Fourier Optics

Appendix A

The Fourier Analysis Approach to Physical Optics

Source: “A. Nussbaum and R. A. Phillips, “Contemporary Optics for Scientists and Engineers”

10

THE FOURIER ANALYSIS APPROACH TO PHYSICAL OPTICS

10-1 Introduction

We pointed out in an earlier chapter that light is one type of electromagnetic wave. If a wave is radiated in free space, it travels at constant velocity and without alteration until it strikes some sort of detector or absorbing surface. On the other hand, if the wave is partially blocked by an opaque barrier or an aperture in a metal plate, the apparent direction of travel will be altered. As we shall see in the development which follows, this bending of the path is called *diffraction.*

When a plane containing an opening is placed in front of a screen and illuminated, a careful examination of the shadow reveals the following:

1. If the screen is far from the opening, an outline of the light source is seen. The edges of this image are fuzzy, consisting of alternating light and dark bands.
2. If the screen is close to the opening, the outline of this aperture is now seen, again with alternating light and dark bands.

Diffraction is actually a kind of interference. Waves from one part of the opening interfere with waves from another part to produce the diffraction pattern; the maxima and minima which appear in the pattern indicate that

interference is taking place. We shall demonstrate that *it is the finite size of an aperture which is responsible for diffraction effects;* thus, an unrestricted wave does not produce a diffraction pattern.

Huyghens was an early proponent of the wave theory of light. In 1678 he explained the occurrence of double refraction in calcite by assuming that light was a wave and proposing a mechanism for wave propagation in crystals. *Huyghens' principle*[10-1),(10-2)] states that each point on a wave front acts as a source of new waves. In a given time, the new waves advance a certain distance, and the envelope of all these secondary wavelets (obtained by drawing a tangent surface) yields the new wave front.

In 1818, Fresnel realized that Huyghens' principle could be used not only to describe double refraction in crystals but also to calculate diffraction patterns of apertures. A knowledge of the wave pattern across a slit (or other aperture) and the employment of this principle permit the pattern on a screen behind the slit to be determined.

Our approach to this topic will be to first give an introductory description of diffraction effects, involving Fresnel's ideas expressed in terms of an integral. Then we will note that the diffraction pattern far from an object is the Fourier transform of the object.* Looking at diffraction from this new point of view, we shall see that an image can be filtered so as to enhance certain of its features. It is remarkable that people have been applying the basic diffraction integral for over one hundred years but that only recently has it been realized that it could be put to active use to process optical images.

There are two classes of diffraction that are normally of interest, corresponding to different approximations used in the evaluation of the diffraction integral. The field distribution far from the object (at infinity) is called the *Fraunhofer diffraction pattern* and is mathematically simpler. The field distribution near the object is called the *Fresnel diffraction pattern.* Its calculation, which is quite complicated, is performed using Fresnel integrals.

The Fraunhofer pattern can be observed by placing a screen at a long distance from the object (as an approximation to infinity). Another way the Fraunhofer pattern can be seen is by placing a double convex lens behind the object; the pattern then appears at the focal plane of the lens. All the diffracted rays at a given angle are focused to a point by the lens, which is equivalent to looking at them from infinity. It is important to realize however that the diffraction pattern arises from the aperture, and not the lens; the lens is used only for convenience in observing the pattern.

10-2 Fraunhofer Diffraction

Let us use Huyghens' principle to find the amplitude $A(x', y')$ of a wave at a point P with coordinates x' and y' on a screen (Fig. 10-1), given its

*An introductory treatment of Fourier series, integrals, and transforms is given in the Appendix.

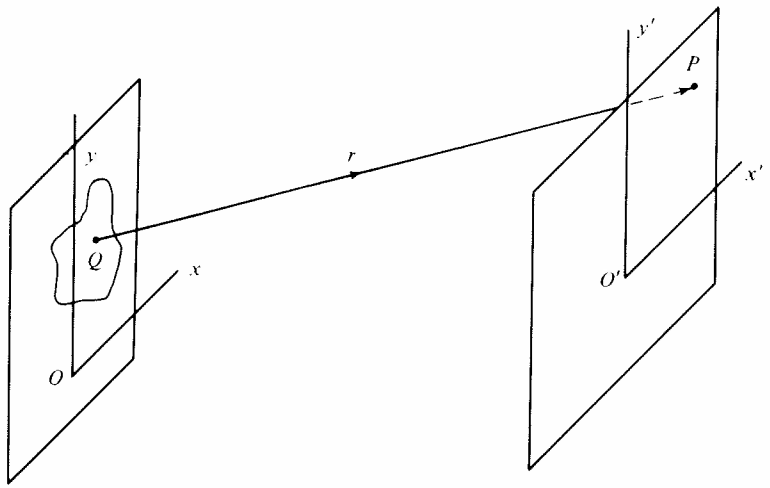


Figure 10-1

amplitude $A(x, y)$ at the point Q with coordinates x and y in an aperture. The screen is located at a long distance from the aperture, so that $r = \overline{QP}$ is large compared to the dimensions of the opening. As shown in connection with Eq. (7-5), the phase change δ at P with respect to Q is

$$\delta = \frac{2\pi}{\lambda} r = kr$$

where the relation between k and λ is given by Eq. (5-2). Because of the large value of r, we can assume that every point (x, y) in the aperture is approximately equidistant from the point (x', y') on the screen, where the complex amplitude will be given by an exponential term, in accordance with Eq. (5-6). Hence, we simply multiply $A(x, y)$ by the area $dS = dx\,dy$ of an elementary section of the aperture and by a phase factor $\exp(i\delta) = \exp(ikr)$ to obtain the expression

$$dA(x', y') = CA(x, y)e^{ikr}\,dx\,dy$$

where the constant C incorporates any necessary dimensional or scale factors. In accordance with Huyghens' principle, the total amplitude on the screen is the sum of all the individual contributions from the aperture, so that

$$A(x', y') = C \int A(x, y)e^{ikr}\,dS \qquad (10\text{-}1)$$

In accordance with our assumption about the relation between the aperture size and r, we can absorb $A(x, y)$ into C and obtain the simpler expression

$$A = C \int e^{ikr}\,dS \qquad (10\text{-}2)$$

where A refers to the amplitude at the screen. Equation (10-2) is the *Fraunhofer diffraction equation.* This expression gives the amplitude of the wave; to find the *intensity* I (as in Section 5-2), we take the square of A, for the intensity is a measure of the energy carried by the wave. This is analogous to the situation in electric circuits: we find the power from the square of the voltage or current, either of which specifies the amplitude.

10-3 Fraunhofer Diffraction by a Single Slit

Suppose we have a slit of width w and of length l (Fig. 10-2). Let us stipulate that the slit is narrow, so that $w \ll l$. We choose axes in the customary manner. Consider two parallel rays in the XOZ plane making an angle θ to the Z axis, as shown. At some point on the upper ray at a distance r from the slit, we write

$$r = r_0 + x \sin \theta \qquad (10\text{-}3)$$

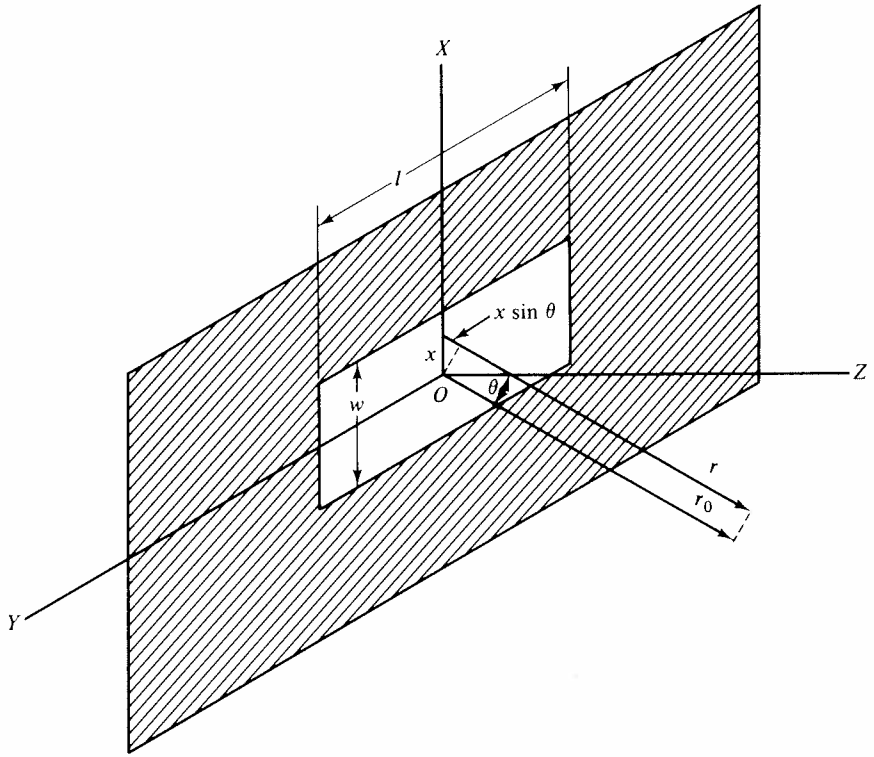


Figure 10-2

where r_0 is the corresponding value of r on the lower ray (the one for which $x = 0$). The element of area in the slit is taken as

$$dS = l\,dx \tag{10-4}$$

Then Eq. (10-2) becomes

$$A = Ce^{ikr_0} l \int_{-w/2}^{w/2} e^{ikx\sin\theta}\,dx$$

Performing the integration gives

$$\begin{aligned} A &= Ce^{ikr_0} l \frac{e^{ikx\sin\theta}}{ik\sin\theta}\Big|_{-w/2}^{w/2} \\ &= Ce^{ikr_0} l \frac{\sin\{(kw\sin\theta)/2\}}{k\sin\theta} \\ &= C'\left(\frac{\sin\alpha}{\alpha}\right) \end{aligned} \tag{10-5}$$

where

$$\alpha = \tfrac{1}{2}kw\sin\theta$$

and

$$C' = Ce^{ikr_0}\frac{wl}{2}$$

The intensity I by Eq. (5-15) is

$$I = \frac{1}{2}AA^* = I_0\left(\frac{\sin^2\alpha}{\alpha^2}\right) \tag{10-6}$$

where $I_0 = \frac{1}{2}C'C'^*$ is the value of I for $\alpha = 0$; $S = wl$ is the area of the slit.

The curve of $(\sin\alpha/\alpha)^2 = (I/I_0)$ vs α is plotted in Fig. 10-3. The principal maximum occurs at $\theta = 0$, or $\alpha = 0$, and there are secondary maxima which

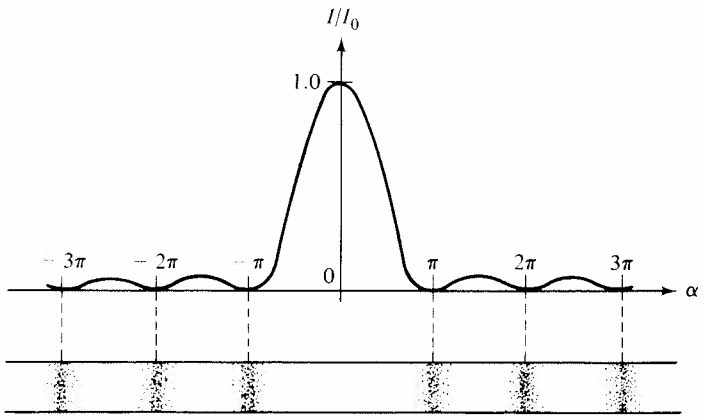


Figure 10-3

can be located by differentiating Eq. (10-6). That is,

$$\frac{dI}{d\alpha} = \frac{2}{I_0}\left(\frac{\sin\alpha}{\alpha}\right)\left(\frac{\alpha\cos\alpha - \sin\alpha}{\alpha^2}\right) = 0$$

or

$$\alpha = \tan\alpha$$

This transcendental equation may be solved numerically or graphically, and we find that the roots are

$$\alpha = 0,\ 1.43\pi,\ 2.46\pi,\ \ldots$$

The minima correspond to $\sin\alpha = 0$ (but $\alpha \neq 0$) which gives

$$\alpha = m\pi, \qquad m = \pm 1, \pm 2, \ldots$$

or

$$\sin\theta = \frac{m\lambda}{w} \tag{10-7}$$

For small angles, this becomes

$$\theta = \frac{m\lambda}{w}$$

and the bright central band has a total angular width of $2\lambda/w$. The diffraction pattern therefore consists of this intense band, with weaker bands on either side (Fig. 10-3). A narrow slit will then produce a wide, but weak, central maximum which shrinks and becomes brighter if w is increased.

For a rectangular or square aperture, we integrate over both x and y and obtain

$$\frac{I}{I_0} = \left(\frac{\sin^2\alpha}{\alpha^2}\right)\left(\frac{\sin^2\beta}{\beta^2}\right) \tag{10-8}$$

where the definition of β is analogous to that of α. The two-dimensional diffraction pattern of a rectangle is therefore a series of orthogonal light and dark bands.

10-4 Fraunhofer Diffraction by a Circular Aperture

Let us change the slit in Section 10-3 to a circular opening of radius R (Fig. 10-4). We take the element of area as a strip parallel to the y axis of width dx and length $2\sqrt{R^2 - x^2}$ (Fig. 10-5). If we use Eq. (10-3) again, Eq. (10-2) becomes

$$A = 2Ce^{ikr_0}\int_{-R/2}^{R/2} e^{ikx\sin\theta}\sqrt{R^2 - x^2}\,dx \tag{10-9}$$

Letting

$$u = \frac{x}{R}$$

and

$$\rho = kR\sin\theta$$

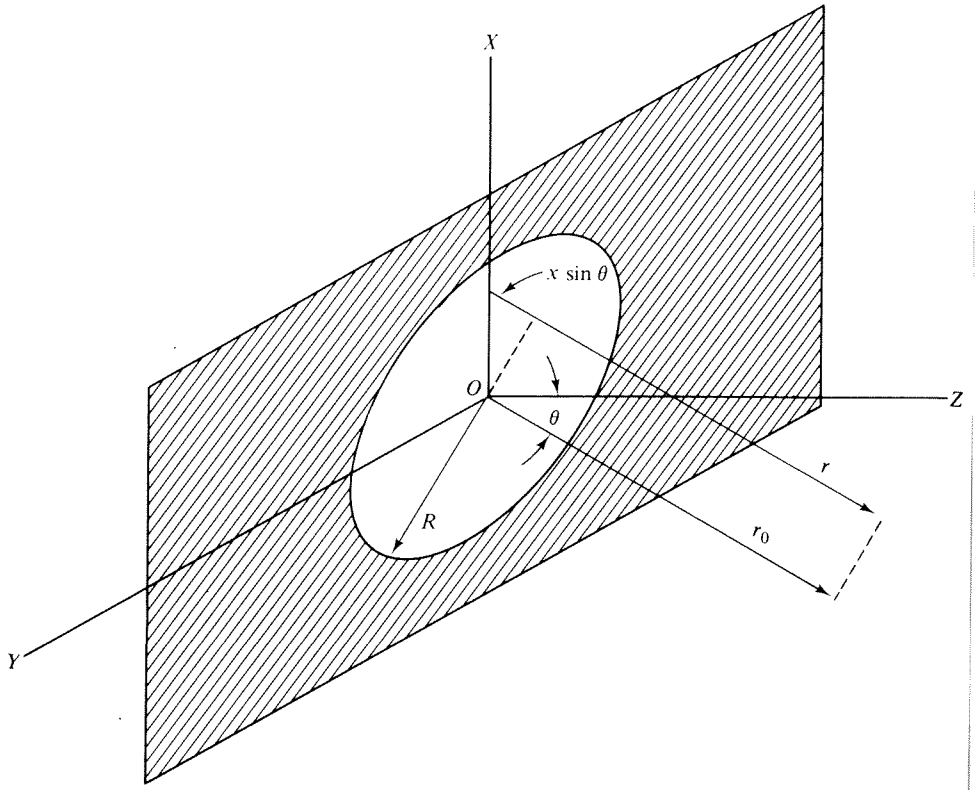


Figure 10-4

Eq. (10-9) is converted to

$$A = 2Ce^{ikr_0}R^2 \int_{-1}^{1} e^{i\rho u}\sqrt{1-u^2}\,du \tag{10-10}$$

The integral in this expression may be written as

$$\int_{-1}^{1} e^{i\rho u}\sqrt{1-u^2}\,du = \int_{-1}^{1} \cos(\rho u)\sqrt{1-u^2}\,du + i\int_{-1}^{1} \sin(\rho u)\sqrt{1-u^2}\,du \tag{10-11}$$

The second term on the right vanishes because $\sin(\rho u)\sqrt{1-u^2}$ is an odd function, and the first term becomes

$$\int_{-1}^{1} \cos(\rho u)\sqrt{1-u^2}\,du = 2\int_{0}^{1} \cos(\rho u)\sqrt{1-u^2}\,du$$

since the integrand is an even function. This new integral can be evaluated numerically, or it may be expressed as a *Bessel function* J_1, defined[10-3] by

$$J_1(z) = \frac{2z}{\pi}\int_0^1 (1-t^2)^{1/2}\cos(zt)\,dt \tag{10-12}$$

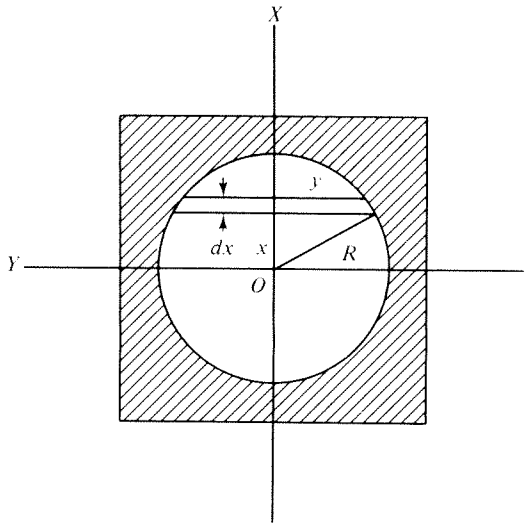


Figure 10-5

Combining this with Eq. (10-10) gives

$$A = \frac{2Ce^{ikr_0}\pi R^2 J_1(\rho)}{\rho} \tag{10-13}$$

Further, when $\theta = 0$, then $\rho = 0$, and Eq. (10-12) shows that

$$\left.\frac{J_1(\rho)}{\rho}\right|_{\rho=0} = \left.\frac{2}{\pi}\int_0^1 (1-u^2)^{1/2}\cos(\rho u)\,du\right|_{\rho=0} = \frac{2}{\pi}\int_0^1 (1-u^2)^{1/2}\,du$$
$$= \frac{2}{\pi}\left[\sqrt{\frac{1-u^2}{2}} + \frac{1}{2}\arcsin u\right]\Bigg|_0^1 = \frac{1}{2} \tag{10-14}$$

Since the intensity is $I = \frac{1}{2}A^2$, then the intensity I_0 for $\theta = 0$ is

$$I_0 = \frac{1}{2}(Ce^{ikr_0}S)^2$$

where S is the area of the aperture. By Eq. (10-13)

$$\frac{I}{I_0} = \left[\frac{2J_1(\rho)}{\rho}\right]^2 \tag{10-15}$$

Problem 10-1

(a) Find $J_1(\rho)$ for $\rho = 0, 0.5, 1.0, \ldots, 10.0$ by numerical integration, verifying Fig. 10-6.

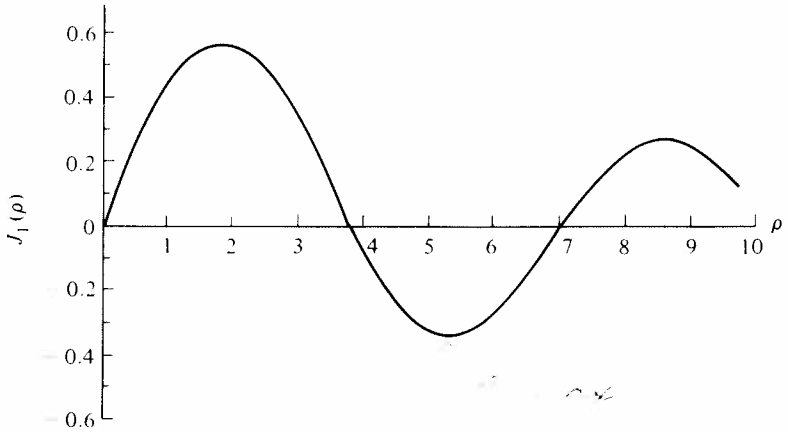


Figure 10-6

(b) Use this result to find $[2J_1(\rho)/\rho]^2$ as part of the same computer program, verifying Fig. 10-6.

(c) Compare Fig. 10-7 with Fig. 10-3. ■

Using the results of Problem 10-1, this expression is plotted in Fig. 10-7, and the alternating light and dark circles corresponding to this intensity pattern are also shown. The bright central area is known as the *Airy disc*. The first dark ring corresponds to the first zero of $J_1(\rho)$, for which $\rho = 3.832$ (Fig. 10-6). Since

$$\sin\theta = \frac{\rho}{kR}$$

we have approximately

$$\theta = \frac{3.83\lambda}{2\pi R} = \frac{1.22\lambda}{D} \tag{10-16}$$

where D is the diameter of the aperture. This formula is of great practical importance because the image formed by a lens in a camera, for example, is actually the Fraunhofer diffraction pattern of each point in the object.

If D is the diameter of the lens (or the entrance pupil), then Eq. (10-16) also gives the minimum angular separation of two points whose Airy discs might hopefully be distinguished in the image plane. This follows from the fact that, if the angular spacing is $1.22\lambda/D$, then the central maximum of one disc corresponds to the first zero of the other one (Fig. 10-8). We realize that this condition is similar to that of Section 7-2. It is called the *Rayleigh criterion*, and depends on the ability of the eye to separate two discs placed so that the center of one is on the poorly defined periphery of another. It is to some extent arbitrary, but represents a reasonable approximation of the eye's image-resolving ability.

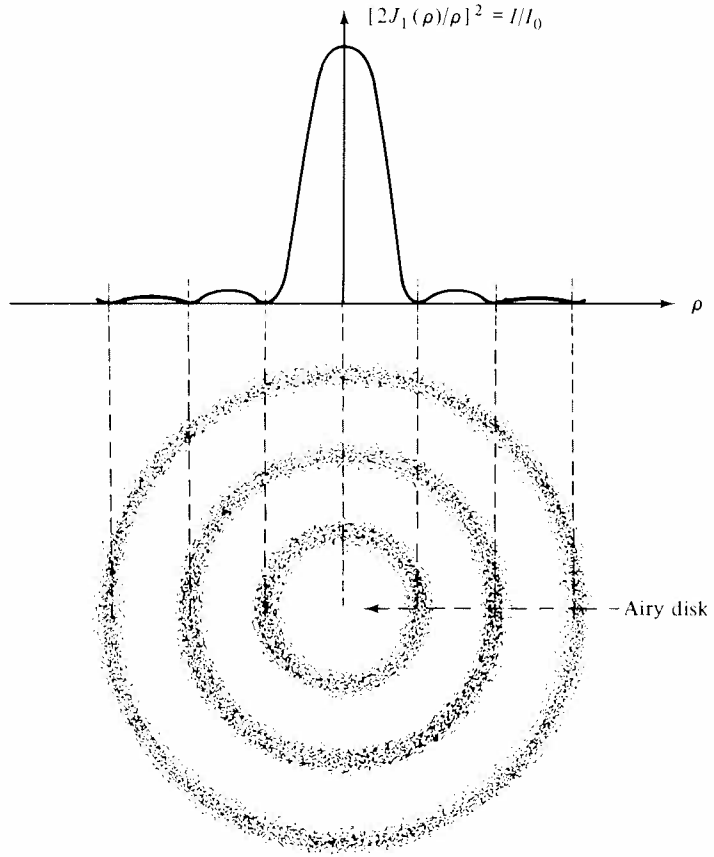


Figure 10-7

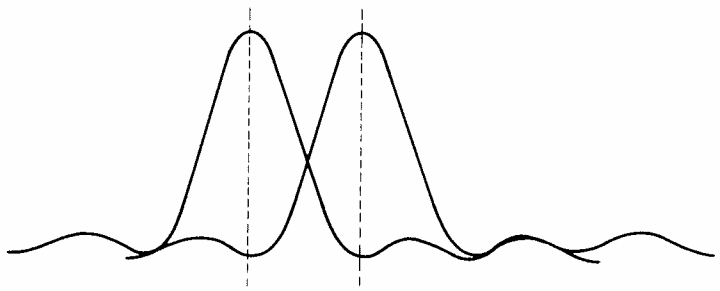


Figure 10-8

Problem 10-2

How big should the diameter of a lens be in order to just resolve the limbs (two points on the ends of a diameter) of the moon? ■

We should realize at this point that diffraction effects place limitations on image quality, in addition to those due to aberrations. For example, a large double convex lens of radius R and focal length f used as the objective of a telescope will have a significant amount of chromatic aberration. In fact, we have seen in Chapter 4 that the transverse chromatic aberration x_C is given by

$$x_C = -\frac{RV}{2} \tag{4-82}$$

where V is the dispersive power. Let us assume that the image spread due to chromatic aberration and the image spread due to diffraction are additive, and find the relation between R and f which gives the sharpest image.

By (10-16), we obtain the spread x_D in the image due to diffraction from

$$\theta = \frac{x_D}{f} = \frac{1.22\lambda}{D} \quad \text{or} \quad x_D = \frac{1.22\lambda f}{2R}$$

Adding the magnitudes of x_D and x_C and then minimizing this sum by differentiating and equating the result to zero gives

$$\frac{dx}{dR} = -\frac{1.22\lambda f}{2R^2} + \frac{V}{2} = 0$$

or

$$R = \sqrt{\frac{1.22\lambda f}{V}} \tag{10-17}$$

Problem 10-3

The light-gathering power of a lens is called its *speed*. This property should depend on the area of a simple lens or the area of the entrance pupil for a system. It should also depend on the solid angle subtended by the image at the lens. This angle increases as the image comes closer to the lens. The speed is inversely proportional to the magnitude f' of the focal length. Hence, the speed S is defined as the ratio of the diameter to the focal length, or

$$S = \frac{d}{f'} \tag{10-18}$$

It is customary to express speed in terms of *f-numbers*, which are the reciprocals of S. For example, a miniature camera typically has a lens for which $f = 5.0$ cm and $d = 1.8$ cm. Then

$$S = \frac{1.8}{5.0} = 0.36$$

and

$$f\text{-number} = \frac{1}{S} = f : 2.8$$

If the diameter is increased to 2.5 cm, the f-number decreases to $f : 2$, which is a rather fast lens. Note that since the area is proportional to the diameter squared, doubling the f-number raises the speed by a factor of 4.0.

Show that the diameter of the Rayleigh disc for a lens, when expressed in units of 10^{-6} m (one micrometer is 10^{-6} m), approximately equals the f-number. ■

Problem 10-4

Lipson and Lipson[10-4] point out that the involved analysis which produced Eq. (10-16) can be replaced with a simple approximation. Consider a rectangular slit whose length is equal to the diameter of the circular opening in Fig. 10-5. Let the two openings have the same area. Show that the first minimum in the diffraction pattern for the rectangular slit is specified by the relation

$$\theta = \frac{1.27\lambda}{D} \tag{10-19}$$

This represents a deviation of about 5% from the value given by Eq. (10-16). ■

Problem 10-5

Let the rectangular opening of Fig. 10-3 be replaced with an infinite sheet of transparent material of nonuniform density. We shall specify the light-transmission properties along the x direction by the Gaussian function

$$f(x) = Be^{-x^2/b^2} \tag{10-20}$$

where B and b are constants. Show that the diffraction pattern is also Gaussian and, therefore, does not show the secondary maxima of the rectangular and circular apertures. This is a consequence of the fact that a Gaussian opening does not terminate abruptly; its transmission properties decay exponentially on either side of the center, and there is no sharply defined edge. ■

Problem 10-6

A metal plate has two rectangular slits arranged as shown in Fig. 10-9. Verify that

$$\frac{I}{I_0} = \left(\frac{\sin\alpha}{\alpha}\right)^2 \cos^2\gamma \tag{10-21}$$

where

$$\alpha = \tfrac{1}{2}kw\sin\theta \tag{10-22}$$

$$\gamma = \tfrac{1}{2}kd\sin\theta \tag{10-23}$$

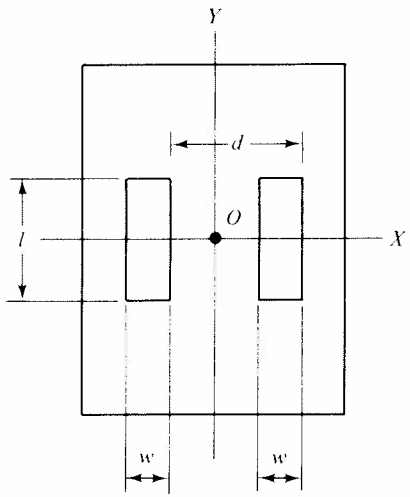


Figure 10-9

Equation (10-13) is the solid curve of Fig. 10-10. This curve may be interpreted as the function $[(\sin \alpha)/\alpha]^2$ of Eq. (10-6) combined with $\cos^2 \gamma$. That is, the behavior for a single slit is the envelope in Fig. 10-10, and the structure in the curve is contributed by the interaction of the two slits. ■

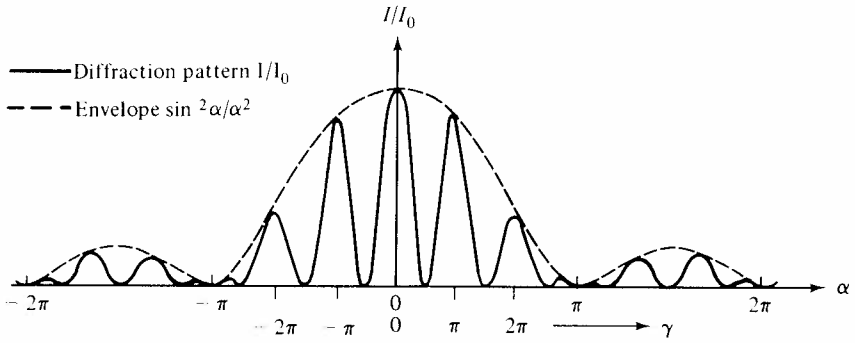


Figure 10-10

10-5 Apodization

The diffraction patterns of Figs. 10-3 and 10-7 show that most of the energy passing through an opening appears in the central portions of the image rather than in the secondary peaks. In situations for which the total usable

energy is critical, such as observations involving very faint stars, it is desirable to increase the energy in the central section at the expense of the secondary peaks. This can be accomplished if the transmission properties of the opening are changed, somewhat along the lines indicated by Problem 10-5. The process of redistributing the energy is called *apodization*, from the Greek word meaning "without feet," and it refers to the reduction of the secondary maxima. As an example, the rectangular slit of Fig. 10-2 can be covered with a glass plate containing a coating of variable density and arranged so that the transmission of light is the greatest at the center (i.e., along the y axis) and falls to zero at the edges in accordance with the relation

$$A(x) = \cos\left(\frac{\pi x}{w}\right) \tag{10-24}$$

where $A(x)$ is the *aperture or pupil function*. It specifies the ratio of the transmitted amplitude to the incident amplitude. As Eq. (10-24) indicates

$$A = \begin{cases} 1 & \text{at} \quad x = 0 \\ 0 & \text{at} \quad x = \pm\dfrac{w}{2} \end{cases}$$

Equation (10-2) for this situation then becomes

$$A = Ce^{ikr_0}l \int_{-w/2}^{w/2} \cos\left(\frac{\pi x}{w}\right) e^{ikx \sin\theta}\, dx \tag{10-25}$$

Problem 10-7

(a) Evaluate the integral in Eq. (10-25) to obtain

$$\frac{I}{I_{0A}} = \left(\frac{\pi^2 \cos\alpha}{\pi^2 - 4\alpha^2}\right)^2 \tag{10-26}$$

where I_{0A} is the value of the intensity for the apodized slit when $\alpha = 0$.

(b) Show that

$$I_{0A} = 1.625 I_0 \tag{10-27}$$

where I_0 refers to the original rectangular slit.

(c) Verify the diffraction pattern of Fig. 10-11 with a numerical calculation. ■

As Fig. 10-11 shows, the use of the absorption plate has put more energy into the central maximum and reduced the height of the secondary maxima. This agrees with what Problem 10-5 indicates: reducing the sharpness of the transition from the transparent to the opaque region has caused a redistribution. Note, however, that only an exponential transition, which requires an infinite distance to go from $A(x) = 1$ to $A(x) = 0$, will put all the energy into the central maximum.

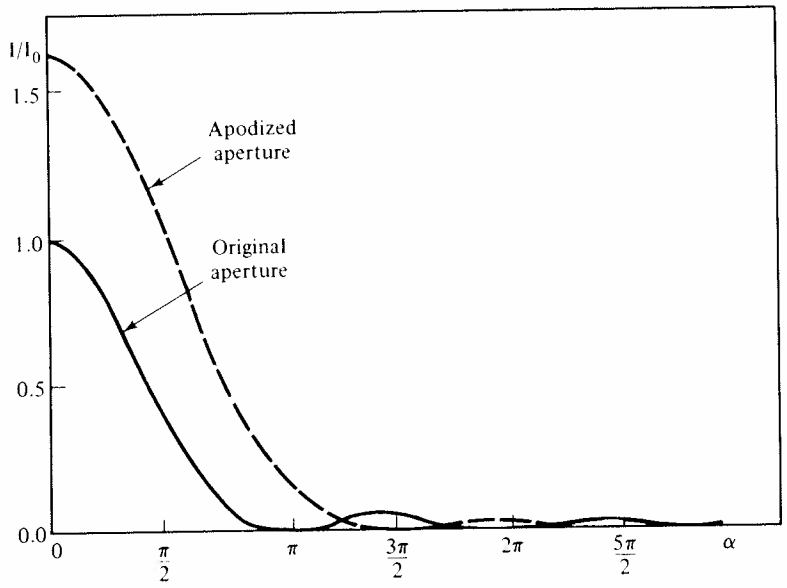


Figure 10-11

10-6 Fresnel Diffraction

Waves whose behavior may be expressed by the relation

$$\mathbf{A} = \mathbf{A}_{\max} e^{\pm i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{5-7}$$

were described as plane waves in Chapter 5. When a wave is generated by a point source and is observed close to the source, it is a *spherical wave* rather than a plane wave and the spatial dependence of its amplitude varies as $\{\exp(ikr)\}/r$. This makes the intensity fall off as $1/r^2$, where r is the distance from the source, agreeing with the results of our study of illumination in Chapter 4. Such a wave results in *Fresnel diffraction*, which occurs whenever the source or the observation point is close to the diffracting body. As an example, let us consider a plane wave passing through a slit. With the aid of Fig. 10-2 we find that the diffraction pattern is a consequence of the phase shift ($x \sin\theta$) between the two rays shown. When we move the observation point close to the source S as shown in Fig. 10-12, the phase shift must be measured by the path difference between the central ray SOS' and the ray STS' at a distance R from the axis. The length of this latter ray is

$$r + r' = \sqrt{h^2 + R^2} + \sqrt{h'^2 + R^2} \tag{10-28}$$

If R is small compared to h and h' (S or S' is close to O, but not so close that

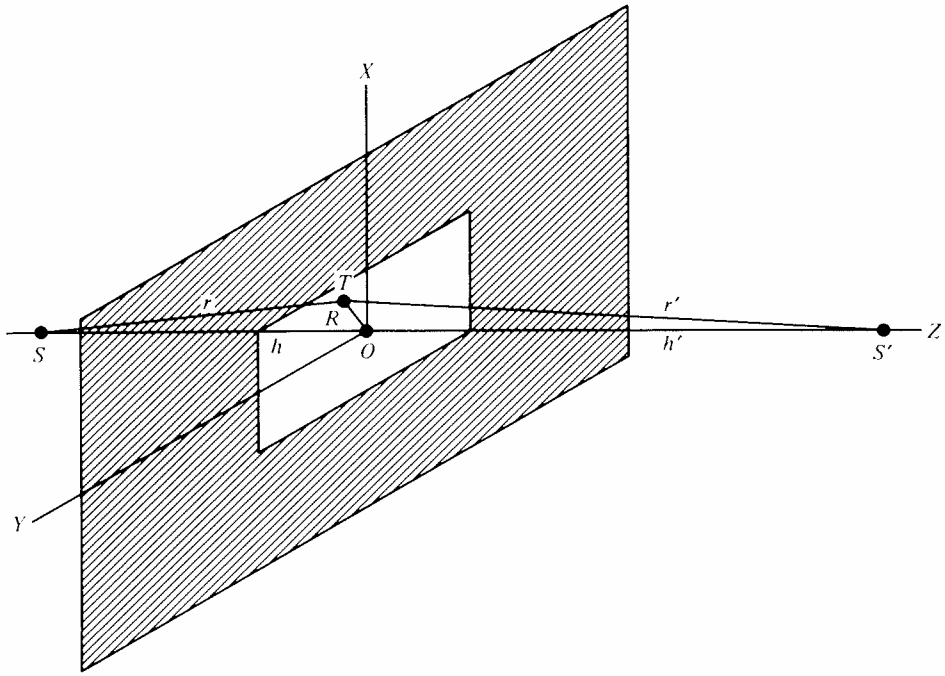


Figure 10-12

R seems large), we may write the first term on the right as

$$\sqrt{h^2 + R^2} = h\sqrt{1 + \left(\frac{R}{h}\right)^2} \sim h\left(1 + \frac{R^2}{2h^2}\right)$$

If we use a similar expression for the second radical, Eq. (10-28) becomes

$$r + r' = (h + h') + \frac{R^2}{2}\left(\frac{1}{h} + \frac{1}{h'}\right)$$

Letting

$$x^2 + y^2 = R^2$$

and defining a length L by

$$\frac{1}{L} = \frac{1}{h} + \frac{1}{h'}$$

we obtain

$$r + r' = (h + h') + \frac{1}{2L}(x^2 + y^2)$$

The integral Eq. (10-2), with a path length of $(r + r')$, becomes

$$A = Ce^{ik(h+h')} \iint e^{ik(x^2+y^2)/2L}\, dS \tag{10-29}$$

The change in the factor $1/(r + r')$ is small compared to the exponential term and can be neglected.

If we introduce the dimensionless variables

$$u = x\sqrt{\frac{k}{\pi L}}, \qquad v = y\sqrt{\frac{k}{\pi L}} \tag{10-30}$$

Eq. (10-29) becomes

$$A = D \int e^{i\pi u^2/2}\, du \int e^{i\pi v^2/2}\, dv \tag{10-31}$$

where D is a new constant. These integrals are converted by the identity

$$\int_0^w e^{i\pi w^2/2}\, dw = \int_0^w \cos\left(\frac{\pi w^2}{2}\right) dw + i \int_0^w \sin\left(\frac{\pi w^2}{2}\right) dw \tag{10-32}$$

The two terms on the right, designated as

$$\left.\begin{aligned} C(w) &= \int_0^w \cos\left(\frac{\pi w^2}{2}\right) dw \\ S(w) &= \int_0^w \sin\left(\frac{\pi w^2}{2}\right) dw \end{aligned}\right\} \tag{10-33}$$

are known as the *Fresnel integrals*, and may be evaluated numerically (See Problem 10-8a). If we plot $S(w)$ vs $iC(w)$, we obtain the *Cornu spiral* shown in Fig. 10-13.

To apply these results to the case of the rectangular slit, we must evaluate Eq. (10-29) over the area enclosed by the aperture. Consider a rectangular opening whose edges are determined by the coordinates x_1, x_2, y_1, and y_2. Then the limits of integration are u_1, u_2 and v_1, v_2 as obtained from Eq. (10-30). In the limiting case of no diffraction at all, we have $u_1 = -\infty$, $u_2 = \infty$ and $v_1 = -\infty$, $v_2 = \infty$, i.e., the opening has been removed. From Fig. 10-13:

$$C(\infty) = S(\infty) = 0.50$$
$$C(-\infty) = S(-\infty) = -0.50$$

Hence

$$\begin{aligned} A &= D \int_{-\infty}^{\infty} e^{i\pi u^2/2}\, du \int_{-\infty}^{\infty} e^{i\pi v^2/2}\, dv \\ &= D[C(u) + iS(u)]\Big|_{-\infty}^{\infty} [C(v) + iS(v)]\Big|_{-\infty}^{\infty} \\ &= D[(1 + i)(1 + i)] \\ &= D(1 + i)^2 \end{aligned}$$

and

$$I_0 = AA^* = 4D^2 \tag{10-34}$$

where I_0 is the intensity for no diffraction and A^* is the complex conjugate of A.

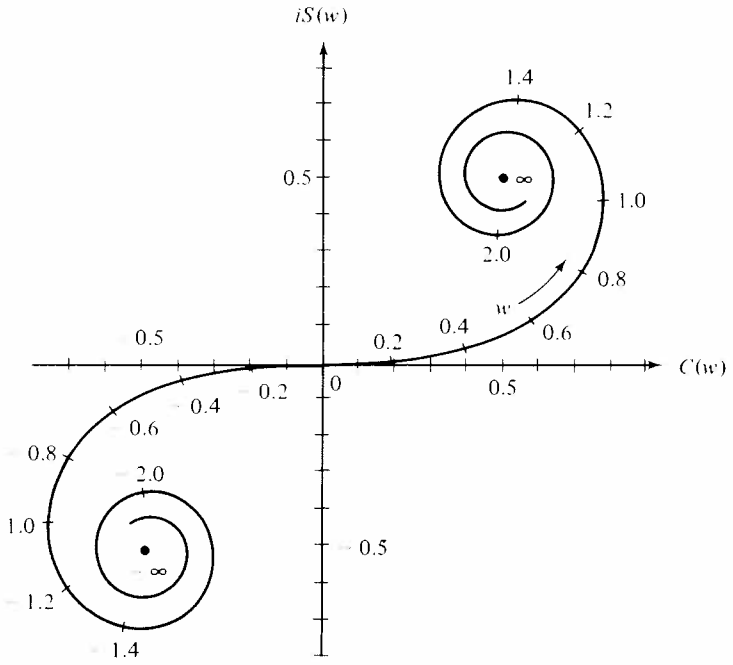


Figure 10-13

For the case of a long slit parallel to the x axis, we keep $u_1 = -\infty$ and $u_2 = \infty$, but let v_1 and v_2 specify the edges of the slit. Finally, we consider a straightedge located at $v_2 = v$, and let $v_1 = -\infty$. Then

$$\begin{aligned} A &= D\int_{-\infty}^{\infty} e^{i\pi u^2/2}\, du \int_{-\infty}^{v} e^{i\pi v^2/2}\, dv \\ &= D(1 + i)[C(v) + iS(v) + 0.5 + 0.5i] \end{aligned}$$

For an observation point S' on the z axis, the value $v = 0$ means that the source S, the diffracting edge, and S' all lie on a straight line, that is, S' lies right on the boundary of the shadow. Using

$$C(0) = S(0) = 0.0$$

gives

$$\begin{aligned} A &= D(1 + i)[0.5(1 + i)] \\ &= 0.5A_0 \end{aligned}$$

where A_0 is the unobstructed amplitude, and

$$I = 0.25I_0 \tag{10-35}$$

Moving the edge so that v becomes positive is equivalent to looking at an area on the screen which is fully illuminated. To find the exact values of

$C(v)$ and $S(v)$ corresponding to this position, we can use the computer output, but if we are satisfied with an estimate, we can compute the definite integrals graphically. For example, if we wish to find the integral

$$\int_{-\infty}^{+1.0} e^{i\pi v^2/2}\,dv = \int_{-\infty}^{1.0}\left[\cos\left(\frac{\pi v^2}{2}\right) + i\sin\left(\frac{\pi v^2}{2}\right)\right]dv$$

we draw a vector on the Cornu spiral (Fig. 10-13) from $w = -\infty$ to $w = 1.0$. The projection $\mathfrak{R}$ on the real axis equals the real part of the integral above, since integrating from $-\infty$ to 0 and from 0 to 1.0 is equivalent to the limits of $-\infty$ to 1.0 and the imaginary part $\mathfrak{I}$ is obtained similarly. This process is indicated in Fig. 10-14. The intensity obtained in this way for values of v

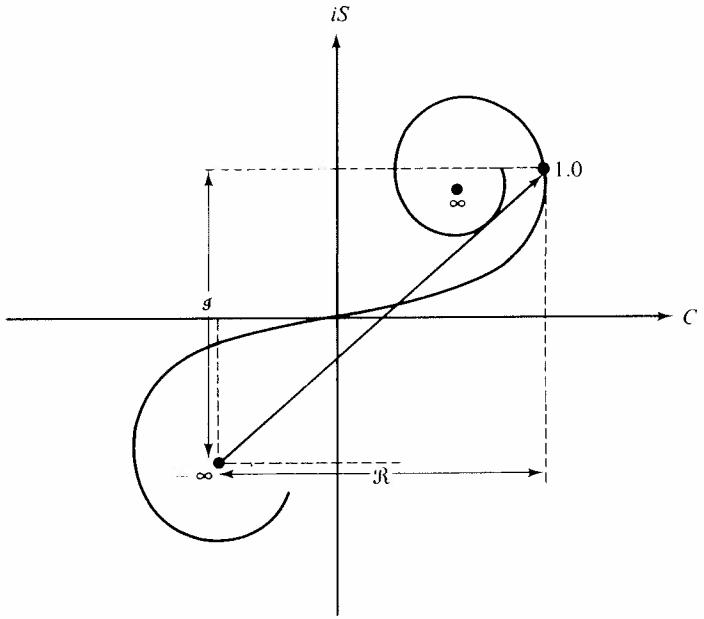


Figure 10-14

corresponding to both the illuminated region ($v > 0$) and the shadow region ($v < 0$) is shown in Fig. 10-15. Note that the diffraction causes a small but finite amount of light to appear in the shadow region.

Problem 10-8

(a) Write a program for evaluating the integrals in Eq. (10-33), verifying Fig. 10-13.

(b) A distant FM transmitter operating at 100 megahertz, a water tower of 10-m diameter, and a radio receiver lie in a straight line. The tower and

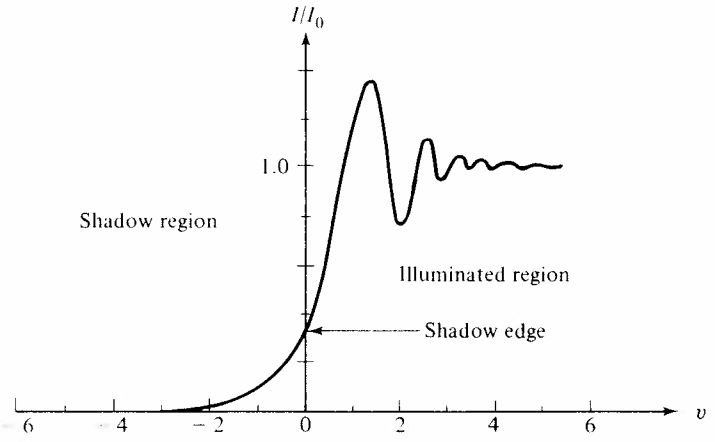


Figure 10-15

receiver are 1000 m apart. Determine the effect of the obstacle on the received signal in an antenna normal to the line joining the transmitter and receiver. ■

Let us make clear again the distinction between Fraunhofer diffraction and Fresnel diffraction. In Fig. 10-2, we considered two parallel rays passing through the slit and we calculated the diffraction pattern *far* from the slit, using the approximation expressed by equation (10-3) and also assuming that the aperture subtended a very small angle at the point of observation. The path difference depended on $x \sin\theta$, that is, it is a linear function of x. *Fraunhofer diffraction corresponds to a plane wave front approaching the object and to a linear exponent in the diffraction integral.* In Fig. 10-12, on the other hand, we dealt with a point source S and a point of observation S', either or both of which were *close* to the aperture. We can think of this as corresponding to a spherical wave. At these distances, $r + r'$ differs from $h + h'$, by an amount $(x^2 + y^2)/2L$. The exponent in the diffraction integral in Eq. (10-29) is now quadratic in x (and y) rather than linear, and the integrals themselves must be evaluated numerically. Hence, Fraunhofer diffrac-

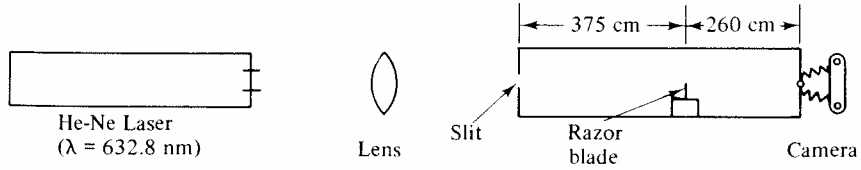


Figure 10-16

tion is observed when both the source and observation point are at long distances from an aperture. (A spherical wave of large radius is approximately equivalent to a plane wave) and Fresnel diffraction is observed when either is a short distance from an aperture.

As an example of the latter, consider the experiment of George[10-5] shown in Fig. 10-16. The lens converts the laser output into a spherical wave, and we are able to see the Fresnel pattern of the edge of the razor blade at a distance which is enormous in comparison to the wavelength ($\lambda = 632.8 \times 10^{-9}$ m) simply because the laser output retains its wave shape almost indefinitely. The intensity of the pattern produced on the film as a function of distance is measured with an instrument called a densitometer. The results shown in Fig. 10-17 agree with the theoretical curve (Fig. 10-15). A photograph of the illuminated region (Fig. 10-18) shows the light and dark bands which correspond to the densitometer curve. A slightly different type of Fresnel experiment was done by Boyer and Fortin.[10-7] They focused a laser with a microscope objective and observed the diffraction pattern created by a wire 0.1 mm in diameter at a distance of 3 m. Their measurements agreed with the theoretical calculation using the Fresnel integrals (rather than the Cornu spiral) to within 4%. An equally simple way of photographing Fraunhofer patterns, using a sodium light source and a camera with a 40-cm telephoto lens, is described by Graham;[10-6] his results are found to agree with Figs. 10-3 and 10-10.

10-7 More Exact Diffraction Theories

The treatment of diffraction given so far has been based on Huyghens' principle and the results we obtained agree with experimental observation. It has long been realized, however, that this theory is lacking in two respects. First, the phase of the reradiated wavelets had to be shifted by $\pi/2$ to give results which were correct under all circumstances, and second, it was predicted that the secondary wavelets would generate the same diffraction pattern in the backward as in the forward direction, an effect which does not occur. If we arbitrarily include an *obliquity factor* $(1 + \cos\theta)$ in Eq. (10-2), where θ is the angle between the direction of propagation of the original wave and the direction of the diffracted wave, then the backward pattern is completely eliminated, for the obliquity factor causes the diffraction integral to vanish when $\theta = \pi$ or $\cos\theta = -1$. Further, it indicates that there is a decrease of amplitude with angle in the forward direction, since $(1 + \cos\theta)$ has its maximum value for $\theta = 0$. Thus, the form of this obliquity factor appears to be very logical.

Kirchhoff in 1882 derived a diffraction integral somewhat like Eq. (10-2), but containing the correct phase and angular terms, by starting with the wave equation, Eq. (5-8). We need not consider his derivation here but simply mention that it involves the use of well-known vector identities (Green's theorems in particular). An alternate approach, involving somewhat more complicated situations, was given by Sommerfeld[10-8]. Neither of these theories, however, takes into account the fact that light is a form of electromagnetic radiation which interacts with material media via its electric and magnetic vectors or fields and that polarization of the light must also be considered. Experiments with microwaves indicated that there are some results which scalar diffraction theory cannot explain and which require the polarization to be taken into account. For details on both the scalar and the vector theory of diffraction, we refer the reader to the standard treatise of Born and Wolf.[5-1]

10-8 Fourier Analysis and Fourier Optics

10-8a Basic Principles

It is generally realized that light waves and radio waves are both forms of electromagnetic radiation; they both can be described by solutions to the wave equation, Eq. (5-8), and possess many properties in common. It is not generally realized, however, that there is a close analogy between the behavior of alternating currents in circuits containing inductance or capacitance and the behavior of light as it passes through an optical system. This parallel situation can be developed on the basis of Fourier analysis, as discussed in the Appendix. We shall briefly review here the aspects of this subject which help clarify the analogy to be developed.

Let us start with the periodic pulse of unit amplitude and wavelength 2π illustrated in Fig. A-3. Equation (A-8), written as

$$y = 0.5 + \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)} \cos(2n+1)x \qquad (10\text{-}36)$$

indicates that this repeating square wave is the sum of a series of cosine terms, each of successively increasing propagation constant n and decreasing amplitude. The decomposition of a train of square waves into its pure sinusoidal components is called a *Fourier analysis* or *spectral decomposition* of the original wave. We can completely specify the nature of the square wave by simply listing the frequencies and amplitudes of the sinusoidal components. This information can be determined electrically, in principle, by passing the wave through a tunable band-pass filter so that each component can be detected and measured separately. By using a rejection filter, we can remove specific components and alter the shape of the wave train.

If we have only a single pulse (Fig. A-1), then the Fourier series in Eq. (10-36) is not valid, and we must express the Fourier decomposition in terms of an integral. That is, instead of summing over a discretely spaced set of

components of the form $\cos x$, $-\frac{1}{3}\cos 3x$, $\frac{1}{5}\cos 5x$, . . . , we use a set of functions whose frequencies are infinitesimally close, and the single pulse is then written as a Fourier integral. To be specific, let the symmetric pulse have unit height and width d, rather than π, so that the definition of this function is

$$f(x) = \begin{cases} 1 & \text{for} \quad -\frac{d}{2} < x < \frac{d}{2} \\ 0 & \text{for} \quad |x| > \frac{d}{2} \end{cases} \tag{10-37}$$

and its Fourier transform is

$$F(k) = \int_{-d/2}^{+d/2} e^{-ikx}\,dx = \frac{1}{-ik} e^{-ikx}\Big|_{-d/2}^{d/2}$$

$$= \frac{2}{k}\sin\left(\frac{kd}{2}\right) = d\frac{\sin(kd/2)}{(kd/2)}$$

$$= d\frac{\sin(\pi d/\lambda)}{(\pi d/\lambda)} \tag{10-38}$$

The resulting function of (d/λ) is known as the *sinc function* and is defined by

$$\operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x} \tag{10-39}$$

as already plotted in Fig. A-13(c). We have previously encountered it in connection with Fraunhofer diffraction by a slit, the amplitude being

$$A = C' \operatorname{sinc} \alpha \tag{10-5}$$

where

$$\alpha = \frac{\pi w \sin\theta}{\lambda}$$

Hence, we expect the amplitude of the diffraction pattern to be the Fourier transform of the aperture function $A(x, y)$, and this we shall see is the case. A slit of finite width causes a plane wave to interfere with itself in such a way that it generates the Fourier spectrum of the aperture function. If the slit is simply an opening, we are finding the transform of $f(x)$ as given by Eq. (10-37), but if there is apodization or phase shift, then the Fourier transform corresponds to a more complicated aperture function.

The diffraction integral that we use Eq. (10-1) can also be written as

$$A(x', y') = C\iint A(x, y)e^{-ikr}\,dx\,dy$$

without changing the resulting intensity. For a rectangular opening, for example, we have

$$A(x', y') = C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x, y)e^{-ik(x+y)}\,dx\,dy \tag{10-40}$$

As shown in the Appendix, Eq. (A-24), the Fourier transform of the function $f(x, y)$ is

$$\mathfrak{F}[f(x, y)] = F(k_x, k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)e^{-i(k_x x + k_y y)}\,dx\,dy \tag{10-41}$$

Hence, Eq. (10-40) will have this form if we let $k_x = k_y = k$ and if $C = 1$. Since the constant does not usually have the value unity, we should say that the intensity of the Fraunhofer diffraction pattern is proportional (rather than identical) to the Fourier transform of the aperture function. Thus, *diffraction phenomena provide a method for optically computing Fourier transforms.*

Problem 10-9

(a) Justify the limits specified in Eq. (10-40).

(b) Eq. (10-40) expresses the amplitude $A(x', y')$ of the diffraction pattern at a point (x', y') on the screen in terms of an integral involving the amplitude $A(x, y)$ over the aperture. On the other hand, the Fourier transform definition, Eq. (10-41), implies that the amplitude of the diffracted wave at the receiving screen should be a function of k. Show that the calculation of Section 10-3 permits us to write $A(x', y')$ as $A(k)$, thus matching the definition. We shall show immediately below that when a lens is part of a diffraction system, both k and the coordinates x', y' are explicitly involved. ■

When we consider the effect of a lens on a plane wave, the situation becomes more complicated. Suppose a ray of light leaves a lens at a point $B(x, y)$ and strikes the focal plane at a point $B'(x', y')$. To simplify the geometrical calculation which follows, assume for the moment that we are dealing with a meridional problem. Let the point B be in the unit plane of a simple lens (Fig. 10-19). The path difference Δr between a ray with a slope

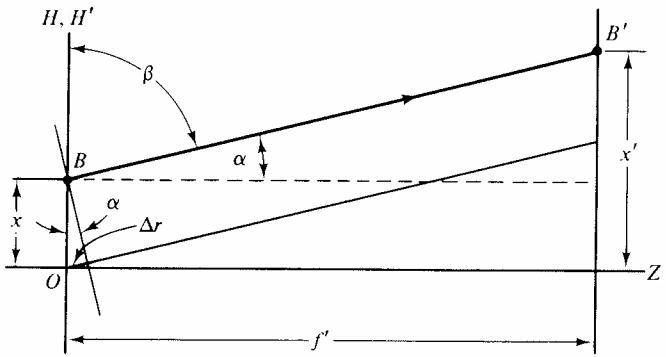


Figure 10-19

α leaving B and another ray from the origin is related to α through the expression

$$\sin \alpha = \frac{\Delta r}{x}$$

or

$$\Delta r = x \cos \beta$$

If x is small, then

$$\tan \alpha = \frac{x' - x}{f'} \sim \frac{x'}{f'}$$

But

$$\tan \alpha \sim \sin \alpha = \cos \beta$$

hence,

$$\Delta r = \frac{xx'}{f'} \tag{10-42}$$

Generalizing to skew rays, Eq. (10-42) will have the form

$$\Delta r = \frac{xx' + yy'}{f'} \tag{10-43}$$

The diffraction integral in Eq. (10-40) then becomes

$$A(x', y') = C \iint P(x, y) e^{-ik(xx' + yy')/f'} \, dx \, dy \tag{10-44}$$

where we have introduced a *pupil function* $P(x, y)$ to explicitly indicate that the lens affects the amplitude of the light passing through it.

Let us define the *spatial frequencies* ν_x and ν_y associated with x' and y', respectively, by

$$\left.\begin{aligned} \nu_x &= \frac{kx'}{f'} = \frac{2\pi x'}{\lambda f'} \\ \nu_y &= \frac{ky'}{f'} = \frac{2\pi y'}{\lambda f'} \end{aligned}\right\} \tag{10-45}$$

so that

$$A(\nu_x, \nu_y) = C \iint P(x, y) e^{-i(\nu_x x + \nu_y y)} \, dx \, dy \tag{10-46}$$

By explicitly showing the functional dependence of A on ν_x and ν_y, we see that *the amplitude of the diffracted wave is the Fourier transform of the pupil function.* Hence, if a lens possessed only the paraxial focusing properties expressed in Eq. (10-43), it would serve as a device which takes a uniform plane wave and uses it to form an image in the focal plane whose amplitude is the two-dimensional Fourier transform of $P(x, y)$. The variables ν_x and ν_y are called spatial frequencies because they play a role analogous to that of the angular frequency ω in the Fourier transform definition

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt \tag{A-28}$$

Although ω has units of radians/second and $\nu = \omega/2\pi$ has units of cycles/second or hertz, the quantities ν_x and ν_y are expressed in terms of cycles/meter, as Eqs. (10-45) indicate.

10-8b An Example of a Phase Object

A simple example which illustrates the Fourier transform interpretation of Eq. (10-40) or Eq. (10-46) is the effect on the diffraction pattern of a phase change in the incident wave. We saw in Chapter 8 that two adjacent waves emerging from a thin film of thickness t and index n will have a phase difference δ given by

$$\delta = \left(\frac{2\pi n}{\lambda}\right) 2t \cos \theta \tag{8-3}$$

where θ is the angle of incidence or reflection. If the beam is normal ($\theta = 0$), then $\delta = 4\pi nt/\lambda$ and for a 180° phase shift (or an integral multiple of this value), the thickness is specified by the relation

$$\frac{(2m + 1)\lambda}{4} = nt$$

where $m = 0, 1, 2, \ldots$. Let us use a square aperture of edge d at the entrance pupil of a lens and cover the central part of it with a film of thickness $t = \lambda/4n$, which produces a 180° phase shift in the incoming wave (Fig. 10-20(a)). Then the pupil function is specified as (Fig. 10-20(b))

$$P(x, y) = \begin{cases} -1 & \text{when} \quad -\dfrac{d}{4} \leqq x, y \leqq \dfrac{d}{4} \\ +1 & \text{when} \quad \begin{cases} -\dfrac{d}{2} \leqq x, y \leqq -\dfrac{d}{4} \\ \dfrac{d}{4} \leqq x, y \leqq \dfrac{d}{2} \end{cases} \end{cases} \tag{10-47}$$

Evaluating Eq. (10-46) for a single variable x or y with $P(x, y)$ as given by Eq. (10-47), we obtain an intensity which varies with distance along either the x' or y' directions in the image plane in accordance with the relation

$$I = C\left[\frac{\sin u}{u} - \frac{\sin (u/2)}{u/2}\right]^2 \tag{10-48}$$

where $u = \nu_x d/2$ or $\nu_y d/2$. Since I vanishes for $u = 0$, the constant cannot be set equal to I_0; instead, we arbitrarily let $C = 1.0$ at the first maximum I_1 in I, as shown in Fig. 10-20(c). The effect of placing a phase-shifting film over the central portion of the aperture is thus opposite to apodization; energy in the diffraction pattern is moved away from the central maximum.

Problem 10-10

Verify Eq. (10-48). ■

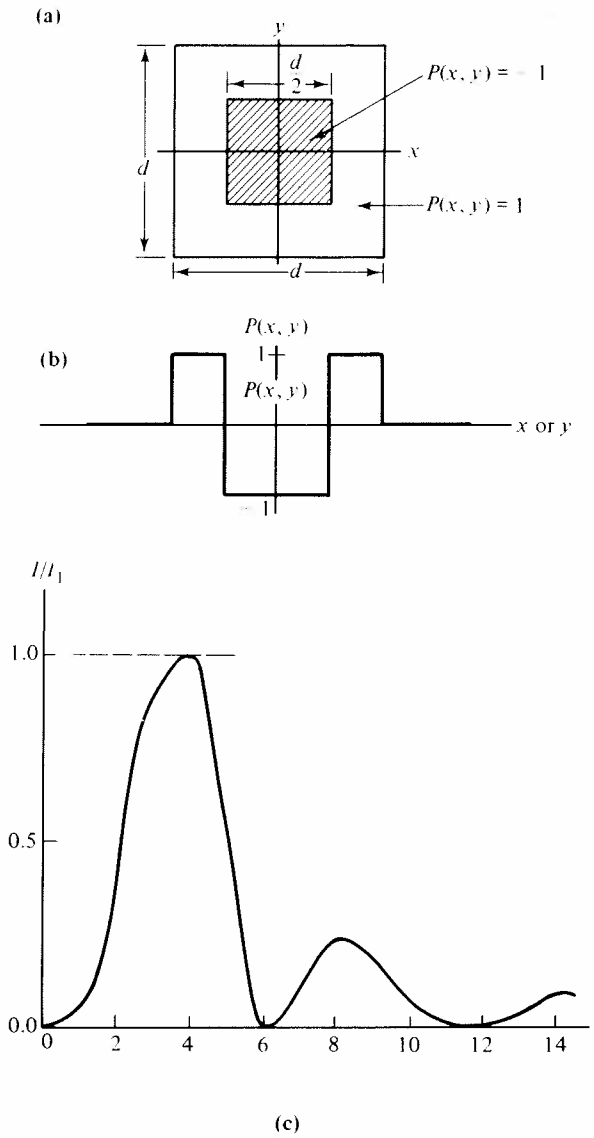


Figure 10-20

It is easy to see that if we had evaluated the diffraction integral Eq. (10-40) for the Fraunhofer pattern of an opening whose aperture function is of the form of Eq. (10-47), we would have obtained a result like Eq. (10-48). That is, the presence or absence of the lens does not change the fact that the 180° phase shift in the central thin film causes the energy to be shifted towards the edges of the diffraction pattern.

10-8c The Imaging Properties of a Lens—The Field Approach

The aim of this section is to derive the mathematical expression which describes how a lens forms an image given the amplitude distribution, $A(x, y)$, of an object. In other words, we will calculate the field behind a lens given the field of an object. This will involve some tedious mathematical manipulations. Unfortunately, they are required for a rigorous development of the theory and a sound understanding of how to apply the theory to real-world situations.

The use of a thin film is not the only way in which an optical system can produce a phase difference; the varying thickness of the lens from axis to margin also causes such a shift. To compute this quantity, consider the asymmetric lens of Fig. 10-21, which shows a ray of light traveling a slant distance T_1' from surface 1 to surface 2, with an inclination angle α_1'. If α_1' is small, then

$$T_1' = t_1' - \overline{V_1A_1} - \overline{A_2V_2}$$

or

$$\begin{aligned} T_1' &= t_1' - \{r_1 - \sqrt{r_1^2 - (x_1^2 + y_1^2)}\} - \{-r_2 - \sqrt{r_2^2 - (x_2^2 + y_2^2)}\} \\ &= t_1' - r_1\left\{1 - \sqrt{1 - \frac{x_1^2 + y_1^2}{r_1^2}}\right\} + r_2\left\{1 - \sqrt{1 - \frac{x_2^2 + y_2^2}{r_2^2}}\right\} \end{aligned} \quad (10\text{-}49)$$

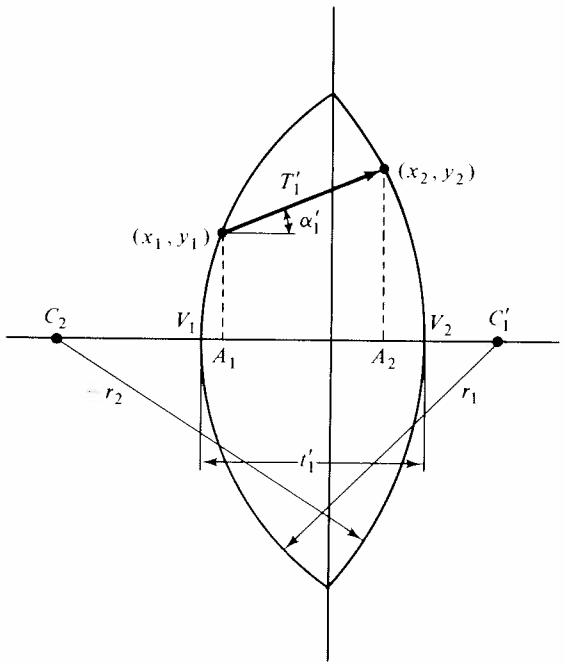


Figure 10-21

Note that the above approximation is not as severe as the paraxial one, where we assumed that $T_1' \sim t_1'$. We may also use the relations

$$\begin{aligned} x_1 &\sim x_2 = x \\ y_1 &\sim y_2 = y \end{aligned} \tag{10-50}$$

and a binomial approximation to obtain

$$\begin{aligned} T_1' &= t_1' - r_1\left\{\frac{x_1^2 + y_1^2}{2r_1^2}\right\} + r_2\left\{\frac{x_2^2 + y_2^2}{2r_2^2}\right\} \\ &= t_1' - \left(\frac{1}{r_1} - \frac{1}{r_2}\right)\left(\frac{x^2 + y^2}{2}\right) \end{aligned} \tag{10-51}$$

Recall from Chapter 1 that the focal length of a thin lens is given by the relation

$$\frac{1}{f'} = (n_1' - 1)\left[\frac{1}{r_1} - \frac{1}{r_2}\right] \tag{1-63}$$

Hence, Eq. (10-51) becomes

$$T_1' = t_1' - \frac{(x^2 + y^2)}{2f'(n_1' - 1)} \tag{10-52}$$

To see how this determines the effect of the lens on the wave passing through it, consider a plane wave striking vertex V_1 (Fig. 10-22). There is a variable phase delay as the wave travels through the lens; the greater the thickness of glass it transverses, the larger the delay. The phase delay inside the lens therefore depends on the index of refraction and on T_1', whereas outside it depends only on the distance $(t_1' - T_1')$ between the wave front entering at V_1 or emerging at V_2 and the glass surfaces. Hence, the plane wave

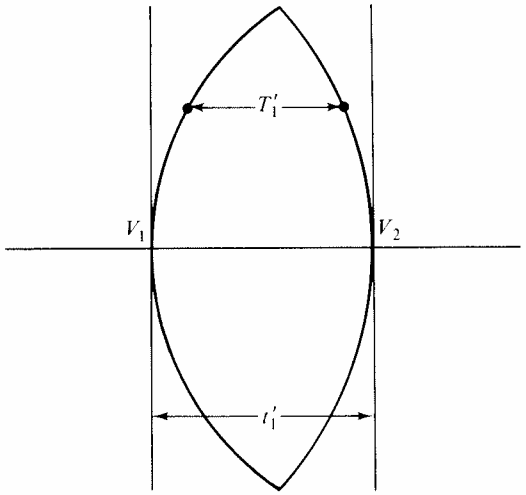


Figure 10-22

has the form $\exp\{ik[n_1'T_1' + (t_1' - T_1')]\}$, and by Eq. (10-52)

$$\begin{aligned} \exp[ik\{n_1'T_1' + (t_1' - T_1')\}] &= \exp[ik\{(n_1' - 1)T_1' + t_1'\}] \\ &= e^{in_1'kt_1'}e^{-ik(x^2 + y^2)/2f'} \end{aligned} \tag{10-53}$$

This result consists of two factors: a constant term which we may drop, since we are concerned with relative rather than absolute intensities, and another term involving $x^2 + y^2$. The amplitude of the wave as it travels through an optical system will be proportional to $\exp(ikr)$ and the variable term in (10-53) will modify this to $\exp[i(kr + \delta)]$, where δ is the phase shift introduced by the lens. Hence, the amplitude of the wave emerging at V_2 is related to that of the incident wave at V_1 by the expression

$$A(x_2, y_2) = A(x_1, y_1)e^{-ik(x_1^2 + y_1^2)/2f'} \tag{10-54}$$

If an apodizing or phase shifting device is present in the entrance pupil of the lens, then a pupil function $P(x_1, y_1)$ as introduced in Eq. (10-44) should be included above, but for simplicity, we shall consider an unmodified lens.

We are now ready to describe the behavior of a wave as it passes from object to lens and through the lens to the image plane, using a method due to Smith[10-9]. The reader is warned at this point that the algebra which follows is formidable and there appears to be no simpler way of obtaining the extremely important conclusions we shall reach. It is convenient to work in the reverse direction; that is, we shall first regard the wave front which passes through the exit pupil of the lens as a Huyghens source whose amplitude is $A(x_2, y_2)$ and from it, we compute the amplitude $A(x', y')$ at the image plane. The diffraction integral (10-1) then becomes

$$A(x', y') = C\int A(x_2, y_2)e^{ikr}\,dS_2 \tag{10-55}$$

The slant distance T_2' between the point (x_2, y_2) and the point (x', y'), for which the spacing in the z direction is t_2', is

$$\begin{aligned} T_2' &= \sqrt{(x' - x_2)^2 + (y' - y_2^2) + t_2'^2} \\ &= t_2'\sqrt{1 + \left(\frac{x' - x_2}{t_2'}\right)^2 + \left(\frac{y' - y_2}{t_2'}\right)^2} \end{aligned}$$

and a binomial expansion gives

$$T_2' = t_2'\left[1 + \frac{1}{2}\left(\frac{x' - x_2}{t_2'}\right)^2 + \frac{1}{2}\left(\frac{y' - y_2}{t_2'}\right)^2\right] \tag{10-56}$$

Since $T_2' = r$ for this situation, Eq. (10-55) becomes

$$A(x', y') = C\int A(x_2, y_2)e^{ik[(x' - x_2)^2 + (y' - y_2)^2]/2t_2'}\,dS_2 \tag{10-57}$$

where the term $e^{ikt_2'}$ has been incorporated into the constant C and the notation dS_2 means that we are integrating over the exit pupil.

Next, we incorporate the phase delay in passing through the lens, as

expressed by Eq. (10-54). Equation (10-57) is then

$$A(x', y') = C \int e^{-ik(x_1^2 + y_1^2)/2f'} A(x_1, y_1) e^{ik[(x'-x_2)^2 + (y'-y_2)^2]/2t_2'} \, dS_2 \tag{10-58}$$

Finally, we use a relation like Eq. (10-57) to connect the amplitude $A(x, y)$ at the object with the amplitude $A(x_1, y_1)$ at the entrance pupil, obtaining the expression

$$A(x', y') = C \iint A(x, y) e^{-ik(x_1^2 + y_1^2)/2f'} \times e^{ik[(x_1-x)^2 + (y_1-y)^2]/2t_1} e^{ik[(x'-x_2)^2 + (y'-y_2)^2]/2t_2'} \, dS_2 \, dS \tag{10-59}$$

where dS refers to an integration in the plane of the object.

For a reasonably thin lens, we can assume that x_1 is approximately equal to x_2 and similarly for y_1 and y_2; thus we express the coordinates at the lens as

$$x_1 = x_2 = x_L$$
$$y_1 = y_2 = y_L$$

Considering just the x terms in Eq. (10-59), we then have

$$e^{-ikx_L^2/2f'} e^{ik(x_L-x)^2/2t_1} e^{ik(x'-x_L)^2/2t_2'} = e^{i(kx_L^2/2)[(1/t_1) + (1/t_2') - (1/f')]} e^{-ikx_L[(x/t_1) + (x'/t_2')]} e^{ikx^2/2t_2'} e^{ikx^2/2t_1} \tag{10-60}$$

This is integrated with respect to the variable dx_L, where

$$dS_2 = dS_L = dx_L \, dy_L$$

and the limits can be taken as $-\infty, \infty$, since there is no contribution to the integral beyond the edges of the lens. Using the standard definite integrals[10-10]

$$\int_0^\infty \begin{Bmatrix} \sin \\ \cos \end{Bmatrix} (ax^2 + 2bx) \, dx = \frac{1}{2}\left(\cos \frac{b^2}{a} \mp \sin \frac{b^2}{a}\right)\sqrt{\frac{\pi}{2a}}$$

we obtain

$$\int_{-\infty}^{\infty} e^{i(ax^2 + 2bx)} \, dx = (1 + i)\sqrt{\frac{\pi}{2a}} e^{-ib^2/a} \tag{10-61}$$

and this in turn gives

$$\begin{aligned} & e^{ikx'^2/2t_2'} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{i(kx_L^2/2)[(1/t_1) + (1/t_2') - (1/f')] - ikx_L[(x/t_1) + (x'/t_2')]} \, dx_L \right] e^{ikx^2/2t_1} \, dx \\ &= e^{ikx'^2/2t_2'} \int_{-\infty}^{\infty} \left[(1 + i)\sqrt{\frac{\pi}{ik[(1/t_1) + (1/t_2') - (1/f')]}} \right. \\ &\quad \left. \times e^{-(ik/2)[(x/t_1) + (x'/t_2')]^2 [(1/t_1) + (1/t_2') - (1/f')]} e^{ikx^2/2t_1} \right] dx \\ &= Ce^{(ikx'^2/2t_2')[1 - [t_1f'/(t_2f' + t_1f' - t_1t_2')]]} \\ &\quad \times \int_{-\infty}^{\infty} e^{(ikx^2/2t_1)[1 - [t_2'f'/(t_2'f' + t_1f' - t_1t_2')]] - ik[xx'f'/(t_2'f' + t_1f' - t_1t_2')]} \, dx \end{aligned} \tag{10-62}$$

where C contains all the constants.

We see that the quadratic exponential terms in front of the integral will cancel if

$$t_1f' = t_2'f' + t_1f' - t_1t_2'$$

or

$$f' = t_1 \tag{10-63}$$

Similarly, the quadratic terms inside the integral cancel if

$$t_2'f' = t_2'f' + t_1f' - t_1t_2'$$

or

$$f' = t_2' \tag{10-64}$$

Under these conditions, the integral represented by Eq. (10-62), with the corresponding integral over the variables in the y direction, will cause Eq. (10-59) to reduce to

$$A(x', y') = C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) e^{-ik(xx' + yy')/f'} \, dx \, dy \tag{10-65}$$

where C again incorporates all the constant factors outside the integral. By Eq. (10-45), this may be written as

$$A(v_x, v_y) = C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) e^{-ik(v_x x + v_y y)/f'} \, dx \, dy \tag{10-66}$$

Thus, if C is taken as unity and the lens does not have an apodization or absorption screen, the field at the focal plane F is converted into its Fourier transform in the focal plane F'. The relation between the fields at other locations is more complicated. As Eq. (10-62) indicates, there are in general quadratic terms present. Physically these lead to Fresnel-type diffraction and much more complicated transformations.

10-8d Summary of Fourier Optics

To summarize, we have demonstrated the following optical methods of producing two-dimensional Fourier transforms:

(a) The Fraunhofer diffraction pattern at a screen at infinity is the Fourier transform of the aperture function of an opening, as shown by Eq. (10-40).

(b) The Fraunhofer diffraction pattern at the focal distance f' from a lens is the Fourier transform of its pupil function as shown by Eq. (10-46).

(c) An object at f is converted into its Fourier transform at f' by a lens as shown by Eqs. (10-63), (10-64), and (10-66).

10-9 Spatial Filtering

10-9a Basic Principles, Periodic Objects

The previous section showed how and why an aperture or a simple lens can compute a Fourier transform, and the similarity with the corresponding electronic problem was discussed. It was also suggested that it should be possible to remove components of the Fourier spectrum and alter the resulting pattern. This process is called *spatial filtering*, and the basic idea goes back to the experiments of Abbe[10-11] and Porter.[10-12]

To understand its significance, let us use the simple approach of Brown.[10-13] It is shown in Problem 10-6 that the diffraction pattern due to a pair of slits of width w and spacing d is

$$\frac{I}{I_0} = \text{sinc}^2 \alpha \cos^2 \gamma \tag{10-21}$$

where

$$\alpha = \tfrac{1}{2} kw \sin\theta, \qquad \gamma = \tfrac{1}{2} kd \sin\theta$$

The function $\text{sinc}^2 \alpha$, which defines the envelope of the curve in Fig. 10-10, is known as the *shape factor* and is due to the individual slits, while the fine structure depends on the spacing of the slits. If we now generalize to a one-dimensional array of N slits of width w, spaced at a distance d, the corresponding diffraction integral is (Fig. 10-23)

$$\begin{aligned}\int_{-\infty}^{\infty} e^{ikx\sin\theta}\,dx &= \int_0^w e^{ikx\sin\theta}\,dx + \int_d^{w+d} e^{ikx\sin\theta}\,dx + \cdots \\ &\quad + \int_{(N-1)d}^{(N-1)d+w} e^{ikx\sin\theta}\,dx \\ &= \left(\frac{e^{ikw\sin\theta}-1}{ik\sin\theta}\right)[1 + e^{ikd\sin\theta} + \cdots + e^{ik(N-1)d\sin\theta}]\end{aligned}$$

Using the formula for the sum of a geometric series with a ratio r and having n terms gives

$$S = \frac{a(1-r^n)}{1-r} = \frac{1-e^{iNkd\sin\theta}}{1-e^{ikd\sin\theta}}$$

for the second term on the right. Using α and γ as in Eq. (10-21), the relative amplitude of the diffraction pattern becomes

$$\frac{A}{A_0} = \left(\frac{e^{2i\alpha}-1}{2i\alpha/w}\right)\left(\frac{1-e^{2iN\gamma}}{1-e^{2i\gamma}}\right)$$

But

$$\begin{aligned}e^{2i\alpha} - 1 &= (e^{i\alpha} - e^{-i\alpha})e^{i\alpha} \\ &= 2e^{i\alpha} i \sin\alpha\end{aligned}$$

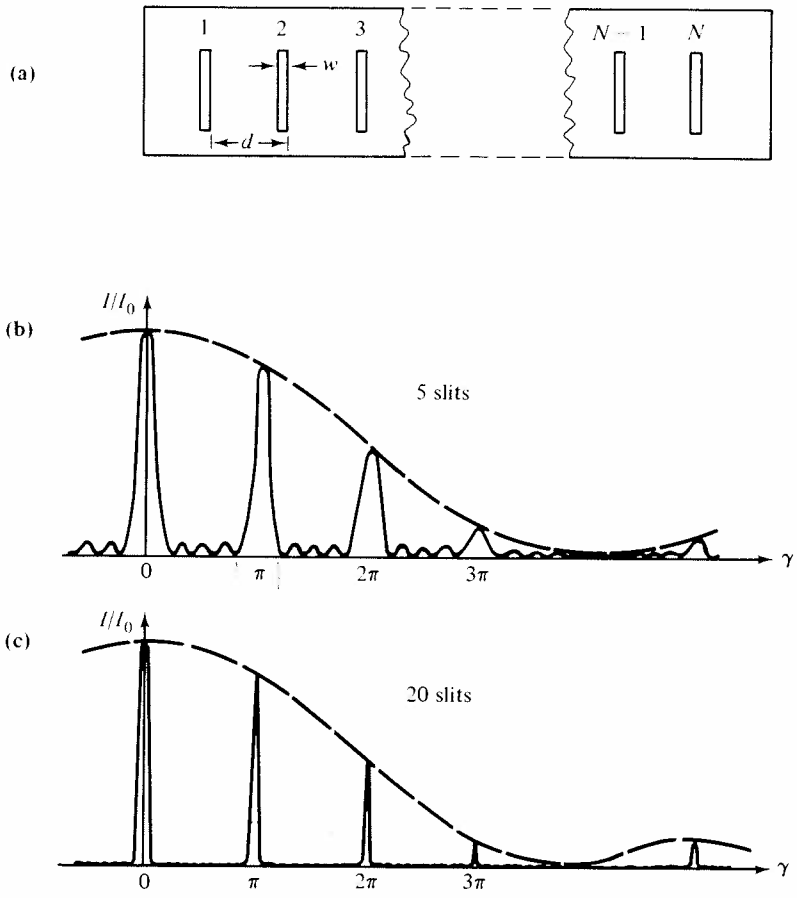


Figure 10-23

with similar expressions for the exponentials involving γ; therefore we obtain

$$\begin{aligned}\frac{A}{A_0} &= \frac{e^{i\alpha}\sin\alpha}{\alpha/w}\,\frac{e^{iN\gamma}\sin N\gamma}{e^{i\gamma}\sin\gamma} \\ &= 2we^{i\alpha}e^{i(N-1)\gamma}\,\frac{\sin\alpha}{\alpha}\,\frac{\sin N\gamma}{\sin\gamma}\end{aligned}$$

To find the intensity, we multiply A by its complex conjugate and absorb all the constants into I_0. In addition, we explicitly incorporate a factor of N^2 into the expression for I/I_0 so that $I \longrightarrow I_0$ when $\theta \longrightarrow 0$. The final result is then

$$\frac{I}{I_0} = \text{sinc}^2\alpha\left(\frac{\sin N\gamma}{N\sin\gamma}\right)^2 \tag{10-67}$$

which has the shape factor of Eq. (10-21), as we might expect.

The other term, the *grating factor* $[(\sin N\gamma)/(N \sin \gamma)]^2$, is a generalization of $\cos^2 \gamma$ in Eq. (10-21), for when $N = 2$, we have

$$\frac{\sin 2\gamma}{2 \sin \gamma} = \cos \gamma$$

which is a well-known identity. Plots of Eq. (10-67) for $N = 5$ and $N = 20$ are shown in Figs. 10-23(b) and (c). We see that the central lobe of sinc α, as shown in Fig. 10-3, is converted into a number of sharp maxima, so that the diffraction pattern consists of a large number of very narrow, closely spaced lines. These principal maxima occur for values of γ which make the denominator $\sin \gamma$ of the grating factor vanish. That is

$$\gamma = m\pi, \qquad m = 0, \pm 1, \pm 2, \ldots \tag{10-68}$$

The maxima are called spectra of order m as used in connection with Eq. (6-5); for example, the central peak is the zero-order spectrum. Equation (10-68) is equivalent to

$$\tfrac{1}{2}kd \sin \theta_m = m\pi$$

where the θ_m are the angles with respect to the normal, and this relation, in turn, becomes

$$d \sin \theta_m = m\lambda \tag{10-69}$$

which we recognize as being similar to the Bragg law as used by crystallographers. This should not be surprising, since a single crystal is simply a diffraction grating for X rays and for electrons. The intensity, as given by Eq. (10-67), and the reinforcement condition as in Eq. (10-69) are approximations, valid only for small angles, since we have neglected the obliquity factor which should appear in the diffraction integral.

We have previously shown that an object at the focal point F of a thin lens results in a Fraunhofer pattern (i.e., a Fourier transform) at a distance f' in image space. In the case of a grating used as the object (Fig. 10-24), the diffraction integral Eq. (10-46) enables us to find how f' determines the size and spacing of the bright bands in the diffraction pattern.

Problem 10-11

(a) Show that the half-width x' of the central bright band in the plane through F' is

$$x' = \frac{f'\lambda}{Nd} \tag{10-70}$$

(b) Show that the maximum corresponding to $\gamma = \pi$ is separated from the central maximum by a distance h_1, where

$$h_1 = \frac{f'\lambda}{d} \tag{10-71}$$

Let a plane wave strike a grating normally, as shown in Fig. 10-24(a). By Eq. (10-69), the zero-order maximum corresponds to rays for which $\theta_0 = 0$, i.e., they are parallel to the optical axis. Such rays should intersect at the focal point F' for either a thin lens or one that is free of aberrations,

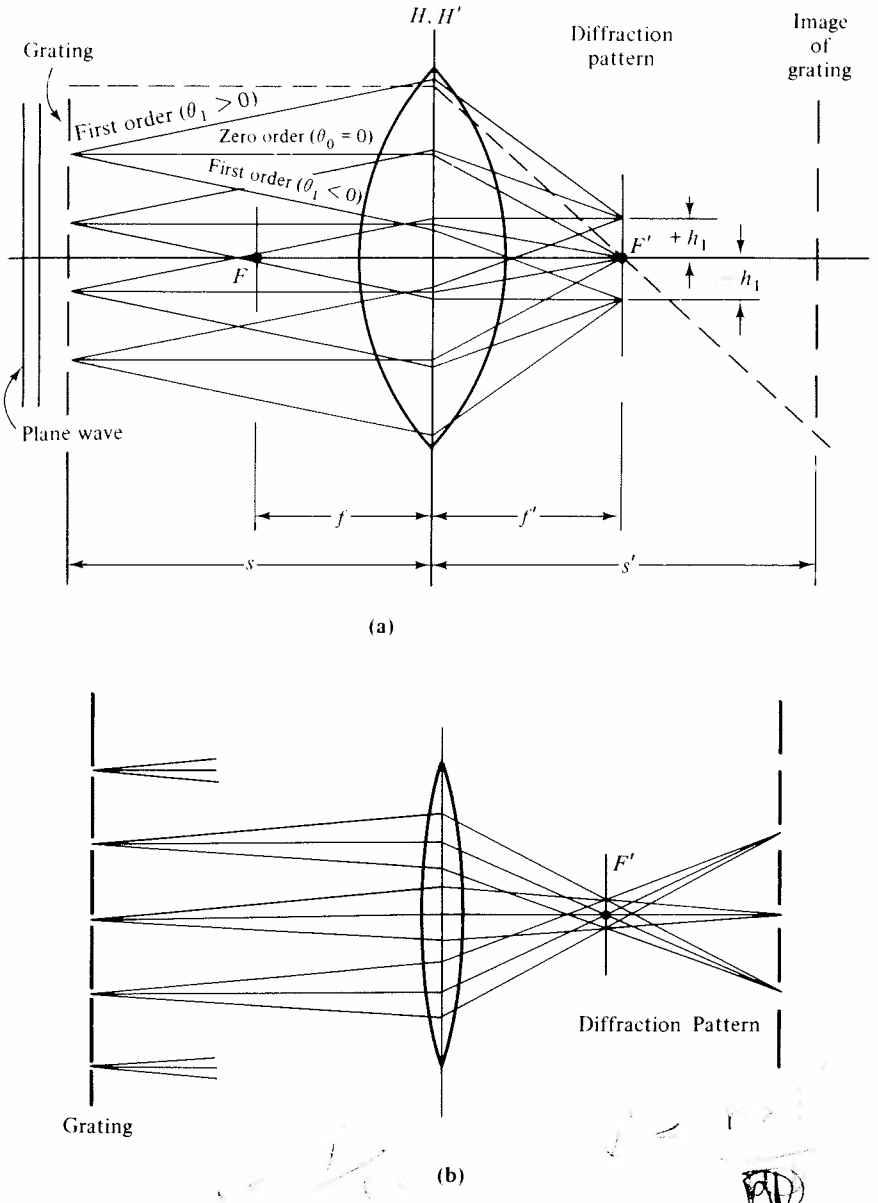


Figure 10-24

as shown. Next, we turn our attention to the rays corresponding to $m = 1$ and traveling with a positive slope $+\theta_1$. Although we demonstrated in the preceding section that the object and the Fraunhofer pattern are located at conjugate foci, the problem just below shows that the position of the grating in object space is arbitrary. That is, *for this special choice of an object*, the fact that the diffracted orders produce sets of parallel rays means that the Fourier transform appears in the focal plane specified by f' regardless of where the object is located. The path of the rays beyond F' was omitted for clarity in Fig. 10-24(a); in Fig. 10-24(b), we see how the three orders from a given object point recombine at the corresponding image point.

Problem 10-12

Use the matrix method to show that a set of parallel rays with slope $+\theta_1$ which strike a thin lens meet in the focal plane F' at a distance $\theta_1 f' = +h_1$ from the axis. ■

To return to spatial filtering, suppose we place an opaque screen with a narrow slit in it at F', choosing the width of the opening to be slightly larger than $2h_1$. Then the two orders $m = 0$ and $m = \pm 1$ shown will pass through the opening and all higher orders—which strike the plate at a distance greater than $\pm h_1$ from the axis—will be blocked. This has a profound effect on the image at s'. Figure 10-25(a) shows the image intensity as we normally expect to observe it; the bright regions have a width $w' = \beta w$, where β is the magnification, and a spacing $d' = \beta d$. The effect of the aperture is shown in Fig. 10-25(b). The rounded corners of the pattern indicate that the higher diffraction orders have been removed; this is a well-known phenomenon in Fourier analysis, as indicated in Fig. A-2 (Appendix) for example. These higher orders correspond to the sharp corners in a step function. To obtain a perfectly square corner, an infinite number of terms in the Fourier series must be included. For the higher values of γ, Eq. (10-45) shows that k is proportional to v_x, and γ is proportional to k by definition. We see therefore that the slit is removing the high spatial frequency components. Exactly the same thing can happen in an electric circuit; the effect of a low-pass filter on a square wave is to round the corners. On the other hand, if we replace the slit with an opaque strip of the same size and permit only the high spatial frequency components or high diffraction orders to reach the image plane, then we have an optical high-pass filter and obtain the image shown in Fig. 10-25(c). Now the edges are sharp and the central sections show the effect of removing the low-frequency components.

10-9b A Numerical Example

We can appreciate in a simple, quantitative way the effect of spatial filtering on an image by using an example of Gerrard.[10-14] Returning to the single

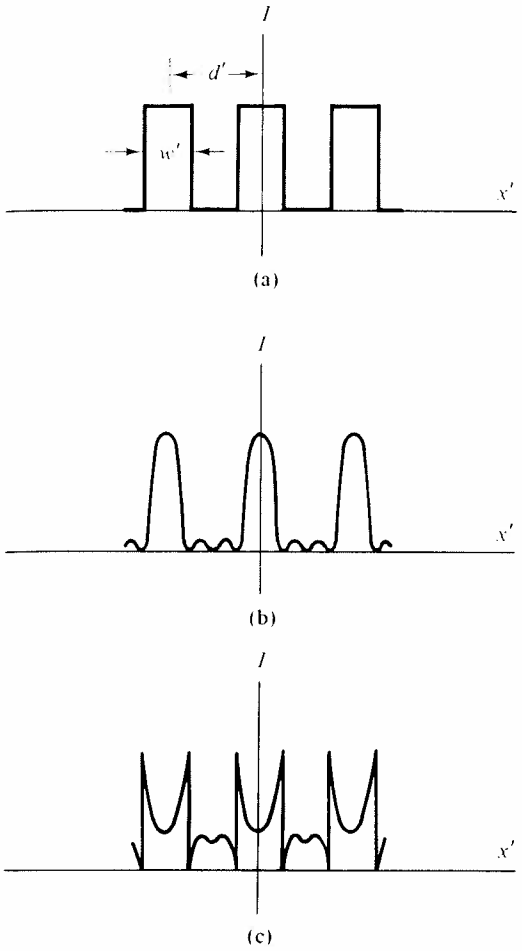


Figure 10-25

slit of width w, we recall that the relative amplitude of the diffraction pattern was given as

$$A = C' \operatorname{sinc} \alpha \tag{10-5}$$

We shall combine this expression for the amplitude of a single slit, the modifications introduced by repeating the slit in a periodic fashion, and the arrangement of the orders indicated in Fig. 10-24 to compute the intensity of the pattern corresponding to the image of a diffraction grating. Since each value of m corresponds to a set of parallel rays striking the lens, we can say that a plane wave striking the grating normally is decomposed into an infinite set of plane waves leaving it, each with a particular amplitude A_m, phase φ_m, and angle of inclination θ_m, where $m = 0, \pm 1, \ldots$. Each A_m will be determined by Eq. (10-5), θ_m by Eq. (10-69), and φ_m must be calculated from the geometry. Before doing so, let us find the total amplitude due to all the individual

contributions through the use of what electrical engineers call a *phasor diagram*. Although these are commonly used to represent complex numbers, for our purposes here we need simply take A_m for $m = 0$ as the phase reference axis, as shown in Fig. 10-26. The terms for $m = \pm 1, \pm 2, \ldots$ add vectorially to this, with the horizontal components giving a contribution $A_m \cos \varphi_m$

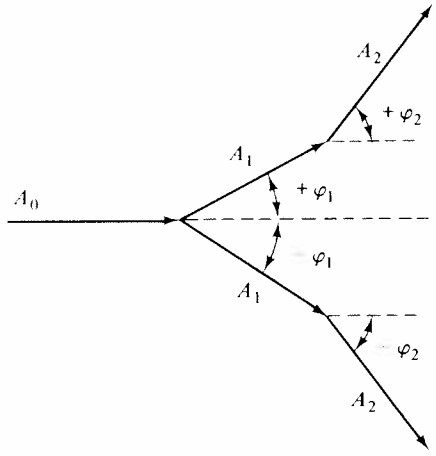


Figure 10-26

and the vertical components cancelling in pairs. The total amplitude at each image point of Fig. 10-24b is then

$$A = \sum_{m=-\infty}^{\infty} A_m \cos \varphi_m \tag{10-72}$$

where $\cos \varphi_0 = 1$ since $\varphi_0 = 0$. Using Eq. (10-5) for the amplitudes gives

$$\frac{A}{C'} = \sum_{m=-\infty}^{\infty} \operatorname{sinc} \alpha_m \cos \varphi_m$$

where

$$\alpha_m = \tfrac{1}{2} kw \sin \theta_m$$

or

$$\frac{A}{C'} = 1 + 2 \sum_{m=1}^{\infty} \operatorname{sinc} \alpha_m \cos \varphi_m \tag{10-73}$$

since $\theta_0 = 0$, $\operatorname{sinc} 0 = 1$, $\varphi_0 = 0$, and $\cos 0 = 1$.

It is convenient to use Eq. (10-69) in paraxial form, so that

$$\theta_m = \frac{m\lambda}{d} \tag{10-74}$$

Then

$$\alpha_m = \frac{\pi}{\lambda} w\theta_m = \frac{mw\pi}{d} \tag{10-75}$$

Turning to φ_m, this introduces a factor $\exp(ik\delta_m)$ in the analytical expression for a plane wave. If x is the distance of one of the slits in the grating from the

axis, then $x \sin \theta_m \sim x\theta_m$ is the difference in path length or phase between a ray of order m leaving this slit and one at the center. This is analogous to the situation illustrated in Fig. 10-2. Hence

$$\varphi_m = k\delta_m = kx\theta_m = \frac{mk\lambda x}{d} = \frac{2\pi mx}{d}$$

Since we wish to find the intensity in the image plane, we use the Gaussian lens equation in the form

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

This corresponds to a magnification

$$\frac{s'}{s} = \beta = \frac{x'}{x} = \frac{(s' - f')}{f'}$$

and

$$\varphi_m = \frac{2\pi mx'f'}{(s' - f')d} \tag{10-76}$$

Inserting Eq. (10-75) and Eq. (10-76) into Eq. (10-73) and squaring gives the image intensity as

$$\frac{I}{I_0} = \left[1 + 2 \sum_{m=1}^{M} \operatorname{sinc} \frac{mw\pi}{d} \cos \frac{2\pi mx'f'}{(s' - f')d}\right]^2 \tag{10-77}$$

where the upper limit M designates the number of terms which we use in the summation.

As a specific example, let the grating have 5000 lines per cm with opaque regions 50% wider than the slits. Then $d = 0.0002$ cm and $w = 0.8\ \mu$m. If the grating is 1.2 cm from a lens for which $f' = 1.0$ cm, then $s' = 6.0$ cm, $\beta = s'/s = 5.0$, and the slit images have a width of 4.0 μm. By Eq. (10-76)

$$\varphi_m = \frac{2\pi mx'}{(5.0)(0.0002)} = 6283.1mx'$$

and from Eq. (10-75)

$$\alpha_m = 1.256m$$

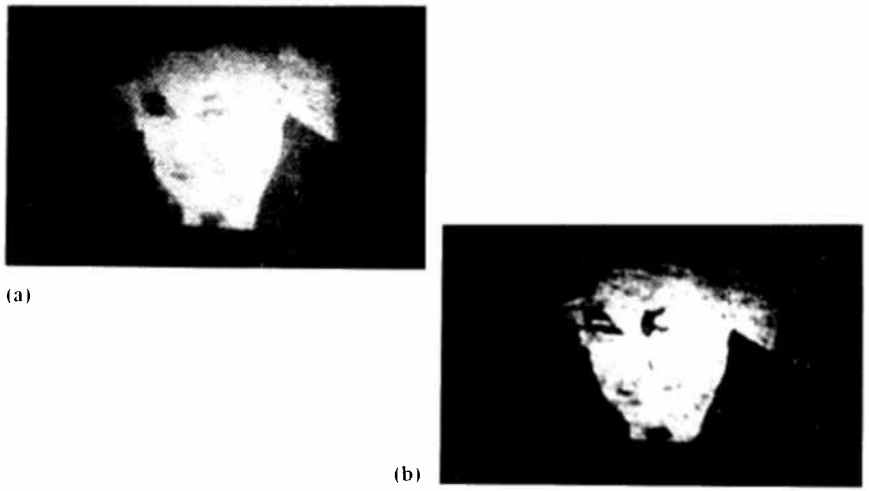
The series to be evaluated is thus

$$\frac{I}{I_0} = \left[1 + 2 \sum_{m=1}^{M} \frac{\sin (1.256m)}{1.256m} \cos (6283.1mx')\right]^2 \tag{10-78}$$

Problem 10-13

(a) Calculate I/I_0 in Eq. (10-78) for $M = 1, 2, 4, 6$, and 11, verifying Figs. 10-27(a)–(e).

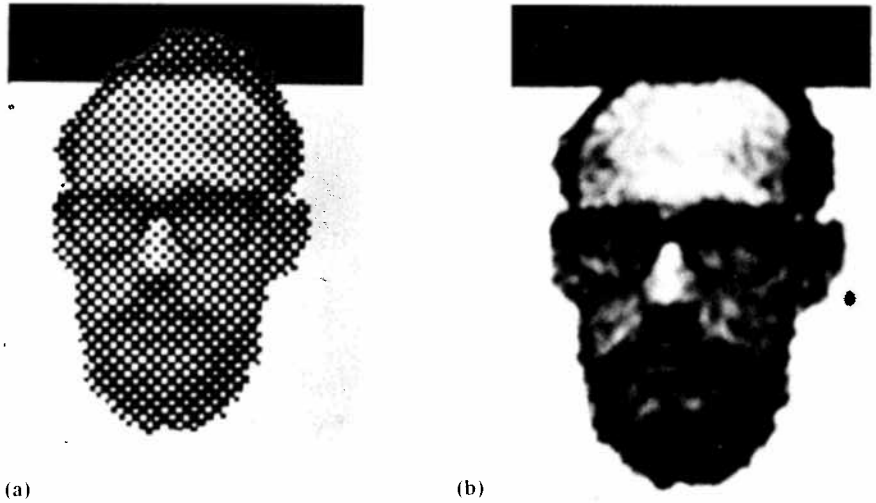
(b) Apply Eq. (10-78) to the case of a high-pass spatial filter, obtaining a few curves such as in Fig. 10-25(c). ■



(a) (b)

Figure 10-29

10-30(a) shows a self-portrait by Minnesota artist Karl E. Bethke, who painted regularly spaced acrylic squares onto canvas. The appearance of shading was obtained by varying the size of the squares, in a manner similar to the halftone process. It is possible to make a filter for the squares by using a pattern of squares having a uniform size and the same spacing as in the portrait. Then a negative photograph could be taken of its transform; this would be the desired filter. When properly positioned in the transform plane, the filter



(a) (b)

Figure 10-30

will block out only those components which make up the pattern of squares, passing both the higher and lower spatial frequency components of the original picture. Unfortunately, such a filter is difficult to construct and align. Therefore a pinhole, as before, is used as the filter, and the result is shown in Fig. 10-30(b). The image does suffer some degradation because high frequencies which make up small details are lost.

A practical application of these ideas is due to Watkins.[10-18] Integrated circuits consisting of a number of diodes, transistors, resistors, and other components are fabricated by photographic masking processes. A master mask is drawn by a draftsman, reduced in several steps to a size of about 1 mm × 1 mm, and then used to prepare a multiple mask by a step-and-repeat process. The final array consists of several hundred identical microcircuit masks, but a few of them may not be perfect copies due to inhomogeneities in the masking materials or failure to achieve exact reproduction, as seen in Fig. 10-31(a). Locating the defective masks by visual observation is an extremely difficult job, since each section has to be compared with the master under a microscope. Watkins used the fact that the diffraction pattern for an array of these masks is similar to that of an $M \times N$ array of rectangular openings so that it consists of two sets of sharp lines intersecting at right angles; that is, it will be an $M \times N$ array of tiny points [Fig. 10-31 (b)].

A filter consisting of an array of spots is placed at the focal plane F' of the lens performing the transform. For a perfect series of masks, the entire pattern is filtered out, and the inverse transform at the image plane is a photograph of very low intensity. However, a nonperiodic error such as a scratch or a contamination gives extra lines in the diffraction pattern which are not removed at the filter plane; this shows up in Fig. 10-31(c), which clearly reveals the defects in Fig. 10-31(a). Hence, a defective mask can be identified very quickly. To locate a particular bad element in an array, the original $M \times N$ matrix can be examined half at a time in the same way, and so on.

Interesting and unexpected effects are observed with multiple apertures. For simplicity, consider an $N \times N$ rectangular array of identical openings. The diffraction integral for any given opening is

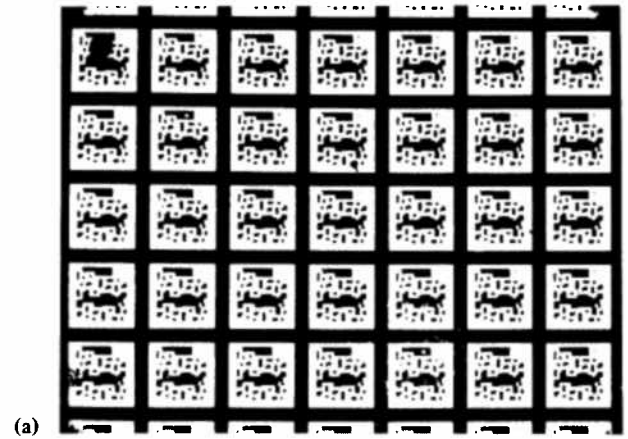
$$A(x', y') = \int A(x, y) e^{ikr}\, dxdy$$

If we transform the coordinates in the plane of the array by the relations

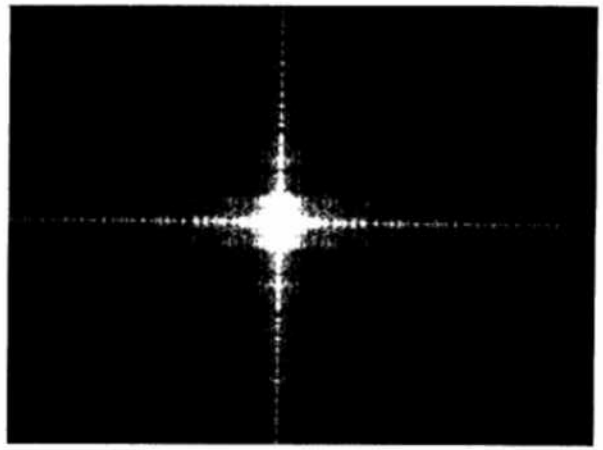
$$x_1 = x + x_0$$
$$y_1 = y + y_0$$

and also

$$r_1 = r + r_0$$



(a)



(b)

(c)

Figure 10-31

the integral becomes

$$A(x', y') = \int A(x_1 - x_0, y_1 - y_0)e^{ik(r_1 - r_0)}\,dx_1 dy_1$$

The phase factor $\exp(-ikr_0) = \exp(i\delta)$ can be removed from the integral, since it does not depend on x_1 and y_1. The function $A(x_1 - x_0, y_1 - y_0)$ is simply the original amplitude, but centered about a new origin. Hence, the diffraction pattern for each individual aperture is unchanged except for a displacement. When we sum over a row of N apertures, we then obtain

$$A(x', y') = \left(\sum_{j=1}^{N} e^{i\delta_j}\right) \int A(x_1 - x_0, y_1 - y_0)e^{ikr}\,dx_1 dy_1$$

If the apertures are randomly spaced, then the phase factors sum to a value of $\sqrt{N}$, as shown in Section 5-6. On the other hand, a regularly spaced array will give values ranging from N to zero, depending on the observation point (x', y').

We thus see that a regular array of N apertures has an intensity which is N^2 times that of a single aperture. Intuitively, we can visualize each opening adding a contribution of the proper phase to all the other openings. Equation (10-67) shows this to be true for the diffraction grating. In deriving this equation, we arbitrarily divided the intensity by a factor of N^2 in order to compare the pattern of the single slit with that of the grating.

On the other hand, the random array increases the intensity only by a factor of N, since there will be out-of-phase components as well. The same results apply to $N \times N$ arrays, the phase factors combining to give an N^2 enhancement for a perfectly periodic structure or a factor of N in the random case. For a periodic arrangement, but with a few shape defects, the result is an intensity increase by approximately N^2 and some extra peaks in the diffraction pattern. These effects are shown in Fig. 10-32, taken from Hoover.[10-19] Figure 10-32a shows the Fraunhofer pattern of 3000 randomly oriented rectangular openings and Fig. 10-32b is a 54×54 regular array of the same kind of opening. Note the clear, sharp spikes in the latter because of the enhanced intensity.

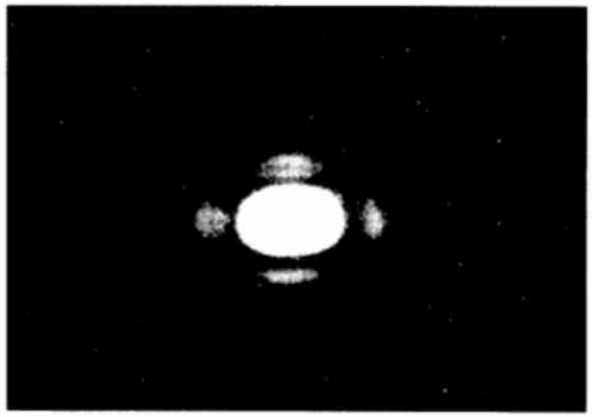
10-9d The Abbe Theory of the Microscope

As another application, we return to the work of Abbe,[10-11] whose motivation was actually the theory of the resolving power of the microscope. Following Stone,[10-20] the relation between transform plane amplitude $A_{F'}(x_{F'})$ and object plane amplitude $A(x)$, from Eq. (10-65), is

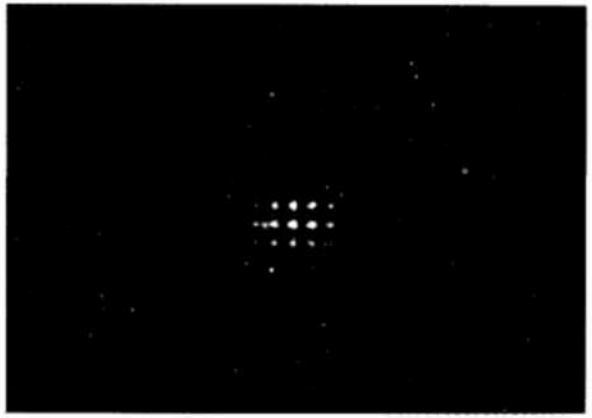
$$A_{F'}(x_{F'}) = \int_{-\infty}^{\infty} A(x)e^{-ikxx_{F'}/f'}\,dx \tag{10-79}$$

and this Fourier transform has an inverse of the form

$$A(x) = \int_{-\infty}^{\infty} A_{F'}(x_{F'})e^{ikxx_{F'}/f'}\,dx_{F'} \tag{10-80}$$



(a)



(b)

Figure 10-32

However, if the image is a faithful reproduction of the object with magnification β, then

$$A'(x') = A\left(\frac{x'}{\beta}\right) \tag{10-81}$$

where A' is the amplitude at the image plane. Eq. (10-80) combined with Eq. (10-81) gives

$$A'(x') = \int_{-\infty}^{\infty} A_{F'}(x_{F'}) e^{ikx'x_{F'}/f'\beta}\, dx_{F'} \tag{10-82}$$

The amplitude in the transform plane for a single slit of width w will be

$$\begin{aligned} A_{F'}(x_{F'}) &= C \operatorname{sinc} \alpha \\ &= w \frac{\sin (kx_{F'}w/2f')}{(kx_{F'}w/2f')} \end{aligned} \tag{10-83}$$

as obtained by integrating Eq. (10-79) and ignoring the constants. Using this in Eq. (10-82) gives

$$A'(x') = w \int_{-\infty}^{\infty} \frac{\sin (kw/2f')x_{F'}}{(kw/2f')x_{F'}} e^{i(kx' f'\beta)x_{F'}}\, dx_{F'} \tag{10-84}$$

Consider the identity

$$\int_{-\infty}^{\infty} \frac{\sin px}{x} e^{iqx}\, dx = \int_{-\infty}^{\infty} \frac{\sin px \cos qx}{x}\, dx + i \int_{-\infty}^{\infty} \frac{\sin px \sin qx}{x}\, dx$$

The second integral, involving an odd function, vanishes; the first integral—from tables[10-10]—becomes

$$2 \int_{0}^{\infty} \frac{\sin px \cos qx}{x}\, dx = \begin{cases} \pi & \text{for} \quad 0 < p < q \\ 0 & \text{for} \quad 0 < q < p \end{cases}$$

Hence

$$\begin{aligned} A'(x') &= w \int_{-\infty}^{\infty} \frac{\sin (kw/2f')x_{F'}}{(kw/2f')x_{F'}} e^{i(kx' f'\beta)x_{F'}}\, dx_{F'} \\ &= \left.\begin{matrix} \dfrac{\pi f'}{k} \\ 0 \end{matrix}\right\} \quad \text{if} \quad \begin{cases} \dfrac{\beta w}{2} < x' \\ \dfrac{\beta w}{2} > x' \end{cases} \end{aligned} \tag{10-85}$$

Equation (10-85) simply specifies the magnified image of the positive half of the grating. Using a similar calculation for the other half, we obtain the sequence illustrated in Fig. 10-33, which shows the original slit, the Fourier transform, and the inverse transform of sinc α.

The calculation just given implies that the width w of the slit is fairly large. The function sinc α has its first minimum when

$$\frac{kx_{F'0}w}{2f'} = \pi$$

or

$$x_{F'0} = \frac{f'\lambda}{w}$$

where $x_{F'0}$ is the position of this minimum. A wide slit means that the entire function sinc α will essentially pass through the lens and reach the focal plane, but if w is very small, then only the central section can get through. That is, the lens edge acts as a low-pass spatial filter, and all detail in the image is lost. The transform at F' is simply a band of reasonably uniform intensity, and the inverse transform of this constant, just as in Eq. (10-83), is

$$A'(x') = R \operatorname{sinc} \alpha'$$

at the image plane, where

$$\alpha' = \frac{kRx'}{2\beta f'}$$

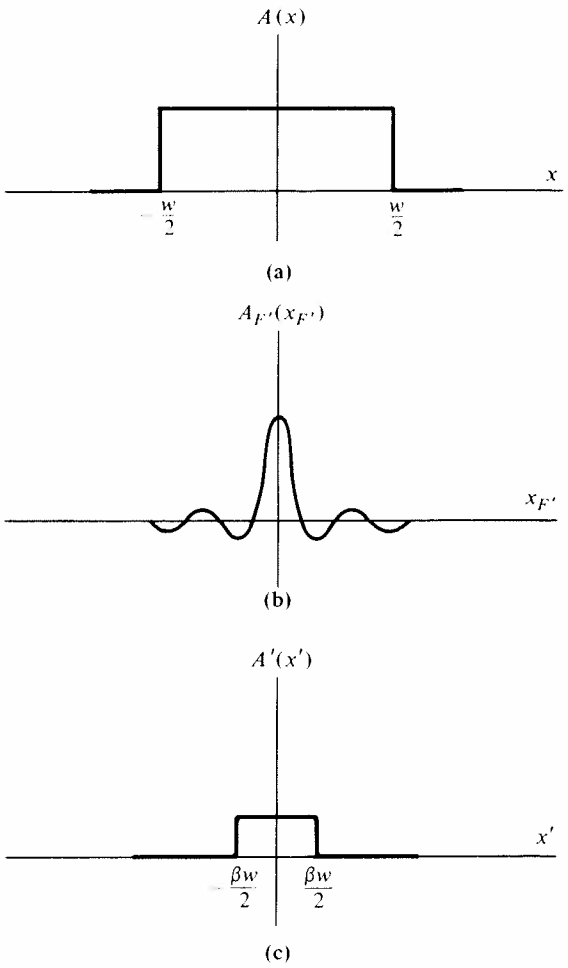


Figure 10-33

and R is the lens radius. The first minimum x'_0 of this pattern is then

$$x'_0 = \frac{2f'\beta\lambda}{R}$$

Hence, the final image is not a faithful reproduction, but a bright band whose size depends on the properties of the microscope objective being used.

Returning to the complete grating, the situation is very similar to what we have already discussed. Figure 10-26(a) represents the image of just one slit when the zero-order and the two first-order spectra are passed; the complete image is a repetition of this single one. We realize that the reproduction is poor, but it is distinguishable as a slit. If the spectra for $m = 1$ were also

filtered out, Eq. (10-78) would reduce to $I = I_0$, and the image would again be a uniformly illuminated screen with no detail. Hence, as Abbe pointed out, for a microscope to resolve fine detail, the objective must be capable of passing at least the first two orders, and the more orders, the better the resolution, as Fig. 10-26 indicates. The spectra for $m = 1$ are spaced a distance $2f'\lambda/d$, and this represents the minimum spacing which the objective must be capable of resolving. If we examine the diffraction pattern with a microscope, then the objective serves as the spatial filter at the plane F' in Fig. 10-24, and its diameter D can be no smaller than $2f'\lambda/d$. Let θ be the half-angle subtended by the objective at the plane H, H'. Then

$$\sin\theta = \frac{R}{f'} = \frac{\lambda}{d} \tag{10-86a}$$

High-power microscope objectives are usually of the oil-immersion type, and the theory of Chapter 1 shows that when the objective is immersed in a liquid of index n, Eq. (10-86) takes the form

$$n\sin\theta = \frac{\lambda}{d}$$

or

$$d = \frac{\lambda}{N.A.} \tag{10-86b}$$

where $n\sin\theta$ is called the *numerical aperture* ($N.A.$) of the objective. Hence, *the minimum spacing d which can be resolved by a good microscope is on the order of the wavelength of light.* If an oil of high index and a large acceptance angle is used, the maximum practical value of $n\sin\theta$ is about 1.6.

It is interesting to note that Abbe's theory, which leads to Eq. (10-86), is based on the idea that at least one order must be passed by the lens. Earlier in Section 10-4, we calculated the diffraction pattern of a circular aperture and used the Rayleigh criteria to determine resolutions, Eq. (10-16). These two theories are based on quite different views of image formation, yet lead to the same result.

Problem 10-14

Use the facts that the smallest angular separation for which the human eye can resolve two points is about 1 minute of arc and the least distance of distinct vision is 25 cm to show that the theoretical maximum magnification for a microscope is on the order of 10^3. ■

10-9e Contrast Improvement

We shall give a precise definition of the concept of contrast later in this chapter [see Section 10-11] but for the time being, let us simply regard it as a mea-

sure of the difference between the bright and dim portions of an illuminated object. For example, alternating black and white bands may be represented by the repeated square wave shown in Fig. 10-34(a), whereas two shades of gray would be represented by a pattern like Fig. 10-34(b). The maximum intensity is less than that in the top portion and the contrast is lower because the change in intensity is smaller.

We realize that Fig. 10-34(b) is like the square wave of Fig. A-3 of the Appendix, but shifted upward by an amount I_C. This is equivalent to changing the constant (or DC) term in Eq. (A-8) from 0.5 to a larger value. If we can eliminate this shift, then the square wave would move down so that its minimum rests on the x-axis, greatly enhancing the contrast. We can even reverse the contrast if the maximum is brought down to the axis and the minimum lies below it. These changes may be accomplished by altering the constant term in the series of Eq. 10-73; we either filter it out or pass it through a film producing a 180° phase reversal, which will change the sign. Fig. 10-24a shows that a filter which will remove the plane wave for which $m = 0$, this being the constant term in the summation, is a disc of radius less than h_1. This term would have been larger if we had started with an object of lower contrast; the grating which we used is assumed to give complete transmission through the slits and none at the opaque regions.

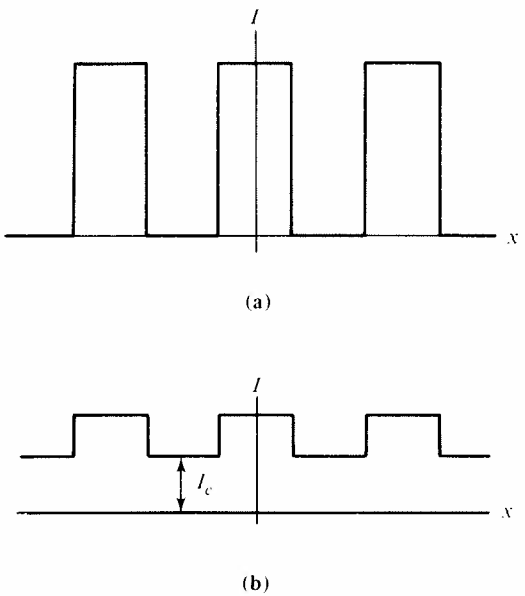


Figure 10-34

10-9f Phase Contrast

The aperture function of the grating we considered at the beginning of this discussion affects the amplitude of the transmitted wave. We now wish to consider a grating which modifies the phase. For example, if we have a periodic array composed of transparent apertures alternating with thin films producing a 90° phase shift, the aperture function is (for the initial section)

$$A(x) = 1 \quad \text{for} \quad 0 \le x \le \frac{d}{4}, \frac{3d}{4} \le x \le d$$

$$A(x) = i \quad \text{for} \quad \frac{d}{4} \le x \le \frac{3d}{4}$$

This grating then has a period of d and each section has a width equal to $d/2$. The Fourier series corresponding to such a function is easily seen to have the form

$$\frac{A}{A_0} = \frac{1}{2}(1 + i) + \frac{2}{\pi}(1 + i)\sum_m a_m \cos\frac{2\pi m x}{d} \qquad \text{(10-86c)}$$

for the real part is like the square wave of Eq. (A-8) and the imaginary part will have the value of zero when the real part equals unity and the value i when the real part vanishes. The term for $m = 0$ is $1 + i$, which is a complex number making an angle of 45° to the real axis. The rest of the series in Eq. (10-86c) is multiplied by $(1 - i)$, representing an angle of −45°, and the two terms are 90° out of phase. Two complex numbers at right angles have no components in each other's direction and hence there can be no interference between the zero order term and any of the others.

Let us next consider the effect of a 270° phase shift on the term $1 + i$. This is accomplished with a thin film at the center of the transform plane and multiplies $(1 + i)$ by $-i$, converting Eq. (10-86) to

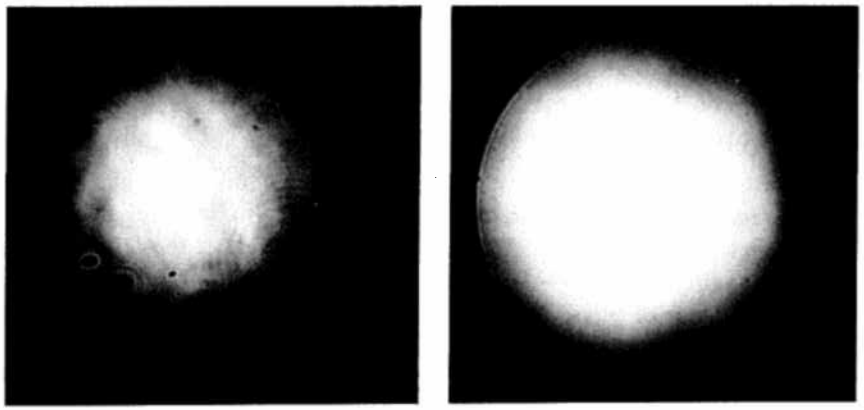
$$\frac{A}{A_0} = (1 - i)\left[\frac{1}{2} + \frac{2}{\pi}\sum_m a_m \cos\frac{2\pi m x}{d}\right] \qquad \text{(10-86d)}$$

Since the term $(1 - i)$ has a magnitude of $\sqrt{2}$, we have essentially the same result as we obtained for the original grating; i.e., the use of a 270° phase shifting filter has converted phase variations into amplitude variations. This is the basis of the *phase contrast microscope*, invented by Frits Zernike in 1935.

Small transparent organisms are difficult to see with an ordinary microscope. It is possible, however, to take advantage of the fact that their index of refraction generally differs from that of the surrounding medium and this results in a phase difference in the transmitted light. A phase shifting film will then produce an amplitude variation which is readily visible.

10-9g Processing the Output of a Laser

Consider the expanded output beam from a typical helium-neon laser (Fig. 10-35). The pattern is irregular because dust particles and imperfections on optical surfaces in the beam produce diffraction patterns. Since the diffraction patterns are objectionable when the beam is used for holography or in photographic work, we would like to eliminate them. Think of the beam as consisting of two parts—the gaussian component (low spatial frequency) and diffraction pattern (higher spatial frequencies). When the beam is transformed by a lens, the gaussian component is located near the center of the transform while the high frequency terms are located away from the center, so that a pinhole placed at the center of the transform plane will transmit only the low frequency component. A second lens of longer focal length then collimates the beam, which is both expanded and smoothed by the assembly (Fig.



(a) (b)

Figure 10-35 (a) Expanded output beam of laser before filtering, and (b) expanded beam after filtering.

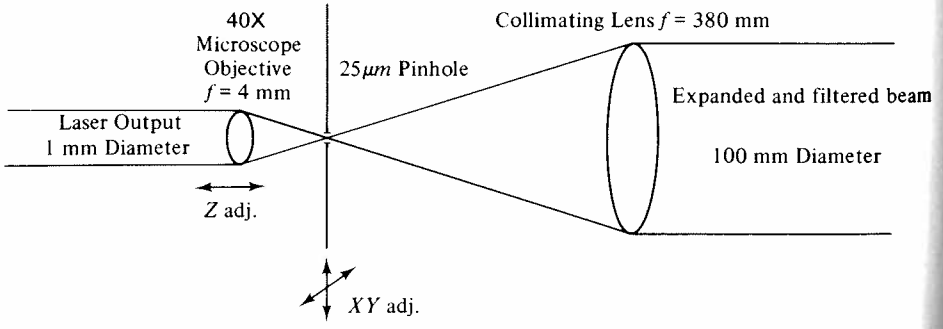


Figure 10-36

10-36). Thus, the principle of spatial filtering is used to obtain a uniform beam which then may be used for a variety of applications in the laboratory.

10-10 The Optical Transfer Function

When we considered the Fourier transform properties of lenses in Section 10-8, we first assumed that the only effect of the lens on a wave passing through it is to introduce a phase shift which depends on the separation of the two glass surfaces. Later we found that in the process of forming an image, a lens cuts out high spatial frequency components of an object. These high spatial frequency components account for fine detail. This was the basic idea in Abbe's theory of the microscope. We might expect a real lens not only to cut out certain spatial frequencies because of diffraction, but also to attenuate others because of aberrations. This suggests that the image-forming quality of a lens could be specified by a function which describes how much each spatial frequency component in an object is attenuated by the lens as it forms an image. As we shall see in this section, there is such a function, called the *optical transfer function*, which specifies lens quality. This approach is in wide use today.

10-10a Derivation of the Optical Amplitude Transfer Function

To derive the optical transfer function, we return to Eq. (10-59) and for simplicity assume:

$$x_1 = x_2 = x_L, \qquad y_1 = y_2 = y_L$$

We include the pupil function $P(x_L, y_L)$ as we did in Eq. (10-44) to explicitly include the effects of the lens on amplitude; in particular, the effects of its finite size on the spatial frequency properties of the transmitted wave. Our integrals will extend from $-$ to $+$. This will facilitate Fourier-transform operations. Equation (10-59) thus becomes

$$A(x', y') = C \iint A(x, y) P(x_L, y_L) e^{-ik(x_L^2 + y_L^2)/2f'} e^{ik[(x_L - x)^2 + (y_L - y)^2]/2t_1} \cdot e^{ik[(x' - x_L)^2 + (y' - y_L)^2]/2t'_2} \, dS_L \, dS \tag{10-87a}$$

As Eq. (10-60) indicates, the quadratic term involving x_L^2 will reduce to unity when we impose the condition that

$$\frac{1}{t_1} + \frac{1}{t'_2} = \frac{1}{f'} \tag{1-6}$$

which we recognize as the Gauss lens equation. For a spherical object producing a spherical image, the quadratic terms in x^2 and x'^2 are constant. For a plane object which is not too large, this condition is approximately true and

absorbing these terms into C, Eq. (10-87a) simplifies to

$$A(x', y') = C \iint A(x, y)P(x_L, y_L)e^{-ikx_L[(x,t_1)+(x',t_2')]} \cdot e^{-iky_L[(y,t_1)+(y',t_2')]}\, dS_L\, dS \tag{10-87b}$$

Using the expression

$$\beta = -\frac{t_2'}{t_1} \tag{1-75}$$

for the magnification (the negative sign follows the conventions of Chapter 1) converts this to

$$A(x', y') = C \iint A(x, y)P(x_L, y_L)e^{-i(k/t_2')[x_L(x'-\beta x)+y_L(y'-\beta y)]}\, dS_L\, dS \tag{10-88}$$

As Goodman[10-21] points out, this relation indicates that the image for $P = 1$ is the transform of the object, or the Fraunhofer pattern of the object as limited by the lens aperture, but centered on image coordinates $x' = \beta x$, $y' = \beta y$. This is what we would expect intuitively; the image has its size determined by paraxial optics, and it lies on the plane towards which the spherical waves originating at (x, y) are being made to converge by the lens. Further, for $P = 1$ over a finite distance the image represents the Airy disc plus the secondary rings of the diffraction pattern obtained by integrating over the lens opening for each point on the object and then integrating over the object to find the total contribution to each image point.

Next, we change variables as specified by

$$\bar{x} = \beta x, \qquad \bar{y} = \beta y \tag{10-89}$$

and define spatial frequencies, following Eq. (10-45), as

$$\left.\begin{aligned} \nu_x &= -\frac{kx_L}{t_2'} \\ \nu_y &= -\frac{ky_L}{t_2'} \end{aligned}\right\} \tag{10-90}$$

where the negative sign is used for convenience. Equation (10-88) becomes

$$A(x', y') = C \int\int\int\int A\left(\frac{\bar{x}}{\beta}, \frac{\bar{y}}{\beta}\right)P\left(-\frac{\nu_x t_2'}{k}, -\frac{\nu_y t_2'}{k}\right) \cdot e^{i[\nu_x(x'-\bar{x})+\nu_y(y'-\bar{y})]}\, d\bar{x}\, d\bar{y}\, d\nu_x\, d\nu_y \tag{10-91}$$

where all constants continue to be absorbed in C. The integration over the lens opening can be expressed in terms of the *point spread function h* defined by

$$h(x', y') = \iint P\left(-\frac{\nu_x t_2'}{k} - \frac{\nu_y t_2'}{k}\right)e^{i(\nu_x x' + \nu_y y')}\, d\nu_x\, d\nu_y \tag{10-92}$$

which we recognize as the Fourier transform of the pupil function. If we use

Eq. (10-92), Eq. (10-91) becomes

$$A(x', y') = C \int A\left(\frac{\bar{x}}{\beta}, \frac{\bar{y}}{\beta}\right)h(x' - \bar{x}, y' - \bar{y})\, d\bar{S} \tag{10-93}$$

which, by Eq. (A-38a), is the convolution of h and the object. Thus, image formation—which we have already shown to be a Fourier transform computation—may also be interpreted as the convolution of the object with the transmission properties of the lens. Lipson and Lipson,[10-4] for example, show the convolution of a two-dimensional object (a photograph) with an aperture such as a slit to obtain a somewhat blurry picture. If just the object and the aperture were used, the result would be a diffraction pattern. To obtain the convolution, they use the aperture to mask the lens of an out-of-focus camera. This essentially performs the integration over the lens opening and gives a convolution rather than a transform.

10-10b Analogy with Electrical Circuits

The function h, which is defined by Eq. (10-92) and used in Eq. (10-93), has an important physical interpretation. As shown in the Appendix, the usual equation governing the behavior of a circuit containing a resistance R and an inductance L in series, by Eq. (A-40), is

$$v(t) = \left(L\frac{d}{dt} + R\right)i(t) \tag{10-94}$$

and by taking the Fourier transform of both sides of this equation, it may be written as

$$V(\omega) = (i\omega L + R)I(\omega) \tag{10-95}$$

or

$$I(\omega) = V(\omega)Y(\omega)$$

where

$$Y(\omega) = \frac{1}{i\omega L + R}$$

In the time domain, $v(t)$ is the excitation and $i(t)$ the response, whereas in the frequency domain, $Y(\omega)$ is known as either the complex admittance or the transfer function. In the special situation where $v(t)$ is a delta function, Eq. (A-59) gives $\mathfrak{F}[\delta(t)] = 1 = V(\omega)$ and $I(\omega)$ is identical to $Y(\omega)$. That is, the transfer function—which we originally identified in terms of the admittance (or impedance) of the circuit—is also seen to be the response to unit voltage or excitation. Furthermore, if $g(t) = v(t) = \delta(t)$ in Eq. (A-26), then Eq. (A-68) with $a = 0$ gives

$$f(t)*g(t) = f(t)*\delta(t) = f(t) = h(t)$$

This shows why $h(t)$, whose Fourier transform is

$$H(\omega) = Y(\omega) = I(\omega) \tag{10-96}$$

is called the *impulse response*.

The analogous situation in optics is the response to a point source of light. By Eq. (10-92), $h(x', y')$ represents the diffraction pattern at the image plane generated by a point source at (x, y), which is why it is called the point spread function. We further see from the Appendix that the solution of Eq. (10-94) is

$$i(t) = \int_{-\infty}^{\infty} g(\alpha)h(t - \alpha)\, d\alpha = v(t)*y(t)$$

[see Eq. (A-38a)] which is analogous to Eq. (10-93); therefore $h(x', y')$—a function of the image coordinates—is the optical equivalent of the time domain admittance $y(t)$, i.e., it is the optical impulse response. Actually, we compute $Y(\omega)$ when we do circuit problems; thus in optics we should correspondingly work with $H(v_x, v_y)$, the transform of $h(x', y')$. Note that in circuit theory we go from the time domain to the frequency domain, whereas in optics we go from two-dimensional space to the spatial frequency domain.

10-10c The Optical Intensity Transfer Function

The function $H(v_x, v_y)$, which is the transform of $h(x', y')$, should thus be called the *optical transfer function* (or optical admittance). But h is the transform of P, so that combining Eq. (10-93) with the definition of H gives

$$H(v_x, v_y) = \mathfrak{F}[h(x', y')] = \mathfrak{F}\left[\mathfrak{F}\left[P\left(-\frac{v_x t_2'}{k}, -\frac{v_y t_2'}{k}\right)\right]\right] \tag{10-97}$$

and since the transform of a Fourier transform is the original function, this reduces to

$$H(v_x, v_y) = P\left(-\frac{v_x t_2'}{k}, -\frac{v_y t_2'}{k}\right) \tag{10-98}$$

This transfer function $H(v_x, v_y)$ specifies the effect of a lens on the amplitude of spatial frequencies passing through, just as $Y(\omega)$ performs a similar function for the angular frequencies in a signal. However, we are usually interested in intensity rather than amplitude, and we define the *optical intensity transfer function* $F(v_x, v_y)$ by the Fourier transform relation

$$F(v_x, v_y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x', y')h^*(x', y')e^{-i(v_x x + v_y y)}\, dx'\, dy'}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x', y')h^*(x', y')\, dx'\, dy'} \tag{10-99}$$

where the denominator is included for normalization purposes. This quantity, rather than H, is what is known as the optical transfer function (*OTF*). Equation (10-99) may be abbreviated as

$$F(v_x, v_y) = \frac{\mathfrak{F}[|h|^2]}{\mathfrak{F}[|h|^2]|_{v_x, v_y = 0}} \tag{10-100}$$

To use this result, we recast it by noting (as explained in the Appendix) that if three functions are related as

$$F(\omega) = G(\omega)H(\omega)$$

then their inverse transforms are connected by the convolution equation

$$f(t) = g(t) * h(t) \tag{A-38b}$$

as shown in the Appendix. Letting

$$g(t) = h^*(t)$$

gives

$$\mathfrak{F}[h * h^*] = HH^*$$

or in two-dimensional form

$$\mathfrak{F}\left[\int\int h(\xi, \eta)h^*(\xi - v_x, \eta - v_y)\, d\xi\, d\eta\right] = H(v_x, v_y)H^*(v_x, v_y) \tag{10-101}$$

where ξ and η are dummy variables. Equation (10-101) has an inverse

$$\mathfrak{F}[h(\xi, \eta)h^*(\xi, \eta)] = \int\int H(\xi, \eta)H^*(\xi + v_x, \eta + v_y)\, d\xi\, d\eta \tag{10-102}$$

as may be verified by writing out the integrals. Combining this with Eq. (10-100) converts the definition of the *OTF* into

$$F(v_x, v_y) = \frac{\int\int H(\xi, \eta)H^*(\xi + v_x, \eta + v_y)\, d\xi\, d\eta}{\int\int |H(\xi, \eta)|^2\, d\xi\, d\eta} \tag{10-103}$$

Letting

$$\xi' = \xi + \frac{v_x}{2}, \qquad \eta' = \eta + \frac{v_y}{2} \tag{10-104}$$

gives

$$F(v_x, v_y) = \frac{\int\int H(\xi' - v_x/2, \eta' - v_y/2)H^*(\xi + v_x/2, \eta' + v_y/2)\, d\xi'\, d\eta'}{\int\int |H(\xi', \eta')|^2\, d\xi'\, d\eta'} \tag{10-105}$$

Then we drop the primes on ξ' and η' and use Eq. (10-98) to write this as

$$F(v_x, v_y) = \int\int P\left(\xi + \frac{t_2' v_x}{2k}, \eta + \frac{t_2' v_y}{2k}\right)P^*\left(\xi - \frac{t_2' v_x}{2k}, \eta - \frac{t_2' v_y}{2k}\right) d\xi\, d\eta \tag{10-106}$$

The denominator has also been dropped, since we can do the normalization with respect to some arbitrary value of F. This extremely important result

indicates that, for an aberration-free lens, the *OTF* is simply the area of overlap of two displaced pupil functions, one centered about $(t_2'v_x/2k, t_2'v_y/2k)$ and the other about the diametrically opposite point $(-t_2'v_x/2k, -t_2'v_y/2k)$.

Example: A Circular Aperture

The most obvious, although not the simplest, example to which Eq. (10-106) should be applied is the circular pupil. For convenience, we will take the centers of the two displaced circles as shown in Fig. 10-37(a). There is no loss in generality, since the overlap depends only on the spacing between centers. The overlap area (shown shaded) is four times the segment labeled S

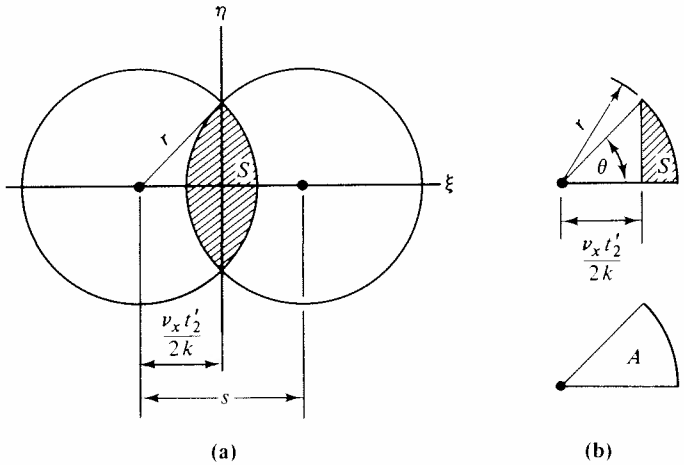


Figure 10-37

in Fig. 10-37(a). To find S, we compute the area A for the sector of angle θ and radius r, which is

$$A = \left(\frac{\theta}{2\pi}\right)(\pi r^2) = \frac{\theta r^2}{2}$$

where

$$\theta = \arccos\left(\frac{s}{2r}\right)$$

and where

$$s = \frac{v_x t_2'}{k}$$

is the spacing between centers. Then, as shown by Fig. 10-37(b),

$$S = A - \frac{1}{2}\left(\frac{s}{2}\right)\sqrt{r^2 - \left(\frac{s}{2}\right)^2}$$

By Eq. (10-106)

$$F(v_x, 0) = \frac{4S}{\pi r^2}$$

$$= \frac{2}{\pi}\left[\arccos\left(\frac{s}{2r}\right) - \left(\frac{s}{2r}\right)\sqrt{1 - \left(\frac{s}{2r}\right)^2}\right] \qquad (10\text{-}107)$$

for

$$0 \le \frac{s}{2r} \le 1$$

Furthermore, $I(v_x, 0) = 0$ when $s \ge 2r$, since this means there is no overlap.

The value v_{xo} of v_x when the circles just touch is called the *cutoff frequency*. It represents the maximum spatial frequency passed by the lens and is given by

$$s = 2r$$

or

$$v_{xo} = \frac{2rk}{t_2'} = \frac{4\pi r}{\lambda t_2'}$$

Plotting Eq. (10-107), we obtain Fig. 10-38(a) (lower curve), noting that $F(0, 0) = 1$. The two-dimensional function, $F(v_x, v_y)$, is shown in Fig. 10-38(b).

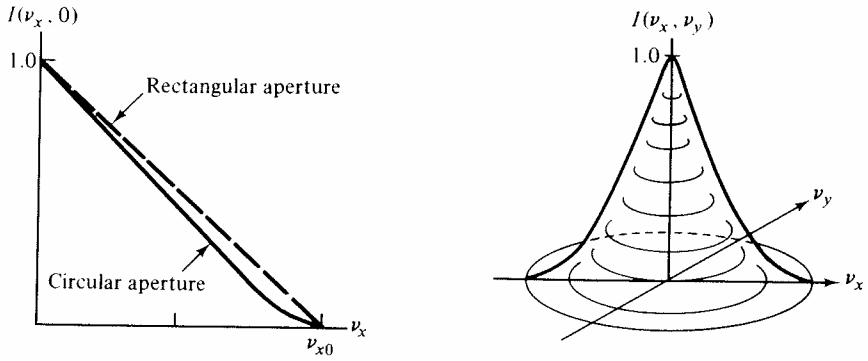


Figure 10-38

Problem 10-15

Find the transfer function for a square aperture of side $d = 2r$, where r is the radius of the circular aperture of Fig. 10-35(a). Verify the upper curve in this figure and the fact that v_{xo} is the same in both cases. ■

10-10d Resolving Power

In order to consider the significance of the *OTF* to the lens designer and user, let us give some further thought to the concept of *resolution* or *resolving*

power $\mathcal{R}$ discussed in a qualitative way in connection with Fig. 10-8. For simplicity, we shall consider a source of light at infinity passing through a narrow slit of length l and width w. If a lens is adjacent to the slit, the diffraction pattern will be focused on a screen at a distance F'. Suppose the object (at infinity) consists of regularly spaced black and white stripes (Fig. 10-39),

Figure 10-39

and the spacing is specified in lines/mm. This quantity is the spatial analogue of the frequency of a wave (in hertz or cycles/second) and hence is the spatial frequency v_x associated with the periodic object. We therefore take it as a measure of the resolving power, since the closer the spacing, the greater the demand on the optical system. Problem 10-15 for a rectangular opening of width w gives a cutoff frequency, for a distant object, of magnitude

$$v_{xo} = \frac{kw}{f'}$$

Using a lens whose diameter d exactly matches the slit width, we then have a maximum resolution given by

$$\mathcal{R} = \frac{k}{f'/w} \tag{10-108}$$

where f'/w is the f-number of Problem 10-3. A perfect $f/2.8$ camera lens in green light (555.0 nm) has a resolving power of about 4000 lines/mm.

Returning to the interpretation of the *OTF*, we realize that Fig. 10-38(a) shows the resolving power of two kinds of apertures used in conjunction with perfect lenses. In fact, Eq. (10-108) indicates that this quantity is directly proportional to the edge of the square (or the diameter of the circle). The graph also shows how the intensity of the diffracted image decreases with closer spacing of the black bars in Fig. 10-39. We recall, however, that we have defined two types of optical transfer function; the amplitude function H is specified by Eq. (10-97) and the intensity function F is given in Eq. (10-99). Another way of distinguishing these two quantities is in terms of coherence. Although we considered this subject at several places in Part II, let us take a very intuitive approach here and stipulate that a coherent plane wave is one for which the phase throughout the wavefront is a known function of position. We can find the intensity received at a screen by combining

amplitudes and phases in the manner of Eq. (5-17) and squaring the result. On the other hand, if the wave is incoherent and the phases are random, then we must work directly with the intensities. This means that the transfer function H can be applied to a completely coherent wave and only to such a wave. If we have an aberration-free lens, the pupil function $P(x_L, y_L)$ has a value of unity when r_L is less than or equal to the lens radius R and it vanishes outside the lens. This condition can be expressed in terms of the *circle function*, defined as

$$\text{circ}\,(\sqrt{x_2 + y_2}) = \begin{cases} 1 \text{ when } \sqrt{x^2 + y^2} \leq 1 \\ 0 \text{ otherwise} \end{cases} \tag{10-109}$$

so that the pupil function becomes

$$P(x_L, y_L) = \text{circ}\left(\frac{\sqrt{x_L^2 + y_L^2}}{R}\right) \tag{10-110}$$

A lens conforming to Eq. (10-110) is said to be *diffraction limited*; its ability to produce a perfect image is restricted only by the diffraction introduced at the edge, which converts each mathematical point in the object to an Airy disc of finite diameter.

The explicit expression for P given by Eq. (10-110) enables us to evaluate the coherent transfer function H through the use of Eq. (10-97), obtaining

$$H(v_x, v_y) = \text{circ}\left(\frac{\sqrt{x_L^2 + y_L^2}}{Rk/t_2'}\right) \tag{10-111}$$

The definition of the circle function means that the cutoff frequency will be

$$v_{co} = \frac{Rk}{t_2'} = \frac{2\pi R}{\lambda t_2'} \tag{10-112}$$

We can compare this result with the Airy disc radius r', as obtained from Eq. (10-16). Since θ is approximately given by r'/t_2', we have that

$$r' = 0.72\frac{t_2'\lambda}{R} \tag{10-113}$$

Let us define a cutoff radius or distance r_{co} as the reciprocal of the cutoff frequency (in analogy with the relation between frequency and period for time-varying functions) so that Eq. (10-112) gives

$$r_{co} = \frac{t_2'\lambda}{2\pi R} = 0.16\frac{t_2'\lambda}{R} \tag{10-114}$$

We thus see that there is fairly good agreement between the size of the Airy disc, as determined by elementary diffraction theory and the resolution limit of the lens specified by the frequency for which the OTF vanishes. This is not surprising, because the former is what determines the Rayleigh criterion for the minimum possible resolution.

Returning to the situation for incoherent light, let r of Eq. (10-107) have a maximum value equal to the lens radius R, so that the incoherent cutoff frequency is

$$\nu_{co} = \frac{4\pi R}{\lambda t_2'} \tag{10-115}$$

which we recognize as twice the value given in Eq. (10-112). We have therefore shown that a lens passing coherent light will have a constant OTF right up to the maximum spatial frequency it can pass and beyond this value, the transfer function vanishes. For incoherent light, on the other hand, the OTF falls monotonically to a cutoff frequency which is half that for coherent light.

The next thing we must worry about is the way in which lens aberrations modify these ideal results. This requires that we have a way of expressing the effect of aberrations on a light wave, and this is the subject of the following section.

10-11 Wave Theory of Aberrations

Consider a group of parallel rays leaving an object and passing through a perfectly corrected lens. These rays will converge to a common image point P' (Fig. 10-40), and they may be regarded as originating from a spherical

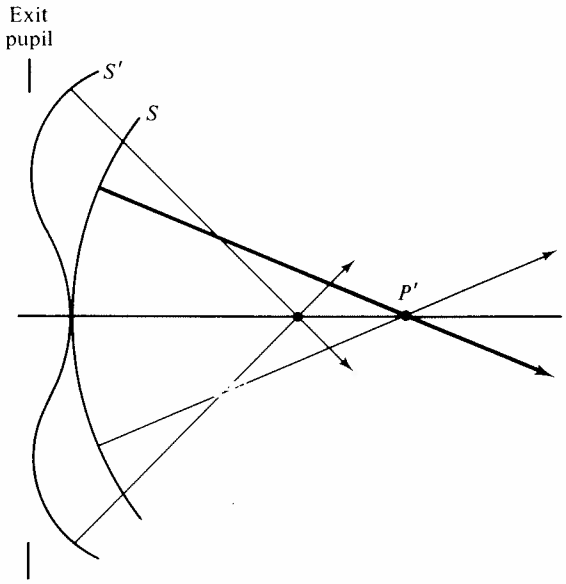


Figure 10-40

surface S which is bounded by the exit pupil. This sphere is the *wave front* producing the image.

10-11a The Wave Aberration

If the lens system producing this spherical wave front has aberrations, then the image may fall at some other point, and it may even be spread out to form a patch of light rather than a sharp point. We then realize that the sphere S is not generating the image; the image may be due to some other sphere and possibly to a distorted surface S', as shown in Fig. 10-40. Aberrations are small for small angles; therefore S' is similar to S near the axis, and the rays from this portion of the distorted wave front meet at the paraxial focus P'. However, rays from the more extreme parts of the wave front meet all along a line rather than forming a sharp image point.

To express the distortion of S into S' in a quantitative way, it is customary to use the distance w (Fig. 10-41) representing the variable spacing between

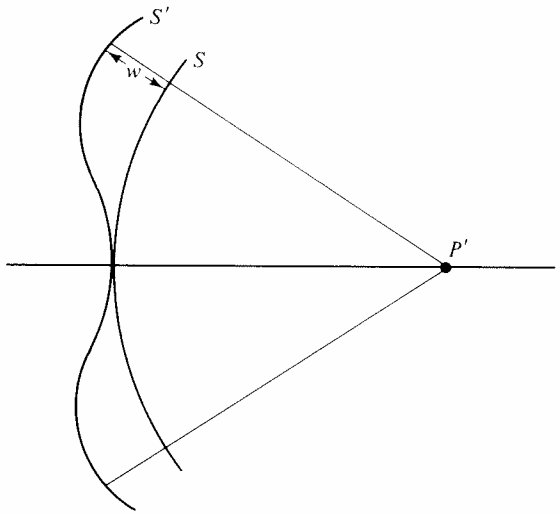


Figure 10-41

the two surfaces as measured along the radius of the sphere. This quantity is called the *wave aberration*. We shall find for spherical aberration, coma, and astigmatism that the wave aberration corresponds to a distorted surface. For curvature of the field and distortion, however, the sphere is unaltered, but displaced; the rays simply meet at the wrong point.

To study this problem quantitatively, consider the situation in Fig. 10-42. The undistorted wave front, called the *reference sphere*, has its center at the arbitrary point $P(x_0, y_0, z_0)$, and it passes through the origin (this is an

arbitrary choice on our part). We choose another point $P(x, y, 0)$ on the distorted wave front S' which lies in the x-y plane. Then the wave front aberration w is

$$\begin{aligned} w &= r - \overline{PP_0} \\ &= [x_0^2 + y_0^2 + z_0^2]^{1/2} - [(x_0 - x)^2 + (y_0 - y)^2 + z_0^2]^{1/2} \end{aligned} \tag{10-116}$$

We write this as

$$w = z_0\left[\left\{1 + \frac{x_0^2 + y_0^2}{z_0^2}\right\}^{1/2} + \left\{1 + \frac{(x_0 - x)^2 + (y_0 - y)^2}{z_0^2}\right\}^{1/2}\right] \tag{10-117}$$

Assuming that

$$z_0^2 \gg x_0^2 + y_0^2$$

we may expand in a binomial series to obtain

$$\begin{aligned} w = {} & z_0\left[1 + \frac{x_0^2 + y_0^2}{2z_0^2} - \frac{x_0^4 + 2x_0^2y_0^2 + y_0^4}{8z_0^4}\right] \\ & - z_0\left[1 + \frac{x_0^2 - 2xx_0 + x^2 + y_0^2 - 2yy_0 + y^2}{2z_0^2}\right. \\ & - \frac{1}{8z_0^4}\{x_0^4 + 4x^2x_0^2 + x^4 + y_0^4 + 4y^2y_0^2 + y^4 \\ & + 2(x_0^2x^2 + x_0^2y_0^2 + x_0^2y^2 + x^2y_0^2 + x^2y^2 + y_0^2y^2) \\ & + 2(4xx_0yy_0 - 2xx_0^3 - 2x^3x_0 - 2xx_0y_0^2 \\ & \left. - 2xx_0y^2 - 2yy_0x_0^2 - 2yy_0x^2 - 2yy_0^3 - 2y^3y_0)\}\right] \end{aligned}$$

or

$$\begin{aligned} w = {} & \left\{\frac{2xx_0 - x^2 + 2yy_0 - y^2}{2z_0}\right\} + \frac{1}{8z_0^3}[x^4 + y^4 + 2x^2y^2 + 4x^3x_0 \\ & + 4y^3y_0 + 4x^2yy_0 + 4xx_0y^2 - 6x^2x_0^2 - 6y^2y_0^2 - 2x^2y_0^2 \\ & - 2y^2x_0^2 - 8xx_0yy_0 + 4xx_0^3 + 4yy_0^3 + 4xx_0y_0^2 + 4yy_0x_0^2] \end{aligned} \tag{10-118}$$

Although this analytical expression for the value of w appears complicated, its interpretation is quite straightforward. Following a development of Rayces,[10-22] we show in Fig. 10-43 a point P on a distorted wave front and a point S on the reference sphere with radius r and center Q. The two surfaces have a common point C as they emerge from the optical system. The coordinate system is chosen so that XOY is the image plane and XOZ is the meridional plane.

Let a ray PT, normal to the actual wave front at P, strike the image plane at T. Then $\overline{QT}$ measures the deviation of the actual image from the center Q of the reference sphere; we therefore call it the *ray aberration*. The wave aberration w on the other hand is specified by the distance

$$\begin{aligned} w &= \overline{QS} - \overline{QP} \\ &= r - v \end{aligned} \tag{10-119}$$

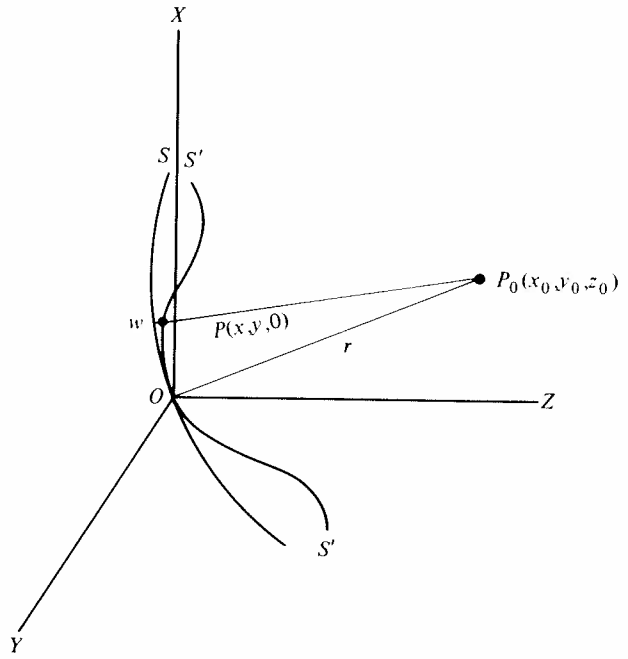


Figure 10-42

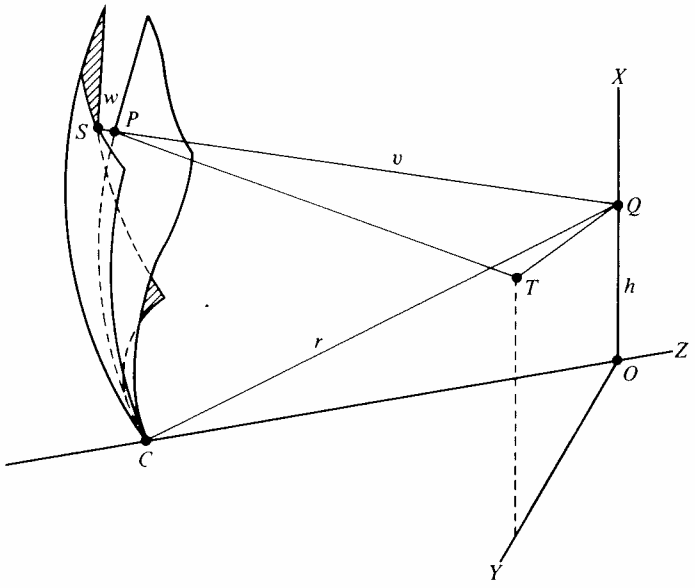


Figure 10-43

where we denote $\overline{QP}$ as v. If P has coordinates x, y, z, then the distance $v = \overline{PQ}$ may be expressed as

$$(x-h)^2 + y^2 + z^2 = v^2 \tag{10-120}$$

which may be written in the alternate form

$$G(x, y, z, v) = 0 \tag{10-121}$$

The surface containing P and C may also be specified by the equation

$$F(x, y, z) = 0 \tag{10-122}$$

or

$$z = f(x, y) \tag{10-123}$$

Further, we may eliminate z from Eqs. (10-121) and (10-122) to obtain

$$v = g(x, y) \tag{10-124}$$

From Eq. (10-123), we have

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy \tag{10-125}$$

and from Eq. (10-124)

$$dv = \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy \tag{10-126}$$

Differentiating the implicit function in Eq. (10-122) gives

$$\frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial y}\,dy + \frac{\partial F}{\partial z}\,dz = 0 \tag{10-127}$$

and doing the same to Eq. (10-121)

$$\frac{\partial G}{\partial x}\,dx + \frac{\partial G}{\partial y}\,dy + \frac{\partial G}{\partial z}\,dz + \frac{\partial G}{\partial v}\,dv = 0 \tag{10-128}$$

If we use Eqs. (10-125) and (10-126), these two equations become

$$\left(\frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x}\right)dz + \left(\frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y}\right)dy = 0 \tag{10-129}$$

and

$$\left(\frac{\partial G}{\partial x} + \frac{\partial G}{\partial z}\frac{\partial z}{\partial x} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial x}\right)dx + \left(\frac{\partial G}{\partial y} + \frac{\partial G}{\partial z}\frac{\partial z}{\partial y} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial y}\right)dy = 0 \tag{10-130}$$

Since dx and dy are independent variables, the last two equations are valid only if the individual coefficients of dx and dy vanish. Hence,

$$\begin{aligned}
\frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} &= 0 \\
\frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial y} &= 0 \\
\frac{\partial G}{\partial x} + \frac{\partial G}{\partial z}\frac{\partial z}{\partial x} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial x} &= 0 \\
\frac{\partial G}{\partial y} + \frac{\partial G}{\partial z}\frac{\partial z}{\partial y} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial y} &= 0
\end{aligned}$$

from which

$$\left.\begin{aligned}
\frac{\partial v}{\partial x} &= \frac{1}{\partial G/\partial v}\left(\frac{\partial G}{\partial z}\frac{\partial F/\partial x}{\partial F/\partial z} - \frac{\partial G}{\partial x}\right) \\
\frac{\partial v}{\partial y} &= \frac{1}{\partial G/\partial v}\left(\frac{\partial G}{\partial z}\frac{\partial F/\partial y}{\partial F/\partial z} - \frac{\partial G}{\partial y}\right)
\end{aligned}\right\} \tag{10-131}$$

The partial derivatives of G, from Eq. (10-120), are

$$\begin{aligned}
\frac{\partial G}{\partial x} &= \frac{\partial}{\partial x}[(x-h)^2 + y^2 + z^2 - v^2] = 2(x-h) \\
\frac{\partial G}{\partial y} &= 2y \\
\frac{\partial G}{\partial z} &= 2z \\
\frac{\partial G}{\partial v} &= -2v
\end{aligned}$$

and Eq. (10-130) becomes

$$\left.\begin{aligned}
\frac{\partial v}{\partial x} &= \frac{(x-h) - z\dfrac{\partial F/\partial x}{\partial F/\partial z}}{v} \\
\frac{\partial v}{\partial y} &= \frac{y - z\dfrac{\partial F/\partial y}{\partial F/\partial z}}{v}
\end{aligned}\right\} \tag{10-132}$$

We know from vector analysis that the gradient of a function F, as in Eq. (10-122), specifies the direction of the normal, and the three components of the gradient are proportional to the direction-cosines α, β, γ of this normal with respect to the axes. Hence

$$\frac{\partial F}{\partial x} : \frac{\partial F}{\partial y} : \frac{\partial F}{\partial z} = \alpha : \beta : \gamma \tag{10-133}$$

Since $\overline{PT}$ in the figure is such a normal, we shall designate its direction-cosines α, β, γ, as used in Eq. (10-133). Also, its end points will have coordinates x', y'. It can be shown that

$$\left.\begin{aligned}
\frac{\alpha}{\gamma} &= \frac{(x-h) - x'}{z} \\
\frac{\beta}{\gamma} &= \frac{y - y'}{z}
\end{aligned}\right\} \tag{10-134}$$

for if we have a vector $\mathbf{r}$ through the origin, then its direction-cosines α, β, γ are related to the components by

$$\alpha = \frac{x}{r}, \qquad \beta = \frac{y}{r}, \qquad \gamma = \frac{z}{r}$$

so that

$$\frac{\alpha}{\beta} = \frac{x}{y}$$

and so on; Eq. (10-134) is simply a generalization of this.

Combining Eqs. (10-133) and (10-134) converts Eq. (10-132) into

$$\frac{\partial v}{\partial x} = \frac{x'}{v}, \qquad \frac{\partial v}{\partial y} = \frac{y'}{v} \tag{10-135}$$

Finally, by Eq. (10-118)

$$\frac{\partial v}{\partial x} = -\frac{\partial w}{\partial x}, \qquad \frac{\partial v}{\partial y} = -\frac{\partial w}{\partial y}$$

and

$$\frac{\partial w}{\partial x} = \frac{x'}{r - w}, \qquad \frac{\partial w}{\partial y} = \frac{v'}{r - w} \tag{10-136}$$

These equations relate the wave aberration to the ray aberration. They are usually used in the approximate form

$$\frac{\partial w}{\partial x} = -\frac{x'}{r}, \qquad \frac{\partial w}{\partial y} = -\frac{y'}{r} \tag{10-137}$$

Returning to Eq. (10-118), we consider the first three terms of the final expression on the right. These are

$$\frac{x^4 + y^4 + 2x^2y^2}{8z_0^3} = \frac{(x^2 + y^2)^2}{8z_0^3} \tag{10-138}$$

We convert to polar coordinates in the plane of the exit pupil by

$$\begin{aligned} x &= \rho \cos\theta \\ y &= \rho \sin\theta \end{aligned} \tag{10-139}$$

Equation (10-138) is then

$$\frac{(x^2 + y^2)^2}{8z_0^3} = C_1\rho^4 \tag{10-140}$$

where $C_1 = 1/8z_0^3$ is a constant as far as the distorted sphere is concerned. Let us assume that this term represents the only contribution to w. Then

$$w = C_1\rho^4$$

and by Eq. (10-137), ignoring minus signs,

$$x' = rC_1\left[4\rho^3\frac{\partial\rho}{\partial x}\right]$$

$$= 4rC_1\rho^3\frac{x}{\rho} = 4rC_1x\rho^2$$

and

$$y' = 4rC_1y\rho^2$$

We then see that the radius ρ' of the image is

$$\rho' = \sqrt{x'^2 + y'^2} = 4rC_1\rho^2\sqrt{x^2 + y^2} = 4rC_1\rho^3 \tag{10-141}$$

so that the image diameter is proportional to the cube of the exit pupil diameter. This is just the behavior we found in Chapter 2 for spherical aberration when we are beyond the paraxial range but not yet into the fifth-power region.

The next four terms, if acting alone, would give a wave aberration

$$\begin{aligned} w &= \frac{4(x^3x_0 + y_1^3y_0 + x^2yy_0 + xx_0y^2)}{8z_0^3} \\ &= \frac{4\rho^3(x_0\cos^3\theta + y_0\sin^3\theta + y_0\cos^2\theta\sin\theta + x_0\cos\theta\sin^2\theta)}{8z_0^3} \\ &= \frac{\rho^3(x_0\cos\theta + y_0\sin\theta)}{2z_0^3} = \frac{(xx_0 + yy_0)\rho^2}{2z_0^3} \end{aligned} \tag{10-142}$$

It is convenient at this point to choose the image plane coordinates so that $x_0 = 0$. This is simply a matter of rotating the system about the z axis of Fig. 10-42 until P_0 lies in the y-z plane and causes no loss in generality. Then Eq. (10-142) becomes

$$w = C_2\rho^2yy_0 = C_2(x^2 + y^2)yy_0 \tag{10-143}$$

where $C_2 = \frac{1}{2}z_0^3$ and by Eq. (10-137)

$$\begin{aligned} x' &= 2y_0C_2xy \\ y' &= y_0C_2(x^2 + 3y^2) \end{aligned} \tag{10-144}$$

or converting back to polar coordinates

$$\begin{aligned} x' &= rhC_2\rho^2\sin 2\theta \\ y' &= rhC_2\rho^2(2 - \cos 2\theta) \end{aligned} \tag{10-145}$$

where we have written y_0 as h, the paraxial image height.

Equations (10-145) have the form of the well-known parametric equations for a circle, which are

$$\begin{aligned} x &= R\cos\varphi \\ y &= R\sin\varphi \end{aligned}$$

Hence, to each zone (ρ = constant) in the plane of the exit pupil, there is a circle in the image plane of radius $rhC_2\rho^2$ and with center on the y axis at the point $(0, -2rhC_2\rho^2)$. As ρ changes, so does the radius and the position of the image circle, which we recognize as a description of coma as the sole aberration.

Jumping to the last four terms in Eq. (10-118), we have

$$w = \frac{(x_0^3 + x_0y_0^2)x + (y_0^3 + y_0x_0^2)y}{2z_0} = \rho_0^2(xx_0 + yy_0) \tag{10-146}$$

Then

$$x' = \frac{\rho_0^2 x_0 r}{2z_0}$$
$$y' = \frac{\rho_0^2 y_0 r}{2z_0} \tag{10-147}$$

or

$$\rho' = \frac{r\rho_0^3}{2z_0} \tag{10-148}$$

Thus, the position of the image point depends on the cube of the distance of the paraxial image point from the axis. This leads to what we have previously identified as distortion.

Problem 10-16

Identify the remaining terms in the coefficient of $(1/8z_0^3)$ with astigmatism and curvature of the field. ■

10-11b The OTF for a Lens with Aberrations

Now that we have an analytical expression for the wave aberration which shows its connection with the third-order geometric aberrations, we can consider the *OTF* for a real lens. Actually, a simple example to start with involves the phenomenon of defocusing, which is not a true aberration. The quadratic terms in the first expression on the right of Eq. (10-118) give a contribution

$$w = \frac{x^2 + y^2}{2z_0} = \frac{\rho^2}{2z_0} \tag{10-149}$$

Then by Eq. (10-137)

$$x' = \frac{xr}{z_0}, \qquad y' = \frac{yr}{z_0}$$

and

$$\rho' = \frac{r}{z_0} \tag{10-150}$$

Hence, each ring of radius ρ in the exit pupil corresponds to a ring of radius ρ' at the image plane. However, recall that we obtained Eq. (10-87b) through the process of eliminating quadratic terms in Eq. (10-62) by satisfying the Gauss equation. In the general case, these terms involve exp $(x_L^2 + y_L^2)$, so that in our present notation, they correspond to $(x^2 + y^2)$, as measured in the exit pupil. Thus, we associate the quadratic expression in Eq. (10-149) with defocusing, and all we need do to eliminate this error is shift the image plane.

The effect of the wave aberration w on the light passing through a lens is to change the path length of each ray, as indicated in Fig. 10-41. This corresponds to a phase shift kw, which we incorporate into the diffraction integrals

and the definition of the *OTF* by the exponential term exp (ikw). Thus the pupil function $P(x, y)$, which is unity for an ideal lens, becomes

$$P(x, y) = e^{ikw(x,y)}$$

and the *OTF* integral in Eq. (10-106) is now

$$F(v_x, v_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ikw[\xi + (t_2' v_x/2k),\, \eta + (t_2' v_y/2k)]} e^{-ikw[\xi - (t_2' v_x/2k),\, \eta - (t_2' v_y/2k)]}\, d\xi\, d\eta \tag{10-151}$$

In the most general case, w will be the power series of Eq. (10-118), but it is easier to consider each aberration individually. Also, it is customary to write Eq. (10-118) in a different form when using it in *OTF* calculations. Let the polar coordinates of Eq. (10-139) be converted into reduced form by the relation

$$\rho' = \frac{\rho}{R}$$

where R is the radius of the exit pupil. We then drop the prime and let ρ designate a relative displacement. In the same way, $h = y_0$ of Fig. 10-41 will be treated as a relative value. If we then substitute cos θ, sin θ, 0, and h for x, y, x_0, and y_0, respectively, in Eq. (10-118), and consolidate all the constants into a set C_{ijk}, we obtain

$$w(h, \rho, \theta) = (C_{020}\rho^2 + C_{111}h\rho\cos\theta) + (C_{040}\rho^4 + C_{131}h\rho^3\cos\theta + C_{222}h^2\rho^2\cos^2\theta + C_{220}h^2\rho^2 + C_{311}h^3\rho\cos\theta) \tag{10-152}$$

where the subscripts on the C_{ijk} designate the exponents i, j, k in the expression $C_{ijk}h^i\rho^j\cos^k\theta$, and the two groups of terms constitute the defocusing effects and the third-order aberrations, respectively.

It has been shown by Longhurst[10-1] that this result, which we have laboriously calculated, is almost intuitively obvious. For example, having located the meridional plane by letting $x_0 = 0$, we see that $w(h, \rho, \theta)$ can depend on θ only through cos θ and not sin θ, since the aberrations are symmetric about this plane. For those aberrations which do not depend on θ—that is, they are symmetric with respect to the axis of the wave front—the terms involving h and ρ must be of the form h^2, ρ^2, or $h^2\rho^2$, since they must be independent of sign.

Problem 10-17

Explain the remaining terms in Eq. (10-152) in similar fashion. ■

As an example of an actual determination of the *OTF*, Hopkins[10-23] has calculated the frequency response of a defocused lens with either a square or a circular exit pupil. Using Eq. (10-151) with only $C_{020} \neq 0$ in Eq. (10-152)

and simplifying to a one-dimensional problem gives

$$F(v_x, 0) = \iint e^{ikC_{020}[(\xi + s/2)^2 + \eta^2]} e^{-ikC_{020}[(\xi - s/2)^2 + \eta^2]}\, d\xi\, d\eta \qquad (10\text{-}153)$$

where $s = v_x t_2'/k$, as before. Using relative coordinates, the overlap of two round openings has a maximum value of π (the area of a unit circle); thus Eq. (10-153), when normalized, is

$$F(v_x, 0) = \frac{1}{\pi} \iint e^{ia\xi}\, d\xi\, d\eta \qquad (10\text{-}154)$$

where

$$a = 2kC_{020}s$$

When the inclination of r in Fig. 10-37 is smaller than θ, we have

$$\left(\frac{s}{2} + \xi\right)^2 + \eta^2 = r^2 = 1$$

so that the limits for the integration over ξ in one quadrant are 0 and $\sqrt{1 - \eta^2} - s/2$. Then

$$F(v_x, 0) = \frac{4}{\pi a} \int_0^{\sqrt{1-(s/2)^2}} \sin\left[a\left\{\sqrt{1 - \eta^2} - \frac{s}{2}\right\}\right] d\eta \qquad (10\text{-}155)$$

This integral can be expressed as a power series in Bessel functions, but it is simpler to do it numerically.

Problem 10-18

Show that Eq. (10-155) reduces to the correct form when $C_{020} = 0$. ■

The behavior of the *OTF* as a function of the normalized spatial frequency v_x/v_{xo} is shown in Fig. 10-44. These plots were generated by letting

$$C_{020} = \frac{n\lambda}{\pi}$$

where $n = 0$, 1, 2, 3, and 4. Then a takes on values

$$a = 2ks\frac{n\lambda}{\pi}$$

$$= 4ns$$

and s goes from 0 to 2, the maximum value permitted in the upper limit.

Problem 10-19

Verify the integration numerically for $C_{020} = n\lambda/4$, $n = 0$ to 4. ■

A very significant feature of the curves for $n = 3$ or 4 is that the *OTF* can actually go negative. This implies a phase reversal in the image, as illustrated by the effect on the radial object of Fig. 10-45(a). The image shows an interchange of black and white areas corresponding to the places where $F(v_x, 0)$ crosses or touches the axis (Fig. 10-45(b)). This is reminiscent of the diffraction pattern due to a circular aperture (Fig. 10-6), and in fact, it is possible to confuse the Airy disc of an aberration-free lens with the first zero in the *OTF* curve of a lens for which the focusing error is about 1 wavelength. This is the well-known effect called *spurious resolution*.

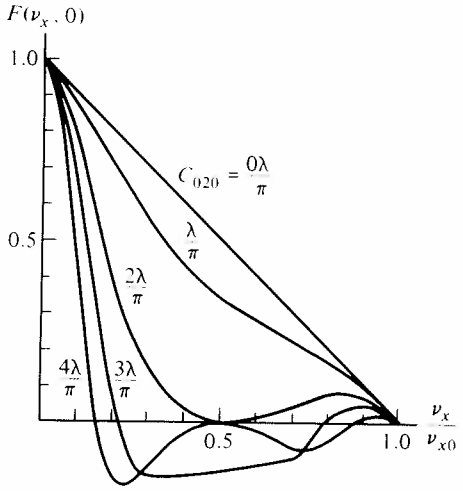


Figure 10-44

Problem 10-20

(a) Evaluate Eq. (10-151) for a unit square aperture, showing that

$$F(v_x, 0) = \frac{2}{a} \sin\left[\frac{(1 - s)a}{2}\right] \qquad (10\text{-}156)$$

(b) Show that when the focusing error is large, the cutoff frequency becomes

$$v_{ox} = \frac{\pi}{f'} C_{020}$$

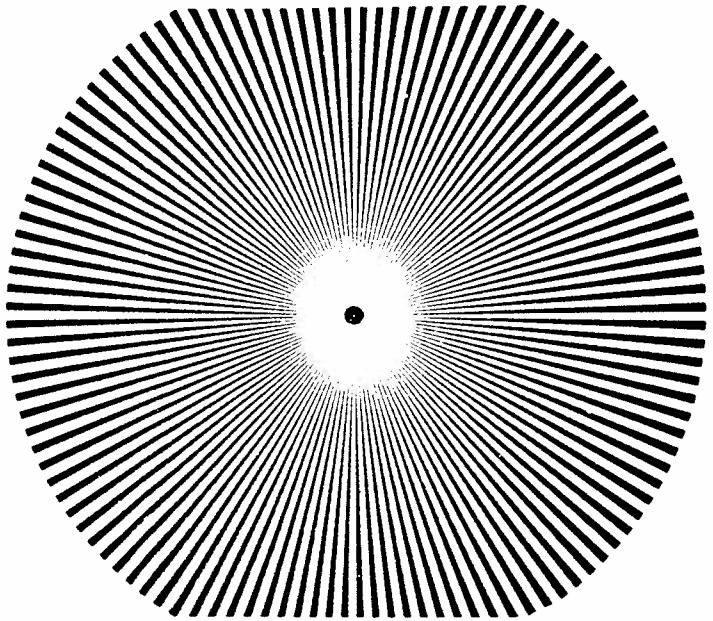
■

The third-order aberrations can be treated in a similar fashion. For example, the spherical aberration term in Eq. (10-152) is $C_{040}\rho^4$. Using this in Eq. (10-151) gives

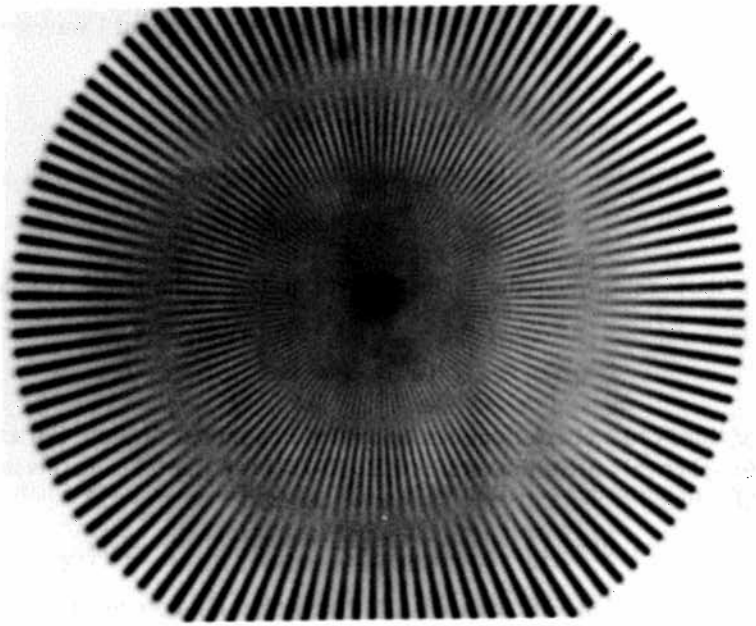
$$F(v_x, 0) = \iint e^{ikC_{040}[(\xi + s/2)^2 + \eta^2]^2} e^{-ikC_{040}[(\xi - s/2)^2 + \eta^2]^2}\, d\xi\, d\eta$$

$$= \iint e^{ikC_{040}(4\xi^3 s + \xi s^3 + 4\xi s\eta^2)}\, d\xi\, d\eta \qquad (10\text{-}157)$$

This integral must be evaluated numerically for both variables, and the results for small values of C_{040} are similar to those for $C_{020} = \lambda/\pi$ in Fig. 10-42.



(a)



(b)

Figure 10-45

10-11c Image Formation as a Convolution Process

The discussion of the *OTF* just given indicated that the diffraction of light by a given optical system implies that there is an upper limit to the ability of the system to resolve a bar target such as in Fig. 10-39. There is another vital piece of information, however, which comes from *OTF* calculations or measurements, and this we shall consider now. Following a development of Francon,[10-24] we shall show in a rather simple way that image formation is a convolution process representing a combination of the energy output of the object and the diffraction properties of the lens. Figure 10-46 shows a flat

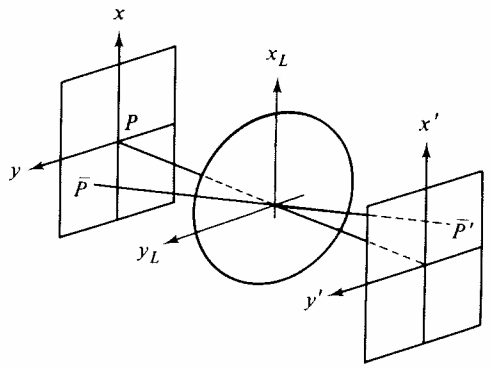


Figure 10-46

object in the x-y plane, the exit pupil of the lens in the x_L-y_L plane, and the image in the x'-y' plane. Let $E(x', y')$ be the illuminance of the image, each point of which is an Airy disc, and let $B(x, y)$ be the luminance (or brightness) of the corresponding point on the object. If we shift the object point P to a to a new position $\bar{P}$, the Airy disc at P' also moves, but the nature of the diffraction effect—except for the associated magnification β—is unchanged. Hence, if we normalize the image plane coordinates with respect to β, we can say that the illumination E anywhere on the image plane is a function only of the relative position of the image point with respect to the object point and may be written as $E(x' - x, y' - y)$. If we also normalize B and E so that we need not worry about units, then the luminous flux or relative intensity I received at the image plane is the integral of all the individual contributions, or

$$I(x', y') = \int\int E(x' - x, y' - y)B(x, y)\, dx\, dy \qquad (10\text{-}158)$$

As seen from the Appendix, this is a convolution relation. It is shown there, in fact, that a convolution arises when two independent physical phenomena are jointly contributing to a particular result. Equation (10-158) indicates that the brightness B of the object, when combined with the diffraction pattern E of each point [this is what we previously identified as the point

spread function or impulse response $h(x', y')$]—determines the nature of the image.

An interesting demonstration of optical convolution, which brings out the meaning of the mathematical processes involved, has been devised by Haskell.[10-25] He takes as an example the convolution of a two-slit pattern with a three-slit pattern (Fig. 10-47(a) and (b)). As described in connection with Fig. A-4 of the Appendix, the convolution is formed by reversing one of the patterns, say $h(\alpha)$, to obtain Fig. 10-47(c). Then we shift $h(-\alpha)$ by an amount t, chosen in this case as $t = -2$ (Fig. 10-47(d)). The corresponding values of $g(\alpha)$ and $h(t - \alpha)$ are then multiplied to obtain the product of Fig. 10-47(e). The total shaded area amounts to 2.0 units, and this is indicated in the convolution $f(t)$ of Fig. 10-47(f).

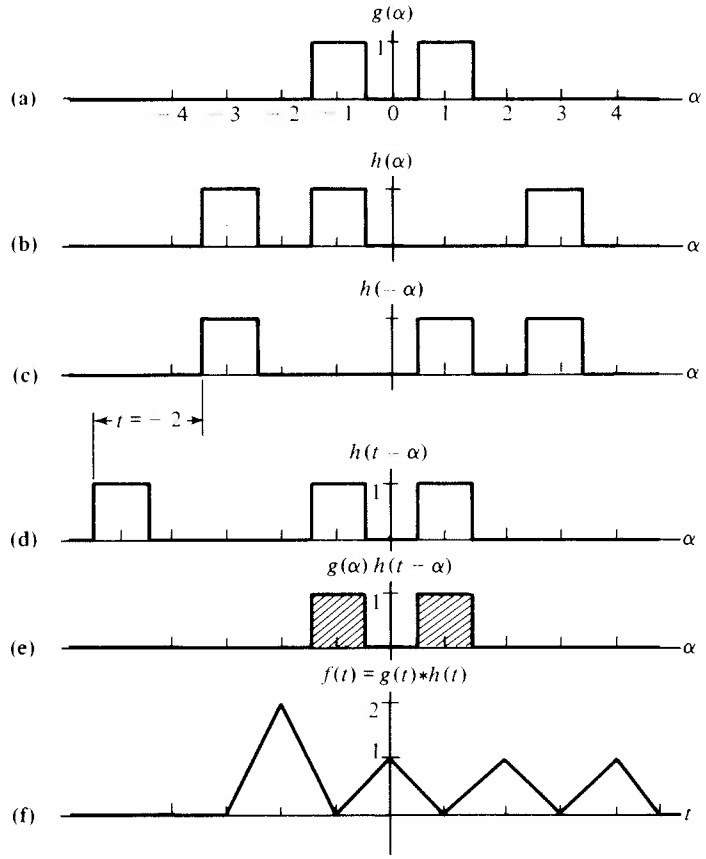
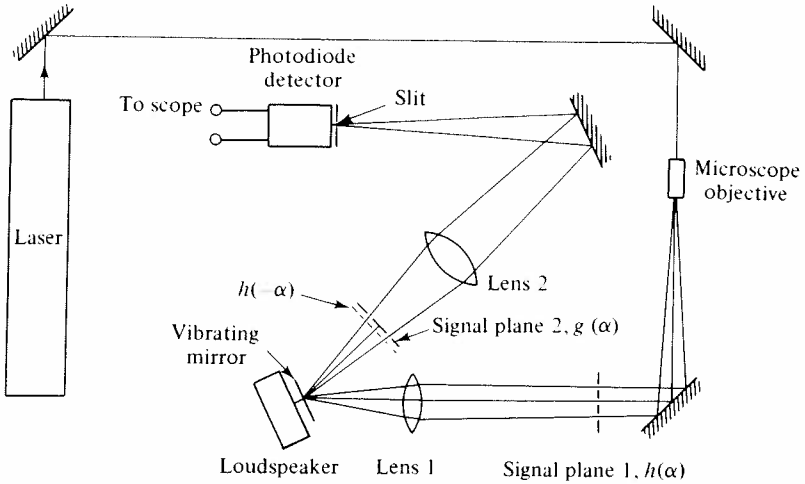


Figure 10-47

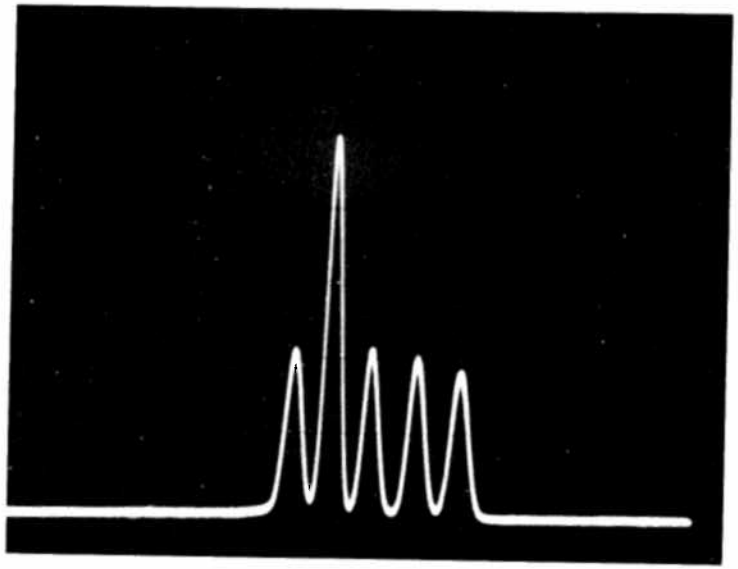
Problem 10-21

Justify the remainder of Fig. 10-47(f). ■

The experimental arrangement is shown in Fig. 10-48(a). A laser beam is expanded and forms an image of a three-slit pattern $h(\alpha)$ at the plane where



(a)



(b)

Figure 10-48

the two-slit pattern $g(\alpha)$ is located. Lens 1 gives an inverted image, which is equivalent to generating $h(-\alpha)$. The vibrating mirror sweeps $h(-\alpha)$ across $g(\alpha)$, thus performing the integration, and the received signal is displayed on an oscilloscope swept by the same signal that drives the speaker. The resulting trace (Fig. 10-48(b)) agrees very well with Fig. 10-47(f).

10-11d The Modulation Transfer Function

Returning to Eq. (10-158), our intuitive look at the meaning of the convolution process should indicate that image formation represents the combination of two independent functions: the variable energy output of the object and the effect of the diffraction process on each point in the object. To talk about the resulting resolution of the lens, we should use the black-and-white bar target as the object. It is convenient, however, to regard the target as alternate regions of high and low (but not zero) luminance and then decompose this pulse function into its Fourier series components. Let a typical target have a luminance of the form

$$B(x) = b_0 + b_1 \cos kx \tag{10-159}$$

It is shown in Fig. 10-49(a). Note that b_0 represents the average value of this component of the total output, and that the maximum and minimum values

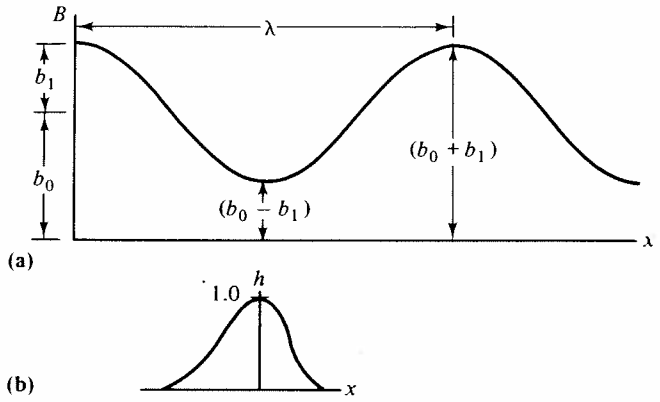


Figure 10-49

are, respectively, $(b_0 + b_1)$ and $(b_0 - b_1)$. Since the intensity I of the radiation is proportional to B, we can define the *contrast* or *modulation* by the relation

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{10-160}$$

which we recognize as identical in form to Eq. (6-14). The quantity M refers to variations in the illuminance received on a screen, whereas V specifies the

corresponding behavior of a fringe system observed by the eye. Using Eq. (10-159), we obtain

$$M_0 = \frac{b_1}{b_0} \tag{10-161}$$

where the subscript refers to the object function.

Let us next postulate some sort of arbitrary diffraction behavior represented by the impulse response curve of Fig. 10-49(b); the skew shape represents the distortion of sinc x by aberrations. We then perform the convolution of h with the cosine curve, but we shall do it analytically rather than graphically. The quantity h is a measure of the illuminance at each point of the image so that this convolution will determine the image intensity I. Equation (10-58) in one dimension, using Eq. (A-46), becomes

$$\begin{aligned} I(x') &= \int h(x)\{b_0 + b_1 \cos k(x' - x)\}\, dx \\ &= b_0 \int h(x)\, dx + b_1 \int h(x) \cos kx' \cos kx\, dx \\ &\quad + b_1 \int h(x) \sin kx' \sin kx\, dx \end{aligned} \tag{10-162}$$

Normalizing $I(x')$ with respect to $\int h(x)\, dx$ (and retaining the same symbol) gives

$$\begin{aligned} I(x') &= b_0 + b_1 \frac{\int h(x) \cos kx\, dx}{\int h(x)\, dx} \cos kx' + b_1 \frac{\int h(x) \sin kx\, dx}{\int h(x)\, dx} \sin kx' \\ &= b_0 + b_1 h_R(k) \cos kx' + b_1 h_I(k) \sin kx' \end{aligned} \tag{10-163}$$

where $k_R(k)$ and $h_I(k)$ are the real and imaginary parts, respectively, of the complex quantity

$$h(k) = \frac{\int h(x) e^{ikx}\, dx}{\int h(x)\, dx} \tag{10-164}$$

We may write Eq. (10-163) as

$$I(x') = b_0 + b_1 |h(k)| \cos (kx' - \varphi) \tag{10-165}$$

where

$$|h(k)| = \sqrt{h_R^2(k) + h_I^2(k)}$$

and

$$\varphi = \arctan \frac{h_I(k)}{h_R(k)}$$

Then the modulation M_i of the image, from Eqs. (10-160) and (10-165), is

$$\begin{aligned} M_i &= \frac{b_1 |h(k)|}{b_0} \\ &= M_0 |h(k)| \end{aligned} \tag{10-166}$$

Thus, we see that the ratio of the image modulation to the object modulation is given by the magnitude of the impulse response. Since we have used $h(x)$ as a measure of intensity, we can say that $|h|$—known as the *modulation transfer function* (*MTF*)—is in fact the absolute magnitude of the spread or impulse response function we have previously used. The places where either the *OTF* or the *MTF* goes to zero represent a complete reversal in contrast since F in Eq. (10-99) depends on h.

The theory based on Eqs. (10-151) and (10-152) can be applied only when we know the aberration coefficients C_{ijk}. Since these are difficult to determine, another approach to the computation of the *MTF* is via the *spot diagram*, as described by Smith.[3-1] Consider a plane object and a group of points distributed uniformly and concentrically about the axis. The places where a parallel set of rays from each of these points strikes the paraxial image plane shown in Fig. 10-50(a) constitute the spot diagram. It may not reproduce the original distribution because of aberrations, and we obtain the spread function by simply combining the number of spots per unit width of the pattern. (Strictly speaking, this is a *line spread function*, which is appropriate for the image of a line in a cylindrical lens, but we are doing a one-dimensional problem in any case.) The function $h(x)$ obtained in this way is normalized with respect to its maximum value (Fig. 10-50(b)) and used to evaluate the integrals in Eq. (10-163), which are approximated by summations. It is also possible to measure the *MTF* directly by using cosine-wave targets and evaluating the integrals with data-processing techniques. A complete survey of the various methods is given by Rosenhauer and Rosenbruch.[10-26]

The significance of the results we have obtained here has been summarized by Smith in the following terms: excellent reproduction (serifs are distinguishable) of a printed page requires the ability to resolve eight line pairs (one black and one white bar of identical width) per height of a lowercase *e*; legible reproduction requires five line pairs; and barely decipherable copy corresponds to three line pairs.

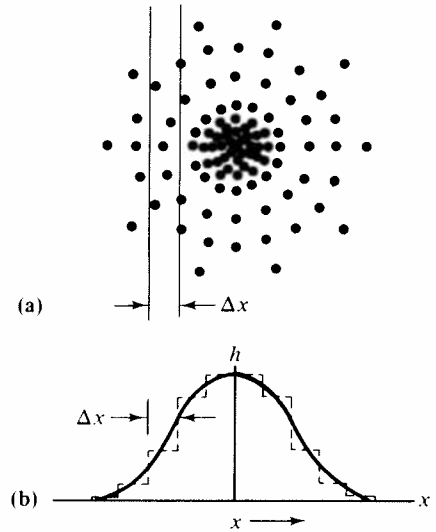


Figure 10-50

Problem 10-22

By regarding a plane wave as a series of wave fronts spaced a distance equal to the wavelength λ, show that the Bragg law of X-ray diffraction (see Halliday and Resnick[10-2]) may be obtained from Huyghens' principle. ■

Problem 10-23

Show that the diffraction pattern of an equilateral triangle is a six-pointed star, each arm of which is very similar to the pattern for a single slit. This result can be verified very easily by forming such an opening from three razor blades taped together with edges of about 1 mm. The aperture is then placed in front of a laser. ■

Problem 10-24

Figure 10-51(a) shows a pattern and Fig. 10-51(b) indicates its appearance after being spatially filtered. Describe the shape of the filter and explain how it produced the filtered image. ■

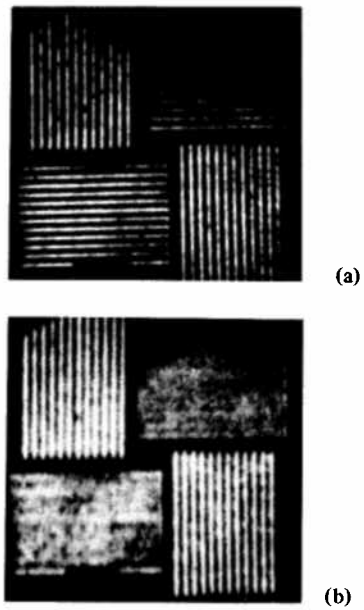


Figure 10-51