OPTICAL HOLOGRAPHY
Student Edition

Robert J. Collier
Christoph B. Burckhardt
Lawrence H.-Lin
Bell Telephone Laboratories
Murray Hill, New Jersey
5. Optical Systems with Spherical Lenses

Chapter 7
LIGHT SOURCES AND OPTICAL TECHNIQUE

As laser sources became available for forming holograms, holography began its transformation to a practical science. Before the laser, the standard sources were mercury discharge lamps which represented at best a compromise between coherence and output intensity. Spatial coherence could be acquired only by reducing the area of the source and thereby drastically reducing the useful output. In contrast, laser radiation can be both highly coherent and intense. Yet, even with a laser source, holography can be difficult and unrewarding if the holographer is unaware of laser properties and requisite optical technique.

In this chapter we examine the coherence properties of continuous wave light sources as they relate to hologram formation and to wavefront reconstruction. Methods of deploying laser beams to obtain the desired subject and reference illumination are discussed, and we present a simple example of an optical arrangement and procedure employed in practical holography.

7.1 Light Sources for Hologram Formation

The requirements placed on light sources to be used for hologram formation depend on parameters determined by the subject and the arrangement of necessary optical components. Figures 7.1a and 7.1b illustrate two common methods of deriving subject and reference waves from a single source. In Fig. 7.1a they are derived by wavefront division and in Fig. 7.1b by amplitude division. In each case the waves reaching the hologram originate in a
source wavefront whose angular extent is $2\theta_0$. As indicated in Fig. 1.18, the degree of spatial or lateral coherence $|\mu_s|$ of a source depends not only on the source size but also on $\theta_0$. If we expect the relative phase for all of the subject and reference wavefronts to remain substantially time independent (permitting the recording of an interference pattern), the source emission must be sufficiently spatially coherent over $2\theta_0$. It must also have sufficient temporal or longitudinal coherence. More precisely, the maximum optical path difference for light rays traveling between source and hologram must be less than the coherence length.

We state in Section 1.9.2, Eq. (1.22) that the complex degree of coherence $\gamma_{12}(\tau)$ encompasses both spatial and temporal coherence. By letting $\tau \to 0$ in Eq. (1.22) we imply that we are concerned only with the lateral aspect of the source coherence. If the correlation of $\nu_{P_1}$ and $\nu_{P_2}$ in Eq. (1.22) is a function of $\tau$ alone, we imply concern only with the longitudinal coherence. This is a somewhat artificial division but useful in discussing the limitations of conventional thermal sources and the advantages of lasers. Our discussion of laser sources will be confined here to continuous-wave gas discharge devices and to He-Ne and argon lasers in particular. Hologram formation using pulsed lasers is treated in Chapter 11. Semiconductor lasers are at present not very useful for holography as they emit only weakly in the visible range, and their coherence properties are poor.

### 7.1.1 Spatial Coherence Requirements

The coherence properties of a laser beam are intimately related to the oscillation mode-structure of the laser. A comprehensive review of the mode-structure can be found in Kogelnik and Li [7.1]. So far as holography is concerned, the radiation from a laser oscillating in any one transverse mode is spatially coherent. It is preferable that the mode of oscillation be the lowest order mode, the TEM$_{00}$, since more uniform illumination can be obtained from this mode than from higher order modes. Higher order modes are also inherently less stable, there being a tendency to oscillate in two or more simultaneously. Almost all commercially available lasers oscillate in only the lowest order mode or can be so adjusted.

In contrast, a conventional nonlaser source is far from being spatially coherent. To increase the spatial coherence of a source of arbitrary cross section, all but a small fraction of its radiating surface must be masked off. If we set some meaningful criterion for minimum spatial coherence, we can use the van Cittert-Zernike theorem (see Section 1.9.3) to estimate the maximum spatial extent of a nonlaser source suitable for forming holograms with the methods illustrated in Fig. 7.1a, b.

For this purpose consider the geometry shown in Fig. 7.2 where a small pinhole $S$, located in the $x'y'$ plane, allows radiation from only a small area of the source to pass to an $xy$ plane a distance $R$ away. We define two sampling positions $P_1$ and $P_2$ on the $xy$ plane with coordinates $(0,0)$ and $(x,y)$, respectively. The angle subtended at the pinhole by the distance $r$ between $P_1$ and $P_2$ is $\theta_0$. Figure 7.2 describes a situation similar to that (via reflections from the subject surface or reference mirrors) must be less than the coherence length.
Light Sources and Optical Technique

7. Light Sources and Optical Technique

Fig. 7.2. Geometry related to discussion of the spatial coherence of a thermal source.

illustrated in Fig. 1.17; Eq. (1.22) defining the complex degree of coherence \( \gamma_{12}(r) \) applies:

\[
\gamma_{12}(r) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} v_{p_1}(t + \tau) v_{p_2}^*(t) \, dt
\]

where \( v_{p_1} \) and \( v_{p_2} \) represent the complex electric fields at \( P_1 \) and \( P_2 \) and \( \tau \) is the transit time difference for light traveling from \( P_1 \) and \( P_2 \) to an observation point \( Q \) (see Fig. 1.17). For the case \( r = 0 \) and with the points \( P_1 \) and \( P_2 \) located as in Fig. 7.2, \( \gamma_{12}(0) = \mu_s(x, y) \) can be written as

\[
\mu_s(x, y) = \frac{\int_{-\infty}^{\infty} v(0, 0, t) v^*(x, y, t) \, dt}{\left( \int_{-\infty}^{\infty} v(x, y, t) v^*(x, y, t) \, dt \right)^{1/2}}.
\]  

(7.1)

The quantity \( \mu_s \) is the complex degree of spatial (lateral) coherence of the apertured source as measured in the \( xy \) plane.

By means of the van Cittert-Zernike theorem [7.2] we can express \( |\mu_s| \), the degree of spatial coherence, as the magnitude of a normalized Fourier transform of the intensity distribution over the pinhole:

\[
|\mu_s(x, y)| = \left| \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x', y') \exp[2\pi i(\xi x' + \eta y')] \, dx' \, dy'}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x', y') \, dx' \, dy'} \right|.
\]  

(7.2)

where \( \xi = x/\lambda R, \eta = y/\lambda R \) [as in Eq. (5.38)], and \( \lambda \) is the mean wavelength of the radiation emitted by the source. The validity of the theorem depends on the following assumptions:

1. The radiation from the source is quasi monochromatic; i.e., the mean wavelength \( \lambda \) is much greater than the deviation \( \Delta \lambda \).
2. The separation \( R \) between the pinhole and the \( xy \) plane is much greater than the extent of the pinhole or the distance \( r \).
3. The radiation inside the pinhole is spatially incoherent.
4. The coherence length of the source radiation \( c/\Delta f = \lambda^2/\Delta \lambda \) is much greater than the maximum difference in optical paths between either of the sampling points and any point in the source. (Here, \( f \) is the temporal frequency.)

The assumptions can all be satisfied in practice; the bandwidth \( \Delta \lambda \) can be made as small as necessary by passing the source radiation through a narrow-band filter.

If we assume the intensity of the source to be uniform over the circular pinhole of radius \( r_0 \), integration of Eq. (7.2) yields

\[
|\mu_s| = \left( \frac{\pi r_0^2}{\lambda^2} \right) \frac{J_1(2\pi r_0 \beta \lambda)}{\pi r_0 \beta \lambda} \]

(7.3)

where we have employed correspondence (4.34). As indicated in Chapter 4, \( J_1(2\pi r_0 \beta \lambda)/\pi r_0 \beta \lambda \) has a maximum value of one. The spatial frequency \( v \) can be expressed as

\[
v = (\xi^2 + \eta^2)^{1/2} = (1/\lambda R)(x^2 + y^2)^{1/2} = r/\lambda R \sim \theta/\lambda
\]

where the angle \( \theta \) is that indicated in Fig. 7.2, and \( |\mu_s| \), the degree of spatial coherence, then becomes

\[
|\mu_s| = \left| \frac{J_1(2\pi r_0 \beta \lambda)}{\pi r_0 \beta \lambda} \right|.
\]  

(7.4)

Figure 7.3 is a plot of \( |\mu_s| \) versus \( r_0 \beta \lambda \).

Suppose we desire to form a hologram using one of the methods of Fig. 7.1 and require \( 2\theta_0 = 30^\circ \) or \( \theta_0 = \pi/12 \) rad. If we also require the degree of spatial coherence to be no less than \( |\mu_s| = 1/\sqrt{2} = 0.707 \) (an arbitrary criterion which will be justified later) and if we set \( \lambda = 0.5 \mu m \), then from Fig. 7.3 we find that \( (r_0 \beta \lambda)_{\max} = 0.25 \), and the maximum allowed diameter for the pinhole is \( 2r_0 = 2.5 \mu m \). When we use a pinhole whose diameter is as small as one micrometer to obtain sufficient spatial coherence from a thermal
source, we greatly reduce the useful power. Among the nonlaser sources, the high-pressure mercury arc lamp radiates the most power per unit area of its emitting surface. Typically this is about 100 W/cm² at \( \lambda = 5461 \) Å, with \( \Delta \lambda \approx 50 \) Å. A pinhole 1 \( \mu \)m in diameter would limit the useful power of the source to 1 \( \mu \)W. Use of a wavelength filter to increase the temporal coherence would cause a further reduction. In contrast, useful power of 100 mW is not difficult to obtain from a continuous-wave laser.

### 7.1.2 Temporal Coherence of Conventional Sources

The temporal coherence of a light source is ultimately determined by the spectral purity of its radiation. When a laser oscillates in only one axial or longitudinal mode (and, of course, one transverse mode), it possesses almost perfect temporal coherence so far as holography is concerned. While most commercially available lasers are adjusted to oscillate in the lowest order transverse mode, they are, however, not constructed for single-longitudinal-mode (single-frequency) operation. In fact, the temporal coherence property of a typical (multiple-mode) gas laser is not much better than that of the spontaneous emission from the same atomic transition under a similar gas discharge condition. In this section we first set up a criterion for temporal coherence and then examine the temporal coherence properties of some conventional gas discharge sources.

Equation (7.4) tells us that the degree of spatial coherence between light at points \( P_1 \) and \( P_2 \) on the \( xy \) plane of Fig. 7.2 depends on the position coordinates of \( P_1 \) and \( P_2 \) only as a function of the angle \( \theta \) between light rays passing from the extended source to each point. Let us consider now the case where \( P_1 \) and \( P_2 \) are spaced along the same ray coming from a point source so that \( \theta = 0 \), \( |\mu_1| = 1 \), and the complex degree of coherence depends only on \( r \). To signify this we replace \( \gamma_1(r) \) in Eq. (1.22) by \( \mu_T(r) \), the complex degree of temporal coherence, and replace \( v_1(r) = v_2(r) \) by \( v(t) \) and \( v_2(t + r) \) by \( v(t + r) \) so that

\[
\mu_T(r) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \frac{v(t + r)v^*(t)}{v(t)v^*(t)} dt = \frac{\int_{-\infty}^{\infty} v(t + r)v^*(t) dt}{\int_{-\infty}^{\infty} v(t)v^*(t) dt} \tag{7.5}
\]

where \( v(t) \) is the complex electric field at \( P_2 \), \( v(t + r) \) is the complex electric field at \( P_1 \) (a point on the same ray between the source and \( P_2 \) but closer to the source). Equation (7.5) holds when a wave is amplitude-divided (Fig. 7.1b) and light from the same original ray travels different paths to an observation point. In this latter arrangement, \( r \), the transit time difference, can be finite though \( P_1 \) and \( P_2 \) are coincident. The transit time difference \( r \) can be expressed in terms of an optical path difference \( \Delta L = cr \) where \( c \) is the speed of light. We now define a coherence length \( \Delta L_{\|} = cr \) such that \( |\mu_T(r_{\|})| = 1/\sqrt{2} \). We shall show in the next section that this criterion is meaningful for holography. For successful hologram formation we require that the difference between any pair of optical paths from light source to any point on the recording medium be less than \( \Delta L_{\|} \). To determine \( \Delta L_{\|} \) for a given source we must be able to plot \( |\mu_T(r)| \) as a function of \( r \). This may be done by rewriting Eq. (7.5) in terms of temporal frequency \( f \) and introducing the power spectrum of the desired source.

Equation (7.5) can be transformed into

\[
\mu_T(r) = \frac{\int_{-\infty}^{\infty} \Phi(f) \exp(2\piifr) df}{\int_{-\infty}^{\infty} \Phi(f) df} \tag{7.6}
\]

where \( \Phi(f) = V(f)V^*(f) \) is the power spectrum and \( V(f) \) is the temporal Fourier transform of \( v(t) \). [See Appendix II for a general definition of \( v(r) \).]

The equivalence of Eqs. (7.5) and (7.6) may be demonstrated simply by considering the result of equating the numerators and denominators separately. Equality of the numerators is merely a statement of the autocorrelation correspondence which holds equally well for temporal variables as for spatial [see correspondence (4.18)]. Equality of the denominators
represents the special case of the autocorrelation correspondence at $r = 0$. Paralleling the spatial Fourier transform relation between $|\mu_\tau(x, y)|$ and $I(x', y')$ in Eq. (7.2), the two functions $|\mu_\tau(r)|$ and $\Phi(f)$ form a temporal Fourier transform pair.

Using Eq. (7.6) let us calculate the degree of temporal coherence $|\mu_\tau(\tau)|$ for some conventional nonlaser light sources. Gas discharge lamps are the most coherent nonlaser sources. An appropriate filter or monochromator can be used to isolate a single spectral line from the discharge. The power spectrum $\Phi(f)$ corresponding to a spectral line emitted by a low-pressure discharge lamp is approximated by the Gaussian expression [7.3]

$$\Phi(f) = \Phi(f) \exp\left\{-\left(\frac{\pi f \Delta f_D}{2\ln 2}\right)^2\right\}$$

where $f$ is the center frequency of the spectral line and $\Delta f_D$ is the full (Doppler) width of the line at half intensity. Upon substitution of Eq. (7.7) into Eq. (7.6) we find in the numerator of Eq. (7.6) the shifted Fourier transform of a Gaussian function. This may be evaluated with the help of temporal versions of correspondences (4.21) and (4.27). The denominator of Eq. (7.6) may be looked up in a table of integrals with the result that

$$|\mu_\tau(\tau)| = \exp\left[-\left(\frac{\pi f \Delta f_D}{2\ln 2}\right)^2\right]$$

and

$$|\mu_\tau(\tau)| = \exp\left[-\left(\frac{\pi f \Delta f_D}{2\ln 2}\right)^2\right].$$

A high-pressure lamp emits a spectral line whose power spectrum is better approximated by a Lorentzian [7.3]

$$\Phi(f) = \Phi(f)\left[1 + \left(\frac{2(f - f')^2}{\Delta f_L}\right)^{-1}\right]$$

where $\Delta f_L$ is the Lorentz line width due to collisions of the various gas and charged particles. Substitution of Eq. (7.9) into Eq. (7.6) and application of (1) the shift theorem, (2) correspondence 444 of Campbell and Foster [7.4], and (3) a table of integrals yields

$$\mu_\tau(\tau) = \exp(-2\pi ifx) \exp(-\pi f \Delta f_L)$$

and

$$|\mu_\tau(\tau)| = \exp(-\pi f \Delta f_L).$$

We can now determine $\Delta L_H = \Delta f_D$ from Eqs. (7.8) and (7.10) or Fig. 7.4, where $|\mu_\tau|$ is plotted as a function of either $\tau \Delta f_D$ or $\tau \Delta f_L$. When $|\mu_\tau| = 1/\sqrt{2} = 0.707$, $\tau_H \Delta f_D = 0.32$ while $\tau_H \Delta f_L = 0.11$. The corresponding coherence lengths are then

$$\Delta L_H = \frac{c\sqrt{2} \ln 2}{\tau \Delta f_D} = \frac{0.32c}{\Delta f_D}, \quad \text{low-pressure discharge},$$

and

$$\Delta L_H = \frac{c\ln 2}{2\tau \Delta f_L} = \frac{0.11c}{\Delta f_L}, \quad \text{high-pressure discharge.}$$

Fig. 7.4. Degree of temporal coherence of a thermal source with a Gaussian or Lorentzian spectral profile.

One of the most coherent of the nonlaser sources is the 6058 Å orange line of 60Kr. The Doppler width $\Delta f_D$ of this line can be made as narrow as $4.5 \times 10^3$ Hz [7.5] and its coherence length is $\Delta L_H = 21$ cm according to Eq. (7.11). If it were not for the low output power per unit area of the source and the small degree of spatial coherence which results when the source is extended, such a source would be useful for holography. A much brighter source is the 5461 Å green line from a high-pressure mercury arc lamp. However, its line width $\Delta f_L$ is about $5 \times 10^4$ Hz ($\Delta \lambda \approx 50$ Å) yielding a coherence length $\Delta L_H \approx 8 \mu$m. To use the light for holography, the line must be narrowed further by passing it through a wavelength filter.

Let us now calculate the coherence lengths of the spontaneous emissions at 6328 Å and at 4880 Å from neon and argon gas discharges, respectively.
We will later compare these with coherence lengths for laser lines at the same wavelengths. The Doppler widths of the two lines are approximately $1.5 \times 10^9$ Hz at 6328 Å and $7.5 \times 10^9$ Hz at 4880 Å under gas discharge conditions common to helium-neon lasers and argon ion lasers. According to Eq. (7.11) the corresponding coherence lengths are approximately 6.4 cm for the 6328 Å red line of atomic neon and 1.5 cm for the 4880 Å blue line of the argon ion.

7.1.3 TEMPORAL COHERENCE OF THE GAS LASER

We now consider the temporal coherence properties of a laser and discuss some methods of obtaining single-longitudinal-mode or single-frequency laser operation under laboratory conditions. We assume, to begin with, that the laser is designed or adjusted for oscillation in the lowest order transverse mode (TEM$_{00}$) but not necessarily for oscillation in a single longitudinal mode. Thus the power spectrum $\Phi(f)$ and hence the temporal coherence properties of the laser depend entirely on the longitudinal mode structure of the laser. We confine our attention to these longitudinal modes.

The resonant cavity of the laser generally consists of a pair of spherical mirrors spaced at a distance $l$ apart. The resonant frequency of the $n$th cavity mode is

$$f_n = n(c/2l)$$

(7.13)

where $n$ is an integer and $c$ is the speed of light. Laser oscillation occurs at all frequencies $f_n$ inside a frequency range $\Delta f_g$ for which the laser active medium has sufficient gain to overcome the power loss (including the loss due to the output coupling) of the cavity. This oscillation frequency range $\Delta f_g$ can be greater or less than the Doppler width $\Delta f_D$. The width $\Delta f_M$ of each mode is determined by the loss and the mechanical and thermal stability of the cavity structure. Typically $\Delta f_M \approx 10^9$ Hz (over an observation period of a few minutes). Let us assume that only one longitudinal mode is allowed to oscillate and that the power spectrum $\Phi(f)$ for the mode can be approximated by a Gaussian having a half-intensity width of $\Delta f_M = 10^9$ Hz. The coherence length of such a single-frequency laser, according to Eq. (7.11), is $\Delta L_H \approx 1 \times 10^8$ cm or 1 km. The optical path lengths encountered in most holographic applications are certainly much less than 1 km. We are therefore justified in considering that $\Delta L_H \to \infty$ and that the power spectrum may be written as a $\delta$ function. If more than one longitudinal mode can oscillate, $\Phi(f)$ can then be represented by series of $\delta$ functions.

What is the coherence length of the laser if more than one longitudinal mode can oscillate? We consider first the case where oscillation can be supported in just two modes. This will happen when the frequency range $\Delta f_n$, within which the laser has sufficient gain for oscillation, spans twice the mode spacing $f_{n+1} - f_n$, i.e.,

$$\Delta f_n/2 = f_{n+1} - f_n = c/2l.$$

The power spectrum of this two-mode laser source can be represented by two delta functions

$$\Phi(f) = I_n \delta(f - f_n) + I_{n+1} \delta(f - f_{n+1})$$

where the $I_n$'s represent the total power in an individual mode. Substituting $\Phi(f)$ into Eq. (7.6) we have

$$\mu_\tau(\tau) = \frac{1}{I_n + I_{n+1}} \cdot \left[ I_n \exp(2\pi i f_n \tau) + I_{n+1} \exp(2\pi i f_{n+1} \tau) \right]$$

$$= \exp(2\pi i f_n \tau) \left[ a_n + a_{n+1} \exp\left(\frac{2\pi i c}{2l \tau} \right) \right]$$

(7.14)

where we have used the expression for the mode spacing $f_{n+1} - f_n = c/2l$ and where

$$a_n = \frac{I_n}{I_n + I_{n+1}}, \quad a_{n+1} = \frac{I_{n+1}}{I_n + I_{n+1}}, \quad \text{and} \quad a_n + a_{n+1} = 1.$$

The degree of coherence $|\mu_\tau(\tau)|$ is given by

$$|\mu_\tau(\tau)| = |a_n + a_{n+1} \exp(i\pi c \tau/l)|$$

$$= \left| a_n^2 + a_{n+1}^2 + 2a_n a_{n+1} \cos(\pi c \tau/l) \right|^{1/2}$$

(7.15)

and is plotted in Fig. 7.5 for two values of the parameter $b = |a_{n+1} - a_n|$. The parameter $b$ represents the minimum value of $|\mu_\tau(\tau)|$. When $b = 0$, $a_n = a_{n+1} = \frac{1}{2}$ and

$$|\mu_\tau(\tau)| = \left| \left(\frac{1}{2} + \frac{1}{2} \cos(\pi c \tau/l) \right)^{1/2} \right| = \left| \cos(\pi c \tau/2l) \right|.$$

When the power in one mode is twice that in the other, $b = \frac{1}{2}$, $a_n = \frac{1}{2}$, $a_{n+1} = \frac{1}{2}$, and

$$|\mu_\tau(\tau)| = \left| \left(\frac{1}{2} \right)^{1/2} \left[ 1 + \frac{1}{2} \cos(\pi c \tau/l) \right]^{1/2} \right|.$$

---

In summary, we have considered the coherence properties of lasers with single-longitudinal-mode operation and have seen how these properties affect the performance of holographic systems.
Regardless of the value of \( b \), the degree of coherence \( |\mu_\tau(\tau)| \) for the two-mode case is periodic in \( \tau \) with a period \( 21/c \). The condition \( |\mu_\tau(\tau)| = 1/\sqrt{2} \) is satisfied by multiple values of \( \tau \) as shown in Fig. 7.5. We define the smallest of these as \( \tau_H \) from which we obtain the coherence length \( \Delta L_{\text{H}} = c\tau_H \). For \( b = 0 \), \( \tau_H = \frac{\delta}{L(1/c)} \), and \( \Delta L_{\text{H}} = \frac{\delta}{L} \). For path length values between 0 and \( 2 \Delta L_{\text{H}} \), the degree of temporal coherence varies monotonically with path difference as in the case of the nonlaser sources indicated in Fig. 7.4. The length of a typical gas laser is \( l = 2 \Delta L_{\text{H}} = 1 \text{ m} \). Path differences required for hologram formation can generally be kept within the limit \( \Delta L_{\text{H}} = 0.5 \text{ m} \).

Due to thermal instabilities in the cavity length \( l \), the value of \( b \) can take on a range of values. When, however, the gain required for oscillation can be obtained only over a frequency range \( \Delta f_0 \) much narrower than the Doppler line width \( \Delta f_D \), the oscillating modes are nearly equal in power and \( b \approx 0 \). Figure 7.5 indicates that even when one mode has twice the power of the other (\( b = \frac{1}{3} \)), the value of \( \tau_H \) (corresponding to \( |\mu_\tau(\tau)| = 0.707 \)) differs very little from that corresponding to \( b = 0 \). We therefore shall restrict ourselves in further discussion to the case where the powers of the modes are equal.

We note that for the case of two oscillating modes and a cavity length of \( 1 \text{ m} \), the coherence length \( \Delta L_{\text{H}} = 0.5 \text{ m} \) represents a drastic reduction from the \( 1 \text{ km} \) computed for the case of single-mode oscillation.

When \( N \) modes of oscillation are permitted, we can write the power spectrum of the laser radiation as

\[
\Phi(f) = I_n \delta(f - f_n) + I_{n+1} \delta(f - f_{n+1}) + \cdots + I_{n+N-1} \delta(f - f_{n+N-1}).
\]

Substitution into Eq. (7.6) yields the degree of coherence

\[
|\mu_\tau(\tau)| = \frac{1}{N} \left| \frac{\exp(2\pi i f_n \tau) + \exp(2\pi i f_{n+1} \tau) + \cdots + \exp(2\pi i f_{n+N-1} \tau)}{\sqrt{N}} \right|,
\]

where we have used the condition that \( I_n = I_{n+1} = \cdots = I_{n+N-1} \). We can write Eq. (7.16) in terms of the mode spacing \( c/2L \) by extracting the factor \( \exp(2\pi i f_n \tau) \) so that

\[
|\mu_\tau(\tau)| = \left| \frac{\exp(2\pi i f_{n-1} \tau)}{N} \times \left[ \exp(\pi i c \tau/L) + \exp(2\pi i c \tau/L) + \cdots + \exp(N\pi i c \tau/L) \right] \right|
\]

where we have used the formula for the sum of a geometric progression of ratio \( \exp(\pi i c \tau/L) \). If we multiply the quantity inside the absolute value sign by its complex conjugate, take its square root, and then take the absolute value of the result, \( |\mu_\tau(\tau)| \) becomes

\[
|\mu_\tau(\tau)| = \frac{1}{N} \left[ \frac{1 - \cos(N\pi c \tau/L)}{1 - \cos(\pi c \tau/L)} \right] = \frac{\sin(N\pi c \tau/L)}{N \sin(\pi c \tau/L)}.
\]

Again, as in the case of two-mode oscillation, the degree of coherence \( |\mu_\tau(\tau)| \) is periodic in \( \tau \). It is plotted in Fig. 7.6 for the cases of a laser oscillating in one, two, three, and four modes. The principal maxima (equal to unity) occur at equal intervals of \( \tau = 21/c \) regardless of the value of \( N \). Zeros occur whenever \( \tau = (m/N)(21/c) \), where \( m = 1, 2, 3, \ldots \), but \( m/N \neq 0, 1, 2, 3, \ldots \).

From the above discussion it is clear that single-frequency operation of a laser is highly desirable. A single-frequency laser offers nearly unlimited coherence length while even a few simultaneously oscillating modes can reduce the coherence length severely (see Fig. 7.6). Short coherence length in hologram formation limits the permissible subject depth and necessitates the often troublesome equalization of the mean optical path lengths of the subject and reference beams. Unfortunately, stable single-frequency operation of a laser is not easily obtained, and its achievement inevitably results in a reduction of output power.
7.6. Degree of temporal coherence for laser operating in one (I), two (II), three (III), and four (IV) longitudinal modes.

7.1.4 METHODS OF ACHIEVING SINGLE-FREQUENCY OPERATION

The most obvious method of achieving single-frequency operation of a laser is to reduce the cavity length \( I \) such that

\[
\Delta f = \frac{c}{2I} = A_f_0.
\]

For this case the gain is sufficient to sustain oscillation only over a band of frequencies \( A_f_0 \) equal to the mode spacing. Hence only a single mode can be supported. If we take \( A_f_0 \) to be approximately the Doppler linewidth \( A_f_0 = 1.5 \times 10^6 \) Hz, the cavity length corresponding to single-mode operation of a He–Ne laser at 6328 Å is \( I = c/(2A_f_0) = 10 \) cm. Only a small volume of the active medium can be enclosed in so short a cavity; consequently only a fraction of a milliwatt of output power can be obtained from the laser [7.5]. The single-frequency cavity length for the argon ion laser operating at 4880 Å is but 2 cm, making it very difficult to build such a laser.

A better approach to single-frequency operation is to couple together two resonant cavities so that only one mode, common to both cavities, can have sufficient gain to sustain laser oscillation. It is frequently desirable to make the optical length \( l_1 \) of one cavity long, to enclose a large volume of active medium while the optical length \( l_2 \) of the second cavity is kept short to widely separate the frequencies of common modes [see Eq. (7.13)]. If the length \( l_2 \) is limited to 10 cm for the He–Ne laser and to 2 cm for the argon ion laser so that the shorter cavity can support only one mode, then single-frequency operation is obtained.

Examples of the two-cavity method are shown in Fig. 7.7a, b. The first configuration [7.6, 7.7] has the virtue of simplicity, but the second configuration [7.8–7.11] is more effective at high power levels. In the former case (Fig. 7.7a) the mode-selecting second cavity is simply a fused-quartz etalon. Its physical length \( L \) is determined by the relation between it and the required optical length \( l_2 = n_1L \), where \( n_1 \) is the index of refraction of fused quartz. A slight tilting of the etalon is usually required to make \( l_2 \) an integral number of half-wavelengths and satisfy Eq. (7.13). Coating of the surfaces of the etalon, to increase its finesse, may not even be necessary in many cases.

In Fig. 7.7b the mode-selecting second cavity is formed by the mirrors \( M_4 \) and \( M_5 \) together with the beam splitter. The configuration produces high
transmission loss at all frequencies except the frequency of the common modes. The curvatures of the mirrors $M_1$ and $M_3$ are chosen to match properly wavefronts at the beam splitter.

Highest reported output powers from single-frequency lasers are about 50 mW at 6328 Å (He-Ne) and near 1 W at 4880 Å (argon ion). The main difficulty in achieving single-frequency operation with two-cavity methods is maintaining the dimensional stability of both cavities. With the increased temperatures accompanying high power levels the difficulty increases. If $l_1 > l_2$, only $l_2$ need be critically adjusted during the operation of the laser. This adjustment can be made by changing the tilt of the etalon in Fig. 7.7a or the position of the mirror $M_3$ in Fig. 7.7b. One can also servo-control the dimension $l_2$ by a feedback system which automatically locks the oscillation frequency to the center of the line $f$. In practice it is not too difficult to maintain stable operation of the laser for a few minutes without any feedback system. This is usually long enough to expose a hologram.

7.2 Fringe Visibility in Hologram Recording

Successful hologram recording depends first of all on the existence of high contrast interference fringes. The intensity for two interfering beams is given by Eq. (1.9) in the form

$$I = I_1 + I_2 + 2a_1 a_2 \cos(\phi_2 - \phi_1).$$

High contrast or large modulation depth implies that the amplitude of the spatially varying cosine term is comparable to the spatially constant term $I_1 + I_2$. The larger the amplitude of the sinusoidal component, the larger the amplitude of the wave diffracted by the hologram during the reconstruction step. Thus, high contrast in the fringe pattern implies bright images, and it is desirable to maximize the contrast. (Diffraction efficiencies of plane and volume holograms are discussed in Chapters 8 and 9, respectively.)

A second concern is for the fidelity with which the subject wave is reconstructed. We restrict our attention here to plane holograms. If, for example, an accurate reconstruction of an original subject wave is to be obtained from an absorption hologram, the hologram recording must have an amplitude transmittance which is linearly related to the intensity of the interference pattern. Let us see how (1) high contrast in the fringe pattern and (2) linear recording of the subject wave can be obtained.

7.2.1 Optimizing Fringe Visibility

Contrast is quantitatively measured by the fringe visibility or standing wave ratio $V$ given in Eq. (1.21) by

$$V = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}$$

(7.18)

where $I_{\text{max}}$ and $I_{\text{min}}$ denote the maximum and minimum intensities of the interference fringes in an observation plane. The fringe visibility may vary over the observation plane; it is defined over any small area larger than a fringe spacing. We shall find that the visibility depends on the degree of coherence $|\gamma_{12}(\tau)|$ between the interfering beams, the angle $\Omega$ between the directions of polarization of the beams, and the ratio $R$ of the intensities of the two beams all measured in the plane of observation. To facilitate obtaining a useful relation between $V$ and the parameters $|\gamma_{12}(\tau)|$, $\Omega$, and $R$, we impose the condition that the latter are all constant over any small area of the observation plane, and we consider the optical arrangement indicated in Fig. 7.8, where both reference and subject wave are derived from the same laser source. Each wave is spherical, the reference diverging from the point source $R$ while the signal appears to diverge from a virtual point source $P$. We are to observe the interference over a very small area $A$ of the observation plane. The area $A$ is large enough for the visibility to be well defined but small enough that the signal and reference waves can be considered plane and that $|\gamma_{12}(\tau)|$ and $\Omega$ can be considered constants.

![Fig. 7.8. Ray directions for the essentially plane wavefronts interfering over the small area $A$.](image-url)
The analysis will be carried out for light propagation vectors lying in the \(xz\) plane only.

We suppose that the laser is operating in a single transverse mode but in several longitudinal modes. Hence, spatial coherence can be considered essentially perfect, and the degree of coherence \(\mu_{x}(r)\) can be replaced by the degree of temporal coherence \(\mu_{t}(r)\). Before beginning the analysis of the interference of the waves indicated in Fig. 7.8, let us rewrite the expression for \(\mu_{t}(r)\) in Eq. (7.5) in a form which separates out the rapidly varying temporal phase factor contained in the complex electric field \(v\). Recall that in Section 1.3 the complex electric field for a monochromatic light wave was defined as

\[
v = a \exp(i \phi) \exp(2\pi ift).
\]

For light emission at several frequencies in a narrow band, \(v\) may be written

\[
v = a \exp(i \phi)[c_0 \exp(2\pi ift) + c_1 \exp(2\pi i(f + \varepsilon_1)t)
+ c_2 \exp(2\pi i(f + \varepsilon_2)t) + \ldots]
= a \exp(i \phi) \exp(2\pi ift)[c_0 + c_1 \exp(2\pi i\varepsilon_1t)
+ c_2 \exp(2\pi i\varepsilon_2t) + \ldots]
\]

or

\[
v(t) = a \exp(i \phi)g(t) \exp(2\pi ift)
\]

(7.19)

where \(g(t) = \sum c_i \exp(2\pi i\varepsilon_it)\) varies at a much slower rate than \(\exp(2\pi ift)\) providing \(\varepsilon_i \ll f\). We see that the complex electric field for narrow-band emission takes the same form as that for monochromatic light if we regard the complex amplitude to be

\[
a(t) = a \exp(i \phi)g(t).
\]

(7.20)

As in the monochromatic case, we may often replace \(v(t)\) with \(a(t)\). If we now substitute Eq. (7.19) into the equation for \(\mu_{t}(r)\), Eq. (7.5), we obtain

\[
\mu_{t}(r) = \exp(2\pi ifr) \cdot \frac{\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} g(t + \tau)g^*(t) \, dt}{\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} g(t)g^*(t) \, dt}
= \exp(2\pi ifr) \cdot \frac{\langle g(t + \tau)g^*(t) \rangle}{\langle g(t)g^*(t) \rangle}
\]

or

\[
\mu_{t}(r) = \mu_{x}(r) \exp(-2\pi ifr) = -\frac{\langle g(t + \tau)g^*(t) \rangle}{\langle g(t)g^*(t) \rangle}.
\]

(7.21)

7.2 Fringe Visibility in Hologram Recording

Let us return now to the interference of the partially coherent reference and signal waves of Fig. 7.8. For the complex amplitude of the reference wave over the area \(A\) we write

\[
r = r \exp(i \phi)g(t + \tau) = r \exp(2\pi i\xi_{x}t)g(t + \tau)
\]

(7.22)

where the spatial frequency \(\xi_{x}\) is defined for the mean wavelength \(\lambda\). The parameter \(\tau\) is defined by the quantity \(\omega \tau\), where \(c\) is the velocity of light and where \(c\) is the difference in optical path from laser source to the area \(A\) when reference and signal routes are compared. The intensity of the reference beam at \(A\) is

\[
I_{r} = \langle rr^* \rangle = r \langle g(t + \tau)g^*(t + \tau) \rangle.
\]

(7.23)

We can divide the signal wave into two components. One is polarized parallel to the reference wave polarization direction and has a complex amplitude at \(A\) given by

\[
a_{||} = a \exp(2\pi i\xi_{x}t)g(t) \cos \Omega
\]

(7.24)

and an intensity

\[
I_{a_{||}} = \langle a_{||}a_{||}^* \rangle = a^2 \cos^2 \Omega \langle g(t)g^*(t) \rangle.
\]

(7.25)

Here \(\Omega\) is the angle between the direction of polarization of reference and signal beams and \(\xi_{x}\) is the spatial frequency of the signal wave defined for the mean wavelength. The other component polarized perpendicular to the reference wave has a complex amplitude at \(A\),

\[
a_{\perp} = a \exp(2\pi i\xi_{x}t)g(t) \sin \Omega
\]

(7.26)

with an intensity

\[
I_{a_{\perp}} = \langle a_{\perp}a_{\perp}^* \rangle = a^2 \sin^2 \Omega \langle g(t)g^*(t) \rangle.
\]

(7.27)

We now desire (1) to substitute Eqs. (7.22)–(7.27) into an expression for the intensity of the interference pattern formed by \(r\), \(a_{||}\), and \(a_{\perp}\); (2) to express the intensity in terms of the degree of temporal coherence \(\mu_{t}(r)\); and (3) to compute the maximum and minimum values of the intensity and thereby the visibility \(V\) of Eq. (7.18). Equation (1.20), the expression for the intensity of the interference of two waves

\[
I = I_{1} + I_{2} + 2 \text{Re}[\langle y_{1}y_{2}^* \rangle]
\]

(7.28)
can be written in terms of complex amplitudes and generalized to express the intensity of the fringe pattern produced by \( r \), \( a \), and \( a' \). We note that the total subject wave intensity \( I_s = I_a + I_{a'} = a^2 \langle g(t)g^*(t) \rangle \) [from Eqs. (7.25) and (7.27)], and that only the subject wave component polarized parallel to the reference wave polarization will produce an interference term. The intensity of the pattern over \( A \) is then the sum of the individual wave intensities plus the interference term:

\[
I = I_a + I_{a'} + 2 \text{Re}\{\langle ra \rangle^* \}
\]

where the beam ratio \( R = (r/a)^2 \). Note that the visibility is independent of the spatial frequencies of either beam and that \( V(R) = V(1/R) \). The visibility has a maximum value of unity when \( |\mu_T|, R, \text{ and } \cos \Omega \) are each unity.

The degree of temporal coherence \( |\mu_T| \) of the light from a laser operating in several longitudinal modes is given in Eq. (7.17) and plotted in Fig. 7.6 as a function of \( r \). We see from the figure that \( |\mu_T| \) approaches unity for small values of \( r \). In particular, \( r \) should be small compared to \( r_H \), the value for which the degree of temporal coherence has the value 0.707. We

The maximum intensity of the fringe pattern \( I_{max} \) is obtained when \( \cos \beta(x,\tau) = +1 \), and the minimum value \( I_{min} \) is obtained when \( \cos \beta(x,\tau) = -1 \). Substituting the extreme values of intensity into Eq. (7.18), the visibility becomes

\[
V = \frac{2ra |\mu_T| \cos \Omega}{r^2 + a^2} = \frac{2 |\mu_T| \sqrt{R \cos \Omega}}{R + 1}
\]
can achieve this condition by ensuring that the difference in optical paths between reference and signal beams, \( A L = c \tau \), is small compared to the coherence length \( A L_\text{H} = c \tau H \). If we set \( R = 1 \) and \( \cos \Omega = 1 \) in Eq. (7.31), observation of the fringe visibility as a function of optical path difference offers a simple method of plotting \( | \mu_\text{e}(r) | \) versus \( A L \). A Michelson interferometer, shown in Fig. 7.9, satisfies the conditions on \( R \) and \( \cos \Omega \) and may be employed for the purpose. Figure 7.10 is an experimental plot of \( V \) versus \( A L \) obtained for a typical 1-m-long He-Ne laser operating at a wavelength of 6328 Å. The coherence length \( A L_\text{H} \), when \( V = 1/\sqrt{2} \), is approximately 12 cm.

When forming a hologram, the signal wave is spatially modulated with subject information, and it is not possible to maximize the visibility by making \( R = 1 \) everywhere over the hologram plane. The best one can do is to adjust the intensity of the reference beam to make \( R = 1 \) at locations of maximum intensity in the signal beam. As we shall see, concern for linearity in the recording generally results in \( R > 1 \) everywhere over the hologram plane.

Light from a typical laser is linearly polarized and it is generally not difficult to maintain \( \cos \Omega = 1 \) and so maximize visibility. Reference to Fig. 7.11, however, shows that it is desirable that both reference and signal beam at the hologram plane be polarized in directions perpendicular to the plane formed by their propagation vectors. (When the beams are kept parallel to an optical table, the desirable polarization direction is perpendicular to the table.) For this case \( \cos \Omega = 1 \) regardless of the angle \( \theta \) between the propagation vectors. On the other hand, when the laser light is polarized parallel to the plane of the propagation vectors, e.g., parallel to the plane of the paper in Fig. 7.11, the angle \( \Omega = \theta \). If in forming the hologram the beams intersect at right angles, \( \cos \Omega \) and the visibility are both zero. The direction of polarization of the light leaving the laser can of course be rotated so that it is perpendicular to the table surface by inserting into the beam a properly oriented half-wave plate (Fig. 7.12).

Certain subjects depolarize the light they reflect. In such cases the undesirable component can be eliminated by means of an analyzer and the visibility maximized by means of the beam ratio.

### 7.2.2 Linear Recording

Let us now consider the relation between linear wavefront recording and the visibility of holographic interference fringes. We assume the recording material is thin so that transmittances can be expressed in terms of \( x \) and \( y \). Suppose the medium is given an exposure (see Section 2.5.1)

\[
E(x, y) = I_\text{p}(x, y) \tau_\text{e} = k_2 I(x, y) \tau_\text{e} \tag{7.32}
\]

where \( \tau_\text{e} \) is the exposure time, \( k_2 \) is the constant of proportionality between the intensities \( I_\text{p} \) and \( I \) defined in Section 1.3, and \( I \), the intensity of the interference pattern, is given by Eq. (7.30). The amplitude transmittance of the exposed and developed record (the hologram) will be dependent in some manner on the exposure \( E \). Suppose that the exposure-dependent part of the transmittance is represented by \( t_\text{p} \). Here we allow the possibility of phase as well as amplitude modulation by the hologram. When \( I \) from Eq. (7.30) is inserted into Eq. (7.32), we have

\[
E(x) = k_2 \tau_\text{e} [r^2 + a^2 + 2ar | \mu_\text{e}(r) | \cos \Omega \cos \beta(x, r)]
= E_0 + E_1(x) \tag{7.33}
\]
where
\[ E_0 = k_1 r_x [r_x^2 + a^2] \]
and
\[
E_1(x) = 2k_1r_x |\mu_x(r)| \cos \beta(x, r) \\
= k_1r_x|\mu_x(r)| \cos \Omega \exp[2\pi i(\xi_x - \xi_a)x] \\
+ k_1r_x|\mu_x(r)| \cos \Omega \exp[-2\pi i(\xi_x - \xi_a)x].
\]  
(7.34)

Along with the above substitution comes the restriction to a small area \( A \) of the hologram plane. If we ensure that the beam ratio \( R > 1 \), then \( E_1 \) will always be less than \( E_0 \) and we can express the amplitude transmittance \( t \) over the area \( A \) as a Taylor series
\[
(7.35)
\]
When the hologram of transmittance \( t \) is illuminated by the original reference wave \( r = r \exp(2\pi i\xi_a x) \) (assumed monochromatic at this stage), the complex amplitude of the modulated light at the back side of area \( A \) is
\[ w = rt. \]

If, in Eq. (7.35), the coefficients of second and higher order terms are negligible, a wave can be reconstructed whose amplitude is linearly proportional to the original subject wave amplitude \( a = a \exp(2\pi i\xi_a x) \). Under this condition the reconstructed wave of interest becomes
\[
(7.36)
\]
over the area \( A \), where we have considered only the second term in Eq. (7.34). In the above,
\[ t_{(B)}(E_0) = \left. \frac{dt}{dE} \right|_{E_0} = \text{constant} \text{ if } \frac{dt}{dE} = 0. \]

We can regroup and write Eq. (7.36) in terms of the visibility \( V \) [see Eq. (7.31)] as
\[
(7.37)
\]
whose value remains independent of the entire range of exposure values found over the hologram plane during formation of the hologram. The range of exposure, \( E_{\text{min}} < E < E_{\text{max}} \), can be obtained from Eqs. (7.33) and (7.31), where \( E_{\text{min}} \) corresponds to \( \cos \beta = -1 \) and \( E_{\text{max}} \) corresponds to \( \cos \beta = +1 \). Expressing the exposure range in terms of the maximum visibility \( V_{\text{max}} \) found on any area of the hologram plane, we have
\[ E_0(1 - V_{\text{max}}) < E < E_0(1 + V_{\text{max}}). \]
(7.40)

Whether the condition \( dt/dE = \text{constant} \) can be satisfied over the range in (7.40) depends on the exposure characteristic of the photosensitive material used to record the hologram (see Chapter 10).

We can make some general observations on the conditions for linear recording, without reference to specific materials, by considering separately plane absorption and plane phase holograms. The complex transmittance of a photosensitive medium suitable for forming a plane hologram can be characterized by
\[ t = t \exp(i\phi) \]
and by its derivative with respect to exposure
\[
\frac{dt}{dE} = \frac{dt}{dE} \exp(i\phi) + it \exp(i\phi) \frac{db}{dE}. \]
(7.41)

When the material is exposed to light, the modulus \( t \) and the phase \( \phi \) of the transmittance may be altered. While to some degree photosensitive
materials exhibit both of these changes, the more important recording materials respond with a significant alteration of only one or the other. If $dt/dE$ is finite and $d\phi/dE = 0$, the material forms an absorption hologram. If $d\phi/dE$ is finite and $dt/dE = 0$, the material forms a phase hologram.

When an absorption hologram is to be formed in a material such as photographic emulsion, the condition for a linear recording, Eq. (7.39), becomes

$$\frac{dt}{dE} = \text{constant}.\tag{7.40}$$

To be assured of linearity we must first obtain a curve of transmittance versus exposure $(t-E)$ for the particular emulsion chosen. The exposure range $E_0(1-V_{\text{max}}) < E < E_0(1+V_{\text{max}})$ must remain within the (approximately) linear portion of the $t-E$ curve, Fig. 7.13. One obtains the curve experimentally by measuring the intensity transmittance $t$ as a function of $E$ and then computing $t = \sqrt{\mathcal{T}}$. Some $t-E$ curves corresponding to emulsions commonly used in holography are found in Chapter 10.

When a phase hologram is formed in an appropriate material such as thermoplastic, Eqs. (7.39) and (7.41) require that

$$\exp(\phi) \frac{d\phi(E)}{dE} = \text{constant} \quad \tag{7.42}$$

A given recording material might have a $\phi(E) - E$ characteristic as indicated in Fig. 7.14. Only for a small exposure range in the straight-line portion of the curve is $\phi(E)$ small and the recording essentially linear. For a larger exposure range and correspondingly larger value of $\phi(E)$ the recording becomes nonlinear. In practice, phase holograms yield high diffraction efficiency, indicative of appreciable exposure range and moderate values of $\phi(E)$, without intolerable nonlinearity effects in the recording.
7.3 Illumination with an Expanded Laser Beam

The beam of light emerging from a laser is typically 1 mm in diameter. For most holographic applications, however, we must illuminate a subject and a hologram recording medium of much larger size. We also desire the illumination to be as uniform as possible. Expansion of a laser beam can be accomplished with one or more lenses or spherical mirrors without significant loss of power or serious alteration of the intensity distribution in the beam. Uniform illumination can be obtained, however, only by sacrificing a portion of the laser output power.

The intensity distribution in a laser beam depends on its transverse mode structure. Except for the lowest order TEM$_{00}$ mode, the intensity vanishes at one or more places over the cross section of the beam. Uniform illumination, therefore, cannot be achieved with higher order modes. In the absence of diffraction effects due to the finite extent of the active laser medium, the radial distribution of intensity in the TEM$_{00}$ mode has the Gaussian form

$$I(r) = \frac{2P_t}{\pi w^2} \exp\left(-\frac{2r^2}{w^2}\right)$$  \hspace{1cm} (7.44)

where $P_t = \int_{0}^{\infty} I(r)2\pi r \, dr$ is the total output power of the laser, $r$ is the radial distance from the beam center, and $w$ is the half-width of the beam. (At the radius $r = w$, the beam intensity drops to $e^{-2}$ times that at the center.) This smooth variation in intensity can be modulated by diffraction effects originating at the edges of the active medium, e.g., at the walls of the gas discharge tube. Diffraction effects are of concern when the laser resonant cavity is formed with two large-radius mirrors. When a hemispherical laser cavity is employed, the effects are usually not discernible.

If the intensity distribution of the beam is a smooth Gaussian, free of diffraction modulation and noise, the beam can be expanded very simply by reflecting it from a front-surface concave or convex mirror. The main advantage of this method is that multiple reflections are avoided. The mirror, however, must be used in an off-axis configuration, introducing aberrations into the expanded beam. To minimize the aberrations the radius of curvature of the mirror should be as large as possible consistent with obtaining the desired beam cross section in a given distance.

If the beam intensity distribution is not a smooth Gaussian but contains variations which we can consider to be noise, the latter can be effectively removed by spatial frequency filtering. The laser beam is brought to a focus by a lens, and a pinhole is placed at the focal point. A simple microscope objective lens can serve both to focus and expand the beam (Fig. 7.15). The pinhole passes only the very lowest spatial frequencies characteristic of the slowly varying Gaussian distribution but blocks the higher spatial frequencies characteristic of the noise. Emerging from the pinhole is an essentially smooth Gaussian intensity distribution superimposed on spherical wavefronts. (A collimating lens can be used to obtain essentially plane wavefronts if desired.)

![Fig. 7.15. Beam expander and spatial filter.](image)

Let us now estimate the size of the pinhole required for effective spatial filtering by considering the incident beam to be Gaussian. If the diameter of the focusing lens is sufficiently large relative to the incident Gaussian beam width, we can neglect the effect of lens-edge diffraction on the Gaussian beam profile. The complex amplitude of the light incident on the lens is the square root of Eq. (7.44) and proportional to $\exp\left(-r^2/w_1^2\right)$, where $w_1$ is the radius or the half-width of the incident beam at the front focal plane of the lens and $r_1$ is the radial distance in that plane. Its Fourier transform is displayed by the lens in the plane of the pinhole and (from the discussion in Section 4.2) given by the correspondence

$$\exp\left(-\frac{r^2}{w_1^2}\right) \Rightarrow \text{const} \cdot \exp\left(-\pi^2 w_1^2 r^2 \frac{v^2}{\lambda f^2}\right)$$

$$\Rightarrow \text{const} \cdot \exp\left(-\frac{\pi^2 w_1^2 r^2}{\lambda f^2}\right)$$  \hspace{1cm} (7.45)

where $\gamma = r_1/\lambda f$, $f$ is the focal length of the lens, and $r_2$ is the radial distance in the plane of the pinhole. The intensity of the Gaussian beam in the back focal plane is the square of the right-hand side of correspondence (7.45)

$$I_g \propto \exp\left(-\frac{2\pi^2 w_1^2}{\lambda f^2} r^2\right)$$
with a beam half-width

$$w_b = \lambda f/nw_1.$$  \hfill (7.46)$$

Substituting typical values of $\lambda = 0.63 \mu m$, $f = 16$ mm, and $w_1 = 1$ mm, we find for the full width of the laser beam near the focus, $2w_2 = 6.4 \mu m$.

If a 10-\(\mu\)m-diameter pinhole is used as the spatial filter, it will easily pass the 6.4 \(\mu\)m laser beam. However, spatial frequencies in excess of $v = r_0/\lambda = 5 \mu m/(16 \text{ mm} \times 0.63 \mu m) = 0.5$ cycles/mm are stopped. The spatial frequencies associated with diffraction modulation, scattering from dust particles, and multiple reflections between lens surfaces are generally of higher value and are effectively filtered out by the 10-\(\mu\)m pinhole.

Let us assume that the laser beam has been expanded and that its intensity distribution remains approximately Gaussian. If the expanded beam illuminates a subject or hologram plate, the uniformity of the illumination will depend on the maximum radius $r_0$ of that portion of the beam intercepted by the subject (or plate). Taking the subject to be a circular disk of radius $r_0$, the intensity of the light beam at the edge of the disk is [see Eq. (7.44)]

$$I_e = I_0 \exp\left(-\frac{2r_0^2}{w^2}\right)$$  \hfill (7.47)$$

where $I_0 = 2P_1/\pi w^2$ is the light intensity at the center of the disk and $P_1$ and $w$ are the total power and half-width of the laser beam at the plane of the disk. The laser power which actually contributes to the illumination of the disk is obtained by integrating Eq. (7.44) over the radius $r$

$$P_i = \frac{2P_1}{\pi w^2} \int_0^{r_0} \exp\left(-\frac{2r^2}{w^2}\right) 2\pi r \, dr$$

$$= P_1\left[1 - \exp\left(-\frac{2r_0^2}{w^2}\right)\right].$$

Substituting Eq. (7.47) for the exponential, we have

$$\frac{P_i}{P_1} + \frac{I_e}{I_0} = 1.$$  \hfill (7.48)$$

If, for example we require that the light intensity at the edge of the disk be at least 50\% of that at the center, the laser power which can be used for the purpose is only 50\% of the total available power.

If we desire to illuminate uniformly a disk of radius $r_0$ with a Gaussian beam we must pass the beam through an absorption filter which reduces the intensity everywhere in the beam to that at the edge. The intensity transmittance $\mathcal{T}$ of the filter should be

$$\mathcal{T} = \exp\left[-\frac{2(r^2 - r_0^2)}{w^2}\right] \quad \text{for} \quad r \leq r_0$$

$$= 0 \quad \text{otherwise}.$$  \hfill (7.49)$$

The laser power $P_i$ which can be used for illumination is now given by the product of $R(r)$ [from Eq. (7.44)], the transmittance $\mathcal{T}$, and the area $\pi r_0^2$:

$$P_i = \frac{2P_1}{\pi w^2} (\pi r_0^2) \exp\left[-\frac{2r_0^2}{w^2}\right] = 2P_1\left[\frac{r_0^2}{w^2}\right] \exp\left(-\frac{2r_0^2}{w^2}\right).$$  \hfill (7.49)$$

The maximum value of $P_i$ can be found by differentiating Eq. (7.49) with respect to $r_0$. That value is

$$P_i,\max = P_1/e = 0.37P_1$$

when $r_0^2 = w^2/2$.

7.4 Division and Attenuation of the Laser Beam

To form a hologram, the output power from a laser must be divided into subject illumination and a reference wave (see Fig. 7.1). If the subject requires illumination from more than one angle (to eliminate shadows), several divisions of the laser beam may be necessary. Amplitude division (Fig. 7.1b) rather than wavefront division is generally preferred since it results in more uniform illumination intensity and requires less beam expansion. The beam may be divided either before or after it is expanded. Two advantages of division before expansion are (1) the aperture of the beam splitter can be small, and (2) diffraction effects related to dust and surface defects on the beam splitter can be spatially filtered with the lens-pinhole combination used for the subsequent beam expansion. An advantage of division after expansion is that only one set of beam expanding optical components is needed.

Partial reflection from a partially silvered (or dielectric-coated) mirror is the simplest method for dividing a laser beam into two parts. The unsilvered surface is antireflection-coated to minimize the interference between reflections from the two surfaces. A variation of the partially reflecting mirror is a cube-type beam splitter made of two right-angle glass prisms cemented together at their hypotenuse surfaces. One of the two hypotenuse
surfaces is partially silvered before the cementing. Cube-type beam splitters are advantageous because they are less likely to be damaged in cleaning and produce no lateral translation of the transmitted beam. On the other hand, the two output beams are fixed approximately perpendicular to each other.

The beams emerging from the beam splitter should have an intensity ratio which optimizes the fringe visibility at the hologram plane. Of course the optimum ratio varies with the circumstances of hologram formation. Therefore a variable-ratio lossless beam splitter is highly desirable, especially in applications of holography to real-time interferometry where the beam ratio must be changed to a new optimum during the reconstruction process.

Perhaps the simplest variable-ratio beam splitter is a partially silvered mirror whose reflectivity varies continuously in an annular ring. A laser beam incident on a selected portion of the ring can thus be divided into two beams having the desired intensity ratio. An obvious undesirable feature of this device is its gradient of reflectivity resulting in nonuniform division across the beam.

A superior variable-ratio beam splitter is a polarization beam splitter made of doubly refracting material such as calcite or crystal quartz. An example [7.12] of this type of beam splitter using a Foster-Seely prism is shown in Fig. 7.16. Half-wave plate \(A\) changes the direction of polarization of a linearly polarized laser beam to any arbitrary direction before the laser beam enters the Foster-Seely prism \(B\). The prism is cut and silvered in such a way that the extraordinary ray, polarized perpendicular to the paper, passes through the prism unhindered; the ordinary ray polarized parallel to the paper, however, suffers a refraction and a reflection leaving the prism at \(90^\circ\) to the extraordinary ray. By rotating the half-wave plate \(A\), we can therefore obtain any desired intensity ratio between the two output beams. A fixed half-wave plate \(C\) changes the direction of polarization of the emerging ordinary ray so that both output beams are polarized perpendicular to the paper as desired in most cases. Of course, compensators or electro-optical crystals may be used instead of the two half-wave plates. These allow use of the beam splitter at any wavelength. Electro-optical crystals offer the possibility of an electrically alterable intensity ratio.

When only one of two output beams is used and the other discarded, a beam splitter also acts as an attenuator. It is generally more economical, however, to use a neutral density filter as a fixed attenuator and a pair of polarizing prisms or sheets as a variable-ratio attenuator.

There are two types of neutral density filters and both are commercially available. In one type the unwanted energy is absorbed by the filter and in the other type unwanted energy is reflected. The absorption type of neutral density filter is superior at low power levels, say below \(1 \text{ W/cm}^2\). Light multiply reflected between the surfaces of the filter is absorbed, thus avoiding unwanted interference. At higher levels, however, an absorption filter can be damaged by the heat generated from the absorbed energy.

### 7.5 Mechanical Stability in Hologram Formation

The highest spatial frequency in the interference fringe pattern to be recorded determines the degree to which the position of the hologram-recording medium must remain fixed relative to the pattern. We assume, for the time being, that the fringe pattern is stationary. The smallest fringe spacing encountered when forming most holograms is approximately the wavelength of the forming light. We must therefore ensure that the hologram recording medium does not move more than a small fraction of a wavelength during its exposure. Under laboratory environment it is not difficult to achieve this relative stability if the holder for the recording medium is carefully designed. On the other hand, the task of maintaining a stable fringe pattern is a more formidable one and is a familiar problem in precision optical interferometry.

To ensure a steady fringe pattern, all optical components as well as the subject, the light source, and the recording medium must be carefully and securely mounted on a massive optical bench or table. Tables made of granite, concrete, steel, or aluminum weighing tons are not uncommon. The mass is used to make the mechanical resonance frequency of the table (about \(1 \text{ Hz}\)) much lower than the frequencies of the building vibration. Isolation of the optical table from the building vibration is adequately

![Fig. 7.16. Polarization beam splitter using a Foster-Seely prism.](image-url)
achieved if the table is supported in a sand bath or by a pneumatic suspension system. For example, one inexpensive suspension system consists of many rubber inner tubes filled with air at a low pressure. An air bearing is another, perhaps more luxurious, method.

Airborne disturbances, both acoustic and thermal, can also cause an unstable interference fringe pattern and are generally more difficult to eliminate than mechanical vibrations. Air-conditioning units and ventilating fans are best turned off during the exposure of a hologram. Sources of heat such as lamps and some electrical equipment should be kept away from the area of the optical table immediately before and during the exposure. The effects of airborne disturbances are minimized if all optical path lengths, especially those between the beam splitter and the hologram recording medium, are made as short as possible.

The problems of both mechanical and airborne disturbances are lessened, of course, if the exposure time can be reduced. It is therefore important to fully utilize the output power of a laser by using the optimum beam intensity ratio and by minimizing the number of optical components and their associated optical loss. An electronic servo-control technique can be devised to stabilize the interference fringe pattern [7.13, 7.14]. This technique is applicable, however, only when the mechanical or airborne disturbance of the fringes can be compensated by a change in the path length of reference beam or signal beam or by a change in the oscillation frequency of the laser.

### 7.6 Light Sources for Hologram Reconstruction

If an original wavefront recorded in a hologram is to be reconstructed with minimum aberration, the direction and radius of curvature of the reconstructing beam must duplicate those of the reference beam used to form the hologram. The resolution of the image produced by the wave reconstructed under this condition is limited only by the size of the hologram and the coherence properties of the reconstructing light source.

To obtain the highest achievable resolution in the image, the light source for wavefront reconstruction should be as coherent as the light source for hologram formation. However, in many cases, this is neither necessary nor desirable. A powerful multimode laser can often be used successfully to illuminate a hologram made with a single-frequency laser and, moreover, to produce a brighter image than the single-mode laser without appreciable loss in image quality. In some cases an arc lamp or an incandescent lamp may be used in the reconstruction step with a not intolerable loss in image resolution. In this section we shall consider how the degree of coherence of the illuminating light source affects the image resolution.

We have seen that the degree of spatial coherence of a nonlaser light source depends on its radial extent according to Eq. (7.4). When the source is a point so that in Eq. (7.4) the radius \( r_0 = 0 \), the degree of spatial coherence \( |\mu_1| \) is a maximum of unity. Suppose, for the purposes of analysis, that we now have a more realizable source of finite extent and thus of lesser spatial coherence, but one which nonetheless has a very high degree of temporal coherence. We consider the extended source to illuminate a hologram which has been formed with a perfect reference point source and inquire as to the effect on the resulting image of this lesser degree of spatial coherence. The effect may be determined for the case of a small source by examining the expressions for the virtual image coordinates in Eq. (3.27),

\[
x_{3v} = \frac{x_0^2 r_z + \mu x_0 z_0 - \mu x_0 z_1}{z_1 + z_2 z_0 - \mu z_1}, \quad z_{3v} = \frac{z_0 z_2 z_1}{z_1 + z_2 - \mu z_0 z_1} (7.50)
\]

where the scaling factor \( m \) in Eq. (3.27) has been set equal to unity, \((x_0, z_0)\) are the coordinates of the illuminating source, \( \mu = \lambda_0 / \lambda_1 \), and the remaining parameters are defined in Fig. 7.17. (Note: The results of Chapter 3 used in the remainder of this section are valid for plane holograms.) Figure 7.17 illustrates the reconstruction situation. Illumination of the hologram \( H \) by the original reference point source \( R \), located a distance \( z_r \) from \( H \), yields the subject image \( P \) in its original location \((x_1, z_1)\). Consider now a second illuminating point source \( R' \) also located a distance \( z_o = z_r \) from the hologram and radiating light of the same wavelength as \( R \) \( |\mu = 1 \) in

![Fig. 7.17. Illustrating the effect of an illuminating source of finite extent on an image point \( P \).](image-url)
Thus the degree of spatial coherence of the illuminating source, determined by the source diameter \( \Delta r \), fixes the minimum resolvable image dimension \( \Delta s \).

As an example, suppose a mercury arc lamp, whose radiation is passed through a narrow-band wavelength filter and whose diameter is 0.3 mm, is used as the illuminating source. If both source and image are at distance \( z_r \) from the hologram, the smallest detail we can expect to resolve in the image is about \( \Delta s = 0.3 \text{ mm} \). At \( z_1 = z_r = 1 \text{ m} \) the resolution approaches that of the eye situated at the hologram and is quite sufficient for many visual displays. Equation \( (7.52) \) implies that the closer the image is to the hologram, the less is the demand for spatial coherence in the illuminating source. If the central plane of the image lies in the hologram plane, the requirement on spatial coherence is least.

We can also employ Eq. \( (3.35) \) to demonstrate the effect on the image when a very small thermal source of finite bandwidth \( \Delta \lambda \) is used to illuminate a hologram formed originally with a reference point source of perfect temporal coherence. If we assume the illuminating and reference beams to be essentially plane waves coming from a single, distant point source at \( (\xi_r, \eta_r, \zeta_r) \), then the virtual image coordinates of Eq. \( (3.35) \) can be written as

\[
x_{sv} - x_1 = \frac{x_r}{\xi_r} \cdot \frac{1 - \mu}{1} = \frac{x_r}{\xi_r} \cdot \frac{\lambda_1 - \lambda_2}{\lambda_2}, \quad z_{sv} = \frac{z_1}{\mu}.
\]

Defining \( x_{sv} - x_1 = \Delta s \), \( x_r/\xi_r = \theta_1 \), and \( \lambda_1 - \lambda_2 = \Delta \lambda \), we have

\[
\Delta s = \theta_1 z_1 (\Delta \lambda/\lambda), \quad z_{sv} - z_1 = (\Delta \lambda/\lambda)z_1.
\]

Suppose we consider the subject to be on axis so that the small change in \( z \) position does not contribute to loss of resolution in a plane transverse to the viewing direction. Since \( x_1 \) corresponds to the original subject position and also to the image position when the original reference wavelength \( \lambda_1 \) illuminates the hologram, \( \Delta s \) represents a displacement from \( x_1 \) due to illumination with a source whose wavelength \( \lambda_2 = \lambda_1 - \Delta \lambda \). Thus a point image would spread to the dimension \( \Delta s \) given in Eq. \( (7.53) \) if the hologram were illuminated by a very small nonmonochromatic source of bandwidth \( \Delta \lambda \).

If we once again consider a high-pressure mercury arc lamp with a bandwidth of 50 \( \text{Å} \) centered at 5461 \( \text{Å} \) and assign some typical values \( \theta_1 = 15^\circ \) and \( z_1 = 100 \text{ mm} \) in Eq. \( (7.53) \), the point image spread \( \Delta s \) becomes 0.24 mm. Note that the bandwidth of the source, representing a relatively low degree of coherence, places a limit on the depth of the image over which one can display small detail and a limit on the offset angle \( \theta_1 \) of the reference beam. If the central plane of a 3D image lies in the hologram plane \( (z_1 = 0) \) and if the angle \( \theta_1 \) is small, white light illumination is possible. A volume of the image surrounding the hologram plate will then appear achromatic, i.e., color dispersion will be negligible (see Section 17.4).

Source size and bandwidth are primary determinants of the smallest resolvable dimension in the image when nonlaser sources are employed to illuminate a hologram, but this is often not the case when a laser source is used. A multilongitudinal-mode He–Ne laser can have sufficient temporal coherence that when it illuminates a hologram formed on a standard 4 \( \times \) 5 inch high-resolution photographic plate, the image resolution is essentially diffraction limited (limited only by the angular aperture of the hologram). Suppose we consider the formation of the real image of a point as indicated in Fig. 7.18. We compute the maximum difference in optical path \( \Delta L \) taken by rays passing from the source (at infinity) to the image via the hologram. We assume that the degree of temporal coherence of the source places no limit on image resolution provided \( \Delta L \) does not exceed the source coherence length \( \Delta L_{\text{h}} \). If we write \( \Delta L \) in terms of the extent of the hologram \( h \), then we can obtain the maximum value \( h_m \) allowed under the condition that the source temporal coherence plays no significant part in determining image resolution. From Fig. 7.18 we have

\[
\Delta L = h \sin \theta_1 + (h^2 + z_1^2)^{1/2} - z_1.
\]

Common values for the parameters \( \theta_1 \) and \( z_1 \) might be \( \theta_1 = 30^\circ \) and \( z_1 = 2h \). These yield \( \Delta L = 0.736h \). Equating \( \Delta L \) and \( \Delta L_{\text{h}} \), we find that

\[
h_m = \Delta L_{\text{h}}/0.736.
\]
7. Light Sources and Optical Technique

A 1-m-long He-Ne laser has a coherence length $\Delta L_{\text{H}} \approx 10$ cm permitting $h_m$ to be somewhat over 13 cm. (Note that a multilongitudinal-mode argon laser with $\Delta L_{\text{H}} \approx 2$ cm restricts $h_m$ to $\frac{1}{4}$ this value.)

7.7 Simple Holographic Technique

In the preceding sections of this chapter we have specified properties of light sources and optical components which are desirable for general application in holography. The requirements placed on the components, of course, vary with the application. In this section we present a simple example of holographic technique which can form good holograms and reconstruct wavefronts yielding high-quality 3D images. Only a minimum number of inexpensive optical components are employed.

The optical arrangement for forming the hologram is shown as a schematic drawing in Fig. 7.19 and as an actual photograph in Fig. 7.20. All optical components are securely fastened to a massive steel optical table which is supported on a number of inflated airplane-tire inner tubes. In the case of the beam splitter and mirrors, tacky wax is used to attach these components to a metal rod which in turn is locked into the collar of a stand and the latter bolted to the table. We show as a light source a helium-neon laser of sufficient length to support several longitudinal modes. It has an output power of 1.8 mW at 6328 Å. A glass plate, approximately 5 mm thick, serves as the beam splitter $B$ and reflects about 5% of the incident power from each (front and back) surface. The front surface reflection is used for the reference beam, and the remaining 90% which passes through the plate is used to illuminate the subject. Two 20x microscope objective lenses (8-mm focal length and 0.5 numerical aperture) expand the beam widths from less than 1 mm to nearly 7 cm near the subject $S$ and

---

**Fig. 7.18.** Laser illumination of a hologram.

**Fig. 7.19.** A simple practical arrangement for forming a hologram.

**Fig. 7.20.** Photograph of the actual arrangement corresponding to Fig. 7.19.
hologram $H$. A pair of front surface mirrors $M_1$ and $M_2$ are so placed that the mean optical path lengths $BM_1SH$ and $BM_2H$ are approximately equal.

To indicate the quality of the results obtainable with the configuration of Fig. 7.20, we have formed a hologram of a set of three ceramic block letters, 2 cm high, arranged one behind the other. The average intensity imparted by this subject to the hologram plane is $0.2 \mu W/cm^2$ while the intensity at the hologram plane due to the reference beam is $1 \mu W/cm^2$. When a Kodak 649 F photographic plate is used to record the hologram, the exposure time is approximately 50 sec. After exposure the plate is developed in photographic developer intended for high-resolution emulsions, such as Kodak D-19 or HRP developer. The remaining processing steps (fixing, clearing, and washing) are normal photographic procedures which are set down in detail in Chapter 10. Figure 7.21 is a photograph of the hologram itself. Note the coarse diffraction patterns characteristic of scattering from dust particles. These are probably located on the microscope objective lenses used to expand the beam (the fringes relevant to the holographic process are, of course, too small to be observed without magnification). Dust particle fringes can, if desired, be eliminated by spatial frequency filtering with a 10-\mu m pinhole placed near the focus of the microscope objective (see Section 7.3). However, their presence does little to degrade the image produced by the hologram.

![Hologram](image1)

**Fig. 7.21.** Photograph of hologram formed with the arrangement of Fig. 7.20.

A simple arrangement for reconstructing the original subject wavefront is shown in Fig. 7.22. A laser and a beam-expanding microscope objective duplicate the original reference wavefront at the hologram. The wavefront reconstructed in this manner generates for the observer a virtual image whose size and position, relative to the hologram, are identical to those of the original subject. Figure 7.23 is a photograph of the image.

![Hologram](image2)

**Fig. 7.23.** Image produced by hologram of Fig. 7.21.

**REFERENCES**


Chapter 8

ANALYSIS OF PLANE HOLOGRAMS

The fringes recorded by small, in-line holograms with nondiffuse light are coarsely spaced relative to the thickness of photographic emulsion. A ray in the wavefront illuminating such a hologram interacts with only one recorded fringe before passing through the hologram. Consequently the hologram response is closely approximated by that of a plane diffraction grating— with focusing properties. Gabor analyzed these properties by considering the hologram to be strictly two dimensional. Predictions of the analysis using the two-dimensional model were in good agreement with the experimental observations.

The off-axis technique, introduced by Leith and Upatnieks, led to holograms whose fringe frequency exceeded that of the in-line hologram by a term proportional to the angle between subject and reference beams [see Eq. (3.15)]. A typical value of the fringe spacing is obtained by considering the interference of two plane waves. Equation (1.10) \( 2d \sin \theta \lambda \), relates the half-angle \( \theta \) between beam directions and the wavelength \( \lambda \) to the fringe spacing \( d \). For \( \theta = 15^\circ \) and \( \lambda = 0.5 \mu m \) (green light), \( d = 1 \mu m \). Photographic emulsions used to record off-axis holograms are often 15 \( \mu m \) in thickness, and the hologram formed therein can no longer be realistically regarded as two dimensional. Nevertheless, a two-dimensional analysis, restated in terms of communication theory, was extended to the off-axis case by Leith and Upatnieks [8.1, 8.2]. Despite the discrepancy between the facts of photographic emulsion and the assumptions of the planar model, it provided a useful framework for the further development of holography. However its application to holograms which might be better described as volume diffraction gratings leads to partially fulfilled predictions and leaves many of the observed properties of holograms unexplained.