

## Chapter 2

### Wave Optics and Holography

In Chapter 1, we presented some mathematical background of Fourier optics as well as some important systems properties including linearity and space invariance. In this chapter, we present some fundamentals of wave optics by starting from Maxwell's equations and deriving the vector wave equation. We will then discuss some simple solutions of the scalar wave equation. Next, we will develop diffraction theory by using the Fresnel diffraction formula, which is uniquely derived by using Fourier transforms. In the process, we will define the spatial frequency transfer function and the spatial impulse response in Fourier optics. In the context of diffraction, we will also develop wavefront transformation by using a lens, show the Fourier transforming properties of the lens, and discuss how spatial filtering is obtained by using a standard two-lens system, leading to the distinction between coherent and incoherent image processing. In the last section of this chapter, we will discuss the basics of holography and show that a Fresnel zone plate is the hologram of a point source object, leading to the concept that the hologram of an arbitrary 3-D object can be considered as a collection of Fresnel zone plates. Finally, we will discuss electronic holography (often called digital holography in literature). This will culminate with the next chapter, which we will discuss a unique holographic recording technique called optical scanning holography.

#### 2.1 Maxwell's Equations and Homogenous Vector Wave Equation

Generally, in the study of optics, we are concerned with four vector quantities called electromagnetic (EM) fields: the electric field strength  $\mathcal{E}$  (V/m), the electric flux density  $\mathcal{D}$  (C/m<sup>2</sup>), the magnetic field strength  $\mathcal{H}$  (A/m), and the magnetic flux density  $\mathcal{B}$  (Wb/m<sup>2</sup>). The fundamental theory of electromagnetic fields is based on *Maxwell's Equations*. In differential form, these equations are expressed as

$$\nabla \cdot \mathcal{D} = \rho_v, \quad (2.1-1)$$

$$\nabla \cdot \mathcal{B} = 0, \quad (2.1-2)$$

$$\nabla \times \mathcal{E} = - \frac{\partial \mathcal{B}}{\partial t}, \quad (2.1-3)$$

$$\nabla \times \mathcal{H} = \mathcal{J} = \mathcal{J}_c + \frac{\partial \mathcal{D}}{\partial t}, \quad (2.1-4)$$

where  $\mathcal{J}_c$  is the current density [A/m<sup>2</sup>] and  $\rho_v$  denotes the electric charge density [C/m<sup>3</sup>].  $\mathcal{J}_c$  and  $\rho_v$  are the sources generating the electromagnetic fields. Maxwell's equations express the physical laws governing the *electric fields*  $\mathcal{E}$  and  $\mathcal{D}$ , *magnetic fields*  $\mathcal{H}$  and  $\mathcal{B}$ , and the *sources*  $\mathcal{J}_c$  and  $\rho_v$ . From Eqs. (2.1-3) and (2.1-4), we see that a time-varying magnetic field produces a time-varying electric field. Conversely, a time-varying electric field produces a time-varying magnetic field. It is precisely this coupling between the electric and magnetic fields that generate electromagnetic waves capable of propagating through a medium or even in free space.

For any given current and charge density distribution, we can solve Maxwell's equations. However, we need to note that Eq. (2.1-1) is not independent of Eq. (2.1-4). Similarly, Eq. (2.1-2) is a consequence of Eq. (2.1-3). By taking the divergence on both sides of Eqs. (2.1-3) and (2.1-4) and using the *continuity equation*:

$$\nabla \cdot \mathcal{J}_c + \frac{\partial \rho_v}{\partial t} = 0, \quad (2.1-5)$$

which is the *principle of conservation of charge*, we can show that  $\nabla \cdot \mathcal{D} = \rho_v$ . Similarly, Eq. (2.1-2) is a consequence of Eq. (2.1-3). Hence, from Eqs. (2.1-1) to (2.1-4), we really have six independent scalar equations (three scalar equations for each curl equation) and twelve unknowns. The unknowns are the  $x$ ,  $y$ , and  $z$  components of  $\mathcal{E}$ ,  $\mathcal{D}$ ,  $\mathcal{H}$ , and  $\mathcal{B}$ . The six more scalar equations required are provided by the *constitutive relations*:

$$\mathcal{D} = \epsilon \mathcal{E}, \quad (2.1-6a)$$

and

$$\mathcal{B} = \mu \mathcal{H}, \quad (2.1-6b)$$

where  $\epsilon$  denotes the permittivity [F/m] and  $\mu$  denotes the permeability [H/m] of the medium. In this book, we take  $\epsilon$  and  $\mu$  to be scalar constants. Indeed, this is true for a *linear*, *homogeneous*, and *isotropic* medium. A medium is *linear* if its properties do not depend on the amplitude of the fields in the medium. It is *homogeneous* if its properties are not functions of space. And the medium is *isotropic* if its properties are the same in all direction from any given point.

Returning our focus to linear, homogeneous, and isotropic media, constants worth remembering are the values of  $\epsilon$  and  $\mu$  for free space (or vacuum):  $\epsilon_0 = (1/36\pi) \times 10^{-9}$  F/m and  $\mu_0 = 4\pi \times 10^{-7}$  H/m.

Using Maxwell's equations and the constitutive relations, we can derive the wave equation, which describes the propagation of the electric and magnetic fields. Example 2.1 shows the derivation of the wave equation for  $\mathcal{E}$ .

### Example 2.1 Derivation of Vector Wave Equation in a Linear, Homogenous, and Isotropic Medium

By taking the curl of both sides of Eq. (2.1-3), we have

$$\begin{aligned}\nabla \times \nabla \times \mathcal{E} &= -\nabla \times \frac{\partial \mathcal{B}}{\partial t} \\ &= -\frac{\partial}{\partial t}(\nabla \times \mathcal{B}) = -\mu \frac{\partial}{\partial t}(\nabla \times \mathcal{H}),\end{aligned}\quad (2.1-7)$$

where we have used the second of the constitutive relations [Eq. (2.1-6b)] and assumed  $\mu$  to be space- and time-independent. Now, by employing Eq. (2.1-4), Eq. (2.1-7) becomes

$$\nabla \times \nabla \times \mathcal{E} = -\mu\epsilon \frac{\partial^2 \mathcal{E}}{\partial t^2} - \mu \frac{\partial \mathcal{J}_c}{\partial t}, \quad (2.1-8)$$

where we have used the first of the constitutive relations [Eq. (2.1-6a)] and assumed  $\epsilon$  to be time-independent. Then, by using the following vector identity ( $\mathcal{A}$  is some arbitrary vector)

$$\nabla \times \nabla \times \mathcal{A} = \nabla(\nabla \cdot \mathcal{A}) - \nabla^2 \mathcal{A}, \quad \nabla^2 = \nabla \cdot \nabla, \quad (2.1-9)$$

in Eq. (2.1-8), we get

$$\nabla^2 \mathcal{E} - \mu\epsilon \frac{\partial^2 \mathcal{E}}{\partial t^2} = \mu \frac{\partial \mathcal{J}_c}{\partial t} + \nabla(\nabla \cdot \mathcal{E}). \quad (2.1-10)$$

If we also assume the permittivity,  $\epsilon$ , to be space-independent, then we can now recast the first of Maxwell's equations [Eq. (2.1-1)] in the form of

$$\nabla \cdot \mathcal{E} = \frac{\rho_v}{\epsilon}, \quad (2.1-11)$$

by using the first of the constitutive relations [Eq. (2.1-6a)]. Incorporating Eq. (2.1-11) into Eq. (2.1-10), we can finally obtain

$$\nabla^2 \mathcal{E} - \mu\epsilon \frac{\partial^2 \mathcal{E}}{\partial t^2} = \mu \frac{\partial \mathcal{J}_c}{\partial t} + \frac{1}{\epsilon} \nabla \rho_v, \quad (2.1-12)$$

which is a *vector wave equation* having source terms on the right-hand side. This is the wave equation for  $\mathcal{E}$  in a linear, homogeneous, and isotropic medium characterized by  $\mu$  and  $\epsilon$ .

For the given  $\mathcal{J}_c$  and  $\rho_v$  in a localized region characterized by  $\mu$  and  $\epsilon$ , say,  $V'$ , we can solve for the electric field,  $\mathcal{E}$ , in the region according to Eq. (2.1-12). Once the generated field reaches the source-free region  $V$  ( $\mathcal{J}_c = 0$ ,  $\rho_v = 0$ ), the field must then satisfy the *homogenous vector wave equation*,

$$\nabla^2 \mathcal{E} - \mu\epsilon \frac{\partial^2 \mathcal{E}}{\partial t^2} = 0. \quad (2.1-13)$$

The situation is delineated in Fig. 2.1. Note that the quantity,  $\mu\epsilon$ , has the unit value of  $(1/\text{velocity})^2$ . We call this velocity  $v$  and define it as

$$v^2 = \frac{1}{\mu\epsilon}. \quad (2.1-14)$$

For free space,  $\mu = \mu_0$ ,  $\epsilon = \epsilon_0$ , and  $v = c$ . We can calculate the value of  $c$  from the values of  $\epsilon_0$  and  $\mu_0$ . This works out to be about  $3 \times 10^8$  m/s. This theoretical value, which was first calculated by Maxwell, was in remarkable agreement with Fizeau's previously measured speed of light (315,300 km/s). This led Maxwell to the conclusion that light is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field based on electromagnetic laws.

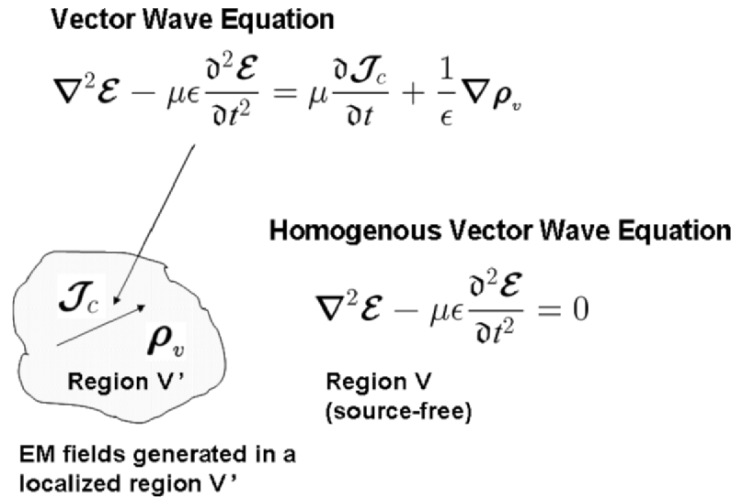


Fig. 2.1 Vector wave equations in a linear, homogeneous, and isotropic medium.

## 2.2 Three-Dimensional Scalar Wave Equation

Equation (2.1-13) is equivalent to three scalar equations - one for every component of  $\mathcal{E}$ . We shall let the field  $\mathcal{E}$  to be of the form

$$\mathcal{E} = \mathcal{E}_x \mathbf{a}_x + \mathcal{E}_y \mathbf{a}_y + \mathcal{E}_z \mathbf{a}_z, \quad (2.2-1)$$

where  $\mathbf{a}_x$ ,  $\mathbf{a}_y$ , and  $\mathbf{a}_z$  denote the unit vectors in the  $x$ ,  $y$ , and  $z$  directions, respectively. Now, the expression for the Laplacian ( $\nabla^2$ ) operator in Cartesian ( $x, y, z$ ) coordinates is given by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (2.2-2)$$

Using the above equation, Equation (2.1-13) becomes

$$\begin{aligned} & \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) (\mathcal{E}_x \mathbf{a}_x + \mathcal{E}_y \mathbf{a}_y + \mathcal{E}_z \mathbf{a}_z) \\ &= \mu\epsilon \frac{\partial^2}{\partial t^2} (\mathcal{E}_x \mathbf{a}_x + \mathcal{E}_y \mathbf{a}_y + \mathcal{E}_z \mathbf{a}_z). \end{aligned} \quad (2.2-3)$$

Comparing the  $\mathbf{a}_x$ -component on both sides of the equation, we have

$$\frac{\partial^2 \mathcal{E}_x}{\partial x^2} + \frac{\partial^2 \mathcal{E}_x}{\partial y^2} + \frac{\partial^2 \mathcal{E}_x}{\partial z^2} = \mu\epsilon \frac{\partial^2 \mathcal{E}_x}{\partial t^2}.$$

Similarly, we end up with the same type of equation shown above for the  $\mathcal{E}_y$  and  $\mathcal{E}_z$  component by comparing the other components in Eq. (2.2-3). Therefore, we can write

$$\nabla^2 \psi = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2}, \quad (2.2-4)$$

where  $\psi$  may represent a component,  $\mathcal{E}_x$ ,  $\mathcal{E}_y$  or  $\mathcal{E}_z$ , of the electric field  $\mathcal{E}$ , and where  $v$  is the velocity of the wave in the medium by using Eq. (2.1-14). Equation (2.2-4) is called the *3-D scalar wave equation*. We shall look at some of its simplest solutions in the next section.

### 2.2.1 Plane Wave Solution

For waves oscillating at the *angular frequency*,  $\omega_0$  (rad/s), one of the simplest solutions to Eq. (2.2-4) is

$$\psi(x, y, z, t) = \exp[j(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{R})]$$

$$= \exp[j(\omega_0 t - k_{0x}x - k_{0y}y - k_{0z}z)], \quad (2.2-5)$$

where  $\mathbf{R} = x\mathbf{a}_x + y\mathbf{a}_y + z\mathbf{a}_z$  is the position vector,  $\mathbf{k}_0 = k_{0x}\mathbf{a}_x + k_{0y}\mathbf{a}_y + k_{0z}\mathbf{a}_z$  is the *propagation vector*, and  $|\mathbf{k}_0| = k_0$  is called the *propagation constant* [rad/m]. With the condition that

$$\frac{\omega_0^2}{k_{0x}^2 + k_{0y}^2 + k_{0z}^2} = \frac{\omega_0^2}{k_0^2} = v^2, \quad (2.2-6)$$

Eq. (2.2-5) is called a plane-wave solution and the wave is called a *plane wave* of unit amplitude. Figure 2.2 shows the direction of propagation of the plane wave, which is determined from the three components  $k_{0x}$ ,  $k_{0y}$ , and  $k_{0z}$ .

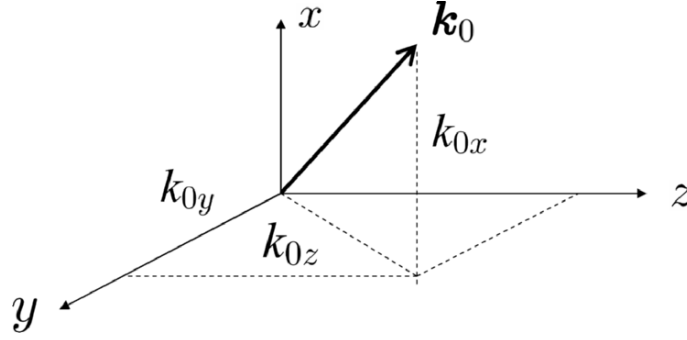


Fig. 2.2 Plane wave propagating along the  $\mathbf{k}_0$  direction.

Since the electromagnetic fields are real functions of space and time, we can define, for example, the electric field by taking the real part of  $\psi$  to obtain a real quantity,

$$\text{Re}[\psi(x, y, z, t)] = \cos(\omega_0 t - k_{0x}x - k_{0y}y - k_{0z}z). \quad (2.2-7)$$

Let us now consider a plane wave propagating along the  $z$ -direction. In one spatial dimension, i.e.,  $\psi(z, t)$ , the wave equation [Eq. (2.2-4)] reads

$$\frac{\partial^2 \psi}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2} \quad (2.2-8)$$

and its plane wave solution then becomes

$$\psi(z, t) = \exp[j(\omega_0 t - k_0 z)] = \exp[j(\omega_0 t)] \exp[-j\theta(z)], \quad (2.2-9)$$

where  $\theta(z) = k_0 z = \frac{2\pi}{\lambda_0} z$  is called the *phase* of the wave with  $\lambda_0$  indicating the wavelength of the wave. Let us take the origin of the coordinates as a

zero-phase position, i.e.,  $\theta(z = 0) = 0$ . In fact, over the whole plane  $z = 0$ , the phase is zero. At  $z = \lambda_0$ , we have  $\theta(z = \lambda_0) = \frac{2\pi}{\lambda_0} \lambda_0 = 2\pi$ . So for every distance of propagation of a wavelength, the phase of the wave gains  $2\pi$ . Therefore, we have what is known as the *planar wavefronts* along the  $z$ -direction. The situation is demonstrated in Fig. 2.3.

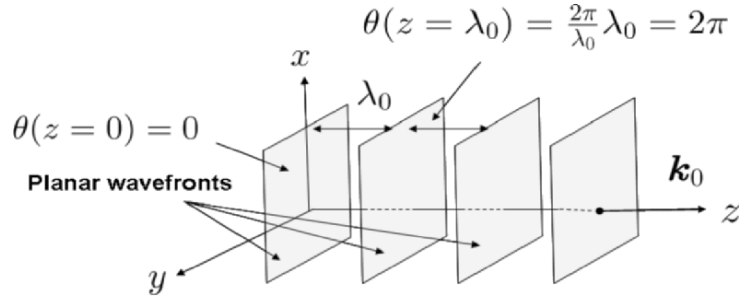


Fig. 2.3 Plane wave propagating along the  $z$ -direction exhibiting planar wavefronts.

### 2.2.2 Spherical Wave Solution

Consider now the spherical coordinates shown in Fig. 2.4.

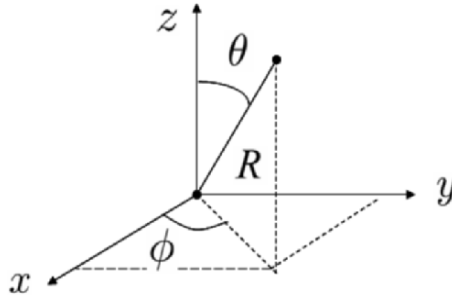


Fig. 2.4 Spherical coordinate system.

The expression for the Laplacian ( $\nabla^2$ ) operator is

$$\nabla^2 = \frac{\partial^2}{\partial R^2} + \frac{2}{R} \frac{\partial}{\partial R} + \frac{1}{R^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} + \frac{1}{R^2} \frac{\partial^2}{\partial \theta^2} + \frac{\cot \theta}{R^2} \frac{\partial}{\partial \theta}. \quad (2.2-10)$$

One of the simplest cases is called spherical symmetry, which requires that  $\psi(R, \theta, \phi, t) = \psi(R, t)$ . Therefore, for spherical symmetry ( $\partial/\partial \phi = 0 = \partial/\partial \theta$ ), the wave equation, Eq. (2.2-4), combined with Eq. (2.2-10) assumes the form

$$\left( \frac{\partial^2 \psi}{\partial R^2} + \frac{2}{R} \frac{\partial \psi}{\partial R} \right) = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2}. \quad (2.2-11)$$

Since

$$R \left( \frac{\partial^2 \psi}{\partial R^2} + \frac{2}{R} \frac{\partial \psi}{\partial R} \right) = \frac{\partial^2 (R\psi)}{\partial R^2},$$

we can re-write Eq. (2.2-11) to become

$$\frac{\partial^2 (R\psi)}{\partial R^2} = \frac{1}{v^2} \frac{\partial^2 (R\psi)}{\partial t^2}. \quad (2.2-12)$$

Now, the above equation is of the same form as that of Eq. (2.2-8). Since Eq. (2.2-9) is the solution to Eq. (2.2-8), we can therefore construct a simple solution to Eq. (2.2-12) as

$$\psi(R, t) = \frac{1}{R} \exp[j(\omega_0 t - k_0 R)], \quad (2.2-13)$$

which is called a *spherical wave*. Again, we can write

$$\psi(R, t) = \frac{1}{R} \exp[j(\omega_0 t - k_0 R)] = \frac{1}{R} \exp(j\omega_0 t) \exp[-j\theta(R)],$$

where  $\theta(R) = k_0 R = \frac{2\pi}{\lambda_0} R$ . We then take the origin of the coordinates as a zero-phase position, i.e.,  $\theta(R = 0) = 0$  and  $\theta(R = \lambda_0) = \frac{2\pi}{\lambda_0} \lambda_0 = 2\pi$ . So, for every distance of propagation of a wavelength, the phase of the wave gains  $2\pi$ . We, therefore, have the so-called *spherical wavefronts* moving along the  $R$ -direction. The situation is shown in Fig. 2.5.

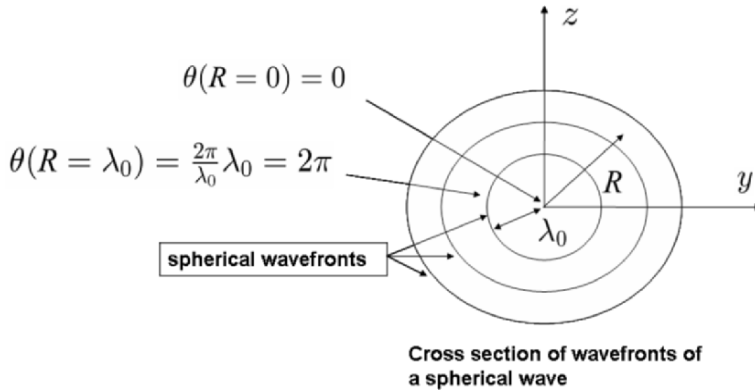


Fig. 2.5 Spherical wavefronts.

While we have previously seen that plane waves and spherical waves are some of the simplest solutions of the 3-D scalar wave equation, we can effectively generate these useful waves in the laboratory. The situation is



shown in Fig. 2.6, where the distance between the two lenses are separated by the sum of their focal lengths,  $f_1 + f_2$ , and we have assumed that the rays emitting from the laser are parallel, i.e., the wave fronts are planar. Note that the parallel rays emerging from the lens of focal length  $f_2$  have a separation of an expansion factor,  $M = f_2/f_1$ , larger than the separation of the rays originally emerging from the laser.

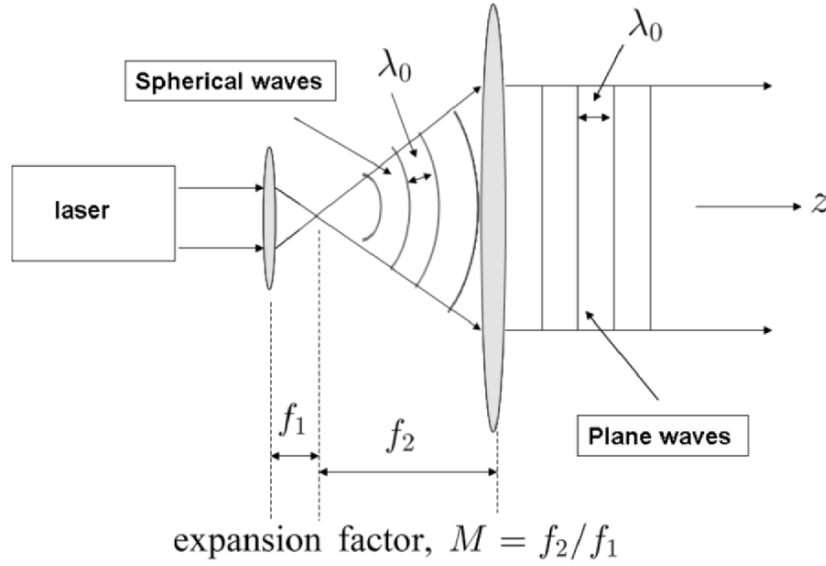


Fig. 2.6 Practical implementation of spherical waves and plane waves.

### 2.3 Scalar Diffraction Theory

Figure 2.7 shows a simple example of diffraction geometry where a plane wave oscillating at  $\omega_0$  is incident on an *aperture* or a *diffracting screen*, located on the plane  $z = 0$ . The problem is to determine the diffracted field distribution after the aperture. To tackle the problem, we will need to solve the 3-D scalar wave equation, which is subject to an initial condition. Let us now formulate the problem mathematically.

Since a plane wave of amplitude  $A$  propagating along the  $z$ -direction is given by  $\psi(z, t) = A \exp[j(\omega_0 t - k_0 z)]$  with the wave's zero-phase position defined at  $z = 0$ , we can then model the field immediately in front of the aperture as  $\psi(z = 0, t) = A \exp(j\omega_0 t)$ . The field immediately after the aperture is then given by  $\psi(x, y, z = 0, t) = \psi_p(x, y; z = 0) \exp(j\omega_0 t)$ .  $\psi_p(x, y; z = 0)$  is called the *initial condition* under consideration. For example, if the aperture has a rectangular opening of width  $x_0$  by  $y_0$ , we can then write  $\psi_p(x, y; z = 0) = \text{Arect}(x/x_0, y/y_0)$ .

It is necessary to find the field distribution at  $z$ , and to do so, we can model the solution as

$$\psi(x, y, z, t) = \psi_p(x, y; z) \exp(j\omega_0 t), \quad (2.3-1)$$

where  $\psi_p(x, y; z)$  is the unknown to be found. In optics,  $\psi_p(x, y; z)$  is called a *complex amplitude*  $\psi_p(x, y; z)$  and we see that it is riding on a *carrier* of frequency  $\omega_0$  ( $\psi_p$  is known as a *phasor* in electrical engineering).

Since the light field must satisfy the wave equation, we therefore substitute this into the 3-D scalar wave equation [Eq. (2.2-4)] to find  $\psi_p(x, y; z)$  under the given initial condition,  $\psi_p(x, y; z = 0)$ .

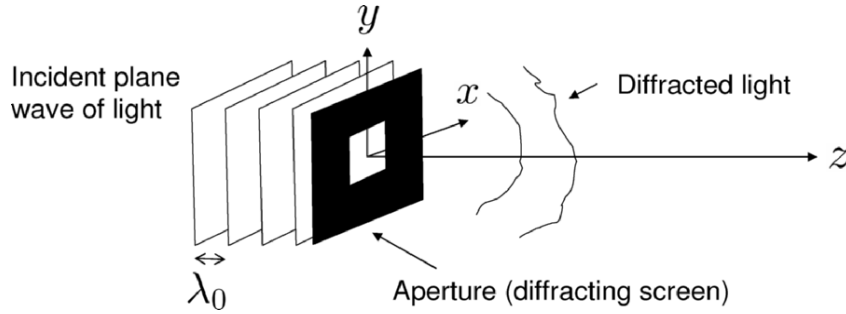


Fig. 2.7 Diffraction geometry.

After substituting Eq. (2.3-1) into Eq. (2.2-4), we get the *Helmholtz equation* for  $\psi_p$ ,

$$\frac{\partial^2 \psi_p}{\partial x^2} + \frac{\partial^2 \psi_p}{\partial y^2} + \frac{\partial^2 \psi_p}{\partial z^2} + k_0^2 \psi_p = 0, \quad k_0 = \frac{\omega_0}{v}. \quad (2.3-2)$$

By taking the 2-D Fourier transform, i.e.,  $\mathcal{F}_{xy}$ , of Eq. (2.3-2) and after further manipulations, we obtain

$$\frac{d^2 \Psi_p}{dz^2} + k_0^2 \left(1 - \frac{k_x^2}{k_0^2} - \frac{k_y^2}{k_0^2}\right) \Psi_p = 0, \quad (2.3-3)$$

where  $\Psi_p(k_x, k_y; z)$  is the Fourier transform of  $\psi_p(x, y; z)$ . We can now readily solve the above equation to get

$$\Psi_p(k_x, k_y; z) = \Psi_{p0}(k_x, k_y) \exp\left[-jk_0 \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} z\right], \quad (2.3-4)$$

where  $\Psi_{p0}(k_x, k_y) = \Psi_p(k_x, k_y; z = 0)$

$$= \mathcal{F}_{xy}\{\psi_p(x, y; z = 0)\} = \mathcal{F}_{xy}\{\psi_{p0}(x, y)\}.$$

We can interpret Eq. (2.3-4) by considering a linear system with  $\Psi_{p0}(k_x, k_y)$  as its input spectrum (i.e., at  $z = 0$ ) and where the output spectrum is  $\Psi_p(k_x, k_y; z)$ . Conclusively, the spatial frequency response of the system is given by

$$\begin{aligned} \frac{\Psi_p(k_x, k_y; z)}{\Psi_{p0}(k_x, k_y)} &= \mathcal{H}(k_x, k_y; z) \\ &= \exp\left[-jk_0\sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} z\right]. \end{aligned} \quad (2.3-5)$$

We call  $\mathcal{H}(k_x, k_y; z)$  the *spatial frequency transfer function of propagation* of light through a distance  $z$  in the medium. Figure 2.8 shows the relationship between the input spectrum and the output spectrum.

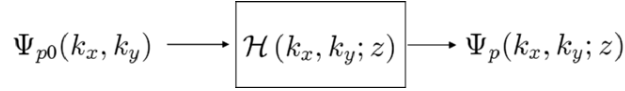


Fig. 2.8 Spatial frequency transfer function of propagation relating input spectrum to output spectrum.

### Example 2.2 Derivation of the Helmholtz Equation

When we substitute  $\psi(x, y, z, t) = \psi_p(x, y; z)\exp(j\omega_0 t)$  into the 3-D scalar wave equation given by Eq. (2.2-4), we have

$$\left[ \frac{\partial^2 \psi_p}{\partial x^2} + \frac{\partial^2 \psi_p}{\partial y^2} + \frac{\partial^2 \psi_p}{\partial z^2} \right] \exp(j\omega_0 t) = \frac{(j\omega_0)^2}{v^2} \psi_p(x, y; z) \exp(j\omega_0 t)$$

or

$$\left[ \frac{\partial^2 \psi_p}{\partial x^2} + \frac{\partial^2 \psi_p}{\partial y^2} + \frac{\partial^2 \psi_p}{\partial z^2} \right] = \frac{-\omega_0^2}{v^2} \psi_p(x, y; z) = -k_0^2 \psi_p$$

which is the Helmholtz equation [Eq. (2.3-2)], where we have incorporated the fact that  $k_0 = \omega_0/v$ . Note that the Helmholtz equation contains no time variable.

### Example 2.3 Derivation of Eq. (2.3-3) and its Solution

By taking the 2-D Fourier transform, i.e.,  $\mathcal{F}_{xy}$ , of Eq. (2.3-2) and by using item #5 of Table 1.1, we can obtain

$$\mathcal{F}_{xy} \left\{ \frac{\partial^2 \psi_p}{\partial x^2} + \frac{\partial^2 \psi_p}{\partial y^2} + \frac{\partial^2 \psi_p}{\partial z^2} + k_0^2 \psi_p \right\} = 0$$

or

$$- (k_x^2 + k_y^2) \Psi_p(k_x, k_y; z) + \frac{d^2 \Psi_p(k_x, k_y; z)}{dz^2} + k_0^2 \Psi_p(k_x, k_y; z) = 0,$$

which can then be re-arranged to become

$$\frac{d^2 \Psi_p}{dz^2} + k_0^2 \left( 1 - \frac{k_x^2}{k_0^2} - \frac{k_y^2}{k_0^2} \right) \Psi_p = 0. \quad (2.3-6)$$

This equation is of the form

$$\frac{d^2 y}{dz^2} + \alpha^2 y = 0,$$

which has the solution  $y(z) = y_0 \exp(-j\alpha z)$  where  $y_0 = y(z=0)$  is given as the initial condition. Using this result, the solution to Eq. (2.3-6) becomes

$$\begin{aligned} \Psi_p(k_x, k_y; z) &= \Psi(k_x, k_y; z=0) \exp \left[ -jk_0 \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} z \right] \\ &= \Psi_{p0}(k_x, k_y) \exp \left[ -jk_0 \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} z \right], \end{aligned} \quad (2.3-7)$$

which is Eq. (2.3-4).

To find the field distribution at  $z$  in the spatial domain, we take the inverse Fourier transform of Eq. (2.3-7):

$$\begin{aligned} \psi_p(x, y; z) &= \mathcal{F}_{xy}^{-1} \left\{ \Psi_p(k_x, k_y; z) \right\} \\ &= \frac{1}{4\pi^2} \int \int \Psi_{p0}(k_x, k_y) \exp \left[ -jk_0 \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} z \right] \\ &\quad \times \exp(-jk_x x - jk_y y) dk_x dk_y. \end{aligned} \quad (2.3-8)$$

Now, by substituting  $\Psi_{p0}(k_x, k_y) = \mathcal{F}_{xy} \left\{ \psi_{p0}(x, y) \right\}$  into Eq. (2.3-8), we can

express  $\psi_p(x, y; z)$  as

$$\begin{aligned}\psi_p(x, y; z) &= \iint \psi_{p0}(x', y') G(x - x', y - y'; z) dx' dy' \\ &= \psi_{p0}(x, y) * G(x, y; z),\end{aligned}\quad (2.3-9)$$

where

$$\begin{aligned}G(x, y; z) &= \frac{1}{4\pi^2} \iint \exp[-jk_0 \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} z] \\ &\quad \times \exp(-jk_x x - jk_y y) dk_x dk_y.\end{aligned}$$

The result of Eq. (2.3-9) indicates that  $G(x, y; z)$  can be considered as the *spatial impulse response of propagation* of light, which can be evaluated to become [Poon and Banerjee (2001)]

$$\begin{aligned}G(x, y; z) &= \frac{jk_0 \exp(-jk_0 \sqrt{x^2 + y^2 + z^2})}{2\pi \sqrt{x^2 + y^2 + z^2}} \\ &\quad \times \frac{z}{\sqrt{x^2 + y^2 + z^2}} \left(1 + \frac{1}{jk_0 \sqrt{x^2 + y^2 + z^2}}\right).\end{aligned}\quad (2.3-10)$$

### 2.3.1 Fresnel Diffraction

Equation (2.3-10) is complicated to use as is, and we shall need to make the following approximations to obtain the well-known Fresnel diffraction formula commonly used in Fourier optics:

(1) For  $z \gg \lambda_0 = 2\pi/k_0$ , i.e., we observe the field distribution many wavelengths away from the diffracting aperture, and we have

$$\left(1 + \frac{1}{jk_0 \sqrt{x^2 + y^2 + z^2}}\right) \approx 1.$$

(2) Using the binomial expansion, the factor

$$\sqrt{x^2 + y^2 + z^2} \approx z + \frac{x^2 + y^2}{2z},$$

provided that  $x^2 + y^2 \ll z^2$ . This condition is called the *paraxial approximation*. If this approximation is used in the more sensitive phase term and only used the first expansion term in the less sensitive denominators of

the first and second terms of Eq. (2.3-10), then  $G(x, y; z)$  becomes the so-called *spatial impulse response in Fourier optics*,  $h(x, y; z)$  [Poon and Banerjee (2001), Goodman (2005)]:

$$h(x, y; z) = \exp(-jk_0 z) \frac{jk_0}{2\pi z} \exp\left[\frac{-jk_0(x^2 + y^2)}{2z}\right]. \quad (2.3-11)$$

If Eq. (2.3-11) is now used in Eq. (2.3-9), we obtain

$$\begin{aligned} \psi_p(x, y; z) &= \psi_{p0}(x, y) * h(x, y; z) \\ &= \exp(-jk_0 z) \frac{jk_0}{2\pi z} \iint \psi_{p0}(x', y') \\ &\quad \times \exp\left\{\frac{-jk_0}{2z}[(x - x')^2 + (y - y')^2]\right\} dx' dy'. \end{aligned} \quad (2.3-12)$$

Equation (2.3-12) is called the *Fresnel diffraction formula* and describes the *Fresnel diffraction* of a beam during propagation and having an arbitrary initial complex profile,  $\psi_{p0}(x, y)$ . To obtain the output field distribution  $\psi_p(x, y; z)$  at a distance  $z$  away from the input (the location of the diffracting screen), we would simply convolve the input field distribution,  $\psi_{p0}(x, y)$ , with the spatial impulse response,  $h(x, y; z)$ .

By taking the 2-D Fourier transform of  $h(x, y; z)$ , we obtain

$$\begin{aligned} H(k_x, k_y; z) &= \mathcal{F}_{xy}\{h(x, y; z)\} \\ &= \exp(-jk_0 z) \exp\left[\frac{j(k_x^2 + k_y^2)z}{2k_0}\right]. \end{aligned} \quad (2.3-13)$$

$H(k_x, k_y; z)$  is known as the *spatial frequency transfer function in Fourier optics*. Indeed, we can derive Eq. (2.3-13) directly if we assume that  $k_x^2 + k_y^2 \ll k_0^2$ , meaning that the  $x$  and  $y$  components of the propagation vector of a wave are relatively small, from Eq. (2.3-5), we have

$$\begin{aligned} \frac{\Psi_p(k_x, k_y; z)}{\Psi_{p0}(k_x, k_y)} &= \mathcal{H}(k_x, k_y; z) \\ &= \exp\left[-jk_0 \sqrt{1 - (k_x^2 + k_y^2)/k_0^2} z\right] \\ &\simeq \exp(-jk_0 z) \exp\left[\frac{j(k_x^2 + k_y^2)z}{2k_0}\right] \\ &= H(k_x, k_y; z) \end{aligned}$$

or

$$\Psi_p(k_x, k_y; z) = \Psi_{p0}(k_x, k_y) H(k_x, k_y; z). \quad (2.3-14)$$

Figure 2.9 summarizes the results of Fresnel diffraction in terms of block diagrams in the spatial domain as well as in the spatial frequency domain.

#### Block diagram representation in spatial domain

$$\psi_{p0}(x, y) \longrightarrow \boxed{h(x, y; z)} \longrightarrow \psi_p(x, y; z)$$

$$\psi_p(x, y; z) = \psi_{p0}(x, y) * h(x, y; z)$$

$$h(x, y; z) = \exp(-jk_0 z) \frac{jk_0}{2\pi z} \exp\left[\frac{-jk_0(x^2 + y^2)}{2z}\right]$$

#### Block diagram representation in spatial frequency domain

$$\Psi_{p0}(k_x, k_y) \longrightarrow \boxed{H(k_x, k_y; z)} \longrightarrow \Psi_p(k_x, k_y; z)$$

$$\Psi_p(k_x, k_y; z) = \Psi_{p0}(k_x, k_y) H(k_x, k_y; z)$$

$$H(k_x, k_y; z) = \mathcal{F}_{xy}\{h(x, y; z)\} = \exp(-jk_0 z) \exp\left[\frac{j(k_x^2 + k_y^2)z}{2k_0}\right]$$

Fig. 2.9 Block diagrams to summarize Fresnel diffraction.

#### Example 2.4 Diffraction of a Point Source

A point source is represented by  $\psi_{p0}(x, y) = \delta(x, y)$ . From Eq. (2.3-12), the complex field at a distance  $z$  away is given by

$$\begin{aligned} \psi_p(x, y, z) &= \delta(x, y) * h(x, y; z) \\ &= \exp(-jk_0 z) \frac{jk_0}{2\pi z} \exp\left[-\frac{jk_0(x^2 + y^2)}{2z}\right]. \end{aligned} \quad (2.3-15)$$

This expression is the paraxial approximation to a *diverging spherical wave*. The variable  $z$  in the argument of the exponential function is called the *radius of curvature* of the spherical wave. The wavefronts are divergent

when  $z > 0$  and convergent when  $z < 0$ . We can re-write Eq. (2.3-15) as

$$\psi_p(x, y, z) = \frac{jk_0}{2\pi z} \exp\left[-jk_0\left(z + \frac{x^2 + y^2}{2z}\right)\right].$$

Now, by considering the argument of the exponential function, we see that by using the binomial expansion  $\sqrt{x^2 + y^2 + z^2} \approx z + \frac{x^2 + y^2}{2z}$ , we can write

$$\begin{aligned} \psi_p(x, y, z) &\simeq \frac{jk_0}{2\pi z} \exp\left[-jk_0(x^2 + y^2 + z^2)^{\frac{1}{2}}\right] \\ &\simeq \frac{jk_0}{2\pi R} \exp(-jk_0 R), \end{aligned} \quad (2.3-16)$$

where we have used  $z \simeq R$  in the less sensitive denominator. Eq. (2.3-16) corresponds to Eq. (2.2-13) for a diverging spherical wave.

### Example 2.5 Diffraction of a Plane Wave

For a plane wave, we can write  $\psi_{p0}(x, y) = 1$ . Then  $\Psi_{p0}(k_x, k_y) = 4\pi^2 \delta(k_x) \delta(k_y)$ . Using Eq. (2.3-14), we have

$$\begin{aligned} \Psi_p(k_x, k_y; z) &= 4\pi^2 \delta(k_x) \delta(k_y) \exp(-jk_0 z) \exp\left[\frac{j(k_x^2 + k_y^2)z}{2k_0}\right] \\ &= 4\pi^2 \delta(k_x) \delta(k_y) \exp(-jk_0 z). \end{aligned}$$

Its inverse transform gives the expression of a plane wave [see Eq. (2.2-9)],

$$\psi_p(x, y, z) = \exp(-jk_0 z).$$

As the plane wave travels, it only acquires phase shift and, as expected, is undiffracted.

### 2.3.2 Diffraction of a Square Aperture

In general, when a light field illuminates a transparency of transmission function given by  $t(x, y)$ , and if the complex amplitude of the light just in front of the transparency is  $\psi_{i,p}(x, y)$ , then the complex field immediately after the transparency is given by  $\psi_{i,p}(x, y)t(x, y)$ . In writing this product result, we assume that the transparency is infinitely thin.



Now, let us consider a simple situation where a plane wave of unit amplitude is incident normally on the transparency  $t(x, y)$ , and the field emerging from the transparency is then  $1 \times t(x, y)$  as  $\psi_{i,p}(x, y) = 1$  in the present case. We want to find the field distribution, which is a distance  $z$  away from the transparency. This corresponds to the Fresnel diffraction of an arbitrary beam profile as the transparency modifies the incident plane wave. The situation is demonstrated in Fig. 2.10.

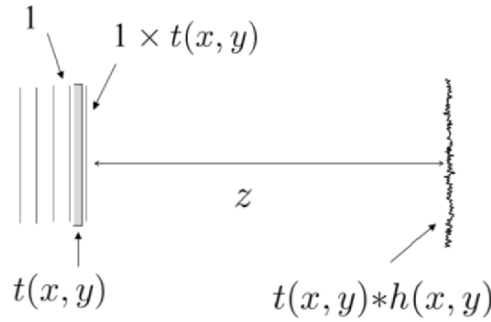
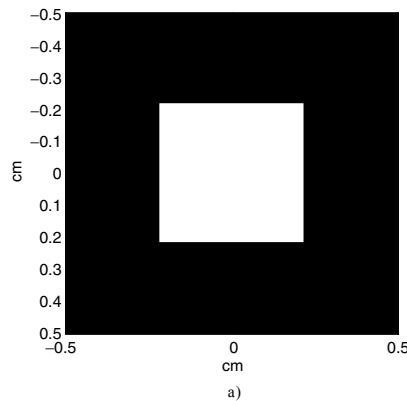


Fig. 2.10 Fresnel diffraction of an arbitrary beam profile  $t(x, y)$ .

Let us further consider a specific case where  $t(x, y) = \text{rect}(x/a, y/a)$ , a square aperture, is used for MATLAB simulations. We then implement  $\psi_p(x, y; z) = \psi_{p0}(x, y) * h(x, y; z)$  in the spatial frequency domain, i.e., using Eq. (2.3-14), where  $\psi_{p0}(x, y)$  is  $t(x, y)$  and is given by  $\text{rect}(x/a, y/a)$  with  $a = 0.4336\text{cm}$ . The m-file, `Fresnel_diffraction.m` shown in Table 2.1, generates the three figures shown below. Figure 2.11a) shows the square aperture,  $\text{rect}(x/a, y/a)$ , which is illuminated by a plane wave of red wavelength ( $\lambda_0 = 0.6328 \times 10^{-4}\text{cm}$ ). Figure 2.11b) and c) show the central cross-section of the square aperture, i.e.,  $|\psi_p(x, 0; 0)|$ , and the Fresnel diffracted magnitude, i.e.,  $|\psi_p(x, 0; z)|$ , at  $z = 5\text{cm}$ , respectively.



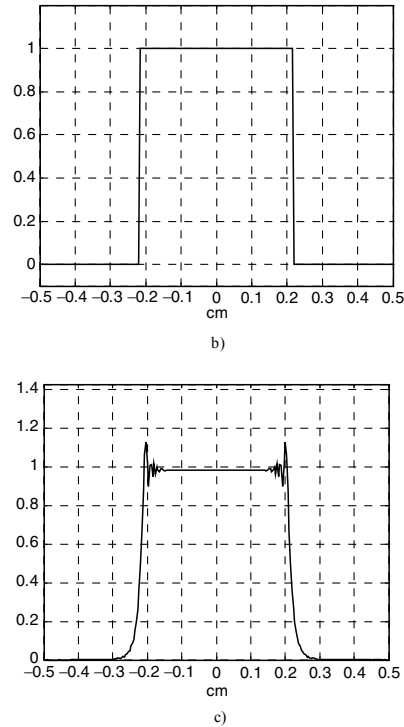


Fig. 2.11 a) Square aperture, b) Central cross-section of a),  
c) Central cross-section of diffracted amplitude at  $z = 5\text{cm}$ .

Table 2.1 Fresnel\_diffraction.m:  
m-file for calculating the Fresnel diffraction of a square aperture.

```
-----
%Fresnel_diffraction.m
%Simulation of Fresnel diffraction of a square aperture
%Adapted from "Contemporary optical image processing with MATLAB®,"
%by T.-C. Poon and P. P. Banerjee, Elsevier 2001, pp. 64-65.
clear

L=1; %L : length of display area
N=256; %N : number of sampling points
dx=L/(N-1); % dx : step size

%Create square image, M by M square, rect(x/a), M=odd number
M=111;
a=M/256
R=zeros(256); %assign a matrix (256x256) of zeros
r=ones(M); %assign a matrix (MxM) of ones
n=(M-1)/2;
R(128-n:128+n,128-n:128+n)=r;
%End of creating input image

%Axis Scaling
```

```

for k=1:256
    X(k)=1/255*(k-1)-L/2;
    Y(k)=1/255*(k-1)-L/2;

    %Kx=(2*pi*k)/((N-1)*dx)
    %in our case, N=256, dx=1/255

    Kx(k)=(2*pi*(k-1))/((N-1)*dx)-((2*pi*(256-1))/((N-1)*dx))/2;
    Ky(k)=(2*pi*(k-1))/((N-1)*dx)-((2*pi*(256-1))/((N-1)*dx))/2;
end

%Fourier transformation of R

FR=(1/256)^2*fft2(R);
FR=fftshift(FR);

%Free space impulse response function
% The constant factor exp(-jk0*z) is not calculated
%sigma=ko/(2*z)=pi/(wavelength*z)
%z=5cm,red light=0.6328*10^-4(cm)
sigma=pi/((0.6328*10^-4)*5);

for r=1:256,
    for c=1:256,
        %compute free-space impulse response with Gaussian apodization against aliasing
        h(r,c)=j*(sigma/pi)*exp(-4*200*(X(r).^2+Y(c).^2))*exp(-j*sigma*(X(r).^2+Y(c).^2));
    end
end

H=(1/256)^2*fft2(h);
H=fftshift(H);
HR=FR.*H;
H=(1/256)^2*fft2(h);
H=fftshift(H);
HR=FR.*H;

hr=ifft2(HR);
hr=(256^2)*hr;
hr=fftshift(hr);

%Image of the rectangle object
figure(1)
image(X,Y,255*HR);
colormap(gray(256));
axis square
xlabel('cm')
ylabel('cm')

% plot of cross section of square
figure(2)
plot(X+dx/2,R(:,127))
grid

```

```
axis([-0.5 0.5 -0.1 1.2])
xlabel('cm')

figure(3)
plot(X+dx/2,abs(hr(:,127)))
grid
axis([-0.5 0.5 0 max(max(abs(hr)))*1.1])
xlabel('cm')
```

-----

## 2.4 Ideal Lens, Imaging Systems, Pupil Functions and Transfer Functions

### 2.4.1 Ideal Lens and Optical Fourier Transformation

In the previous section, we have discussed light diffraction by apertures. In this section, we will discuss the passage of light through an *ideal lens*. An ideal lens is a phase object. When an ideal focusing (or convex) lens has focal length  $f$ , its phase transformation function,  $t_f(x, y)$ , is given by

$$t_f(x, y) = \exp \left[ j \frac{k_0}{2f} (x^2 + y^2) \right], \quad (2.4-1)$$

where we have assumed that the ideal lens is infinitely thin. For a uniform plane wave incident upon the lens, the wavefront behind the lens is a converging spherical wave (for  $f > 0$ ) that converges ideally to a point source (a distance of  $z = f$ ) behind the lens. We can see that this is the case when we apply the Fresnel diffraction formula [see Eq. (2.3-12)] :

$$\psi_p(x, y, z = f) = \psi_{p0}(x, y) * h(x, y; z = f), \quad (2.4-2)$$

where  $\psi_{p0}(x, y)$  is now given by  $1 \times t_f(x, y)$ . The constant, 1, in front of  $t_f(x, y)$  signifies that we have a plane wave (of unit amplitude) incident. For example, if we have an incident *Gaussian beam* of the profile given by  $\exp[-a(x^2 + y^2)]$ , then  $\psi_{p0}(x, y)$  will be given by  $\exp[-a(x^2 + y^2)] \times t_f(x, y)$ . Let us now return to Eq. (2.4-2) where  $\psi_{p0}(x, y) = t_f(x, y)$ , and by using Eq. (2.3-12) we have

$$\begin{aligned} \psi_p(x, y; f) &= \exp(-jk_0 f) \frac{jk_0}{2\pi f} \iint t_f(x', y') \\ &\quad \times \exp \left\{ \frac{-jk_0}{2f} [(x - x')^2 + (y - y')^2] \right\} dx' dy' \\ &\propto \iint 1 \exp \left[ \frac{jk_0}{2f} (x'^2 + y'^2) \right] \end{aligned}$$

$$\begin{aligned} & \times \exp\left[\frac{-jk_0}{2f}(x'^2 + y'^2 - 2xx' - 2yy')\right] dx' dy' \\ & = \iint 1 \exp\left[\frac{jk_0}{f}(xx' + yy')\right] dx' dy', \end{aligned}$$

which is recognized to be proportional to a 2-D Fourier transform of 1, i.e.,  $\delta(x, y)$ .

Let us now investigate the effect of placing a transparency,  $t(x, y)$ , against the ideal lens, which is shown in Fig. 2.12. In general,  $t(x, y)$  is a complex function such that if a complex field,  $\psi_{i,p}(x, y)$ , is incident on it, then the field immediately behind the transparency-lens combination is

$$\psi_{i,p}(x, y) t(x, y) \exp[j\frac{k_0}{2f}(x^2 + y^2)].$$

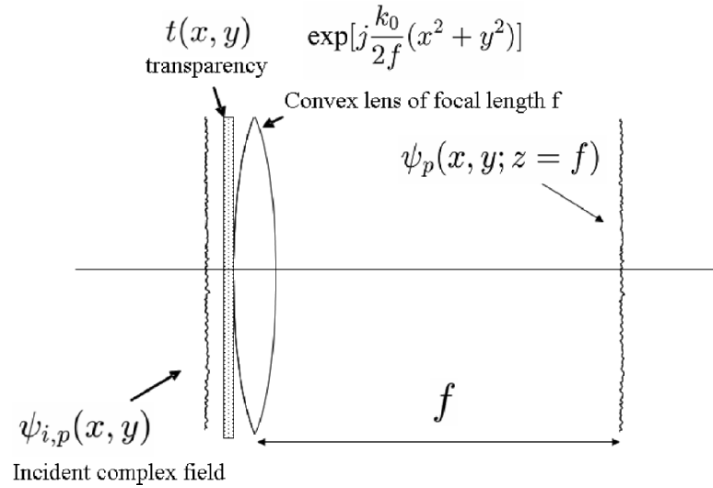


Fig. 2.12 A transparency immediately before an ideal lens under complex field illumination.

Again, for brevity, when illuminated by a unit-amplitude plane wave, the field immediately behind the combination is given by  $1 \times t(x, y) \exp[j\frac{k_0}{2f}(x^2 + y^2)]$ . We can then find the field distribution at a distance  $z = f$  by using the Fresnel diffraction formula, Eq. (2.3-12), as

$$\psi_p(x, y; z = f) = \exp(-jk_0 f) \frac{jk_0}{2\pi f} \exp\left[\frac{-jk_0}{2f}(x^2 + y^2)\right]$$

$$\begin{aligned}
& \times \iint t(x', y') \exp \left[ j \frac{k_0}{f} (xx' + yy') \right] dx' dy' \\
& = \exp(-jk_0 f) \frac{jk_0}{2\pi f} \exp \left[ \frac{-jk_0}{2f} (x^2 + y^2) \right] \\
& \times \mathcal{F}_{xy} \left\{ t(x, y) \right\} \bigg|_{\substack{k_x = k_0 x/f \\ k_y = k_0 y/f}}, \tag{2.4-3}
\end{aligned}$$

where  $x$  and  $y$  denote the transverse coordinates at  $z = f$ . Hence, the complex field on the focal plane ( $z = f$ ) is proportional to the Fourier transform of  $t(x, y)$ , but has *phase curvature* term  $\exp[\frac{-jk_0}{2f}(x^2 + y^2)]$ . Note that if  $t(x, y) = 1$ , i.e., the transparency is completely clear, then we have  $\psi_p(x, y, z = f) \propto \delta(x, y)$ , which corresponds to the focusing of a plane wave by a lens, as discussed earlier.

### Example 2.5 Transparency in front of a Lens

Suppose that a transparency,  $t(x, y)$ , is located at a distance,  $d_0$ , in front of an ideal convex lens and is illuminated by a plane wave with a unit strength shown in Fig. 2.13. The physical situation is shown in Fig. 2.13a), which can be represented by a block diagram given by Fig. 2.13b). According to the block diagram, we write

$$\psi_p(x, y; f) = \{ [t(x, y) * h(x, y; d_0)] t_f(x, y) \} * h(x, y; f), \tag{2.4-4}$$

which, apart from some constant, can be evaluated to obtain

$$\begin{aligned}
\psi_p(x, y; f) & = \frac{jk_0}{2\pi f} \exp[-jk_0(d_0 + f)] \exp \left[ -j \frac{k_0}{2f} \left(1 - \frac{d_0}{f}\right) (x^2 + y^2) \right] \\
& \times \mathcal{F}_{xy} \{ t(x, y) \} \bigg|_{\substack{k_x = k_0 x/f \\ k_y = k_0 y/f}}. \tag{2.4-5}
\end{aligned}$$

As in Eq. (2.4-3), note that a phase curvature factor as a function of  $x$  and  $y$  again precedes the Fourier transform, which represents the phase error if one wishes to compute the optical Fourier transformation. However, the phase curvature vanishes for the special case of  $d_0 = f$ . That is, from Eq. (2.4-5) and by disregarding some inessential constant, we now have

$$\psi_p(x, y; f) = \mathcal{F}_{xy}\{t(x, y)\} \Big|_{\substack{k_x=k_0x/f \\ k_y=k_0y/f}} = T(k_0x/f, k_0y/f). \quad (2.4-6)$$

Therefore, when the transparency is placed in the front focal plane of the convex lens, the phase curvature disappears and we recover the exact Fourier transform on the back focal plane. *Fourier-plane processing* of an “input” transparency located on the front focal plane can now be performed on the back focal plane. This is the essence of Fourier optics to perform *coherent image processing*.

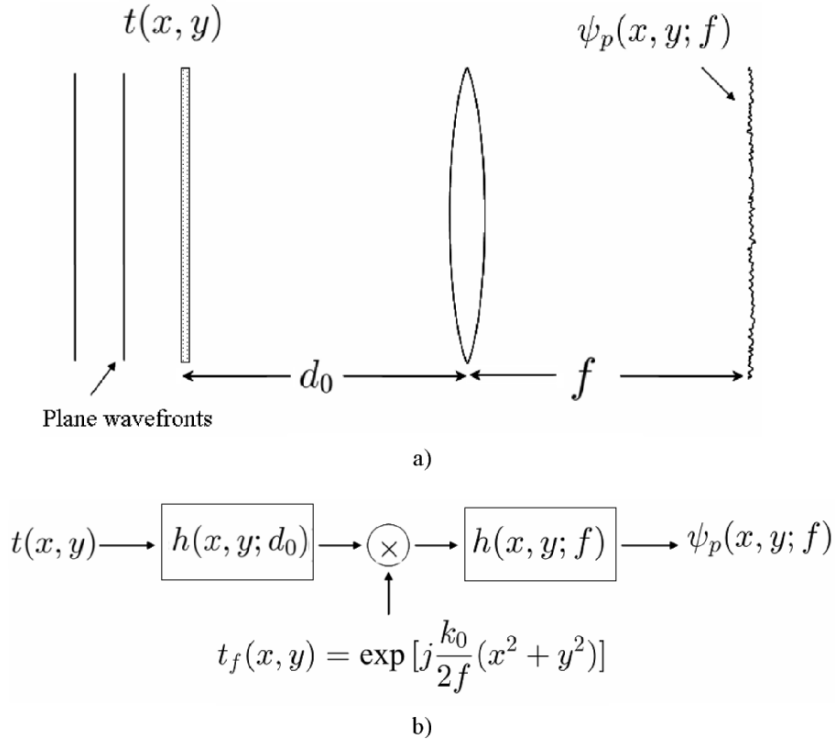


Fig. 2.13 Plane-wave illumination of a transparency  $t(x, y)$  located a distance  $d_0$  in front of a convex lens of focal length  $f$ : a) Physical situation, b) Block diagram.

### 2.4.2 Coherent Image Processing

The two-lens system is traditionally attractive for coherent image processing because, in the configuration shown in the Fig. 2.14, the Fourier transform of the input transparency,  $t(x, y)$ , appears on the common focal plane, or *Fourier plane*. In order to perform Fourier-plane processing of the input

transparency, we can insert a transparency on the Fourier plane that will suitably modify the Fourier transform of the input transparency.

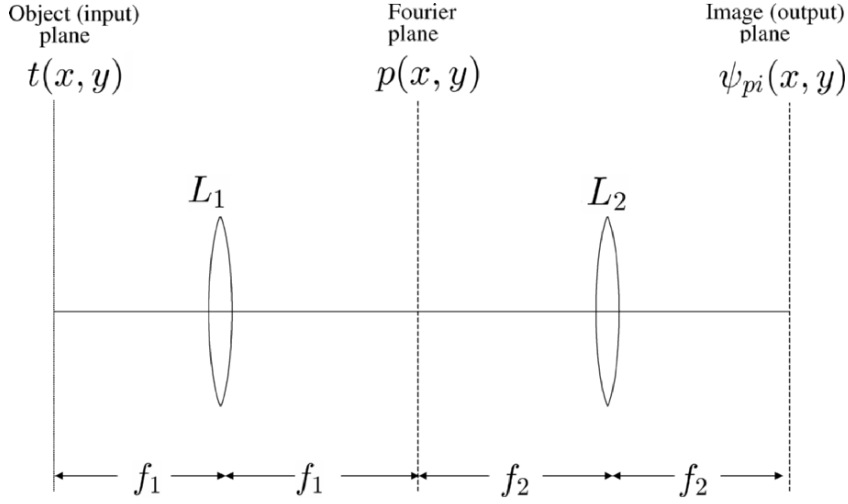


Fig. 2.14 Standard two-lens imaging processing system.

The Fourier plane transparency is commonly called a *spatial filter*,  $p(x, y)$ . According to Eq. (2.4-6), when a transparency  $t(x, y)$  is placed on the front focal plane of lens  $L_1$  as shown in Fig. 2.14, the field distribution on the common focal plane is given by  $T(k_0x/f_1, k_0y/f_1)$ , where we assume that the transparency is illuminated by a plane wave. After this field distribution is modified by the spatial filter, we can finally find the field distribution on the back focal plane of lens  $L_2$ ,  $\psi_{pi}$ , by again using Eq. (2.4-6) and neglecting some constant, as

$$\psi_{pi}(x, y) = \mathcal{F}_{xy} \left\{ T(k_0x/f_1, k_0y/f_1) p(x, y) \right\} \Big|_{\substack{k_x=k_0x/f_2 \\ k_y=k_0y/f_2}},$$

which can be evaluated, in terms of convolution, to give

$$\begin{aligned} \psi_{pi}(x, y) &= t(x/M, y/M) * \mathcal{F}_{xy} \left\{ p(x, y) \right\} \Big|_{\substack{k_x=k_0x/f_2 \\ k_y=k_0y/f_2}} \\ &= t(x/M, y/M) * P\left(\frac{k_0x}{f_2}, \frac{k_0y}{f_2}\right), \end{aligned} \quad (2.4-7)$$

where  $M = -f_2/f_1$  is the magnification factor and  $P$  is the Fourier transform of  $p$ . By comparing Eq. (2.4-7) with Eq. (1.2-2), we can describe the impulse response of the two-lens system, or the *coherent point spread function* (CPSF), as



$$h_c(x, y) = \mathcal{F}_{xy} \left\{ p(x, y) \right\} \Big|_{\substack{k_x = k_0 x / f_2 \\ k_y = k_0 y / f_2}} = P\left(\frac{k_0 x}{f_2}, \frac{k_0 y}{f_2}\right). \quad (2.4-8)$$

$p(x, y)$  is often called the *pupil function* of the system. We can see that the coherent PSF is given by the Fourier transform of the pupil function as shown in Eq. (2.4-8). By definition, the corresponding *coherent transfer function* is the Fourier transform of the coherent PSF:

$$\begin{aligned} H_c(k_x, k_y) &= \mathcal{F}_{xy} \left\{ h_c(x, y) \right\} \\ &= \mathcal{F}_{xy} \left\{ P\left(\frac{k_0 x}{f_2}, \frac{k_0 y}{f_2}\right) \right\} = p\left(\frac{-f_2 k_x}{k_0}, \frac{-f_2 k_y}{k_0}\right). \end{aligned} \quad (2.4-9)$$

We observe that *spatial filtering* is directly proportional to the functional form of the pupil function in coherent image processing.

The complex field on the image plane can then be written as

$$\psi_{pi}(x, y) \propto t(x/M, y/M) * h_c(x, y), \quad (2.4-10)$$

and hence the corresponding *image intensity* is

$$I_i(x, y) = |\psi_{pi}(x, y)|^2 \propto |t(x/M, y/M) * h_c(x, y)|^2. \quad (2.4-11)$$

### 2.4.3 Incoherent Image Processing

So far, we have discussed that the illumination of an object is *spatially coherent* - an example being the use of a laser. This means that the complex amplitudes of light falling on all parts of an object vary in unison, meaning that any two points on an object receive light that has a fixed relative phase and does not vary with time. On the other hand, an object may be illuminated with light having the property that the complex amplitudes on all parts of the object vary randomly, so that any two points on the object receive light of illumination is termed *spatially incoherent*. Light from extended sources, such as fluorescent tube lights, is incoherent. As it turns out, a *coherent system is linear with respect to the complex fields* and hence Eqs. (2.4-10) and (2.4-11) hold for *coherent optical systems*. On the other hand, *an incoherent optical system is linear with respect to the intensities*. To find the image intensity, we perform convolution with the given intensity quantities as follows:

$$I_i(x, y) \propto |t(x/M, y/M)|^2 * |h_c(x, y)|^2. \quad (2.4-12)$$

This equation is the basis for incoherent image processing.  $|h_c(x, y)|^2$  is the impulse response of the incoherent optical system and is often called the *intensity point spread function* (IPSF) of the optical system. Note that the IPSF is real and non-negative, which particularly means that it is not possible to directly implement even the simplest enhancement and restoration algorithms (e.g., highpass, derivative, etc.), which requires a *bipolar PSF* [Lohmann and Rhodes (1978)].

As usual, the Fourier transform of an impulse response will give a transfer function known as the *optical transfer function* (OTF) of the incoherent imaging system. For this case, it is given by

$$OTF(k_x, k_y) = \mathcal{F}_{xy}\{|h_c(x, y)|^2\} = H_c(k_x, k_y) \otimes H_c(k_x, k_y), \quad (2.4-13)$$

which can be explicitly written in terms of the coherent transfer function  $H_c$ :

$$OTF(k_x, k_y) = \int \int H_c^*(k'_x, k'_y) H_c(k'_x + k_x, k'_y + k_y) dk'_x dk'_y. \quad (2.4-14)$$

Note that one of the most important properties of the OTF, which follows a property of correlation, is that

$$|OTF(k_x, k_y)| \leq |OTF(0, 0)|. \quad (2.4-15)$$

This property states that the OTF always has a central maximum, which always signifies *lowpass filtering* regardless of the pupil function used in the system [Lukosz (1962)].

### Example 2.6 Coherent Transfer Functions and OTFs

Consider a two-lens system as shown in Fig. 2.14 with  $f_1 = f_2 = f$  and  $p(x, y) = \text{rect}(x/X)$ , i.e., a slit of width  $X$  along the  $y$ -direction. Using Eq. (2.4-9), the coherent transfer function becomes

$$\begin{aligned} H_c(k_x, k_y) &= \text{rect}\left(\frac{x}{X}\right) \Big|_{x=-fk_x/k_0} \\ &= \text{rect}\left(\frac{k_x}{Xk_0/f}\right), \end{aligned} \quad (2.4-16)$$

which is plotted in Fig. 2.15a). Now, the OTF is the autocorrelation of  $H_c$  as calculated by Eq. (2.4-13) and is plotted in Fig. 2.15b). Observe that both

situation perform *lowpass filtering* of spatial frequencies on an input image. Under incoherent illumination, it is possible to transmit twice the range of spatial frequency of an image as compared to the use of coherent illumination. However, the spectrum of an image transmitted through the pass-band is modified by the shape of the OTF.

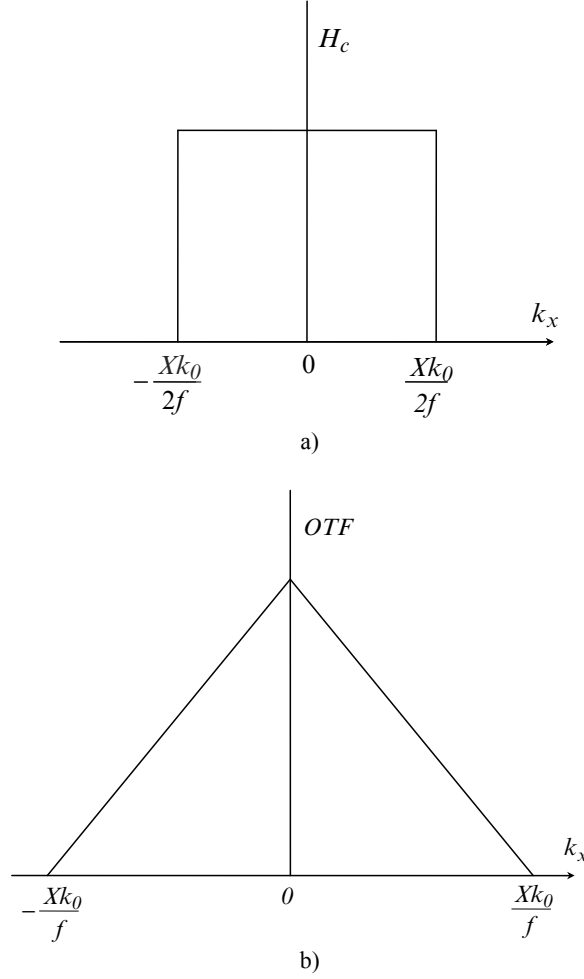


Fig. 2.15 a) The coherent transfer function, and b) the OTF for the pupil function  $p(x, y) = \text{rect}(x/X)$ .

Now, let us consider

$$p(x, y) = \left[ \text{rect}\left(\frac{x - x_0}{X}\right) + \text{rect}\left(\frac{x + x_0}{X}\right) \right], \quad x_0 > \frac{X}{2},$$

which is a two-slit object aligned along the  $y$ -direction. The coherent transfer function is

$$H_c(k_x, k_y) = \left[ \text{rect}\left(\frac{x - x_0}{X}\right) + \text{rect}\left(\frac{x + x_0}{X}\right) \right] \Big|_{x = -f k_x / k_0}$$

$$= \left[ \text{rect}\left(\frac{-k_x - x_0 k_0 / f}{X k_0 / f}\right) + \text{rect}\left(\frac{-k_x + x_0 k_0 / f}{X k_0 / f}\right) \right].$$

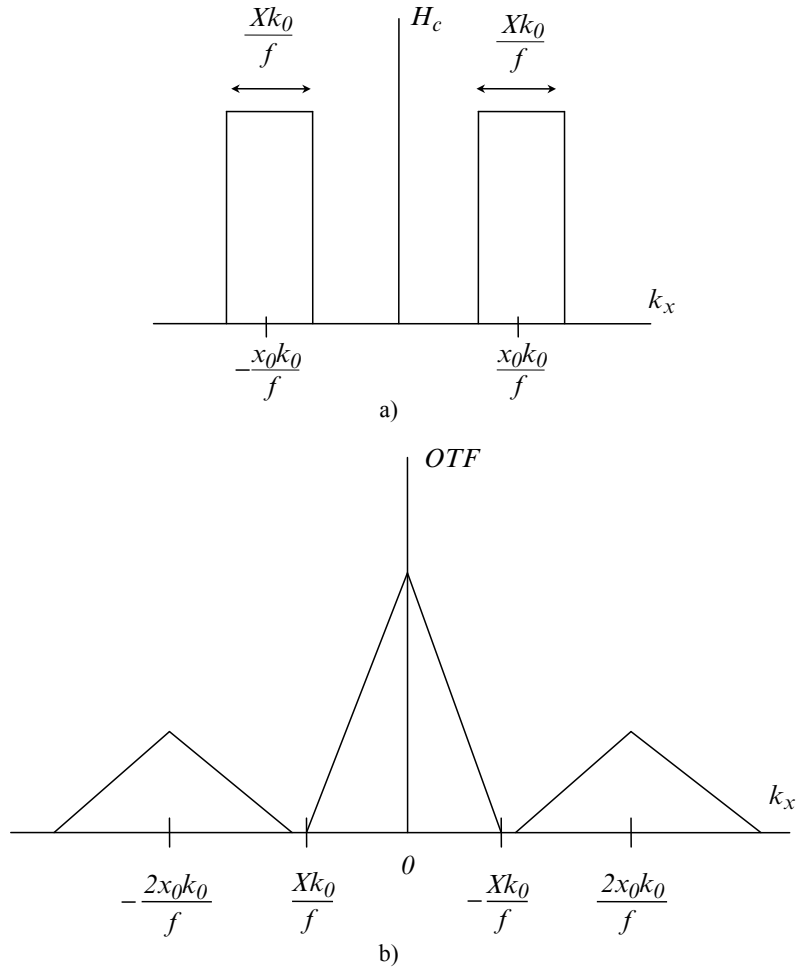


Fig. 2.16 a) The coherent transfer function, and b) the OTF for the pupil function  $p(x, y) = \text{rect}[(x - x_0)/X] + \text{rect}[(x + x_0)/X]$ .

We plot  $H_c(k_x, k_y)$  in Fig. 2.16a) along with the OTF in Fig. 2.16b). Note that even though it may be possible to achieve *band-pass filtering* with

coherent illumination, incoherent processing always gives rise to inherently low-pass characteristics because its point spread function is real and positive [see Eq. (2.4-13)]. A large amount of attention has been focused on devising methods to realize band-pass characteristics by using novel incoherent image processing techniques [see, e.g., Lohmann and Rhodes (1978), Stoner (1978), Poon and Korpel (1979), Mait (1987)], where the synthesis of bipolar or even complex point spread functions (PSFs) in incoherent optical systems is possible. Such techniques are called *bipolar incoherent image processing*. The article by Indebetouw and Poon [1992] provides a comprehensive review of bipolar incoherent image processing.

## 2.5 Holography

### 2.5.1 Fresnel Zone Plate as a Point-Source Hologram

A photograph is a 2-D recording of a 3-D scene. What is actually recorded is the light intensity at the plane of the photographic recording film - the film being light sensitive only to the intensity variations. Hence, the developed film's amplitude transparency is  $t(x, y) \propto I(x, y) = |\psi_p|^2$ , where  $\psi_p$  is the complex field on the film. As a result of this intensity recording, all the information on the relative phases of light waves from the original 3-D scene is lost. This loss of phase information on the light field destroys the 3-D character of the scene, i.e., we cannot change the perspective of the image in the photograph by viewing it from a different angle (i.e., *parallax*) and we cannot interpret the depth of the original 3-D scene.

As an example, let us take the photographic recording of a point source located at the origin, but with a distance of  $z_0$  away from the film. The situation is shown in Fig. 2.17a). Now, according to Eq. (2.3-15), the complex field just before the film is given by

$$\begin{aligned}\psi_p(x, y; z_0) &= \delta(x, y) * h(x, y; z_0) \\ &= \exp(-jk_0 z_0) \frac{jk_0}{2\pi z_0} \exp\left[-\frac{jk_0(x^2 + y^2)}{2z_0}\right].\end{aligned}$$

Hence, the developed film's amplitude transparency is

$$t(x, y) \propto I(x, y) = |\psi_p(x, y; z_0)|^2 = \left(\frac{k_0}{2\pi z_0}\right)^2. \quad (2.5-1)$$

Note that the phase information of  $\psi_p(x, y; z_0)$  is completely lost. Now, for a point source located at  $(x_0, y_0)$ , as shown in Fig. 2.17b), the complex field just before the film is given by

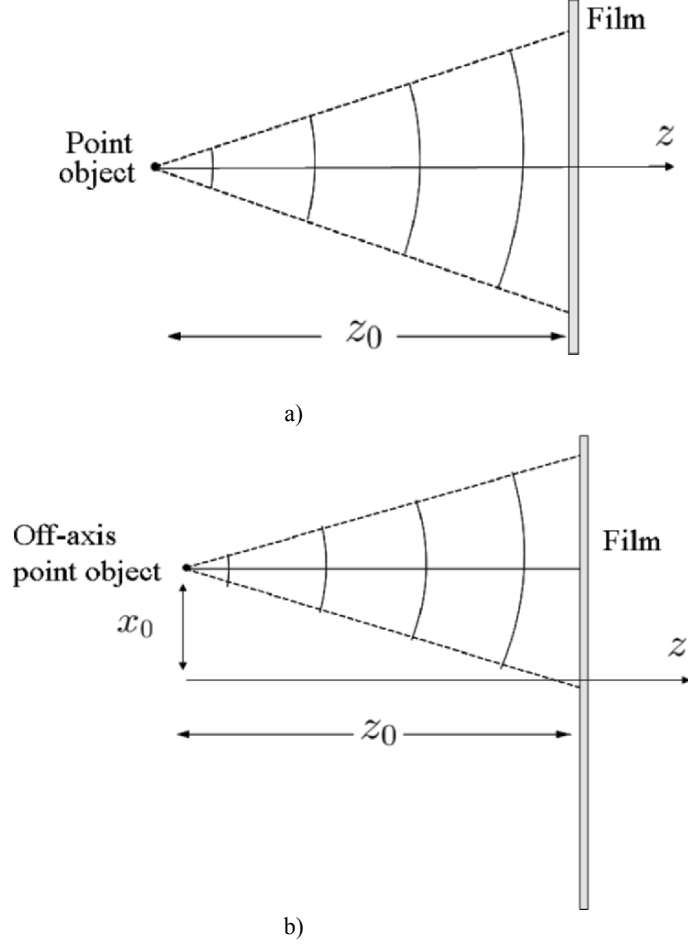


Fig. 2.17 Photographic recording of a point source: a) located at the origin  $(0, 0)$ , and b) located at  $(x_0, y_0)$ , both  $z_0$  away from the film.

$$\begin{aligned}\psi_p(x, y; x_0, y_0, z_0) &= \delta(x - x_0, y - y_0) * h(x, y; z_0) \\ &= \exp(-jk_0 z_0) \frac{jk_0}{2\pi z_0} \exp\left[-\frac{jk_0[(x - x_0)^2 + (y - y_0)^2]}{2z_0}\right],\end{aligned}$$

and what is recorded is

$$t(x, y) \propto I(x, y) = |\psi_p(x, y; x_0, y_0, z_0)|^2 = \left(\frac{k_0}{2\pi z_0}\right)^2, \quad (2.5-2)$$

which is identical to the result given by Eq. (2.5-1). Again the phase information of  $\psi_p(x, y; x_0, y_0, z_0)$  is lost, and we also notice that the 3-D location of the point source, i.e.,  $x_0, y_0$ , and  $z_0$ , is mostly lost.

Holography is an extraordinary technique that was invented by Gabor [1948], where not only the amplitude, but also the phase of a light field can be recorded. The word “holography” combines parts of two Greek words: *holos*, meaning “complete,” and *graphein*, meaning “to record.” Thus, holography means the recording of complete information. Hence, in the holographic process, the film records both the amplitude and phase of a light field. The resulting recorded film is called a “*hologram*.” When a hologram is properly illuminated, an exact replica of the original 3-D wave field is reconstructed. We shall discuss the *holographic recording* of a point object as an example. Once we know how a single point is recorded, the recording of a complicated object can be regarded as the recording of a collection of points.

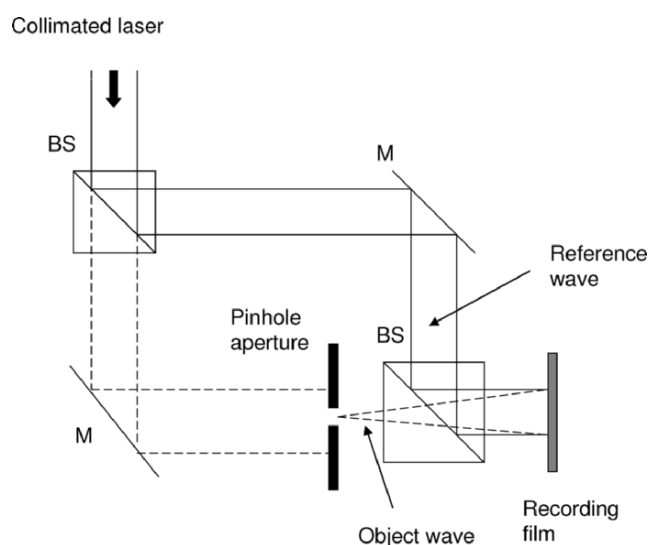


Fig. 2.18 Holographic recording of a point source object.

Figure 2.18 shows a collimated laser which is split into two plane waves and then recombined by using two mirrors (M) and two beam splitters (BS). One plane wave is used to illuminate the pinhole aperture (our point object), and the other is used to illuminate directly the recording film. The plane wave that is diffracted by the pinhole aperture generates a diverging spherical wave. In holography, this diverging wave is known as an *object wave*. The plane wave that directly illuminates the photographic plate is known as a *reference wave*. Let  $\psi_o$  represent the field distribution of the object wave on the plane of the recording film, and similarly, let  $\psi_r$  represent the field distribution of the reference wave on the plane of the recording film. The film now records the interference of the reference wave and the object wave, i.e., what is recorded is given by  $|\psi_r + \psi_o|^2$ , provided that the reference wave and the object wave are mutually coherent over the film. The

coherency of the light waves is guaranteed by the use of a laser source and ensures that the difference between the two paths is less than the coherent length of the laser. This kind of recording is known as *holographic recording*, and is distinct from a photographic recording where the reference wave does not exist and, therefore, only the object wave is recorded.

Let us now consider the recording of an off-axis point object at a distance of  $z_0$  from the recording film. The pinhole aperture is then modeled as  $\delta(x - x_0, y - y_0)$ . According to Fresnel diffraction, the object wave arises from the point object on the film and is given by

$$\begin{aligned}\psi_o &= \delta(x - x_0, y - y_0) * h(x, y; z_0) \\ &= \exp(-jk_0 z_0) \frac{jk_0}{2\pi z_0} \exp\{-jk_0[(x - x_0)^2 + (y - y_0)^2]/2z_0\}.\end{aligned}$$

This object wave is a *spherical wave*.

For the reference plane wave, we assume that the plane wave has the same initial phase as the point object at a distance of  $z_0$  away from the film. Therefore, its field distribution on the film is  $\psi_r = a \exp(-jk_0 z_0)$ , where  $a$  is the amplitude of the plane wave. Hence, the intensity distribution that is being recorded on the film, or the transmittance of the hologram, is given by

$$\begin{aligned}t(x, y) &\propto |\psi_r + \psi_o|^2 \\ &= |a + \frac{jk_0}{2\pi z_0} \exp\{-jk_0[(x - x_0)^2 + (y - y_0)^2]/2z_0\}|^2 \\ &= A + B \sin\{\frac{k_0}{2z_0}[(x - x_0)^2 + (y - y_0)^2]\} \\ &= FZP(x - x_0, y - y_0; z_0)\end{aligned}\tag{2.5-3}$$

where  $A = a^2 + (\frac{k_0}{2\pi z_0})^2$ ,  $B = \frac{k_0}{\pi z_0}$  and  $k_0 = 2\pi/\lambda_0$ .

The expression in Eq. (2.5-3) is called the sinusoidal *Fresnel zone plate (FZP)*, which is a hologram of the point source object. Note that the center of the zone plate specifies the location,  $x_0$  and  $y_0$ , of the point object, and the spatial variation of the zone plate is governed by a sine function with a quadratic spatial dependence. For an on-axis point source, i.e.,  $x_0 = y_0 = 0$  in Eq. (2.5-3), located a distance of  $z_0$  away from the film, we have an on-axis Fresnel zone plate as



$$\begin{aligned}
t(x, y) &\propto |\psi_r + \psi_o|^2 \\
&= |a + \frac{jk_0}{2\pi z_0} \exp[-\frac{jk_0}{2z_0}(x^2 + y^2)]|^2 \\
&= A + B \sin[\frac{k_0}{2z_0}(x^2 + y^2)] \\
&= FZP(x, y; z_0).
\end{aligned} \tag{2.5-4}$$

Let us now investigate the quadratic spatial dependence of  $FZP(x, y; z_0)$ . The spatial rate of change of the phase on the zone plate along the  $x$ -direction is

$$f_{local} = \frac{1}{2\pi} \frac{d}{dx} (\frac{k_0}{2z_0} x^2) = \frac{x}{\lambda_0 z_0}. \tag{2.5-5}$$

This is a local fringe frequency that increases linearly with the spatial coordinate,  $x$ . The farther it is away from the origin of the zone, the higher the frequency will be. So, for a fixed point (local) on the hologram, we can deduce the depth information,  $z_0$ , by finding the local fringe frequency for a given wavelength of light,  $\lambda_0$ . Therefore, we see that the depth information is encoded within the phase of the FZP. Figure 2.19 shows the dependence of the Fresnel zone plate characteristic as a function of the depth parameter  $z$  (for  $z = z_0$  and  $2z_0$ ). As the point source becomes further away from the recording film, the recorded FZP has a lower local fringe frequency.

Figure 2.20 shows us that as the point source moves to a new location  $x_0, y_0$ , the center of the zone plate translates accordingly. Hence, we see that the zone contains the complete 3-D information of the point source. The center of the zone,  $x_0$  and  $y_0$ , defines the transverse location of the point object, and the fringe variation defines the depth location,  $z_0$ . Table 2.2 shows the MATLAB code used to generate Fresnel zone plates that are presented in Figs. 2.19 and 2.20. For an arbitrary 3-D object, we can think of the object as a collection of points, and therefore, we can envision that we have a collection of zones on the hologram, where each zone carries the transverse location as well as the depth information of each individual point. In fact, a hologram has been considered as a type of Fresnel zone plate, and the holographic imaging process has been discussed previously in terms of zone plates [Rogers (1950), Siemens-Wapniarski and Parker Givens (1968)].

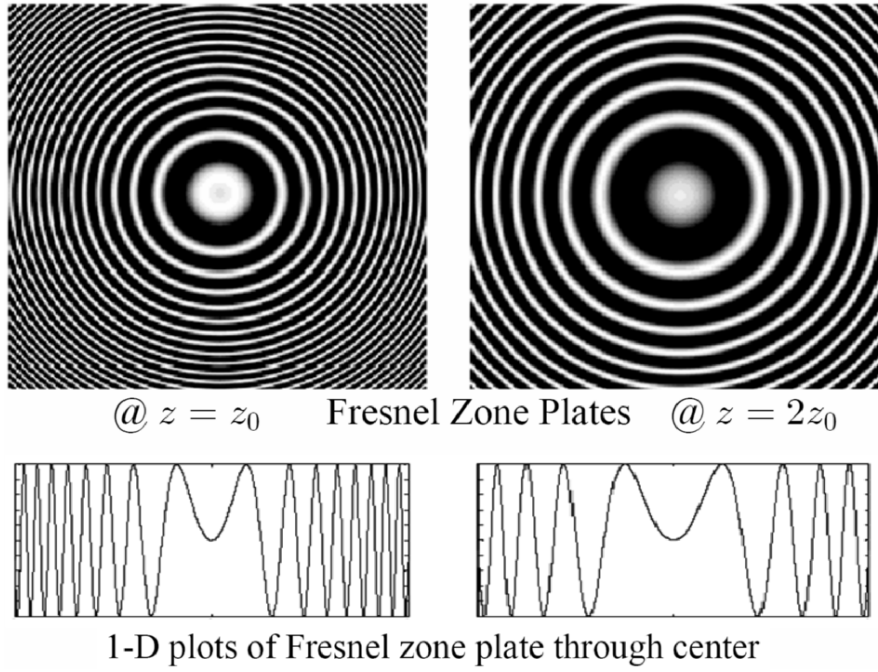


Fig. 2.19 On-axis Fresnel zone plate as a function of depth,  $z$ .

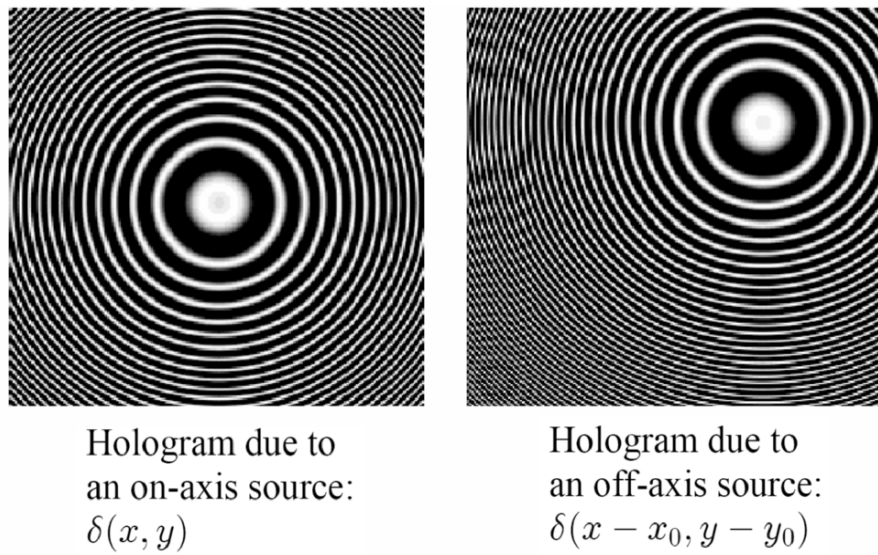


Fig. 2.20 Fresnel zone plate due to point sources at different transverse locations but otherwise located at the same depth  $z_0$ .

Table 2.2 Fresnel\_zone\_plate.m:  
m-file for calculating FZPs illustrated in Figs. 2.19 and 2.20.

```
-----
%Fresnel_zone_plate.m
%Adapted from "Contemporary optical image processing with MATLAB®,"
%by T.-C. Poon and P. P. Banerjee, Elsevier 2001, pp.177-178.
%
%display function is 1+sin(sigma*((x-x0)^2+(y-y0)^2)). All scales are arbitrary.
%sigma=pi/(wavelength*z)
clear;

z0=input('z0, distance from the point object to film, enter z0 (from 2 to 10)=');
x0=input('Inputting the location of the center of the FZP x0=y0,enter x0 (from -8 to 8) =');

ROWS=256;
COLS=256;
colormap(gray(255))
sigma=1/z0;
y0=-x0;
y=-12.8;

for r=1:COLS,
    x=-12.8;
    for c=1:ROWS, %compute Fresnel zone plate
        fFZP(r,c)=exp(j*sigma*(x-x0)*(x-x0)+j*sigma*(y-y0)*(y-y0));
        x=x+.1;
    end
    y=y+.1;
end

%normalization
max1=max(fFZP);
max2=max(max1);
scale=1.0/max2;
fFZP=fFZP.*scale;
R=127*(1+imag(fFZP));
figure(1)
image(R);
axis square on
axis off
-----
```

So far, we have discussed the transformation of a point object to a zone plate on the hologram, which corresponds to a *recording* or *coding process*. In order to retrieve the point object from the hologram, we need a *reconstruction* or *decoding process*. This can be done by simply illuminating the hologram with a *reconstruction wave*. Figure 2.21 corresponds to the reconstruction of a hologram of the point object located on-axis, i.e., the reconstruction of the hologram given by Eq. (2.5-4).

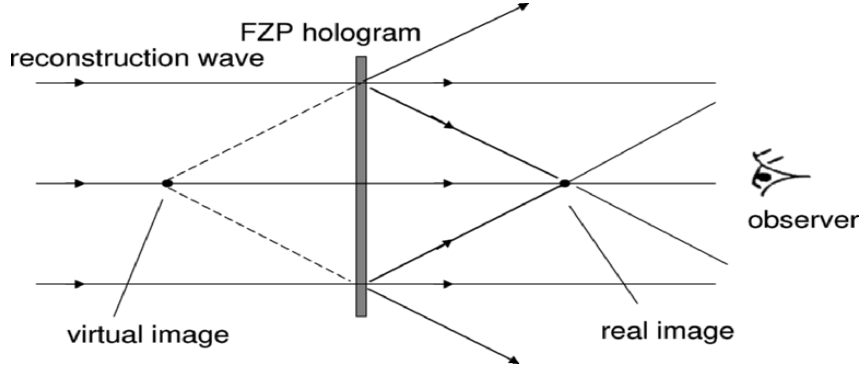


Fig. 2.21 Holographic reconstruction of a point source object.

Note that in practice as is shown in Fig. 2.21, the reconstruction wave is usually identical to the reference wave. Therefore, we take the reconstruction wave to have a field distribution on the plane of the hologram given by  $\psi_{rc}(x, y) = a$ . Hence, the field distribution of the transmitted wave immediately after the hologram is  $\psi_{rc}t(x, y) = at(x, y)$  and the field at arbitrary distance of  $z$  away, according to Fresnel diffraction, is given by the evaluation of

$$at(x, y) * h(x, y; z).$$

For the point-object hologram given by (2.5-4), after we expand the sine term of the hologram  $t(x, y)$ , we obtain

$$t(x, y) = A + \frac{B}{2j} \left\{ \exp\left[j \frac{k_0}{2z_0} (x^2 + y^2)\right] - \exp\left[-j \frac{k_0}{2z_0} (x^2 + y^2)\right] \right\}.$$

Therefore, as a result of the illumination of the hologram by the reconstruction wave, we have three waves. These waves, according to the convolution operation,  $at(x, y) * h(x, y; z)$ , are as follows:

*Zero-order beam:*

$$aA * h(x, y; z = z_0) = aA. \quad (2.5-6a)$$

*Real image (or the twin image):*

$$\sim \exp\left[j \frac{k_0}{2z_0} (x^2 + y^2)\right] * h(x, y; z = z_0) \sim \delta(x, y). \quad (2.5-6b)$$

*Virtual image:*

$$\sim \exp\left[-j\frac{k_0}{2z_0}(x^2 + y^2)\right] * h(x, y; z = -z_0) \sim \delta(x, y). \quad (2.5-6c)$$

By writing Eq. (2.5-6c), we have back-propagated the field immediately behind the hologram by a distance of  $z_0$  to demonstrate that a virtual image will form behind the hologram. As illustrated in Fig. 2.21, optical fields from this virtual image correspond to a diverging wave behind the hologram. We notice that the zero-order beam is caused by the bias in the hologram and the virtual image is the reconstructed original point object. The real image is located at a distance of  $z_0$  in front of the hologram, which is known as the *twin image*.

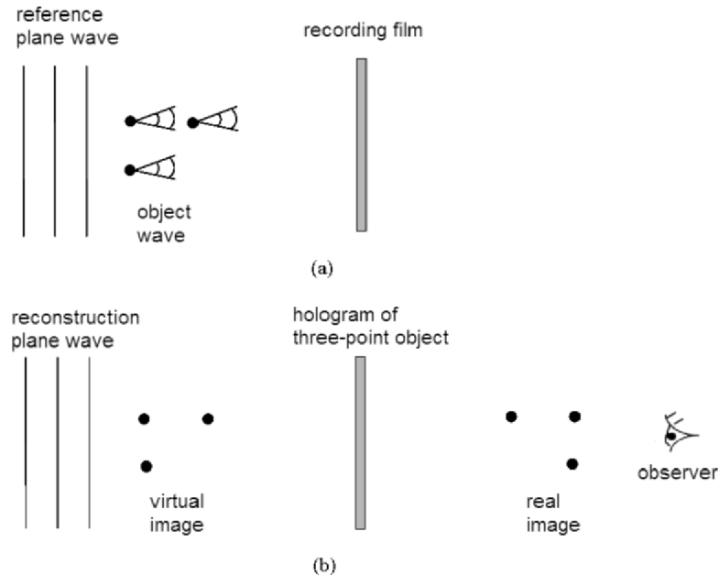


Fig. 2.22 Holographic recording and reconstruction of a three-point object.

Figure 2.22 shows the holographic recording of a 3-point object and its reconstruction. Note that the virtual image appears at the correct 3-D location as the original object, while the real image (the twin image) is the mirror-image of the original object, with the axis of reflection on the plane of the hologram.

### 2.5.2 Off-Axis Holography

In the last section, we discussed the so-called *on-axis holography*. The term “on-axis” refers to the use of a reference wave that is coaxially illuminating the hologram with the object wave. Although this technique can record 3-D

information of an object, it also create an annoying effect when we view the reconstructed virtual image. The real image (or the twin image) will also be reconstructed along the viewing direction [see Figs. 2.21 and 2.22]. In holography, this is infamously known as the “*twin-image problem*.”

*Off-axis holography* is a method that was devised by Leith and Upatnieks [1964] to separate the twin-image and the zero-order beam from the desired image. To achieve off-axis recording, the reference plane wave will need to be incident on the recording film off-axis. Referring back to Fig. 2.18, this can be done by simply, for example, rotating the beamsplitter (BS) between the pinhole aperture and the film in a clockwise direction so that the reference plane wave is incident on the film at an angle. The situation is shown in Fig. 2.23, where the plane reference wave is incident at an angle,  $\theta$ .  $\theta$  is called the *recording angle* in off-axis holographic recording.

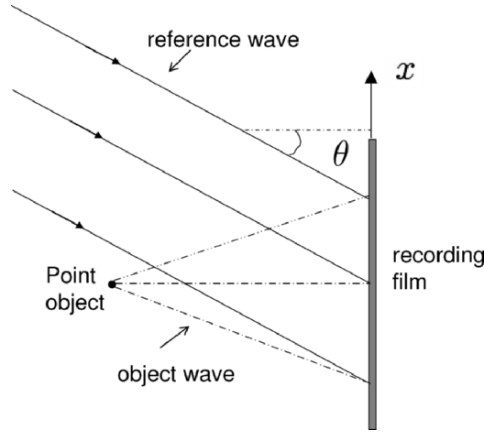


Fig. 2.23 Recording with off-axis reference plane wave.  
The point object is  $z_0$  away from the film.

For off-axis recording, we have  $t(x, y) = |\psi_r + \psi_o|^2$ , where the reference plane wave,  $\psi_r$ , is now an off-axis plane wave given by  $a \exp(jk_0 x \sin \theta)$ . The object wave,  $\psi_o$ , is the spherical wave generated by the on-axis point source. Similar to Eq. (2.5-3), where  $x_0 = y_0 = 0$  for an on-axis point object, we now have

$$\begin{aligned}
 t(x, y) &= \left| a \exp(jk_0 x \sin \theta) + \frac{jk_0}{2\pi z_0} \exp[-jk_0(x^2 + y^2)/2z_0] \right|^2 \\
 &= A + B \sin\left[\frac{k_0}{2z_0}(x^2 + y^2) + k_0 x \sin \theta\right], \quad (2.5-7)
 \end{aligned}$$

where  $A = a^2 + \left(\frac{k_0}{2\pi z_0}\right)^2$  and  $B = \frac{ak_0}{\pi z_0}$ .  $t(x, y)$  given by Eq. (2.5-7) is called

an *off-axis hologram*. Eq. (2.5-7) can be expanded into three terms as

$$t(x, y) = A + \frac{B}{2j} \left\{ \exp \left[ j \left[ \frac{k_0}{2z_0} (x^2 + y^2) + k_0 x \sin \theta \right] \right] \right. \\ \left. - \exp \left[ -j \left[ \frac{k_0}{2z_0} (x^2 + y^2) + k_0 x \sin \theta \right] \right] \right\}.$$

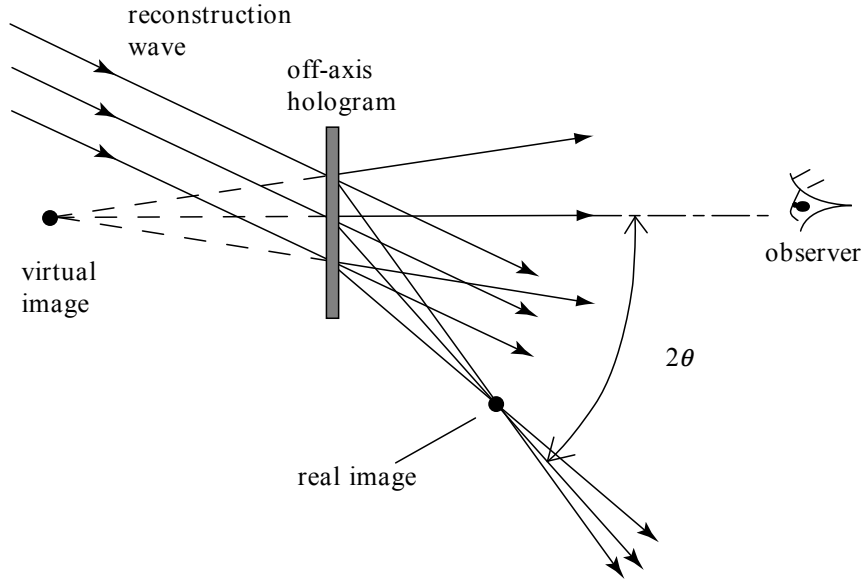


Fig. 2.24 Holographic reconstruction of off-axis hologram.  
The twin image (or the real image) is not observed if  $\theta$  is large enough.

By illuminating the hologram with a reconstruction wave identical to the reference wave, we have  $\psi_{rc}t(x, y)$  immediately after the hologram, where  $\psi_{rc} = a \exp(jk_0 x \sin \theta) = \psi_r$ . As in the case of on-axis holography, by performing Fresnel diffraction, we have  $\psi_{rc}t(x, y) * h(x, y; z)$ , thereby creating three waves as follows:

*Zero-order beam:*

$$A a \exp(jk_0 x \sin \theta) * h(x, y; z = z_0) \\ \sim \exp(jk_0 x \sin \theta). \quad (2.5-8a)$$

*Real image (or the twin image):*

$$a \exp(jk_0 x \sin \theta) \exp \left\{ j \left[ \frac{k_0}{2z_0} (x^2 + y^2) + k_0 x \sin \theta \right] \right\} * h(x, y; z = z_0) \\ \sim \delta(x + 2z_0 \sin \theta, y) . \quad (2.5-8b)$$

*Virtual image:*

$$a \exp(jk_0 x \sin \theta) \exp \left\{ -j \left[ \frac{k_0}{2z_0} (x^2 + y^2) + k_0 x \sin \theta \right] \right\} * h(x, y; z = -z_0) \\ \sim \delta(x, y) . \quad (2.5-8c)$$

The situation is depicted in Fig. 2.24.

### 2.5.3 Digital Holography

As discussed in the last section, in regards to off-axis holographic reconstruction, the three reconstructed beams propagate along different directions, and if the recording angle is sufficiently large, the virtual image can be viewed without any disturbances from the zero-order beam and the real image. This technique of off-axis recording is also known as *carrier-frequency holography*. We can re-write Eq. (2.5-7) as

$$t(x, y) = A + B \sin \left[ \frac{k_0}{2z_0} (x^2 + y^2) + 2\pi f_c x \right], \quad (2.5-9)$$

where  $f_c = k_0 \sin \theta / 2\pi = \sin \theta / \lambda_0$  is the *spatial carrier*. For realistic parameter values,  $\theta = 45^\circ$  and  $\lambda_0 = 0.6 \mu m$  for red laser light, we have  $\sin \theta / \lambda_0 \sim 1,000$  cycle/mm. This technique translates to a film resolution of at least 1000 lp/mm [or line-pair/mm] in order to employ this technique for holographic recording. Common holographic films have a resolution of about 5000 lp/mm. For comparison, standard black and white film resolution is about 80-100 lp/mm and color film is about 40-60 lp/mm. But can we use electronic devices such as CCD cameras for holographic recording? If we can do it, we can bypass the darkroom preparation of films and, therefore, we can perform real-time or electronic recording of holographic information. Some of the best CCD camera in the market, such as Canon D60 (3072x2048 pixels, 67.7 lp/mm,  $7.4 \mu m$  pixel size), could not record off-axis holograms efficiently because its resolution is about a couple of orders of magnitude worse than the resolution of holographic films. We can see that off-axis recording places a stringent resolution requirement on electronic recording



media. We can relax the resolution requirement by making the recording angle smaller, but this approach requires a very small recording angle that often makes it impractical. Because of this reason, on-axis holography seems to be prevalent in digital holography [Piestun, Shamir, Wekamp, and Bryngdahl (1997)]. On the other hand, twin-image problems need to be tackled when on-axis holography is employed. Indeed twin-image elimination is an important research topic [Poon et al. (2000)].

While we have discussed holographic recording electronically or digitally, for reconstruction, we can also perform it digitally. Once the holographic information is in the electronic or the digital domain, we can digitally evaluate Fresnel diffraction by performing the convolution, which is  $at(x, y) * h(x, y; z)$ , where  $a$  is some constant amplitude of the reconstruction beam,  $t(x, y)$  is the recorded hologram, and  $h(x, y; z)$  is the spatial impulse response in Fourier optics. For various values of  $z = z_1, z_2$ , etc., we can reconstruct different planes normal to the hologram. The whole 3-D volume of the object is then constructed plane by plane. The situation is shown in Fig. 2.25.

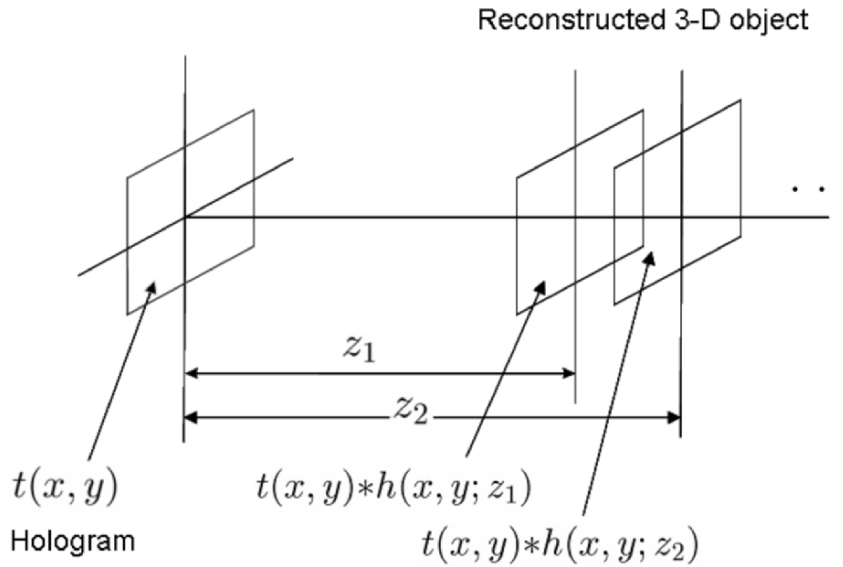


Fig. 2.25 Digital holographic reconstruction.

An alternative way to utilize electronically or digitally recorded hologram is to have it displayed on some sort of *spatial light modulator* (SLM) for real-time coherent reconstruction. A 2-D spatial light modulator is a device with which one can imprint a 2-D pattern on a laser beam by passing the laser beam through it (or by reflecting the laser beam off the device). A *liquid*

*crystal television* (LCTV) (upon suitably modification) is a good example of spatial light modulators. In fact, we can think of a spatial light modulator as a real-time transparency because one can update 2-D images or holograms upon the spatial light modulator in real time without developing films into transparencies. Again, off-axis holographic recording places stringent resolution requirement on SLMs and we will come back to this later when discussing 3-D display applications in Chapter 4.

All in all, in this section we mention electronic or digital recording and manipulation of holographic information. This type of research is commonly known as *digital* (or *electronic*) *holography*. The reader may find a pioneering contribution in the work of Goodman and Lawrence [1967]. Ever since, digital holography has become a practical tool with an increasing number of applications [Schnars and Juptner (2002)]. Most recently, an edited book on the subject organizes a collection of key chapters that covers digital holography and 3-D display techniques so as to provide the reader with the state-of-the-art developments in these important areas around the world [Poon (2006)]. Starting in the next chapter, we will discuss a unique electronic holographic recording technique called *optical scanning holography*.

## References

- 2.1 Banerjee, P.P. and T.-C. Poon (1991). *Principles of Applied Optics*. Irwin, Illinois.
- 2.2 Gabor, D. (1948). "A new microscopic principle," *Nature*
- 2.3 Goodman, J. W. and R.W. Lawrence (1967). "Digital image formation from electronically detected holograms," *Applied Physics Letters* 11, 77-79.
- 2.4 Goodman, J. W. (2005). *Introduction to Fourier Optics*. 3rd. ed., Roberts and Company Publishers, Englewood, Colorado.
- 2.5 Indebetouw, G. and T.-C. Poon (1992). "Novel approaches of incoherent image processing with emphasis on scanning methods," *Optical Engineering* 31, 2159–2167.
- 2.6 E. N. Leith, E.N and J. Upatnieks (1962). "Reconstructed wavefronts and communication theory," *Journal of the Optical Society of America* 52, 1123–1130.
- 2.7 Lohmann A.W. and W. T. Rhodes (1978). "Two-pupil synthesis of optical transfer functions," *Applied Optics* 17, 1141-1150.
- 2.8 Lukosz, W. (1962). "Properties of linear low pass filters for non-negative signals," *Journal of the Optical Society of America* 52, 827-829.
- 2.9 Mait, J. N. (1987). "Pupil-function design for complex incoherent spatial filtering," *Journal of the Optical Society of America A* 4, 1185- 1193.
- 2.10 Piestun R., J. Shamir, B. Wekamp, and O. Bryngdahl (1997). "On-axis computer-generated holograms for three-dimensional display," *Optics Letters* 22, 922-924.
- 2.11 Poon, T.-C. and A. Korpel (1979). "Optical transfer function of an acousto-optic heterodyning image processor," *Optics Letters* 4, 317-319.
- 2.12 Poon, T.-C., T. Kim, G. Indebetouw, B. W. Schilling, M. H. Wu, K. Shinoda, and Y. Suzuki (2000). "Twin-image elimination experiments for three-dimensional images in optical scanning holography," *Optics Letters* 25, 215-217.

- 2.13** Poon T.-C. and P. P. Banerjee (2001). *Contemporary Optical Image Processing with MATLAB®*. Elsevier, Oxford, UK.
- 2.14** Poon, T.-C., ed., (2006). *Digital Holography and Three-Dimensional Display: Principles and Applications*. Springer, New York, USA.
- 2.15** Rogers, G.L. (1950). "The black and white holograms," *Nature* 166, 1027.
- 2.16** Siemens-Wapniarski, W.J. and M. Parker Givens (1968). "The experimental production of synthetic holograms," *Applied Optics* 7, 535-538.
- 2.17** Stoner, W. (1978). "Incoherent optical processing via spatially offset pupil masks," *Applied Optics* 17, 2454-2466.
- 2.18** Schnars, U. and W. P. O. Juptner (2002). "Digital recording and numerical reconstruction of holograms," *Meas. Sci. Technol.* 13, R85–R101.