An example of double exposure holographic interferometry. A hologram of the c-clamp is exposed. The light source is then shuttered and the clamp tightened. A hologram of the strained c-clamp is exposed onto the same recording material. After processing, illumination of the doubly exposed hologram results in the reconstruction of two wavefronts—one corresponding to each amount of strain on the c-clamp. These two waves interfere so as to form the fringes pictured here.
because the optical path difference between a typical pair of reflected (diffracted) rays, such as (1) and (2) in the figure, is given by

\[ \text{OPD} = 2d \cos \left( \frac{1}{2} \theta \right). \]  

(2.11)

Since \( d = \lambda_1/2 \cos \left( \frac{1}{2} \theta \right) \), rays (1) and (2) are in phase and interfere constructively, so there is strong reflection in this direction for this wavelength. If the hologram is rotated to a viewing angle of \( \frac{1}{2} \theta' \) (Fig. 2.19c), only the wavelength satisfying the condition

\[ 2d \cos \left( \frac{1}{2} \theta' \right) = \lambda_2 \]  

(2.12)

or

\[ \frac{\lambda_1}{\lambda_2} = \frac{\cos \left( \frac{1}{2} \theta \right)}{\cos \left( \frac{1}{2} \theta' \right)} \]  

(2.13)

will be strongly reflected.

Color holograms may be made using this effect. A full-color object is illuminated with red, green, and blue light, and the reference beam is composed of the same combination. Each color then forms its own set of standing waves in the medium, each with different spacing. Readout can be accomplished by illuminating with white light. Each set of Bragg planes filters out the color with which it was made and a full-color reconstruction results. A more detailed analysis of color holograms, along with other possible arrangements, will be given in Chapter 7.

The arrangements described in this chapter do not indicate all that are possible, of course. We have shown only typical examples of arrangements that fall in one or the other of the two major classifications, Fresnel holograms and Fraunhofer holograms. Generally, a Fraunhofer hologram is formed when the object is a great distance away from the recording medium. A Fresnel hologram is formed when the object is near the recording medium. A more accurate mathematical distinction between these two types will be made in Chapter 3.

3 General Theory of Plane Holograms

3.0 Introduction

In this chapter the basic notation and coordinate systems to be used throughout the book will be introduced. The basic mathematics of making the hologram and reconstructing the wavefronts for each of the basic hologram types will then be discussed. In each case the fewest possible space and frequency coordinates consistent with a full understanding of the principles involved will be used. In this way the notation and mathematics will be kept as simple as possible. Extreme rigor will be avoided, but assumptions and their validity will be noted.

3.1 Notation

The origin of a Cartesian \((x, y, z)\) coordinate system is centered in the recording medium (Fig. 3.1). Two-dimensional recording media (thermoplastics, thin photographic emulsions, etc.) define the \(x-y\) plane. A typical object point will be denoted by \((x_o, y_o, z_o)\). If the reference beam is derived from a point source the coordinates of this point will be denoted by \((x_R, y_R, z_R)\). If \(z_R = -\infty\), the reference beam is a plane wave and \(\alpha_\beta\) will be the angle between the projection of the propagation vector \(k\) onto the \(x-z\) plane and the \(z\)-axis. Points in the image formed with the reconstructed wavefront will be denoted by the image coordinates \((x_i, y_i, z_i)\) (Fig. 3.1b). During reconstruction the illuminating wave may be derived from a point source which is not identical to the reference point. We will denote the coordinates of this point by \((x_c, y_c, z_c)\). If the reconstruction is with a plane wave, \(z_c = -\infty\) and we will define \(\alpha_\beta\) as the angle between the projection of
the k vector of the illuminating wave onto the x-z plane and the z-axis, in the same way as with the reference wave.

The amplitude distribution of the light at the object will be denoted by \( F(x_0, y_0, z_0) \). This field gives rise to a disturbance \( O(x, y, z) \) at the hologram. The reference wave at the hologram can be described by \( R(x, y, z) \). The time dependence of these fields has been omitted, and \( O \) and \( R \) are considered to be complex amplitudes. For reconstruction the hologram is illuminated by a wave \( C(x, y, z) \). The wave transmitted by the hologram in general gives rise to an image \( G(x_1, y_1, z_1) \), or \( G_c(x_1, y_1, z_1) \), denoting either the primary or conjugate image, respectively. Note that both object \( (F) \) and image \( (G) \) are also complex amplitudes. The final processed hologram will have a complex amplitude transmission \( t(x, y, z) \) as a result of exposure with irradiance \( I(x, y, z) = |H(x, y, z)|^2 \), where

\[
H(x, y, z) = O(x, y, z) + R(x, y, z) \tag{3.1}
\]

is the total field at the hologram. This notation is shown in Figs. 3.2a and b.

3.2 ANALYSIS

3.2.1 Fresnel Holograms

Having settled on a suitable notation, we can now describe mathematically the recording of the hologram and the subsequent wavefront reconstruction. For the following discussion in this chapter, the recording medium will be considered two-dimensional. These holograms will be called plane holograms. Surface-deformation thermoplastics and photographic emulsions which are thin compared to the spacing of the highest frequency exposure variations are good approximations to plane holograms. A three-dimensional recording medium will be discussed in Chapter 4; these holograms will be called volume holograms.

Consider the arrangement shown in Fig. 3.3. The reference point at
and this is the form which we will use throughout most of this book in order to simplify many of the equations. For the purpose of classifying hologram types, however, we return to (3.2) and write the distance \( r \) as

\[
r = \sqrt{(x_o - x)^2 + (y_o - y)^2 + (z_o - z)^2}.
\]

The reference wave at the hologram plane is given by

\[
R(x, y, z) = R_0 e^{i k z}.
\]

with

\[
s = \sqrt{(x_R - x)^2 + (y_R - y)^2 + (z_R - z)^2}.
\]

For plane holograms, we are interested only in the plane \( z = 0 \), therefore

\[
r = \sqrt{(x_o - x)^2 + (y_o - y)^2 + z_o^2}.
\]

In many cases in holography, \( z_o^2 \sim x_o^2 + y_o^2 \), and we may not use the binomial expansion for (3.6). In this case, the integral (3.2) is extremely difficult and analysis is usually not performed explicitly. When \( z_o^2 > x_o^2, y_o^2 \), we may expand Eq. 3.6 as

\[
r \approx z_o + \frac{x_o^2 + y_o^2}{2z_o} + \frac{x^2 + y^2}{2z_o} - \frac{s x_o + y y_o}{z_o} - \frac{(x_o - x)^4}{8z_o^3} - \frac{(y_o - y)^4}{8z_o^3} + \ldots
\]

When we can neglect quadratic and higher order terms in \( x_o \) and \( y_o \), we speak of Fraunhofer holograms (for plane reference waves); when the quadratic terms cannot be neglected, we speak of Fresnel holograms. We will consider the latter first.

If we write \( O(x, y) = O_o(x, y)e^{i \phi_o(x, y)} \) and \( R(x, y) = R_o(x, y)e^{i \phi_R(x, y)} \), where \( O_o \) and \( R_o \) are real and \( \phi_o \) and \( \phi_R \) describe the spatial phase variation of the object and reference beams, respectively, then Eq. 3.1 becomes

\[
H(x, y) = O_o e^{i \phi_o} + R_o e^{i \phi_R}.
\]

From (2.2), this field produces an exposure

\[
E(x, y) = |H(x, y)|^2
\]

which gives rise to a final amplitude transmittance

\[
I(x, y) = \beta |H(x, y)|^2
\]

\[
= \beta O_o^2 + R_o^2 + \bar{O_o} R_o e^{i(\phi_o - \phi_R)} + \bar{O_o} R_o e^{-i(\phi_o - \phi_R)}. \quad (3.10)
\]

If we illuminate the hologram with a wave

\[
C(x, y) = C_o(x, y)e^{i \phi_c(x, y)},
\]

\[C(x, y) = C_o(x, y)e^{i \phi_c(x, y)}\]
the transmitted wave at the hologram will be
\[ \psi(x, y) = C(x, y) \cdot t(x, y) \]
\[ = B [C_o O e^{i \varphi_o} + C_o R e^{i \varphi_r} + C_o O R e^{i (\varphi_+ \varphi_{-\varphi})} + C_o O R e^{i (\varphi_+ \varphi_{-\varphi})}] \]
\[ + C_o O R e^{i (\varphi_+ \varphi_{-\varphi})} \]. (3.12)

This is the basic equation of holography. The way in which the phases of the various terms are expressed determines the type of hologram, either Fresnel or Fraunhofer.

\[ \Phi(x, y) = C(x, y) \cdot t(x, y) \]

Fig. 3.4 An elementary example: recording a spherical object wave using a spherical reference wave.

To see more clearly how the wavefront reconstruction process works, let us analyze in detail a specific example. A suitable simple arrangement is shown in Fig. 3.4. Here we are considering a single point object and a single point reference source. Both sources emit spherical waves, therefore
\[ O(x, y) = B e^{i k r} \]
and
\[ R(x, y) = A e^{i k r} \]. (3.13)

\[ \Phi(x, y) = C(x, y) \cdot t(x, y) \]

A and B are the initial amplitude of the reference and object waves, respectively. This result can be obtained with the use of (3.2); e.g., see [3], p. 369. Assume that A and B are essentially constant over the area of the hologram. The distances \( r_o, r, s_o \), and \( s \) are given by
\[ r = [(x - x_o)^2 + (y - y_o)^2 + z_o^2]^{1/2} \]
\[ r_o = [(x_o^2 + y_o^2 + z_o^2)^{1/2}] \]
\[ s = [(x - x_R)^2 + (y - y_R)^2 + z_R^2]^{1/2} \]
\[ s_o = [(x_R^2 + y_R^2 + z_R^2)^{1/2}] \]. (3.15)

From (3.13) and (3.15) we have
\[ \phi_o(x, y) = k(r - r_o) \]
\[ = \frac{2\pi}{\lambda} \left\{ ((x - x_o)^2 + (y - y_o)^2 + z_o^2) - ((x_R^2 + y_R^2 + z_R^2) / \lambda) \right\} \] (3.19)

which is the phase of the object wave in the hologram plane relative to the origin. Expansion of the square roots yields
\[ \phi_o(x, y) = \frac{2\pi}{\lambda} \left\{ \frac{1}{2z_o} (x^2 + y^2 - 2zx_o - 2yy_o) + \ldots \right\} \]. (3.20)

The terms that have been omitted in (3.20) contain \((1/z_o)^3\), \((1/z_o)^5\), and so on. For our present purposes we are interested only in the first-order terms; the higher order terms in some cases are not negligible, however, and they give rise to aberrations of the reconstructed wave. These higher order terms will be discussed in Chapter 5.

A similar expression is obtained for the phase of the reference wave
\[ \phi_R(x, y) = \frac{2\pi}{\lambda} \left\{ \frac{1}{2z_R} (x^2 + y^2 - 2zx_R - 2yy_R) + \ldots \right\} \]. (3.21)

Similarly, for the illuminating wave,
\[ \phi_I(x, y) = \frac{2\pi}{\lambda} \left\{ \frac{1}{2z_e} (x^2 + y^2 - 2zx_e - 2yy_e) + \ldots \right\} \]. (3.22)

In the third term of Eq. 3.12, these phases add in the combination
\[ \varphi = \varphi_R + \varphi_o = \frac{\pi}{\lambda} \left\{ (x^2 + y^2) \left( \frac{1}{z_o} - \frac{1}{z_R} + \frac{1}{z_e} \right) - 2x \left( \frac{x_o}{z_o} - \frac{x_R}{z_R} + \frac{x_e}{z_e} \right) - 2y \left( \frac{y_o}{z_o} - \frac{y_R}{z_R} + \frac{y_e}{z_e} \right) \right\} \]. (3.23)

Following Meier [4], we now consider this as representing the first-order term of the reconstructed spherical wave,
\[ \psi^{(1)} = \frac{\pi}{\lambda} \frac{\left( x^2 + y^2 - 2xX_p - 2yY_p \right)}{Z_p} \]
with \( Z_p \) the radius and \( X_p, Y_p \) the center coordinates of the spherical wave.
The subscript \( p \) indicates that we are referring to the primary wave. From (3.23) and (3.24),

\[
\begin{align*}
Z_p &= \frac{Z_o z_R \epsilon_c}{Z_R \epsilon - z_c \epsilon + Z_o \epsilon_R} \\
X_p &= \frac{X_o z_R \epsilon_c - X_o z_c \epsilon + X_o \epsilon \epsilon_R}{Z_R \epsilon - z_c \epsilon + Z_o \epsilon_R} \\
Y_p &= \frac{Y_o z_R \epsilon_c - Y_o z_c \epsilon + Y_o \epsilon \epsilon_R}{Z_R \epsilon - z_c \epsilon + Z_o \epsilon_R}
\end{align*}
\]

(3.25)

If the illuminating wave is identical with the reference wave, then \( z_c = z_R \), \( x_c = x_R \), \( y_c = y_R \) and

\[
\begin{align*}
Z_p &= z_o; \quad X_p = x_o; \quad Y_p = y_o
\end{align*}
\]

(3.26)

so that the reconstructed wave exactly corresponds to the original object wave.

Performing the same set of operations for the fourth term of Eq. 3.12 yields

\[
\begin{align*}
Z_c &= -\frac{Z_o \epsilon - z_c \epsilon + Z_o \epsilon_R}{Z_R \epsilon - z_c \epsilon + Z_o \epsilon_R} \\
X_c &= \frac{X_o \epsilon - x_c \epsilon + X_o \epsilon \epsilon_R}{Z_R \epsilon - z_c \epsilon + Z_o \epsilon_R} \\
Y_c &= \frac{Y_o \epsilon - y_c \epsilon + Y_o \epsilon \epsilon_R}{Z_R \epsilon - z_c \epsilon + Z_o \epsilon_R}
\end{align*}
\]

(3.27)

where the subscript \( c \) means that we are referring to the conjugate wave. For the illuminating wave identical with the reference wave, \( z_R = z_o \), \( x_R = x_o \), \( y_R = y_o \), then

\[
\begin{align*}
Z_c &= -z_o \\
X_c &= \frac{2 x_R \epsilon - x_o \epsilon_R}{2 z_o - z_R} \\
Y_c &= \frac{2 y_R \epsilon - y_o \epsilon_R}{2 z_o - z_R}
\end{align*}
\]

(3.28)

so that this wave yields an image on the opposite side of the hologram plane from the object, and displaced from the axis by an amount which depends on the original object and reference point coordinates.

Whether the images are real or virtual depends on the sign of \( Z_o \) and \( Z_c \).

Since \( z_o \) is always negative (because of the way in which we have set up the coordinate system) a negative \( Z_o \) or \( Z_c \) means a virtual image. Since \( z_R \) and \( z_c \) are arbitrary, both images may be real or virtual.

When the illuminating wave is not identical to the reference wave, the first-order image positions are given by (3.25) and (3.27). We shall see later that this gives rise to magnification along with aberrations. Magnification and wavefront aberrations also occur when the wavelength of illuminating beam is not the same as the wavelength of the reference beam.

One very common arrangement used in holography involves a plane reference wave. In this case \( z_R = \infty \) and we denote the direction of the wave by \( \alpha_R \), the angle between the propagation vector \( k \) and the \( z \)-axis in the \( x \)-\( z \) plane. Analogously, we reconstruct with a plane wave \( (z_c = \infty) \) at an angle \( \alpha_c \) (Fig. 3.5). In this situation we have

\[
\begin{align*}
\varphi_R &= \frac{2 \pi}{\lambda} x \sin \alpha_R \\
\varphi_c &= \frac{2 \pi}{\lambda} x \sin \alpha_c
\end{align*}
\]

(3.29)

so that

\[
\begin{align*}
\varphi_o - \varphi_R + \varphi_c &= \frac{2 \pi}{\lambda} \left[ \frac{x^2 + y^2}{z_o} - 2 x \left( \frac{x_o}{z_o} - \sin \alpha_R + \sin \alpha_c \right) - 2 y \frac{y_o}{z_o} \right]
\end{align*}
\]

(3.30)

and

\[
\begin{align*}
Z_p &= z_o; \\
x_c &= x_o - z_o \sin \alpha_R + z_o \sin \alpha_c; \\
y_c &= y_o.
\end{align*}
\]

(3.31)

For the illuminating beam identical with the reference beam \((\sin \alpha_R = \sin \alpha_c)\), one of the reconstructed waves exactly corresponds to the original object wave. For the conjugate image we have

\[
\begin{align*}
(x_o, z_o)
\end{align*}
\]

Fig. 3.5 Recording a spherical object wave using a plane reference wave.
42 General Theory of Plane Holograms

\[ \varphi_R - \varphi_0 + \varphi_c = \frac{2\pi}{\lambda} \left[ x \sin \alpha_R - \frac{1}{2z_o} (x^2 + y^2 - 2xx_o - 2yy_o) + x \sin \alpha_c \right] \] (3.32)

and

\[ \begin{align*}
Z_o &= -z_o \\
x_c &= x_o + z_o \sin \alpha_R + z_o \sin \alpha_c \\
y_c &= y_o;
\end{align*} \] (3.33)

therefore both images are displaced from the axis and from each other by amounts which depend on the offset angle of reference and illuminating beams.

### 3.2.2 Fraunhofer Holograms

The designation *Fresnel hologram* has been applied where the terms of order two or greater in \( x_o \) and \( y_o \) may be neglected in the expansion (3.7). Generally, a Fresnel hologram is formed when the object is near the hologram plane. When the object size is small compared to its distance \( z_o \) from the hologram plane, we need consider only the first-order terms in \( x_o \) and \( y_o \) in Eq. 3.7. In this case, we speak of a Fraunhofer hologram if the reference wave is plane. The obliquity factor of Eq. 3.2 is nearly unity and

\[ r \sim z_o + \frac{x^2 + y^2}{2z_o} - \frac{xx_o}{z_o} - \frac{yy_o}{z_o}. \] (3.34)

The denominator \( r \) in the integral of (3.2) may be written as \( z_o \) and taken outside the integral; therefore (3.2) becomes

\[ O(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{E(x_o, y_o)}{z_o} \exp \left[ -ik \frac{(xx_o + yy_o)}{2z_o} \right] dx_o dy_o \] (3.35)

for a *plane* hologram. Strictly speaking, we can neglect the second and higher order terms in \( x \) and \( y \) only in the limiting case \( z_o \to \infty \). Equation 3.35, however, is valid for the situation where a large, well-corrected lens is placed a focal length away from the object and the \( x, y \) plane is in the rear focal plane of the lens. To understand why this is true, consider the two situations shown in Fig. 3.6. In Fig. 3.6a we see that the disturbance in the \( x, y \) plane, a great distance from the object, can be considered as arising from the superposition of plane waves originating from each point of the object and traveling in the direction defined by the angle \( \alpha_o \). If a well-corrected lens is placed a focal length away from the object (Fig. 3.6b), however, all of the light from the object traveling at an angle \( \alpha_o \) will come to a focus at a point \( x \) in the focal plane of the lens. Since the optical path from the wavefront of the wave traveling at an angle \( \alpha_o \) to \( x \) is the same for all rays, we obtain essentially the same interference effects predicted by (3.35). The reason for placing the lens a distance \( f \) (one focal length) from the object is discussed in Appendix A.

Hence we can write Eq. 3.35 as

\[ O(x, y) = C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ -i \frac{k}{f} (xx_o + yy_o) \right] dx_o dy_o \] (3.36)

**Fig. 3.6** Imaging the far field with a lens. (a) Illustrating the desired far field distribution. (b) Showing how the desired field distribution forms in the focal plane of a lens.
where \( C = -(i/2z_o) e^{i\beta} \).

By suitably defining the object function \( F(x_o, y_o) \) so that it goes to zero beyond the object, the limits on the integral can be extended to \( \pm \infty \). In this case, we see that, apart from a constant, the amplitude distribution of the hologram plane from the object is very nearly the two-dimensional Fourier transform of the object distribution (see Appendix A). This same result is obtained, of course, when \( z_o \) is very large compared with \( (x^2 + y^2)^{1/2} \). Holograms of this sort are often called “Fourier transform” holograms, when the reference wave is plane.

This type of hologram has some very interesting properties. Suppose the object consists of a single point at \( x_o = a \) (considering only the two-dimensional example as in Fig. 3.7). Then \( F(x_o, y_o) = \delta(x_o - a) \), where \( \delta(x) \) denotes the Dirac delta function (see ref. [3], p. 752), and

\[
O(x) = \text{const} \times e^{-ikf/\lambda x} \tag{3.37}
\]

This represents a plane wave incident on the hologram plane at an angle \( \alpha_o = \sin^{-1}(a/f) \).

If the reference wave is now a plane wave striking the hologram plane at an angle \( \alpha_R \), we have

\[
\delta(x) = \varphi_R(x) - \varphi_o(x) = kx(\sin \alpha_R - \sin \alpha_o) \tag{3.38}
\]

for the phase difference between the two waves. There will be a bright fringe for \( \delta = 2m\pi \) where \( m \) is an integer, or for

\[
x = \frac{m(2\pi/k)}{\sin \alpha_R - \sin \alpha_o} \tag{3.39}
\]

where \( k = 2\pi/\lambda \). Thus a single object point in a Fourier transform system forms a set of straight fringes of spacing

\[
\Delta x_f = \Delta m \frac{\lambda}{\sin \alpha_R - \sin \alpha_o} = \frac{\lambda}{\sin \alpha_R - \sin \alpha_o} \tag{3.40}
\]

The inverse of this fringe spacing, \( 1/\Delta x_f \), is the basic spatial frequency recorded on the hologram. Figure 3.8 shows how the spatial frequency \( v_f \) varies as a function of the total angle \( \varphi \) between the two beams for two common situations. In (1), the normal to the recording plane bisects the angle formed by the two beams so that \( \alpha_o = -\alpha_R \); in (2), one of the beams is incident normally, \( \alpha_o = 0 \). The wavelength used for the computation is the He-Ne laser wavelength, .63 \( \mu \). If the hologram is a Fresnel hologram, there are, of course, many values of \( \alpha_o \)—in this case the curves of Fig. 3.8 are still useful if we call \( \alpha_o \) the angle subtended at the hologram plane by the center of the object—\( v_f \) will then be the average spatial frequency of the hologram.

Returning to the problem at hand, we have, as usual, the total field at the hologram given by (in one dimension)

\[
H(x) = O_d(x)e^{i\varphi(x)} + R_d(x)e^{i\varphi_o(x)} \tag{3.40}
\]
which can be written

\[ H(x) = \frac{O_e e^{-i k \sin \alpha_o}}{2} + \frac{R_o e^{-i k \sin \alpha_o}}{2}, \]

(3.41)

so that the exposure is [by (3.9)]

\[ E(x) = \left[ O_e^2 + R_o^2 + O_e R_o e^{-(ik \sin \alpha_o - ik \sin \alpha_R)} + O_e R_o e^{-(ik \sin \alpha_R - ik \sin \alpha_o)} \right]. \]

(3.42)

We will assume the amplitude transmittance to be given as in Eq. 2.2:

\[ I(x) = \beta \frac{E(x)}{\cos(\phi(x))} = C e^{-i k \sin \alpha_o}, \]

(3.43)

then the transmitted wave is of the form

\[ \psi(x) = \beta \left[ C_o O_o e^{-i k \sin \alpha_o} + C_o R_o e^{-i k \sin \alpha_R} \right. \]

\[ + C_o O_o R_o e^{-i k (\sin \alpha_o - \sin \alpha_R)} \]

\[ \left. + C_o O_o R_o e^{-i k (\sin \alpha_R - \sin \alpha_o)} \right]. \]

(3.44)

Each of these terms represents a plane wave, since the phase advances linearly with \( x \). The two waves represented by the first two terms are traveling in the same direction as the illuminating wave. This is the zero-order wave from the gratinglike pattern on the hologram. The third and fourth terms are equal amplitude plane waves traveling in two different directions which are symmetrical about the zero order. These represent the two first orders of the grating and they both image the object point at infinity. If \( \alpha_o = \alpha_R \), then we see that the third term represents the reconstructed object wave. If the recording is linear, as has been assumed, there are no higher diffracted orders since the grating transmittance is sinusoidal. This can be seen from (3.42) which can be written

\[ E(x) = \left[ O_e^2 + R_o^2 + 2 O_e R_o \cos \left( kx \sin \alpha_o - \sin \alpha_R \right) \right]. \]

(3.45)

Figures 3.9a and b show schematically the recording and reconstructing of a single-point Fourier transform hologram.

The extension to more than a single object point is simple if we extend the diffraction grating analogy. Each object point yields a plane wave at the hologram plane which interferes with a plane reference wave, producing a multiplicity of coherently superposed sinusoidal gratings. Each grating then forms two diffracted first orders which form the two images at infinity of each object point. The modulation, or fringe visibility, of each of the component gratings decreases as the number of object points increases, but as we will see in Chapter 8 when discussing information storage capabilities, the allowable object size (or number of object points) becomes quite large before the modulation decreases to a point where no hologram can be recorded.

There is one other special recording arrangement for producing Fourier transform holograms which deserves mention here. If the reference beam is derived from a point source that is in the same plane as the object, and no lenses are used, the sphericities of the two spherical waves at the hologram plane (one from an object point, the other from the reference point) tend to cancel. Thus an approximately constant-spatial-frequency grating is produced for each object point and we have a "lensless Fourier transform hologram" [5]. To see how this comes about, consider the arrangement of Fig. 3.10. For an object point at \( x_o \) and the reference point at \( x_R \), the angle between the two rays at the hologram is \( \beta(x) \), where

\[ \beta(x) = \tan^{-1} \left( \frac{x_o - x}{z_o} \right) + \tan^{-1} \left( \frac{x - x_R}{z_o} \right) \]

\[ = \frac{x_o - x_R}{z_o} - \frac{1}{3 z_o^3} \left[ x_o^3 - x_R^3 - 3 x(x_o^2 - x_R^2) + 3 x_R^2(x_o - x_R) \right] + \cdots \]

(3.46)
which does not vary strongly with \( x \) for \( x/z_x \ll 1 \). Since the angle between the two interfering beams is approximately constant across the hologram, so is the spacing of the fringes: hence we have an approximate Fourier transform hologram, although not in the sense that the field distribution at the hologram, caused by the object is the Fourier transform of the object distribution.

### 3.3 PHASE HOLOGRAMS

#### 3.3.1 Introduction

Thus far we have considered only absorption holograms, for which the light of the illuminating beam is absorbed in correspondence with the recording exposure. We have assumed that the transmission function \( t(x) \) was a real valued function. We will now show that holograms can also be formed on recording media for which \( t(x) \) is a complex function, that is, the hologram alters the phase of the illuminating wave in correspondence with the recording exposure. Such a hologram is termed a “phase hologram” and has some interesting and important properties.

In order to demonstrate most clearly the principles of phase holograms, we will consider only the case of pure phase modulation wherein \( t(x) \) is pure imaginary. Holograms can be produced with \( t(x) \) having both real and imaginary parts—some absorption and some phase modulation (indeed most of the holograms produced are probably of this type since it is difficult to eliminate phase modulation completely).

Pure phase holograms can be made by many methods. The earliest reference to phase holograms describes their preparation by contact printing of Gabor holograms, employing the Carbro process to produce relief images in transparent gelatin on glass substrates [6]. Phase holograms may be produced by contact printing (or by direct recording with blue or green laser light) onto a resist material. They may be produced by using a thermoplastic material as the recording medium [7]. It is also possible to make phase holograms with conventional silver halide photographic emulsions—either by utilizing the relief image or index change, or both. Relief images are an imagewise emulsion thickness variation caused by the differential tanning action of the developer [8]. These thickness variations result in the desired spatial phase modulation of the transmitted light. To obtain a pure phase modulation using this technique, one can bleach out the exposed silver grains so that no density variation remains [9], or one can metallize the emulsion surface and view the hologram image in reflection [10]. Phase holograms may be produced by utilizing a change in the index of refraction of the recording medium. In the photosensitive materials a change in index results from a photopolymerization of the resist material upon exposure. There are now no exposurewise thickness variations, but the signal is recorded in the form of index variations that impose the desired phase modulation onto the illuminating beam. A bleached photographic hologram also shows evidence of index change in the exposed areas.

These are just some of the methods for producing phase holograms which are known at this time. There will undoubtedly be more methods. The main interest in phase holograms stems from their very high refraction efficiency—the ratio of incident to usable diffracted flux. Burckhardt has shown that the efficiency of a thick dielectric grating can be as high as 100% [11]. For a thin phase grating, the efficiency can be as high as 34% [12]. Compare these numbers with a little over 6% for the usual absorption type of gratings (or hologram).

#### 3.3.2 Analysis

For simplicity, we will restrict our attention to pure phase holograms—those for which only the phase and not the amplitude of the illuminating wave is affected.

To begin, we suppose that the recording medium is such that an exposure \( E(x) \) results in either a change in index \( n(x) \) or a change in thickness \( h(x) \) of the medium. If the exposed hologram is illuminated with a wave \( C(x) \), then the transmitted wave will have a phase variation imposed on it of the form \( knh(x) \) or \( kni(x) \) for the relief or index types, respectively. This im-
posed phase modulation yields various diffracted orders, the \( \pm 1 \)st orders producing the desired reconstructions. For thick holograms, one can arrange to produce only one diffracted order, and one can eliminate the zero order by a suitable adjustment of the phase shift.

Next we expose the recording medium to an object wave described by

\[
O(x) = O_0 e^{i\phi(x)}
\]

and a plane reference wave

\[
R(x) = R_0 e^{i\beta x}
\]

where \( \beta = k \sin \alpha_R \), as shown in Fig. 3.11. Let us further assume that the final phase modulation to be imposed on the illuminating wave \( C(x) \) is simply equal to the exposure \( E(x) \). Call this phase modulation \( \varphi(x) \) so that we can write

\[
\varphi(x) = E(x),
\]

where \( E(x) \) can be written as (ignoring the proportionality constant)

\[
E(x) = |O(x) + R(x)|^2
\]

\[
= O_0^2 + R_0^2 + O_0 R_0 e^{i(\varphi_0 - \beta x)} + O_0 R_0 e^{-i(\varphi_0 - \beta x)}
\]

\[
= O_0^2 + R_0^2 + 2 O_0 R_0 \cos(\varphi_0 - \beta x).
\]

The complex transmittance function then becomes

\[
l(x) = E(x) = e^{i\varphi(x)}
\]

\[
= e^{iO_0^2 e^{i(\varphi_0 - \beta x)}}
\]

\[
= t_o e^{iO_0^2 e^{i2\varphi_0 e^{i(\varphi_0 - \beta x)}}}
\]

For simplification of notation, we write

\[
K = t_o e^{iO_0^2 e^{i2\varphi_0}}
\]

\[
a = 2O_0 R_0
\]

\[
\theta = \varphi_0 - \beta x
\]

so that

\[
l(x) = K e^{i\alpha_c \theta}
\]

Using the Bessel function expansions

\[
\cos(a \cos \theta) = J_0(a) + 2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(a) \cos(2n\theta)
\]

\[
\sin(a \cos \theta) = 2 \sum_{n=0}^{\infty} (-1)^n t^{2n+1} J_{2n+1}(a) \cos((2n+1)\theta)
\]

and the relation

\[
ed^{ix} = \cos x + i \sin x
\]

we can write

\[
l(x) = K \left\{ J_0(a) + 2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(a) \cos(2n\theta)
\]

\[
+ 2i \sum_{n=0}^{\infty} (-1)^n J_{2n+1}(a) \cos((2n+1)\theta) \right\}
\]

The \( J_0(a) \) are Bessel functions of the first kind. Each \( J_n(a) \) is the amplitude of the \( n \)th diffracted order. Note that in general all orders are present, unlike the case of an absorption hologram where the sinusoidal amplitude modulation led to only the \( \pm 1 \)st orders. The term of (3.56) leading to the images of interest is

\[
K[2J_1(a) \cos \theta],
\]

which can be written as

\[
2iKJ_1(a) \left( e^{i\theta} + e^{-i\theta} \right) = iKJ_1(a) [e^{i(\varphi_0 - \beta x)} + e^{-i(\varphi_0 - \beta x)}].
\]

These two terms represent the primary and conjugate image waves. The transmission term leading to the primary image is

\[
t_p(x) = t_p \exp \left[ i \left( O_0^2 + R_0^2 + \frac{\pi}{2} \right) \right] J_0(2O_0 R_0) e^{i(\varphi_0 - \beta x)}
\]

If the hologram is illuminated with a wave \( C(x) = R(x) \) (Fig. 3.12) we obtain the primary image wave...
\[ \psi_p(x) = C(x) \psi(x) = R_o e^{i \alpha x} \psi(x) \]
\[ = \int e^{i(\alpha_0^2 + R_o^2 + \pi/2) x} R_o \int_j(2O_o R_o) e^{i\varphi} \]  
\[ \text{(3.60)} \]

Since

\[ J_1(x) = \frac{x}{2} - \frac{(\frac{x}{2})^3}{1!2} + \frac{(\frac{x}{2})^5}{5!2^5} - \cdots \]
\[ = \frac{x}{2}, \quad \text{for small } x, \]

we have, if \( 2O_o R_o \) is small,

\[ \psi_p(x) \approx \frac{1}{2} e^{i(\alpha_0^2 + R_o^2 + \pi/2)} O_o e^{i\varphi} \]
\[ \text{(3.61)} \]

which is seen to be, aside from the unimportant phase and amplitude factors, just the original object wave. If the product \( 2O_o R_o \) is large, some amplitude distortion will be present. From (3.49) and (3.50), we see that \( 2O_o R_o \) is the amplitude of the phase modulation

\[ \varphi(x) = \alpha_0^2 + R_o^2 + 2O_o R_o \cos(\varphi_0 - \beta x) \]
\[ \text{(3.63)} \]

\[ \varphi(x) = kn \left[ h_o + \frac{h_1}{2} \cos(\varphi_0 - \beta x) \right] \]
\[ \text{(3.65)} \]

where \( h_1 \) is the peak-to-peak variation of the hologram thickness and the average thickness is \( h_o \) as shown in Fig. 3.13. A distortion-free primary (or conjugate) image may be formed with a phase hologram if the phase modulation is small; in the foregoing example, \( h_1 \) should be kept small.

REFERENCES


Hence if the phase hologram is of the relief image type we might have, for example,

\[ \varphi(x) = kn h(x) = \alpha_0^2 + R_o^2 + 2O_o R_o \cos(\varphi_0 - \beta x) \]
\[ \text{(3.64)} \]

which can be written in the form