

and photorefractive crystals are now being used, to an increasing extent, for specific applications for which they offer definite advantages. Very interesting possibilities have been opened up in dynamic holography, including the generation of phase-conjugate waves by Brillouin scattering, and holography in resonant media [Denisyuk, 1980].

The striking realism of holographic images has always been something that has fascinated scientists as well as laymen. However, as a result of the developments of the last decade, holography has ceased to be a novelty and has become a well-established technique with many invaluable applications. As always in scientific research, some of these advances were essentially unanticipated; we can fully expect many more such surprises in the next few years.

2

The wavefront reconstruction process

The concepts of holography outlined in Chapter 1 can now be formulated and discussed in more specific terms.

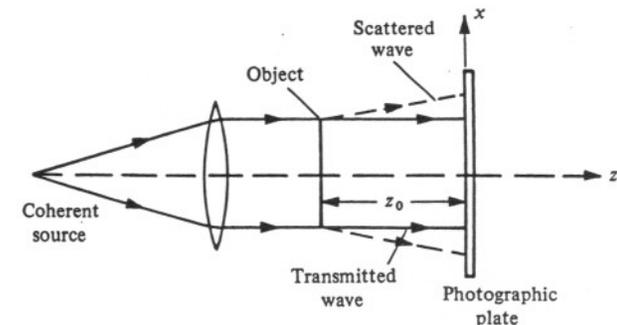
2.1. The in-line (Gabor) hologram

Consider the optical system shown in fig. 2.1, which is essentially that used by Gabor [1948] to demonstrate holographic imaging. In this setup, the coherent light source as well as the object, which is a transparency containing small opaque details on a clear background, are located along the axis normal to the photographic plate.

When the object is illuminated with a uniform parallel beam, the light transmitted by it consists of two parts. The first is a relatively strong, uniform plane wave corresponding to the directly transmitted light. This constitutes the reference wave, and, since its amplitude and phase do not vary across the photographic plate, its complex amplitude can be written as a real constant r . The second is a weak scattered wave due to the transmittance variations in the object. The complex amplitude of this wave, which varies across the photographic plate, can be written as $o(x, y)$, where $|o(x, y)| \ll r$.

The resultant complex amplitude at any point on the photographic plate

Fig. 2.1. Optical system for recording an in-line (Gabor) hologram.



is the sum of these two, so that the intensity at this point (see Appendix A3.1) is

$$I(x, y) = |r + o(x, y)|^2 = r^2 + |o(x, y)|^2 + r o(x, y) + r o^*(x, y), \quad (2.1)$$

where $o^*(x, y)$ is the complex conjugate of $o(x, y)$.

A positive transparency is made from this recording. For simplicity, it is assumed that this has been processed so that its amplitude transmittance t (the ratio of the transmitted amplitude to that incident on it) is a linear function of the intensity and can be written as

$$t = t_0 + \beta T I, \quad (2.2)$$

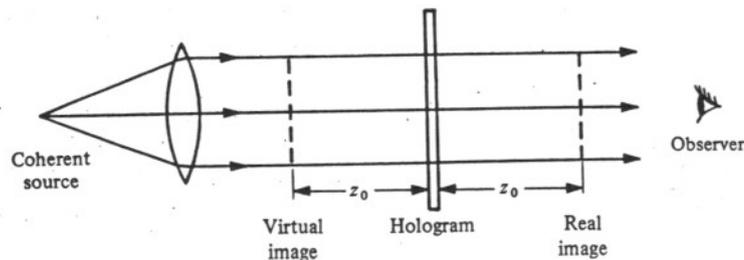
where t_0 is a constant background transmittance, T is the exposure time and β is a parameter determined by the photographic material used and the processing conditions. The amplitude transmittance of the transparency is, accordingly,

$$t(x, y) = t_0 + \beta T [r^2 + |o(x, y)|^2 + r o(x, y) + r o^*(x, y)]. \quad (2.3)$$

To view the reconstructed image, this transparency is replaced in the same position as the original photographic plate and illuminated once again with the same parallel beam of monochromatic light used to record the hologram, as shown in fig. 2.2. Since the complex amplitude at any point in this beam is, apart from a constant factor, the same as that in the reference beam, the complex amplitude transmitted by the hologram can be written as

$$u(x, y) = r t(x, y) = r(t_0 + \beta T r^2) + \beta T r |o(x, y)|^2 + \beta T r^2 o(x, y) + \beta T r^2 o^*(x, y). \quad (2.4)$$

Fig. 2.2. Optical system for reconstructing the image from an in-line (Gabor) hologram, showing the formation of the twin images.



The complex amplitude of the transmitted wave consists, therefore, of four distinct components.

The first of these, $r(t_0 + \beta T r^2)$, is a uniformly attenuated plane wave, the directly transmitted beam.

The second term, $\beta T r |o(x, y)|^2$, is extremely small in comparison to the other terms, since it has been assumed initially that $|o(x, y)| \ll r$. Under these conditions, it can be neglected.

The third term, $\beta T r^2 o(x, y)$, is, except for a constant factor, the same as the original wavefront from the object incident on the photographic plate. Hence, it gives rise to a reconstructed image of the object in its original position. Since this is located behind the transparency at a distance z_0 from it, and the reconstructed wave appears to diverge from it, it is a virtual image.

In the same manner, the fourth term corresponds to a wavefront which resembles the wavefront diverging from the object, except that it is of opposite curvature. Hence, it converges to form a real image, the conjugate image, at the same distance z_0 in front of the hologram.

It is apparent that, with such a system, an observer focusing on one image sees it superposed on the out-of-focus twin image as well as a strong coherent background. This constitutes its most serious limitation.

Another major limitation is the need for the object to have a high average transmittance, if the second term on the right hand side of (2.4), which has been taken as negligible, is not to interfere with the reconstructed image. Typically, it is possible to form images of fine opaque lines on a transparent background, but not *vice versa*.

Finally, it should be noted that the hologram used to reconstruct the image must be a positive transparency. Since the image-forming waves interfere with the background in the process of reconstruction, if the plate which has been exposed in the recording step is used directly (in which case β in (2.2) is negative), the reconstructed image will also be a negative.

2.2. The off-axis (Leith-Upatnieks) hologram

The first successful technique for separating the twin images was developed by Leith & Upatnieks [1962, 1963, 1964]. As shown in fig. 2.3, a separate reference beam derived from the same coherent source is allowed to fall on the photographic plate, during the recording process, at an offset angle θ to the beam from the object. For simplicity, this reference beam can be assumed to be a collimated beam of uniform intensity.

The complex amplitude due to the object beam at any point (x, y) on the photographic plate can then be written as

$$o(x, y) = |o(x, y)| \exp[-i\phi(x, y)], \quad (2.5)$$

while that due to the reference beam is

$$r(x, y) = r \exp(i2\pi\xi_r x), \quad (2.6)$$

where $\xi_r = (\sin \theta)/\lambda$, since the reference beam has uniform intensity and only its phase varies across the photographic plate.

The resultant intensity at the photographic plate is

$$\begin{aligned} I(x, y) &= |r(x, y) + o(x, y)|^2 \\ &= |r(x, y)|^2 + |o(x, y)|^2 \\ &\quad + r|o(x, y)| \exp[-i\phi(x, y)] \exp(-i2\pi\xi_r x) \\ &\quad + r|o(x, y)| \exp[i\phi(x, y)] \exp(i2\pi\xi_r x), \\ &= r^2 + |o(x, y)|^2 + 2r|o(x, y)| \cos[2\pi\xi_r x + \phi(x, y)]. \end{aligned} \quad (2.7)$$

As can be seen from (2.7), the amplitude and phase of the object wave are encoded, respectively, as amplitude and phase modulation of a set of interference fringes equivalent to a spatial carrier with a spatial frequency equal to ξ_r .

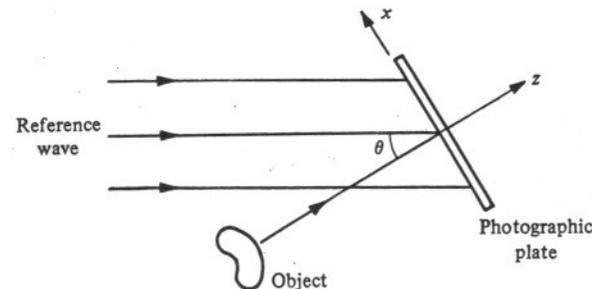
If, as in (2.2), it is assumed that the amplitude transmittance of the photographic plate after processing is linearly related to the intensity in the interference pattern, the amplitude transmittance of the hologram can be written as

$$\begin{aligned} t(x, y) &= t_0 + \beta T \{ |o(x, y)|^2 \\ &\quad + r|o(x, y)| \exp[-i\phi(x, y)] \exp(-i2\pi\xi_r x) \\ &\quad + r|o(x, y)| \exp[i\phi(x, y)] \exp(i2\pi\xi_r x) \}, \end{aligned} \quad (2.8)$$

where β is the slope (in this case, negative) of the amplitude transmittance versus exposure characteristic of the photographic material, T is the exposure time and t_0 is a constant background transmittance.

To reconstruct the image, the hologram is illuminated once again, as shown in fig. 2.4 with the same reference beam as was used to record it. The

Fig. 2.3. Hologram recording with an off-axis reference beam.



complex amplitude $u(x, y)$ of the transmitted wave is, in this case also, the sum of four terms, each corresponding to one of the terms of (2.8), and can be written as

$$\begin{aligned} u(x, y) &= r(x, y)t(x, y), \\ &= u_1(x, y) + u_2(x, y) + u_3(x, y) + u_4(x, y), \end{aligned} \quad (2.9)$$

where

$$u_1(x, y) = t_0 r \exp(i2\pi\xi_r x), \quad (2.10)$$

$$u_2(x, y) = \beta T r |o(x, y)|^2 \exp(i2\pi\xi_r x), \quad (2.11)$$

$$u_3(x, y) = \beta T r^2 o(x, y), \quad (2.12)$$

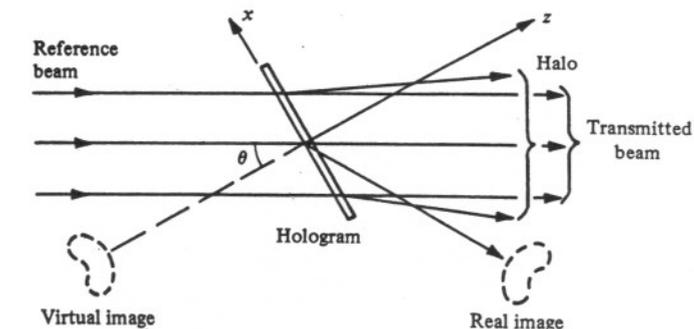
$$u_4(x, y) = \beta T r^2 o^*(x, y) \exp(i4\pi\xi_r x). \quad (2.13)$$

The first term on the right hand side of (2.9), $u_1(x, y)$, is, as before, merely the attenuated reference beam, which is a plane wave directly transmitted through the hologram. This directly transmitted beam is surrounded by a halo due to the second term, $u_2(x, y)$, which is spatially varying. The angular spread of this halo is determined by the angular extent of the object.

The third term, $u_3(x, y)$ is identical with the original object wave, except for a constant factor, and generates a virtual image of the object in its original position; this wave makes an angle θ with the directly transmitted wave. Similarly, the fourth term, $u_4(x, y)$, gives rise to the conjugate real image. However, in this case, the fourth term includes an exponential factor, $\exp(i4\pi\xi_r x)$, which indicates that the conjugate wave is deflected off the axis at an angle approximately twice that which the reference wave makes with it.

Thus, even though two images – one real and one virtual – are still reconstructed in this setup, they are angularly separated from the directly

Fig. 2.4. Image reconstruction by a hologram recorded with an off-axis reference beam.



transmitted beam and from each other, and the offset angle θ of the reference beam is made large enough, it is possible to ensure that there is no overlap. This method therefore eliminates at one stroke all the major drawbacks of Gabor's original in-line arrangement. In addition, it is also interesting that the sign of β in this case only affects the phase of the reconstructed image – a 'positive' image is obtained even if the hologram recording is a photographic negative.

The minimum value of the offset angle θ required to ensure that each of the images can be observed without any interference from its twin, as well as from the directly transmitted beam and the halo of scattered light surrounding it, is determined by the minimum spatial carrier frequency (ξ_r) for which there is no overlap between the angular spectra of the third and fourth terms and those of the first and second terms.

These angular spectra are the Fourier transforms (see Appendix A1) of these terms and can be written as follows.

$$U_1(\xi, \eta) = \mathcal{F}\{t_0 r \exp(i2\pi\xi_r x)\}, \\ = t_0 r \delta(\xi + \xi_r, \eta). \quad (2.14)$$

$$U_2(\xi, \eta) = \mathcal{F}\{\beta T r |o(x, y)|^2 \exp(i2\pi\xi_r x)\}, \\ = \beta T r [O(\xi, \eta) \star O(\xi, \eta) \ast \delta(\xi + \xi_r, \eta)], \quad (2.15)$$

where $O(\xi, \eta) = \mathcal{F}\{o(x, y)\}$ is the spatial frequency spectrum of the object beam, and the symbols \star and \ast denote, respectively, the operations of correlation and convolution.

$$U_3(\xi, \eta) = \mathcal{F}\{\beta T r^2 o(x, y)\}, \\ = \beta T r^2 O(\xi, \eta). \quad (2.16)$$

$$U_4(\xi, \eta) = \mathcal{F}\{\beta T r^2 o^*(x, y) \exp(i4\pi\xi_r x)\}, \\ = \beta T r^2 O^*(\xi, \eta) \ast \delta(\xi + 2\xi_r, \eta). \quad (2.17)$$

As can be seen from fig. 2.5, which shows these spectra schematically, the term $|U_3|$ is merely the object-beam spectrum multiplied by a constant and is centred at the origin of the spatial frequency plane. The term $|U_1|$ corresponds to the spatial frequency of the carrier fringes and is a δ function located at $(-\xi_r, 0)$, while $|U_2|$ is centred on this δ function and, being proportional to the auto-correlation function of $O(\xi, \eta)$, has twice the extent of the object-beam spectrum. Finally, $|U_4|$ is similar to $|U_3|$ but is displaced to a centre frequency $(-2\xi_r, 0)$.

Evidently, $|U_3|$ and $|U_4|$ will not overlap $|U_1|$ and $|U_2|$ if the spatial carrier frequency ξ_r is chosen so that

$$\xi_r \geq 3\xi_m, \quad (2.18)$$

where ξ_m is the highest frequency in the spatial frequency spectrum of the object beam.

2.3. Fourier Holograms

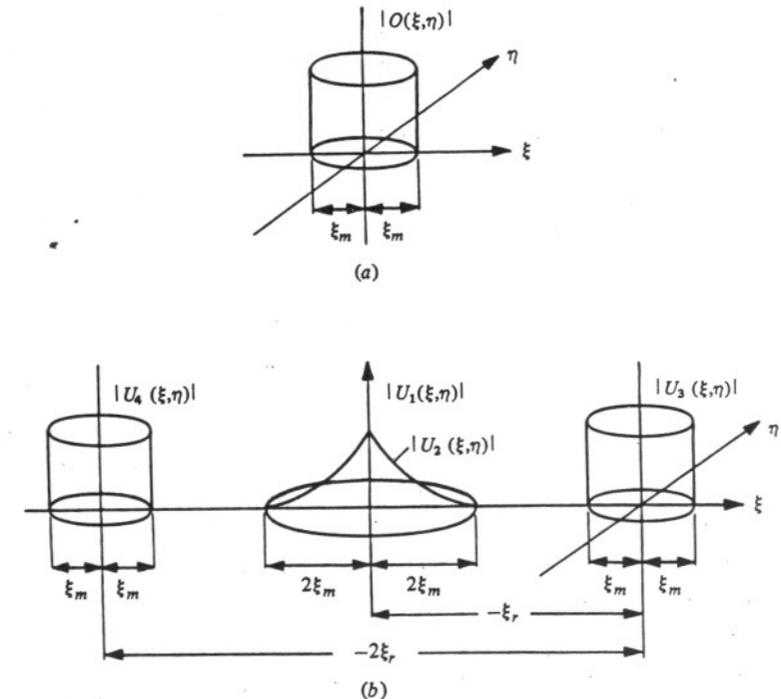
Another interesting category of hologram recording configurations is one in which the waves that interfere at the hologram plane are the Fourier transforms of the object and reference waves. Normally, this implies an object which lies in a single plane, or is of limited thickness.

A typical optical arrangement for recording such a hologram (Vander Lugt, 1964) is shown in fig. 2.6. The object, a transparency located in the front focal plane of a lens, is illuminated by a parallel beam of monochromatic light. If the complex amplitude leaving the object plane is $o(x, y)$, the complex amplitude of the object wave at the photographic plate located in the back focal plane of the lens is

$$O(\xi, \eta) = \mathcal{F}\{o(x, y)\}. \quad (2.19)$$

The reference beam is derived from a point source also located in the front focal plane of the lens. If $\delta(x + b, y)$ is the complex amplitude of this point

Fig. 2.5. Spatial frequency spectra of (a) the object beam and (b) a hologram recorded with an off-axis reference beam.



source, the complex amplitude of the reference wave at the hologram plane is

$$R(\xi, \eta) = \exp(-i2\pi\xi b). \quad (2.20)$$

The intensity in the interference pattern formed by these two waves is therefore

$$I(\xi, \eta) = 1 + |O(\xi, \eta)|^2 + O(\xi, \eta) \exp(i2\pi\xi b) + O^*(\xi, \eta) \exp(-i2\pi\xi b). \quad (2.21)$$

To reconstruct the image, the processed hologram is placed in the front focal plane of the lens and illuminated with a parallel beam of monochromatic light of unit amplitude as shown in fig. 2.7. If it is assumed, as before, that the amplitude transmittance of the processed hologram is a linear function of $I(\xi, \eta)$, the complex amplitude of the light transmitted by the hologram is

$$U(\xi, \eta) = t_0 + \beta TI(\xi, \eta). \quad (2.22)$$

The complex amplitude in the back focal plane of the lens is then the

Fig. 2.6. Optical system for recording a Fourier hologram.

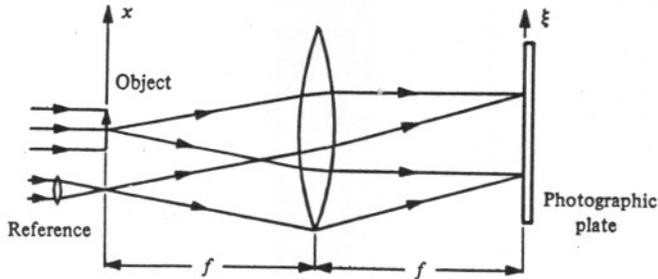
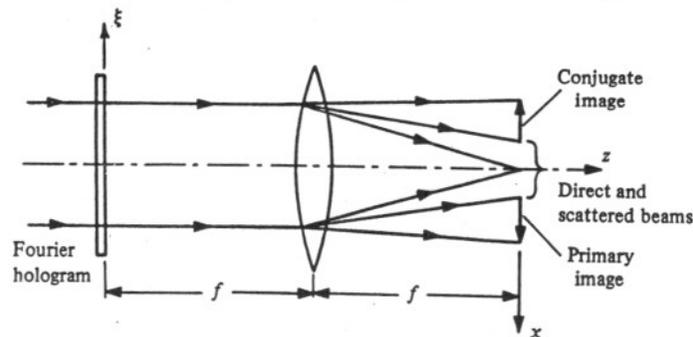


Fig. 2.7. Reconstruction setup for a Fourier hologram.



Fourier transform of $U(\xi, \eta)$,

$$u(x, y) = \mathcal{F}\{U(\xi, \eta)\}, \\ = (t_0 + \beta T)\delta(x, y) + \beta T o(x, y) \star o(x, y) + \beta T o(x - b, y) + \beta T o^*(-x + b, -y). \quad (2.23)$$

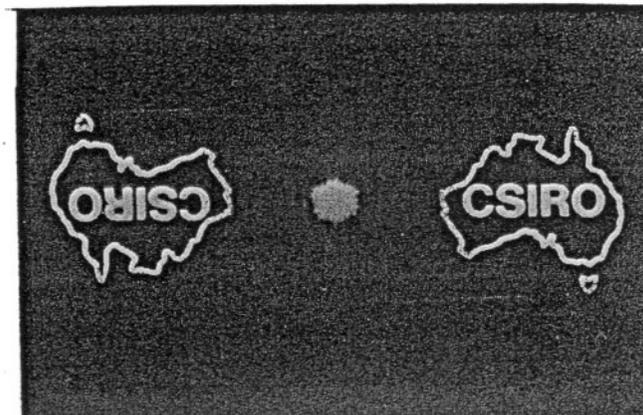
As shown in fig. 2.7, the first term on the right-hand side of (2.23) comes to a focus on the axis, while the second term forms a halo around it. The third term corresponds to the original object wave shifted downwards by a distance b , while the fourth term is the conjugate of the original object wave inverted and shifted upwards by the same amount b . Both the images are real and can be recorded on a photographic film placed in the back focal plane of the lens. Since the film records only the intensity distribution in the image, it is possible, in this case, to identify the conjugate image only by the fact that it is inverted (see fig. 2.8).

Fourier holograms have the useful property that the reconstructed image is stationary even when the hologram is translated in its own plane. This is because a shift of a function in the spatial domain only results in its Fourier transform being multiplied by a phase factor which is a linear function of the spatial frequency (see Appendix A1). This phase factor has no effect on the intensity distribution in the image.

2.4 The lensless Fourier hologram

A hologram with the same properties as a Fourier hologram can be obtained even without a lens to produce the Fourier transform of the object wave, provided the reference wave is produced by a point source in the plane of the object [Stroke, 1965; Stroke, Brumm & Funkhouser, 1965].

Fig. 2.8. Twin images reconstructed by a Fourier hologram.



means of a small moving pupil [Dainty & Weir, 1971]. As the pupil uncovers different parts of the hologram, the speckle pattern in the image plane changes so that the exposure corresponds to the sum of the irradiance distributions in a number of uncorrelated speckle patterns. This method was extended by Yu & Wang [1973], who proposed the use of a moving mask with a random distribution of openings covering the entire hologram, and by Som & Budhiraja [1975], who claimed that a further reduction in speckle could be obtained by the use of a moving mask at some distance from the hologram, both in recording and reconstruction. However, it can be shown that while there is a reduction of speckle, it is always accompanied by a reduction in either the resolution or the contrast of the image [Hariharan & Hegedus, 1974a, b; McKechnie, 1975a, b; Östlund & Biedermann, 1977].

4

Types of holograms

A hologram recorded on a photographic plate and processed normally is equivalent to a grating with a spatially varying transmittance. However, with suitable processing, it is possible to produce a spatially varying phase shift. In addition, if the thickness of the recording medium is large compared with the fringe spacing, volume effects are important. In an extreme case, it is even possible to produce holograms in which the fringes are planes running almost parallel to the surface of the recording material and which reconstruct an image in reflected light.

Based on these characteristics, holograms recorded in a thin recording medium can be divided into amplitude holograms and phase holograms. Holograms recorded in relatively thick recording media can be classified either as transmission amplitude holograms, transmission phase holograms, reflection amplitude holograms or reflection phase holograms.

In the next few sections we shall examine some of the principal characteristics of these six types of holograms.

4.1. Thin holograms

Any hologram in which the thickness of the recording material is small compared with the average spacing of the interference fringes can be classified as a thin hologram. Such a hologram can be characterized by a spatially varying complex amplitude transmittance

$$t(x, y) = |t(x, y)| \exp[-i\phi(x, y)]. \quad (4.1)$$

4.1.1. Thin amplitude holograms

In an amplitude hologram, $\phi(x, y)$ is essentially constant while $|t(x, y)|$ varies over the hologram. To calculate the complex amplitude of the diffracted waves from such a hologram and, hence, its diffraction efficiency, consider a grating formed in a suitable thin recording medium by a plane object wave and a plane reference wave.

If we assume that the resulting amplitude transmittance is linearly related to the intensity in the interference pattern, the amplitude transmittance of

the grating can be written as

$$|t(x)| = t_0 + t_1 \cos Kx. \quad (4.2)$$

where t_0 is the average amplitude transmittance of the grating, t_1 is the amplitude of the spatial variation of $|t(x)|$ and

$$K = 2\pi/\Lambda, \quad (4.3)$$

where Λ is the spacing of the fringes.

Since the values of $|t(x)|$ are limited to the range $0 \leq |t(x)| \leq 1$, and the amplitudes of the diffracted waves are linearly proportional to the amplitude of the spatial variation of $|t(x)|$, the diffracted amplitude is a maximum when

$$\begin{aligned} |t(x)| &= (1/2) + (1/2) \cos Kx, \\ &= (1/2) + (1/4) \exp(iKx) + (1/4) \exp(-iKx). \end{aligned} \quad (4.4)$$

The maximum amplitude in each of the diffracted orders is one fourth of that in the wave used to illuminate the hologram, so that the peak diffraction efficiency is

$$\varepsilon_{\max} = (1/16), \quad (4.5)$$

or 0.0625.

In practice, no recording medium has a linear response over the full range of transmittance values from 0 to 1; hence this value of ε cannot be achieved without running into nonlinear effects.

4.1.2. Thin phase holograms

For a lossless phase grating, $|t(x)| = 1$, so that the complex amplitude transmittance is

$$t(x) = \exp[-i\phi(x)]. \quad (4.6)$$

If the phase shift produced by the recording medium is linearly proportional to the intensity in the interference pattern,

$$\phi(x) = \phi_0 + \phi_1 \cos(Kx), \quad (4.7)$$

and the complex amplitude transmittance of the grating is

$$t(x) = \exp(-i\phi_0) \exp[-i\phi_1 \cos(Kx)]. \quad (4.8)$$

If we neglect the constant phase factor $\exp(-i\phi_0)$, the right hand side of (4.8) can be expanded as a Fourier series to give

$$t(x) = \sum_{n=-\infty}^{\infty} i^n J_n(\phi_1) \exp(inKx), \quad (4.9)$$

where J_n is the Bessel function of the first kind of order n .

Such a thin phase grating diffracts a wave incident on it into a large number of orders, the diffracted amplitude in the n th order being proportional to the value of the Bessel function $J_n(\phi_1)$. Only the wave of order 1 contributes to the desired image. As shown in fig. 4.1, the amplitude diffracted into this order, which is proportional to $J_1(\phi_1)$, increases at first with the phase modulation and then decreases.

The diffraction efficiency of the grating is, accordingly,

$$\varepsilon = J_1^2(\phi_1), \quad (4.10)$$

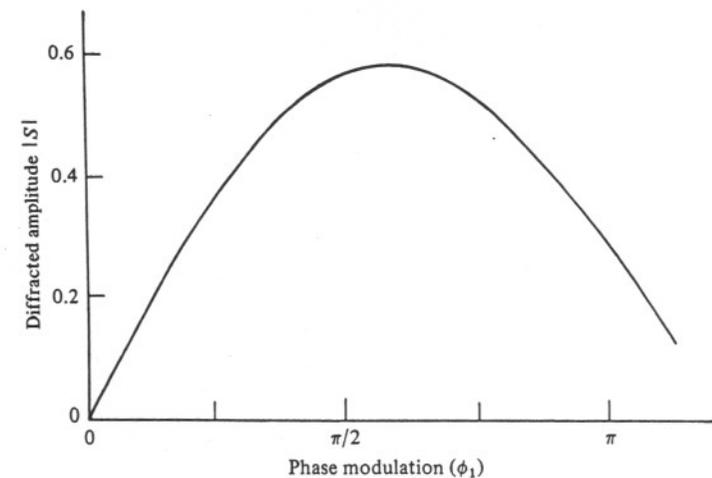
and its maximum value is

$$\varepsilon_{\max} = 0.339. \quad (4.11)$$

4.2. Volume holograms

The medium in which a hologram is recorded can have a thickness of as much as a few millimetres, while the fringe spacing may only be of the order of $1 \mu\text{m}$. The hologram is then a three-dimensional system of layers corresponding to a periodic variation of absorption or refractive index, and the diffracted amplitude is a maximum only when the Bragg condition is satisfied. The characteristics of generalized volume gratings have been discussed in detail by Russell [1981] and by Solymar & Cooke [1981]. However, for simplicity, we will only consider a grating produced by recording the interference of two infinite plane wavefronts in a thick

Fig. 4.1. Diffraction efficiency of a thin phase grating as a function of the phase modulation [Kogelnik, 1967].



recording medium. We will also assume that, initially, the recording medium is perfectly transparent but, after processing, develops a sinusoidal variation of the absorption or the refractive index in the direction perpendicular to the interference surfaces. In addition, while the interference surfaces can assume any orientation, only two limiting cases will be considered, in which they are either perpendicular or parallel to the hologram plane.

The first case arises when the two interfering wavefronts make equal but opposite angles to the surface of the recording medium and are incident on it from the same side. Holograms recorded in this fashion give a reconstructed image in transmitted light. The second case arises when the wavefronts are symmetrical with respect to the surface of the recording medium but are incident on it from opposite sides. Holograms of this type give a reconstructed image by reflection.

The spacing between successive fringe planes is a minimum, and the volume effects are most pronounced, when the angle between the two interfering wavefronts is a maximum ($\approx 180^\circ$). This has made it possible to produce reflection holograms with a wavelength selectivity high enough to reconstruct an image of acceptable quality when illuminated with white light.

4.3. The coupled wave theory

When analysing the diffraction of light by such thick gratings, it is necessary to take into account the fact that the amplitude of the diffracted wave increases progressively, while that of the incident wave decreases, as they propagate through the grating. One way of doing this is by means of a coupled wave approach such as that developed by Kogelnik [1967, 1969].

Consider a coordinate system in which, as shown in fig. 4.2, the z -axis is perpendicular to the surfaces of the recording medium and the x -axis is in the plane of incidence, while the fringe planes are oriented perpendicular to the plane of incidence. The grating vector \mathbf{K} is perpendicular to the fringe planes. It is of length $|\mathbf{K}| = 2\pi/\Lambda$, where Λ is the grating period, and makes an angle ψ ($\psi = 90^\circ$ or 0° in the cases shown) with the z axis. The refractive index n and the absorption constant α are assumed to vary sinusoidally, their values at any point being given by the relations

$$n = n_0 + n_1 \cos \mathbf{K} \cdot \mathbf{x}, \tag{4.12}$$

$$\alpha = \alpha_0 + \alpha_1 \cos \mathbf{K} \cdot \mathbf{x}, \tag{4.13}$$

where the radius vector $\mathbf{x} = (x, y, z)$. For simplicity, the refractive index of the surrounding medium is also assumed to be n_0 .

If monochromatic light is incident on the hologram grating at, or near, the Bragg angle, and if the thickness of the medium is large enough, only two waves in the grating need be taken into consideration; these are the incoming reference wave R and the outgoing signal wave S . Since the other diffraction orders violate the Bragg condition strongly, they are severely attenuated and can be neglected. If we also assume that these waves are polarized with their electric vector perpendicular to the plane of incidence, their interaction in the grating can be described by the scalar wave equation

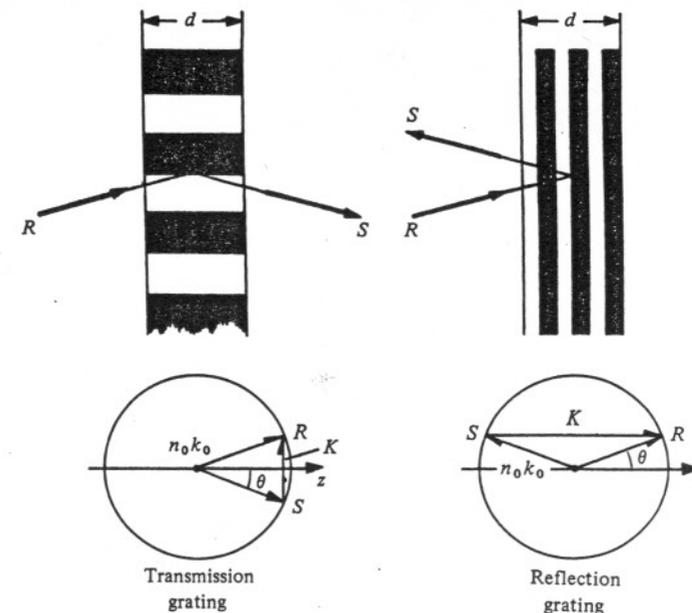
$$\nabla^2 E + k^2 E = 0, \tag{4.14}$$

where E is the total electric field and k is the (spatially varying) propagation constant in the grating.

We assume that the absorption per wavelength as well as the relative variations in refractive index of the medium are small, so that

$$\begin{aligned} n_0 k_0 &\gg \alpha_0, \\ n_0 k_0 &\gg \alpha_1, \\ n_0 &\gg n_1, \end{aligned} \tag{4.15}$$

Fig. 4.2. Volume transmission and reflection gratings and their associated vector diagrams for Bragg incidence [Kogelnik, 1967].



where $k_0 = 2\pi/\lambda$. The propagation constant can then be written in the form

$$k^2 = B^2 - 2i\alpha_0 B + 4\kappa B \cos \mathbf{K} \cdot \mathbf{x}, \quad (4.16)$$

where $B = n_0 k_0$ is the average propagation constant, and κ is the coupling constant defined as

$$\kappa = (\pi n_1 / \lambda) - i\alpha_1 / 2. \quad (4.17)$$

This coupling constant describes the interaction between the reference wave R and the signal wave S . If $\kappa = 0$, there is no modulation of the refractive index or the absorption and hence, no diffraction.

The propagation of the two coupled waves through the grating can be described by their complex amplitudes $R(z)$ and $S(z)$, which vary along z as a result of the energy interchange between them as well as energy losses due to absorption. The total electric field E in the grating is then the superposition of the fields due to these two waves so that

$$E = R(z) \exp(-i\rho \cdot \mathbf{x}) + S(z) \exp(-i\sigma \cdot \mathbf{x}), \quad (4.18)$$

where ρ and σ are the propagation vectors for the two waves; these are defined by the propagation constants and the directions of propagation of R and S . The quantity ρ is assumed to be equal to the propagation vector of the free reference wave in the absence of coupling, while σ is determined by the grating and is related to ρ and the grating vector by the expression

$$\sigma = \rho - \mathbf{K}. \quad (4.19)$$

For the special case of incidence at the Bragg angle θ_0 , the lengths of both ρ and σ are equal to the free propagation constant $n_0 k_0$, and the Bragg condition, which can be written in the form

$$\cos(\psi - \theta_0) = K / 2n_0 k_0, \quad (4.20)$$

is obeyed. If (4.20) is differentiated, we obtain

$$(d\theta_0 / d\lambda_0) = K / 4\pi n_0 \sin(\psi - \theta_0). \quad (4.21)$$

It follows from (4.21) that small changes in the angle of incidence or the wavelength have similar effects.

A useful parameter for evaluating the effects of deviations from the Bragg condition is the dephasing measure ζ , which can be defined as

$$\begin{aligned} \zeta &= (|\rho|^2 - |\sigma|^2) / 2|\rho|, \\ &= (B^2 - |\sigma|^2) / 2B, \\ &= K \cos(\psi - \theta) - K^2 \lambda / 4\pi n_0, \end{aligned} \quad (4.22)$$

from (4.19). For small deviations $\Delta\theta$ and $\Delta\lambda$ from the Bragg condition, this becomes

$$\zeta = \Delta\theta \cdot K \sin(\psi - \theta_0) - \Delta\lambda \cdot K^2 / 4\pi n_0. \quad (4.23)$$

To derive the coupled wave equations, (4.14) and (4.16) are combined, and (4.18) and (4.19) are inserted. If then, the terms involving $\exp(-i\rho \cdot \mathbf{x})$ and $\exp(-i\sigma \cdot \mathbf{x})$ are compared, we get

$$R'' - 2iR'\rho_z - 2i\alpha BR + 2\kappa BS = 0, \quad (4.24)$$

and

$$S'' - 2iS'\sigma_z - 2i\alpha BS + (B^2 - |\sigma|^2)S + 2\kappa BR = 0, \quad (4.25)$$

where the primes denote differentiation with respect to z . If, in addition, it is assumed that the energy interchange between S and R , as well as the energy absorption in the medium are slow, the second differentials R'' and S'' can be neglected. From (4.23) these equations can then be rewritten in the form

$$R' \cos \theta + \alpha R = i\kappa S, \quad (4.26)$$

$$[\cos \theta - (K/B) \cos \psi] S' + (\alpha + i\zeta) S = -i\kappa R. \quad (4.27)$$

The coupled wave equations (4.26) and (4.27) show that the amplitude of a wave changes along z because of coupling to the other wave (κR , κS) or absorption (αR , αS). For deviations from the Bragg condition, S is forced out of synchronism with R , due to the term involving ζS , and the interaction decreases.

The coupled wave equations, (4.26) and (4.27), can be solved for the appropriate boundary conditions. These are $R(0) = 1$, $S(0) = 1$, for transmission gratings, and $R(0) = 1$, $S(d) = 0$, for reflection gratings.

In the next few sections we will discuss the solutions for the most important cases, namely lossless phase gratings and pure absorption gratings, the grating planes being assumed to run either normal to the surface (for transmission gratings) or parallel to the surface (for reflection gratings). The method of solution of the coupled wave equations, as well as solutions for the cases of slanted gratings, lossy phase gratings and mixed gratings are to be found in the original paper by Kogelnik [1969]. This paper also gives an extension of the theory to light polarized with the electric vector in the plane of incidence.

4.4. Volume transmission holograms

4.4.1. Phase gratings

In a lossless phase grating $\alpha_0 = \alpha_1 = 0$. Diffraction is caused by the spatial variation of the refractive index. The diffracted amplitude is then

$$S(d) = \frac{-i \exp(-i\chi) \sin(\Phi^2 + \chi^2)^{1/2}}{(1 + \chi^2 / \Phi^2)^{1/2}}, \quad (4.28)$$

4.11. Imaging properties of volume holograms

In the case of volume holograms, which diffract strongly at the Bragg angle, the amplitude of the reconstructed wavefront is affected by any changes of wavelength or geometry between recording and reconstruction. In addition, changes in the thickness of a photographic emulsion due to processing can result in a rotation of the fringe planes as well as a change in their spacing [Vilkomerson & Bostwick, 1967] (see fig. 4.11).

With gratings recorded with plane wavefronts it is possible to compensate for these by changing either the angle of illumination or the wavelength [Belvaux, 1975]. However, complete compensation is not possible with a hologram of a point at a finite distance, or an extended object, and this results in variations of amplitude across the reconstructed wavefront and reduced diffraction efficiency. The effects of emulsion shrinkage in volume holograms have been considered by Latta [1971c] and by Forshaw [1973], who have shown that it is possible, in this case also, to define a pupil function; this differs from the pupil function for a thin hologram defined in section 3.4, in that it involves spatial modulation of the amplitude as well as the phase.

5

Optical systems and light sources

A typical optical system for recording transmission holograms of a diffusely reflecting object is shown in fig. 5.1, while one for recording a reflection hologram is shown in fig. 5.2.

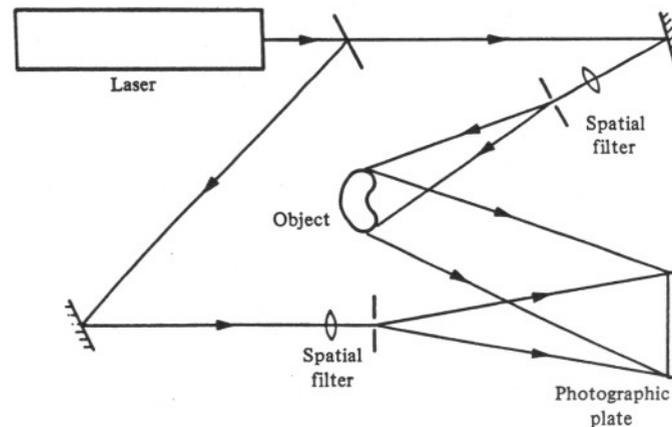
A simpler arrangement for making reflection holograms is shown in fig. 5.3. This is essentially the same as that described originally by Denisjuk [1965] in which, instead of using separate object and reference beams, the portion of the reference beam transmitted by the photographic plate is used to illuminate the object. It gives good results with specular reflecting objects and with a recording medium such as dichromated gelatin, which scatters very little light.

Making a hologram involves recording a two-beam interference pattern. The principal factors which must be taken into account in a practical setup to obtain good results are discussed in the next few sections.

5.1. Vibration isolation

Any change in the phase difference between the two beams during the exposure will result in a movement of the fringes and reduced modulation in the hologram [Neumann, 1968].

Fig. 5.1. Optical system for recording a transmission hologram.



In some situations, the effects of object movement can be minimized by means of an optical system in which the reference beam is reflected from a mirror mounted on the object [Mottier, 1969]. Alternatively, if the consequent loss in resolution can be tolerated, a portion of the laser beam can be focused to a spot on the object, producing a diffuse reference beam [Waters, 1972]. Stability requirements for reflection holograms can be minimized with a setup similar to that shown in fig. 5.3 in which the surface of the object is painted with a retroreflective paint and the hologram plate is rigidly attached to it ('piggyback' holography) [Neumann & Penn, 1972].

Fig. 5.2. Typical optical arrangement for recording a reflection hologram.

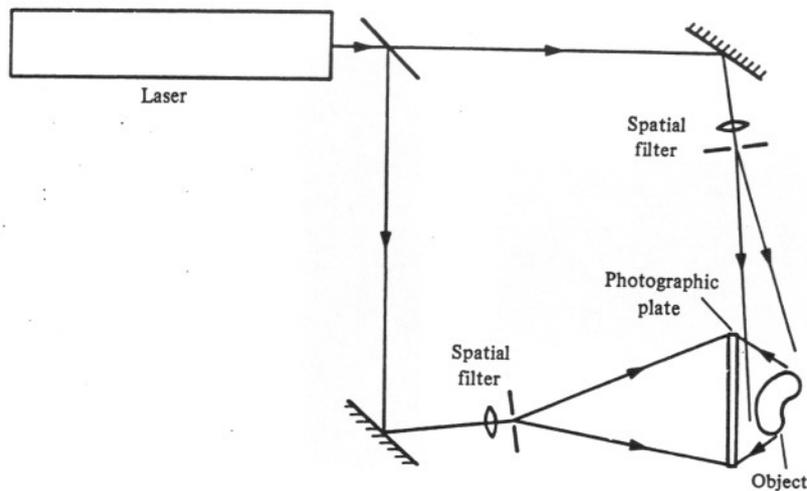
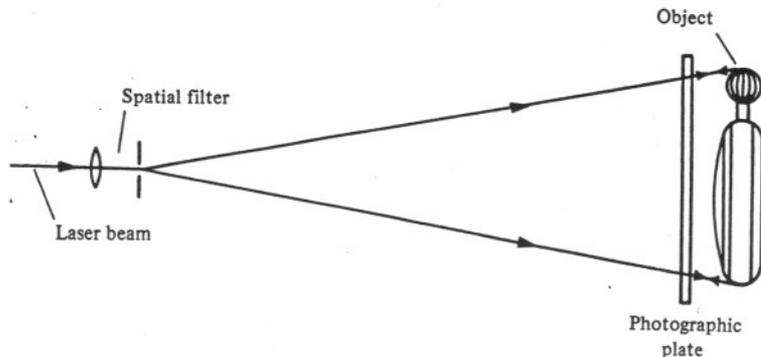


Fig. 5.3. Simple setup for making reflection holograms.



Relative motion is virtually eliminated since the plate moves with the object.

Most commonly, to avoid mechanical disturbances, all the optical components as well as the object and the recording medium are mounted on a stable surface. Acceptable results can be obtained with a concrete slab resting on inflated scooter inner tubes, but most laboratories now use a rigid optical table supported by a pneumatic suspension system, so that it has a low natural frequency of vibration (≈ 1 Hz). This has the advantage that components can be mounted on magnetic bases or bolted down to its surface.

Air currents, acoustic waves and temperature changes also cause major problems. Their effects are usually minimized by enclosing the working area.

Residual disturbances can be eliminated almost completely by a feedback system in which any motion of the interference fringes in the hologram plane is picked up by a photodetector [Neumann & Rose, 1967; MacQuigg, 1977]. Variations in its output are amplified and applied to a piezoelectric element which controls the position of a mirror in the reference-beam path to restore the path difference between the two beams to its original value.

5.2. Fringe visibility

To produce a hologram that reconstructs a bright image, the interference pattern formed at the recording medium by the object and reference waves should have as high a contrast as possible. This is because the amplitude of the diffracted wave increases with the modulation depth of the interference pattern that is recorded.

The contrast of the interference pattern at any point in the hologram plane is measured by the fringe visibility (see Appendix A3.1) which is given by the relation

$$\mathcal{V} = (I_{\max} - I_{\min}) / (I_{\max} + I_{\min}), \quad (5.1)$$

where I_{\max} and I_{\min} are the local maximum and minimum values of the intensity.

5.3. Beam polarization

Most gas lasers used for holography have Brewster angle windows on the plasma tube so that the output beam is linearly polarized. Maximum visibility of the fringes will then be obtained if the angle ψ between the electric vectors in the two interfering beams is zero (see Appendix A3.1).

This condition is automatically satisfied, irrespective of the angle between the two beams, if they are both polarized with the electric vector normal to the plane of the optical table. On the other hand, if they are polarized with the electric vector parallel to the surface of the table, the angle ψ between the electric vectors is equal to the angle θ between the two beams, and, in the extreme case where the two beams intersect at right angles, the visibility of the fringes drops to zero.

With a diffusely reflecting or metallic object it is also necessary to take into account the polarization changes in the light scattered by it [Rogers, 1966; Ghandeharian & Boerner, 1978] which can result in a significant decrease in the visibility of the fringes. This can be minimized either by rotating the polarization of the reference beam so that the cross-polarized light can also interfere with it, or by using a sheet polarizer in front of the hologram plate to eliminate the cross-polarized component. Another alternative is to use a circularly polarized reference beam [Vanin, 1979].

5.4. Beam splitters

If we assume that the interfering beams are polarized with the electric vector perpendicular to the plane of incidence, the fringe visibility is given by the relation

$$\mathcal{V} = 2|\gamma_{12}(\tau)|or/(o^2 + r^2), \quad (5.2)$$

where r and o are the amplitudes of the reference and object beams, and $\gamma_{12}(\tau)$ is the degree of coherence between them. If the beam ratio, $R = (r/o)^2$, is defined as the ratio of the irradiances of the reference and object beams, (5.2) can be rewritten as

$$\mathcal{V} = 2|\gamma_{12}(\tau)|R^{1/2}/(1 + R). \quad (5.3)$$

The fringe visibility is obviously a maximum when $R = 1$. However, the wave scattered by a diffusely reflecting object exhibits quite strong variations in amplitude (see Appendix A4.1). Hence, in hologram recording it is usually necessary to work with a value of $R \gg 1$ to avoid nonlinear effects.

To optimize the visibility of the fringes at the hologram plane it should be possible to vary the ratio of the power in the beam illuminating the object to that in the reference beam.

A convenient way to do this is to use a beam splitter consisting of a disc coated with a thin aluminium film whose reflectivity is a linear function of the azimuth. Such a beam splitter must be used in the unexpanded laser beam because of the gradient of reflectivity across a larger beam. Since an

aluminium film typically absorbs about 30 per cent of the energy incident on it, this limits its use to moderate laser powers (< 500 mW).

A much better variable-ratio beam splitter, with a uniform field and very low insertion loss, uses a polarizing prism to divide the incident beam into two orthogonally polarized components [Caulfield & Beyen, 1967]. A typical setup for this is shown in fig. 5.4. The ratio of the transmitted and reflected powers is given by the relation

$$I_{\text{trans}}/I_{\text{refl}} = \cot^2 \psi, \quad (5.4)$$

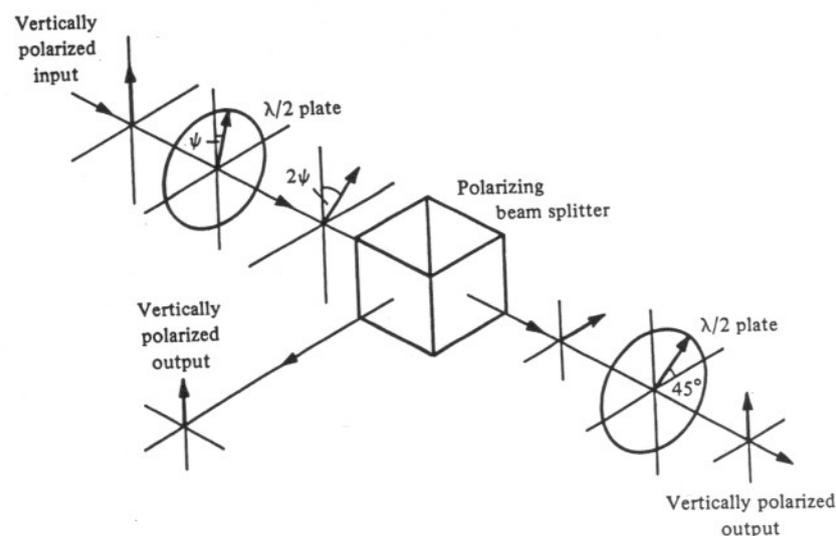
where ψ is the angle which the incident electric vector makes with the vertical. This ratio can be conveniently controlled by using a half-wave plate to rotate the direction of polarization of the input beam. Another fixed half-wave plate in the transmitted beam is used to bring the electric vector of this beam back to vertical.

5.5. Beam expansion

The laser beam has to be expanded to illuminate the object and the plate on which the hologram is recorded. Usually this is done with microscope objectives.

If the laser is oscillating in the TEM_{00} mode, the beam has a Gaussian profile so that the amplitude at a point at a radial distance r from the centre

Fig. 5.4. Variable-ratio beam splitter using a polarizing prism.



of the beam is

$$a(r) = a(0) \exp(-r^2/w^2), \quad (5.5)$$

where $a(0)$ is the amplitude at the centre of the beam and w is the beam radius (the radius at which the amplitude drops to $1/e$ of its maximum value).

Accordingly, $P(r)$ the laser power within a circle of radius r is given by the relation

$$P(r) = \int_0^r I(r) 2\pi r dr, \quad (5.6)$$

where $I(r) = |a(r)|^2$ is the intensity at a radial distance r from the centre of the beam; this can be written as

$$P(r) = P_{\text{tot}} [1 - \exp(-2r^2/w^2)], \quad (5.7)$$

where P_{tot} is the total power in the beam.

From (5.5) and (5.7) it follows that

$$P(r)/P_{\text{tot}} = 1 - [I(r)/I(0)]. \quad (5.8)$$

This is a useful relation giving the loss in power that must be tolerated for a given degree of uniformity of illumination.

Due to the high coherence of laser light, the expanded beam invariably exhibits diffraction patterns (spatial noise) due to scattering from dust particles on the optical surfaces in the beam path. To eliminate these, a pinhole is placed at the focus of the microscope objective. If this pinhole has a radius ρ , spatial frequencies higher than $\xi = \rho/\lambda f$, where λ is the wavelength of the light and f is the focal length of the microscope objective, are blocked. These higher spatial frequencies mainly represent noise, so that the transmitted beam has a smooth Gaussian profile.

A proper choice of the size of the pinhole can ensure that the power loss is minimal. Thus, it follows that the amplitude in the focal plane of the microscope objective, obtained from the two-dimensional Fourier transform of (5.5), is

$$A(\rho) = A(0) \exp(-\pi^2 w^2 \rho^2 / \lambda^2 f^2), \quad (5.9)$$

where ρ is the radial distance from the centre of the beam.

Typically, with an argon-ion (Ar^+) laser, $\lambda = 514 \text{ nm}$ and $w = 0.8 \text{ mm}$, and, with a 16 mm microscope objective and a $10 \mu\text{m}$ diameter pinhole, the amplitude at the edge of the pinhole is only 0.08 of that at the centre. Hence, from (5.8), more than 99 per cent of the total power in the beam is transmitted through the pinhole.

5.6. Exposure control

An accurate spot photometer is required to measure the irradiances due to the object and reference beams in the hologram plane, so as to set the beam ratio R at a suitable value. Because of the limited dynamic range of photographic materials used for holography, the object illumination should be adjusted so that the irradiance in the hologram plane due to it is reasonably uniform. In addition, precise control of the exposure is required to ensure good diffraction efficiency and avoid nonlinear effects. It is not enough to maintain a specified exposure time because the laser output can fluctuate during the exposure. To overcome this problem it is convenient to use an electronic exposure-control unit which integrates the irradiance in the hologram recording plane and closes the shutter at a preset value of radiant exposure [Lin & Beauchamp, 1970a]. A more sophisticated system which is suitable for multicolour holography has been described by Oreb & Hariharan [1981].

5.7. Coherence requirements

In order to obtain maximum fringe visibility, it is also essential to use coherent illumination. Gas lasers provide an intense source of highly coherent light and are therefore used almost universally in optical holography.

Spatial coherence is automatically ensured if the laser oscillates in a single transverse mode, preferably the lowest order or TEM_{00} mode, since this is inherently the most stable and gives most uniform illumination over the field. Normally, this is no problem since most gas lasers are designed to operate in this mode [Bloom, 1968]. However, they are not usually designed for single-frequency operation, which would imply that they should also oscillate in only one longitudinal mode. The temporal coherence of the light from most lasers is therefore limited by their longitudinal mode structure.

To obtain a satisfactory hologram, the maximum optical path difference between the object and reference beams in the recording setup must be less than the coherence length (see Appendix A3.4) of the light from the laser. With an extended object, the holodiagram [Abramson, 1969] (see section 14.9) can be used to optimize the layout.

5.8. Temporal coherence of laser light

The simplest form of resonant cavity for a laser is made up of a pair of mirrors separated by a distance L , though in some cases, where operation

on more than one line is possible, a wavelength selective prism may also be necessary as shown in fig. 5.5. The resonant frequencies of such a cavity are given by the expression

$$\nu_n = n(c/2L), \quad (5.10)$$

where n is an integer and c is the speed of light. However, as shown in fig. 5.6, the laser can oscillate only at those frequencies within the gain curve of the active medium at which the gain is adequate to overcome the cavity losses.

The width of the individual modes depends on the losses as well as the mechanical stability of the cavity structure and is typically about 3 MHz. Accordingly, if a laser is made to oscillate in a single longitudinal mode, the coherence length would be of the order of 100 metres.

Since, in frequency space, the width of the individual modes is much less

Fig. 5.5. Optical system of a typical Ar⁺ laser.

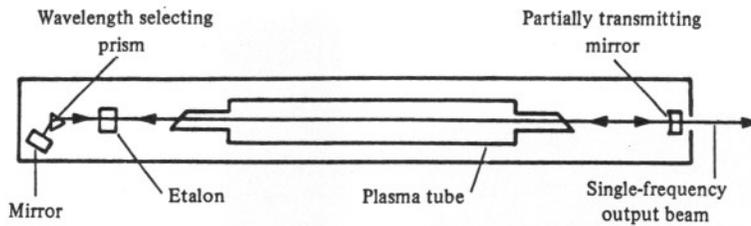
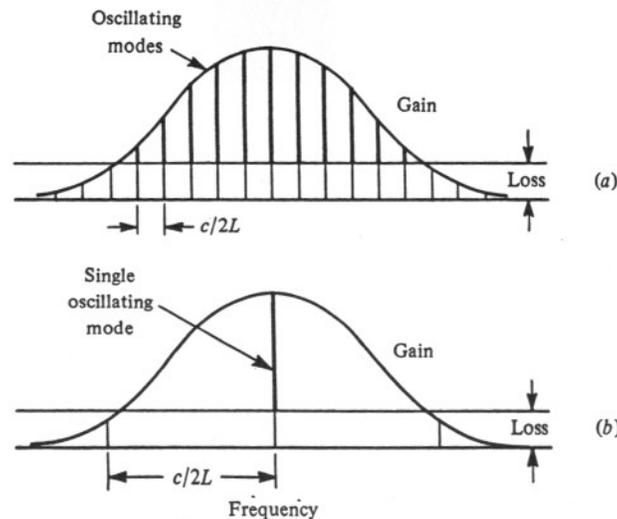


Fig. 5.6. Oscillation frequencies of a laser with (a) a long resonant cavity and (b) a short resonant cavity.



than their separation (which may range from 600 MHz for a 25 cm cavity down to 75 MHz for a 2 m cavity), the power spectrum of a laser oscillating in N longitudinal modes can be represented by N equally spaced delta functions. If we assume that the total power is equally divided between the N modes, the power spectrum can be written as

$$S(\nu) = \sum_n^{n+N-1} \delta(\nu - \nu_n). \quad (5.11)$$

The degree of temporal coherence can then be evaluated from (A3.14), and, for a path difference p , is

$$\gamma_{12}(p) = \frac{|\sin(N\pi p/2L)|}{N \sin(\pi p/2L)}. \quad (5.12)$$

This is a periodic function whose first zero occurs when

$$p = 2L/N, \quad (5.13)$$

corresponding to the effective coherence length. Fringes with acceptable visibility can usually be obtained for path differences that are less than half the coherence length.

It is apparent from (5.13) that the existence of more than one longitudinal mode in a laser reduces the coherence length severely. A short coherence length is troublesome, since it makes it essential to equalize the mean optical paths of the object and reference beams and restricts the maximum depth of the object field that can be recorded.

The simplest way to force a laser to operate in a single longitudinal mode is to use a very short cavity so that the spacing of the longitudinal modes is greater than the width of the gain profile over which oscillation is possible. This occurs when

$$c/2L \geq \Delta\nu, \quad (5.14)$$

where $\Delta\nu$ is the width of the gain profile (typically 1.7 GHz for a helium-neon (He-Ne) laser, 3.5 GHz for an Ar⁺ laser). However, the power available from such a short cavity is extremely limited.

The most common method of ensuring single frequency operation is to use an intracavity etalon as shown in fig. 5.5. In effect the laser is now made up of two resonant cavities, and only those modes which are common to both cavities, as shown in fig. 5.7, have low enough losses for oscillation to be possible. If the length of the etalon is made short enough that it satisfies (5.14), it can support only one mode, and single-frequency operation is obtained.

If the etalon is to act as a simple transmission filter, it must be tilted to decouple it from the laser cavity. It can then be tuned to maximize the

output power by tilting it further so that its resonant frequency corresponds to the peak of the gain curve [Hercher, 1969]. An alternative method of tuning which gives better frequency stability and efficiency is to mount the etalon within an oven which can be maintained at any desired temperature. Decoupling is easier if an etalon with concentric surfaces is used [Hariharan, 1982].

5.9. Gas lasers

The light source most commonly used for holography is a gas laser. Gas lasers, in general, are cheaper, easier to operate and have better coherence characteristics than other types of lasers. The range of useful wavelengths and typical output powers available at these wavelengths with the most commonly used gas lasers are summarized in Table 5.1.

When working with any laser, adequate safety measures must be taken to avoid eye damage (see ANSI, 1980; BSI, 1983). Even with relatively low-power gas lasers (≈ 1 mW) the beam should not enter the eye directly. This is because the beam is focused into a very small spot on the retina resulting in a power density about 10^5 times that at the cornea. With medium power lasers (< 100 mW) care must also be taken to avoid stray reflections, but the

risk of eye damage is minimal once the beam has been diffused or expanded. However, at power levels > 1 W, viewing even a light-coloured diffusing surface illuminated by the unexpanded beam is dangerous, since the surface intensity is comparable to that of the sun. With such lasers, as well as with pulsed lasers (see section 5.10) safety glasses must be worn.

For a simple holographic set up, the He-Ne laser is by far the most economical choice. It operates on a single line at 633 nm, does not require water cooling and has a long life. However, depending on their power, commercial He-Ne lasers oscillate in three to five longitudinal modes, and the coherence length is limited.

In contrast to the He-Ne laser, the Ar⁺ laser essentially has a multiline output but can be made to operate on a single line by replacing the reflecting end mirror by a prism and mirror assembly. It is also relatively easy to obtain single-frequency operation with an etalon. Argon-ion lasers can give high power output and an extended coherence length in the blue or green regions of the spectrum, the two strongest lines being at 488 nm and 514 nm.

The krypton-ion (Kr⁺) laser is very similar in its construction and characteristics to the Ar⁺ laser and is a useful replacement for the He-Ne laser, where high output power and an extended coherence length are required at the red end of the spectrum (647 nm).

The helium-cadmium (He-Cd) laser provides a stable source at a relatively short wavelength (442 nm). It is very useful with recording materials such as photoresists (see section 7.3), whose increased sensitivity at this wavelength makes up for the lower power available.

Fig. 5.7. Normal multifrequency output, etalon transmittance and single-frequency output with an intracavity etalon for a typical Ar⁺ laser.

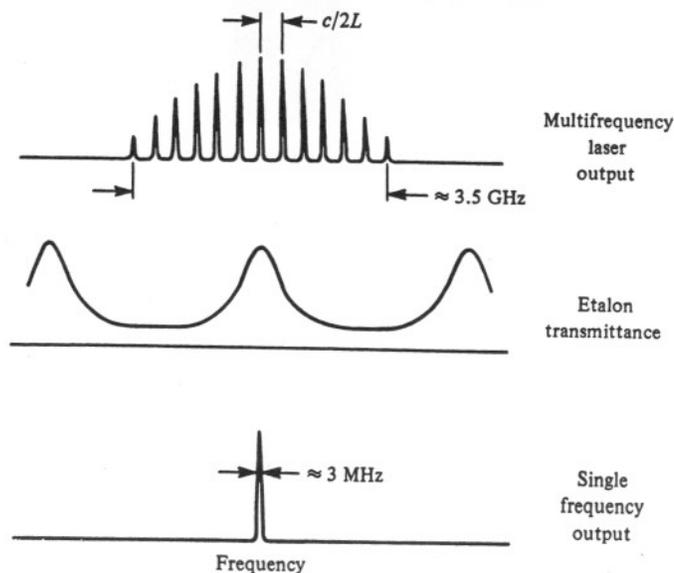


Table 5.1. Output wavelength and output power of gas lasers

Wavelength nm	Laser	Typical power mW	Colour
442	He-Cd	25	Violet
458	Ar ⁺	200	Blue-violet
476	Kr ⁺	50	Blue
477	Ar ⁺	400	Blue
488	Ar ⁺	1000	Green-blue
514	Ar ⁺	1400	Green
521	Kr ⁺	70	Green
633	He-Ne	2-50	Red
647	Kr ⁺	500	Red

The recording medium

Before we look at the different recording materials which have been used for optical holography, it is necessary to define the most important characteristics of a holographic recording medium.

6.1. Macroscopic characteristics

Any material used to record a hologram must respond to exposure to light (after additional processing, where necessary) with a change in its optical properties. The complex amplitude transmittance of such a material can be written in the most general manner as

$$\begin{aligned} t &= \exp(-\alpha d) \exp[-i(2\pi nd/\lambda)], \\ &= |t| \exp(-i\phi), \end{aligned} \quad (6.1)$$

where α is the absorption constant of the material, d is its thickness and n is its refractive index. Accordingly, as mentioned earlier, holographic recording materials can be classified, for convenience, either as pure amplitude-modulating materials if only α changes with the exposure, or as pure phase-modulating materials if $\alpha \approx 0$ and either n or d changes with the exposure.

The response of the recording material, defined by (6.1), can be described in these two limiting cases by curves of amplitude transmittance *versus* exposure ($|t| - E$ curves) as shown in fig. 6.1, or curves of the effective phase shift against exposure ($\Delta\phi - E$ curves) as shown in fig. 6.2.

6.2. The modulation transfer function

While curves such as those in figs. 6.1 and 6.2 describe the behaviour of a recording material quite satisfactorily on a macroscopic scale, they are not, by themselves, adequate to predict the response of the material on a microscopic scale.

As discussed in section 5.2 the intensity at any point on the hologram recording medium is given by a relation of the form

$$I = \langle I \rangle [1 + \mathcal{V} \cos(\phi_r - \phi_o)], \quad (6.2)$$

where $\langle I \rangle$ is the average intensity, \mathcal{V} is the visibility of the interference

fringes formed by the object and reference beams, and ϕ_o and ϕ_r are their phases at this point.

The actual intensity distribution to which the material is exposed is different, because there is always a certain amount of lateral spreading of light within the recording medium, determined by its scattering properties and its coefficient of absorption. Because of this, the actual modulation of the intensity within the recording material is always less than that in the original interference pattern.

For any spatial frequency s , the ratio of $\mathcal{V}'(s)$, the actual modulation of

Fig. 6.1. Curve of amplitude transmittance $|t|$ against exposure (E) for a recording material.

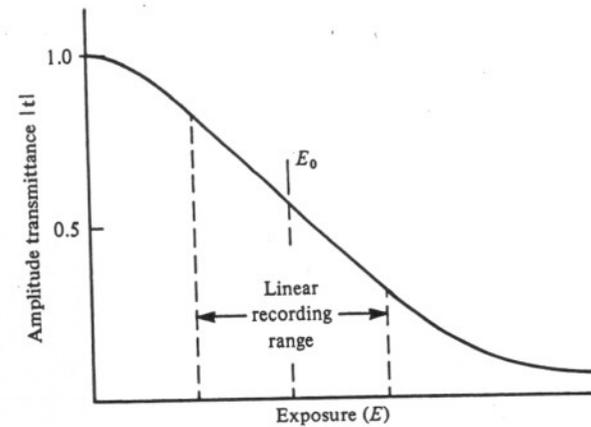
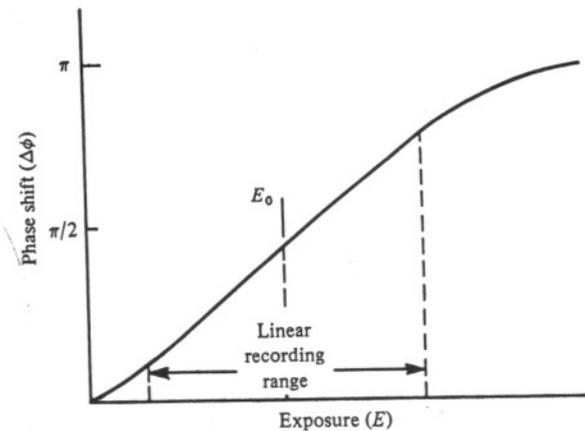


Fig. 6.2. Curve of phase shift ($\Delta\phi$) against exposure (E) for a recording material.



the intensity distribution within the material, to $\mathcal{V}(s)$, the output modulation, is termed the modulation transfer function (or MTF) $M(s)$, so that

$$M(s) = \mathcal{V}'(s)/\mathcal{V}(s). \quad (6.3)$$

This parameter, which is normally less than unity, serves to characterize the relative response of the material at different spatial frequencies.

6.3. Diffraction efficiency

The diffraction efficiency ε of a hologram has been defined (see section 3.6) as the ratio of the power diffracted into the desired image to that illuminating the hologram.

For an exposure time T , the effective exposure of the recording material is, from (6.2) and (6.3),

$$\begin{aligned} E &= TI, \\ &= T\langle I \rangle [1 + \mathcal{V}(s)M(s) \cos(\phi_r - \phi_o)], \\ &= \langle E \rangle [1 + \mathcal{V}(s)M(s) \cos(\phi_r - \phi_o)]. \end{aligned} \quad (6.4)$$

If we assume linear recording, as in (2.2), the amplitude transmittance of the recording medium can be written as

$$t = \langle t \rangle + \beta dE, \quad (6.5)$$

where $\langle t \rangle$ is the amplitude transmittance for the average exposure $\langle E \rangle$ and β is the value of (dt/dE) at $\langle E \rangle$. Accordingly, the amplitude transmittance of the hologram is

$$t = \langle t \rangle + \beta \langle E \rangle \mathcal{V}(s)M(s) \cos(\phi_r - \phi_o). \quad (6.6)$$

When the hologram is illuminated once again with the original reference wave r , the amplitude of the reconstructed image is

$$u_3 = (1/2)r\beta \langle E \rangle \mathcal{V}(s)M(s), \quad (6.7)$$

and its intensity is

$$I_3 = (1/4)|r|^2 [\beta \langle E \rangle \mathcal{V}(s)M(s)]^2. \quad (6.8)$$

Accordingly, the diffraction efficiency of the hologram is

$$\varepsilon = (1/4)[\beta \langle E \rangle \mathcal{V}(s)M(s)]^2. \quad (6.9)$$

Now, since $\beta = (dt/dE)$,

$$\begin{aligned} \beta E &= E(dt/dE), \\ &= \log_{10} e \cdot (dt/d \log E), \\ &= 0.434\Gamma, \end{aligned} \quad (6.10)$$

where $\Gamma = (dt/d \log E)$.

The diffraction efficiency is proportional to the square of Γ , the gradient

of the t versus $\log E$ curve, as well as to the squares of $\mathcal{V}(s)$, the input modulation, and $M(s)$, the MTF [Biedermann, 1969]. The maximum diffraction efficiency is obtained where the slope of the t versus $\log E$ curve is steepest. This is usually at a slightly higher exposure than that corresponding to the steepest part of the t versus E curve.

A method of characterizing recording materials for holography based on (6.9) has also been described by Lin [1971]. For an ideal recording material, plots of $\sqrt{\varepsilon}$ versus $\langle E \rangle$ with $\mathcal{V}(s)$ as a parameter and plots of $\sqrt{\varepsilon}$ versus $\mathcal{V}(s)$ with $\langle E \rangle$ as a parameter should be straight lines. The departure of the characteristics of any real material from the ideal can therefore be easily seen by comparing the actual measured curves with these. Apart from the maximum diffraction efficiency, or the exposure needed to obtain a given diffraction efficiency, these curves also make it possible to determine the range of fringe visibility $\mathcal{V}(s)$ or beam ratio R for which the hologram recording is linear (indicated by the straight line portion of the $\sqrt{\varepsilon}$ versus $\mathcal{V}(s)$ curve at constant $\langle E \rangle$) and the value of average exposure representing the best compromise between linearity and efficiency.

6.4. Image resolution

With an ideal recording material, the resolution of the image is determined only by the dimensions of the hologram. However, with any practical recording material, its MTF will, in general, affect the resolution as well as the intensity of the reconstructed image [Kozma & Zelenka, 1970].

This is because, in most holographic systems, the spatial frequency s varies over the hologram. It then follows from (6.9) that, due to the corresponding variations in the MTF of the recording medium, the diffraction efficiency of the hologram will vary over its aperture. Wherever this variation is appreciable, it must be taken into account in evaluating the imaging properties of the hologram.

The only exceptions to this situation are when the image is formed at or near the hologram or with a Fourier hologram (see section 2.3). In the latter case, the spatial frequency s is constant over the entire hologram. As a result, the MTF of the recording medium only affects the intensity of the image; its resolution is determined by the aperture of the hologram.

At the other extreme, where the object and the reference source are at different distances from the hologram, and the hologram is quite large, it is possible for the spatial frequency s to exceed the resolution limit of the recording material over part of the hologram. This limits the useful aperture of the hologram and, hence, the resolution of the image.

noise distribution for a diffusely illuminated object. Unfortunately, the expressions obtained are quite complicated. This difficulty is partly overcome in an alternative treatment by Tischer [1970] based on the use of Chebyshev polynomials. With simple objects such as lines and points, the terms of the series represent corresponding pairs of ghost images, so that, in principle, the optimum operational conditions can be evaluated readily.

However, the polynomial method has limitations, as pointed out by Velzel [1973], who was able to develop a complete theory of nonlinear holographic image formation using a modified transform method, and derive analytical expressions for the efficiency and the image contrast for recording media with an exponential characteristic (a linear phase hologram) and a binary characteristic. A generalization of this theory which can be used where the transfer function of the recording medium is complex has been outlined by Ghandeharian & Boerner [1977].

6.6. Effect of hologram thickness

Intermodulation effects would make it almost impossible to produce bright holographic images of good quality but for the fact that, in a volume hologram, intermodulation noise is reduced significantly by the angular selectivity of the hologram (see sections 4.4 and 4.5).

Consider a simple object consisting of only two points, which gives rise to two sets of interference fringes with the reference beam having spatial frequencies s_{01} and s_{02} . The intermodulation terms then have the general form $ps_{01} \pm qs_{02}$, where p and q are positive integers. Most of the resultant spatial frequencies are therefore significantly different from those corresponding to the object points. Qualitatively, it is apparent that when the hologram is illuminated once again with the same reference beam used to record it, the Bragg condition will not be satisfied for these intermodulation frequencies [Upatnieks & Leonard, 1970].

A detailed analysis of intermodulation noise in a volume hologram due to nonlinear recording has been made by Guther & Kusch [1974], using the coupled wave theory (see section 4.3). Their results show that the noise is effectively limited by the thickness of the recording medium as well as by the aperture of the hologram. A simpler analysis by Hariharan [1979a] shows that if the angle between the two beams in the recording setup is large enough for the diffracted beams corresponding to different orders not to overlap, the signal-to-noise ratio should improve by a factor approximately equal to $(\psi/\Delta\theta)$ where $2\Delta\theta$ is the width of the passband of the angular selectivity function, and ψ is the angular extent of the object.

6.7. Noise

In optical holography, noise is a convenient term for non-image-forming light which is diffracted or scattered in the same general direction as the reconstructed image.

If we exclude noise generated by nonlinear effects, which has been discussed in section 6.5, the other source of noise with recording materials such as photographic emulsions is scattered light from the individual grains, which are distributed in the emulsion layer in a random manner [Goodman, 1967; Kozma, 1968a].

The noise spectrum of a photographic material can be studied with the optical system shown in fig. 6.3. If a uniformly exposed photographic plate with an amplitude transmittance $t(x, y)$ is placed in front of the lens and illuminated with a collimated beam of monochromatic light, the diffracted amplitude at any point (x_f, y_f) in the back focal plane of the lens is (see Appendix A2.3)

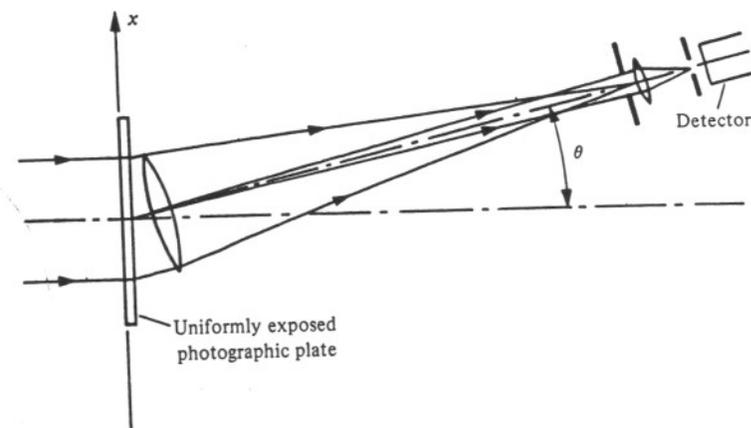
$$a(x_f, y_f) = (ia/\lambda f) \exp [(-i\pi/\lambda f)(x_f^2 + y_f^2)] T(\xi, \eta), \quad (6.15)$$

where a is the amplitude of the plane wave illuminating the plate, f is the focal length of the lens and $t(x, y) \leftrightarrow T(\xi, \eta)$; ξ and η are spatial frequencies defined by the relations $\xi = x_f/\lambda f$, $\eta = y_f/\lambda f$.

The intensity at this point is then

$$I(x_f, y_f) = |a(x_f, y_f)|^2 = (a^2/\lambda^2 f^2) |T(\xi, \eta)|^2. \quad (6.16)$$

Fig. 6.3. Arrangement for measuring the noise spectrum of photographic materials used for holography.



If this is averaged over a large enough area to eliminate local fluctuations due to speckle, it can be written as

$$I(\xi, \eta) = (a^2/\lambda^2 f^2) S(\xi, \eta), \quad (6.17)$$

where $S(\xi, \eta)$ is the power spectrum of the transmittance of the plate.

To eliminate effects due to diffraction at the edges of the sample, a second lens is normally used to image the sample on to another aperture placed in front of the detector. The effective area of the sample is then the projection of this aperture on the sample. Since this is proportional to $1/\cos \theta$, it also compensates for the decrease in the effective passband of the collecting aperture [Biedermann & Johansson, 1975].

6.8. Signal-to-noise ratio with coherent illumination

In computing the signal-to-noise ratio in the image reconstructed by a hologram, it is necessary to take into account the fact that it is the amplitudes of the signal and the noise which have to be added, since they are both encoded on the same coherent carrier [Goodman, 1967].

Consider the reconstructed image of an object consisting of a bright patch on a dark background. Let the intensity due to the nominally uniform signal be I_i , while that of the randomly varying background is I_N . The noise N in the bright area can be defined as the variance of the resulting fluctuations of the intensity, and is given by the relation

$$N = [\langle I^2 \rangle - \langle I \rangle^2]^{1/2}, \quad (6.18)$$

where I is the intensity at any point and $\langle I \rangle$ is the average intensity.

Since a_N the complex amplitude of the background has circular statistics (see Appendix A4), the mean value of terms involving a_N or a_N^* as well as powers of a_N and a_N^* , other than those involving only $|a_N|^2$, is zero. Hence, if a_i is the complex amplitude of the signal, the average intensity is

$$\begin{aligned} \langle I \rangle &= \langle (a_i + a_N)(a_i^* + a_N^*) \rangle, \\ &= \langle I_i + I_N + a_i a_N^* + a_i^* a_N \rangle, \\ &= I_i + \langle I_N \rangle. \end{aligned} \quad (6.19)$$

Similarly,

$$\begin{aligned} \langle I^2 \rangle &= \langle (I_i + I_N + a_i a_N^* + a_i^* a_N)^2 \rangle, \\ &= \langle I_i^2 + I_N^2 + 4I_i I_N \rangle, \\ &= I_i^2 + \langle I_N^2 \rangle + 4I_i \langle I_N \rangle. \end{aligned} \quad (6.20)$$

However, from (A4.4) and (A4.5), $\langle I_N^2 \rangle = 2\langle I_N \rangle^2$, so that

$$\langle I^2 \rangle = I_i^2 + 2\langle I_N \rangle^2 + 4I_i \langle I_N \rangle. \quad (6.21)$$

Accordingly, from (6.18), (6.19) and (6.21) we have

$$\begin{aligned} N &= [\langle I_N^2 \rangle + 2I_i \langle I_N \rangle]^{1/2}, \\ &= \langle I_N \rangle [1 + (2I_i/\langle I_N \rangle)]^{1/2}, \end{aligned} \quad (6.22)$$

so that the signal-to-noise ratio is

$$I_i/N = I_i/\langle I_N \rangle [1 + (2I_i/\langle I_N \rangle)]^{1/2}. \quad (6.23)$$

In the limiting case when $I_i \gg \langle I_N \rangle$ (which is usually the situation for a hologram recording of good quality), the signal-to-noise ratio becomes

$$I_i/N = (I_i/2\langle I_N \rangle)^{1/2}. \quad (6.24)$$

With coherent illumination, the signal-to-noise ratio is proportional to the square root of the ratio of the intensities of the signal and the scattered background. Even a weak scattered background leads to relatively large fluctuations in intensity in the bright areas of the image.

Practical recording media

The ideal recording material for holography should have a spectral sensitivity well matched to available laser wavelengths, a linear transfer characteristic, high resolution and low noise. In addition, it should either be indefinitely recyclable or relatively inexpensive.

While several materials have been studied [Smith, 1977; Hariharan, 1980b], none has been found so far that meets all these requirements. However, a few have significant advantages for particular applications. This chapter reviews, in the light of the general considerations discussed in Chapter 6, the properties of some of these materials (see Table 7.1 for a summary of their principal characteristics).

7.1. Silver halide photographic emulsions

Silver halide photographic emulsions are still the most widely used recording material for holography, mainly because of their relatively high sensitivity and because they are commercially available. In addition, they can be dye sensitized so that their spectral sensitivity matches the most commonly used laser wavelengths.

An apparent drawback of photographic materials is that they need wet processing and drying; however, development is actually an amplification process, with a gain of the order of 10^6 , which yields high sensitivity as well as a stable hologram. Another advantage of the formation of a latent image is that the optical properties of the recording medium do not change during the exposure, unlike materials in which the image is formed in real time. This makes it possible to record several holograms in the same photographic emulsion without any interaction between them.

Data on some of the silver halide photographic emulsions available for holography are summarized in Table 7.2, while typical $|t|$ versus E and spectral sensitivity curves are presented in figs. 7.1 and 7.2. Most of these emulsions are available coated on glass plates or film in a range of sizes and normally have an antihalation backing. Plates without any antihalation backing are available for making reflection holograms, though it is also possible to remove the antihalation backing with alcohol.

Table 7.1. Recording materials for holography

Material	Reusable	Processing	Type of hologram	Exposure required J/m^2	Spectral sensitivity nm	Resolution limit mm^{-1}	Max. diffraction efficiency (sine grating)
Photographic emulsions	No	Wet chemical	Amplitude (normal) Phase (bleached)	5×10^{-3} – 5×10^{-1}	400–700	1000–10 000	0.05 0.60
Dichromated gelatin	No	Wet chemical	Phase	10^2	350–580	> 10 000	0.90
Photoresists	No	Wet chemical	Phase	10^2	uv–500	3000	0.30
Photopolymers	No	Post exposure	Phase	10 – 10^4	uv–650	200–1500	0.90
Photochromics	Yes	None	Amplitude	10^2 – 10^3	300–700	> 5000	0.02
Photothermoplastics	Yes	Charge and heat	Phase	10^{-1}	400–650	500–1200 (bandpass)	0.30
Photorefractive							
LiNbO ₃	Yes	None	Phase	10^4	350–500	> 1500	0.20
Bi ₁₂ SiO ₂₀	Yes	None	Phase	10	350–550	> 10 000	0.25

12.8.2. Multiple imaging using lensless Fourier holograms

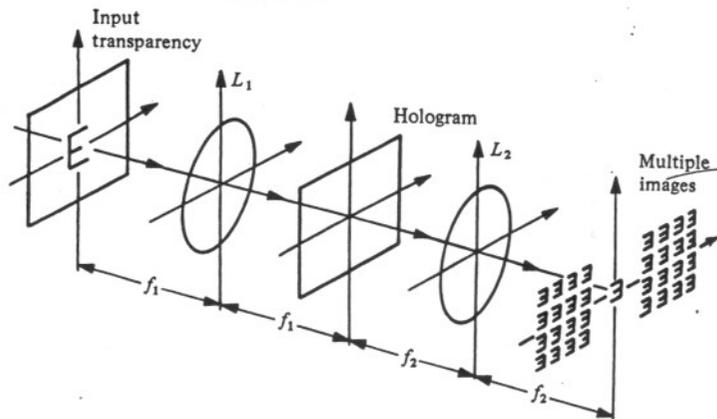
This technique, developed by Groh [1968], uses a much simpler optical setup. In the first step, a lensless Fourier hologram of an array of point sources $P_1 \dots P_n$ is recorded with a point reference source R .

To use this hologram to generate multiple images, it is illuminated with the conjugate to the original reference wave by means of a lens placed behind it, as shown in fig. 12.9. The hologram then produces real images of the array of object points $P_1 \dots P_n$ in their original positions. If then, the point source is replaced by an illuminated transparency, an array of images of the transparency are formed, centred on the positions of the original point sources $P_1 \dots P_n$.

A variation of these methods [Kalestynski, 1973, 1976] is to record a hologram of the transparency using multiple reference beams. When illuminated with a single reference beam, this hologram then produces an array of images.

Problems arise with all these techniques, due to cross-talk, if the hologram recording is not strictly linear. These can be avoided, at the expense of a considerable reduction in diffraction efficiency, if the hologram is produced by successive exposures using individual object beams separately, rather than all of them together. Another problem is that only the centres of the images are free from aberrations. Accordingly, the individual images must subtend only a small angle at the hologram.

Fig. 12.8. Multiple-image generation by a Fourier hologram [Lu, 1968].



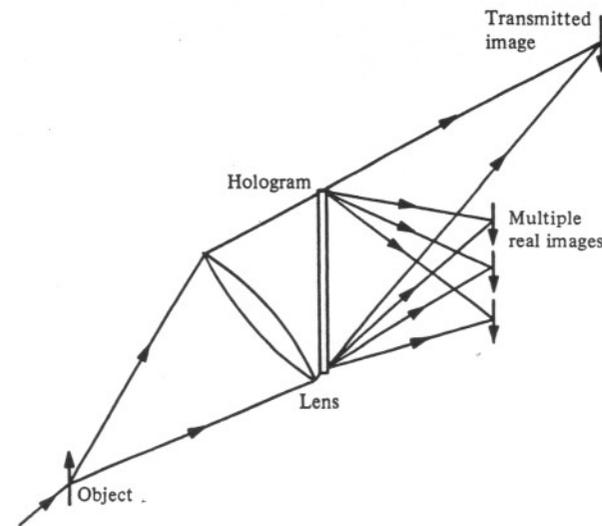
12.9. Holographic diffraction gratings

Diffraction gratings formed by recording an interference pattern in a suitable light-sensitive medium (now commonly called holographic diffraction gratings) are rapidly replacing conventional ruled gratings in spectroscopy. While Burch & Palmer [1961] first showed that transmission gratings could be made by photographing interference fringes using silver halide emulsions, it was the use of photoresist layers coated on optically worked blanks which finally led to the production of spectrographic gratings of high quality [Rudolph & Schmahl, 1967; Labeyrie & Flamand, 1969]. After processing, these yield a relief image (see section 7.3) which can be coated with an evaporated metal layer and used as a reflection grating.

Holographic gratings have several advantages over ruled gratings. Besides being cheaper and simpler to produce, they are free from periodic and random errors and exhibit much less scattered light. In addition, it is possible to produce much larger gratings of finer pitch, as well as gratings on substrates of varying shapes, and gratings with curved grooves and varying pitch. This makes it possible to produce gratings with unique focusing properties and opens up the possibility of new designs of spectrometers (see, for example, the review by Namioka, Seya & Noda [1976]).

Against this, their main disadvantage is that the groove profile cannot be

Fig. 12.9. Multiple imaging by means of a hologram of an array of point sources [Groh, 1968].



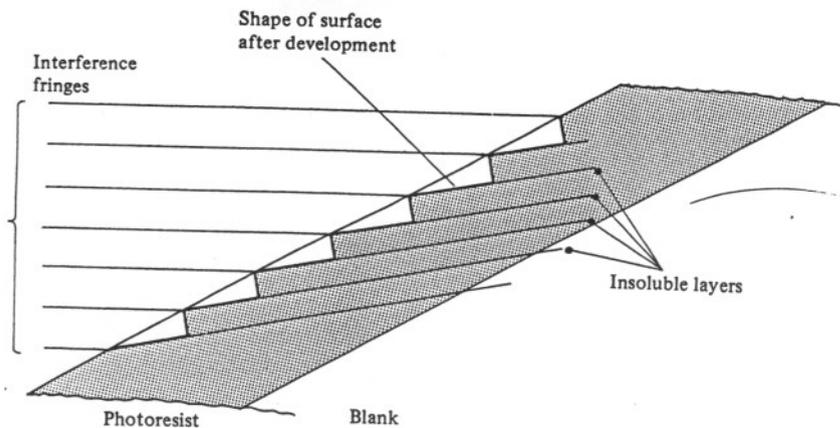
controlled as easily as in ruled gratings. While even a sinusoidal profile can give quite high diffraction efficiencies for very small grating spacings ($\approx \lambda$) [Loewen, Maystre, McPhedran & Wilson, 1975] it is usually necessary to produce a triangular groove profile for maximum diffraction efficiency. Accordingly, a number of methods have been proposed for the production of blazed holographic gratings (see, for example, the reviews by Schmahl & Rudolph [1976] and by Hutley [1976, 1982]).

One method of achieving this result is to expose the photoresist to a sawtooth irradiance distribution, which is built up by a process of Fourier synthesis, either by using more than two beams to produce the fringes, or by making multiple exposures to fringe patterns of appropriate periodicities and phases [McPhedran, Wilson & Waterworth, 1973; Schmahl, 1975; Breidne, Johansson, Nilsson & Ahlen, 1979].

Another possibility which has been explored is to start with a sinusoidal or partially blazed profile and modify it by ion-beam etching to produce a well-formed triangular profile [Aoyagi & Namba, 1976; Aoyagi, Sano & Namba, 1979].

In the most widely used method, however, [Sheridon, 1968; Hutley, 1975] the photoresist layer is aligned obliquely to the fringe pattern, as shown in fig. 12.10. This produces, within the thickness of the resist, layers which are alternately soluble and insoluble. After development, the surface profile is determined by the shape of the insoluble layers near the surface. The only disadvantage of this technique is that one of the beams is incident through the back of the blank, which must therefore be of optical quality.

Fig. 12.10. The production of blazed gratings in photoresist [Hutley, 1982].



An optical system for this purpose is shown in fig. 12.11. In this, light from an Ar^+ laser ($\lambda = 458 \text{ nm}$) is split into two beams of equal intensity which are focused by microscope objectives on pinholes. Each pinhole is located at the focus of an off-axis parabolic mirror, so that a collimated beam is obtained upon reflection from the mirror. As shown in fig. 12.10, the photoresist-coated blank is placed in the interference field at a small angle to the standing waves.

To produce gratings of good quality, the optics must produce wavefronts plane to $\lambda/10$. The liquid photoresist is applied to the optically worked blank, which is then spun rapidly to produce a uniform layer, about $0.5 \mu\text{m}$ thick. In addition to the precautions normally taken to ensure stability of the fringes during the exposure, a closed-loop servo system is used to maintain the optical path difference in the interferometer stable to better than $\lambda/50$.

12.10. Holographic scanners

Holographic scanners are a relatively new development which could solve many of the problems associated with mirror scanners. Their most promising applications are in point-of-sale terminals and, with a modulated laser beam, for high-speed non-impact printing.

A simple disc holographic scanner [Cindrich, 1967; McMahon, Franklin & Thaxter, 1969] is shown in fig. 12.12. The disc has a number of holograms

Fig. 12.11. Optical system used to produce blazed holographic gratings [courtesy I. G. Wilson, CSIRO Division of Chemical Physics, Melbourne, Australia].

