

---

# FOURIER OPTICS

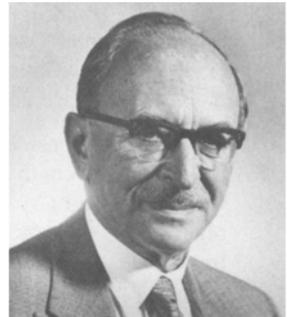
- 4.1 PROPAGATION OF LIGHT IN FREE SPACE
  - A. Correspondence Between the Spatial Harmonic Function and the Plane Wave
  - B. Transfer Function of Free Space
  - C. Impulse-Response Function of Free Space
- 4.2 OPTICAL FOURIER TRANSFORM
  - A. Fourier Transform in the Far Field
  - B. Fourier Transform Using a Lens
- 4.3 DIFFRACTION OF LIGHT
  - A. Fraunhofer Diffraction
  - \*B. Fresnel Diffraction
- 4.4 IMAGE FORMATION
  - A. Ray-Optics Description of Image Formation
  - B. Spatial Filtering
  - C. Single-Lens Imaging System
- 4.5 HOLOGRAPHY



**Josef von Fraunhofer (1787–1826)** developed diffraction gratings and contributed to the understanding of light diffraction. His epitaph reads “*Approximavit sidera*; he brought the stars nearer.”



**Jean-Baptiste Joseph Fourier (1768–1830)** recognized that periodic functions can be considered as sums of sinusoids. Harmonic analysis is the basis of Fourier optics.

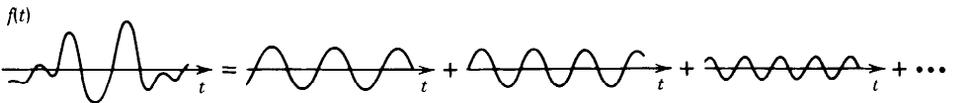


**Dennis Gabor (1900–1979)** made the first hologram in 1947. He received the Nobel Prize in 1971.

Fourier optics provides a description of the propagation of light waves based on harmonic analysis (the Fourier transform) and linear systems. The methods of harmonic analysis have proven to be useful in describing signals and systems in many disciplines. Harmonic analysis is based on the expansion of an arbitrary function of time  $f(t)$  as a superposition (a sum or an integral) of harmonic functions of time of different frequencies (see Appendix A, Sec. A.1). The harmonic function  $F(\nu)\exp(j2\pi\nu t)$ , which has frequency  $\nu$  and complex amplitude  $F(\nu)$ , is the building block of the theory. Several of these functions, each with its own value of  $F(\nu)$ , are added to construct the function  $f(t)$ , as illustrated in Fig. 4.0-1. The complex amplitude  $F(\nu)$ , as a function of frequency, is called the Fourier transform of  $f(t)$ . This approach is useful for the description of linear systems (see Appendix B, Sec. B.1). If the response of the system to each harmonic function is known, the response to an arbitrary input function is readily determined by the use of harmonic analysis at the input and superposition at the output.

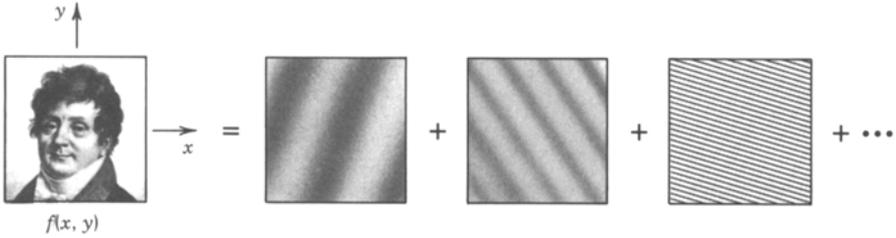
An arbitrary function  $f(x, y)$  of the two variables  $x$  and  $y$ , representing the spatial coordinates in a plane, may similarly be written as a superposition of harmonic functions of  $x$  and  $y$  of the form  $F(\nu_x, \nu_y)\exp[-j2\pi(\nu_x x + \nu_y y)]$ , where  $F(\nu_x, \nu_y)$  is the complex amplitude and  $\nu_x$  and  $\nu_y$  are the **spatial frequencies** (cycles per unit length; typically cycles/mm) in the  $x$  and  $y$  directions, respectively.<sup>†</sup> The harmonic function  $F(\nu_x, \nu_y)\exp[-j2\pi(\nu_x x + \nu_y y)]$  is the two-dimensional building block of the theory. It can be used to generate an arbitrary function of two variables  $f(x, y)$ , as illustrated in Fig. 4.0-2 (see Appendix A, Sec. A.3).

The plane wave  $U(x, y, z) = A\exp[-j(k_x x + k_y y + k_z z)]$  plays an important role in wave optics. The coefficients  $(k_x, k_y, k_z)$  are components of the wavevector  $\mathbf{k}$  and  $A$  is a complex constant. At points in an arbitrary plane,  $U(x, y, z)$  is a spatial harmonic function. In the  $z = 0$  plane, for example,  $U(x, y, 0)$  is the harmonic function  $f(x, y) = A\exp[-j2\pi(\nu_x x + \nu_y y)]$ , where  $\nu_x = k_x/2\pi$  and  $\nu_y = k_y/2\pi$  are the spatial frequencies (cycles/mm) and  $k_x$  and  $k_y$  are the spatial angular frequencies (radians/mm). There is a one-to-one correspondence between the plane wave  $U(x, y, z)$  and the spatial harmonic function  $f(x, y) = U(x, y, 0)$ , provided that the spatial frequency does not exceed the inverse wavelength  $1/\lambda$ . Since an arbitrary function  $f(x, y)$  can be analyzed as a superposition of harmonic functions, an arbitrary traveling

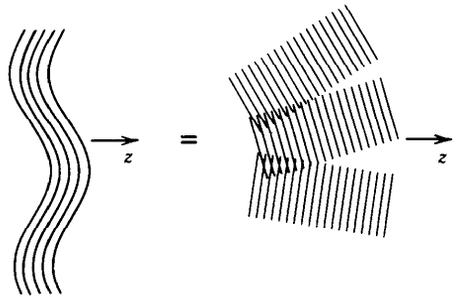


**Figure 4.0-1** An arbitrary function  $f(t)$  may be analyzed as a sum of harmonic functions of different frequencies and complex amplitudes.

<sup>†</sup>The spatial harmonic function is defined with a minus sign in the exponent, in contrast to the plus sign used in the definition of the temporal harmonic function (see Appendix A, Sec. A.3). These signs match those of a forward-traveling plane wave.



**Figure 4.0-2** An arbitrary function  $f(x, y)$  may be analyzed as a sum of harmonic functions of different spatial frequencies and complex amplitudes.

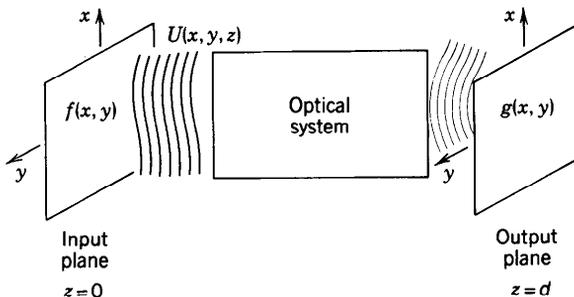


**Figure 4.0-3** The principle of Fourier optics: an arbitrary wave in free space can be analyzed as a superposition of plane waves.

wave  $U(x, y, z)$  may be analyzed as a sum of plane waves (Fig. 4.0-3). The plane wave is the building block used to construct a wave of arbitrary complexity. Furthermore, if it is known how a linear optical system modifies plane waves, the principle of superposition can be used to determine the effect of the system on an arbitrary wave.

Because of the important role Fourier analysis plays in describing linear systems, it is useful to describe the propagation of light through linear optical components, including free space, using a linear-system approach. The complex amplitudes in two planes normal to the optic ( $z$ ) axis are regarded as the input and output of the system (Fig. 4.0-4). A linear system may be characterized by either its **impulse-response function** (the response of the system to an impulse, or a point, at the input) or by its **transfer function** (the response to spatial harmonic functions), as described in Appendix B.

The chapter begins with a Fourier description of the propagation of light in free space (Sec. 4.1). The transfer function and impulse-response function of the free-space



**Figure 4.0-4** The transmission of an optical wave  $U(x, y, z)$  through an optical system between an input plane  $z = 0$  and an output plane  $z = d$ . This is regarded as a linear system whose input and output are the functions  $f(x, y) = U(x, y, 0)$  and  $g(x, y) = U(x, y, d)$ , respectively.

propagation system are determined. In Sec. 4.2 we show that a lens may perform the operation of the spatial Fourier transform. The transmission of light through apertures is discussed in Sec. 4.3; this is a Fourier-optics approach to the diffraction of light. Section 4.4 is devoted to image formation and spatial filtering. Finally, an introduction to holography, the recording and reconstruction of optical waves, is presented in Sec. 4.5. Knowledge of the basic properties of the Fourier transform and linear systems in one and two dimensions (reviewed in Appendices A and B) is necessary for understanding this chapter.

### 4.1 PROPAGATION OF LIGHT IN FREE SPACE

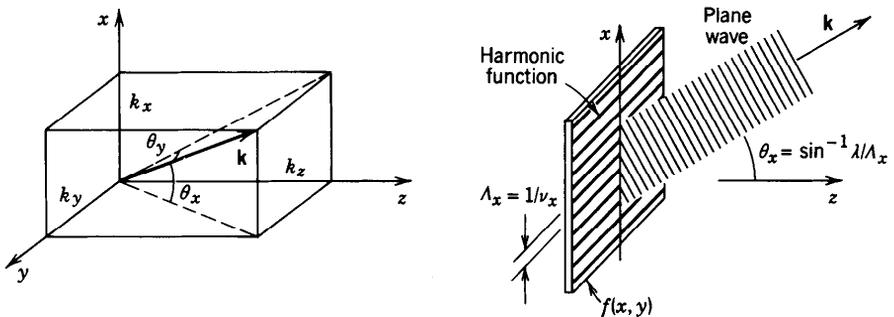
#### A. Correspondence Between the Spatial Harmonic Function and the Plane Wave

Consider a plane wave of complex amplitude  $U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)]$  with wavevector  $\mathbf{k} = (k_x, k_y, k_z)$ , wavelength  $\lambda$ , wavenumber  $k = (k_x^2 + k_y^2 + k_z^2)^{1/2} = 2\pi/\lambda$ , and complex envelope  $A$ . The vector  $\mathbf{k}$  makes angles  $\theta_x = \sin^{-1}(k_x/k)$  and  $\theta_y = \sin^{-1}(k_y/k)$  with the  $y$ - $z$  and  $x$ - $z$  planes, respectively, as illustrated in Fig. 4.1-1. The complex amplitude in the  $z = 0$  plane,  $U(x, y, 0)$ , is a spatial harmonic function  $f(x, y) = A \exp[-j2\pi(\nu_x x + \nu_y y)]$  with spatial frequencies  $\nu_x = k_x/2\pi$  and  $\nu_y = k_y/2\pi$  (cycles/mm). The angles of the wavevector are therefore related to the spatial frequencies of the harmonic function by

$$\begin{aligned} \theta_x &= \sin^{-1} \lambda \nu_x \\ \theta_y &= \sin^{-1} \lambda \nu_y. \end{aligned}$$

(4.1-1)  
Correspondence Between  
Spatial Frequencies and  
Angles

Recognizing  $\Lambda_x = 1/\nu_x$  and  $\Lambda_y = 1/\nu_y$  as the periods of the harmonic function in the  $x$  and  $y$  directions, we see that the angles  $\theta_x = \sin^{-1}(\lambda/\Lambda_x)$  and  $\theta_y = \sin^{-1}(\lambda/\Lambda_y)$  are governed by the ratios of the wavelength of light to the period of the harmonic function in each direction. These geometrical relations follow from matching the wavefronts of the wave to the periodic pattern of the harmonic function in the  $z = 0$  plane, as illustrated in Fig. 4.1-1.



**Figure 4.1-1** A harmonic function of spatial frequencies  $\nu_x$  and  $\nu_y$  at the plane  $z = 0$  is consistent with a plane wave traveling at angles  $\theta_x = \sin^{-1} \lambda \nu_x$  and  $\theta_y = \sin^{-1} \lambda \nu_y$ .

If  $k_x \ll k$  and  $k_y \ll k$ , so that the wavevector  $\mathbf{k}$  is paraxial, the angles  $\theta_x$  and  $\theta_y$  are small ( $\sin \theta_x \approx \theta_x$  and  $\sin \theta_y \approx \theta_y$ ) and

$$\begin{aligned} \theta_x &\approx \lambda \nu_x \\ \theta_y &\approx \lambda \nu_y. \end{aligned}$$

(4.1-2)

Spatial Frequencies and Angles  
(Paraxial Approximation)

Thus the angles of inclination of the wavevector are directly proportional to the spatial frequencies of the corresponding harmonic function.

Apparently, there is a one-to-one correspondence between the plane wave  $U(x, y, z)$  and the harmonic function  $f(x, y)$ . Given one, the other can be readily determined (if the wavelength  $\lambda$  is known). Given the wave  $U(x, y, z)$ , the harmonic function  $f(x, y)$  is obtained by sampling in the  $z = 0$  plane,  $f(x, y) = U(x, y, 0)$ . Given the harmonic function  $f(x, y)$ , on the other hand, the wave  $U(x, y, z)$  is constructed by using the relation  $U(x, y, z) = f(x, y) \exp(-jk_z z)$  with

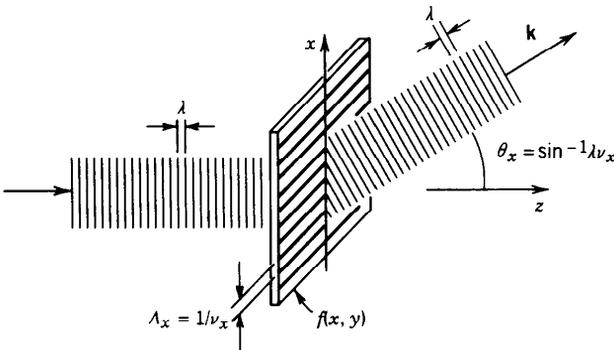
$$k_z = \pm (k^2 - k_x^2 - k_y^2)^{1/2}, \quad k = 2\pi/\lambda. \tag{4.1-3}$$

A condition of validity of this correspondence is that  $k_x^2 + k_y^2 < k^2$ , so that  $k_z$  is real. This condition implies that  $\lambda \nu_x < 1$  and  $\lambda \nu_y < 1$ , so that the angles  $\theta_x$  and  $\theta_y$  defined by (4.1-1) exist. The + and - signs in (4.1-3) represent waves traveling in the forward and backward directions, respectively. We shall be concerned with forward waves only.

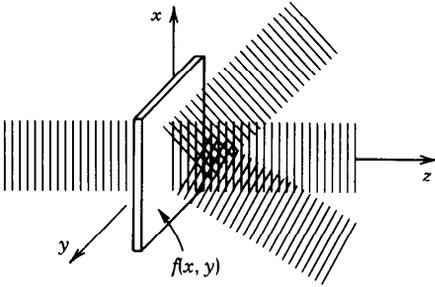
**Spatial Spectral Analysis**

When a plane wave of unity amplitude traveling in the  $z$  direction is transmitted through a thin optical element with complex amplitude transmittance  $f(x, y) = \exp[-j2\pi(\nu_x x + \nu_y y)]$  the wave is modulated by the harmonic function, so that  $U(x, y, 0) = f(x, y)$ . The incident wave is then converted into a plane wave with a wavevector at angles  $\theta_x = \sin^{-1} \lambda \nu_x$  and  $\theta_y = \sin^{-1} \lambda \nu_y$  (see Fig. 4.1-2). The optical element is a diffraction grating which acts like a prism (see Exercise 2.4-5).

If the transmittance of the optical element  $f(x, y)$  is the sum of several harmonic functions of different spatial frequencies, the transmitted optical wave is also the sum of an equal number of plane waves dispersed into different directions; each spatial frequency is mapped into a corresponding direction, in accordance with (4.1-1). The



**Figure 4.1-2** A thin element whose amplitude transmittance is a harmonic function of spatial frequency  $\nu_x$  (period  $\Lambda_x = 1/\nu_x$ ) bends a plane wave of wavelength  $\lambda$  by an angle  $\theta_x = \sin^{-1} \lambda \nu_x = \sin^{-1}(\lambda/\Lambda_x)$ .



**Figure 4.1-3** A thin optical element of amplitude transmittance  $f(x, y)$  decomposes an incident plane wave into many plane waves. The plane wave traveling at the angles  $\theta_x = \sin^{-1} \lambda \nu_x$  and  $\theta_y = \sin^{-1} \lambda \nu_y$  has a complex envelope  $F(\nu_x, \nu_y)$ , the Fourier transform of  $f(x, y)$ .

amplitude of each wave is proportional to the amplitude of the corresponding harmonic component of  $f(x, y)$ .

More generally, if  $f(x, y)$  is a superposition integral of harmonic functions,

$$f(x, y) = \iint_{-\infty}^{\infty} F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)] d\nu_x d\nu_y, \quad (4.1-4)$$

with frequencies  $(\nu_x, \nu_y)$  and amplitudes  $F(\nu_x, \nu_y)$ , the transmitted wave  $U(x, y, z)$  is the superposition of plane waves,

$$U(x, y, z) = \iint_{-\infty}^{\infty} F(\nu_x, \nu_y) \exp[-j(2\pi\nu_x x + 2\pi\nu_y y)] \exp(-jk_z z) d\nu_x d\nu_y,$$

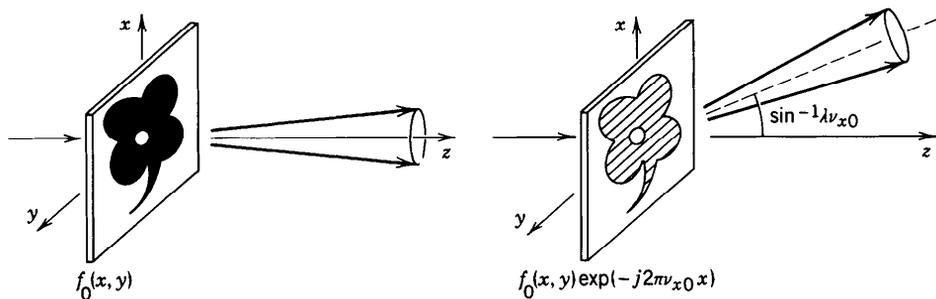
with complex envelopes  $F(\nu_x, \nu_y)$ , where  $k_z = (k^2 - k_x^2 - k_y^2)^{1/2} = 2\pi(1/\lambda^2 - \nu_x^2 - \nu_y^2)^{1/2}$ . Note that  $F(\nu_x, \nu_y)$  is the Fourier transform of  $f(x, y)$  (see Appendix A, Sec. A.3).

Since an arbitrary function may be Fourier analyzed as a superposition integral of the form (4.1-4), the light transmitted through a thin optical element of arbitrary transmittance may be written as a superposition of plane waves (see Fig. 4.1-3), provided that  $\nu_x^2 + \nu_y^2 < 1/\lambda^2$ . This process of “spatial spectral analysis” is akin to the angular dispersion of different temporal-frequency components (wavelengths) provided by a prism. Free-space propagation serves as a natural “spatial prism,” sensitive to the spatial instead of the temporal frequencies of the optical wave.

### Amplitude Modulation

Consider a transparency with complex amplitude transmittance  $f_0(x, y)$ . If the Fourier transform  $F_0(\nu_x, \nu_y)$  extends over widths  $\pm \Delta\nu_x$  and  $\pm \Delta\nu_y$  in the  $x$  and  $y$  directions, the transparency will deflect an incident plane wave by angles  $\theta_x$  and  $\theta_y$  in the range  $\pm \sin^{-1}(\lambda \Delta\nu_x)$  and  $\pm \sin^{-1}(\lambda \Delta\nu_y)$ , respectively.

Consider a second transparency of complex amplitude transmittance  $f(x, y) = f_0(x, y) \exp[-j2\pi(\nu_{x0}x + \nu_{y0}y)]$ , where  $f_0(x, y)$  is slowly varying compared to  $\exp[-j2\pi(\nu_{x0}x + \nu_{y0}y)]$  so that  $\Delta\nu_x \ll \nu_{x0}$  and  $\Delta\nu_y \ll \nu_{y0}$ . We may regard  $f(x, y)$  as an amplitude-modulated function with a carrier frequency  $\nu_{x0}$  and  $\nu_{y0}$  and modulation function  $f_0(x, y)$ . The Fourier transform of  $f(x, y)$  is  $F_0(\nu_x - \nu_{x0}, \nu_y - \nu_{y0})$ , in accordance with the frequency-shifting property of the Fourier transform (see Appendix A). The transparency will deflect a plane wave to directions centered about the angles  $\theta_{x0} = \sin^{-1} \lambda \nu_{x0}$  and  $\theta_{y0} = \sin^{-1} \lambda \nu_{y0}$  (Fig. 4.1-4). This can also be readily seen by regarding  $f(x, y)$  as a transparency of transmittance  $f_0(x, y)$  in contact with a grating or prism of transmittance  $\exp[-j2\pi(\nu_{x0}x + \nu_{y0}y)]$  that provides the angular deflection  $\theta_{x0}$  and  $\theta_{y0}$ .



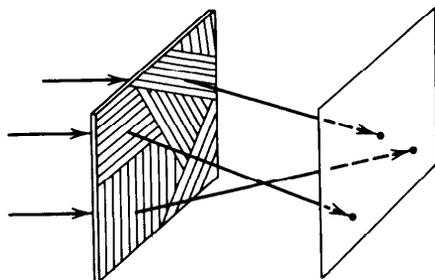
**Figure 4.1-4** Deflection of light by the transparencies  $f_0(x, y)$  and  $f_0(x, y)\exp(-j2\pi\nu_{x0}x)$ . The “carrier” harmonic function  $\exp(-j2\pi\nu_{x0}x)$  acts as a prism that deflects the wave by an angle  $\theta_{x0} = \sin^{-1} \lambda\nu_{x0}$ .

This idea may be used to record two images  $f_1(x, y)$  and  $f_2(x, y)$  on the same transparency using the *spatial-frequency multiplexing* scheme  $f(x, y) = f_1(x, y)\exp[-j2\pi(\nu_{x1}x + \nu_{y1}y)] + f_2(x, y)\exp[-j2\pi(\nu_{x2}x + \nu_{y2}y)]$ . The two images may be easily separated by illuminating the transparency with a plane wave, whereupon the two images are deflected at different angles and are thus separated. This principle will prove useful in holography (Sec. 4.5), where it is often desired to separate two images recorded on the same transparency.

**Frequency Modulation**

We now examine the transmission of a plane wave through a transparency made of a “collage” of several regions, the transmittance of each of which is a harmonic function of some spatial frequency, as illustrated in Fig. 4.1-5. If the dimensions of each region are much greater than the period, each region acts as a grating or a prism that deflects the wave in some direction, so that different portions of the incident wavefront are deflected into different directions. This principle may be used to create maps of optical interconnections, which may be used in optical computing applications, as described in Sec. 21.5.

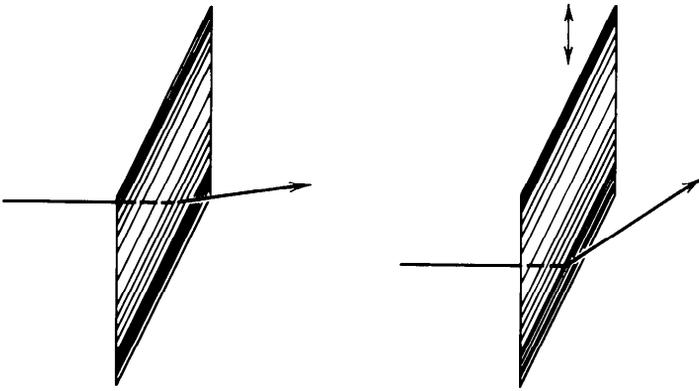
A transparency may also have a harmonic transmittance with a spatial frequency that varies continuously and slowly with position (in comparison with  $\lambda$ ), much as the frequency of a frequency-modulated (FM) signal varies slowly with time. Consider, for example, the phase function  $f(x, y) = \exp[-j2\pi\phi(x, y)]$ , where  $\phi(x, y)$  is a continuous slowly varying function of  $x$  and  $y$ . In the neighborhood of a point  $(x_0, y_0)$ , we may use the Taylor’s series expansion  $\phi(x, y) \approx \phi(x_0, y_0) + (x - x_0)\nu_x + (y - y_0)\nu_y$ , where the derivatives  $\nu_x = \partial\phi/\partial x$  and  $\nu_y = \partial\phi/\partial y$  are evaluated at the position  $(x_0, y_0)$ . The local variation of  $f(x, y)$  with  $x$  and  $y$  is therefore proportional to the quantity  $\exp[-j2\pi(\nu_x x + \nu_y y)]$ , which is a harmonic function with spatial frequencies



**Figure 4.1-5** Deflection of light by a transparency made of several harmonic functions (phase gratings) of different spatial frequencies.

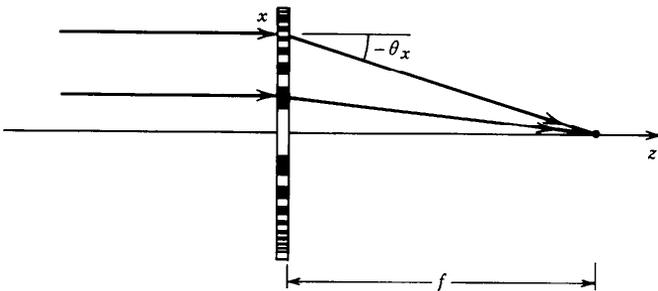
$\nu_x = \partial\phi/\partial x$  and  $\nu_y = \partial\phi/\partial y$ . Since the derivatives  $\partial\phi/\partial x$  and  $\partial\phi/\partial y$  vary with  $x$  and  $y$ , so do the spatial frequencies. The transparency  $f(x, y) = \exp[-j2\pi\phi(x, y)]$  therefore deflects the portion of the wave at the position  $(x, y)$  by the position-dependent angles  $\theta_x = \sin^{-1}(\lambda \partial\phi/\partial x)$  and  $\theta_y = \sin^{-1}(\lambda \partial\phi/\partial y)$ .

**EXAMPLE 4.1-1. Scanning.** A thin transparency with complex amplitude transmittance  $f(x, y) = \exp(j\pi x^2/\lambda f)$  introduces a phase shift  $2\pi\phi(x, y)$  where  $\phi(x, y) = -x^2/2\lambda f$ , so that the wave is deflected at the position  $(x, y)$  by the angles  $\theta_x = \sin^{-1}(\lambda \partial\phi/\partial x) = \sin^{-1}(-x/f)$  and  $\theta_y = 0$ . If  $|x/f| \ll 1$ ,  $\theta_x \approx -x/f$  and the deflection angle  $\theta_x$  is directly proportional to the transverse distance  $x$ . This transparency may be used to deflect a narrow beam of light. If the transparency is moved at a uniform speed, the beam is deflected by a linearly increasing angle as illustrated in Fig. 4.1-6.



**Figure 4.1-6** Using a frequency-modulated transparency to scan an optical beam.

**EXAMPLE 4.1-2. Imaging.** If the transparency in Example 4.1-1 is illuminated by a plane wave, each part of the wave is deflected by a different angle and as a result the wavefront is altered. The local wavevector at position  $x$  bends by an angle  $-x/f$  so that all wavevectors meet at a single point on the optical axis a distance  $f$  from the transparency, as illustrated in Fig. 4.1-7. The transparency acts as a cylindrical lens with focal length  $f$ . Similarly, a transparency with the transmittance  $f(x, y) = \exp[j\pi(x^2 + y^2)/\lambda f]$  acts as a



**Figure 4.1-7** A transparency with transmittance  $f(x, y) = \exp(j\pi x^2/\lambda f)$  bends the wave at position  $x$  by an angle  $\theta_x \approx -x/f$  so that it acts as a cylindrical lens with focal length  $f$ .

spherical lens with focal length  $f$ . Indeed, this is the expression for the transmittance of a thin lens [see (2.4-6)].

### EXERCISE 4.1-1

#### The Fresnel Zone Plate

- (a) Use harmonic analysis near the position  $x$  to show that a transparency with complex amplitude transmittance

$$f(x, y) = \begin{cases} 1, & \text{if } \cos\left(\pi \frac{x^2}{\lambda f}\right) > 0 \\ 0, & \text{otherwise} \end{cases}$$

acts as a cylindrical lens with multiple focal lengths.

- (b) A circularly symmetric transparency of complex amplitude transmittance

$$f(x, y) = \begin{cases} 1, & \text{if } \cos\left(\pi \frac{x^2 + y^2}{\lambda f}\right) > 0 \\ 0, & \text{otherwise} \end{cases}$$

is known as a Fresnel zone plate (see Fig. 4.1-8). Show that it acts as a spherical lens with multiple focal lengths.

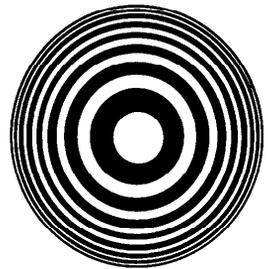
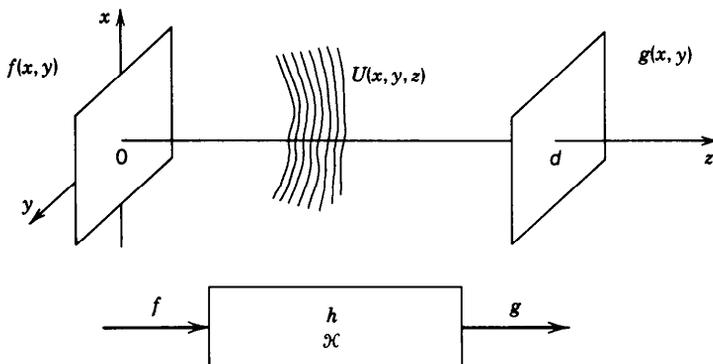


Figure 4.1-8 The Fresnel zone plate.

## B. Transfer Function of Free Space

We now examine the propagation of a monochromatic optical wave of wavelength  $\lambda$  and complex amplitude  $U(x, y, z)$  in the free space between the planes  $z = 0$  and  $z = d$ , called the input and output planes, respectively (see Fig. 4.1-9). Given the complex amplitude of the wave at the input plane,  $f(x, y) = U(x, y, 0)$ , we shall determine the complex amplitude at the output plane,  $g(x, y) = U(x, y, d)$ .

We regard  $f(x, y)$  and  $g(x, y)$  as the input and output of a linear system. The system is linear since the Helmholtz equation, which  $U(x, y, z)$  must satisfy, is linear. The system is shift-invariant because of the invariance of free space to displacement of the coordinate system. A linear shift-invariant system is characterized by its impulse



**Figure 4.1-9** Propagation of light between two planes is regarded as a linear system whose input and output are the complex amplitudes of the wave in the two planes.

response function  $h(x, y)$  or by its transfer function  $\mathcal{H}(\nu_x, \nu_y)$ , as explained in Appendix B, Sec. B.2. We now proceed to determine expressions for these functions.

The transfer function  $\mathcal{H}(\nu_x, \nu_y)$  is the factor by which an input spatial harmonic function of frequencies  $\nu_x$  and  $\nu_y$  is multiplied to yield the output harmonic function. We therefore consider a harmonic input function  $f(x, y) = A \exp[-j2\pi(\nu_x x + \nu_y y)]$ . As explained earlier, this corresponds to a plane wave  $U(x, y, z) = A \exp[-j(k_x x + k_y y + k_z z)]$  where  $k_x = 2\pi\nu_x$ ,  $k_y = 2\pi\nu_y$ , and

$$k_z = (k^2 - k_x^2 - k_y^2)^{1/2} = 2\pi \left( \frac{1}{\lambda^2} - \nu_x^2 - \nu_y^2 \right)^{1/2}. \quad (4.1-5)$$

The output  $g(x, y) = A \exp[-j(k_x x + k_y y + k_z d)]$ , so that we can write  $\mathcal{H}(\nu_x, \nu_y) = g(x, y)/f(x, y) = \exp(-jk_z d)$ , from which

$$\mathcal{H}(\nu_x, \nu_y) = \exp \left[ -j2\pi \left( \frac{1}{\lambda^2} - \nu_x^2 - \nu_y^2 \right)^{1/2} d \right]. \quad (4.1-6)$$

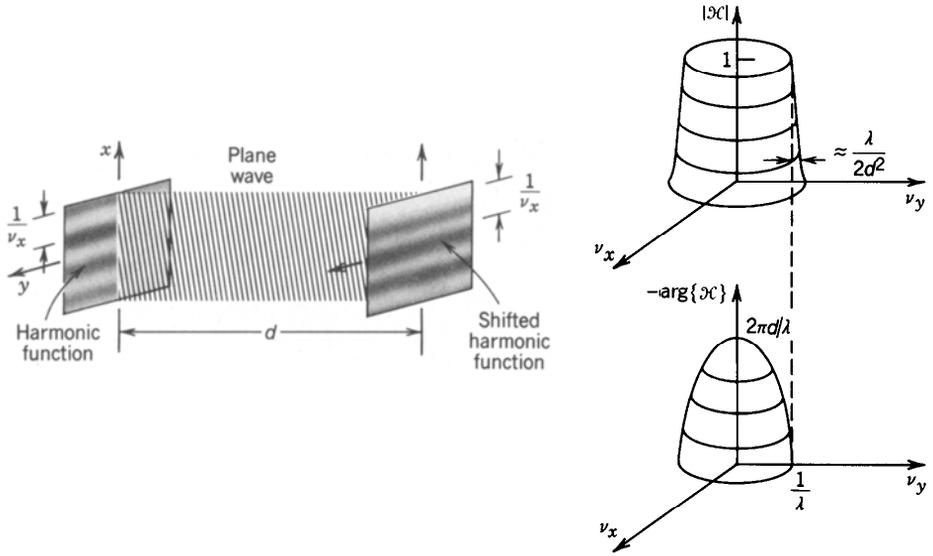
Transfer Function  
of Free Space

The transfer function  $\mathcal{H}(\nu_x, \nu_y)$  is therefore a circularly symmetric complex function of the spatial frequencies  $\nu_x$  and  $\nu_y$ . Its magnitude and phase are sketched in Fig. 4.1-10.

For spatial frequencies for which  $\nu_x^2 + \nu_y^2 \leq 1/\lambda^2$  (i.e., frequencies lying within a circle of radius  $1/\lambda$ ) the magnitude  $|\mathcal{H}(\nu_x, \nu_y)| = 1$  and the phase  $\arg\{\mathcal{H}(\nu_x, \nu_y)\}$  is a function of  $\nu_x$  and  $\nu_y$ . A harmonic function with such frequencies therefore undergoes a spatial phase shift as it propagates, but its magnitude is not altered.

At higher spatial frequencies,  $\nu_x^2 + \nu_y^2 > 1/\lambda^2$ , the quantity under the square root in (4.1-6) is negative so that the exponent is real and the transfer function  $\exp[-2\pi(\nu_x^2 + \nu_y^2 - 1/\lambda^2)^{1/2} d]$  represents an attenuation factor.<sup>†</sup> The wave is then called an **evanescent wave**. When  $\nu_p = (\nu_x^2 + \nu_y^2)^{1/2}$  exceeds  $1/\lambda$  slightly, i.e.,  $\nu_p \approx 1/\lambda$ , the attenuation factor is  $\exp[-2\pi(\nu_p^2 - 1/\lambda^2)^{1/2} d] = \exp[-2\pi(\nu_p - 1/\lambda)^{1/2}(\nu_p + 1/\lambda)^{1/2} d] \approx \exp[-2\pi(\nu_p - 1/\lambda)^{1/2}(2d^2/\lambda)^{1/2}]$ , which equals  $\exp(-2\pi)$  when  $(\nu_p - 1/\lambda) \approx \lambda/2d^2$ , or  $(\nu_p - 1/\lambda)/(1/\lambda) \approx \frac{1}{2}(\lambda/d)^2$ . For  $d \gg \lambda$  the attenuation factor drops sharply when the spatial frequency slightly exceeds  $1/\lambda$ , as illustrated in Fig. 4.1-10.

<sup>†</sup>The  $-$  sign in (4.1-3) was used since the  $+$  sign would have resulted in an exponentially growing function, which is physically unacceptable.



**Figure 4.1-10** Magnitude and phase of the transfer function  $\mathcal{H}(\nu_x, \nu_y)$  for free-space propagation between two planes separated by a distance  $d$ .

We may therefore regard  $1/\lambda$  as the cutoff spatial frequency (the spatial bandwidth) of the system. Thus

*the spatial bandwidth of light propagation  
in free space is approximately  $1/\lambda$  cycles/mm.*

Features contained in spatial frequencies greater than  $1/\lambda$  (corresponding to details of size finer than  $\lambda$ ) cannot be transmitted by an optical wave of wavelength  $\lambda$  over distances much greater than  $\lambda$ .

**Fresnel Approximation**

The expression for the transfer function in (4.1-6) may be simplified if the input function  $f(x, y)$  contains only spatial frequencies that are much smaller than the cutoff frequency  $1/\lambda$ , so that  $\nu_x^2 + \nu_y^2 \ll 1/\lambda^2$ . The plane-wave components of the propagating light then make small angles  $\theta_x \approx \lambda\nu_x$  and  $\theta_y \approx \lambda\nu_y$  corresponding to paraxial rays.

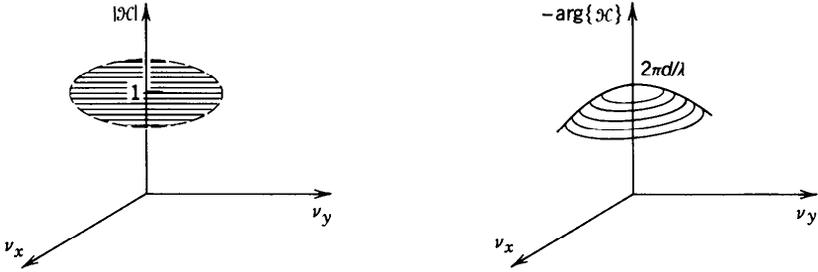
Denoting  $\theta^2 = \theta_x^2 + \theta_y^2 \approx \lambda^2(\nu_x^2 + \nu_y^2)$ , where  $\theta$  is the angle with the optical axis, the phase factor in (4.1-6) is

$$\begin{aligned}
 2\pi \left( \frac{1}{\lambda^2} - \nu_x^2 - \nu_y^2 \right)^{1/2} d &= 2\pi \frac{d}{\lambda} (1 - \theta^2)^{1/2} \\
 &= 2\pi \frac{d}{\lambda} \left( 1 - \frac{\theta^2}{2} + \frac{\theta^4}{8} - \dots \right). \tag{4.1-7}
 \end{aligned}$$

Neglecting the third and higher terms of this expansion, (4.1-6) may be approximated by

$$\mathcal{H}(\nu_x, \nu_y) \approx \mathcal{H}_0 \exp \left[ j\pi\lambda d (\nu_x^2 + \nu_y^2) \right],$$

(4.1-8)  
Transfer Function  
of Free Space  
(Fresnel Approximation)



**Figure 4.1-11** The transfer function of free-space propagation for low spatial frequencies (much less than  $1/\lambda$  cycles/mm) has a constant magnitude and a quadratic phase.

where  $\mathcal{H}_C = \exp(-jk d)$ . In this approximation, the phase is a quadratic function of  $\nu_x$  and  $\nu_y$ , as illustrated in Fig. 4.1-11. This approximation is known as the **Fresnel approximation**.

The condition of validity of the Fresnel approximation is that the third term in (4.1-7) is much smaller than  $\pi$  for all  $\theta$ . This is equivalent to

$$\frac{\theta^4 d}{4\lambda} \ll 1. \quad (4.1-9)$$

If  $a$  is the largest radial distance in the output plane, the largest angle  $\theta_m \approx a/d$ , and (4.1-9) may be written in the form

$$\boxed{\frac{N_F \theta_m^2}{4} \ll 1,} \quad (4.1-10)$$

Condition of Validity of  
Fresnel Approximation

where  $N_F = a^2/\lambda d$  is the Fresnel number. For example, if  $a = 1$  cm,  $d = 100$  cm, and  $\lambda = 0.5 \mu\text{m}$ , then  $\theta_m = 10^{-2}$  radian,  $N_F = 200$ , and  $N_F \theta_m^2/4 = 5 \times 10^{-3}$ . In this case the Fresnel approximation is applicable.

### Input-Output Relation

Given the input function  $f(x, y)$ , the output function  $g(x, y)$  may be determined as follows: (1) We determine the Fourier transform

$$F(\nu_x, \nu_y) = \iint_{-\infty}^{\infty} f(x, y) \exp[j2\pi(\nu_x x + \nu_y y)] dx dy, \quad (4.1-11)$$

which represents the complex envelopes of the plane-wave components in the input plane; (2) the product  $\mathcal{H}(\nu_x, \nu_y)F(\nu_x, \nu_y)$  gives the complex envelopes of the plane-wave components in the output plane; and (3) the complex amplitude in the output plane is the sum of the contributions of these plane waves,

$$g(x, y) = \iint_{-\infty}^{\infty} \mathcal{H}(\nu_x, \nu_y) F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)] d\nu_x d\nu_y.$$

Using the Fresnel approximation for  $\mathcal{H}(\nu_x, \nu_y)$ , which is given by (4.1-8), we have

$$g(x, y) = \mathcal{H}_0 \iint_{-\infty}^{\infty} F(\nu_x, \nu_y) \exp[j\pi\lambda d(\nu_x^2 + \nu_y^2)] \exp[-j2\pi(\nu_x x + \nu_y y)] d\nu_x d\nu_y.$$

(4.1-12)

Equations (4.1-12) and (4.1-11) serve to relate the output function  $g(x, y)$  to the input function  $f(x, y)$ .

### C. Impulse-Response Function of Free Space

The impulse-response function  $h(x, y)$  of the system of free-space propagation is the response  $g(x, y)$  when the input  $f(x, y)$  is a point at the origin  $(0, 0)$ . It is the inverse Fourier transform of the transfer function  $\mathcal{H}(\nu_x, \nu_y)$ . Using Sec. A.3 and Table A.1-1 in Appendix A and  $k = 2\pi/\lambda$ , the inverse Fourier transform of (4.1-8) is

$$h(x, y) \approx h_0 \exp\left[-jk \frac{x^2 + y^2}{2d}\right],$$

(4.1-13)

Impulse-Response  
Function of Free Space  
(Fresnel Approximation)

where  $h_0 = (j/\lambda d) \exp(-jkd)$ . This function is proportional to the complex amplitude at the  $z = d$  plane of a paraboloidal wave centered about the origin  $(0, 0)$  [see (2.2-16)]. Thus each point in the input plane generates a paraboloidal wave; all such waves are superposed at the output plane.

#### Free-Space Propagation as a Convolution

An alternative procedure for relating the complex amplitudes  $f(x, y)$  and  $g(x, y)$  is to regard  $f(x, y)$  as a superposition of different points (delta functions), each producing a paraboloidal wave. The wave originating at the point  $(x', y')$  has an amplitude  $f(x', y')$  and is centered about  $(x', y')$  so that it generates a wave with amplitude  $f(x', y')h(x - x', y - y')$  at the point  $(x, y)$  in the output plane. The sum of these contributions is the two-dimensional convolution

$$g(x, y) = \iint_{-\infty}^{\infty} f(x', y') h(x - x', y - y') dx' dy',$$

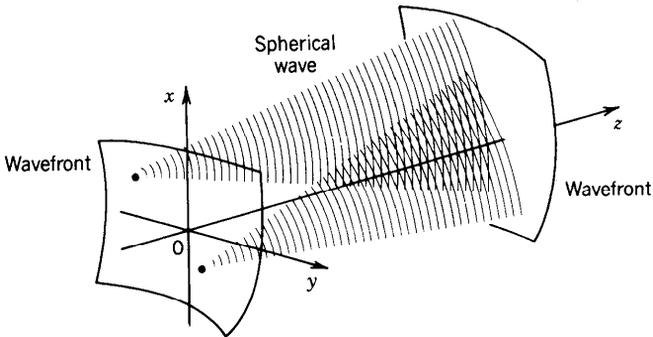
which, in the Fresnel approximation, becomes

$$g(x, y) = h_0 \iint_{-\infty}^{\infty} f(x', y') \exp\left[-j\pi \frac{(x - x')^2 + (y - y')^2}{\lambda d}\right] dx' dy',$$

(4.1-14)

where  $h_0 = (j/\lambda d) \exp(-jkd)$ .

In summary: Within the Fresnel approximation, there are two approaches to determining the complex amplitude  $g(x, y)$  in the output plane, given the complex amplitude  $f(x, y)$  in the input plane: (1) Equation (4.1-14) is based on a space-domain



**Figure 4.1-12** The Huygens–Fresnel principle. Each point on a wavefront generates a spherical wave.

approach in which the input wave is expanded in terms of paraboloidal elementary waves; and (2) Equation (4.1-12) is a frequency-domain approach in which the input wave is expanded as a sum of plane waves.

### EXERCISE 4.1-2

**Gaussian Beams Revisited.** If the function  $f(x, y) = A \exp[-(x^2 + y^2)/W_0^2]$  represents the complex amplitude of an optical wave  $U(x, y, z)$  in the plane  $z = 0$ , show that  $U(x, y, z)$  is the Gaussian beam discussed in Chap. 3, (3.1-7). Use both the space- and frequency-domain methods.

### Huygens – Fresnel Principle

The **Huygens–Fresnel principle** states that each point on a wavefront generates a spherical wave (Fig. 4.1-12). The envelope of these secondary waves constitutes a new wavefront. Their superposition constitutes the wave in another plane. The system's impulse-response function for propagation between the planes  $z = 0$  and  $z = d$  is

$$h(x, y) \propto \frac{1}{r} \exp(-jkr), \quad r = (x^2 + y^2 + d^2)^{1/2}. \quad (4.1-15)$$

In the paraxial approximation, the spherical wave given by (4.1-15) is approximated by the paraboloidal wave in (4.1-13) (see Sec. 2.2B). Our derivation of the impulse response function is therefore consistent with the Huygens–Fresnel principle.

## 4.2 OPTICAL FOURIER TRANSFORM

As has been shown in Sec. 4.1, the propagation of light in free space is described conveniently by Fourier analysis. If the complex amplitude of a monochromatic wave of wavelength  $\lambda$  in the  $z = 0$  plane is a function  $f(x, y)$  composed of harmonic components of different spatial frequencies, each harmonic component corresponds to a plane wave: The plane wave traveling at angles  $\theta_x = \sin^{-1} \lambda \nu_x$ ,  $\theta_y = \sin^{-1} \lambda \nu_y$  corresponds to the components with spatial frequencies  $\nu_x$  and  $\nu_y$  and has an amplitude

$F(\nu_x, \nu_y)$ , the Fourier transform of  $f(x, y)$ . This suggests that light can be used to compute the Fourier transform of a two-dimensional function  $f(x, y)$ , simply by making a transparency with amplitude transmittance  $f(x, y)$  through which a uniform plane wave of unity magnitude is transmitted.

Because each of the plane waves has an infinite extent and therefore overlaps with the other plane waves, however, it is necessary to find a method of separating these waves. It will be shown that at a sufficiently long distance, only a single plane wave contributes to the total amplitude at each point in the output plane, so that the Fourier components are eventually separated naturally. A more practical approach is to use a lens to focus each of the plane waves into a single point.

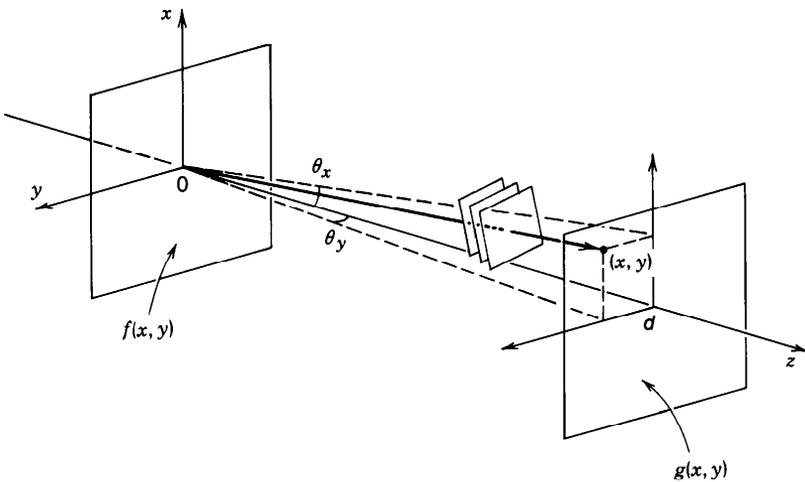
### A. Fourier Transform in the Far Field

We now proceed to show that if the propagation distance  $d$  is sufficiently long, the only plane wave that contributes to the complex amplitude at a point  $(x, y)$  in the output plane is the wave with direction making angles  $\theta_x \approx x/d$  and  $\theta_y \approx y/d$  with the optical axis (see Fig. 4.2-1). This is the wave with wavevector components  $k_x \approx (x/d)k$  and  $k_y \approx (y/d)k$  and amplitude  $F(\nu_x, \nu_y)$  with  $\nu_x = x/\lambda d$ , and  $\nu_y = y/\lambda d$ . The complex amplitudes  $g(x, y)$  and  $f(x, y)$  of the wave at the  $z = d$  and  $z = 0$  planes are related by

$$g(x, y) \approx h_0 F\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right), \tag{4.2-1}$$

Free-Space Propagation  
(Fraunhofer Approximation)

where  $F(\nu_x, \nu_y)$  is the Fourier transform of  $f(x, y)$  and  $h_0 = (j/\lambda d) \exp(-jkd)$ . Contributions of all other waves cancel out as a result of destructive interference. This approximation is known as the **Fraunhofer approximation**. Two proofs of (4.2-1) are provided.



**Figure 4.2-1** When the distance  $d$  is sufficiently long, the complex amplitude at point  $(x, y)$  in the  $z = d$  plane is proportional to the complex amplitude of the plane-wave component with angles  $\theta_x \approx x/d \approx \lambda \nu_x$  and  $\theta_y \approx y/d \approx \lambda \nu_y$ , i.e., to the Fourier transform  $F(\nu_x, \nu_y)$  of  $f(x, y)$ , with  $\nu_x = x/\lambda d$  and  $\nu_y = y/\lambda d$ .

**Proof 1.** We begin with the relation between  $g(x, y)$  and  $f(x, y)$  in (4.1-14). The phase in the argument of the exponent is  $(\pi/\lambda d)[(x - x')^2 + (y - y')^2] = (\pi/\lambda d)[(x^2 + y^2) + (x'^2 + y'^2) - 2(xx' + yy')]$ . If  $f(x, y)$  is confined to a small area of radius  $b$ , and if the distance  $d$  is sufficiently large so that the Fresnel number  $N_F' = b^2/\lambda d$  is small,

$$N_F' \ll 1, \quad (4.2-2)$$

Condition of Validity  
of Fraunhofer Approximation

then the phase factor  $(\pi/\lambda d)(x'^2 + y'^2) \leq \pi(b^2/\lambda d)$  is negligible and (4.1-14) may be approximated by

$$g(x, y) = h_0 \exp\left(-j\pi \frac{x^2 + y^2}{\lambda d}\right) \iint_{-\infty}^{\infty} f(x', y') \exp\left(j2\pi \frac{xx' + yy'}{\lambda d}\right) dx' dy'. \quad (4.2-3)$$

The factors  $x/\lambda d$  and  $y/\lambda d$  may be regarded as the frequencies  $\nu_x = x/\lambda d$  and  $\nu_y = y/\lambda d$ , so that

$$g(x, y) = h_0 \exp\left(-j\pi \frac{x^2 + y^2}{\lambda d}\right) F\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right), \quad (4.2-4)$$

where  $F(\nu_x, \nu_y)$  is the Fourier transform of  $f(x, y)$ . The phase factor given by  $\exp[-j\pi(x^2 + y^2)/\lambda d]$  in (4.2-4) may also be neglected and (4.2-1) obtained if we also limit our interest to points in the output plane within a circle of radius  $a$  centered about the  $z$ -axis so that  $\pi(x^2 + y^2)/\lambda d \leq \pi a^2/\lambda d \ll \pi$ . This is applicable when the Fresnel number  $N_F = a^2/\lambda d \ll 1$ .

The **Fraunhofer approximation** is therefore valid whenever the Fresnel numbers  $N_F$  and  $N_F'$  are small. The Fraunhofer approximation is more difficult to satisfy than the Fresnel approximation, which requires that  $N_F \theta_m^2/4 \ll 1$  [see (4.1-10)]. Since  $\theta_m \ll 1$  in the paraxial approximation, it is possible to satisfy the Fresnel condition  $N_F \theta_m^2/4 \ll 1$  for Fresnel numbers  $N_F$  not necessarily  $\ll 1$ .

## Summary

In the Fraunhofer approximation, the complex amplitude  $g(x, y)$  of a wave of wavelength  $\lambda$  in the  $z = d$  plane is proportional to the Fourier transform  $F(\nu_x, \nu_y)$  of the complex amplitude  $f(x, y)$  in the  $z = 0$  plane, evaluated at the spatial frequencies  $\nu_x = x/\lambda d$  and  $\nu_y = y/\lambda d$ . The approximation is valid if  $f(x, y)$  is confined to a circle of radius  $b$  satisfying,  $b^2/\lambda d \ll 1$ , and at points in the output plane within a circle of radius  $a$  satisfying  $a^2/\lambda d \ll 1$ .

## EXERCISE 4.2-1

**Conditions of Validity of the Fresnel and Fraunhofer Approximations: A Comparison.** Demonstrate that the Fraunhofer approximation is more restrictive than the Fresnel approximation by taking  $\lambda = 0.5 \mu\text{m}$ , assuming that the object points  $(x, y)$  lie within a circle of radius  $b = 1 \text{ cm}$ , and determining the range of distances  $d$  for which the two approximations are applicable.

**\*Proof 2.** The complex amplitude  $g(x, y)$  in (4.1-12) is expressed as an integral of plane waves of different frequencies. If  $d$  is sufficiently large so that the phase in the integrand is much greater than  $2\pi$ , it can be shown using the method of stationary phase<sup>†</sup> that only one value of  $\nu_x$  contributes to the integral. This is the value for which the derivative of the phase  $\pi\lambda d\nu_x^2 - 2\pi\nu_x x$  with respect to  $\nu_x$  vanishes; i.e.,  $\nu_x = x/\lambda d$ . Similarly, the only value of  $\nu_y$  that contributes to the integral is  $\nu_y = y/\lambda d$ . This proves the assertion that only one plane wave contributes to the far field at a given point.

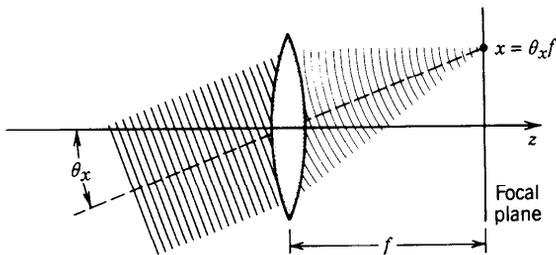
**B. Fourier Transform Using a Lens**

The plane-wave components that constitute a wave may also be separated by use of a lens. A thin spherical lens transforms a plane wave into a paraboloidal wave focused to a point in the lens focal plane (see Sec. 2.4 and Exercise 2.4-3). If the plane wave arrives at small angles  $\theta_x$  and  $\theta_y$ , the paraboloidal wave is centered about the point  $(\theta_x f, \theta_y f)$ , where  $f$  is the focal length (see Fig. 4.2-2). The lens therefore maps each direction  $(\theta_x, \theta_y)$  into a single point  $(\theta_x f, \theta_y f)$  in the focal plane and thus separates the contributions of the different plane waves.

In reference to the optical system shown in Fig. 4.2-3, let  $f(x, y)$  be the complex amplitude of the optical wave in the  $z = 0$  plane. Light is decomposed into plane waves, with the wave traveling at small angles  $\theta_x = \lambda\nu_x$  and  $\theta_y = \lambda\nu_y$  having a complex amplitude proportional to the Fourier transform  $F(\nu_x, \nu_y)$ . This wave is focused by the lens into a point  $(x, y)$  in the focal plane where  $x = \theta_x f = \lambda f\nu_x$ ,  $y = \theta_y f = \lambda f\nu_y$ . The complex amplitude at point  $(x, y)$  in the output plane is therefore proportional to the Fourier transform of  $f(x, y)$  evaluated at  $\nu_x = x/\lambda f$  and  $\nu_y = y/\lambda f$ , so that

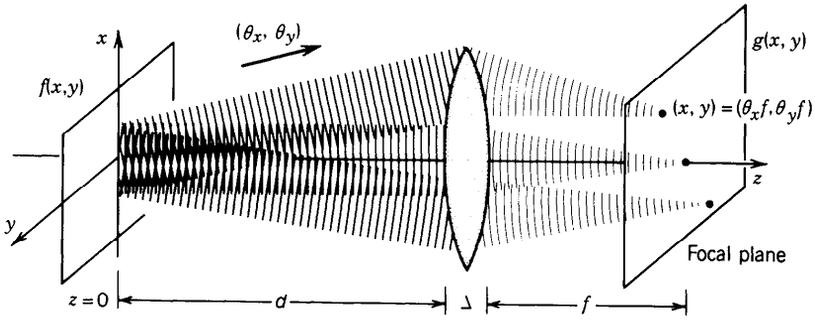
$$g(x, y) \propto F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right). \tag{4.2-5}$$

To determine the proportionality factor in (4.2-5), we analyze the input function  $f(x, y)$  into its Fourier components and trace the plane wave corresponding to each component through the optical system. We then superpose the contributions of these waves at the output plane to obtain  $g(x, y)$ . All these waves will be assumed to be



**Figure 4.2-2** Focusing of a plane wave into a point. A direction  $(\theta_x, \theta_y)$  is mapped into a point  $(x, y) = (\theta_x f, \theta_y f)$ .

<sup>†</sup>See, e.g., Appendix III in M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, New York, 6th ed. 1980.



**Figure 4.2-3** Focusing of the plane waves associated with the harmonic Fourier components of the input function  $f(x, y)$  into points in the focal plane. The amplitude of the plane wave with direction  $(\theta_x, \theta_y) = (\lambda\nu_x, \lambda\nu_y)$  is proportional to the Fourier transform  $F(\nu_x, \nu_y)$  and is focused at the point  $(x, y) = (\theta_x f, \theta_y f) = (\lambda f \nu_x, \lambda f \nu_y)$ .

paraxial and the Fresnel approximation will be used. The procedure takes the following four steps.

1. The plane wave with angles  $\theta_x = \lambda\nu_x$  and  $\theta_y = \lambda\nu_y$  has a complex amplitude  $U(x, y, 0) = F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)]$  in the  $z = 0$  plane and  $U(x, y, d) = \mathcal{H}(\nu_x, \nu_y) F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)]$  in the  $z = d$  plane, immediately before crossing the lens, where  $\mathcal{H}(\nu_x, \nu_y) = \mathcal{H}_0 \exp[j\pi\lambda d(\nu_x^2 + \nu_y^2)]$  is the transfer function of a distance  $d$  of free space and  $\mathcal{H}_0 = \exp(-jk d)$ .
2. Upon crossing the lens, the complex amplitude is multiplied by the lens phase factor  $\exp[j\pi(x^2 + y^2)/\lambda f]$  [the phase factor  $\exp(-jk\Delta)$ , where  $\Delta$  is the width of the lens, has been ignored]. Thus

$$U(x, y, d + \Delta) = \mathcal{H}_0 \exp\left(j\pi \frac{x^2 + y^2}{\lambda f}\right) \times \exp\left[j\pi\lambda d(\nu_x^2 + \nu_y^2)\right] F(\nu_x, \nu_y) \exp[-j2\pi(\nu_x x + \nu_y y)].$$

This expression is simplified by writing  $-2\nu_x x + x^2/\lambda f = (x^2 - 2\nu_x \lambda f x)/\lambda f = [(x - x_0)^2 - x_0^2]/\lambda f$ , with  $x_0 = \lambda\nu_x f$ ; a similar relation for  $y$  is written with  $y_0 = \lambda\nu_y f$ , so that

$$U(x, y, d + \Delta) = A(\nu_x, \nu_y) \exp\left[j\pi \frac{(x - x_0)^2 + (y - y_0)^2}{\lambda f}\right], \quad (4.2-6)$$

where

$$A(\nu_x, \nu_y) = \mathcal{H}_0 \exp\left[j\pi\lambda(d - f)(\nu_x^2 + \nu_y^2)\right] F(\nu_x, \nu_y). \quad (4.2-7)$$

Equation (4.2-6) is recognized as the complex amplitude of a paraboloidal wave converging toward the point  $(x_0, y_0)$  in the lens focal plane,  $z = d + \Delta + f$ .

3. We now examine the propagation in the free space between the lens and the output plane to determine  $U(x, y, d + \Delta + f)$ . We apply (4.1-14) to (4.2-6), use

the relation  $\int \exp[j2\pi(x - x_0)x'/\lambda f] dx' = \lambda f \delta(x - x_0)$ , and obtain

$$U(x, y, d + \Delta + f) = h_0(\lambda f)^2 A(\nu_x, \nu_y) \delta(x - x_0) \delta(y - y_0)$$

where  $h_0 = (j/\lambda f) \exp(-jkf)$ . Indeed, the plane wave is focused into a single point at  $x_0 = \lambda \nu_x f$  and  $y_0 = \lambda \nu_y f$ .

- The last step is to integrate over all the plane waves (all  $\nu_x$  and  $\nu_y$ ). By virtue of the sifting property of the delta function,  $\delta(x - x_0) = \delta(x - \lambda f \nu_x) = (1/\lambda f) \delta(\nu_x - x/\lambda f)$ , this integral gives  $g(x, y) = h_0 A(x/\lambda f, y/\lambda f)$ . Substituting from (4.2-7) we finally obtain

$$g(x, y) = h_l \exp \left[ j\pi \frac{(x^2 + y^2)(d - f)}{\lambda f^2} \right] F \left( \frac{x}{\lambda f}, \frac{y}{\lambda f} \right), \quad (4.2-8)$$

where  $h_l = \partial \mathcal{C}_0 h_0 = (j/\lambda f) \exp[-jk(d + f)]$ . Thus the coefficient of proportionality in (4.2-5) contains a phase factor that is a quadratic function of  $x$  and  $y$ .

Since  $|h_l| = 1/\lambda f$  it follows from (4.2-8) that the optical intensity at the output plane is

$$I(x, y) = \frac{1}{(\lambda f)^2} \left| F \left( \frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \right|^2. \quad (4.2-9)$$

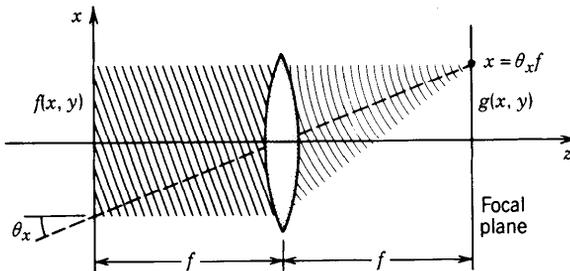
The intensity of light at the output plane (the back focal plane of the lens) is therefore proportional to the squared absolute value of the Fourier transform of the complex amplitude of the wave at the input plane, regardless of the distance  $d$ .

The phase factor in (4.2-8) vanishes if  $d = f$ , so that

$$g(x, y) = h_l F \left( \frac{x}{\lambda f}, \frac{y}{\lambda f} \right), \quad (4.2-10)$$

Fourier Transform  
Property of a Lens

where  $h_l = (j/\lambda f) \exp(-j2kf)$ . This geometry is shown in Fig. 4.2-4.



**Figure 4.2-4** Fourier transform system. The Fourier component of  $f(x, y)$  with spatial frequencies  $\nu_x$  and  $\nu_y$  generates a plane wave at angles  $\theta_x = \lambda \nu_x$  and  $\theta_y = \lambda \nu_y$  and is focused by the lens to the point  $(x, y) = (f\theta_x, f\theta_y) = (\lambda f \nu_x, \lambda f \nu_y)$  so that  $g(x, y)$  is proportional to the Fourier transform  $F(x/\lambda f, y/\lambda f)$ .

In summary: The complex amplitude of light at a point  $(x, y)$  in the back focal plane of a lens of focal length  $f$  is proportional to the Fourier transform of the complex amplitude in the front focal plane evaluated at the frequencies  $\nu_x = x/\lambda f$ ,  $\nu_y = y/\lambda f$ . This relation is valid in the Fresnel approximation. Without the lens, the Fourier transformation is obtained only in the Fraunhofer approximation, which is more restrictive.

### EXERCISE 4.2-2

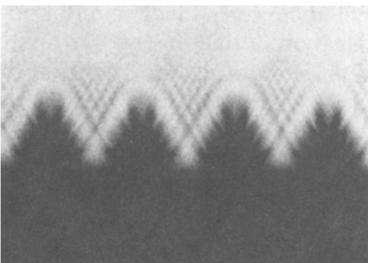
**The Inverse Fourier Transform.** Verify that the optical system in Fig. 4.2-4 performs the inverse Fourier transform operation if the coordinate system in the front focal plane is inverted, i.e.,  $(x, y) \rightarrow (-x, -y)$ .

## 4.3 DIFFRACTION OF LIGHT

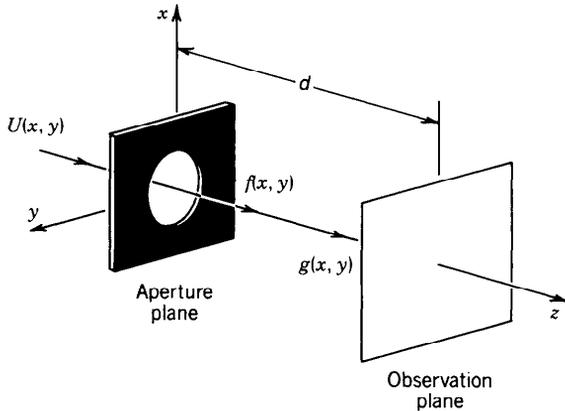
When an optical wave is transmitted through an aperture in an opaque screen and travels some distance in free space, its intensity distribution is called the diffraction pattern. If light were treated as rays, the diffraction pattern would be a shadow of the aperture. Because of the wave nature of light, however, the diffraction pattern may deviate slightly or substantially from the aperture shadow, depending on the distance between the aperture and observation plane, the wavelength, and the dimensions of the aperture. An example is illustrated in Fig. 4.3-1. It is difficult to determine exactly the manner in which the screen modifies the incident wave, but the propagation in free space beyond the aperture is always governed by the laws described earlier in this chapter.

The simplest theory of diffraction is based on the *assumption* that the incident wave is transmitted without change at points within the aperture, but is reduced to zero at points on the back side of the opaque part of the screen. If  $U(x, y)$  and  $f(x, y)$  are the complex amplitudes of the wave immediately to the left and right of the screen (Fig. 4.3-2), then in accordance with this assumption,

$$f(x, y) = U(x, y)p(x, y), \quad (4.3-1)$$



**Figure 4.3-1** Diffraction pattern of the teeth of a saw. (From M. Cagnet, M. Françon, and J. C. Thierré, *Atlas of Optical Phenomena*, Springer-Verlag, Berlin, 1962.)



**Figure 4.3-2** A wave  $U(x, y)$  is transmitted through an aperture of amplitude transmittance  $p(x, y)$ , generating a wave of complex amplitude  $f(x, y) = U(x, y)p(x, y)$ . After propagation a distance  $d$  in free space the complex amplitude is  $g(x, y)$  and the diffraction pattern is the intensity  $I(x, y) = |g(x, y)|^2$ .

where

$$p(x, y) = \begin{cases} 1 & \text{inside the aperture} \\ 0, & \text{outside the aperture} \end{cases} \quad (4.3-2)$$

is called the **aperture function**.

Given  $f(x, y)$ , the complex amplitude  $g(x, y)$  at an observation plane a distance  $d$  from the screen may be determined using the methods described in Secs. 4.1 and 4.2. The diffraction pattern  $I(x, y) = |g(x, y)|^2$  is known as **Fraunhofer diffraction** or **Fresnel diffraction**, depending on whether free-space propagation is described using the Fraunhofer approximation or the Fresnel approximation, respectively.

Although this approach gives reasonably accurate results in most cases, it is not exact. The validity and self-consistency of the assumption that the complex amplitude  $f(x, y)$  vanishes at points outside the aperture on the back of the screen are questionable since the transmitted wave propagates in all directions and reaches those points. A theory of diffraction based on the exact solution of the Helmholtz equation under the boundary conditions imposed by the aperture is mathematically difficult. Only a few geometrical structures have yielded exact solutions. However, different diffraction theories have been developed using a variety of assumptions, leading to results with varying accuracies. Rigorous diffraction theory is beyond the scope of this book.

## A. Fraunhofer Diffraction

Fraunhofer diffraction is the theory of transmission of light through apertures under the assumption that the incident wave is multiplied by the aperture function and using the Fraunhofer approximation to determine the propagation of light in the free space beyond the aperture. The Fraunhofer approximation is valid if the propagation distance  $d$  between the aperture and observation planes is sufficiently large so that the Fresnel number  $N'_F = b^2/\lambda d \ll 1$ , where  $b$  is the largest radial distance within the aperture.

Assuming that the incident wave is a plane wave of intensity  $I_i$  traveling in the  $z$  direction so that  $U(x, y) = I_i^{1/2}$ , then  $f(x, y) = I_i^{1/2}p(x, y)$ . In the Fraunhofer approx-

imation [see (4.2-1)],

$$g(x, y) \approx I_i^{1/2} h_0 P\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right), \quad (4.3-3)$$

where

$$P(\nu_x, \nu_y) = \iint_{-\infty}^{\infty} p(x, y) \exp[j2\pi(\nu_x x + \nu_y y)] dx dy$$

is the Fourier transform of  $p(x, y)$  and  $h_0 = (j/\lambda d) \exp(-jk d)$ . The diffraction pattern is therefore

$$I(x, y) = \frac{I_i}{(\lambda d)^2} \left| P\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right) \right|^2. \quad (4.3-4)$$

In summary: The Fraunhofer diffraction pattern at the point  $(x, y)$  is proportional to the squared magnitude of the Fourier transform of the aperture function  $p(x, y)$  evaluated at the spatial frequencies  $\nu_x = x/\lambda d$  and  $\nu_y = y/\lambda d$ .

### EXERCISE 4.3-1

**Fraunhofer Diffraction from a Rectangular Aperture.** Verify that the Fraunhofer diffraction pattern from a rectangular aperture, of height and width  $D_x$  and  $D_y$  respectively, observed at a distance  $d$  is

$$I(x, y) = I_o \operatorname{sinc}^2 \frac{D_x x}{\lambda d} \operatorname{sinc}^2 \frac{D_y y}{\lambda d}, \quad (4.3-5)$$

where  $I_o = (D_x D_y / \lambda d)^2 I_i$  is the peak intensity and  $\operatorname{sinc}(x) = \sin(\pi x) / (\pi x)$ . Verify that the first zeros of this pattern occur at  $x = \pm \lambda d / D_x$  and  $y = \pm \lambda d / D_y$ , so that the angular divergence of the diffracted light is given by

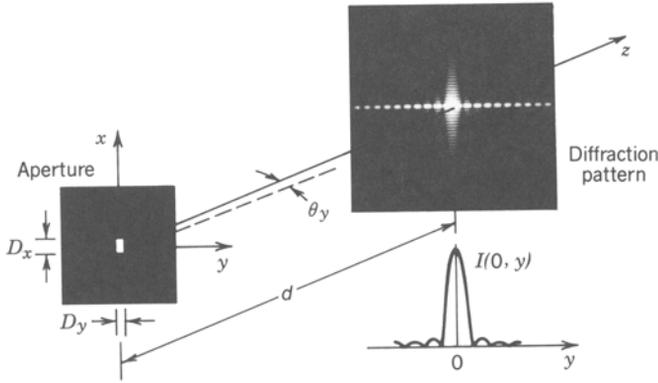
$$\theta_x = \frac{\lambda}{D_x}, \quad \theta_y = \frac{\lambda}{D_y}. \quad (4.3-6)$$

If  $D_y < D_x$ , the diffraction pattern is wider in the  $y$  direction than in the  $x$  direction, as illustrated in Fig. 4.3-3.

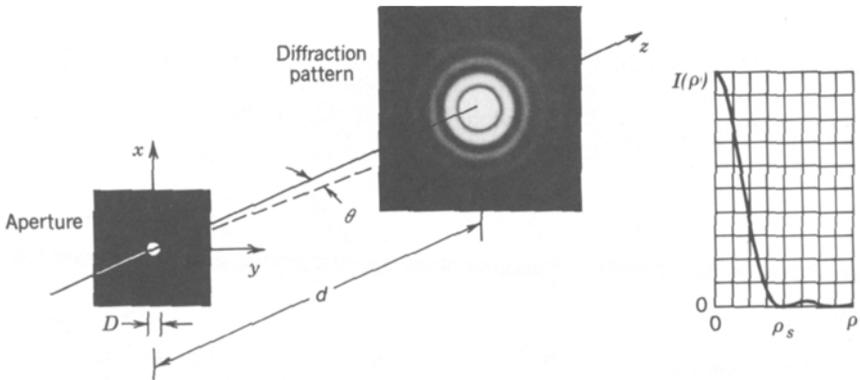
### EXERCISE 4.3-2

**Fraunhofer Diffraction from a Circular Aperture.** Verify that the Fraunhofer diffraction pattern from a circular aperture of diameter  $D$  (Fig. 4.3-4) is

$$I(x, y) = I_o \left[ \frac{2J_1(\pi D \rho / \lambda d)}{\pi D \rho / \lambda d} \right]^2, \quad \rho = (x^2 + y^2)^{1/2}, \quad (4.3-7)$$



**Figure 4.3-3** Fraunhofer diffraction from a rectangular aperture. The central lobe of the pattern has half-angular widths  $\theta_x = \lambda/D_x$  and  $\theta_y = \lambda/D_y$ .



**Figure 4.3-4** The Fraunhofer diffraction pattern from a circular aperture produces the Airy pattern with the radius of the central disk subtending an angle  $\theta = 1.22\lambda/D$ .

where  $I_o = (\pi D^2/4\lambda d)^2 I_i$  is the peak intensity and  $J_1(\cdot)$  is the Bessel function of order 1. The Fourier transform of circularly symmetric functions is discussed in Appendix A, Sec. A.3. The circularly symmetric pattern (4.3-7), known as the **Airy pattern**, consists of a central disk surrounded by rings. Verify that the radius of the central disk, known as the **Airy disk**, is  $\rho_s = 1.22\lambda d/D$  and subtends an angle

$$\theta = 1.22 \frac{\lambda}{D}.$$

(4.3-8)  
Half-Angle Subtended  
by the Airy Disk

The Fraunhofer approximation is valid for distances  $d$  that are usually extremely large. They are satisfied in applications of long-distance free-space optical communication such as laser radar (lidar) and satellite communication. However, as shown in Sec. 4.2B, if a lens of focal length  $f$  is used to focus the diffracted light, the intensity pattern in the focal plane is proportional to the squared magnitude of the Fourier transform of

$p(x, y)$  evaluated at  $v_x = x/\lambda f$  and  $v_y = y/\lambda f$ . The observed pattern is therefore identical to that obtained from (4.3-4), with the distance  $d$  replaced by the focal length  $f$ .

### EXERCISE 4.3-3

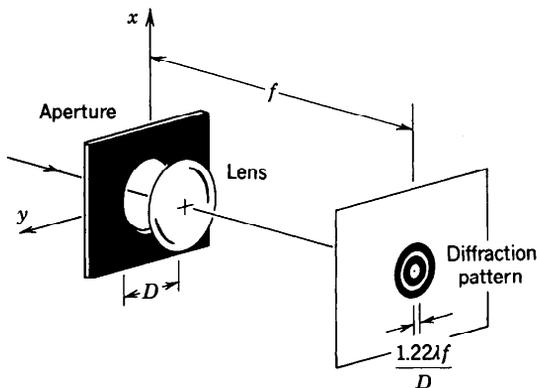
**Spot Size of a Focused Optical Beam.** A beam of light is focused using a lens of focal length  $f$  with a circular aperture of diameter  $D$  (Fig. 4.3-5). If the beam is approximated by a plane wave at points within the aperture, verify that the pattern of the focused spot is

$$I(x, y) = I_o \left[ \frac{2J_1(\pi D \rho / \lambda f)}{\pi D \rho / \lambda f} \right]^2, \quad \rho = (x^2 + y^2)^{1/2}, \quad (4.3-9)$$

where  $I_o$  is the peak intensity. Compare the radius of the focused spot,

$$\rho_s = 1.22\lambda \frac{f}{D}, \quad (4.3-10)$$

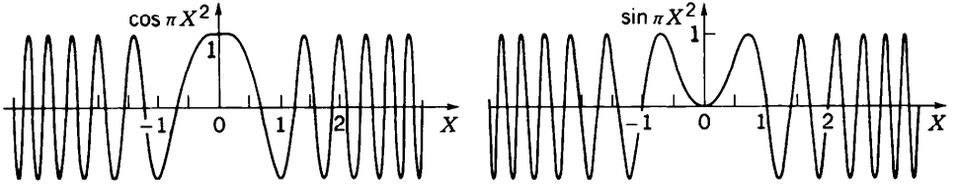
to the spot size obtained when a Gaussian beam of waist radius  $W_0$  is focused by an ideal lens of infinite aperture [see (3.2-15)].



**Figure 4.3-5** Focusing of a plane wave transmitted through a circular aperture of diameter  $D$ .

### \*B. Fresnel Diffraction

The theory of Fresnel diffraction is based on the assumption that the incident wave is multiplied by the aperture function  $p(x, y)$  and propagates in free space in accordance with the Fresnel approximation. If the incident wave is a plane wave traveling in the  $z$ -direction with intensity  $I_i$ , the complex amplitude immediately after the aperture is  $f(x, y) = I_i^{1/2} p(x, y)$ . Using (4.1-14), the diffraction pattern  $I(x, y) = |g(x, y)|^2$  at a



**Figure 4.3-6** The real and imaginary parts of  $\exp(-j\pi X^2)$ .

distance  $d$  is

$$I(x, y) = \frac{I_i}{(\lambda d)^2} \left| \iint_{-\infty}^{\infty} p(x', y') \exp\left[-j\pi \frac{(x-x')^2 + (y-y')^2}{\lambda d}\right] dx' dy' \right|^2. \quad (4.3-11)$$

It is convenient to normalize all distances using  $(\lambda d)^{1/2}$  as a unit of distance, so that  $X = x/(\lambda d)^{1/2}$  and  $X' = x'/(\lambda d)^{1/2}$  are the normalized distances (and similarly for  $y$  and  $y'$ ). Equation (4.3-11) then gives

$$I(X, Y) = I_i \left| \iint_{-\infty}^{\infty} p(X', Y') \exp\{-j\pi[(X-X')^2 + (Y-Y')^2]\} dX' dY' \right|^2. \quad (4.3-12)$$

The integral in (4.3-12) is the convolution of  $p(X, Y)$  and  $\exp[-j\pi(X^2 + Y^2)]$ . The real and imaginary parts of  $\exp(-j\pi X^2)$ ,  $\cos \pi X^2$  and  $\sin \pi X^2$ , are plotted in Fig. 4.3-6. They oscillate at an increasing frequency and their first lobes lie in the intervals  $|X| < 1/\sqrt{2}$  and  $|X| < 1$ , respectively. The total area under the function  $\exp(-j\pi X^2)$  is 1, with the main contribution to the area coming from the first few lobes, since subsequent lobes cancel out. If  $a$  is the radius of the aperture, the radius of the normalized function  $p(X, Y)$  is  $a/(\lambda d)^{1/2}$ . The result of the convolution, which depends on the relative size of the two functions, is therefore governed by the Fresnel number  $N_F = a^2/\lambda d$ .

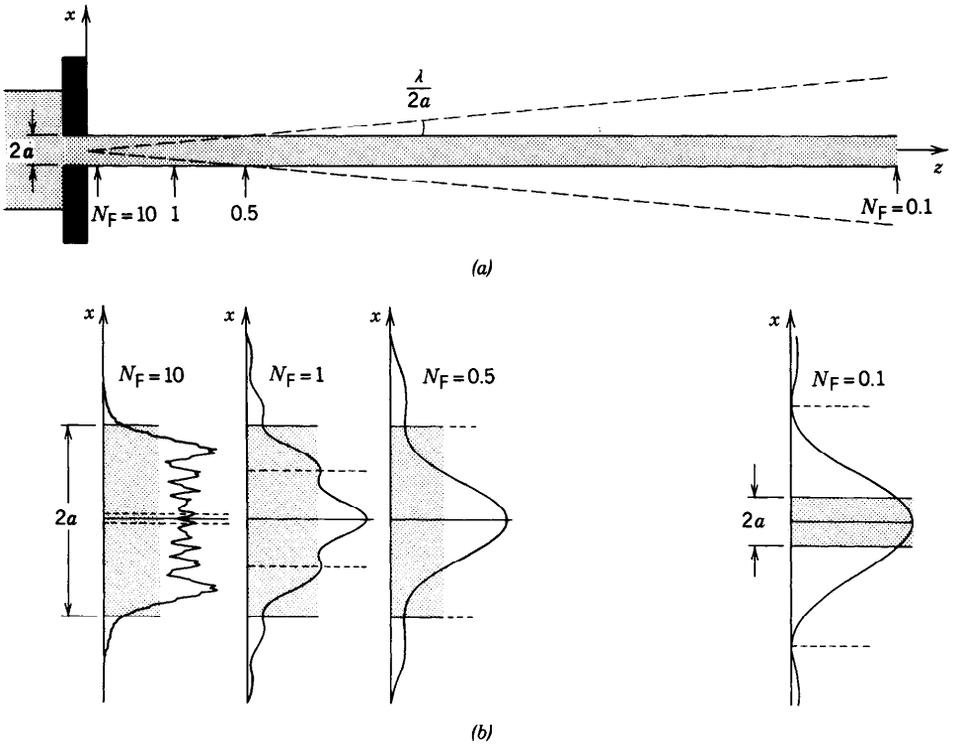
If the Fresnel number is large, the normalized width of the aperture  $a/(\lambda d)^{1/2}$  is much greater than the width of the main lobe, and the convolution yields approximately the wider function  $p(X, Y)$ . Under this condition the Fresnel diffraction pattern is a shadow of the aperture, as would be expected from ray optics. Note that ray optics is applicable in the limit  $\lambda \rightarrow 0$ , which corresponds to the limit  $N_F \rightarrow \infty$ . In the opposite limit, when  $N_F$  is small, the Fraunhofer approximation becomes applicable and the Fraunhofer diffraction pattern is obtained.

**EXAMPLE 4.3-1. Fresnel Diffraction from a Slit.** Assume that the aperture is a slit of width  $D = 2a$ , so that  $p(x, y) = 1$  when  $|x| \leq a$ , and 0 elsewhere. The normalized coordinate is  $X = x/(\lambda d)^{1/2}$  and

$$p(X, Y) = \begin{cases} 1, & |X| \leq \frac{a}{(\lambda d)^{1/2}} = N_F^{1/2} \\ 0, & \text{elsewhere,} \end{cases} \quad (4.3-13)$$

where  $N_F = a^2/\lambda d$  is the Fresnel number. Substituting into (4.3-12), we obtain  $I(X, Y) = I_i |g(X)|^2$ , where

$$g(X) = \int_{-\sqrt{N_F}}^{\sqrt{N_F}} \exp[-j\pi(X-X')^2] dX' = \int_{X-\sqrt{N_F}}^{X+\sqrt{N_F}} \exp(-j\pi X'^2) dX'. \quad (4.3-14)$$



**Figure 4.3-7** Fresnel diffraction from a slit of width  $D = 2a$ . (a) Shaded area is the geometrical shadow of the aperture. The dashed line is the width of the Fraunhofer diffracted beam. (b) Diffraction pattern at four axial positions marked by the arrows in (a) and corresponding to the Fresnel numbers  $N_F = 10, 1, 0.5$ , and  $0.1$ . The shaded area represents the geometrical shadow of the slit. The dashed lines at  $|x| = (\lambda/D)d$  represent the width of the Fraunhofer pattern in the far field. Where the dashed lines coincide with the edges of the geometrical shadow, the Fresnel number  $N_F = a^2/\lambda d = 0.5$ .

This integral is usually written in terms of the Fresnel integrals

$$C(x) = \int_0^x \cos \frac{\pi \alpha^2}{2} d\alpha, \quad S(x) = \int_0^x \sin \frac{\pi \alpha^2}{2} d\alpha,$$

which are available in the standard computer mathematical libraries.

The complex function  $g(X)$  may also be evaluated using Fourier-transform techniques. Since  $g(x)$  is the convolution of a rectangular function of width  $N_F^{1/2}$  and  $\exp(-j\pi X^2)$ , its Fourier transform  $G(\nu_x) \propto \text{sinc}(N_F^{1/2}\nu_x) \exp(j\pi\nu_x^2)$  (see Table A.1-1 in Appendix A). Thus  $g(X)$  may be computed by determining the inverse Fourier transform of  $G(\nu_x)$ . If  $N_F \gg 1$ , the width of  $\text{sinc}(N_F^{1/2}\nu_x)$  is much narrower than the width of the first lobe of  $\exp(j\pi\nu_x^2)$  (see Fig. 4.3-6) so that  $G(\nu_x) \approx \text{sinc}(N_F^{1/2}\nu_x)$  and  $g(X)$  is the rectangular function representing the aperture shadow.

The diffraction pattern from a slit is plotted in Fig. 4.3-7 for different Fresnel numbers corresponding to different distances  $d$  from the aperture. At very small distances (very large  $N_F$ ), the diffraction pattern is a perfect shadow of the slit. As the distance increases ( $N_F$  decreases), the wave nature of light is exhibited in the form of small oscillations around the edges of the aperture (see also the diffraction pattern in Fig. 4.3-1). For very small  $N_F$ , the Fraunhofer pattern described by (4.3-5) is obtained. This is a sinc function with the first zero subtending an angle  $\lambda/D = \lambda/2a$ .

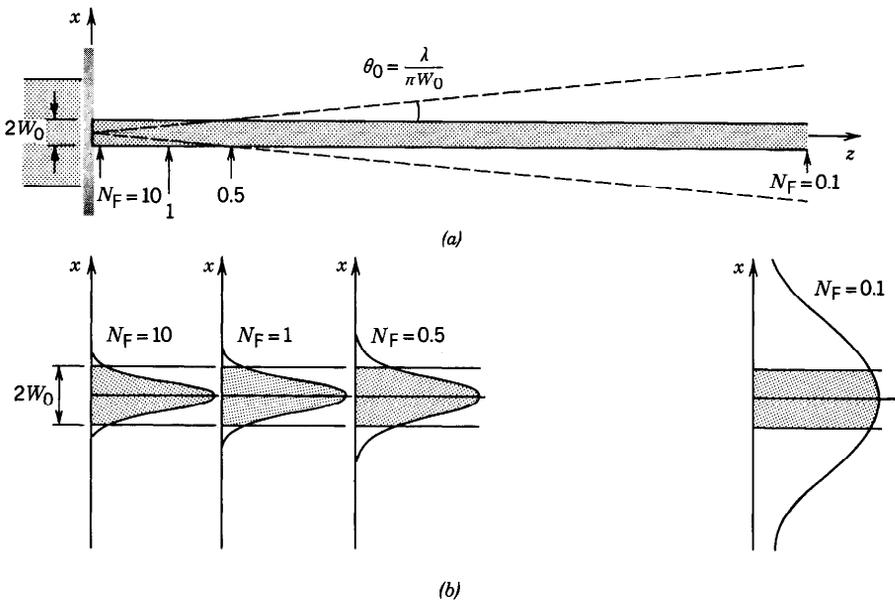
**EXAMPLE 4.3-2. Fresnel Diffraction from a Gaussian Aperture.** If the aperture function  $p(x, y)$  is the Gaussian function  $p(x, y) = \exp[-(x^2 + y^2)/W_0^2]$ , the Fresnel diffraction equation (4.3-11) may be evaluated exactly by finding the convolution of

$\exp[-(x^2 + y^2)/W_0^2]$  with  $h_0 \exp[-j\pi(x^2 + y^2)/\lambda d]$  using, for example, Fourier transform techniques (see Appendix A). The resultant diffraction pattern is

$$I(x, y) = I_i \left[ \frac{W_0}{W(d)} \right]^2 \exp \left[ -2 \frac{x^2 + y^2}{W^2(d)} \right],$$

where  $W^2(d) = W_0^2 + \theta_0^2 d^2$  and  $\theta_0 = \lambda/\pi W_0$ .

The diffraction pattern is a Gaussian function of  $1/e^2$  half-width  $W(d)$ . For small  $d$ ,  $W(d) \approx W_0$ ; but as  $d$  increases,  $W(d)$  increases and approaches  $W(d) \approx \theta_0 d$  when  $d$  is sufficiently large for the Fraunhofer approximation to be applicable, so that the angle subtended by the Fraunhofer diffraction pattern is  $\theta_0$ . These results are illustrated in Fig. 4.3-8, which is analogous to the illustration in Fig. 4.3-7 for diffraction from a slit. The wave diffracted from a Gaussian aperture is the Gaussian beam described in detail in Chap. 3.



**Figure 4.3-8** Fresnel diffraction pattern for a Gaussian aperture of radius  $W_0$  at distances  $d$  such that the parameter  $(\pi/2)W_0^2/\lambda d$ , which is analogous to the Fresnel number  $N_F$  in Fig. 4.3-7, is 10, 1, 0.5, and 0.1. These values correspond to  $W(d)/W_0 = 1.001, 1.118, 1.414, \text{ and } 5.099$ , respectively. The diffraction pattern is Gaussian at all distances.

## Summary

In the order of increasing distance from the aperture, the diffraction pattern is:

1. A shadow of the aperture.
2. A Fresnel diffraction pattern, which is the convolution of the normalized aperture function with  $\exp[-j\pi(X^2 + Y^2)]$ .
3. A Fraunhofer diffraction pattern, which is the squared-absolute value of the Fourier transform of the aperture function. The far field has an angular divergence proportional to  $\lambda/D$ , where  $D$  is the diameter of the aperture.

## 4.4 IMAGE FORMATION

An ideal image formation system is an optical system that replicates the distribution of light in one plane, the object plane, into another, the image plane. Since the optical transmission process is never perfect, the image is never an exact replica of the object. Aside from image magnification, there is also blur resulting from imperfect focusing and from the diffraction of optical waves. This section is devoted to the description of image formation systems and their fidelity. Methods of linear systems, such as the impulse-response function and the transfer function (Appendix B), are used to characterize image formation. A simple ray-optics approach is presented first, then a treatment based on wave optics is subsequently developed.

### A. Ray-Optics Description of Image Formation

Consider an imaging system using a lens of focal length  $f$  at distances  $d_1$  and  $d_2$  from the object and image planes, respectively, as shown in Fig. 4.4-1. When  $1/d_1 + 1/d_2 = 1/f$ , the system is focused so that paraxial rays emitted from each point in the object plane reach a single corresponding point in the image plane. Within the ray theory of light, the imaging is “ideal,” with each point of the object producing a single point of the image. The impulse-response function of the system is an impulse function.

Suppose now that the system is not in focus, as illustrated in Fig. 4.4-2, and assume that the focusing error is

$$\epsilon = \frac{1}{d_2} + \frac{1}{d_1} - \frac{1}{f}. \quad (4.4-1)$$

A point in the object plane generates a patch of light in the image plane that is a shadow of the lens aperture. The distribution of this patch is the system’s impulse-response function. For simplicity, we shall consider an object point lying on the optical axis and determine the distribution of light  $h(x, y)$  it generates in the image plane.

Assume that the plane of the focused image lies at a distance  $d_{2o}$  satisfying the imaging equation  $1/d_{2o} + 1/d_1 = 1/f$ . The shadow of a point on the edge of the aperture at a radial distance  $\rho$  is a point in the image plane with radial distance  $\rho_s$ , where the ratio  $\rho_s/\rho = (d_{2o} - d_2)/d_{2o} = 1 - d_2/d_{2o} = 1 - d_2(1/f - 1/d_1) = 1 - d_2(1/d_2 - \epsilon) = \epsilon d_2$ . If  $p(x, y)$  is the aperture function, also called the **pupil**

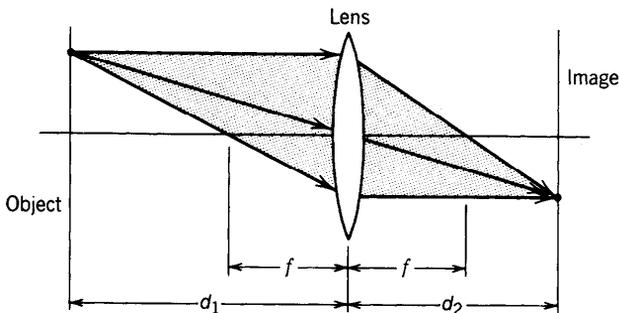
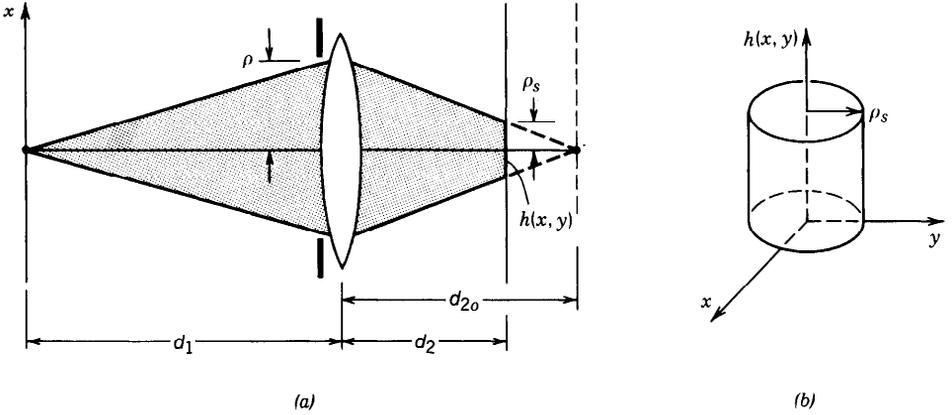


Figure 4.4-1 Rays in a focused imaging system.



**Figure 4.4-2** (a) Rays in a defocused imaging system. (b) The impulse-response function of an imaging system with a circular aperture of diameter  $D$  is a circle of radius  $\rho_s = \epsilon d_2 D/2$ , where  $\epsilon$  is the focusing error.

**function** [ $p(x, y) = 1$  for points inside the aperture, and 0 elsewhere], then  $h(x, y)$  is a scaled version of  $p(x, y)$  magnified by a factor  $\rho_s/\rho = \epsilon d_2$ , so that

$$h(x, y) \propto p\left(\frac{x}{\epsilon d_2}, \frac{y}{\epsilon d_2}\right). \tag{4.4-2}$$

Impulse-Response Function  
of a Defocused System  
(Ray-Optics Theory)

As an example, a circular aperture of diameter  $D$  corresponds to an impulse-response function confined to a circle of radius

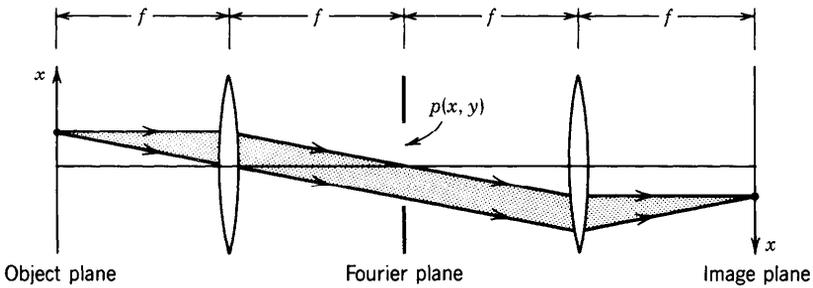
$$\rho_s = \frac{\epsilon d_2 D}{2}, \tag{4.4-3}$$

Radius of Blur Spot

as illustrated in Fig. 4.4-2. The radius  $\rho_s$  of this “blur spot” is an inverse measure of resolving power and image quality. A small value of  $\rho_s$  means that the system is capable of resolving fine details. Since  $\rho_s$  is proportional to the aperture diameter  $D$ , the image quality may be improved by use of a small aperture. A small aperture corresponds to a reduced sensitivity of the system to focusing errors, so that it corresponds to an increased “depth of focus.”

**B. Spatial Filtering**

Consider now the two-lens imaging system illustrated in Fig. 4.4-3. This system, called the **4- $f$  system**, serves as a focused imaging system with unity magnification, as can be easily verified by ray tracing.

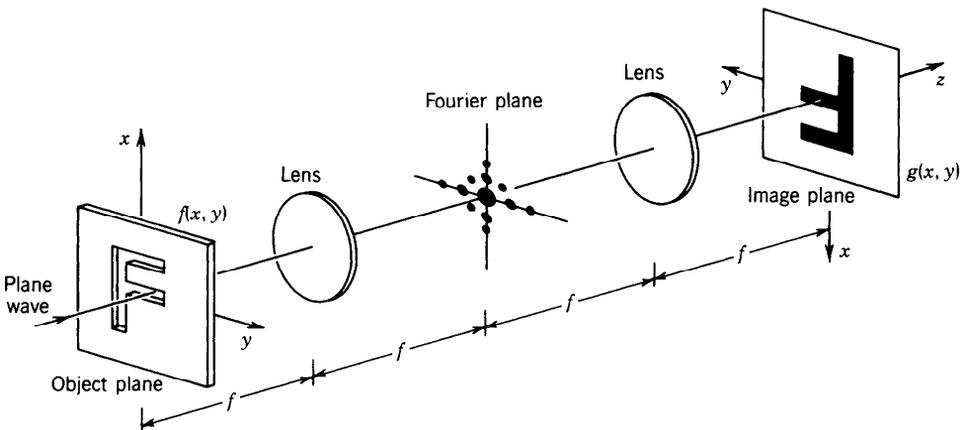


**Figure 4.4-3** The  $4-f$  imaging system. If an inverted coordinate system is used in the image plane, the magnification is unity.

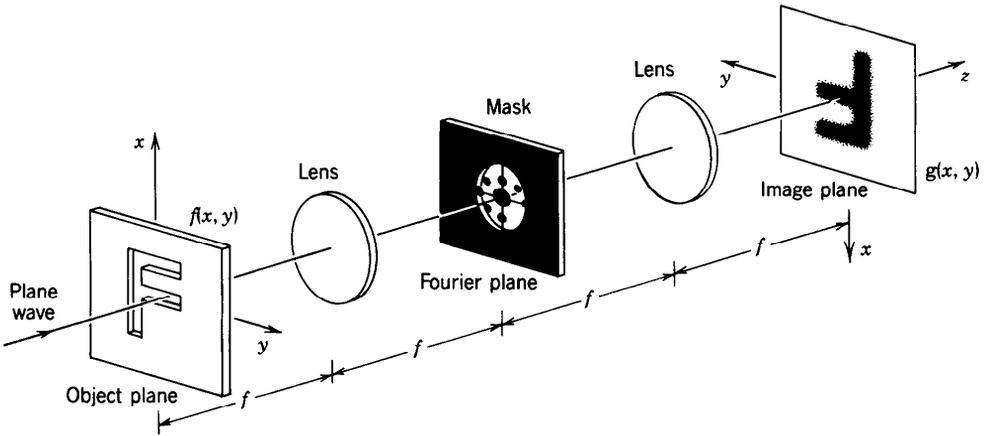
The analysis of wave propagation through this system becomes simple if we recognize it as a cascade of two Fourier-transforming subsystems. The first subsystem (between the object plane and the Fourier plane) performs a Fourier transform, and the second (between the Fourier plane and the image plane) performs an inverse Fourier transform since the coordinate system in the image plane is inverted (see Exercise 4.2-2). As a result, in the absence of an aperture the image is a perfect replica of the object.

Let  $f(x, y)$  be the complex amplitude transmittance of a transparency placed in the object plane and illuminated by a plane wave  $\exp(-jkz)$  traveling in the  $z$  direction, as illustrated in Fig. 4.4-4, and let  $g(x, y)$  be the complex amplitude in the image plane. The first lens system analyzes  $f(x, y)$  into its spatial Fourier transform and separates its Fourier components so that each point in the Fourier plane corresponds to a single spatial frequency. These components are then recombined by the second lens system and the object distribution is perfectly reconstructed.

The  $4-f$  imaging system can be used as a spatial filter in which the image  $g(x, y)$  is a filtered version of the object  $f(x, y)$ . Since the Fourier components of  $f(x, y)$  are available in the Fourier plane, a mask may be used to adjust them selectively, blocking some components and transmitting others, as illustrated in Fig. 4.4-5. The Fourier component of  $f(x, y)$  at the spatial frequency  $(\nu_x, \nu_y)$  is located in the Fourier plane at the point  $x = \lambda f \nu_x, y = \lambda f \nu_y$ . To implement a filter of transfer function  $\mathcal{H}(\nu_x, \nu_y)$ , the



**Figure 4.4-4** The  $4-f$  system performs a Fourier transform followed by an inverse Fourier transform, so that the image is a perfect replica of the object.



**Figure 4.4-5** Spatial filtering. The transparencies in the object and Fourier planes have complex amplitude transmittances  $f(x, y)$  and  $p(x, y)$ . A plane wave traveling in the  $z$  direction is modulated by the object transparency, Fourier transformed by the first lens, multiplied by the transmittance of the mask in the Fourier plane and inverse Fourier transformed by the second lens. As a result, the complex amplitude in the image plane  $g(x, y)$  is a filtered version of  $f(x, y)$ . The system has a transfer function  $\mathcal{H}(\nu_x, \nu_y) = p(\lambda f \nu_x, \lambda f \nu_y)$ .

complex amplitude transmittance  $p(x, y)$  of the mask must be proportional to  $\mathcal{H}(x/\lambda f, y/\lambda f)$ . Thus the transfer function of the filter realized by a mask of transmittance  $p(x, y)$  is

$$\mathcal{H}(\nu_x, \nu_y) = p(\lambda f \nu_x, \lambda f \nu_y),$$

(4.4-4)  
Transfer Function of the 4-f  
Spatial Filter With Mask  
Transmittance  $p(x, y)$

where we have ignored the phase factor  $j \exp(-j2kf)$  associated with each Fourier transform operation [the argument of  $h_l$  in (4.2-10)]. The Fourier transforms  $G(\nu_x, \nu_y)$  and  $F(\nu_x, \nu_y)$  of  $g(x, y)$  and  $f(x, y)$  are related by  $G(\nu_x, \nu_y) = \mathcal{H}(\nu_x, \nu_y)F(\nu_x, \nu_y)$ .

This is a rather simple result. *The transfer function has the same shape as the pupil function.* The corresponding impulse-response function  $h(x, y)$  is the inverse Fourier transform of  $\mathcal{H}(\nu_x, \nu_y)$ ,

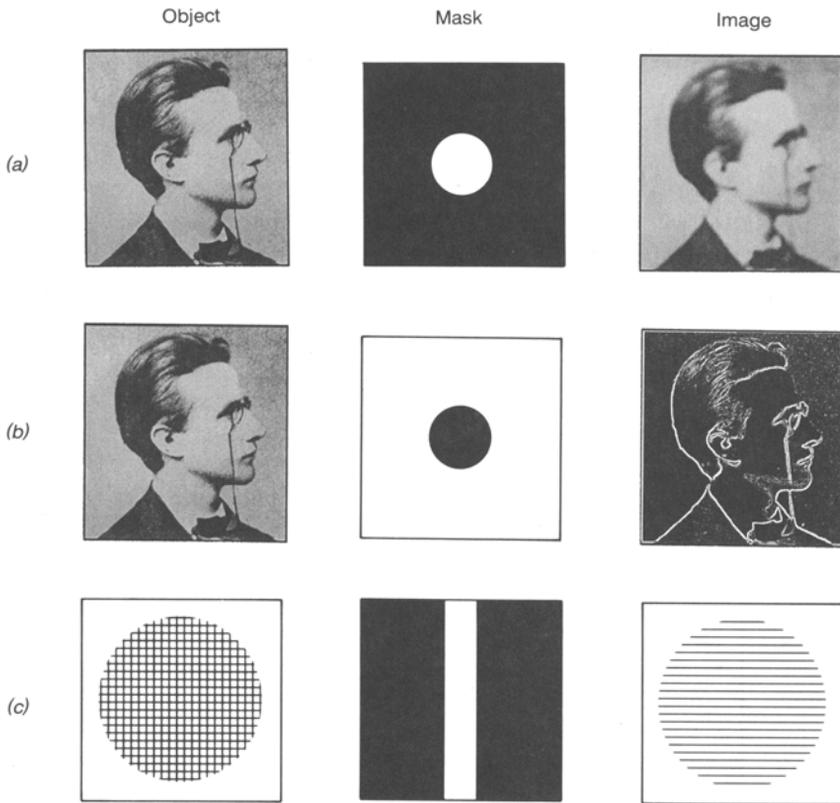
$$h(x, y) = \frac{1}{(\lambda f)^2} P\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right),$$

(4.4-5)

where  $P(\nu_x, \nu_y)$  is the Fourier transform of  $p(x, y)$ .

**Examples of Spatial Filters**

- The ideal circularly symmetric *low-pass filter* has a transfer function  $\mathcal{H}(\nu_x, \nu_y) = 1$ ,  $\nu_x^2 + \nu_y^2 < \nu_s^2$  and  $\mathcal{H}(\nu_x, \nu_y) = 0$ , otherwise. It passes spatial frequencies that are smaller than the cutoff frequency  $\nu_s$  and blocks higher frequencies. This filter is implemented by a mask in the form of a circular aperture of diameter  $D$ , with  $D/2 = \nu_s \lambda f$ . For example, if  $D = 2$  cm,  $\lambda = 1 \mu\text{m}$ , and  $f = 100$  cm, the cutoff



**Figure 4.4-6** Examples of object, mask, and filtered image for three spatial filters: (a) low-pass filter; (b) high-pass filter; (c) vertical-pass filter. Black means the transmittance is zero and white means the transmittance is unity.

frequency (spatial bandwidth)  $\nu_s = D/2\lambda f = 10$  lines/mm. This filter eliminates spatial frequencies that are greater than 10 lines/mm, so that the smallest size of discernible detail in the filtered image is approximately 0.1 mm.

- The *high-pass filter* is the complement of the low-pass filter. It blocks low frequencies and transmits high frequencies. The mask is a clear transparency with an opaque central circle. The filter output is high at regions of large rate of change and small at regions of smooth or slow variation of the object. The filter is therefore useful for edge enhancement in image-processing applications.
- The *vertical-pass filter* blocks horizontal frequencies and transmits vertical frequencies. Only variations in the  $x$  direction are transmitted. If the mask is a vertical slit of width  $D$ , the highest transmitted frequency is  $\nu_y = (D/2)/\lambda f$ .

Examples of these filters and their effects on images are illustrated in Fig. 4.4-6.

### C. Single-Lens Imaging System

We now consider image formation in the single-lens imaging system shown in Fig. 4.4-7 using a wave-optics approach. We first determine the impulse-response function, and then derive the transfer function.

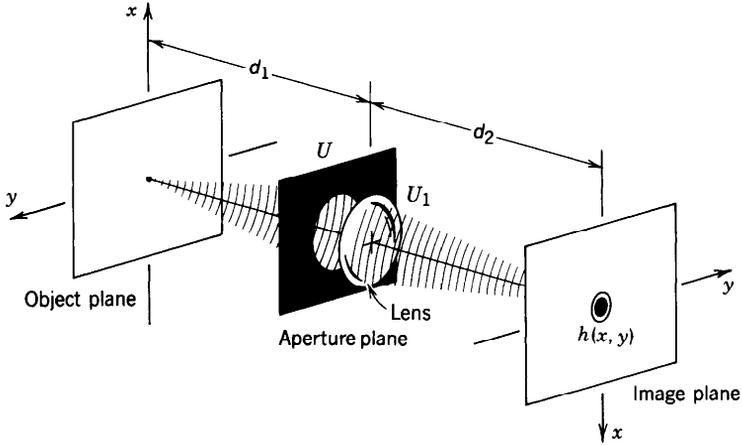


Figure 4.4-7 Single-lens imaging system.

**Impulse-Response Function**

To determine the impulse-response function we consider an object composed of a single point (an impulse) on the optical axis at the point  $(0, 0)$ , and follow the emitted optical wave as it travels to the image plane. The resultant complex amplitude is the impulse-response function  $h(x, y)$ .

An impulse in the object plane produces in the aperture plane a spherical wave approximated by [see (4.1-13)]

$$U(x, y) \approx h_1 \exp \left[ -jk \frac{x^2 + y^2}{2d_1} \right], \tag{4.4-6}$$

where  $h_1 = (j/\lambda d_1) \exp(-jk d_1)$ . Upon crossing the aperture and the lens,  $U(x, y)$  is multiplied by the pupil function  $p(x, y)$  and the lens quadratic phase factor  $\exp[jk(x^2 + y^2)/2f]$ , becoming

$$U_1(x, y) = U(x, y) \exp \left( jk \frac{x^2 + y^2}{2f} \right) p(x, y). \tag{4.4-7}$$

The resultant field  $U_1(x, y)$  then propagates in free space a distance  $d_2$ . In accordance with (4.1-14) it produces the amplitude

$$h(x, y) = h_2 \iint_{-\infty}^{\infty} U_1(x', y') \exp \left[ -j\pi \frac{(x - x')^2 + (y - y')^2}{\lambda d_2} \right] dx' dy', \tag{4.4-8}$$

where  $h_2 = (j/\lambda d_2) \exp(-jk d_2)$ . Substituting from (4.4-6) and (4.4-7) into (4.4-8) and casting the integrals as a Fourier transform, we obtain

$$h(x, y) = h_1 h_2 \exp \left( -j\pi \frac{x^2 + y^2}{\lambda d_2} \right) P_1 \left( \frac{x}{\lambda d_2}, \frac{y}{\lambda d_2} \right), \tag{4.4-9}$$

where  $P_1(\nu_x, \nu_y)$  is the Fourier transform of the function

$$p_1(x, y) = p(x, y) \exp \left( -j\pi \epsilon \frac{x^2 + y^2}{\lambda} \right), \tag{4.4-10}$$

known as the **generalized pupil function**. The factor  $\epsilon$  is the focusing error given by (4.4-1).

For a high-quality imaging system, the impulse-response function is a narrow function, extending only over a small range of values of  $x$  and  $y$ . If the phase factor  $\pi(x^2 + y^2)/\lambda d_2$  in (4.4-9) is much smaller than 1 for all  $x$  and  $y$  within this range, it can be neglected, so that

$$h(x, y) = h_0 P_1\left(\frac{x}{\lambda d_2}, \frac{y}{\lambda d_2}\right), \tag{4.4-11}$$

Impulse-Response  
Function

where  $h_0 = h_1 h_2$  is a constant of magnitude  $(1/\lambda d_1)(1/\lambda d_2)$ . It follows that the system's impulse-response function is proportional to the Fourier transform of the generalized pupil function  $p_1(x, y)$  evaluated at  $\nu_x = x/\lambda d_2$  and  $\nu_y = y/\lambda d_2$ .

If the system is focused ( $\epsilon = 0$ ), then  $p_1(x, y) = p(x, y)$ , and

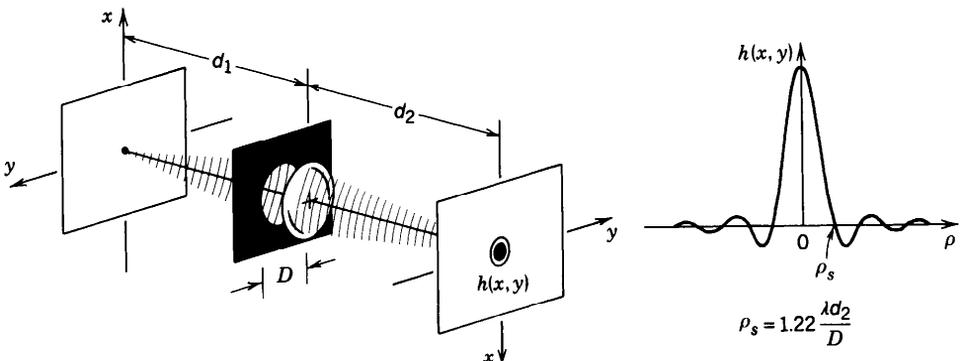
$$h(x, y) \approx h_0 P\left(\frac{x}{\lambda d_2}, \frac{y}{\lambda d_2}\right), \tag{4.4-12}$$

where  $P(\nu_x, \nu_y)$  is the Fourier transform of  $p(x, y)$ . This result is similar to the corresponding result in (4.4-5) for the 4- $f$  system.

**EXAMPLE 4.4-1. Impulse-Response Function of a Focused Imaging System with a Circular Aperture.** If the aperture is a circle of diameter  $D$  so that  $p(x, y) = 1$  if  $\rho = (x^2 + y^2)^{1/2} \leq D/2$ , and zero otherwise, then the impulse-response function is

$$h(x, y) = h(0, 0) \frac{2J_1(\pi D \rho / \lambda d_2)}{\pi D \rho / \lambda d_2}, \quad \rho = (x^2 + y^2)^{1/2}, \tag{4.4-13}$$

and  $|h(0, 0)| = (\pi D^2 / 4 \lambda^2 d_1 d_2)$ . This is a circularly symmetric function whose cross section is shown in Fig. 4.4-8. It drops to zero at a radius  $\rho_s = 1.22 \lambda d_2 / D$  and oscillates slightly before it vanishes. The radius  $\rho_s$  is therefore a measure of the size of the blur circle. If the system is focused at  $\infty$ ,  $d_1 = \infty$ ,  $d_2 = f$ , and  $\rho_s = 1.22 \lambda F_\#$ , where  $F_\# = f/D$  is the lens



**Figure 4.4-8** Impulse-response function of an imaging system with a circular aperture.

*F*-number. Thus systems of smaller  $F_{\#}$  (larger apertures) have better image quality. This assumes, of course, that the larger lens does not introduce geometrical aberrations.

**Transfer Function**

The transfer function of a linear system can only be defined when the system is shift invariant (see Appendix B). Evidently, the single-lens imaging system is not shift invariant since a shift  $\Delta$  of a point in the object plane is accompanied by a *different* shift  $M\Delta$  in the image plane, where  $M = -d_2/d_1$  is the magnification.

The image is different from the object in two ways. First, the image is a magnified replica of the object, i.e., the point  $(x, y)$  of the object is located at a new point  $(Mx, My)$  in the image. Second, every point is smeared into a patch as a result of defocusing or diffraction. We can therefore think of image formation as a cascade of two systems—a system of ideal magnification followed by a system of blur, as depicted in Fig. 4.4-9. By its nature, the magnification system is shift variant. For points near the optical axis, the blur system is approximately shift invariant and therefore can be described by a transfer function.

The transfer function  $\mathcal{H}(\nu_x, \nu_y)$  of the blur system is determined by obtaining the Fourier transform of the impulse-response function  $h(x, y)$  in (4.4-11). The result is

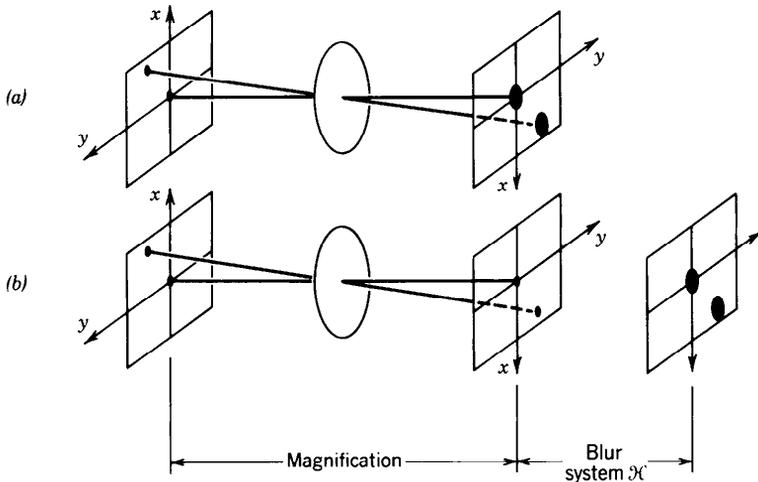
$$\mathcal{H}(\nu_x, \nu_y) \approx p_1(\lambda d_2 \nu_x, \lambda d_2 \nu_y), \tag{4.4-14}$$

Transfer Function

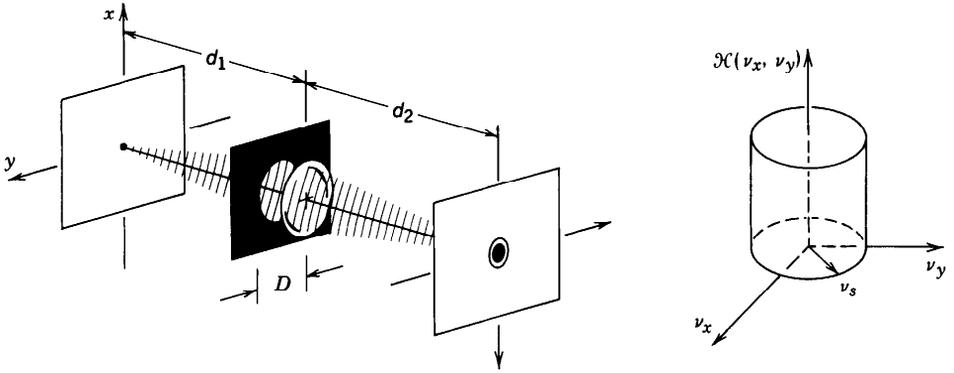
where  $p_1(x, y)$  is the generalized pupil function and we have ignored a constant phase factor  $\exp(-jk d_1) \exp(-jk d_2)$ . If the system is focused, then

$$\mathcal{H}(\nu_x, \nu_y) \approx p(\lambda d_2 \nu_x, \lambda d_2 \nu_y), \tag{4.4-15}$$

where  $p(x, y)$  is the pupil function. This result is identical to that obtained for the 4-*f* imaging system [see (4.4-4)]. If the aperture is a circle of diameter  $D$ , for example, then



**Figure 4.4-9** The imaging system in (a) is regarded in (b) as a combination of an ideal imaging system with only magnification, followed by shift-invariant blur in which each point is blurred into a patch with a distribution equal to the impulse-response function.



**Figure 4.4-10** Transfer function of a focused imaging system with a circular aperture of diameter  $D$ . The system has a spatial bandwidth  $\nu_s = D/2\lambda d_2$ .

the transfer function is constant within a circle of radius  $\nu_s$ , where

$$\nu_s = \frac{D}{2\lambda d_2}, \tag{4.4-16}$$

and vanishes elsewhere, as illustrated in Fig. 4.4-10.

If the lens is focused at infinity, i.e.,  $d_2 = f$ ,

$$\nu_s = \frac{1}{2\lambda F_{\#}},$$

(4.4-17)  
Spatial Bandwidth

where  $F_{\#} = f/D$  is the lens  $F$ -number. For example, for an  $F$ -2 lens ( $F_{\#} = f/D = 2$ ) and for  $\lambda = 0.5 \mu\text{m}$ ,  $\nu_s = 500$  lines/mm. The frequency  $\nu_s$  is the spatial bandwidth, i.e., the highest spatial frequency that the imaging system can transmit.

### 4.5 HOLOGRAPHY

**Holography** involves the recording and reconstruction of optical waves. A **hologram** is a transparency containing a coded record of the optical wave.

Consider a monochromatic optical wave whose complex amplitude in some plane, say the  $z = 0$  plane, is  $U_o(x, y)$ . If, somehow, a thin optical element (call it a transparency) with complex amplitude transmittance  $t(x, y)$  equal to  $U_o(x, y)$  were able to be made, it would provide a complete record of the wave. The wave could then be reconstructed simply by illuminating the transparency with a uniform plane wave of unit amplitude traveling in the  $z$  direction. The transmitted wave would have a complex amplitude in the  $z = 0$  plane  $U(x, y) = 1 \cdot t(x, y) = U_o(x, y)$ . The original wave would then be reproduced at all points in the  $z = 0$  plane, and therefore reconstructed everywhere in the space  $z > 0$ .

As an example, we know that a uniform plane wave traveling at an angle  $\theta$  with respect to the  $z$  axis in the  $x$ - $z$  plane has a complex amplitude  $U_o(x, y) = \exp[-jk \sin \theta x]$ . A record of this wave would be a transparency with complex amplitude transmittance  $t(x, y) = \exp[-jk \sin \theta x]$ . Such a transparency acts as a prism that

bends an incident plane wave  $\exp(-jkz)$  by an angle  $\theta$  (see Sec. 2.4B), thus reproducing the original wave.

The question is how to make a transparency  $t(x, y)$  from the original wave  $U_o(x, y)$ . One key impediment is that optical detectors, including the photographic emulsions used to make transparencies, are responsive to the optical intensity,  $|U_o(x, y)|^2$ , and are therefore insensitive to the phase  $\arg\{U_o(x, y)\}$ . Phase information is obviously important and cannot be disregarded, however. For example, if the phase of the oblique wave  $U_o(x, y) = \exp[-jk \sin\theta x]$  were not recorded, neither would the direction of travel of the wave. To record the phase of  $U_o(x, y)$ , a code must be found that transforms phase into intensity. The recorded information could then be optically decoded in order to reconstruct the wave.

**The Holographic Code**

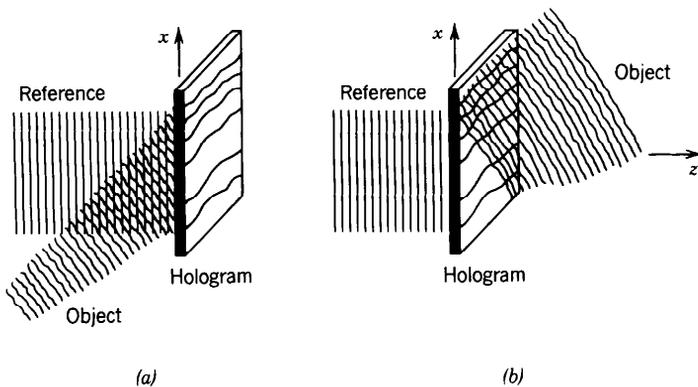
The holographic code is based on mixing the original wave (hereafter called the **object wave**)  $U_o$  with a known **reference wave**  $U_r$  and recording their interference pattern in the  $z = 0$  plane. The intensity of the sum of the two waves is photographically recorded and a transparency of complex amplitude transmittance  $t$ , proportional to the intensity, is made [Fig. 4.5-1(a)]. The transmittance is therefore given by

$$\begin{aligned} t &\propto |U_o + U_r|^2 = |U_r|^2 + |U_o|^2 + U_r^*U_o + U_rU_o^*, \\ &= I_r + I_o + U_r^*U_o + U_rU_o^*, \\ &= I_r + I_o + 2(I_rI_o)^{1/2} \cos[\arg\{U_r\} - \arg\{U_o\}], \end{aligned} \tag{4.5-1}$$

where  $I_r$  and  $I_o$  are, respectively, the intensities of the reference wave and the object wave in the  $z = 0$  plane.

The transparency, called a **hologram**, clearly carries coded information pertinent to the magnitude and phase of the wave  $U_o$ . In fact, as an interference pattern the transmittance  $t$  is highly sensitive to the difference between the phases of the two waves, as was shown in Sec. 2.5.

To decode the information in the hologram and reconstruct the object wave, the reference wave  $U_r$  is again used to illuminate the hologram [Fig. 4.5-1(b)]. The result is



**Figure 4.5-1** (a) A hologram is a transparency on which the interference pattern between the original wave (object wave) and a reference wave is recorded. (b) The original wave is reconstructed by illuminating the hologram with the reference wave.

a wave with complex amplitude

$$U = iU_r \propto U_r I_r + U_r I_o + I_r U_o + U_r^2 U_o^* \quad (4.5-2)$$

in the hologram plane  $z = 0$ . The third term on the right-hand side is the original wave multiplied by the intensity  $I_r$  of the reference wave. If  $I_r$  is uniform (independent of  $x$  and  $y$ ), this term constitutes the desired reconstructed wave. But it must be separated from the other three terms. The fourth term is a conjugated version of the original wave modulated by  $U_r^2$ . The first two terms represent the reference wave, modulated by the sum of the intensities of the two waves.

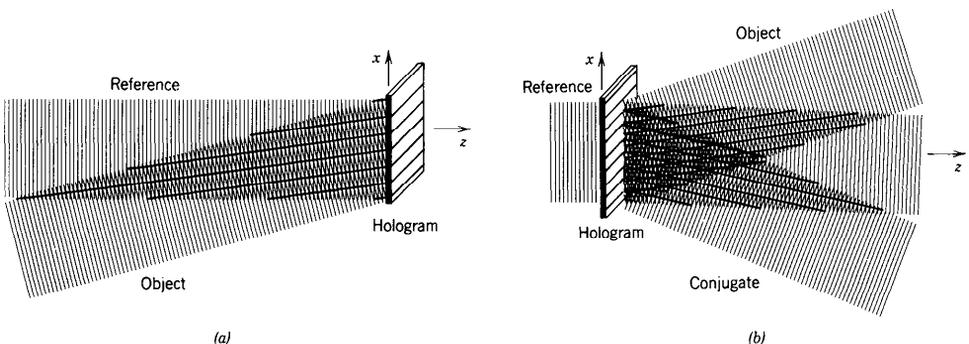
If the reference wave is selected to be a uniform plane wave propagating along the  $z$  axis,  $I_r^{1/2} \exp(-jkz)$ , then in the  $z = 0$  plane  $U_r(x, y) = I_r^{1/2}$  is a constant independent of  $x$  and  $y$ . Dividing (4.5-2) by  $U_r = I_r^{1/2}$  gives

$$U(x, y) \propto I_r + I_o(x, y) + I_r^{1/2} U_o(x, y) + I_r^{1/2} U_o^*(x, y). \quad (4.5-3)$$

Reconstructed Wave  
in Plane of Hologram

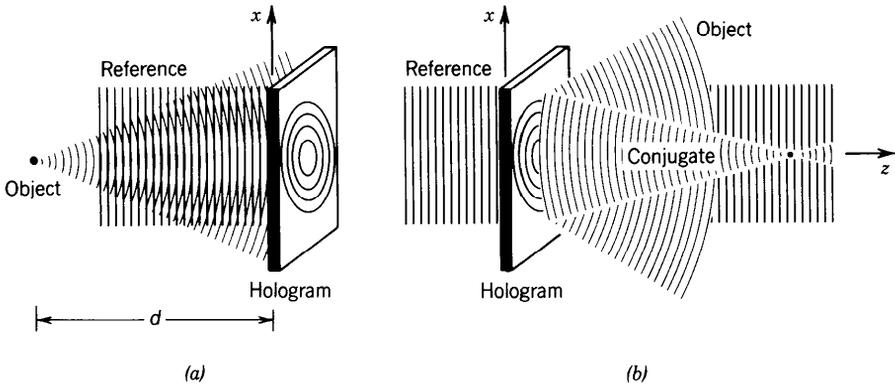
The significance of the various terms in (4.5-3), and the methods of extracting the original wave (the third term), are clarified by means of a number of examples.

**EXAMPLE 4.5-1. Hologram of an Oblique Plane Wave.** If the object wave is an oblique plane wave at angle  $\theta$  [Fig. 4.5-2(a)],  $U_o(x, y) = I_o^{1/2} \exp(-jk \sin \theta x)$ , then (4.5-3) gives  $U(x, y) \propto I_r + I_o + (I_r I_o)^{1/2} \exp(-jk \sin \theta x) + (I_r I_o)^{1/2} \exp(jk \sin \theta x)$ . Since the first two terms are constant, they correspond to a wave propagating in the  $z$  direction (the continuance of the reference wave). The third term corresponds to the original object wave, whereas the fourth term represents the **conjugate wave**, a plane wave traveling at an angle  $-\theta$ . The object wave is therefore separable from the other waves. In fact, this hologram is nothing but a recording of the interference pattern formed from two oblique plane waves at an angle  $\theta$  (Sec. 2.5A). It serves as a sinusoidal diffraction grating that splits an incident reference wave into three waves at angles  $0$ ,  $\theta$ , and  $-\theta$  [see Fig. 4.5-2(b) and Sec. 2.4B].



**Figure 4.5-2** The hologram of an oblique plane wave is a sinusoidal diffraction grating: (a) recording; (b) reconstruction.

**EXAMPLE 4.5-2. Hologram of a Point Source.** Here the object wave is a spherical wave originating at the point  $\mathbf{r}_0 = (0, 0, -d)$ , as illustrated in Fig. 4.5-3, so that  $U_o(x, y) \propto \exp(-jk|\mathbf{r} - \mathbf{r}_0|)/|\mathbf{r} - \mathbf{r}_0|$ , where  $\mathbf{r} = (x, y, 0)$ . The first term of (4.5-3) corresponds to a plane wave traveling in the  $z$  direction, whereas the third is proportional to the amplitude of the original spherical wave originating at  $(0, 0, -d)$ . The fourth term is proportional to the amplitude of the conjugate wave  $U_o^*(x, y) \propto \exp(jk|\mathbf{r} - \mathbf{r}_0|)/|\mathbf{r} - \mathbf{r}_0|$ , which is a converging spherical wave centered at the point  $(0, 0, d)$ . The second term is proportional to  $1/|\mathbf{r} - \mathbf{r}_0|^2$  and its corresponding wave therefore travels in the  $z$  direction with very small angular spread since its intensity varies slowly in the transverse plane.



**Figure 4.5-3** Hologram of a spherical wave originating from a point source: (a) recording; (b) reconstruction. The conjugate wave forms a real image of the point.

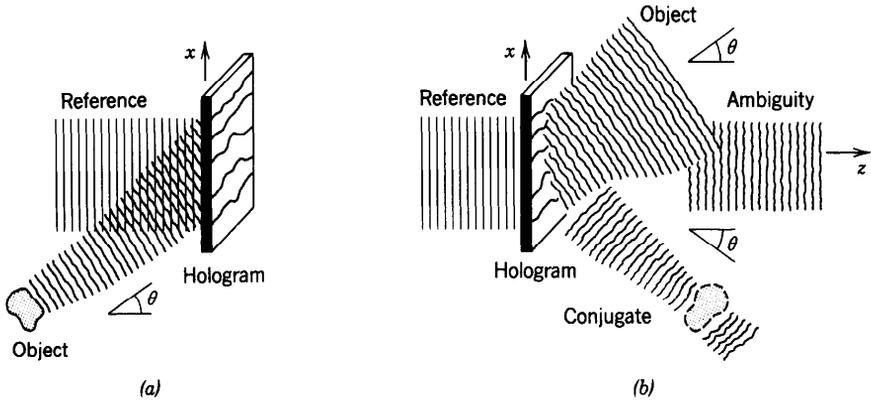
### Off-Axis Holography

One means of separating the four components of the reconstructed wave is to ensure that they vary at well-separated spatial frequencies, so that they have well-separated directions. This form of spatial frequency multiplexing (see Sec. 4.1A) is assured if the object and reference waves are offset so that they arrive from well-separated directions.

Assume that the object wave has a complex amplitude  $U_o(x, y) = f(x, y) \exp(-jk \sin \theta x)$ . This is a wave of complex envelope  $f(x, y)$  modulated by a phase factor equal to that introduced by a prism with deflection angle  $\theta$ . It is assumed that  $f(x, y)$  varies slowly so that its maximum spatial frequency  $\nu_s$  corresponds to an angle  $\theta_s = \sin^{-1} \lambda \nu_s$  much smaller than  $\theta$ . The object wave therefore has directions centered about the angle  $\theta$ , as illustrated in Fig. 4.5-4. Equation (4.5-3) gives

$$U(x, y) \propto I_r + |f(x, y)|^2 + I_r^{1/2} f(x, y) \exp(-jk \sin \theta x) + I_r^{1/2} f^*(x, y) \exp(+jk \sin \theta x).$$

The third term is evidently a replica of the object wave, which arrives from a direction at an angle  $\theta$ . The presence of the phase factor  $\exp(jk \sin \theta x)$  in the fourth term indicates that it is deflected in the  $-\theta$  direction. The first term corresponds to a plane wave traveling in the  $z$  direction. The second term, usually known as the **ambiguity term**, corresponds to a nonuniform plane wave in directions within a cone of small angle  $2\theta_s$  around the  $z$  direction. The offset of the directions of the object and reference waves results in a natural angular separation of the object and conjugate waves from each other and from the other two waves if  $\theta > 3\theta_s$ , thus allowing the original wave to be recovered unambiguously.



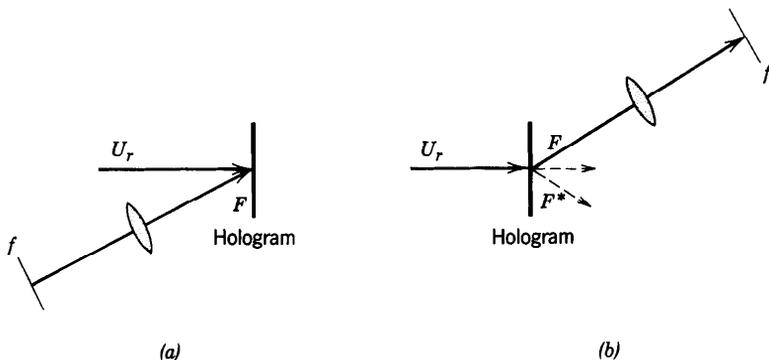
**Figure 4.5-4** Hologram of an off-axis object wave: (a) recording; (b) reconstruction. The object wave is separated from both the reference and conjugate waves.

An alternative method of reducing the effect of the ambiguity wave is to make the intensity of the reference wave much greater than that of the object wave. The ambiguity wave [second term of (4.5-3)] is then much smaller than the other terms since it involves only object waves; it is therefore relatively negligible.

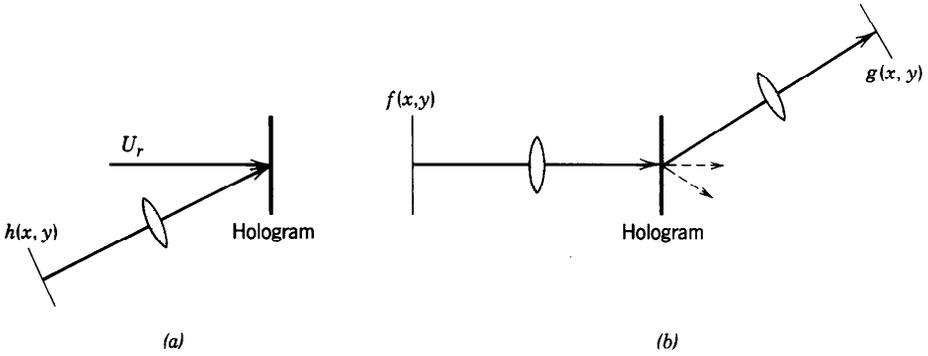
### Fourier-Transform Holography

The Fourier transform  $F(\nu_x, \nu_y)$  of a function  $f(x, y)$  may be computed optically by use of a lens (see Sec. 4.2). If  $f(x, y)$  is the complex amplitude in one focal plane of the lens, then  $F(x/\lambda f, y/\lambda f)$  is the complex amplitude in the other focal plane, where  $f$  is the focal length of the lens and  $\lambda$  is the wavelength. Since the Fourier transform is usually a complex-valued function, it cannot be recorded directly.

The Fourier transform  $F(x/\lambda f, y/\lambda f)$  may be recorded holographically by regarding it as an object wave,  $U_o(x, y) = F(x/\lambda f, y/\lambda f)$ , mixing it with a reference wave  $U_r(x, y)$ , and recording the superposition as a hologram [Fig. 4.5-5(a)]. Reconstruction is achieved by illumination of the hologram with the reference wave as usual. The reconstructed wave may be inverse Fourier transformed using a lens so that the original function  $f(x, y)$  is recovered [Fig. 4.5-5(b)].



**Figure 4.5-5** Hologram of a wave whose complex amplitude represents the Fourier transform of a function  $f(x, y)$ : (a) recording; (b) reconstruction.



**Figure 4.5-6** The Vander Lugt holographic filter. (a) A hologram of the Fourier transform of  $h(x, y)$  is recorded. (b) The Fourier transform of  $f(x, y)$  is transmitted through the hologram and inverse Fourier transformed by a lens. The result is a function  $g(x, y)$  proportional to the convolution of  $f(x, y)$  and  $h(x, y)$ . The overall process provides a spatial filter with impulse-response function  $h(x, y)$ .

**Holographic Spatial Filters**

A spatial filter of transfer function  $\mathcal{H}(\nu_x, \nu_y)$  may be implemented by use of a 4- $f$  optical system with a mask of complex amplitude transmittance  $p(x, y) = \mathcal{H}(x/\lambda f, y/\lambda f)$  placed in the Fourier plane (see Sec. 4.4B). Since the transfer function  $\mathcal{H}(\nu_x, \nu_y)$  is usually complex-valued, the mask transmittance  $p(x, y)$  has a phase component and is difficult to fabricate using conventional printing techniques. If the filter impulse-response function  $h(x, y)$  is real-valued, however, a Fourier-transform hologram of  $h(x, y)$  may be created by holographically recording the Fourier transform  $U_o(x, y) = \mathcal{H}(x/\lambda f, y/\lambda f)$ .

Using the Fourier transform of the input  $f(x, y)$  as a reference,  $U_r(x, y) = F(x/\lambda f, y/\lambda f)$ , the hologram constructs the wave

$$U_r(x, y)U_o(x, y) = F(x/\lambda f, y/\lambda f)\mathcal{H}(x/\lambda f, y/\lambda f).$$

The inverse Fourier transform of the reconstructed object wave, obtained with a lens of focal length  $f$  as illustrated in Fig. 4.5-6(b), therefore yields a complex amplitude  $g(x, y)$  with a Fourier transform  $G(\nu_x, \nu_y) = \mathcal{H}(\nu_x, \nu_y)F(\nu_x, \nu_y)$ . Thus  $g(x, y)$  is the convolution of  $f(x, y)$  with  $h(x, y)$ . The overall system, known as the **Vander Lugt filter**, performs the operation of convolution, which is the basis of spatial filtering.

If the conjugate wave  $U_r(x, y)U_o^*(x, y) = F(x/\lambda f, y/\lambda f)\mathcal{H}^*(x/\lambda f, y/\lambda f)$  is, instead, inverse Fourier transformed, the correlation, instead of the convolution, of the functions  $f(x, y)$  and  $h(x, y)$  is obtained. The operation of correlation is useful in image-processing applications, including pattern recognition.

**The Holographic Apparatus**

An essential condition for the successful fabrication of a hologram is the availability of a monochromatic light source with minimal phase fluctuations. The presence of phase fluctuations results in the random shifting of the interference pattern and the washing out of the hologram. For this reason, a coherent light source (usually a laser) is a necessary part of the apparatus. The coherence requirements for the interference of light waves are discussed in Chap. 10.

Figure 4.5-7 illustrates a typical experimental configuration used to record a hologram and reconstruct the optical wave scattered from the surface of a physical object. Using a beamsplitter, laser light is split into two portions, one is used as the reference wave, whereas the other is scattered from the object to form the object wave. The

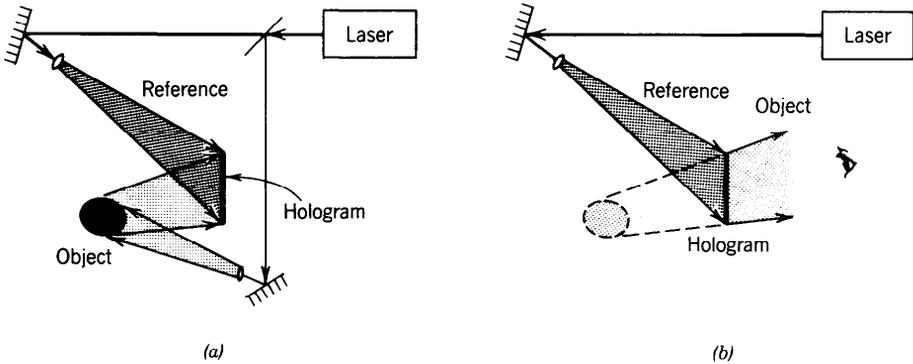


Figure 4.5-7 Holographic recording (a) and reconstruction (b).

optical path difference between the two waves should be as small as possible to ensure that the two beams maintain a nonrandom phase difference [the term  $\arg\{U_r\} - \arg\{U_o\}$  in (4.5-1)].

Since the interference pattern forming the hologram is composed of fine lines separated by distances of the order of  $\lambda / \sin \theta$ , where  $\theta$  is the angular offset between the reference and object waves, the photographic film must be of high resolution and the system must not vibrate during the exposure. The larger  $\theta$ , the smaller the distances between the hologram lines, and the more stringent these requirements are. The object wave is reconstructed when the recorded hologram is illuminated with the reference wave, so that a viewer sees the object as if it were actually there, with its three-dimensional character preserved.

**Volume Holography**

It has been assumed so far that the hologram is a thin planar transparency on which the interference pattern of the object and reference waves is recorded. We now consider recording the hologram in a relatively thick medium and show that this offers an advantage. Consider the simple case when the object and reference waves are plane waves with wavevectors  $\mathbf{k}_r$  and  $\mathbf{k}_o$ . The recording medium extends between the planes  $z = 0$  and  $z = \Delta$ , as illustrated in Fig. 4.5-8. The interference pattern is now a function of  $x$ ,  $y$ , and  $z$ :

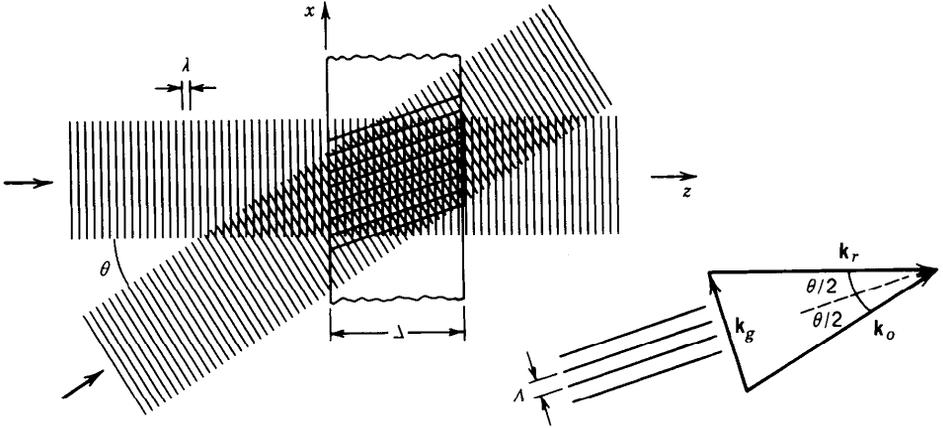
$$\begin{aligned} I(x, y, z) &= |I_r^{1/2} \exp(-j\mathbf{k}_r \cdot \mathbf{r}) + I_o^{1/2} \exp(-j\mathbf{k}_o \cdot \mathbf{r})|^2 \\ &= I_r + I_o + 2(I_r I_o)^{1/2} \cos(\mathbf{k}_o \cdot \mathbf{r} - \mathbf{k}_r \cdot \mathbf{r}) \\ &= I_r + I_o + 2(I_r I_o)^{1/2} \cos(\mathbf{k}_g \cdot \mathbf{r}), \end{aligned}$$

where  $\mathbf{k}_g = \mathbf{k}_o - \mathbf{k}_r$ . This is a sinusoidal pattern of period  $\Lambda = 2\pi/|\mathbf{k}_g|$  and with the surfaces of constant intensity normal to the vector  $\mathbf{k}_g$ .

For example, if the reference wave points in the  $z$  direction and the object wave makes an angle  $\theta$  with the  $z$  axis,  $|\mathbf{k}_g| = 2k \sin(\theta/2)$  and the period is

$$\Lambda = \frac{\lambda}{2 \sin(\theta/2)}, \tag{4.5-4}$$

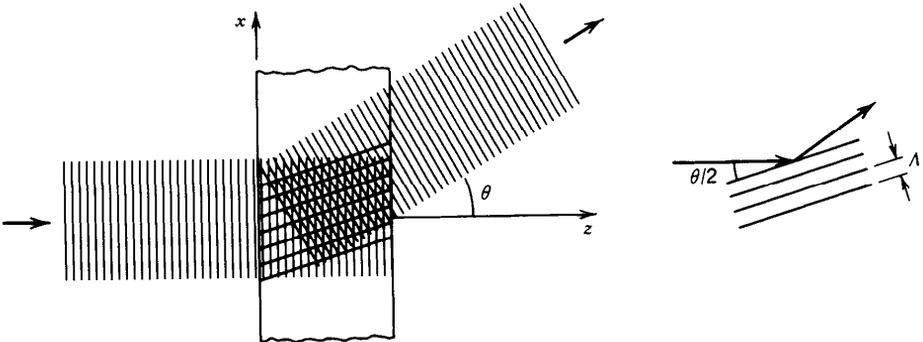
as illustrated in Fig. 4.5-8.



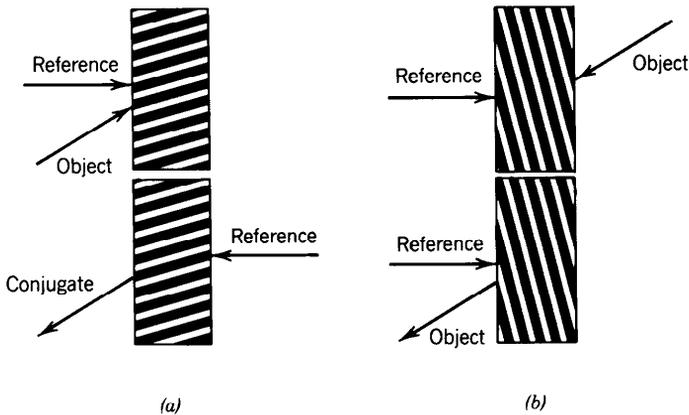
**Figure 4.5-8** Interference pattern when the reference and object waves are plane waves. Since  $|k_r| = |k_o| = 2\pi/\lambda$  and  $|k_g| = 2\pi/\Lambda$ , from the geometry of the vector diagram  $2\pi/\Lambda = 2(2\pi/\lambda) \sin(\theta/2)$ , so that  $\Lambda = \lambda/2 \sin(\theta/2)$ .

If recorded in an emulsion, this pattern serves as a thick diffraction grating, a **volume hologram**. The vector  $k_g$  is called the **grating vector**. When illuminated with the reference wave as illustrated in Fig. 4.5-9, the parallel planes of the grating reflect the wave only when the Bragg condition  $\sin \phi = \lambda/2\Lambda$  is satisfied, where  $\phi$  is the angle between the planes of the grating and the incident reference wave (see Exercise 2.5-3). In our case  $\phi = \theta/2$ , so that  $\sin(\theta/2) = \lambda/2\Lambda$ . In view of (4.5-4), the Bragg condition is indeed satisfied, so that the reference wave is indeed reflected. As evident from the geometry, the reflected wave is an extension of the object wave, so that the reconstruction process is successful.

Suppose now that the hologram is illuminated with a reference wave of different wavelength  $\lambda'$ . Evidently, the Bragg condition,  $\sin(\theta/2) = \lambda'/2\Lambda$ , will not be satisfied and the wave will not be reflected. It follows that the object wave is reconstructed only if the wavelength of the reconstruction source matches that of the recording source. If light with a broad spectrum (white light) is used as a reconstruction source, only the “correct” wavelength would be reflected and the reconstruction process would be successful.



**Figure 4.5-9** The reference wave is Bragg reflected from the thick hologram and the object wave is reconstructed.



**Figure 4.5-10** Two geometries for recording and reconstruction of a volume hologram. (a) This hologram is reconstructed by use of a reversed reference wave; the reconstructed wave is a conjugate wave traveling in a direction opposite to the original object wave. (b) A reflection hologram is recorded with the reference and object waves arriving from opposite sides; the object wave is reconstructed by reflection from the grating.

Although the recording process must be done with monochromatic light, the reconstruction can be achieved with white light. This provides a clear advantage in many applications of holography. Other geometries for recording and reconstruction of a volume hologram are illustrated in Fig. 4.5-10.

Another type of hologram that may be viewed with white light is the **rainbow hologram**. This hologram is recorded through a narrow slit so that the reconstructed image, of course, also appears as if seen through a slit. However, if the wavelength of reconstruction differs from the recording wavelength, the reconstructed wave will appear to be coming from a displaced slit since a magnification effect will be introduced. If white light is used for reconstruction, the reconstructed wave appears as the object seen through many displaced slits, each with a different wavelength (color). The result is a rainbow of images seen through parallel slits. Each slit displays the object with parallax effect in the direction of the slit, but not in the orthogonal direction. Rainbow holograms have many commercial uses as displays.

## READING LIST

### *Fourier Optics and Optical Signal Processing*

- G. Reynolds, J. B. DeVelis, G. B. Parrent, and B. J. Thompson, *The New Physical Optics Notebook: Tutorials in Fourier Optics*, SPIE—The International Society for Optical Engineering, Bellingham, WA, and American Institute of Physics, New York, 1989.
- J. L. Horner, ed., *Optical Signal Processing*, Academic Press, San Diego, CA, 1987.
- F. T. S. Yu, *White-Light Optical Signal Processing*, Wiley, New York, 1985.
- E. G. Steward, *Fourier Optics: An Introduction*, Halsted Press, New York, 1983.
- P. M. Duffieux, *Fourier Transform and Its Applications to Optics*, Wiley, New York, 2nd ed. 1983.
- F. T. S. Yu, *Optical Information Processing*, Wiley, New York, 1983.
- H. Stark, ed., *Applications of Optical Fourier Transforms*, Academic Press, New York, 1982.
- S. H. Lee, ed., *Optical Information Processing Fundamentals*, Springer-Verlag, New York, 1981.
- J. D. Gaskill, *Linear Systems, Fourier Transforms and Optics*, Wiley, New York, 1978.
- F. P. Carlson, *Introduction to Applied Optics for Engineers*, Academic Press, New York, 1978.

- D. Casasent, ed., *Optical Data Processing: Applications*, Springer-Verlag, New York, 1978.
- W. E. Kock, G. W. Stroke, and Yu. E. Nesterikhin, *Optical Information Processing*, Plenum Press, New York, 1976.
- G. Harburn, C. A. Taylor, and T. R. Welberry, *Atlas of Optical Transforms*, Cornell University Press, Ithaca, NY, 1975.
- T. Cathey, *Optical Information Processing and Holography*, Wiley, New York, 1974.
- H. S. Lipson, ed., *Optical Transforms*, Academic Press, New York, 1972.
- M. Cagnet, M. Françon, and S. Mallick, *Atlas of Optical Phenomena*, Springer-Verlag, New York, 1971.
- A. R. Shulman, *Optical Data Processing*, Wiley, New York, 1970.
- J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
- A. Papoulis, *Systems and Transforms with Applications in Optics*, McGraw-Hill, New York, 1968.
- G. W. Stroke, *An Introduction to Coherent Optics and Holography*, Academic Press, New York, 1966.
- L. Mertz, *Transformations in Optics*, Wiley, New York, 1965.
- C. A. Taylor and H. Lipson, *Optical Transforms*, Cornell University Press, Ithaca, NY, 1964.
- E. L. O'Neill, *Introduction to Statistical Optics*, Addison-Wesley, Reading, MA, 1963.

### **Diffraction**

- S. Solimeno, B. Crosignani, and P. DiPorto, *Guiding, Diffraction, and Confinement of Optical Radiation*, Academic Press, New York, 1986.
- J. M. Cowley, *Diffraction Physics*, North-Holland, New York, 1981, 3rd ed. 1984.
- M. Françon, *Diffraction: Coherence in Optics*, Pergamon Press, New York, 1966.

### **Image Formation**

- C. S. Williams and O. A. Becklund, *Introduction to the Optical Transfer Function*, Wiley, New York, 1989.
- M. Françon, *Optical Image Formation and Processing*, Academic Press, New York, 1979.
- J. C. Dainty and R. Shaw, *Image Science*, Academic Press, New York, 1974.
- K. R. Barnes, *The Optical Transfer Function*, Elsevier, New York, 1971.
- E. H. Linfoot, *Fourier Methods in Optical Image Evaluation*, Focal Press, New York, 1964.

### **Holography**

- G. Saxby, *Practical Holography*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- J. E. Kasper, *Complete Book of Holograms: How They Work and How to Make Them*, Wiley, New York, 1987.
- W. Schumann, J.-P. Zurcher, and D. Cuche, *Holography and Deformation Analysis*, Springer-Verlag, New York, 1985.
- N. Abramson, *The Making and Evaluation of Holograms*, Academic Press, New York, 1981.
- Y. I. Ostrovsky, M. M. Butusov, and G. V. Ostrovskaya, *Interferometry by Holography*, Springer-Verlag, New York, 1980.
- H. J. Caulfield, ed., *Handbook of Optical Holography*, Academic Press, New York, 1979.
- W. Schumann and M. Dubas, *Holographic Interferometry*, Springer-Verlag, New York, 1979.
- C. M. Vest, *Holographic Interferometry*, Wiley, New York, 1979.
- G. Bally, ed., *Holography in Medicine and Biology*, Springer-Verlag, New York, 1979.
- L. M. Soroko, *Holography and Coherent Optics*, Plenum Press, New York, 1978.
- R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic Press, New York, 1971, paperback edition 1977.
- H. M. Smith, *Principles of Holography*, Wiley, New York, 1969, 2nd ed. 1975.
- M. Françon, *Holography*, Academic Press, New York, 1974.
- H. J. Caulfield and L. Sun, *The Applications of Holography*, Wiley-Interscience, New York, 1970.

J. B. DeVelis and G. O. Reynolds, *Theory and Applications of Holography*, Addison-Wesley, Reading, MA, 1967.

## PROBLEMS

- 4.1-1 **Correspondence Between Harmonic Functions and Plane Waves.** The complex amplitudes of a monochromatic wave of wavelength  $\lambda$  in the  $z = 0$  and  $z = d$  planes are  $f(x, y)$  and  $g(x, y)$ , respectively. Assuming that  $d = 10^4\lambda$ , use harmonic analysis to determine  $g(x, y)$  in the following cases:
- (a)  $f(x, y) = 1$ ;
  - (b)  $f(x, y) = \exp[(-j\pi/\lambda)(x + y)]$ ;
  - (c)  $f(x, y) = \cos(\pi x/2\lambda)$ ;
  - (d)  $f(x, y) = \cos^2(\pi y/2\lambda)$ ;
  - (e)  $f(x, y) = \sum_m \text{rect}[(x/10\lambda) - 2m]$ ,  $m = 0, \pm 1, \pm 2, \dots$ , where  $\text{rect}(x) = 1$  if  $|x| \leq \frac{1}{2}$  and 0, otherwise.
- Describe the physical nature of the wave in each case.
- 4.1-2 In Problem 4.1-1, if  $f(x, y)$  is a circularly symmetric function with a maximum spatial frequency of 200 lines/mm, determine the angle of the cone within which the wave directions are confined. Assume that  $\lambda = 633$  nm.
- 4.1-3 **Logarithmic Interconnection Map.** A transparency of amplitude transmittance  $t(x, y) = \exp[-j2\pi\phi(x)]$  is illuminated with a uniform plane wave of wavelength  $\lambda = 1 \mu\text{m}$ . The transmitted light is focused by an adjacent lens of focal length  $f = 100$  cm. What must  $\phi(x)$  be so that the ray that hits the transparency at position  $x$  is deflected and focused to a position  $x' = \ln(x)$  for all  $x > 0$ ? (Note that  $x$  and  $x'$  are measured in millimeters.) If the lens is removed, how should  $\phi(x)$  be modified so that the system performs the same function? This system may be used to perform a logarithmic coordinate transformation, as discussed in Chap. 21.
- 4.2-1 **Proof of the Lens Fourier-Transform Property.** (a) Show that the convolution of  $f(x)$  and  $\exp(-j\pi x^2/\lambda d)$  may be obtained in three steps: Multiply  $f(x)$  by  $\exp(-j\pi x^2/\lambda d)$ ; evaluate the Fourier transform of the product at the frequency  $\nu_x = x/\lambda d$ ; and multiply the result by  $\exp(-j\pi x^2/\lambda d)$ . (b) The Fourier transform system in Fig. 4.2-4 is a cascade of three systems—propagation a distance  $f$  in free space, transmission through a lens of focal length  $f$ , and propagation a distance  $f$  in free space. Noting that propagation a distance  $d$  in free space is equivalent to convolution with  $\exp(-j\pi x^2/\lambda d)$  [see (4.1-14)], and using the result in (a), derive the lens' Fourier transform equation (4.2-10). For simplicity ignore the  $y$  dependence.
- 4.2-2 **Fourier Transform of Line Functions.** A transparency of amplitude transmittance  $t(x, y)$  is illuminated with a plane wave of wavelength  $\lambda = 1 \mu\text{m}$  and focused with a lens of focal length  $f = 100$  cm. Sketch the intensity distribution in the plane of the transparency and in the lens focal plane in the following cases (all distances are measured in mm):
- (a)  $t(x, y) = \delta(x - y)$ ;
  - (b)  $t(x, y) = \delta(x + a) + \delta(x - a)$ ,  $a = 1$  mm;
  - (c)  $t(x, y) = \delta(x + a) + j\delta(x - a)$ ,  $a = 1$  mm,
- where  $\delta(\cdot)$  is the delta function (see Appendix A, Sec. A.1).
- 4.2-3 **Design of an Optical Fourier-Transform System.** A lens is used to display the Fourier transform of a two-dimensional function with spatial frequencies between 20 and 200 lines/mm. If the wavelength of light is  $\lambda = 488$  nm, what should be the

focal length of the lens so that the highest and lowest spatial frequencies are separated by a distance of 9 cm in the Fourier plane?

- 4.3-1 **Fraunhofer Diffraction from a Diffraction Grating.** Derive an expression for the Fraunhofer diffraction pattern for an aperture made of  $M = 2L + 1$  parallel slits of infinitesimal widths separated by equal distances  $a = 10\lambda$ ,

$$p(x, y) = \sum_{m=-L}^L \delta(x - ma).$$

Sketch the pattern as a function of the observation angle  $\theta = x/d$ , where  $d$  is the observation distance.

- 4.3-2 **Fraunhofer Diffraction with an Oblique Incident Wave.** The diffraction pattern from an aperture with aperture function  $p(x, y)$  is  $(1/\lambda d)^2 |P(x/\lambda d, y/\lambda d)|^2$ , where  $P(\nu_x, \nu_y)$  is the Fourier transform of  $p(x, y)$  and  $d$  is the distance between the aperture and observation planes. What is the diffraction pattern when the direction of the incident wave makes a small angle  $\theta_x \ll 1$ , with the  $z$ -axis in the  $x$ - $z$  plane?
- \*4.3-3 **Fresnel Diffraction from Two Pinholes.** Show that the Fresnel diffraction pattern from two pinholes separated by a distance  $2a$ , i.e.,  $p(x, y) = [\delta(x - a) + \delta(x + a)]\delta(y)$ , at an observation distance  $d$  is the periodic pattern,  $I(x, y) = (2/\lambda d)^2 \cos^2(2\pi ax/\lambda d)$ .
- \*4.3-4 **Relation Between Fresnel and Fraunhofer Diffraction.** Show that the Fresnel diffraction pattern of the aperture function  $p(x, y)$  is equal to the Fraunhofer diffraction pattern of the aperture function  $p(x, y) \exp[-j\pi(x^2 + y^2)/\lambda d]$ .
- 4.4-1 **Blurring a Sinusoidal Grating.** An object  $f(x, y) = \cos^2(2\pi x/a)$  is imaged by a defocused single-lens imaging system whose impulse-response function  $h(x, y) = 1$  within a square of width  $D$ , and  $= 0$  elsewhere. Derive an expression for the distribution of the image  $g(x, 0)$  in the  $x$  direction. Derive an expression for the contrast of the image in terms of the ratio  $D/a$ . The contrast  $= (\max - \min)/(\max + \min)$ , where  $\max$  and  $\min$  are the maximum and minimum values of  $g(x, 0)$ .
- 4.4-2 **Image of a Phase Object.** An imaging system has an impulse-response function  $h(x, y) = \text{rect}(x)\delta(y)$ . If the input wave is

$$f(x, y) = \begin{cases} \exp\left(j\frac{\pi}{2}\right) & \text{for } x > 0 \\ \exp\left(-j\frac{\pi}{2}\right) & \text{for } x \leq 0, \end{cases}$$

determine and sketch the intensity  $|g(x, y)|^2$  of the output wave  $g(x, y)$ . Verify that even though the intensity of the input wave  $|f(x, y)|^2 = 1$ , the intensity of the output wave is not uniform.

- 4.4-3 **Optical Spatial Filtering.** Consider the spatial filtering system shown in Fig. 4.4-5 with  $f = 1000$  mm. The system is illuminated with a uniform plane wave of unit amplitude and wavelength  $\lambda = 10^{-3}$  mm. The input transparency has amplitude transmittance  $f(x, y)$  and the mask has amplitude transmittance  $p(x, y)$ . Write an expression relating the complex amplitude  $g(x, y)$  of light in the image plane to  $f(x, y)$  and  $p(x, y)$ . Assuming that all distances are measured in mm, sketch  $g(x, 0)$

in the following cases:

(a)  $f(x, y) = \delta(x - 5)$  and  $p(x, y) = \text{rect}(x)$ ;

(b)  $f(x, y) = \text{rect}(x)$  and  $p(x, y) = \text{sinc}(x)$ .

Determine  $p(x, y)$  such that  $g(x, y) = \nabla_T^2 f(x, y)$ , where  $\nabla_T^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$  is the transverse Laplacian operator.

4.4-4 **Optical Cross-Correlation.** Show how a spatial filter may be used to perform the operation of cross-correlation (defined in Appendix A) between two images described by the real-valued functions  $f_1(x, y)$  and  $f_2(x, y)$ . Under what conditions would the complex amplitude transmittances of the masks and transparencies used be real-valued?

\*4.4-5 **Impulse-Response Function of a Severely Defocused System.** Using wave optics, show that the impulse-response function of a severely defocused imaging system (one for which the defocusing error  $\epsilon$  is very large) may be approximated by  $h(x, y) = p(x/\epsilon d_2, y/\epsilon d_2)$ , where  $p(x, y)$  is the pupil function. *Hint:* Use the method of stationary phase described on page 124 (proof 2) to evaluate the integral that results from the use of (4.4-11) and (4.4-10). Note that this is the same result predicted by the ray theory of light [see (4.4-2)].

4.4-6 **Two-Point Resolution.** (a) Consider the single-lens imaging system discussed in Sec. 4.4C. Assuming a square aperture of width  $D$ , unit magnification, and perfect focus, write an expression for the impulse-response function  $h(x, y)$ .

(b) Determine the response of the system to an object consisting of two points separated by a distance  $b$ , i.e.,

$$f(x, y) = \delta(x)\delta(y) + \delta(x - b)\delta(y).$$

(c) If  $\lambda d_2/D = 0.1$  mm, sketch the magnitude of the image  $g(x, 0)$  as a function of  $x$  when the points are separated by a distance  $b = 0.5, 1,$  and  $2$  mm. What is the minimum separation between the two points such that the image remains discernible as two spots instead of a single spot, i.e., has two peaks.

4.4-7 **Ring Aperture.** (a) A focused single-lens imaging system, with magnification  $M = 1$  and focal length  $f = 100$  cm has an aperture in the form of a ring

$$p(x, y) = \begin{cases} 1, & a \leq (x^2 + y^2)^{1/2} \leq b, \\ 0, & \text{otherwise,} \end{cases}$$

where  $a = 5$  mm and  $b = 6$  mm. Determine the transfer function  $H(\nu_x, \nu_y)$  of the system and sketch its cross section  $H(\nu_x, 0)$ . The wavelength  $\lambda = 1$   $\mu\text{m}$ .

(b) If the image plane is now moved closer to the lens so that its distance from the lens becomes  $d_2 = 25$  cm, with the distance between the object plane and the lens  $d_1$  as in (a), use the ray-optics approximation to determine the impulse-response function of the imaging system  $h(x, y)$  and sketch  $h(x, 0)$ .

4.5-1 **Holography with a Spherical Reference Wave.** The choice of a uniform plane wave as a reference wave is not essential to holography; other waves can be used. Assuming that the reference wave is a spherical wave centered about the point  $(0, 0, -d)$ , determine the hologram pattern and examine the reconstructed wave when:

(a) the object wave is a plane wave traveling at an angle  $\theta_x$ ;

(b) the object wave is a spherical wave centered at  $(-x_0, 0, -d_1)$ .

Approximate spherical waves by paraboloidal waves.

- 4.5-2 **Optical Correlation.** A transparency with an amplitude transmittance given by  $f(x, y) = f_1(x - a, y) + f_2(x + a, y)$  is Fourier transformed by a lens and the intensity is recorded on a transparency (hologram). The hologram is subsequently illuminated with a reference wave and the reconstructed wave is Fourier transformed with a lens to generate the function  $g(x, y)$ . Derive an expression relating  $g(x, y)$  to  $f_1(x, y)$  and  $f_2(x, y)$ . Show how the correlation of the two functions  $f_1(x, y)$  and  $f_2(x, y)$  may be determined with this system.