(a)

tion of the diffracting edge. The result is virtually the same as a diffraction pattern. From the graph it can be seen that the irradiance falls off rapidly and monotonically in the shadow zone ($v_2 < 0$) as $v_2 \rightarrow -\infty$. On the other hand, in the illuminated zone ($v_2 > 0$) the irradiance oscillates with diminishing amplitude about the unobstructed value U_0 as $v_2 \rightarrow +\infty$. The highest irradiance occurs just inside the illuminated region at the point $v_2 \approx 1.25$, where I_p is 1.37 times the irradiance of the unobstructed wave. This is seen as a bright fringe next to the geometrical shadow.

5.6 Applications of the Fourier Transform to Diffraction

Let us return to the discussion of Fraunhofer diffraction. We now consider the general problem of diffraction by an aperture having not only an arbitrary shape, but also an arbitrary transmission including phase retardation, which may vary over different parts of the aperture.


(b)

Figure 5.25. Fresnel diffraction by a straightedge. (a) Points on the Cornu spiral; (b) corresponding points on the intensity curve; $v = 0$ defines the geometrical shadow edge. A photograph of the diffraction pattern is shown below.
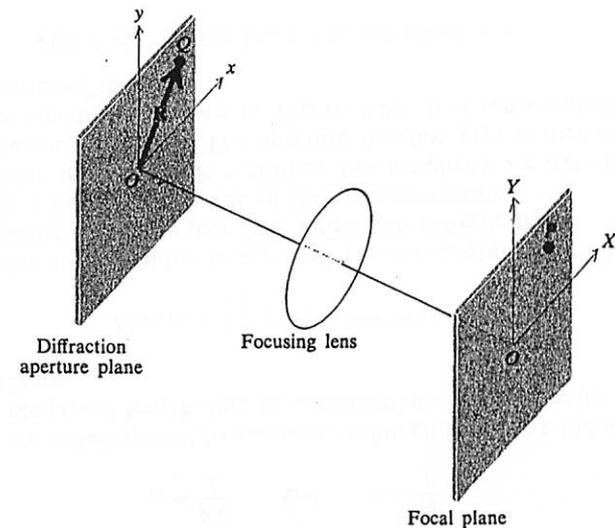
Figure 5.26. Geometry of the general diffraction problem.

We choose coordinates as indicated in Figure 5.26. The diffracting aperture lies in the xy plane, and the diffraction pattern appears in the XY plane, which is the focal plane of the focusing lens. According to elementary geometrical optics, all rays leaving the diffracting aperture in a given direction, specified by direction cosines α, β, and γ, are

brought to a common focus. This focus is located at the point $P(X,Y)$ where $X \approx L\alpha$ and $Y \approx L\beta$, L being the focal length of the lens. The assumption is made here that α and β are small, so that $\beta \approx \tan \beta$. We also assume that $\gamma \approx 1$.

Now the path difference δr, between a ray starting from the point $Q(x,y)$ and a parallel ray starting from the origin O, is given by $\mathbf{R} \cdot \hat{\mathbf{n}}$, (Figure 5.27), where $\mathbf{R} = \hat{\mathbf{i}}x + \hat{\mathbf{j}}y$ and $\hat{\mathbf{n}}$ is a unit vector in the direction
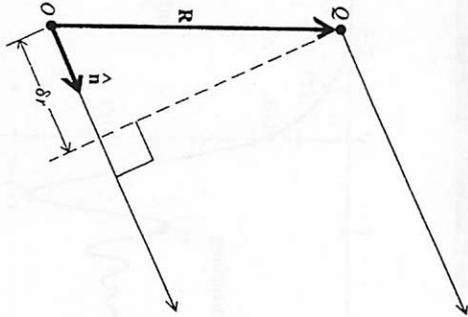

Figure 5.27. Path difference between two parallel rays of light originating from points O and Q in the xy plane.

of the ray. Since $\hat{\mathbf{n}}$ can be expressed as $\hat{\mathbf{n}} = \hat{\mathbf{i}}\alpha + \hat{\mathbf{j}}\beta + \hat{\mathbf{k}}\gamma$, then

$$\delta r = \mathbf{R} \cdot \hat{\mathbf{n}} = x\alpha + y\beta = x\frac{X}{L} + y\frac{Y}{L} \qquad (5.50)$$

It follows that the fundamental diffraction integral [Equation (5.16)] giving the diffraction pattern in the XY plane is, aside from a constant multiplying factor, expressible in the form

$$U(X,Y) = \iint e^{ik\delta r}\, d\mathscr{A} = \iint e^{ik(Xx+Yy)/L}\, dx\, dy \qquad (5.51)$$

This is the case for a uniform aperture.

For a uniform rectangular aperture the double integral reduces to the product of two one-dimensional integrals. The result is stated earlier in Section 5.4.

For a nonuniform aperture we introduce a function $g(x,y)$ called the *aperture function*. This function is defined such that $g(x,y)\, dx\, dy$

is the amplitude of the diffracted wave originating from the element of area $dx\, dy$. Thus instead of Equation (5.51), we have the more general formula

$$U(X,Y) = \iint g(x,y)\, e^{ik(Xx+Yy)}\, dx\, dy \qquad (5.52)$$

It is convenient at this point to introduce the quantities

$$\mu = \frac{kX}{L} \quad \text{and} \quad \nu = \frac{kY}{L} \qquad (5.53)$$

μ and ν are called *spatial frequencies*, although they have the dimensions of reciprocal length, that is, wavenumber. We now write Equation (5.52) as

$$U(\mu,\nu) = \iint g(x,y)\, e^{i(\mu x + \nu y)}\, dx\, dy \qquad (5.54)$$

We see that the functions $U(\mu,\nu)$ and $g(x,y)$ constitute a two-dimensional Fourier transform pair. The diffraction pattern, in this context, is actually a Fourier resolution of the aperture function.

Consider as an example a grating. For simplicity we treat it as a one-dimensional problem. The aperture function $g(y)$ is then a periodic step function as shown in Figure 5.28. It is represented by a Fourier series of the form

$$g(y) = g_0^2 + g_1 \cos(\nu_0 y) + g_2 \cos(2\nu_0 y) + \cdots \qquad (5.55)$$
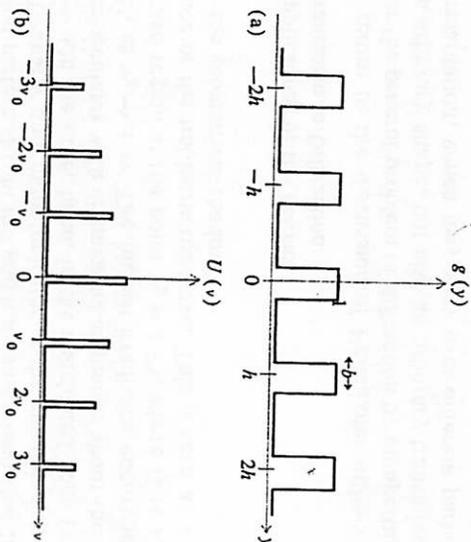
Figure 5.28. Aperture function for a grating and its Fourier transform.

The fundamental spatial frequency ν_0 is given by the periodicity of the grating, namely,

$$\nu_0 = \frac{2\pi}{h} \qquad (5.56)$$

where h is the grating spacing. This dominant spatial frequency appears in the diffraction pattern as the first-order maximum, the amplitude of which is proportional to g_1. Maxima of higher order correspond to higher Fourier components of the aperture function $g(y)$. Thus if the aperture function were of the form of a cosine function $g_0 + g_1 \cos(\nu_0 y)$ instead of a periodic step function, then the diffraction pattern would consist only of the central maximum and the two first-order maxima. Second or higher diffraction orders would not appear.

Apodization   Apodization (literally "to remove the feet") is the name given to any process by which the aperture function is altered in such a way as to produce a redistribution of energy in the diffraction pattern. Apodization is usually employed to reduce the intensity of the secondary diffraction maxima.

It is perhaps easiest to explain the theory of apodization by means of a specific example. Let the aperture consist of a single slit. The aperture function in this case is a single step function: $g(y) = 1$ for $-b/2 < y < b/2$ and $g(y) = 0$ otherwise (Figure 5.29). The corresponding diffraction pattern, expressed in terms of spatial frequencies, is

$$U(\nu) = \int_{-b/2}^{+b/2} e^{i\nu y}\, dy = b\,\frac{\sin\left(\frac{1}{2}\nu b\right)}{\left(\frac{1}{2}\nu b\right)} \qquad (5.57)$$

This is equivalent to the normal case already discussed in Section 5.5.

Suppose now that the aperture function is altered by apodizing in such a way that the resultant aperture transmission is a cosine function: $g(y) = \cos(\pi y/b)$ for $-b/2 < y < b/2$ and zero otherwise, as shown in Figure 5.29. This could be accomplished, for example, by means of a suitably coated-glass plate placed over the aperture. The new diffraction pattern is given by

$$U(\nu) = \int_{-b/2}^{+b/2} \cos\left(\frac{\pi y}{b}\right) e^{i\nu y}\, dy$$

$$= \cos(\nu b/2)\left(\frac{1}{\nu - \pi/b} - \frac{1}{\nu + \pi/b}\right) \qquad (5.58)$$

A comparison of the two diffraction patterns is shown graphically in the figure. The result of apodization in this case is a substantial reduction in the secondary maxima relative to the central maximum; in
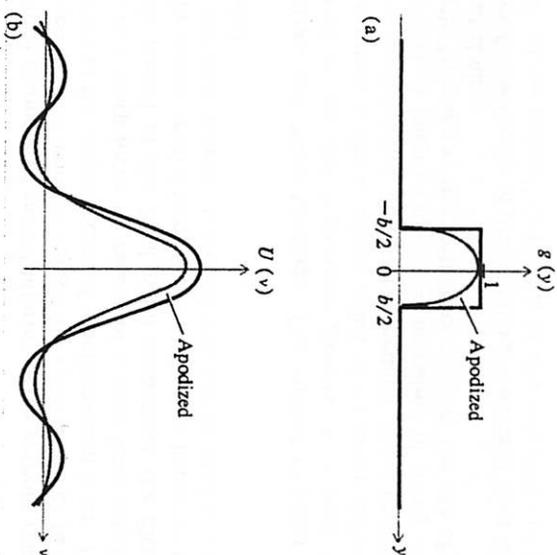
Figure 5.29.   (a) Aperture functions for a slit and an apodized slit; (b) the Fourier transforms.

other words, apodization has suppressed the higher spatial frequencies.

In a similar way it is possible to apodize the circular aperture of a telescope so as to reduce greatly the relative intensities of the diffraction rings that appear around the images of stars (discussed in Section 5.5). This enhances the ability of the telescope to resolve the image of a dim star near that of a bright one.

Spatial Filtering   Consider the diagram shown in Figure 5.30. Here the xy plane represents the location of some *coherently* illuminated object.[3] This object is imaged by an optical system (not shown), the image appearing in the $x'y'$ plane. The diffraction pattern $U(\mu,\nu)$ of the object function $g(x,y)$ appears in the $\mu\nu$ plane. This plane is analogous to the XY plane in Figure 5.26. Hence, from Equation (5.54) $U(\mu,\nu)$ is the Fourier transform of $g(x,y)$. The image function $g'(x',y')$ that appears in the $x'y'$ plane is, in turn, the Fourier transform of $U(\mu,\nu)$. Now if *all* spatial frequencies in the range $\mu = \pm\infty$, $\nu = \pm\infty$ were transmitted equally by the optical system, then, from the properties of the Fourier transform, the image function $g'(x',y')$

³ For a discussion of the theory of spatial filtering with incoherent illumination see Reference [10].
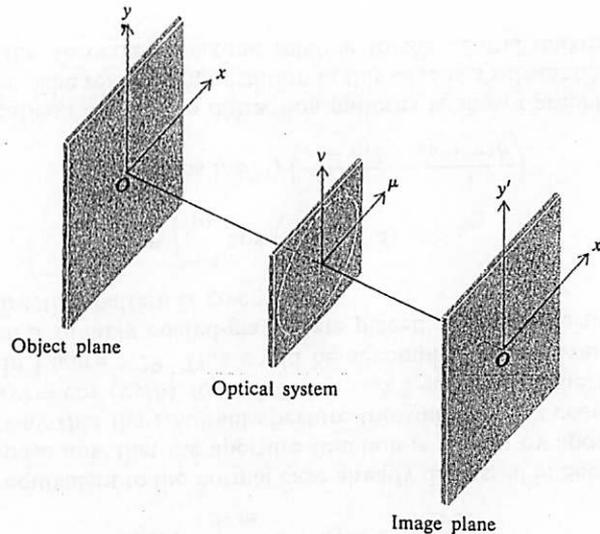
Figure 5.30. Geometry for the general problem of image formation by an optical system.

would be *exactly* proportional to the object function $g(x,y)$; that is, the image would be a true reproduction of the object. However, the finite size of the aperture at the $\mu\nu$ plane limits the spatial frequencies that are transmitted by the optical system. Furthermore there may be lens defects, aberrations, and so forth, which result in a modification of the function $U(\mu,\nu)$. All of these effects can be incorporated into one function $T(\mu,\nu)$ called the *transfer function* of the optical system. This function is defined implicitly by the equation

$$U'(\mu,\nu) = T(\mu,\nu)\, U(\mu,\nu)$$

Thus

$$g'(x', y') = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} T(\mu,\nu)U(\mu,\nu)\, e^{-i(\mu x' + \nu y')}\, d\mu\, d\nu \qquad (5.59)$$

that is, the image function is the Fourier transform of the product $T(\mu,\nu) \cdot U(\mu,\nu)$. The limits of integration are $\pm\infty$ in a formal sense only. The actual limits are given by the particular form of the transfer function $T(\mu,\nu)$.

The transfer function can be modified by placing various screens and apertures in the $\mu\nu$ plane. This is known as *spatial filtering*. The situation is quite analogous to the filtering of an electrical signal by means of a passive electrical network. The object function is the input signal, and the image function is the output signal. The optical system

acts like a filter that allows certain spatial frequencies to be transmitted but rejects others.

Suppose, for example, that the object is a grating so that the object function is a periodic step function. This case can be treated as a one-dimensional problem. The object function $f(y)$ and its Fourier transform $U(\nu)$ are then just those shown in Figure 5.28. Now let the aperture in the $\mu\nu$ plane be such that only those spatial frequencies that lie between $-\nu_{max}$ and $+\nu_{max}$ are transmitted. This means that we have low-pass filtering. From Equation (5.53) we have $\nu_{max} = kb/f$, where $2b$ is the physical width of the aperture in the $\mu\nu$ plane. The transfer function for this case is a step function: $T(\nu) = 1$, $-\nu_{max} < \nu < +\nu_{max}$, and zero otherwise. The image function is, accordingly,

$$g'(y') = \int_{-\nu_{max}}^{+\nu_{max}} U(\nu)e^{-i\nu y'}\, d\nu \qquad (5.60)$$

Without going into the details of the calculation of $g'(y')$, we show in Figure 5.31(a) a graphical plot for some arbitrary choice of ν_{max}. Instead of the sharp step function that constitutes the object, the image is rounded at the corners and also shows small periodic variations.

A high-pass optical filter is obtained by placing in the $\mu\nu$ plane a screen that blocks off the central part of the diffraction pattern. This part of the diffraction pattern corresponds to the low frequencies. The approximate form of the resulting image function is shown in Figure 5.31(b). Only the edges of the grating steps are now visible in the image plane. *The edge detail comes from the higher spatial frequencies.*

A practical example of spatial filtering is the *pinhole spatial filter* which is used in laser work to reduce the spurious fringe pattern that always occurs in the output beam of a helium–neon laser. The beam is brought to a sharp focus by means of a short-focal-length lens. A fine pinhole placed at the focal point constitutes the filter, which removes the higher spatial frequencies and hence improves the beam quality of the laser output. A second lens can be used to render the beam parallel.

Phase Contrast and Phase Gratings The method of phase contrast was invented by the Dutch physicist Zernike. It is used to render visible a transparent object whose index of refraction differs slightly from that of a surrounding transparent medium. Phase contrast is particularly useful in microscopy for examination of living organisms, and so forth. In essence, the method consists of the use of a special type of spatial filter.

To simplify the theory of phase contrast, we shall treat the case of a so-called "phase grating" consisting of alternate strips of high- and

Object function
(grating)

$g(y)$

Transfer function
(low pass)

$-3v_0 \quad -2v_0 \quad -v_0 \quad 0 \quad v_0 \quad 2v_0 \quad 3v_0$

Image function

$g'(y')$

Spatial
frequency
v

(a)

$-6v_0 \quad -5v_0 \quad -4v_0$

$g'(y')$

Transfer function
(high pass)

$4v_0 \quad 5v_0 \quad 6v_0$

Image function

Spatial
frequency
v

(b)
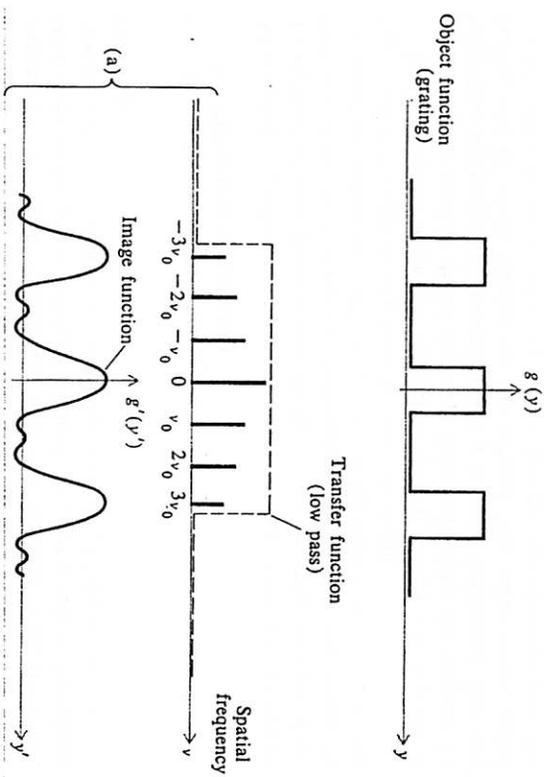
Figure 5.31. Graphs illustrating spatial filtering. (a) Low-pass filtering; (b) high-pass filtering.

low-index material, all strips being perfectly transparent. The grating is coherently illuminated and constitutes the object. The object function is thus represented by the exponential

$$g(y) = e^{i\phi(y)} \qquad (5.61)$$

where the phase factor $\phi(y)$ is a periodic step function as shown in Figure 5.32(a). The "height" of the step is the optical-phase difference between the two kinds of strips; that is, $\Delta\phi = kz\,\Delta n$, where z is the thickness and Δn is the difference between the two indices of

(a)

$\phi(y)$

$-h$ ⟶ h ⟶ y

$U_1(v)$

$U_2(v)$

v

(b)
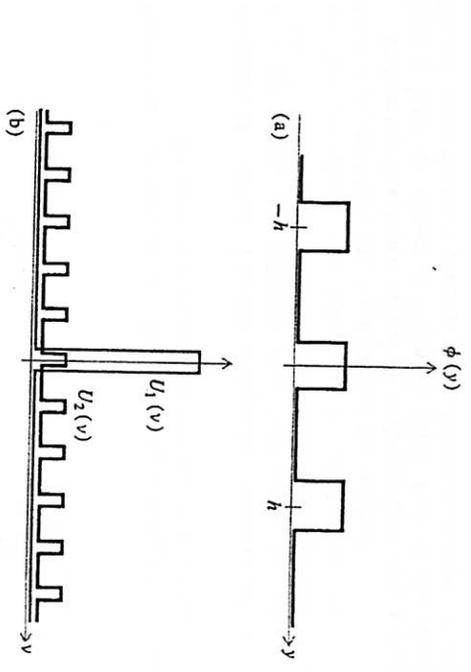
Figure 5.32. (a) The phase function of a periodic phase grating; (b) Fourier transforms of the aperture U_1 and the grating U_2.

refraction. If we assume that this phase difference is very small, then to a good approximation, we can write

$$g(y) = 1 + i\phi(y) \qquad (5.62)$$

The Fourier transform of the above function is

$$U(v) = \int_{-\infty}^{\infty} [1 + i\phi(y)] e^{ivy}\, dy = \int_{-b/2}^{+b/2} e^{ivy}\, dy + i \int_{-b/2}^{+b/2} \phi(y)\, e^{ivy}\, dy$$
$$= U_1(v) + iU_2(v) \qquad (5.63)$$

Here $U_1(v)$ represents the diffraction pattern of the whole-object aperture. It is essentially zero everywhere except for $v \approx 0$; that is, $U_1(v)$ contains only very low spatial frequencies. On the other hand, $U_2(v)$ represents the diffraction pattern of the periodic step function $\phi(y)$. The two functions are plotted in Figure 5.32(b).

By virtue of the factor i in the result, $U_1 + iU_2$, the two components U_1 and iU_2 are 90 degrees out of phase. The essential trick in the phase-contrast method consists of inserting a spatial filter in the μv plane, which has the property of shifting the phase of iU_2 by an additional 90 degrees. In practice this is accomplished by means of a device known as a *phase plate*. The physical arrangement is shown in Figure 5.33. The phase plate is just a transparent-glass plate having a small section whose optical thickness is ¼ wavelength greater than the remainder of the plate. This thicker section is located in the central part of the μv plane, that is, in the region of low spatial frequencies. The result of inserting the phase plate is to change the function
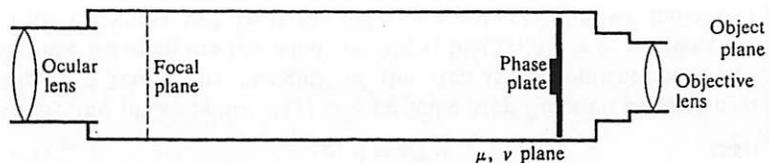
Figure 5.33. Physical arrangement of the optical elements for phase contrast microscopy.

$U_1 + iU_2$ to $U_1 + U_2$. The new image function is given by the Fourier transform of the new $U(\nu)$, namely,

$$g'(y') = \int U_1(\nu)e^{-i\nu y'} \, d\nu + \int U_2(\nu)e^{-i\nu y'} \, d\nu$$
$$= g_1(y') + g_2(y')$$

(5.64)

Now the first function g_1 is just the image function of the whole-object aperture. It represents the constant background. The second function g_2 is the image function for a regular grating of alternate transparent and opaque strips. This means that the phase grating has been rendered visible. It appears in the image plane as alternate bright and dark strips. Although the above analysis has been for a periodic grating, a similar argument can be applied to a transparent-phase object of any shape.

The method of optical-phase contrast has a close analogy in electrical communications. A phase-modulated signal is converted into an amplitude-modulated signal by introduction of a phase shift of 90 degrees to the carrier frequency. This is essentially what the phase plate does in the phase-contrast method. The net result is that phase modulation in the object is converted into amplitude modulation in the image.

5.7 Reconstruction of the Wave Front by Diffraction. Holography

An unusual and interesting method of producing an image—known as the method of *wave-front reconstruction*—has recently become of importance in the field of optics. Although the basic idea was originally proposed by Gabor in 1947 [12], it attracted little attention until the highly coherent light of the laser became available.

In this method a special diffraction screen, called a *hologram*, is used to reconstruct in detail the wave field emitted by the subject. To make the hologram the output from a laser is separated into two beams, one of which illuminates the subject. The other beam, called the *reference beam,* is reflected onto a fine-grained photographic film
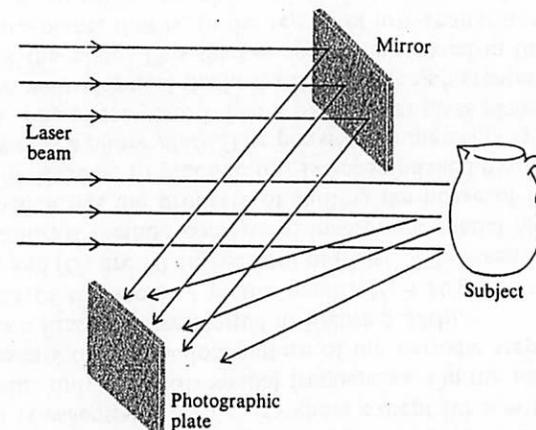
(a)
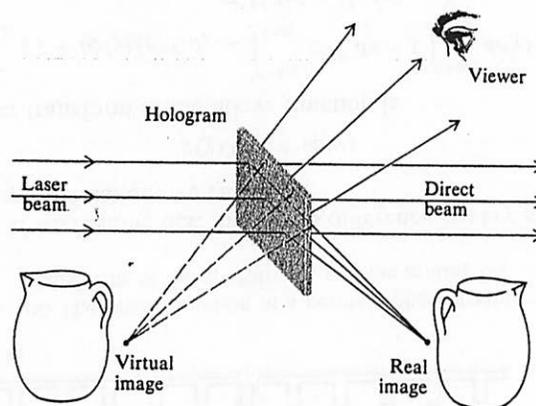
(b)

Figure 5.34. (a) Arrangement for producing a hologram; (b) use of the hologram in producing the real and virtual images.

by means of a mirror. The film is exposed simultaneously to the reference beam and the reflected laser light from the subject [Figure 5.34(a)]. The resulting complicated interference pattern recorded by the film constitutes the hologram. It contains all the information needed to reproduce the wave field of the subject.

In use the developed hologram is illuminated with a single beam from a laser as shown in Figure 5.34(b). Part of the resulting diffracted wave field is a precise, three-dimensional copy of the original wave reflected by the subject. The viewer looking at the hologram

sees the image in depth and by moving his head can change his perspective of the view.

In order to simplify the discussion of the theory of holography, we shall assume that the reference beam is collimated; that is, it consists of plane waves, although this is not actually necessary in practice. Let x and y be the coordinates in the plane of the recording photographic plate, and let $U(x,y)$ denote the complex amplitude of the reflected wave front in the xy plane. Since $U(x,y)$ is a complex number, we can write it as

$$U(x,y) = a(x,y)e^{i\phi(x,y)} \tag{5.65}$$

where $a(x,y)$ is real.

Similarly, let $U_0(x,y)$ denote the complex amplitude of the reference beam. Since this beam is plane, we can write

$$U_0(x,y) = a_0 \, e^{i(\mu x + \nu y)} \tag{5.66}$$

where a_0 is a constant and μ and ν are the spatial frequencies of the reference beam in the xy plane. They are given by

$$\mu = k \sin \alpha \qquad \nu = k \sin \beta \tag{5.67}$$

in which k is the wave number of the laser light, and α and β specify the direction of the reference beam.

The irradiance $I(x,y)$ that is recorded by the photographic film is thus given by the expression

$$I(x,y) = \|U + U_0\|^2 = a^2 + a_0^2 + aa_0 e^{i[\phi(x,y) - \mu x - \nu y]} + aa_0 e^{-i[\phi(x,y) - \mu x - \nu y]}$$
$$= a^2 + a_0^2 + 2aa_0 \cos [\phi(x,y) - \mu x - \nu y] \tag{5.68}$$

This is actually an interference pattern. It contains information in the form of amplitude and phase modulations of the spatial frequencies of the reference beam. The situation is somewhat analogous to the impression of information on the carrier wave of a radio transmitter by means of amplitude or phase modulation.

When the developed hologram is illuminated with a single beam U_0 similar to the reference beam, the resulting transmitted wave U_T is proportional to U_0 times the transmittance of the hologram at the point (x,y). The transmittance will be proportional to $I(x,y)$. Hence, except for a constant proportionality factor that we ignore,

$$U_T(x,y) = U_0 I = a_0(a^2 + a_0^2)e^{i(\mu x + \nu y)} + a_0^2 a \, e^{i\phi} + a_0^2 a e^{-i(\phi - 2\mu x - 2\nu y)}$$
$$= (a^2 + a_0^2) U_0 + a_0^2 U + a^2 U^{-1} U_0^{-2} \tag{5.69}$$

The hologram acts somewhat like a diffraction grating. It produces a direct beam and two first-order diffracted beams on either side of the direct beam [Figure 5.34(b)]. The term $(a^2 + a_0^2)U_0$ in Equation

(5.69) comprises the direct beam. The term $a_0^2 U$ represents one of the diffracted beams. Since it is equal to a constant times U, this beam is the one that reproduces the reflected light from the subject and forms the virtual image. The last term represents the other diffracted beam and gives rise to a real image.

We shall not attempt to prove the above statements in detail. They can be verified by considering a very simple case, namely, that in which the subject is a single white line on a dark background. In this case the hologram turns out to be, in fact, a simple periodic grating. The zero order of the diffracted light is the direct beam, whereas the two first orders on either side comprise the virtual and the real images.

In holography the viewer always sees a positive image whether a positive or a negative photographic transparency is used for the hologram. The reason for this is that a negative hologram merely produces a wave field that is shifted 180 degrees in phase with respect to that of a positive hologram. Since the eye is insensitive to this phase difference, the view seen by the observer is identical in the two cases.

Remarkable technical advances have been made in the field of holography in recent years. Holography in full color is possible by using three different laser wavelengths instead of just one, the holographic record being on black-and-white film. The holographic principle has been extended to include the use of acoustic waves for imaging in optically opaque media and to microwaves for long-distance holography.

Holographic Interferometry One of the most notable applications of holography is in the field of interferometry. In this application the surface to be tested can be irregular and diffusely reflecting instead of smooth and highly polished as is required for the ordinary Michelson and Twyman-Green type of interferometric work. In *double-exposure* holographic interferometry two separate exposures are made on a single recording film. If the surface under study undergoes any deformation or movement during the time interval between exposures, such movement is revealed on the reconstructed image in the form of interference fringes. In *double-pulse* holography the two exposures are produced by short, intense laser pulses from a high-power pulsed laser. These pulses are closely spaced in time so that the holographic image fringes can show motion, vibration patterns, and so on. The method is especially useful for nondestructive testing. For more information on the subject of holography, the reader is encouraged to consult a text such as *An Introduction to Coherent Optics and Holography* by G. W. Stroke [38].

that the resolving power of a Fabry-Perot instrument can be expressed as

$$RP = N\mathscr{F} = N\pi\left(\frac{\sqrt{R}}{1-R}\right) \qquad (4.21)$$

and is thus determined by the order of interference and the reflectance.

By increasing the order of interference, the resolving power with a given reflectance can be made as large as desired. This is easily accomplished by increasing the mirror separation, because $N = 2nd/\lambda_0$, approximately, as given by Equation (4.10). However, the free spectral range then diminishes, and so a compromise must be chosen in any given application. For a given value of the mirror separation the resolving power, in principle, can be increased indefinitely by making the reflectance closer and closer to unity. However, a practical limit is imposed by absorption in the reflecting surface which reduces the intensity of the transmitted fringes, as discussed in Section 4.1.

Silver and aluminum films, deposited by vacuum evaporation, have long been used for Fabry-Perot instruments. The useful reflectance with these metal films, as limited by absorption, is only about 80 to 90 percent. More recently, multilayer dielectric films have been used for Fabry-Perot work. With such films, discussed in the next section, useful reflectances approaching 99 percent can be achieved. A good Fabry-Perot instrument can easily have a resolving power of 1 million, which is 10 to 100 times that of a prism or small-grating spectroscope. For a more complete discussion of Fabry-Perot interferometry, and of other instruments used in high-resolution spectroscopy, see References [5] and [41] listed at the end of the book.

4.4 Theory of Multilayer Films

Multilayer films are widely used in science and industry for control of light. Optical surfaces having virtually any desired reflectance and transmittance characteristics may be produced by means of thin film coatings. These films are usually deposited on glass or metal substrates by high-vacuum evaporation. The well-known use of antireflecting coatings for camera lenses and other optical instruments is only one of the many practical applications of thin films. Other applications include such things as heat-reflecting and heat-transmitting mirrors ("hot" and "cold" mirrors), one-way mirrors, optical filters, and so forth.

First consider the case of a single layer of dielectric of index n_1 and thickness l between two infinite media of indices n_0 and n_T (Figure 4.7). For simplicity we shall develop the theory for normally incident light. The modifications for the general case of oblique in-
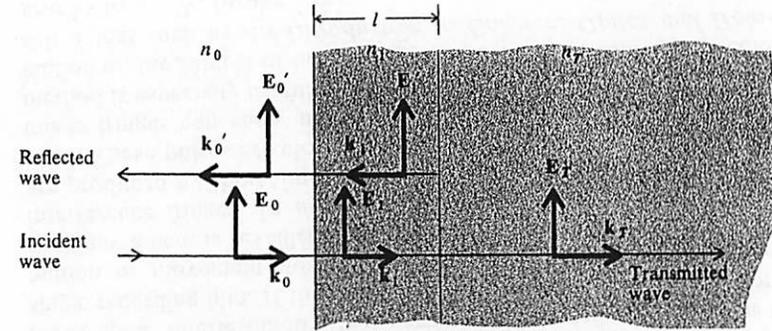
Figure 4.7. Wave vectors and their associated electric fields for the case of normal incidence on a single dielectric layer.

cidence are easily made. The amplitude of the electric vector of the incident beam is E_0. That of the reflected beam is E_0', and that of the transmitted beam is E_T. The electric-field amplitudes in the film are E_1 and E_1' for the forward and backward traveling waves, respectively, as indicated in the figure.

The boundary conditions require that the electric and magnetic fields be continuous at each interface. These conditions are expressed as follows:

First Interface	*Second Interface*
Electric: $E_0 + E_0' = E_1 + E_1'$	$E_1 e^{ikl} + E_1' e^{-ikl} = E_T$
Magnetic: $H_0 - H_0' = H_1 - H_1'$	$H_1 e^{ikl} - H_1' e^{-ikl} = H_T$
or $n_0 E_0 - n_0 E_0' = n_1 E_1 - n_1 E_1'$	$n_1 E_1 e^{ikl} - n_1 E_1' e^{-ikl} = n_T E_T$

The relations for the magnetic fields follow from the theory developed in Section 2.7. The phase factors e^{ikl} and e^{-ikl} result from the fact that the wave travels through a distance l from one interface to the other.

If we eliminate the amplitudes E_1 and E_1', we obtain

$$1 + \frac{E_0'}{E_0} = (\cos kl - i\frac{n_T}{n_1}\sin kl)\frac{E_T}{E_0}$$

$$n_0 - n_0\frac{E_0'}{E_0} = (-in_1\sin kl + n_T\cos kl)\frac{E_T}{E_0}$$

(4.22)

or, in matrix form,

$$\begin{bmatrix} 1 \\ n_0 \end{bmatrix} + \begin{bmatrix} 1 \\ -n_0 \end{bmatrix}\frac{E_0'}{E_0} = \begin{bmatrix} \cos kl & \dfrac{-i}{n_1}\sin kl \\ -in_1\sin kl & \cos kl \end{bmatrix}\begin{bmatrix} 1 \\ n_T \end{bmatrix}\frac{E_T}{E_0} \qquad (4.23)$$

which can be abbreviated as

$$\begin{bmatrix} 1 \\ n_0 \end{bmatrix} + \begin{bmatrix} 1 \\ -n_0 \end{bmatrix} r = M \begin{bmatrix} 1 \\ n_T \end{bmatrix} t \qquad (4.24)$$

We have here introduced the reflection coefficient

$$r = \frac{E_0'}{E_0} \qquad (4.25)$$

and the transmission coefficient

$$t = \frac{E_T}{E_0} \qquad (4.26)$$

The matrix, known as the *transfer matrix*

$$M = \begin{bmatrix} \cos kl & -\dfrac{i}{n_1} \sin kl \\ -i n_1 \sin kl & \cos kl \end{bmatrix} \qquad (4.27)$$

where n_1 is the index of refraction, and $k = 2\pi/\lambda = 2\pi n_1/\lambda_0$.

Now suppose that we have N layers numbered 1, 2, 3, ... N having indices of refraction $n_1, n_2, n_3, \ldots n_N$ and thicknesses $l_1, l_2, l_3, \ldots l_N$, respectively. In the same way that we derived Equation (4.24), we can show that the reflection and transmission coefficients of the multilayer film are related by a similar matrix equation:

$$\begin{bmatrix} 1 \\ n_0 \end{bmatrix} + \begin{bmatrix} 1 \\ -n_0 \end{bmatrix} r = M_1 M_2 M_3 \cdots M_N \begin{bmatrix} 1 \\ n_T \end{bmatrix} t = M \begin{bmatrix} 1 \\ n_T \end{bmatrix} t \quad (4.28)$$

where the transfer matrices of the various layers are denoted by M_1, $M_2, M_3, \ldots, M_N$. Each transfer matrix is of the form given by Equation (4.27) with appropriate values of n, l, and k. The overall transfer matrix M is the product of the individual transfer matrices. Let the elements of M be A, B, C, and D, that is

$$M_1 M_2 M_3 \cdots M_N = M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad (4.29)$$

We can then solve Equation (4.28) for r and t in terms of these elements. The result is

$$r = \frac{An_0 + Bn_T n_0 - C - Dn_T}{An_0 + Bn_T n_0 + C + Dn_T} \qquad (4.30)$$

$$t = \frac{2n_0}{An_0 + Bn_T n_0 + C + Dn_T} \qquad (4.31)$$

The reflectance R and the transmittance T are then given by $R = |r|^2$ and $T = |t|^2$, respectively.

Antireflecting Films   The transfer matrix for a single film of index n_1 and thickness l is given by Equation (4.27). Suppose this film is placed on a glass substrate of index n_T. The coefficient of reflection of the combination, in air, is then given by Equation (4.30) with $n_0 = 1$. The result is

$$r = \frac{n_1(1 - n_T) \cos kl - i(n_T - n_1{}^2) \sin kl}{n_1(1 + n_T) \cos kl - i(n_T + n_1{}^2) \sin kl} \qquad (4.32)$$

If the optical thickness of the film is $\frac{1}{4}$ wavelength, then $kl = \pi/2$. The reflectance for a quarter-wave film is therefore

$$R = |r|^2 = \frac{(n_T - n_1{}^2)^2}{(n_T + n_1{}^2)^2} \qquad (4.33)$$

In particular, the reflectance is zero if

$$n_1 = \sqrt{n_T} \qquad (4.34)$$

Magnesium fluoride, whose index is 1.35, is commonly used for coating lenses. Although this does not exactly satisfy the above requirement for ordinary glass, $n_T \approx 1.5$, the reflectance of glass coated with a quarter-wave layer of magnesium fluoride is reduced to about 1 percent, which is one fourth that of uncoated glass. This can result in a considerable saving of light in the case of optical instruments having many elements, such as high-quality camera lenses that may have as many as five or six components, that is, ten or twelve reflecting surfaces.

By using two layers, one of high index and one of low, it is possible to obtain zero reflectance (at one wavelength) with available coating materials. More layers obviously afford greater latitude and more extensive possibilities. Thus with three suitably chosen layers the reflectance can be reduced to zero for two wavelengths and can be made to average less than $\frac{1}{4}$ percent over almost the entire visible spectrum. Some curves are shown in Figure 4.8.
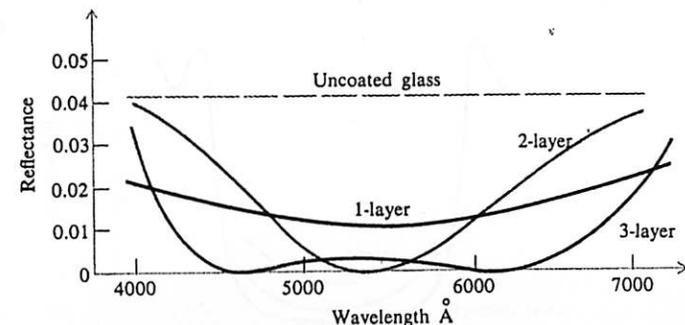

Figure 4.8.   Curves of reflectance versus wavelength of antireflecting films.

Figure 4.9. Multilayer stack for producing high reflectance. The stack consists of alternate quarter-wave layers of high and low index material. (Note: λ is the wavelength in the material.)

High-Reflectance Films In order to obtain a high value of reflectance in a multilayer film, a stack of alternate layers of high index, n_H, and low index, n_L, materials is used, the thickness of each layer being $\frac{1}{4}$ wavelength, Figure 4.9. In this case the transfer matrices are all of the same form, and the product of two adjacent ones is

$$\begin{bmatrix} 0 & \dfrac{-i}{n_L} \\ -in_L & 0 \end{bmatrix}\begin{bmatrix} 0 & \dfrac{-i}{n_H} \\ -in_H & 0 \end{bmatrix} = \begin{bmatrix} \dfrac{-n_H}{n_L} & 0 \\ 0 & \dfrac{-n_L}{n_H} \end{bmatrix} \qquad (4.35)$$

If the stack consists of 2N layers, then the transfer matrix of the complete multilayer film is

$$M = \begin{bmatrix} \dfrac{-n_H}{n_L} & 0 \\ 0 & \dfrac{-n_L}{n_H} \end{bmatrix}^N = \begin{bmatrix} \left(\dfrac{-n_H}{n_L}\right)^N & 0 \\ 0 & \left(\dfrac{-n_L}{n_H}\right)^N \end{bmatrix} \qquad (4.36)$$

Assuming, for simplicity, that n_0 and n are both unity, the reflectance of a multilayer stack is given by Equation (4.30) as follows:

$$R = |r|^2 = \left[\frac{(-n_H/n_L)^N - (-n_L/n_H)^N}{(-n_H/n_L)^N + (-n_L/n_H)^N}\right]^2 = \left[\frac{(n_H/n_L)^{2N} - 1}{(n_H/n_L)^{2N} + 1}\right]^2 \qquad (4.37)$$

The reflectance thus approaches unity for large N. For instance, an eight-layer stack (N = 4) of zinc sulfide (n_H = 2.3) and magnesium fluoride (n_L = 1.35) gives a reflectance of about 0.97, which is higher than the reflectance of pure silver in the visible region of the spectrum. A 30-layer stack results in a reflectance of better than 0.999. This maximum reflectance, of course, occurs only at 1 wavelength, but it is possible to broaden the region of high reflectance by combinations of different thicknesses. Figure 4.10 shows some approx-

Figure 4.10. Reflectance curves for some multilayer high-reflectance films.

Figure 4.11. Multilayer Fabry-Perot interference filter.

imate curves of reflectance as a function of wavelength for multilayer films such as are used in laser work.

Fabry-Perot Interference Filter   A Fabry-Perot type of filter consists of a layer of dielectric having a thickness of $\frac{1}{2}$ wavelength (for some wavelength λ_0) and bounded on both sides by partially reflecting surfaces. In effect, this is a Fabry-Perot etalon with a very small spacing. The result is a filter that has a transmission curve given by the Airy function Equation (4.8), where the peak occurs at the wavelength λ_0. Higher-order peaks also occur for wavelengths $\frac{1}{2}\lambda_0$, $\frac{1}{3}\lambda_0$, $\frac{1}{4}\lambda_0$, . . . , and so forth. The spectral width of the transmission band depends on the reflectance of the bounding surfaces. Fabry-Perot filters were first made with silver films to produce the necessary high reflectance, but now they are usually made entirely of multilayer dielectric films. The latter are superior to the metal films because of their higher reflectance and lower absorption. Figure 4.11 shows a typical design of a multilayer Fabry-Perot interference filter together with a transmission curve.

PROBLEMS

4.1   The plates of a Fabry-Perot interferometer are coated with silver of such a thickness that for each plate the reflectance is 0.9, the transmittance is 0.05, and the absorption is 0.05. Find the maximum and minimum transmittance of the interferometer. What is the value of the reflecting finesse and of the coefficient of finesse?

4.2   Find the resolving power of the interferometer in Problem 4.1 if the plate separation is 1 cm and the wavelength is 500 nm.

4.3   The mirrors of a Fabry-Perot resonator for a laser are coated to give a reflectance of 0.99 and they are separated a distance of 1 m. Find the value of the fringe width in wavelength and in frequency at a wavelength of 633 nm.

4.4   Show that the radii of the interference fringes of a plane-parallel Fabry-Perot interferometer, as shown in Figure 4.5(a), are approximately proportional to $\sqrt{0}$, $\sqrt{1}$, $\sqrt{2}$, . . . $\sqrt{N'}$, where N' is an integer, provided that there is a zero-radius fringe. [Hint: Use Equation (4.10) and assume that θ is small, so that $\cos\theta \approx 1 - \theta^2/2$.]

4.5   A collimated beam of white light falls at normal incidence on a plate of glass of index n and thickness d. Develop a formula for the transmittance as a function of wavelength. Show that maxima occur at those wavelengths such that

$$\lambda_N = \frac{2nd}{N}$$

where λ_N is the vacuum wavelength and N is an integer. The transmission function is thus periodic (in wavenumber or frequency) and is called a "channeled spectrum."

4.6   Using the result of Problem 4.5, find the maximum and minimum transmittance of the channels for a plate of index of refraction 2.5. If the plate is 1 mm thick, find the wavelength separation between adjacent channels at a vacuum wavelength of 600 nm.

4.7   Calculate the reflectance of a quarter-wave antireflecting film of magnesium fluoride ($n = 1.35$) coated on an optical glass surface of index 1.52.

4.8   Find the peak reflectance of a high-reflecting multilayer film consisting of eight layers of high-low index material (four of each) of index $n_L = 1.4$ and $n_H = 2.8$.

4.9   Fill in the steps in the derivation of the expression of the transfer matrix of a single film

$$M = \begin{bmatrix} \cos kl & -\dfrac{i}{n_1}\sin kl \\ -in_1 \sin kl & \cos kl \end{bmatrix}$$

4.10   Show that in the case of oblique incidence the transfer matrix of a single film is

$$M = \begin{bmatrix} \cos\beta & -\dfrac{i}{p}\sin\beta \\ -ip\sin\beta & \cos\beta \end{bmatrix}$$

where

$$\beta = kl \cos\theta$$

and

$$p = n_1 \cos\theta \qquad (TE \text{ polarization})$$
$$p = \frac{n_1}{\cos\theta} \qquad (TM \text{ polarization})$$

The angle θ is the angle between the wave vector inside and the normal to the surface of the film.