This is OPTIONAL material, for your education only. It will not be on any exam or homework.

Where the factor of 1.22 comes from, for diffraction through a circular aperture (in $\theta_{min} \simeq 1.22\lambda/A$).

θ_{min} small and in radians

cf. Optics, 2nd ed., by Eugene Hecht, the text for Phys 110 (Physical Optics), offered Fall semester.

10.2.5 The Circular Aperture

Fraunhofer diffraction at a circular aperture is an effect of great practical significance in the study of optical instrumentation. Envision a typical arrangement: plane waves impinging on a screen Σ containing a circular aperture and the consequent far-field diffraction pattern spread across a distant observing screen σ. By using a focusing lens L_2, we can bring σ in close to the aperture without changing the pattern. Now, if L_2 is positioned within and exactly fills the diffracting opening in Σ, the form of the pattern is essentially unaltered. The lightwave reaching Σ is cropped, so that only a circular segment propagates through L_2 to form an image in the focal plane. This is obviously the same process that takes place in an eye, telescope, microscope, or camera lens. The image of a distant point source, as formed by a perfectly aberration-free converging lens, is never a point but rather some sort of diffraction pattern. We are essentially collecting only a fraction of the incident wavefront and therefore cannot hope to form a perfect image. As shown in the last section, the expression for the optical disturbance at P, arising from an arbitrary aperture in the far-field case, is

$$E = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \iint_{\text{Aperture}} e^{ik(Yy+Zz)/R}\, dS. \qquad [10.41]$$

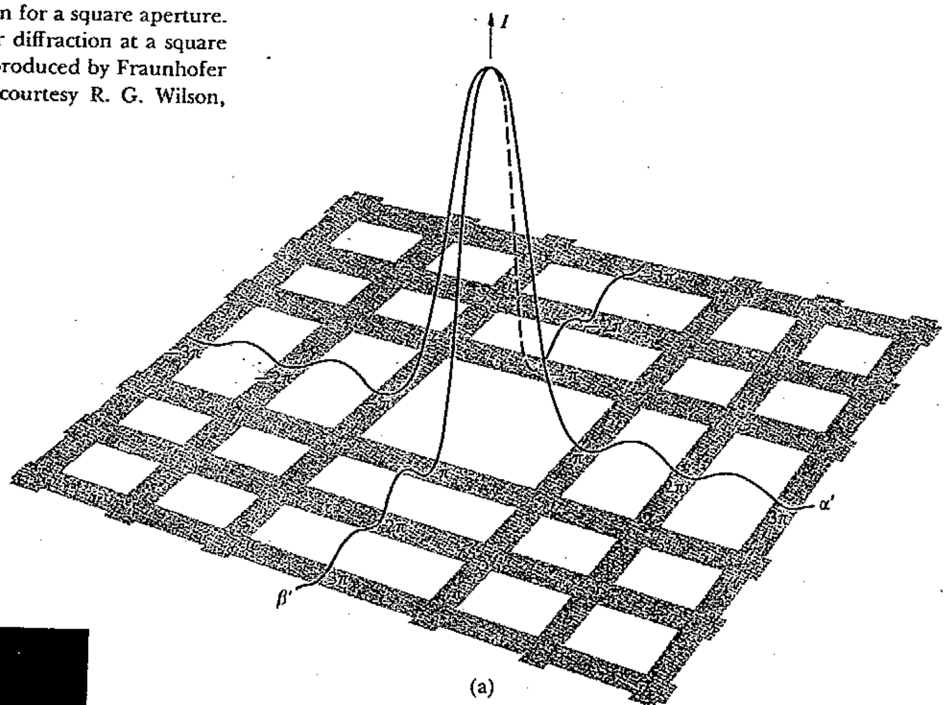
For a circular opening, symmetry would suggest introducing spherical polar coordinates in both the plane of the aperture and the plane of observation, as shown in Fig. 10.26. Therefore, let

$$z = \rho \cos\phi \qquad y = \rho \sin\phi$$
$$Z = q \cos\Phi \qquad Y = q \sin\Phi.$$

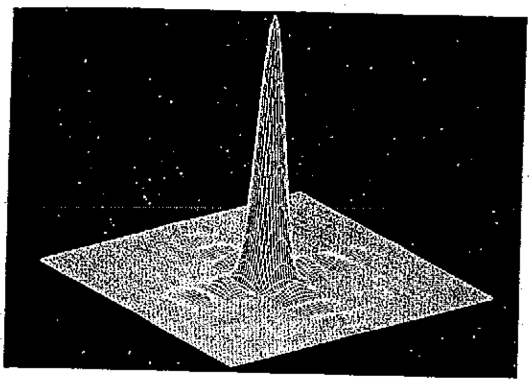
The differential element of area is now

$$dS = \rho\, d\rho\, d\phi.$$

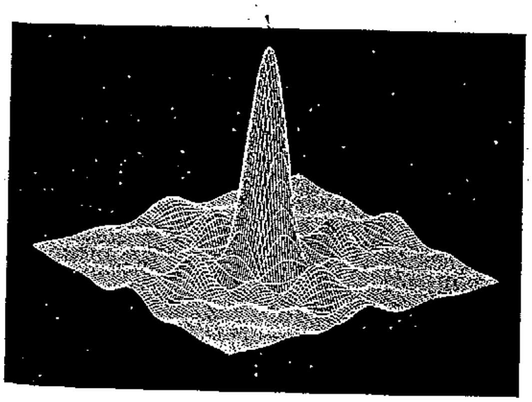
Figure 10.25 (a) The irradiance distribution for a square aperture. (b) The irradiance produced by Fraunhofer diffraction at a square aperture. (c) The electric field distribution produced by Fraunhofer diffraction via a square aperture. (Photos courtesy R. G. Wilson, Illinois Wesleyan University.)



(a)



(b)



(c)

Substituting these expressions into Eq. (10.41), it becomes

$$E = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \int_{\rho=0}^{a} \int_{\phi=0}^{2\pi} e^{i(k\rho q/R)\cos(\phi-\Phi)} \rho \, d\rho \, d\phi. \tag{10.46}$$

Because of the complete axial symmetry, the solution must be independent of Φ. We might just as well solve Eq. (10.46) with $\Phi = 0$ as with any other value, thereby simplifying things slightly.

The portion of the double integral associated with the variable ϕ,

$$\int_0^{2\pi} e^{i(k\rho q/R)\cos\phi} \, d\phi,$$

is one that arises quite frequently in the mathematics of physics. It is a unique function in that it cannot be reduced to any of the more common forms, such as the various hyperbolic, exponential, or trigonometric functions, and indeed with the exception of these, it is

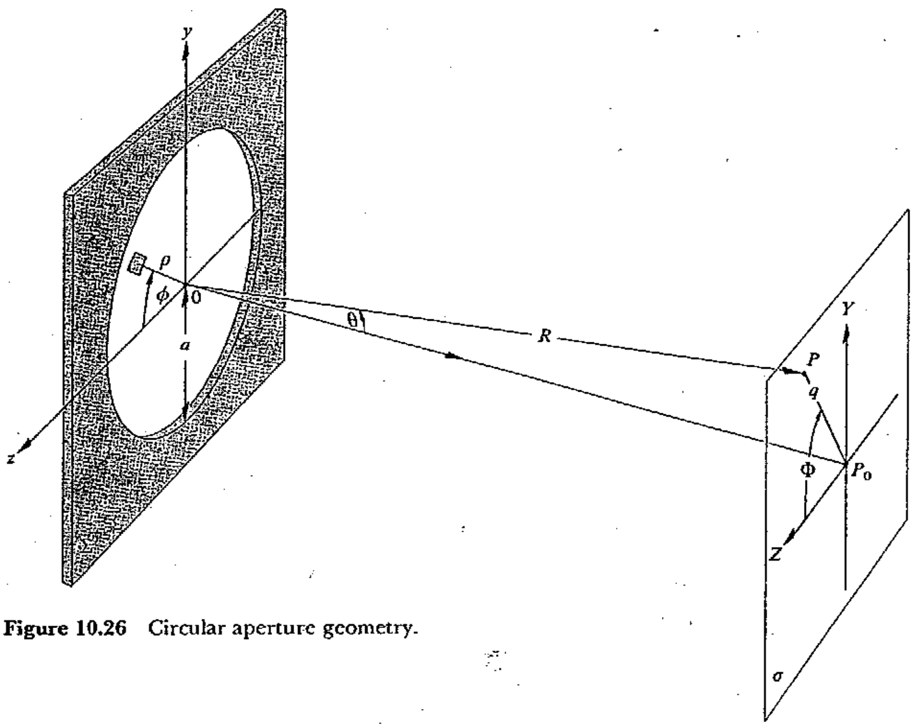


Figure 10.26 Circular aperture geometry.

perhaps the most often encountered. The quantity

$$J_0(u) = \frac{1}{2\pi}\int_0^{2\pi} e^{iu\cos v}\, dv \qquad (10.47)$$

is known as the *Bessel function* (of the first kind) of order zero. More generally,

$$J_m(u) = \frac{i^{-m}}{2\pi}\int_0^{2\pi} e^{i(mv+u\cos v)}\, dv \qquad (10.48)$$

represents the Bessel function of order m. Numerical values of $J_0(u)$ and $J_1(u)$ are tabulated for a large range of u in most mathematical handbooks. Just like sine and cosine, the Bessel functions have series expansions and are certainly no more esoteric than these familiar childhood acquaintances. As seen in Fig. 10.27, $J_0(u)$ and $J_1(u)$ are slowly decreasing oscillatory functions that do nothing particularly dramatic.

Equation (10.46) can be rewritten as

$$E = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R}\, 2\pi \int_0^a J_0(k\rho q/R)\rho\, d\rho. \qquad (10.49)$$

Another general property of Bessel functions, referred to as a recurrence relation, is

$$\frac{d}{du}[u^m J_m(u)] = u^m J_{m-1}(u).$$

When $m = 1$, this clearly leads to

$$\int_0^u u' J_0(u')\, du' = uJ_1(u), \qquad (10.50)$$

with u' just serving as a dummy variable. If we now return to the integral in Eq. (10.49) and change the variable such that $w = k\rho q/R$, then $d\rho = (R/kq)\, dw$ and

$$\int_{\rho=0}^{\rho=a} J_0(k\rho q/R)\rho\, d\rho = (R/kq)^2 \int_{w=0}^{w=kaq/R} J_0(w)w\, dw.$$

Making use of Eq. (10.50), we get

$$E(t) = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R}\, 2\pi a^2 (R/kaq) J_1(kaq/R). \qquad (10.51)$$

The irradiance at point P is $\langle(\mathrm{Re}\, E)^2\rangle$ or $\frac{1}{2}EE^*$, that is,

$$I = \frac{2\mathcal{E}_A^2 A^2}{R^2}\left[\frac{J_1(kaq/R)}{kaq/R}\right]^2, \qquad (10.52)$$

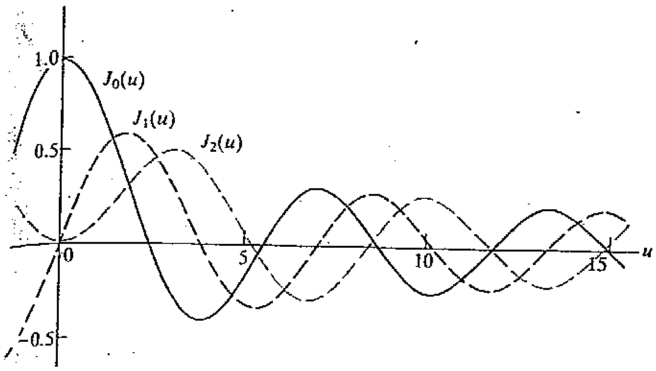


Figure 10.27 Bessel functions.

where A is the area of the circular opening. To find the irradiance at the center of the pattern (i.e., at P_0), set $q = 0$. It follows from the above recurrence relation ($m = 1$) that

$$J_0(u) = \frac{d}{du} J_1(u) + \frac{J_1(u)}{u}. \tag{10.53}$$

From Eq. (10.47) we see that $J_0(0) = 1$, and from Eq. (10.48), $J_1(0) = 0$. The ratio of $J_1(u)/u$ as u approaches zero has the same limit (L'Hospital's rule) as the ratio of the separate derivatives of its numerator and denominator, namely, $dJ_1(u)/du$ over 1. But this means that the right-hand side of Eq. (10.53) is twice that limiting value, so that $J_1(u)/u = \frac{1}{2}$ at $u = 0$. The irradiance at P_0 is therefore

$$I(0) = \frac{\mathcal{E}_A^2 A^2}{2R^2}, \tag{10.54}$$

which is the same result obtained for the rectangular opening (10.43). If R is assumed to be essentially constant over the pattern, we can write

$$I = I(0)\left[\frac{2J_1(kaq/R)}{kaq/R}\right]^2. \tag{10.55}$$

Since $\sin\theta = q/R$, the irradiance can be written as a function of θ,

$$I(\theta) = I(0)\left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta}\right]^2, \tag{10.56}$$

and as such is plotted in Fig. 10.28. Because of the axial symmetry, the towering central maximum corresponds to a high-irradiance circular spot known as the **Airy disk.** It was Sir George Biddell Airy (1801–1892), Astronomer Royal of England, who first derived Eq. (10.56). The central disk is surrounded by a dark ring that corresponds to the first zero of the function $J_1(u)$. From Table 10.1 $J_1(u) = 0$ when $u = 3.83$, that is, $kaq/R = 3.83$. The radius q_1 drawn to the center of this first dark ring can be thought of as the extent of the Airy disk. It is given by

$$q_1 = 1.22\frac{R\lambda}{2a}. \tag{10.57}$$

Table 10.1 Bessel functions.*

x	$J_1(x)$*	x	$J_1(x)$	x	$J_1(x)$
0.0	0.0000	3.0	0.3391	6.0	−0.2767
0.1	0.0499	3.1	0.3009	6.1	−0.2559
0.2	0.0995	3.2	0.2613	6.2	−0.2329
0.3	0.1483	3.3	0.2207	6.3	−0.2081
0.4	0.1960	3.4	0.1792	6.4	−0.1816
0.5	0.2423	3.5	0.1374	6.5	−0.1538
0.6	0.2867	3.6	0.0955	6.6	−0.1250
0.7	0.3290	3.7	0.0538	6.7	−0.0953
0.8	0.3688	3.8	0.0128	6.8	−0.0652
0.9	0.4059	3.9	−0.0272	6.9	−0.0349
1.0	0.4401	4.0	−0.0660	7.0	−0.0047
1.1	0.4709	4.1	−0.1033	7.1	0.0252
1.2	0.4983	4.2	−0.1386	7.2	0.0543
1.3	0.5220	4.3	−0.1719	7.3	0.0826
1.4	0.5419	4.4	−0.2028	7.4	0.1096
1.5	0.5579	4.5	−0.2311	7.5	0.1352
1.6	0.5699	4.6	−0.2566	7.6	0.1592
1.7	0.5778	4.7	−0.2791	7.7	0.1813
1.8	0.5815	4.8	−0.2985	7.8	0.2014
1.9	0.5812	4.9	−0.3147	7.9	0.2192
2.0	0.5767	5.0	−0.3276	8.0	0.2346
2.1	0.5683	5.1	−0.3371	8.1	0.2476
2.2	0.5560	5.2	−0.3432	8.2	0.2580
2.3	0.5399	5.3	−0.3460	8.3	0.2657
2.4	0.5202	5.4	−0.3453	8.4	0.2708
2.5	0.4971	5.5	−0.3414	8.5	0.2731
2.6	0.4708	5.6	−0.3343	8.6	0.2728
2.7	0.4416	5.7	−0.3241	8.7	0.2697
2.8	0.4097	5.8	−0.3110	8.8	0.2641
2.9	0.3754	5.9	−0.2951	8.9	0.2559

*$J_1(x) = 0$ for $x = 0, 3.832, 7.016, 10.173, 13.324, \ldots$

Adapted from E. Kreyszig, *Advanced Engineering Mathematics*, Wiley.

For a lens focused on the screen σ, the focal length $f \approx R$, so

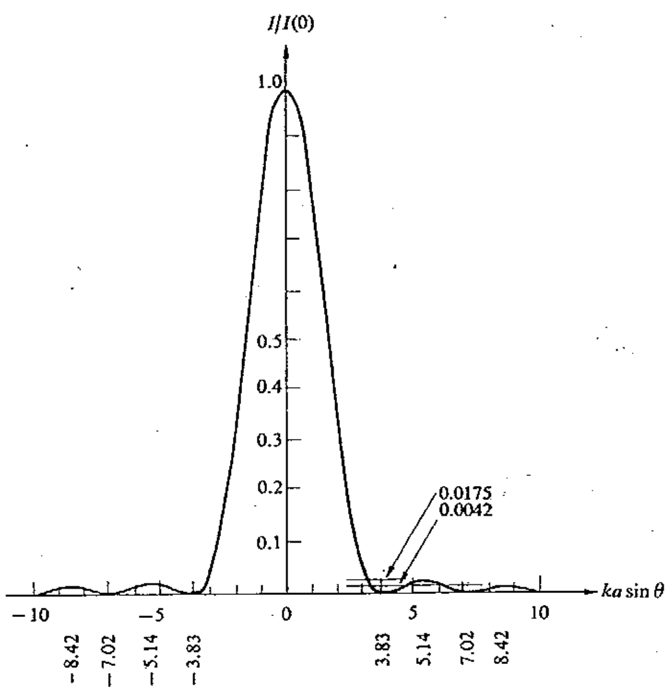
$$q_1 \approx 1.22 \frac{f\lambda}{D}, \tag{10.58}$$

where D is the aperture diameter, in other words, $D = 2a$. (The *diameter* of the Airy disk in the visible spectrum is *very roughly* equal to the $f/\#$ of the lens in millionths of a meter.) As shown in Figs. 10.29 to 10.31, q_1 varies inversely with the hole's diameter. As D approaches λ, the Airy disk can be very large indeed, and the circular aperture begins to resemble a point source of spherical waves.

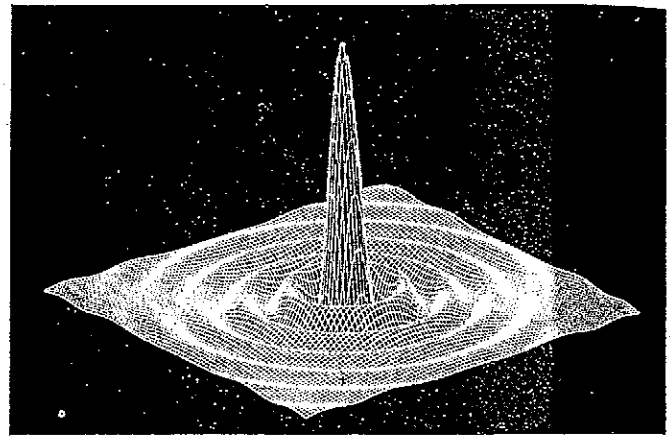
The higher-order zeros occur at values of kaq/R equal to 7.02, 10.17, and so forth. The secondary maxima are located where u satisfies the condition

$$\frac{d}{du}\left[\frac{J_1(u)}{u}\right] = 0,$$

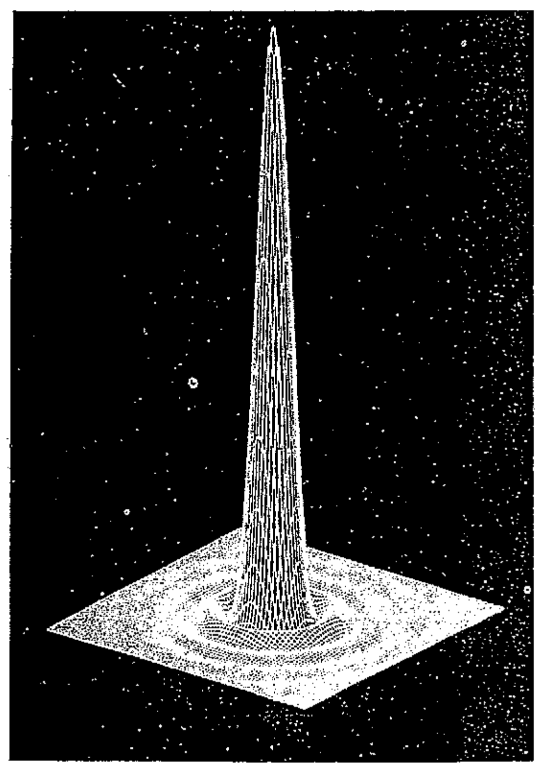
which is equivalent to $J_2(u) = 0$. From the tables then,



(a)



(b)



(c)

Figure 10.28 (a) The Airy pattern. (b) Electric field created by Fraunhofer diffraction at a circular aperture. (c) Irradiance resulting from Fraunhofer diffraction at a circular aperture. (Photos courtesy R. G. Wilson, Illinois Wesleyan University.)

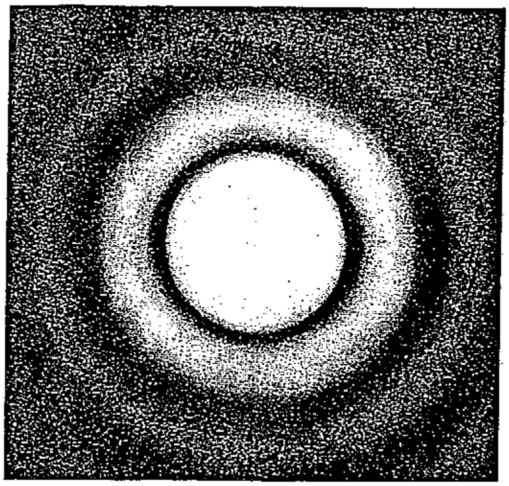


Figure 10.29 Airy rings (0.5-mm hole diameter). (Photo by E. H.)

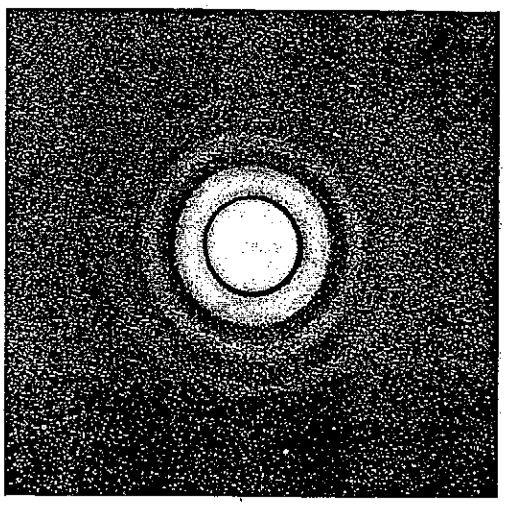


Figure 10.30 Airy rings (1.0-mm hole diameter). (Photo by E. H.)

these secondary peaks occur when kaq/R equals 5.14, 8.42, 11.6, and so on, whereupon $I/I(0)$ drops from 1 to 0.0175, 0.0042, and 0.0016, respectively (Problem 10.22).

Circular apertures are preferable to rectangular ones, as far as lens shapes go, since the circle's irradiance curve is broader around the central peak and drops off more rapidly thereafter. Exactly what fraction of the total light energy incident on σ is confined to within

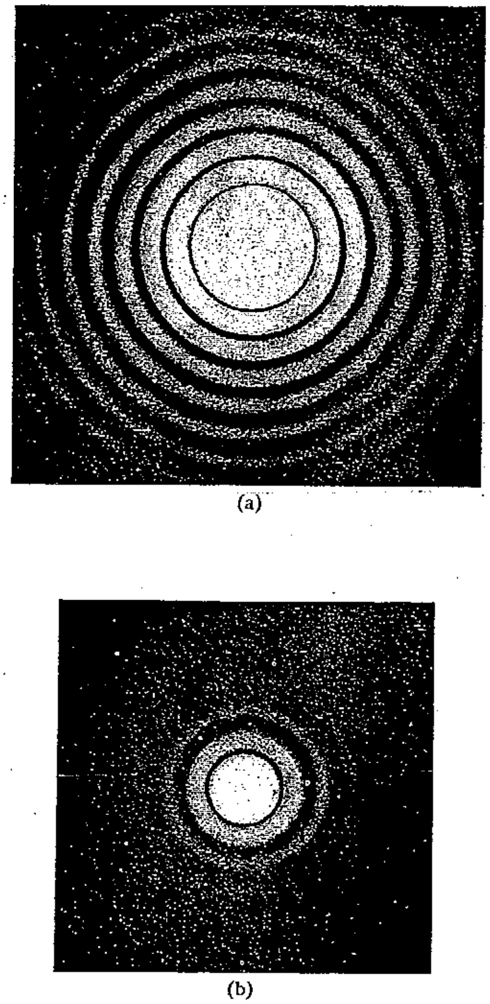


Figure 10.31 (a) Airy rings—long exposure (1.5-mm hole diameter). (b) Central Airy disc—short exposure with the same aperture. (Photos by E. H.)

the various maxima is a question of interest, but one somewhat too involved to solve here.* On integrating the irradiance over a particular region of the pattern, one finds that 84% of the light arrives within the Airy disk, and 91% within the bounds of the second dark ring.

*See Born and Wolf, *Principles of Optics*, p. 398, or the very fine elementary text by Towne, *Wave Phenomena*, p. 464.

10.2.6 Resolution of Imaging Systems

Imagine that we have some sort of lens system that forms an image of an extended object. If the object is self-luminous, it is likely that we can regard it as made up of an array of incoherent sources. On the other hand, an object seen in reflected light will surely display some phase correlation between its various scattering points. When the point sources are in fact incoherent, the lens system will form an image of the object, which consists of a distribution of partially overlapping, yet independent, Airy patterns. In the finest lenses, which have negligible aberrations, the spreading out of each image point due to diffraction represents the ultimate limit on image quality.

Suppose that we simplify matters somewhat and examine only two equal-irradiance, incoherent, distant point sources. For example, consider two stars seen through the objective lens of a telescope, where the entrance pupil corresponds to the diffracting aperture. In the previous section we saw that the radius of the Airy disk was given by $q_1 = 1.22f\lambda/D$. If $\Delta\theta$ is the corresponding angular measure, then $\Delta\theta = 1.22\lambda/D$, inasmuch as $q_1/f = \sin\Delta\theta \approx \Delta\theta$. The Airy disk for each star will be spread out over an angular half-width $\Delta\theta$ about its geometric image point, as shown in Fig. 10.32. If the angular separation of the stars is $\Delta\varphi$ and if $\Delta\varphi \gg \Delta\theta$, the images will be distinct and easily resolved. As the stars approach each other, their respective images come together, overlap, and commingle into a single blend of fringes. If Lord Rayleigh's criterion is applied, the stars are said to be *just resolved* when the center of one Airy disk falls on the first minimum of the Airy pattern of the other star. (We can certainly do a bit better than this, but Rayleigh's criterion, however arbitrary, has the virtue of being particularly uncomplicated.*) The *minimum resolvable angular separation* or *angular limit of resolution* is

$$(\Delta\varphi)_{\min} = \Delta\theta = 1.22\lambda/D, \qquad (10.59)$$

as depicted in Fig. 10.33. If $\Delta\ell$ is the center-to-center separation of the images, the **limit of resolution** is

$$(\Delta\ell)_{\min} = 1.22f\lambda/D. \qquad (10.60)$$

The **resolving power** for an image-forming system is generally defined as either $1/(\Delta\varphi)_{\min}$ or $1/(\Delta\ell)_{\min}$.

If the smallest resolvable separation between images is to be reduced (i.e., if the resolving power is to be increased), the wavelength, for instance, might be made smaller. Using ultraviolet rather than visible light in microscopy allows for the perception of finer detail. The electron microscope utilizes equivalent wavelengths of about 10^{-4} to 10^{-5} that of light. This makes it possible to examine objects that would otherwise be completely obscured by diffraction effects in the visible spectrum. On the other hand, the resolving power of a telescope can be increased by increasing the diameter of the objective lens or mirror. Besides collecting more of the incident radiation, this will also result in a smaller Airy disk and therefore a sharper, brighter image. The Mount Palomar 200-in telescope has a mirror 5 m in diameter (neglecting the obstruction of a small region at its center). At 550 nm it has an angular limit of resolution of 2.7×10^{-2} s of arc. In contrast, the Jodrell Bank radio telescope, with a 250-ft diameter, operates at a rather long, 21-cm wavelength. It therefore has a limit of resolution of only about 700 s of arc. The human eye has a pupil diameter that of course varies. Taking it, under bright conditions, to be about 2 mm, with $\lambda = 550$ nm, $(\Delta\varphi)_{\min}$ turns out to be roughly 1 min of arc. With a focal length of about 20 mm, $(\Delta\ell)_{\min}$ on the retina is 6700 nm. This is roughly twice the mean spacing between receptors. The human eye should therefore be able to resolve two points, an inch apart, at a distance of some 100 yards. You will probably not be able to do quite that well; one part in one thousand is more likely.

A more appropriate criterion for resolving power has been proposed by C. Sparrow. Recall that at the Rayleigh limit there is a central minimum or saddle point between adjacent peaks. A further decrease in the distance between the two point sources will cause the central dip to grow shallower and ultimately disappear. The angular separation corresponding to that configuration is Sparrow's limit. The resultant

* In Rayleigh's own words: "This rule is convenient on account of its simplicity and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution." See Section 9.6.1. for further discussion.

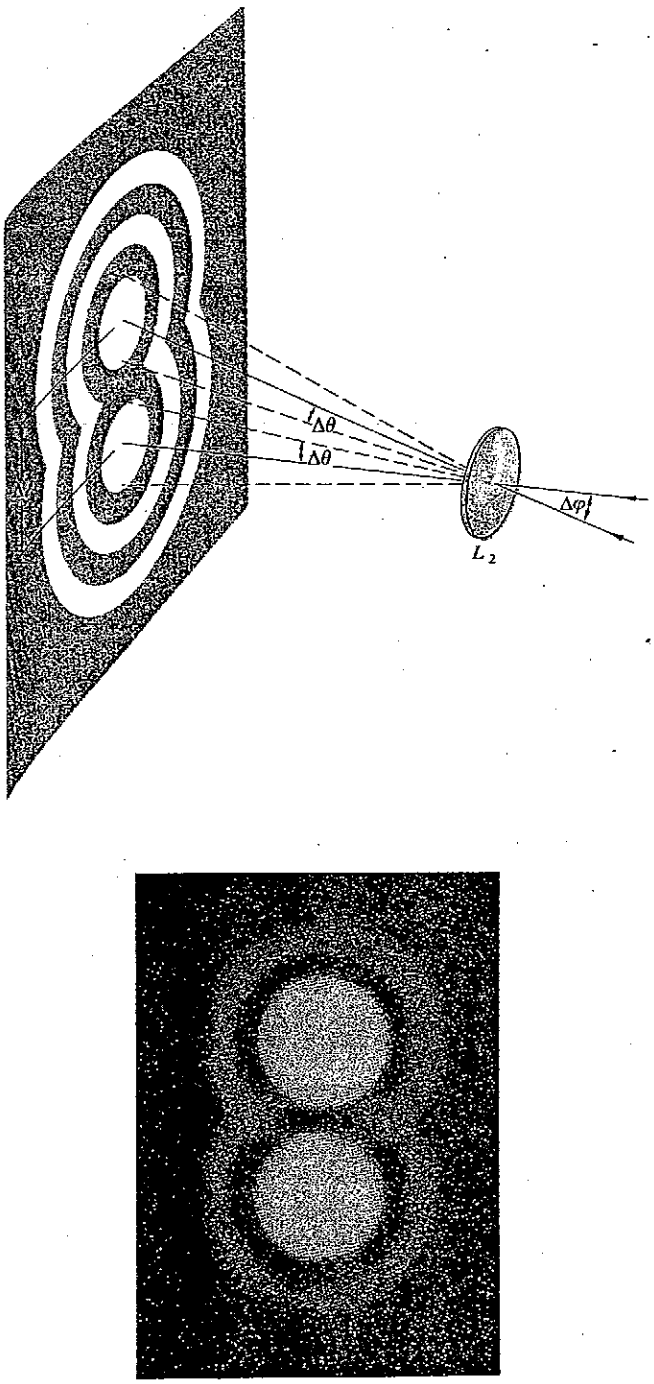


Figure 10.32 Overlapping images.

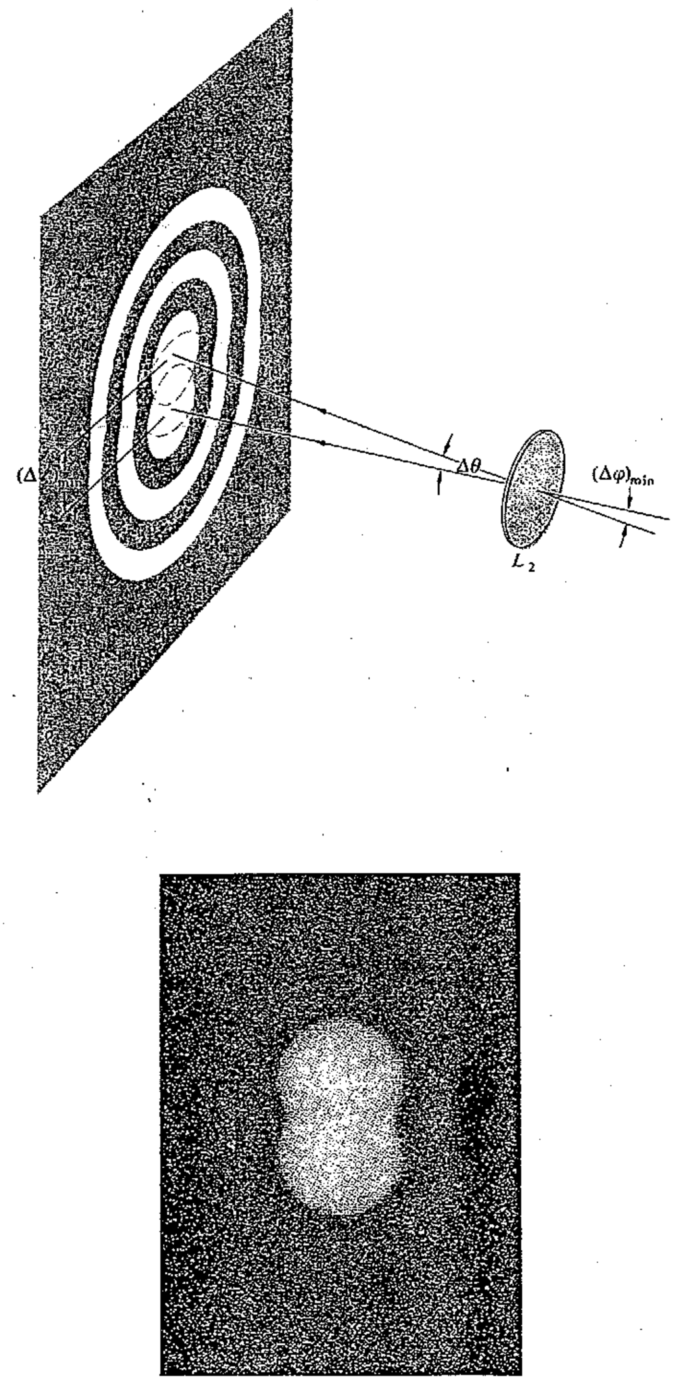


Figure 10.33 Overlapping images.

maximum has a broad flat top; in other words, at the origin, which is the center of the peak, the second derivative of the irradiance function is zero; there is no change in slope (Fig. 10.40).

Unlike the Rayleigh rule, which rather tacitly assumes incoherence, the Sparrow condition can readily be generalized to coherent sources. In addition, astronomical studies of equal-brightness stars have shown that Sparrow's criterion is by far the more realistic.

10.2.7 The Diffraction Grating

A repetitive array of diffracting elements, either apertures or obstacles, that has the effect of producing periodic alterations in the phase, amplitude, or both of an emergent wave is said to be a **diffraction grating.** One of the simplest such arrangements is the multiple-slit configuration of Section 10.2.3. It seems to have been invented by the American astronomer David Rittenhouse in about 1785. Some years later Joseph von Fraunhofer independently rediscovered the principle and went on to make a number of important contributions to both the theory and technology of gratings. The earliest devices were indeed multiple-slit assemblies, usually consisting of a grid of fine wire or thread wound about and extending between two parallel screws, which served as spacers. A wavefront, in passing through such a system, is confronted by alternate opaque and transparent regions, so that it undergoes a modulation in *amplitude.* Accordingly, a multiple-slit configuration is said to be a *transmission amplitude grating.* Another, more common form of transmission grating is made by ruling or scratching parallel notches into the surface of a flat, clear glass plate [Fig. 10.34(a)]. Each of the scratches serves as a source of scattered light, and together they form a regular array of parallel line sources. When the grating is totally transparent, so that there is negligible amplitude modulation, the regular variations in the optical thickness across the grating yield a modulation in *phase,* and we have what is known as a *transmission phase grating* (Fig. 10.35). In the Huygens–Fresnel representation you can envision the wavelets as radiated with different phases over the grating surface. An emerging wavefront therefore contains

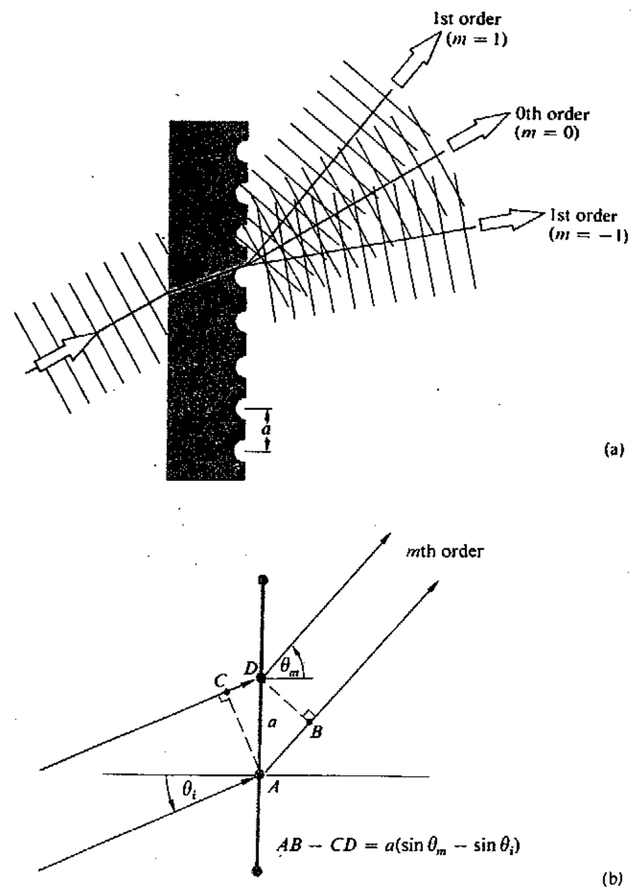


Figure 10.34 A transmission grating.

periodic variations in its shape rather than its amplitude. This in turn is equivalent to an angular distribution of constituent plane waves.

On reflection from this kind of grating, light scattered by the various periodic surface features will arrive at some point P with a definite phase relationship. The consequent interference pattern generated after reflection is quite similar to that arising from transmission. Gratings designed specifically to function in this fashion are known as *reflection phase gratings* (Fig. 10.36). Contemporary gratings of this sort are generally ruled in thin films of aluminum that have been evaporated onto optically flat glass blanks. The aluminum, being fairly