With most new or advancing technologies, the period of discovery and initial investigation of fundamental principles is immediately followed by a period of intensive search for materials. The recent discoveries of the transistor and laser are two prime examples of this type of technological evolution. A third example is the invention and recent renaissance of holography. Following the revival in 1962–64, there came a prolonged (and indeed, ongoing) search for an ideal recording material. The purpose of this monograph is to present the results of this search.

1 History

The history of holography begins with its invention by Gabor [1.1] in 1948. His ideas and demonstrations form the foundation of modern holography. Because of the severe limitations imposed by having to work with a totally inadequate light source, among other reasons, few practical uses were found for holography, even though it represented a fundamentally different method of image formation.

However, in the period 1962–64 Leith and Upatnieks [1.2, 3] published a series of papers describing methods for overcoming two of the major problems associated with the earlier techniques—the limited coherence and intensity of the light and the very problematical mixing of the signal (image forming) and illuminating waves during readout. These papers created a great deal of excitement and the whole field of holography underwent a rebirth. The coherence problem was solved by the invention of the laser. Its use as a source for holography allowed for the diffuse illumination of large, three-dimensional objects and thus the ability to view the reconstruction, complete with parallax and depth effects, with the unaided eye. The use of the laser also permitted great latitude and flexibility in the choice of object to be used and of the physical layout of the recording setup—a degree of freedom undreamed of by the early workers in the 1950’s [1.4–11].

The second problem, that of the mixing of the signal and illuminating waves, was solved by Leith and Upatnieks by the use of an off-axis (noncollinear) reference beam to record the hologram. This allowed reconstruction with a similarly off-axis illuminating beam. Since this light was traveling in a substantially different direction, it no longer obscured the image-forming light.
By this simple technique they were able to solve one of the major problems with the earlier forms of holography.

At about the same time, Denisyuk [1.12] in Russia and Van Heerden [1.13] in the USA were expanding the basic 1948 concepts to include the recording of waves that could be reconstructed by reflection of white light and waves that could be recorded throughout a large volume of the recording material.

As a result of all of the aforementioned innovations, new life was given to the old idea of holography. Not only was its practicality well demonstrated, but the striking three-dimensional images that could be produced fired the imaginations of many thousands of students and scientists throughout the world. The vast amount of activity in the field during the time span of roughly 1964 through 1970 resulted in many new ideas for application, creating a demand for increasingly better recording materials, a demand that in large part has been satisfied.

1.2 Basic Description

If \( O \) represents a monochromatic wave from the object, and \( R \) a reference wave coherent with \( O \), the total field at the recording medium is \( O + R \) so that a square-law recording medium responds to the irradiance \( |O + R|^2 \). After processing, the hologram recording material has a certain complex amplitude transmittance \( T(x) \) that can be expressed as a function of the exposure \( E(x) = |O + R|^2 \) (\( t \) is the exposure time):

\[
T(x) = \beta t |O + R|^2 = \beta t (|O|^2 + |R|^2 + OR^* + O^*R)
\]

where the * denotes complex conjugate and \( \beta \) depends on the recording material involved. When the hologram with this amplitude transmittance is illuminated with a wave \( C \), the transmitted wave at the hologram is

\[
\psi(x) = CT(x) = \beta t (C|O|^2 + C|R|^2 + CR^*O + CRO^*)
\]

If the illuminating wave \( C \) is sufficiently uniform so that \( CR^* \) is approximately constant across the hologram, the third term of (1.2) becomes simply a constant \( \times O \). Thus this term represents a reconstructed wave that is identical to the original object wave \( O \). The other terms represent the zero-order wave \( \beta t (|O|^2 + |R|^2) \) and the other first-order wave \( \beta t RO^* \). The latter is known as the “conjugate” wave. These terms represent the waves that caused the problems referred to above with the early Gabor holograms. But because of the offset reference beam associated with this type of hologram, these waves are spatially separated from the desired reconstruction wave. This is the main advantage of the off-axis technique and is the principal reason for the resurgence of interest in holography in 1964.
1.3 Classification

Holograms are classified as to type with the recording geometry, the type of modulation imposed on the illuminating wave, the thickness of the recording material, and the mode of image formation as factors.

If the two interfering beams are traveling in substantially the same direction, the recording of the interference pattern is said to be a Gabor hologram or an in-line hologram. If the two interfering beams arrive at the recording medium from substantially different directions, the recording is a Leith-Upatnieks or off-axis hologram. If the two interfering beams are traveling in essentially opposite directions, the recorded hologram is said to be a Lippmann or reflection hologram.

These hologram types are further classified, depending on recording geometry, as Fresnel, Fourier Transform, Fraunhofer, or Lensless Fourier Transform. Generally speaking, if the object is reasonably close to the recording medium, say 10 or so hologram or object diameters distant, the field at the hologram plane is the Fresnel diffraction pattern of the object, and the hologram thus recorded is a Fresnel hologram. If, on the other hand, the object and recording medium are separated by many hologram or object diameters, the field at the recording plane is the Fraunhofer or far field diffraction pattern of the object and a Fraunhofer hologram results. If a lens is used to produce the far field pattern at the recording plane, with the lens a focal length distant from both the object and recording plane, then a Fourier Transform hologram results, provided the reference wave is plane. Lastly, if the reference beam originates from a point that is the same distance from the recording plane as the object, then a so-called Lensless Fourier Transform hologram results.

Any of the foregoing hologram types may also be recorded as a thick or thin hologram. A thin (or plane) hologram is one for which the thickness of the recording medium is small compared to the fringe spacing. A thick (or volume) hologram is one for which the thickness is of the order of or greater than the fringe spacing. The distinction between thick and thin holograms is usually made with the aid of the $Q$-parameter defined as

$$ Q \equiv \frac{2\pi \lambda d}{n \Lambda^2}, $$

where $\lambda$ is the illuminating wavelength, $n$ the index of the recording material, $d$ its thickness, and $\Lambda$ the spacing of the recorded fringes. The hologram is considered thick when $Q \geq 10$ and thin otherwise, although it has recently been shown [1.14] that thick hologram theory (coupled wave theory) is quite accurate even for $Q$ values of order 1.

Finally, holograms are also classified by the mechanism by which the illuminating light is diffracted. In an amplitude hologram the interference pattern is recorded as a density variation of the recording medium, and the amplitude of the illuminating wave is accordingly modulated. In a phase hologram, a phase modulation is imposed on the illuminating wave.
The amplitude transmittance of a hologram is a complex function that describes the change in the amplitude and phase of an illuminating wave upon transmission through the material. If the illuminating field is described by the complex function \( C(x) = C_0(x) \exp[i\phi_0(x)] \) and the transmitted field by \( \psi(x) = b_t(x) \exp[i\phi_t(x)] \) then the amplitude transmittance of the hologram is defined as the ratio of these two quantities:

\[
T(x) = \frac{\psi(x)}{C(x)} = \left| \frac{b_t(x)}{C_0(x)} \right| e^{i(\phi_t - \phi_0)}
\]

In most instances, the illuminating wave \( C \) is constant or nearly so, so that the complex amplitude transmittance is given by

\[
T(x) = b_t(x) e^{i\phi_t(x)}
\]

If \( b_t(x) \) is not constant but \( \phi_t(x) \) is, it is the amplitude of the illuminating wave that is modulated and the hologram is an amplitude hologram. The magnitude of \( b_t(x) \) depends on the absorption constant \( a \) and the thickness of the recording material according to

\[
b_t(x) = C_0(x) e^{-a(x)d}
\]

Thus for an amplitude hologram the transmitted amplitude is modulated in accordance with the exposure, since the exposure causes a spatial variation of the absorption constant \( a \). The two most common recording materials that can accomplish this are silver halide photographic emulsions (Chap. 2) and photochromic materials (Chap. 5).

If, on the other hand, \( b_t(x) \) is constant and \( \phi_t(x) \) varies, then the phase of the illuminating wave is modulated and the hologram is a phase hologram. The phase of the illuminating wave can be modulated in one or both of two different ways. If a surface relief is found as the result of exposure to the interference fringe pattern, then the thickness of the material is a function of \( x \). If the index of refraction of the material is changed in accordance with the exposure, then the optical path of the transmitted light is a function of \( x \). In either case, the phase of the transmitted light will vary with \( x \) because the optical path \( nd \) will be a function of \( x \). The phase \( \phi_t \) of the transmitted light is related to the optical path according to

\[
\phi_t = \frac{2\pi}{\lambda} nd
\]

A phase hologram results when \( \phi_t \) is a function of the exposure—such as when either the index of refraction \( n \) or the thickness \( d \), or both, is exposure dependent. From (1.7) it is seen that a change \( \Delta \phi_t \) in \( \phi_t \) is given by

\[
\Delta \phi_t = \frac{2\pi}{\lambda} [d \Delta n + (n - 1) \Delta d]
\]
where the $n-1$ factor appears assuming the hologram to be in air. If the hologram is thin, $d$ is very small and the contribution to $\Delta \varphi_i$ from the term $d\Delta n$ is negligible and

$$\Delta \varphi_i = \frac{2\pi}{\lambda} (n-1)d$$

(1.9)

This would be a surface relief hologram, achieved mainly by embossing, etching a photo resist material, or with a thermoplastic film (Chap. 3). For a thick hologram, on the other hand, having a negligible surface relief but a varying index, we have

$$\Delta \varphi_i = \frac{2\pi}{\lambda} d\Delta n$$

Typical materials for this type of hologram are bleached photographic emulsions (Chap. 2), dichromated gelatin (Chap. 3), and ferroelectric crystals (Chap. 4). All of these materials (and others) are discussed in detail in the following chapters.

### 1.4 Diffraction Efficiency

Diffraction efficiency is defined as the ratio of the total, useful, image-forming light flux diffracted by the hologram to the total light flux used to illuminate the hologram. If some of the illuminating light is diffracted into an unwanted conjugate (or primary) image, it is still only the light in the primary (or conjugate) image that is pertinent to the stated diffraction efficiency. The importance of producing a hologram with a reasonable diffraction efficiency is obvious, whether the application is simply the viewing of a three-dimensional scene, holographic interferometry, or data storage; efficient use of the illuminating light allows for smaller and more economical systems. Being able to use a smaller and cheaper laser for illumination or readout, for example, could be the deciding factor in determining the feasibility of a holographic data storage system.

However, one must approach the question of diffraction efficiency with some careful thought. In many cases it is the output signal-to-noise ratio (s/n) that is the more important consideration. The noise in a holographic image arises principally from three factors: intrinsic nonlinearity, such as in phase holograms, that causes unwanted light to be diffracted into the image region; the nonlinearity of the amplitude transmittance-exposure characteristic of the recording medium; and the noise caused by the buildup of the granular structure of the recording material. (A discussion of the sources of noise in holographic images follows in Sect. 5.) It is, however, possible to increase the diffraction efficiency without seriously impairing the s/n characteristics of the recording by a judicious choice of recording material and processing technique.
The equations describing the diffraction efficiency of the various hologram types are divided into two groups, those for plane holograms and those for volume holograms. This is because the description of the diffraction process for the two cases is fundamentally different, even though it has been shown that in the common case of amplitude holograms, for most practical recording situations the resulting diffraction efficiencies are equivalent [115]. This is definitely not the case for phase holograms, however. The distinction between thick and thin holograms is made with the aid of the $Q$-parameter defined earlier [(1.3)]. The hologram is considered to be thick when $Q \gtrsim 10$ and thin otherwise.

1.4.1 Plane Holograms

Amplitude Holograms

We begin by assuming an irradiance distribution at the hologram (for a diffuse object) given by

$$H(x) = |O_o(x)e^{i\phi_0(x)} + R_o e^{i\omega x}|^2
= O_o^2(x) + R_o^2 + 2O_o R_o \cos(\omega x - \phi_0)$$

where $O_o(x)$ and $R_o$ are the (real) amplitudes and $\phi_0(x)$ and $\omega x$ are the phase of the object and reference waves, respectively. The object is assumed to be diffuse so that both the phase and amplitude of the object wave are functions of the space coordinate $x$. The average spatial frequency of this distribution is given by

$$\bar{v} = \frac{1}{2\pi} \left\langle \frac{d}{dx} [\omega x - \phi_0(x)] \right\rangle,$$

where the angle brackets indicate a spatial average. If we take into account the MTF of the recording material, the effective exposure is

$$E = tH = t[O_o^2(x) + R_o^2][1 + mM(\bar{v}) \cos(\omega x - \phi_0)],$$

where $t$ is the exposure time and $m$ is the exposure modulation given by

$$m = \frac{2[\langle O_o^2(x)R_o^2 \rangle]^{1/2}}{\langle O_o^2(x) \rangle + R_o^3}$$

and $M(\bar{v})$ is the MTF. In (1.14) we have used $\langle O_o^2(x) \rangle$ rather than the more rigorous $O_o^2(x)$, and so we are neglecting the self-interference or speckle caused by the diffuse object. However, the effects of the speckle are not excessive as long as the modulation $m$ is not too large. The modulation of the fringes represented by the cosine term of (1.13) is reduced by the factor $M(\bar{v})$, the MTF of the recording material at the average frequency $\bar{v}$. This is a good approximation as
long as the object size is such that it subtends a very small angle at the hologram. If this is the case, \( \varphi_0(x) \) is nearly a constant, or a constant plus a term proportional to \( x \), and the bandwidth of the recorded signal is small and centered around \( \bar{v} \). If we assume the amplitude transmittance-log exposure curve \( T_a - \log E \) to be essentially linear with slope \( \alpha \) in the vicinity of the average exposure \( E_0 = t[\langle O_0^2(x) \rangle + R_0^2] \), we can write for the amplitude transmittance

\[
T_a(E) = T_a(E_0) + \alpha \log(E/E_0),
\]

so that

\[
T_a(x) = T_a(E_0) + \alpha \log[1 + mM(\bar{v})\cos(\omega x - \varphi_0)]
\]

Expanding the logarithmic function, we have

\[
T_a(x) = T_a(E_0) + 0.43\alpha[mM(\bar{v})\cos(\omega x - \varphi_0) - \frac{m^2M^2(\bar{v})}{2}\cos^2(\omega x - \varphi_0) + \frac{m^3M^3(\bar{v})}{3}\cos^3(\omega x - \varphi_0) - \ldots]
\]

The \( T_a - \log E \) curve is linear only in the vicinity of \( E_0 \), which implies a moderately small exposure modulation \( m \). Therefore, we must assume the \( m \) is small enough that we can neglect all except the first-order term in (1.17) and arrive at

\[
T_a(x) = T_a(E_0) + 0.43\alpha mM(\bar{v})\cos(\omega x - \varphi_0)
\]

The object wave is reconstructed by illuminating the hologram with the reference wave \( R_0 \exp(i\omega x) \), yielding a conjugate image term

\[
\psi'_i(x) = \frac{0.434}{2} R_0^2\alpha mM(\bar{v})e^{i\varphi_0}
\]

The irradiance in the reconstructed field is

\[
II_e = R_0^2 \left| \frac{0.434}{2} \alpha mM(\bar{v}) \right|^2
\]

and the diffraction efficiency is

\[
\eta = \frac{II_e}{R_0^2} = \left| \frac{0.434}{2} \alpha mM(\bar{v}) \right|^2
\]
Now
\[ m = 2\sqrt{\frac{\langle O_0^2(x) \rangle R_0^2}{\langle O_0^2(x) \rangle + R_0^2}} = \frac{2 \sqrt{K}}{1 + K} \]  

where \( K \) is the beam ratio \( \frac{R_0^2}{\langle O_0^2(x) \rangle} \). Since we are already restricted to moderately small values of \( m \), \( K \) is somewhat greater than unity, and so \( m \approx \frac{1}{2} \sqrt{K} \) giving for the diffraction efficiency

\[ \eta = 0.188 \frac{\alpha^2}{K} M^2(\nu) \]  

This equation shows that the diffraction efficiency will be a maximum at an average exposure for which the \( T_a - \log E \) curve of the recording material has maximum slope.

Because we have assumed that the exposure modulation \( m \) is not too large, \( (1.23) \) does not indicate the maximum diffraction efficiency obtainable with thin amplitude holograms. To get an idea of this upper limit, we write for the amplitude transmittance of the exposed and processed hologram

\[ T_a(x) = T_a(E_0) + T_1 \cos(\omega x - \varphi_0) \]  

(1.24)

where \( T_a(E_0) \) is the average transmittance due to the average exposure \( E_0 = t[\langle O_0^2(x) \rangle + R_0^2] \) and \( T_1 \) is the amplitude of the spatially varying portion of the exposure.

\[ t[\langle O_0^2(x) \rangle + R_0^2] m M(\nu) \cos(\omega x - \varphi_0) \]

To achieve the maximum possible diffraction efficiency, \( T_a(x) \) should vary between 0 and 1. Thus

\[ T_a(x) = \frac{1}{2} + \frac{1}{2} \cos(\omega x - \varphi_0) \]

\[ = \frac{1}{2} + \frac{1}{4} e^{i(\omega x - \varphi_0)} + \frac{1}{4} e^{-i(\omega x - \varphi_0)} \]  

(1.25)

From this equation it is evident that the maximum diffracted amplitude in either the primary or conjugate image is only one-fourth the amplitude of the illuminating wave. The useful light flux in either image is therefore only one-sixteenth of the illuminating flux, yielding a maximum diffraction efficiency of 0.0625. This is somewhat artificial, however, since there are really no practical materials that are linear over an exposure range so large that \( T_a(x) \) will vary from 0 to 1. Some nonlinearity or clipping will always occur when the exposure modulation gets large. Therefore, the maximum efficiency of 0.0625 cannot be achieved when it is necessary that the reconstructed wave amplitude be proportional to the object wave amplitude.
Phase Holograms

If the phase of the transmitted wave is proportional to the exposure
\[ \varphi_t(x) = \gamma E(x) \]  
then a phase hologram results. If
\[ E(x) \propto |O + R|^2 \]
and
\[ O(x) = O_0 e^{i\varphi_0(x)} \]
\[ R(x) = R_0 e^{i\beta x} \]
then
\[ T_a(x) = h e^{i\varphi_t(x)} \]
\[ = h \exp(i\gamma O_0^2) \exp(i\gamma R_0^2) \]
\[ \cdot \exp\{2i\gamma O_0 R_0 [\cos(\varphi_0 - \beta x)]\} . \]  
(1.29)

Simplifying the notation we set
\[ K \equiv h \exp[i\gamma(O_0^2 + R_0^2)] \]
\[ a \equiv 2\gamma O_0 R_0 \]
\[ \Theta \equiv \varphi_0 - \beta x \]
so that
\[ T_a(x) = K e^{ia\cos \Theta} \]

Now by using the Bessel function expansions
\[ \cos(a \cos \Theta) = J_0(a) + 2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(a) \cos 2n\Theta \]
\[ \sin(a \cos \Theta) = 2 \sum_{n=0}^{\infty} (-1)^{n+1} J_{2n+1}(a) \cos[(2n + 1)\Theta] \]  
(1.32)
and the relation
\[ e^{ix} = \cos x + i \sin x \]
(1.33)
we can write
\[ T_a(x) = K \{J_0(a) + 2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(a) \cos 2n\Theta \]
\[ + 2i \sum_{n=0}^{\infty} (-1)^{n+1} J_{2n+1}(a) \cos[(2n + 1)\Theta]\} \]  
(1.34)
The $J_n(a)$ are Bessel functions of the first kind. Each is the amplitude of the $n$-th diffracted order. The term leading to the images of interest is $K [2iJ_1(a) \cos \Theta]$ which can be written as

$$2iKJ_1(a) \left( \frac{e^{i\Theta} + e^{-i\Theta}}{2} \right) = iKJ_1(a) [e^{i(\phi_0 - \beta x)} - e^{-i(\phi_0 - \beta x)}]$$  \hspace{1cm} (1.35)

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

$$T'_0(x) = h_i \exp [i(\gamma O^2 + \gamma R^2_o + \pi/2)]$$
$$\cdot J_1(2\gamma O_o R_o) \exp[i(\phi_0 - \beta x)].$$  \hspace{1cm} (1.36)

When the hologram is illuminated with a wave of unit amplitude, the first-order diffracted wave has an amplitude $T'_0$ given by (1.36). Since we are assuming for simplicity that we have a pure phase hologram, that is, no absorption, $h_i = 1.0$ and the diffracted wave amplitude is simply $J_1(2\gamma O_o R_o)$. The diffraction efficiency is equal to the square of this quantity. The maximum value is 0.339, which is considerably higher than the maximum efficiency for amplitude holograms. It will be seen from the following sections that the efficiency of thick phase grating is even higher yet, which explains the very great interest in producing low noise phase holograms.

1.4.2 Volume Holograms

Coupled Wave Theory

The preceding theories for thin holograms cannot apply when the diffraction efficiency becomes high, because for high diffraction efficiencies the illuminating wave will be strongly depleted as it passes through the grating. In some way one must take into account the fact that at some point within the grating there will be two mutually coherent waves of comparable magnitude traveling together. Such an account is the basis of the coupled wave theory, so aptly applied to the problem of diffraction from thick holograms by Kogelnik [1.16]. This elegant analysis gives closed form results for the angular and wavelength sensitivities for all of the possible hologram types—transmission and reflection, amplitude or phase, with and without loss, and with slanted or unslanted fringe planes. The equations also give, of course, the maximum diffraction efficiency achievable when the grating is illuminated at the Bragg angle and the fringe planes are unslanted. The equations are the result of a theory that assumes that the gratings are relatively thick so that there are only two waves in the medium to be considered, that is, that the Bragg effects are rather strong. Nevertheless, the equations are surprisingly accurate over a very large range of $Q$-values, including values considerably less than 10.

For a complete treatment of the theory one should refer to [1.16]. Here we will only outline briefly the underlying ideas of the theory and give the principal results.
The coupled wave theory assumes that there are only two waves present in the grating: the illuminating wave $C$ and the diffracted signal wave $S$. It is assumed that the Bragg condition is approximately satisfied by these two waves and that all other orders strongly violate the Bragg condition and hence are not present. The equations that are derived express the coherent interaction between the waves $C$ and $S$.

Figure 1.1 defines the grating assumed by Kogelnik for his analysis. The $z$-axis is perpendicular to the surfaces of the medium, the $x$-axis is in the plane of incidence and parallel to the medium boundaries, and the $y$-axis is perpendicular to the page. The fringe planes are oriented perpendicularly to the plane of incidence and slanted respect to the medium boundaries at an angle $\phi$. The grating vector $K$ is oriented perpendicularly to the fringe planes and is of length $|K| = 2\pi/\lambda$, where $\lambda$ is the period of the grating. The angle of incidence for the illuminating wave $C$ in the medium is $\Theta$. The Bragg angle is $\Theta_0$. It is assumed that the fringes are sinusoidal variations of the index of refraction or of the absorption constant, or both for the case of mixed gratings. The assumed equations are therefore

$$n_0 = n + n_1 \cos(K \cdot X)$$

and

$$a_0 = a + a_1 \cos(K \cdot X),$$

where $K \cdot X = \omega x$ and $\omega$ is $2\pi$ divided by the fringe spacing in the $x$-direction. The quantities $n$ and $a$ are the average values of the index and absorption constant, respectively. The slant of the fringe planes is described by the constants $c_R$ and $c_S$, which are given by

$$c_R = \cos \Theta$$

(1.39)

$$c_S = \cos \Theta \left( \frac{K}{\beta} \right) \cos \Theta$$

(1.40)

where $\beta = 2\pi n/\lambda$, and $\lambda$ is the free space wavelength.
When the Bragg condition is satisfied we have
\[
\cos(\varphi - \Theta_0) = K/2\beta
\]
and
\[
c_R = \cos \Theta_0, c_S = -\cos(2\varphi - \Theta_0).
\]
In the case where the Bragg condition is not satisfied, either by a deviation \(\Delta \Theta\) from the Bragg angle \(\Theta_0\) or by a deviation \(\Delta \lambda\) from the correct wavelength \(\lambda_0\), where both \(\Delta \Theta\) and \(\Delta \lambda\) are assumed to be small, Kogelnik introduces a dephasing measure \(\vartheta\). This parameter is a measure of the rate at which the illuminating wave and the diffracted wave get out of phase, resulting in a destructive interaction between them. The dephasing measure is given by
\[
\vartheta = \Delta \Theta \cdot K \cdot \sin(\varphi - \Theta_0) - \Delta \lambda K^2/4\pi n,
\]
where
\[
\Delta \Theta = \Theta - \Theta_0
\]
and
\[
\Delta \lambda = \lambda - \lambda_0
\]
The coupled wave equations that result from this theory are
\[
\begin{align*}
c_R R' + aR &= -ikS \\
c_S S' + (a + i\vartheta)S &= -ikR
\end{align*}
\]
where \(k\) is the constant defined by
\[
k = \frac{\pi n_1}{\lambda} - \frac{ia_4}{2}
\]
and the primes denote differentiation with respect to \(z\), and \(i = \sqrt{-1}\). The physical interpretation of these equations as described by Kogelnik is as follows:

As the illuminating and diffracted waves travel through the grating in the \(z\)-direction their amplitudes are changing. These changes are caused by the absorption of the medium as described by the terms \(a_R\) and \(a_S\), or by the coupling of the waves through the terms \(kR\) and \(kS\). When the Bragg condition is not satisfied, the dephasing measure \(\vartheta\) is nonzero and the two waves become out of phase and interact destructively.

The solutions to these equations take on various forms depending on the type of hologram (grating) considered. When transmission holograms are considered, we say the fringe planes are unslanted when they are perpendicular
Basic Holographic Principles

Fig. 1.2. Diffraction efficiency as a function of the optical path variation $n_1d/cos\theta_0$ in units of the wavelength for a thick phase hologram viewed in transmission.

to the surface. This condition is described by $c_R = c_S = \cos \Theta$ since $\phi = \pi/2$ in this case. If the Bragg condition is also satisfied, then of course $\Theta = \Theta_0$. For reflection holograms, on the other hand, no slant means that the fringe planes are parallel to the surface, so $\phi = 0$. If the Bragg condition also holds, then $c_R = -c_S = \cos \Theta_0$.

Transmission Holograms

Pure Phase Holograms. For a pure phase, transmission hologram the absorption constant $a_0 = 0$ and the solution of the coupled wave equations leads to a diffraction efficiency, for the general case of slanted fringes and Bragg condition not satisfied, given by

$$\eta = \sin^2(v^2 + \zeta^2)^{1/2}/(1 + \zeta^2/v^2)$$

$$v = \pi n_1 d / \lambda (c_R c_S)^{1/2}$$

$$\zeta = \xi d / 2c_S = \Delta \Theta K d \sin(\phi - \Theta_0) / 2c_S$$

$$= -\Delta \lambda K^2 d / 8\pi n c_S.$$ 

In the case for which there is no slant and the Bragg condition is satisfied, the formula reduces to the well-known equation

$$\eta = \sin^2(\pi n_1 d / \lambda \cos \Theta_0).$$

A graph of this equation as a function of the optical path in units of the free space wavelength $\lambda$ is shown in Fig. 1.2. It is seen that diffraction efficiencies of 1.0 are possible with this type of hologram.

Pure Phase Holograms with Loss. For this case we consider the effect of some loss on a phase hologram. The loss takes the form of some residual absorption, such as might be caused by a chemical stain that has not been washed out or perhaps
some intrinsic absorption at the readout wavelength. For this sort of grating, $a_1$ is zero but $a$ is not. The equation for the diffraction efficiency takes the form
\begin{align}
\eta &= e^{-2ad/cos \Theta} \sin^2(v^2 + \zeta^2)^{1/2}/(1 + \zeta^2/v^2)^{1/2} \\
v &= \pi n_1 d/\lambda \cos \Theta \\
\zeta &= \theta d/2 \cos \Theta = \Delta \Theta \beta d \sin \Theta_0,
\end{align}
(1.52)
where we assume that the grating is unslanted and that the Bragg condition is not satisfied. The absorption evidenced by the exponential term in front and leading to an overall decrease in the efficiency to Bragg mismatch is not much different from that of the lossless hologram. In the case where the Bragg condition is satisfied, the equation simplifies to
\begin{align}
\eta &= e^{-2ad/cos \Theta_0} \sin^2(\pi n_1 d/\lambda \cos \Theta_0)
\end{align}
The effect of the absorption is simply to decrease the overall efficiency by the exponential factor in front. A curve of this function is shown in Fig. 1.3 for several values of the residual density $D_0 = 0.868d/\cos \Theta_0$, which is just the optical density measured in the direction of the illuminating wave.

Amplitude Holograms. For a pure amplitude hologram there is no modulation of the refractive index, so $n_1 = 0$. The solution to the coupled wave equations in this case is
\begin{align}
\eta &= \frac{c_R}{c_S} \exp \left[ -ad \left( \frac{1}{c_R} - \frac{1}{c_S} \right) \right] \sinh^2(v^2 + \zeta^2)^{1/2}/(1 + \zeta^2/v^2) \\
v &= a_1 d/2(c_R c_S)^{1/2} \\
\zeta &= \frac{1}{2} ad \left( \frac{1}{c_R} - \frac{1}{c_S} \right)
\end{align}
Fig. 1.4. Diffraction efficiency as a function of the parameter $a_1 d/2 \cos \theta_0$ for thick amplitude holograms viewed in transmission. The curves are plotted for several values of the reference-to-object beam ratio $K$, which is related to the absorption constants $a$ and $a_1$ through $a/a_1 = 2\sqrt{K/(1+K)}$. The average specular transmittance of the hologram is $\exp(-2ad/\cos \theta_0)$.

For the case where the Bragg condition is fulfilled but the fringe planes are slanted. For the case where the fringes are unslanted but the Bragg condition is not satisfied, the equation for the diffraction efficiency becomes:

$$\eta = e^{-2ad/\cos \theta} \sinh^2(v^2)/(1 - v)$$

$$v = a_1 d/2 \cos \Theta$$

$$= \theta d/2 \cos \Theta \approx A \Theta / d \sin \Theta_0$$

and in case the Bragg condition is satisfied this simplifies to:

$$\eta = e^{-2ad/\cos \theta_0} \sinh^2(a_1 d/2 \cos \Theta_0)$$

which is plotted in Fig. 1.4. The maximum diffraction efficiency is achieved when $a_1 = a$ and $ad/\cos \Theta_0 = \ln 3 = 1.1$, and is 0.037. This is considerably lower than the theoretical maximum for a thin amplitude hologram (0.0652). The maximum is reached when $ad/\cos \Theta_0 = 1.1$, which corresponds to an average density of 0.955 (measured in the direction $\Theta_0$) or an amplitude transmittance of 0.34.

**Mixed Holograms.** A mixed hologram is one for which there is a phase modulation in addition to the amplitude modulation. In this case both $a_1$ and $n_1$ are nonzero. This situation is of some practical importance because any real hologram will be, more or less, a mixed hologram. For example, for a grating produced on a photographic emulsion there is quite likely to be an exposure-dependent variation in the refractive index or a surface relief image in addition to the transmittance variation. This then would result in a mixed hologram. The
relative contribution of the phase and amplitude portions can be predicted from coupled wave theory. Interestingly enough, it turns out that the contributions of the phase and amplitude parts are simply additive, at least for the case of an unslanted grating and Bragg incidence. The equation for the diffraction efficiency is

$$\eta = \left[ \sin^2 \left( \frac{\pi n_1 d}{\lambda \cos \Theta_0} \right) + \sinh^2 \left( \frac{\alpha_1/2 \cos \Theta_0}{\lambda} \right) \right] e^{-2 \alpha_1 \cos \Theta_0}$$  \hspace{1cm} (57)

where all of the symbols have been previously defined.

**Reflection Holograms**

*Pure Phase Holograms.* The general equation for the diffraction efficiency of a pure phase reflection hologram with slanted fringe planes and non-Bragg illumination is

$$\eta = \left[ 1 + \left( 1 - \frac{\zeta^2}{v^2} \right) \sinh^2 \left( v^2 - \zeta^2 \right) \right]^{1/2}$$

$$v = \frac{i\pi n_1 d}{\lambda (c_R c_s)^{1/2}}$$

$$\zeta = -\frac{\varphi d}{2c_s}$$

$$= \Delta \Theta K d \sin (\Theta_0 - \varphi)/2c_s$$

$$= \Delta \lambda K^2 d/8\pi n c_s.$$ 

Note that $v$ is real since $c_s$ is negative. For an unslanted grating and Bragg incidence (1.57) becomes

$$\eta = \tanh^2 \left( \frac{\pi n_1 d}{\lambda \cos \Theta_0} \right).$$

Figure 1.5 shows a plot of this equation. It is seen that it is possible to have a diffraction efficiency of 1.0.

*Amplitude Holograms.* For a reflection hologram of the amplitude type, where $n_1 = 0$ and $a_1$ is nonzero, the diffraction efficiency for unslanted fringes ($\varphi = 0$) and Bragg incidence is given by

$$\eta = \frac{c_R}{c_s} \left[ \frac{v}{v^2 + \left( \frac{\zeta^2}{v^2} - 1 \right)^{1/2}} \right]^{-1} \cos (\zeta^2 - v^2)^{1/2}$$

$$v = \frac{ia_1 d}{2(c_R c_s)^{1/2}}$$

$$\zeta = D_0 - i\zeta_0$$

$$D_0 = \frac{ad}{\cos \Theta_0}$$

$$\zeta_0 = \Delta \Theta \beta d \sin \Theta_0 = \frac{1}{2} \frac{\Delta \lambda}{\lambda} K d.$$  \hspace{1cm} (1.60)
If the Bragg condition is satisfied, then this equation simplifies to

$$\eta = D_1^2 \left\{ 2[D_0^2 + (D_0^2 - D_{1/4}^2)^{1/2} \coth(D_0^2 - D_{1/4}^2)^{1/2}] \right\}^{-2}$$

(1.61)

where $D_1 = a_1 d/\cos \Theta_0$. This equation is plotted in Fig. 1.6. The largest allowable modulation is for $D_1 = D_0$ or equivalently, $a_1 = a$. In this case the maximum diffraction efficiency is 0.072, which is achieved in the limiting case $D_1 = D_0 = \infty$.

**Summary**

Figure 1.7 summarizes the principal results. The numbers shown are the maximum theoretical diffraction efficiency obtainable for each type of hologram. With the exception of the thick, phase reflection hologram, each of the theoretical maxima has been very nearly achieved experimentally. For a reason that is yet unexplained, no holographic material and process have yet come close to the theoretical 1.0 diffraction efficiency for this type of hologram.
THEORETICAL MAXIMUM DIFFRACTION EFFICIENCY

<table>
<thead>
<tr>
<th>Hologram Type:</th>
<th>Thin Transmission</th>
<th>Thick Transmission</th>
<th>Thick Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modulation:</td>
<td>Amplitude</td>
<td>Phase</td>
<td>Amplitude</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.0625</td>
<td>0.339</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Fig. 7. Table of maximum diffraction efficiency for various hologram types.

Clearly the highest efficiencies are achieved for phase-type holograms, and these can be formed quite readily on many types of holographic recording materials. Also, for each of the hologram types considered, it was assumed that some sort of linear relationship existed between input and output. For the case of the thick holograms it was between the amplitude transmittance and the exposure. Unfortunately, it is impossible in any real system to maintain this linearity over the exposure range required to produce maximum efficiency. Thus in most real situations one has to balance an increase in nonlinearity noise against an increase in diffraction efficiency. According to Kogelnik [1.16], the efficiencies that are possible while maintaining the assumed linear response are smaller by about 50% than the corresponding maximum efficiencies, depending on how strictly linearity is defined.

1.5 Noise

The noise associated with a holographic image is any unwanted light that is scattered or diffracted into the direction of the desired image during the reconstruction process. This noise light can take the form of a veiling glare that reduces the contrast of the image, or it can be in the form of an unwanted, spurious image. All of the forms of noise present in holography are signal dependent in that they depend either on the amount or form of the object light; because of the coherent nature of the reconstruction process, the noise light adds coherently to the signal or image light.

The principal sources of noise light are as follows:

1) Granularity. When a photographic emulsion is used as the recording material, the information recorded on the hologram is built up from millions of tiny clumps of silver called grains. These photographic grains appear microscopically as fine black specks on a clear background. When light is transmitted through this pattern it is scattered over a wide range of directions, and this is the light that forms a veiling glare around the desired holographic image. There are some recording materials that are either grainless or essentially grainless, and hence do not scatter light. A thermoplastic material, for example, utilizes the surface relief to diffract the light and the material itself is grainless, and so one would expect no noise of this type. The photochromic materials have grains that are molecular in size. These are so small that the material is essentially grainless and scatters no light.
2) **Scattering from the Support and/or Binder Material.** Scattering can arise even in the grainless materials because of the optical imperfections of the host material. The photochromics are crystalline materials that are subject to imperfections that scatter light. In the case of the thermoplastics it would be the plastic itself plus the photoconductor layer plus the support material that would scatter the light. In the case of the photographic types of materials one often has to consider the scattering from the support material and from the gelatin binder itself.

3) **Phase Noise.** Phase noise is intrinsic to phase holograms, but occurs only in the case of diffuse objects. The self-interference of the object light is recorded as a low-frequency pattern on the hologram. This low-frequency phase pattern diffracts unwanted light into and around the holographic image.

4) **Nonlinearity Noise.** Nonlinearity noise arises in cases where the final amplitude transmittance of the hologram is not strictly proportional to the recording exposure. The presence of recorded information on the hologram that is proportional to the second and higher orders of the exposure gives rise to unwanted light diffracted in the image direction. This type of noise occurs mainly when the exposure modulation is high so that the range of exposures exceeds the dynamic range of the recording material.

5) **Speckle Noise.** This is a type of noise associated with holographic images, but it does not strictly fit the definition given in the first paragraph above. In a strict sense speckle noise is not really noise because it arises precisely because of an accurate reconstruction of the object wave. However, it is noise in the sense that the image is definitely degraded by its presence. It arises whenever a diffuse wavefront of coherent light is limited in extent by the apertures of the system. When this happens the phase undulations of the diffuse wavefront become amplitude variations, which have a granular or salt-and-pepper appearance. However, since this noise is not a function of the recording material, it will be ignored in this volume.

A more detailed discussion of the various types of noise is presented with the discussions of the different recording materials in the following chapters and so we will not elaborate more here.

**References**

1.1 D. Gabor: Nature 161, 777 (1948)
1.4 M. J. Beurger: J. Appl. Phys. 21, 909 (1950)
1.5 W. L. Bragg, G. L. Rogers: Nature 167, 190 (1951)
1.6 H. M. A. El-Sum: *Reconstructed Wavefront Microscopy*, Ph. D. Thesis (Stanford Univ. 1952)
1.7 M. E. Haine, J. Dyson: Nature 166, 315 (1950)
1.10 G. L. Rogers: Nature 166, 237 (1950)