 Plane waves propagating beyond a spherical interface—the object focus.

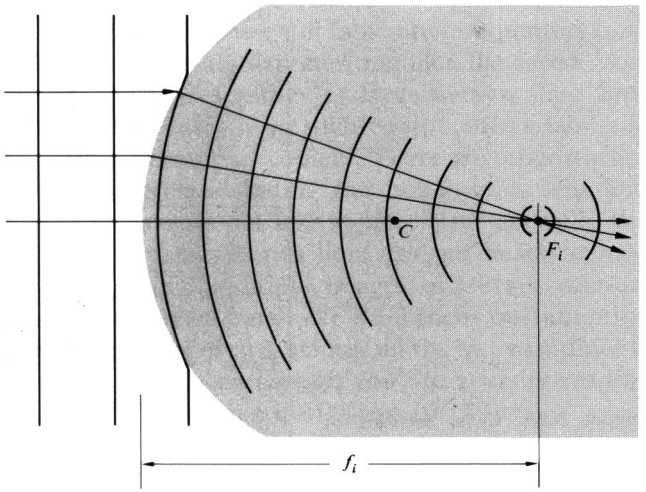


Figure 5.11 The reshaping of plane into spherical waves at a spherical interface—the image focus.

Defining the *second or image focal length* f_i as equal to s_i in this special case (Fig. 5.11), we have

$$f_i = \frac{n_2}{n_2 - n_1} R. \qquad (5.10)$$

Recall that an image is virtual when the rays diverge from it (Fig. 5.12). Analogously, *an object is virtual when the rays converge toward it* (Fig. 5.13). Observe that the virtual object is now on the right-hand side of the vertex, and therefore s_o will be a negative quantity. Moreover, the surface is concave, and its radius will also be negative, as required by Eq. (5.9), since f_o would be negative. In the same way the virtual image distance appearing to the left of V is negative.

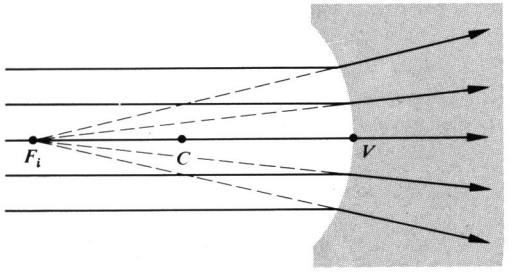


Figure 5.12 A virtual image point.

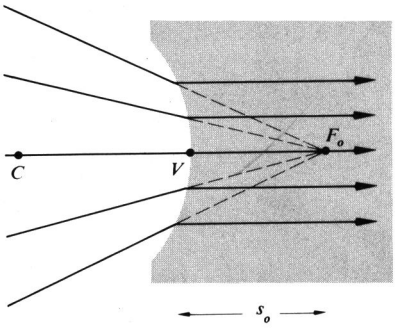


Figure 5.13 A virtual object point.

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. 5.14). In the traditional sense, *a lens is an optical system consisting of two or more refracting interfaces, at least one of which is curved.* Generally the nonplanar surfaces are centered on a common axis. These surfaces are most frequently spherical segments and are often coated with thin dielectric films to control their transmission properties (see Section 9.9). A lens that consists of one element (i.e., it has only two refracting surfaces) is a *simple lens.* The presence of more than one element makes it a *compound lens.* A lens is also classified as to whether it is *thin* or *thick*, that is, whether its thickness is effectively negligible or not. 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 diverse forms shown in Fig. 5.15. 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,

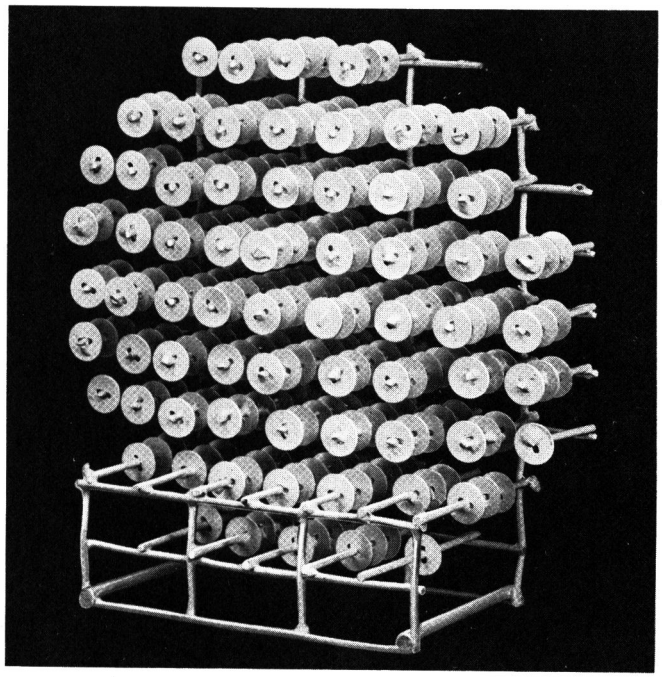
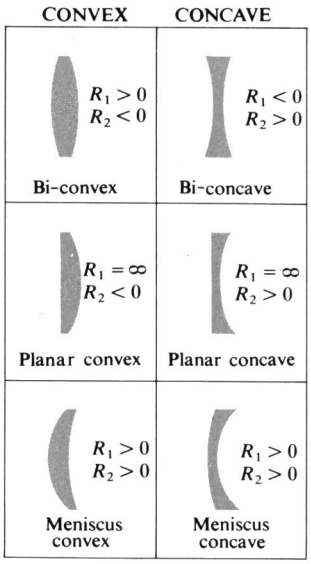
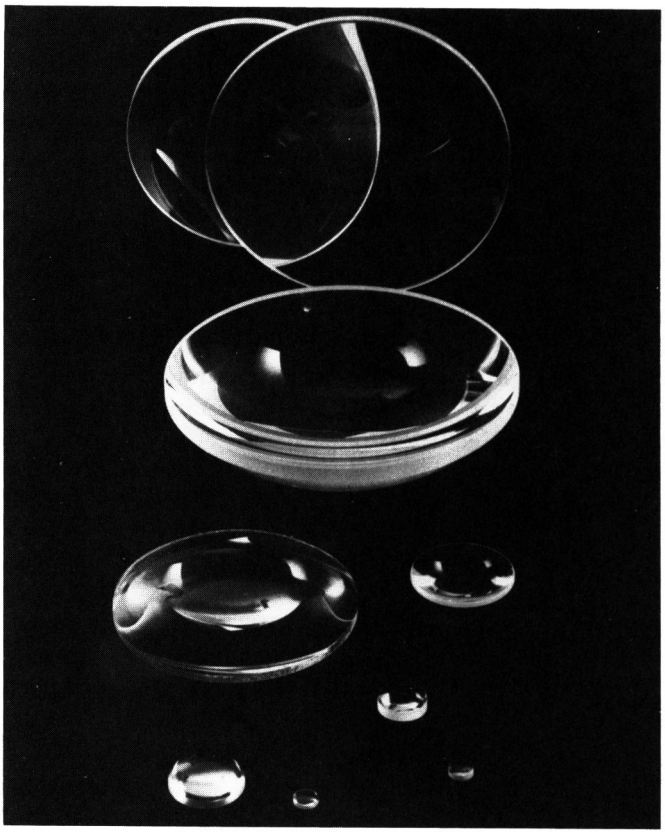


Figure 5.14 A lens for short-wavelength radiowaves. The disks serve to refract these waves much as rows of atoms refract light. (Photo courtesy Optical Society of America.)



(a)

Figure 5.15 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.)



(b)

are thinner at the center and tend to advance that portion of the wavefront, causing it to diverge more then it did upon entry.

In the broadest sense, *a lens is a refracting device that is used to reshape wavefronts in a controlled manner.* Although this is usually done by passing the wave through at least one specially shaped interface separating two different homogeneous media, it is not the only approach available. For example, it is also possible to reconfigure a wavefront by passing it through an inhomogeneous medium. A *gradient-index*, or GRIN, lens is one where the desired effect is accomplished by using a medium in which the index of refraction varies in a prescribed fashion. Different portions of the wave propagate at different speeds, and the front changes shape as it progresses. In the commercial GRIN material (available only since 1976) the index varies radially, decreasing parabolically out from the central axis.

Today GRIN lenses are still fabricated in quantity only in the form of small-diameter, parallel, flat-faced rods. Usually grouped together in large arrays, they have been used extensively in such equipment as facsimile machines and compact copiers. There are other unconventional lenses, including the holographic lens and even the gravitational lens (where, for example, the gravity of a galaxy bends light passing in its vicinity, thereby forming multiple images of distant celestial objects, such as quasars). We shall focus our attention in the remainder of this chapter on the more traditional types of lenses, even though you are actually reading these words through a GRIN lens (p.179).

i) Thin-Lens Equations

Return for a moment 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. 5.16). Let's now locate the conjugate points for the lens of index n_l surrounded by a medium of index n_m, as in Fig. 5.17, where another end has simply been ground on the piece in Fig. 5.16(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

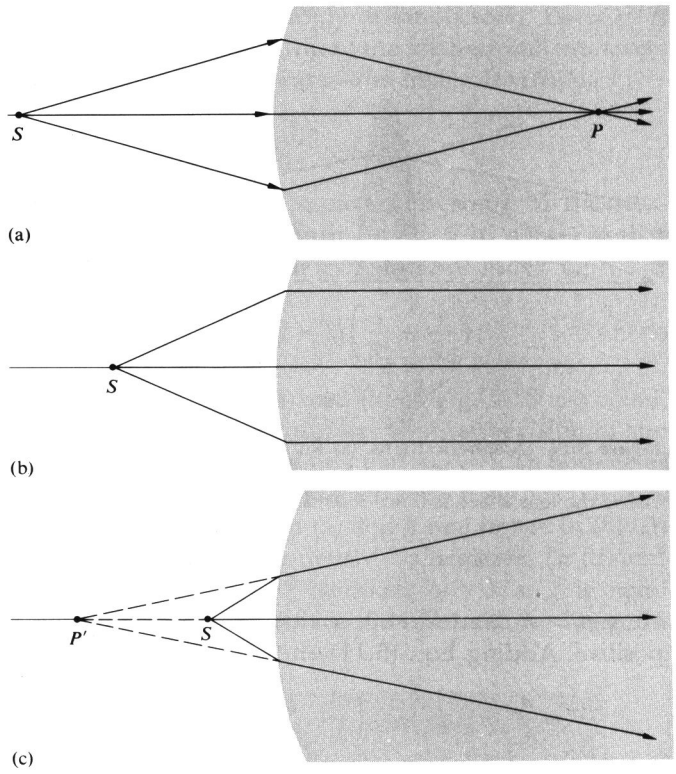


Figure 5.16 Refraction at a spherical interface.

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

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

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

$$s_{o2} = -s_{i1} + d. \qquad (5.12)$$

Thus at the second surface Eq. (5.8) yields

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

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

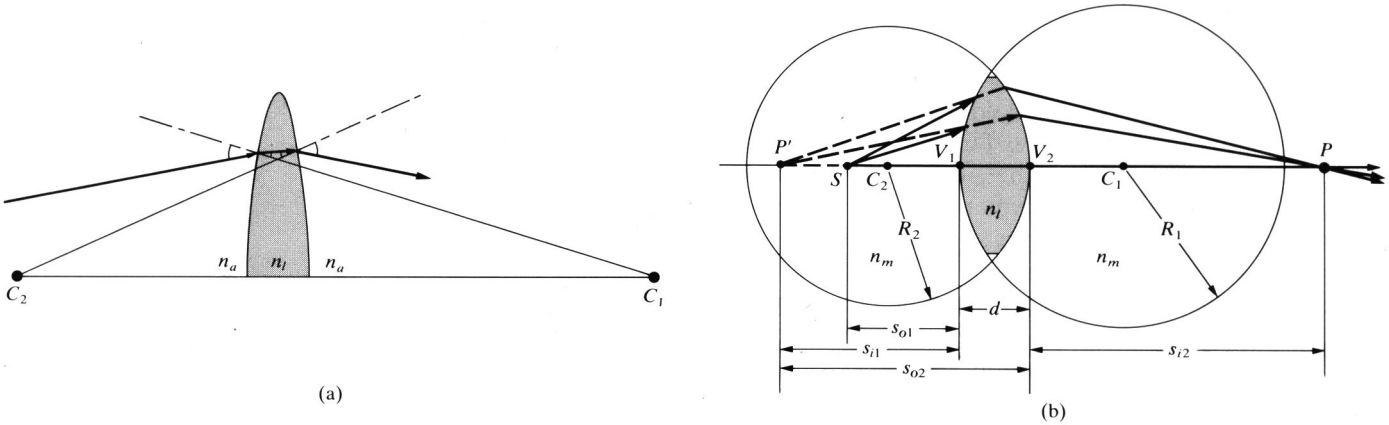


Figure 5.17 A spherical lens. (a) Refraction at the interfaces. The radius drawn from C_1 is normal to the first surface, and as the ray enters the lens it bends down *toward* that normal. The radius from C_2 is normal to the second surface; and as the ray emerges, since $n_l > n_a$, the ray bends down *away* from that normal. (b) The geometry.

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

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

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

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

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

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

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

Similarly

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

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

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

and

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

which is the famous **Gaussian lens formula**. 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). Hence

$$\frac{1}{600} + \frac{1}{s_i} = \frac{1}{100}$$

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 that uses auto lamps is probably the most convenient. Standing as far as you can from the source, project a clear image of it onto a white sheet of paper. You should be able to see the lamp quite clearly and not just as 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 filament 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.

ii) 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 in a medium of index n_m,

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

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

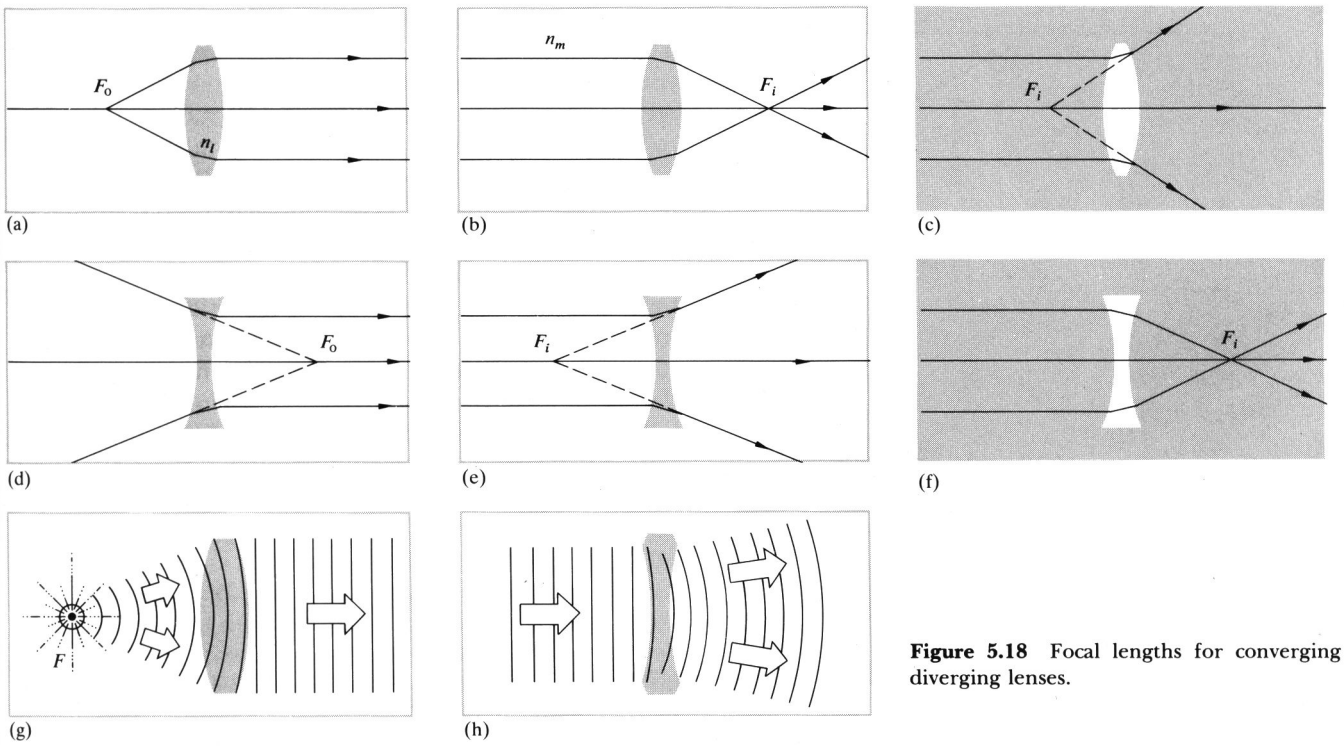
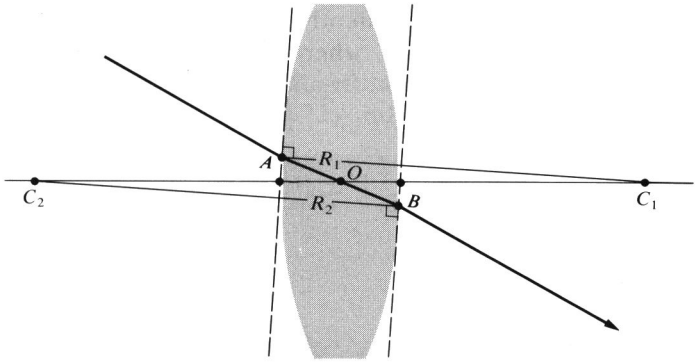


Figure 5.18 Focal lengths for converging and diverging lenses.

the other are virtual. The last situation shows $n_{lm} < 1$, 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, however, 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* of the lens O. To see this, draw two parallel planes, one on each side tangent to the lens at any pair of points A and B. This can easily be done by selecting A and B such that the radii $\overline{AC_1}$ and $\overline{BC_2}$ are themselves parallel. It is to be shown that the paraxial ray traversing $\overline{AB}$ enters and leaves the lens in the same direction. It is 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 4.19 and Fig. 4.55), 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.



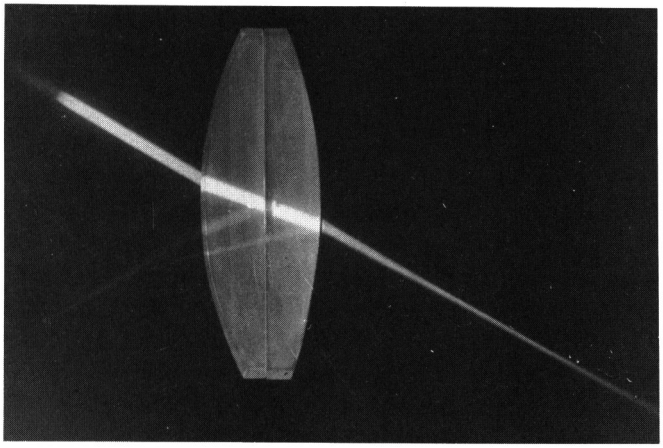


Figure 5.19 The optical center of a lens. (Photo by E.H.)

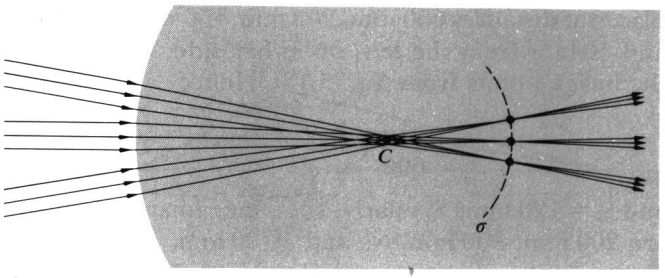


Figure 5.20 Focusing of several ray bundles.

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.

* 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. 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.

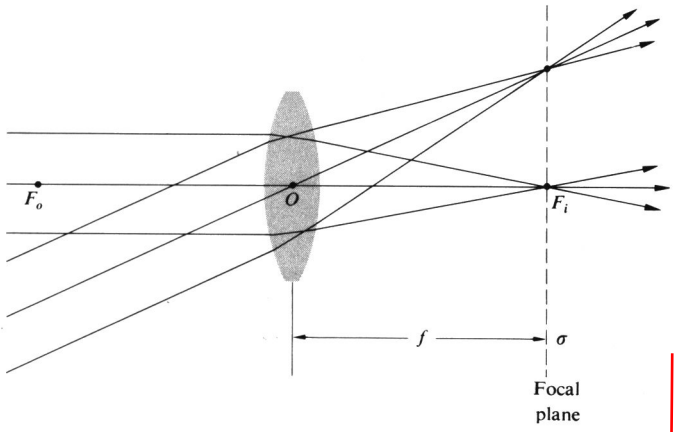


Figure 5.21 The focal plane of a lens.

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 next—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. Since 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 optical axis and vice versa; the third is the undeviated ray through

iii) Finite Imagery

Thus far we've dealt with the mathematical abstraction of a single-point source, but now let's suppose that a great many such points combine to form a continuous finite object. For the moment, imagine the object to be a segment of a 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. It should be noted 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, just as was done in Section (i). 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

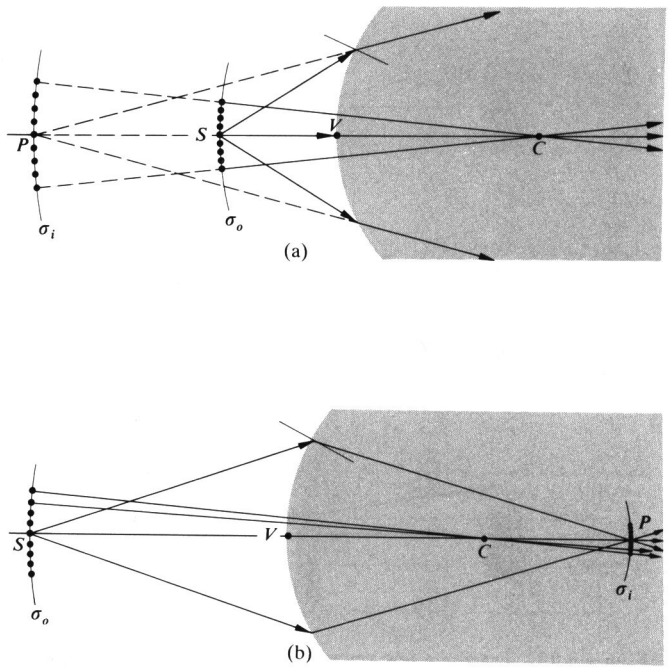


Figure 5.22 Finite imagery.

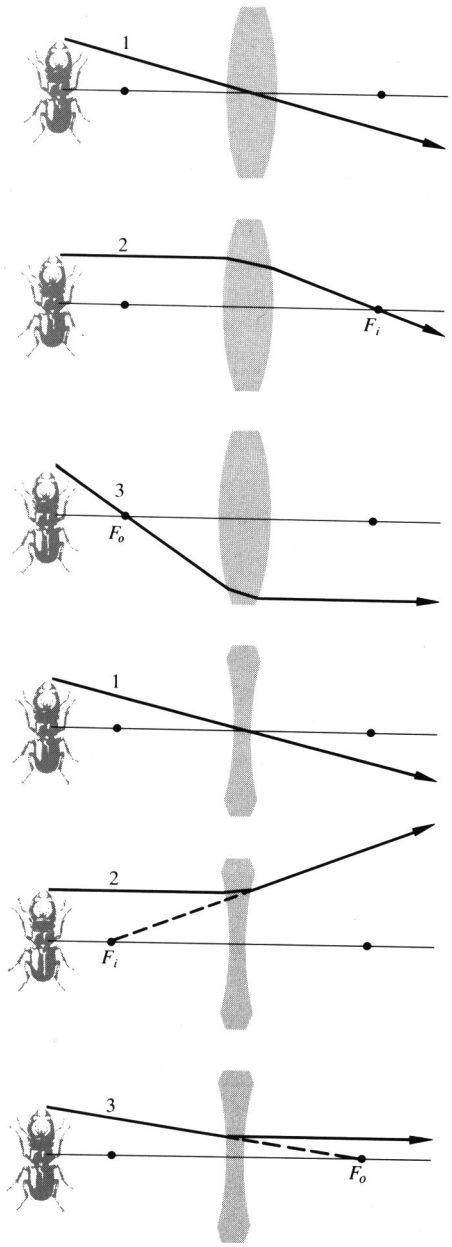


Figure 5.23 Tracing a few key rays through a positive and negative lens.

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. Thus 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 will see later, this is tantamount to saying that the two principal planes of a thin lens coincide.)

In accord with convention, transverse distances above the optical axis are taken as positive quantities, and those below the axis are given negative numerical values. Therefore in Fig. 5.25 $y_o > 0$ and $y_i < 0$. Here the image is said to be *inverted*, whereas if $y_i > 0$ when $y_o > 0$, it is *erect*. Observe that triangles AOF_i and $P_2P_1F_i$ are similar. Ergo

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

Likewise, 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

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

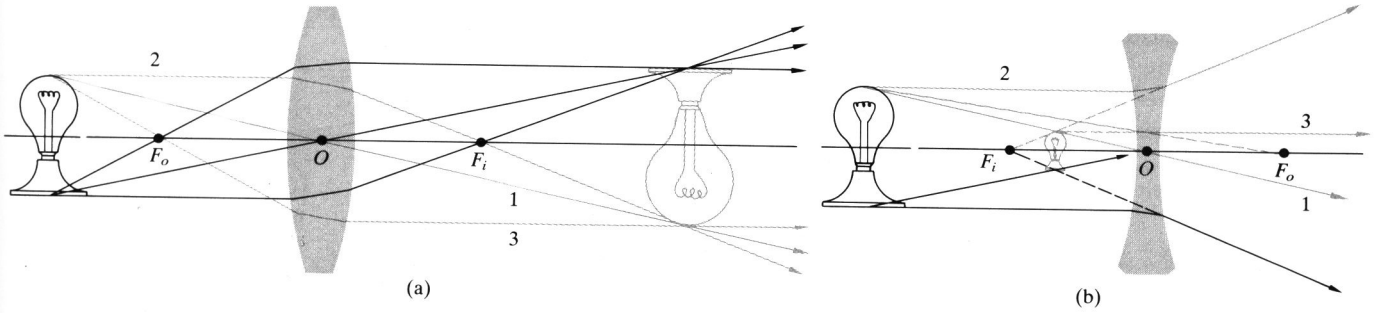
and

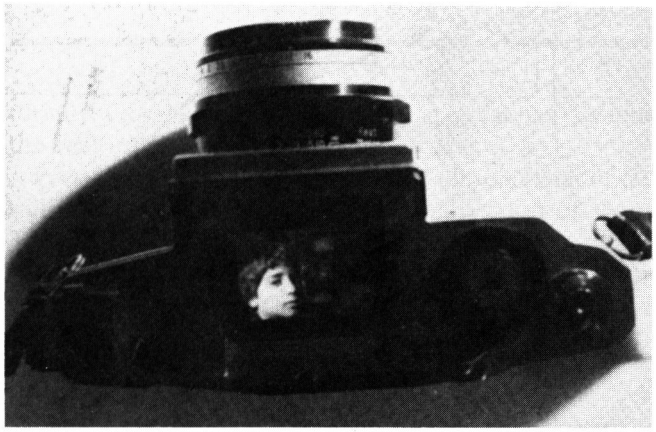
$$\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i},$$

which is, of course, the Gaussian lens equation (5.17). Furthermore, triangles $S_2S_1F_o$ and BOF_o are similar and

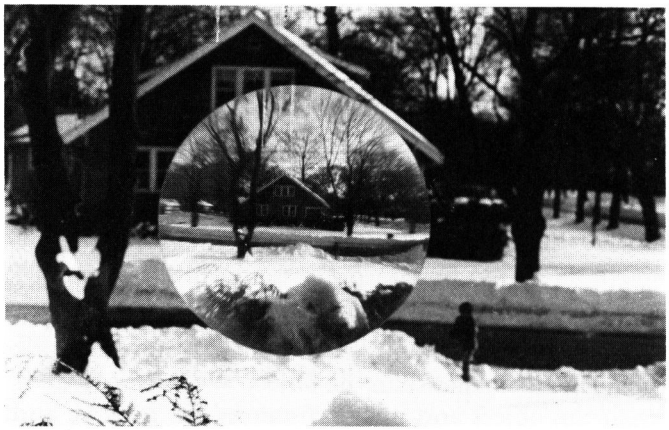
$$\frac{f}{(s_o - f)} = \frac{|y_i|}{y_o}. \tag{5.22}$$

Using the distances measured from the focal points and





(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 much as the eye projects its image on the retina. (d) The minified, rightside-up, virtual image formed by a negative lens.

combining this information with Eq. (5.19), we have

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

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. To be sure, 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

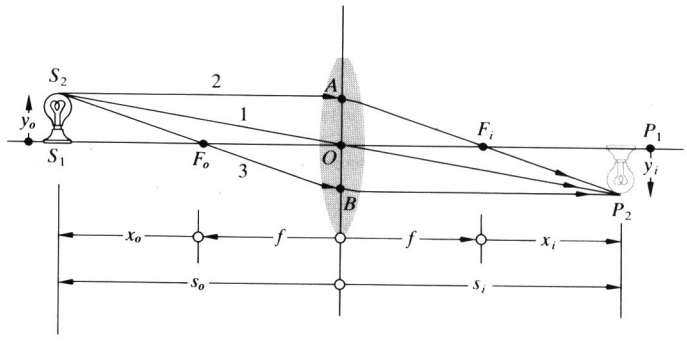


Figure 5.25 Object and image location for a thin lens.

when making those hasty freehand ray diagrams for which he is already infamous.

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

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

Or from Eq. (5.20)

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

Thus *a positive M_T connotes an erect image, while a negative value means the image is inverted* (see Table 5.2). Bear in mind that s_i and s_o are both positive for real objects and images. *Clearly, then, all such 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. 5.24, whence

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

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 (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 5.6). Table 5.3 summarizes a number of image configurations resulting from the juxtaposition of a thin lens and a real object. Figure 5.26 illustrates the behavior pic-

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

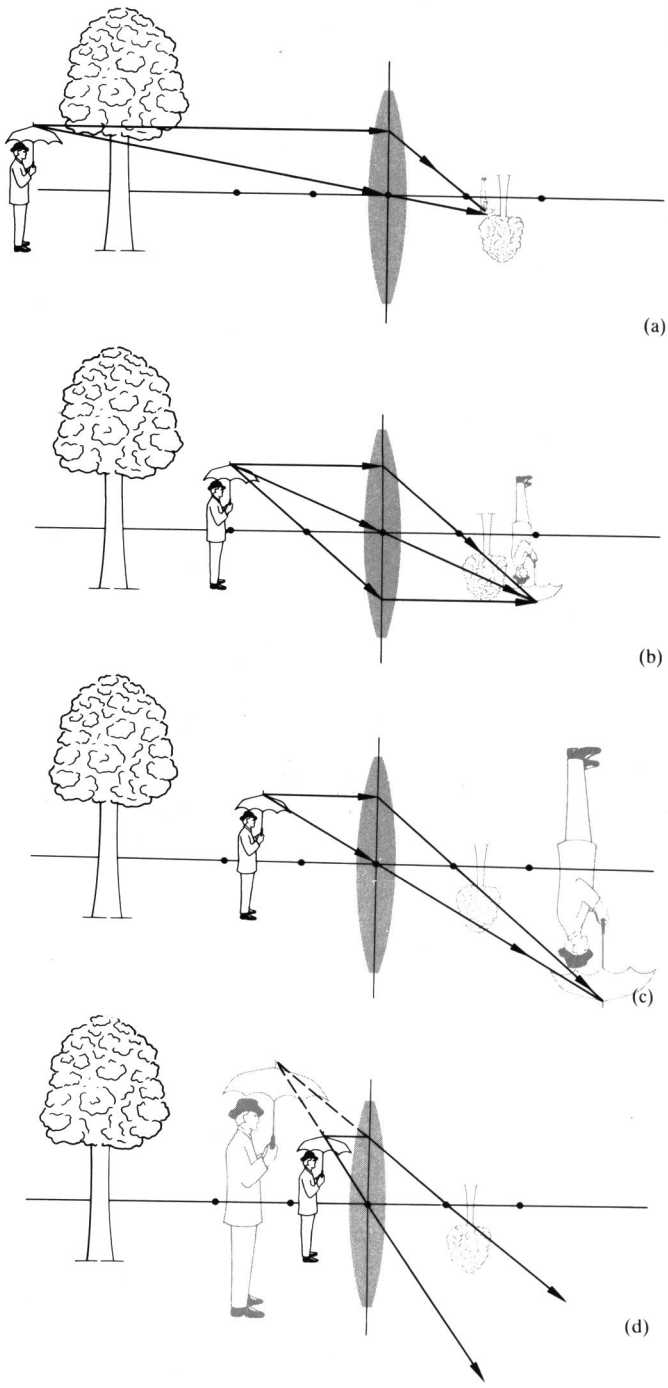


Figure 5.26 The image-forming behavior of a thin positive lens.

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

torially. Observe that as the object approaches the lens, the real image moves away from it.

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

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

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

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

for a thin lens in a single medium (Fig. 5.27). 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. 5.28).

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.

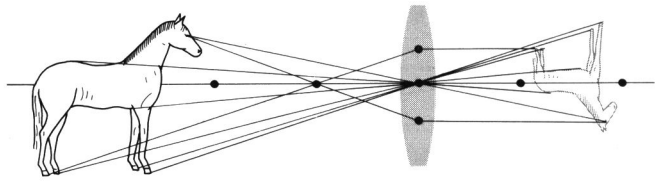


Figure 5.27 The transverse magnification is different from the longitudinal magnification.

iv) Thin-Lens Combinations

Our purpose here is not to become proficient in the subtle intricacies of modern lens design, but rather to begin to appreciate, utilize, and adapt those systems already available.

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 electronic digital 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. Almost all

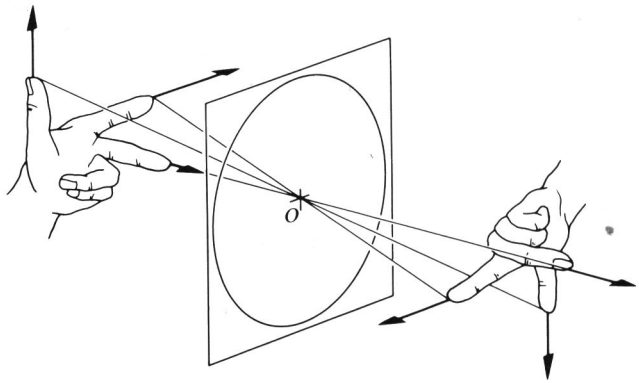


Figure 5.28 Image orientation for a thin lens.

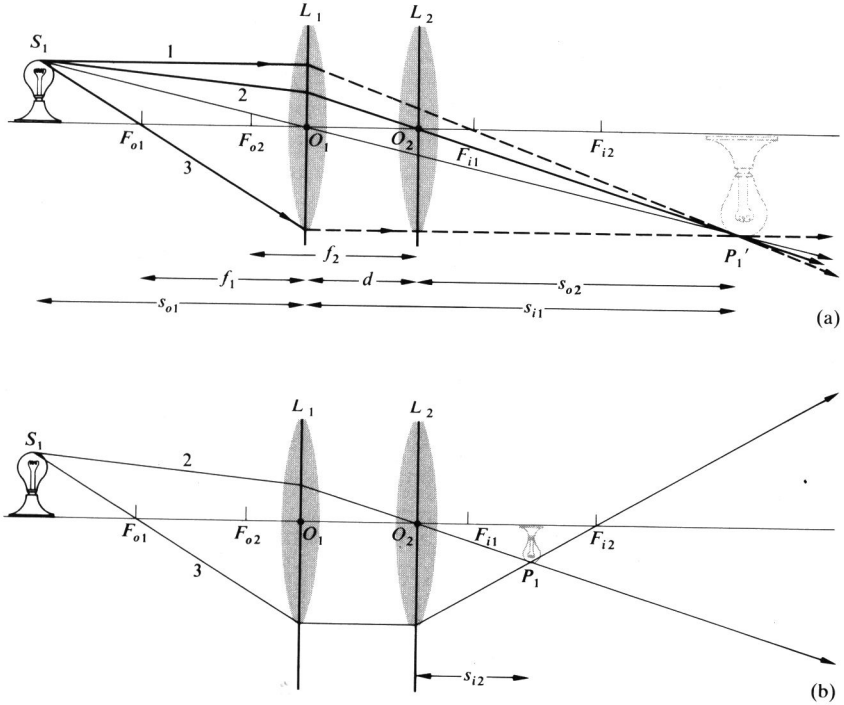


Figure 5.29 Two thin lenses separated by a distance smaller than either focal length.

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

Suppose we have two thin positive lenses L_1 and L_2 separated by a distance d, which is smaller than either focal length, as in Fig. 5.29. The resulting image can be located graphically as follows. If we overlook L_2 for a moment, the image formed exclusively by L_1 is constructed with 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.

A similar pair of lenses is illustrated in Fig. 5.30, 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, we have for L_1

$$\frac{1}{s_{i1}} = \frac{1}{f_1} - \frac{1}{s_{o1}} \tag{5.29}$$

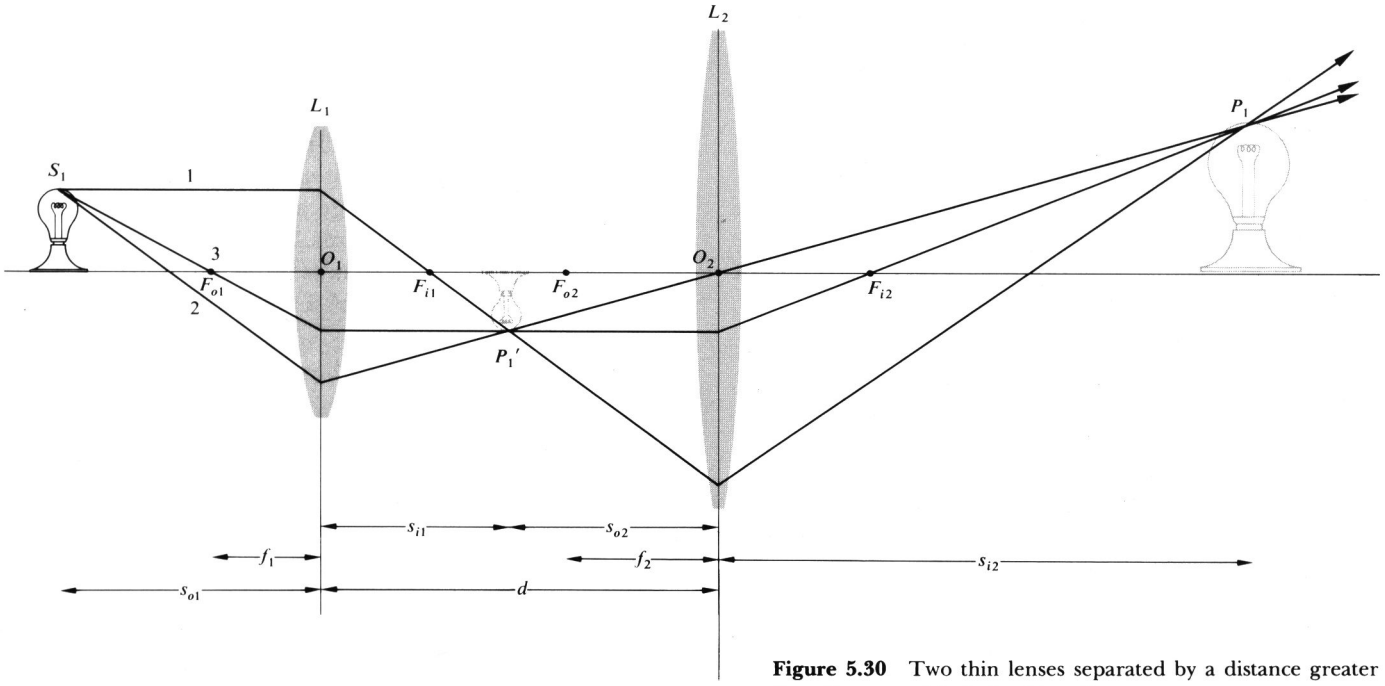


Figure 5.30 Two thin lenses separated by a distance greater than the sum of their focal lengths.

or

$$s_{i1} = \frac{s_{o1} f_1}{s_{o1} - f_1}. \tag{5.30}$$

This is positive, and the intermediate image is to the right of L_1, when $s_{o1} > f_1$ and $f_1 > 0$. For L_2

$$s_{o2} = d - s_{i1}, \tag{5.31}$$

and if $d > s_{i1}$, the object for L_2 is real (as in Fig. 5.30), whereas if $d < s_{i1}$, it is virtual ($s_{o2} < 0$, as in Fig. 5.29). 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), we get

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

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 cm from the first of two positive lenses. These

in turn are separated by 20 cm and have focal lengths of 30 cm and 50 cm, respectively. By direct substitution (5.33)

$$s_{i2} = \frac{50(20) - 50(50)(30)/(50 - 30)}{20 - 50 - 50(30)/(50 - 30)} = 26.2 \text{ cm},$$

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 (5.25) 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. 5.29, 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. Likewise, 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

$$\text{b.f.l.} = \frac{f_2(d - f_1)}{d - (f_1 + f_2)}. \tag{5.36}$$

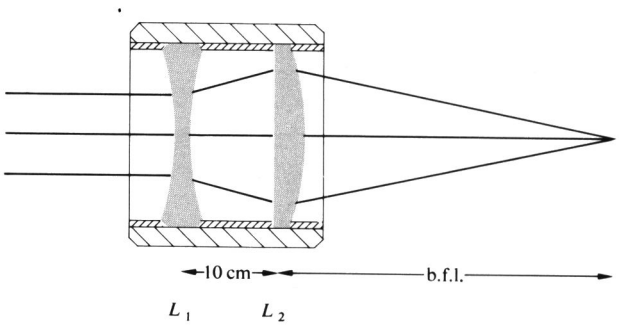
To see how this works numerically, let's find both the b.f.l. and f.f.l. for the thin-lens system in Fig. 5.31(a), where $f_1 = -30$ cm and $f_2 = +20$ cm. Then

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

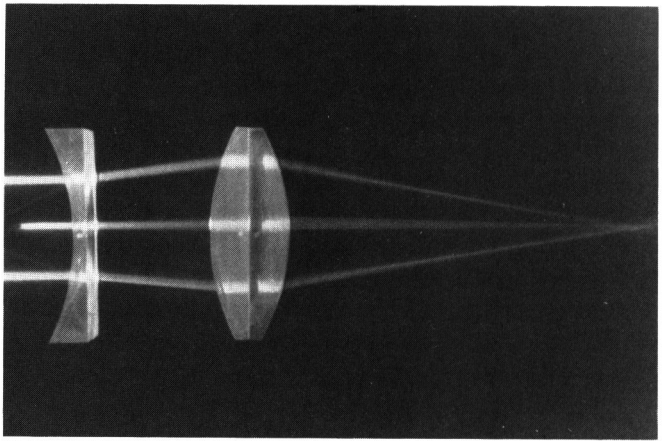
and similarly f.f.l. = 15 cm. Incidentally, notice that if $d = f_1 + f_2$, plane waves entering the compound lens from either side will emerge as plane waves (Problem 5.27), as in telescopic systems.

Observe that if $d \to 0$, that is, if 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 5.31 A positive and negative thin-lens combination. (Photo by 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 5.29 is quite easy to duplicate, and the procedure should be self-evident, whereas Fig. 5.30 requires a bit more care. First determine the focal lengths of the two lenses by imaging a distant source. Then hold one of the lenses (L_2) at a fixed distance *slightly greater than its focal length* from the plane of observation (i.e., a piece of white paper). Now comes the maneuver that requires some effort if you don't have an optical bench. Move the second lens (L_1) toward the source, keeping it reasonably centered. Without any attempts to block out light entering L_2 directly, you will probably see a blurred image of your hand holding L_1. Position the lenses so that the region on the screen corresponding to L_1 is as bright as possible. The scene spread across L_1 (i.e., its image within the image) will become clear and erect, as in Fig. 5.30.

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 as A.S. The adjustable leaf diaphragm that is usually located behind

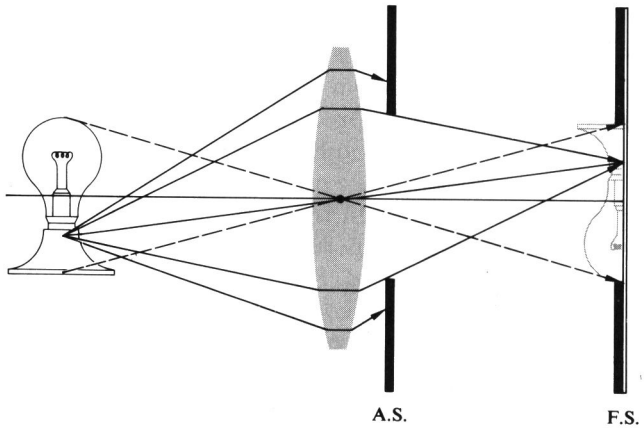


Figure 5.32 Aperture stop and field stop.

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. 5.32, 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. 5.32) 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 very top nor the bottom of the object in Fig. 5.32 passes the field stop. Opening the circular aperture stop would cause the system to accept a larger energy cone and in so doing increase the irradiance at each image point. In contrast, opening the field stop would allow the extremities of the object, which were previously blocked, to be imaged.

5.3.2 Entrance and Exit Pupils

Another concept, quite 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. 5.33, 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. 5.33 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.

If you wanted 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 5.34 represents a similar arrangement in which the entrance and exit pupil locations should be self-evident. The last two diagrams

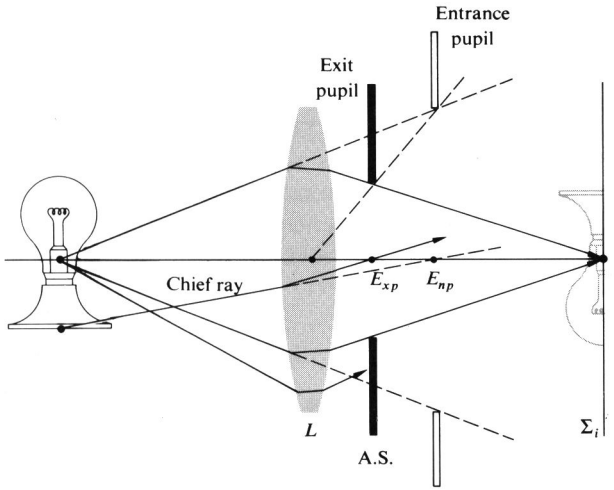


Figure 5.33 Entrance pupil and exit pupil.

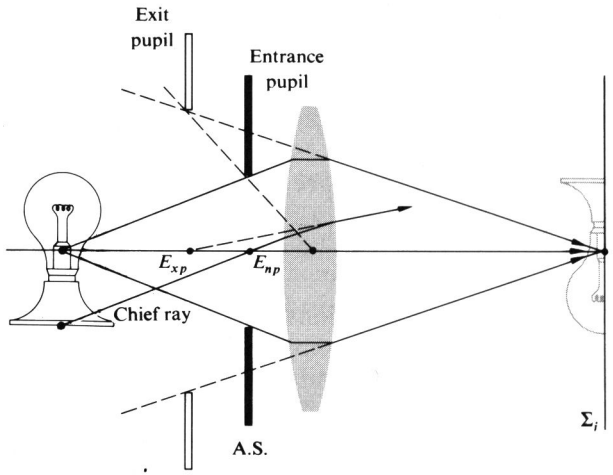


Figure 5.34 A front aperture stop.

included a ray labeled the *chief ray.* It is defined to be *any ray from an off-axis object point that passes through the center of the aperture stop. The chief ray enters the optical system along a line directed toward the midpoint of the entrance pupil, E_{np}, and leaves the system along a line passing through the center of the exit pupil, E_{xp}.* The chief ray, associated with a conical bundle of rays from a point on the object, effectively behaves as the central ray of the bundle and is representative of it. Chief rays are of particular importance when the aberrations of a lens design are being corrected.

Figure 5.35 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 5.30 deals with just this

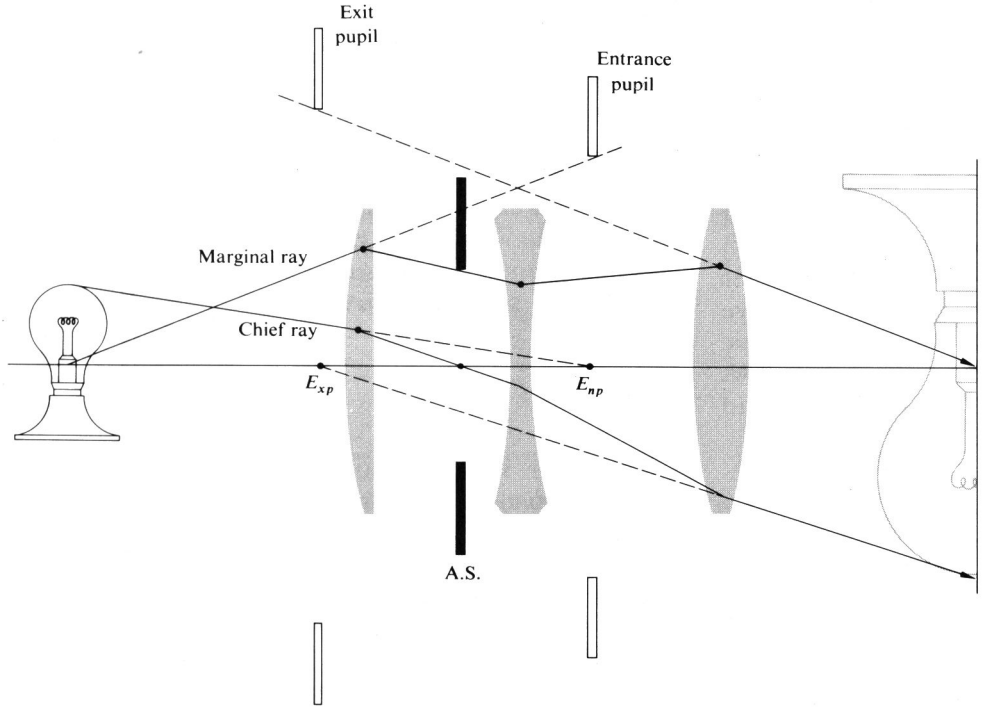


Figure 5.35 Pupils and stops for a three-lens system.

kind of calculation.

Notice how the cone of rays, in Fig. 5.36, 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 b The result is a gradual fading out of the imag near its periphery, a process known as *vigneti*

The locations and sizes of the pupils of an optical system are of considerable practical importance. In

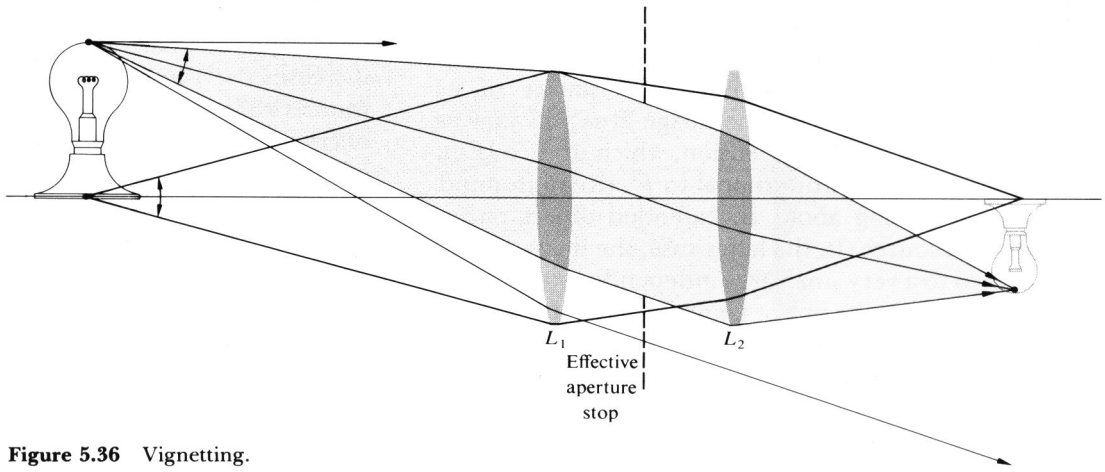


Figure 5.36 Vignetting.

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 will be 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.

5.3.3 Relative Aperture and *f*-Number

Suppose we wish to 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 reflections, absorption, and so forth, the incoming energy will be spread across a corresponding region of the image. Thus 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 said to be the **f-number,** or $f/\#$, that is,

$$f/\# \equiv \frac{f}{D}, \tag{5.40}$$

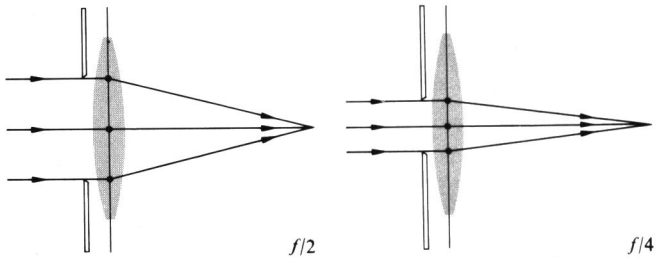


Figure 5.37 Stopping down a lens to change the *f-number.*

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 5.37 illustrates the point by showing a thin lens behind a variable iris diaphragm operating at either $f/2$ or $f/4$. A smaller *f-number* clearly permits more light to reach the image plane.

Camera lenses are usually specified by their focal lengths and largest possible apertures; for example, you might see "50 mm, $f/1.4$" on the barrel of a lens. Since the photographic exposure time is proportional to the square of the *f-number* the latter is sometimes spoken of as the *speed* of the lens. An $f/1.4$ lens is said to be twice as fast as an $f/2$ lens. Usually lens diaphragms have *f-number* markings of 1, 1.4, 2, 2.8, 4, 5.6, 8, 11, 16, 22, and so on. The largest relative aperture in this case corresponds to $f/1$, and that's a fast lens—$f/2$ is more typical. Each consecutive diaphragm setting increases the *f-number* by a multiplicative factor of $\sqrt{2}$ (numerically 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.

For precise work, in which reflection and absorption losses in the lens itself must be taken into consideration, the *T-number* is highly useful. In effect, it is a modified (increased) *f-number* that a given real lens would actually have to have were it to transmit an amount of light corresponding to a particular value of f/D.

5.4 MIRRORS

Mirror systems are being used in increasingly extensive applications, 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 bandwidth, 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. 3.29). As we will see later, when we consider their aberrations, mirrors have other attributes that 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, the latter being chosen because of its high efficiency in the UV and IR (see Fig. 4.42), and the former because of its rigidity. In recent times, 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 whole new generation of lightweight precision mirrors is being 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 is the kind 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. 5.38).

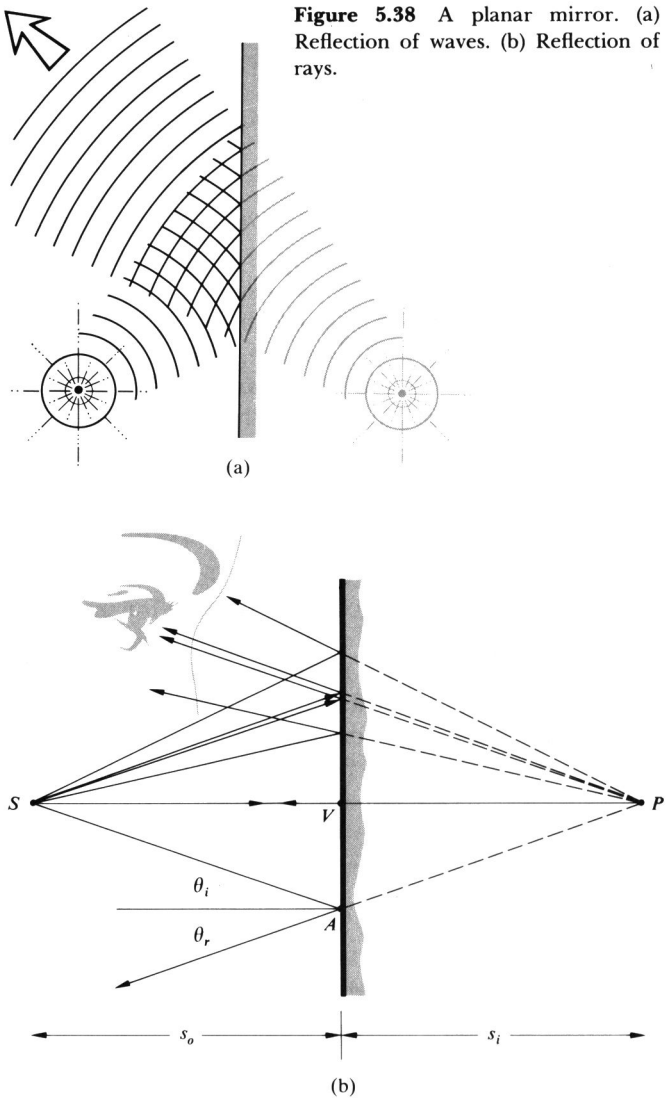


Figure 5.38 A planar mirror. (a) Reflection of waves. (b) Reflection of rays.