

8

Coherence and Fringe Localization

T. D. Milster and N. A. Beaudry

A simple plane wave is sometimes used to describe a laser beam's electromagnetic disturbance, such as

$$U(z, t) = A \exp[j(kz - \omega t)] , \quad (8.1)$$

where the disturbance is known over all time and at every location in space. In Eq. (8.1), $k = 2\pi/\lambda$ is the propagation constant, ω is the radian frequency and t is time. Many theoretical formulations in physical optics assume this ideal form of a monochromatic electromagnetic wave. In fact, much insight is gained and many phenomena are adequately described using ideal waves. While an ideal wave can be a close approximation to a very good laser beam, it is *never* a completely true representation of nature. Real wave fields exhibit phases, frequencies and wavelengths that vary randomly with respect to time. The random nature may be a very small fraction of the average frequency, as in a high-quality laser beam, or it may be much more significant, like the field emitted from a tungsten light bulb. Due to this random nature, the statistical description of light plays an extremely important role in determining the behavior of many optical systems.

While it is clear that some form of statistical description is necessary, in most cases a complete statistical description is not required, and a second-order model is sufficient [5.1]. The second-order average of the statistical properties of an optical field is referred to as a *coherence function*. The field of study known as *coherence theory* is simply the study of the coherence function. A large amount of information is derived from these second-order statistics, including the ability of the optical field to create interference patterns.

As a simple example, consider the interference pattern created when a screen is placed in a complex-valued optical field $U(\mathbf{r}, t)$, as shown in Fig. 8.1. The screen has two small pinholes that transmit light, but it is otherwise opaque. The pinholes are spaced by distance d . Cartesian vectors s_1 and s_2 describe the positions and distances of the pinholes from a point on the source. Pinholes P1 and P2 are small enough so that they act as point sources on transmission, and an interference pattern is produced on the observation plane. The interference pattern consists of bright and dark fringes, where the irradiance of the total field is the measurable quantity. The optical path length s'_1 between P1 and (x_0, y_0) is vt_1 , where v is wave velocity in the observation space and t_1 is the time required for the wave to travel from P1 to (x_0, y_0) . Similarly, the optical path length s'_2 between P2 and (x_0, y_0) is vt_2 . The optical path difference OPD_0 is the difference in optical path lengths, or $OPD_0 = v(t_1 - t_2)$. If the source is point-like and limited to a small band of wavelengths around $\bar{\lambda}$, a bright fringe is produced where $OPD_0 = m\bar{\lambda}$ and m is an integer. A dark fringe is produced where $OPD_0 = (m + 1/2)\bar{\lambda}$.

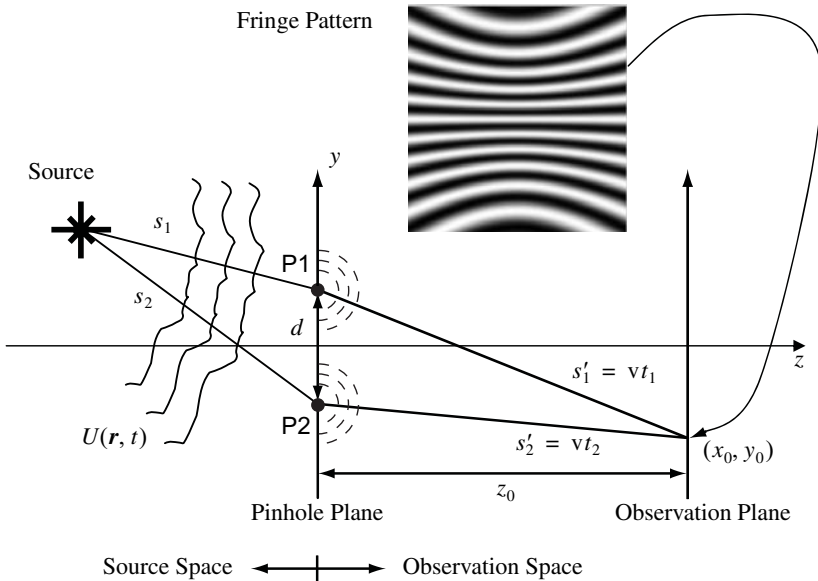


Fig. 8.1. **General geometry for coherence analysis.** A fringe pattern is produced when a light source illuminates two pinholes in an opaque screen. Properties of the fringe pattern depend on the type and location of the source.

Characteristics of the fringe pattern on the observation plane depend on the type of source used to create field $U(\mathbf{r}, t)$. The most dramatic characteristic of the

fringe pattern that changes as a result of the type of source is the fringe visibility.¹ The fringe visibility is measured locally, and it is given by

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \quad (8.2)$$

where I_{max} and I_{min} are the maximum and minimum irradiances of the local fringe modulation, as shown in irradiance profiles along $(0, y_0)$ in Fig. 8.2. A high-visibility fringe pattern has $V \approx 1$, where I_{min} is much smaller than I_{max} . Fringes are clearly observed for $V > 0.5$, while it is difficult to discern the bright and dark portions of a fringe for $V < 0.2$. Note that $V = 0$ does not imply zero local irradiance, but it does imply that the irradiance in the region is void of modulation along y_0 .

If the source of the optical field $U(\mathbf{r}, t)$ is a high quality laser beam, the two pinholes in Fig. 8.1 produce high-visibility fringes along y_0 . The analysis of this problem is identical to the analysis of the fields radiating from two coherent point sources, where a hyperboloidal fringe field is generated at the observation plane. If the laser beam illuminates the pinholes symmetrically (lengths $s_1 = s_2$ in Fig. 8.1), the fringes are hyperboloids that simplify to straight, equally spaced fringes along the central region of y_0 when $z_0 \gg d$, as shown in Fig. 8.2(A).

If the source is extended but nearly monochromatic, like light from a highly filtered tungsten light bulb, the fringe pattern exhibits a significant reduction in visibility, as shown in Fig 8.2(B). The visibility is inversely proportional to the source size. That is, as the source size increases, visibility reduces. Notice that fringe visibility is not a function of y_0 . The behavior of visibility reduction as a function of source size is a result of the effect known as *spatial coherence*.

If the source used to create field $U(\mathbf{r}, t)$ is a centered polychromatic point source, like a distant arc lamp, fringe visibility varies as a function of y_0 as shown in Fig. 8.2(C). Notice that fringe visibility is higher near $y_0 = 0$ and decreases as y_0 increases. Behavior of fringe visibility as a function of the wavelength distribution is the result of the effect known as *temporal coherence*.

In this chapter, the characteristics of both temporal and spatial coherence are explored. Also, derivations for basic coherence formulas are provided, which usually result in simple Fourier transform relationships. Scalar representations of the wave field are used, in order to simplify the discussion. No polarization-dependent analysis is provided. The medium in which the fringes are observed is linear, stationary, homogeneous, nondispersive and isotropic.

The chapter is divided into three sections. Section 8.1 is an introduction to coherence theory with a simplified development. Section 8.2 discusses the topic of *fringe localization*, which concerns the spatial location of high-visibility fringes

1. Fringe visibility is also called fringe *contrast* in some reference material.

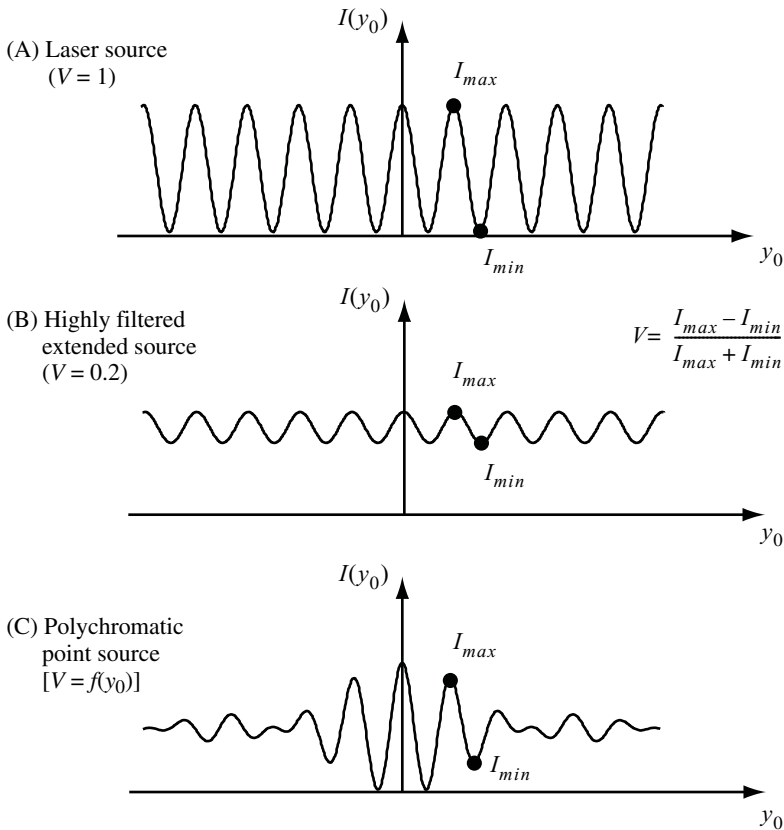


Fig. 8.2. **Various characteristics of fringe visibility.** A fringe pattern is produced when a light source illuminates two pinholes in an opaque screen. Properties of the fringe pattern depend on the type and location of the source.

in different interferometers. Section 8.3 extends the mathematical development of coherence theory. For further reading, suggested references are [5.2,5.3, and 5.4].

8.1 Basic coherence

8.1.1 The mutual coherence function

Mathematically, the instantaneous electric field arriving at some position (x_0, y_0) in the observation plane is given by

$$U_{\text{obs}}(\mathbf{r}_1, \mathbf{r}_2, t, t_1, t_2) = \sqrt{K_1}U(\mathbf{r}_1, t + t_1) + \sqrt{K_2}U(\mathbf{r}_2, t + t_2) , \quad (8.3)$$

where K_1 and K_2 are constants that describe the fraction of the energy transmitted by each pinhole measured at the observation plane, Cartesian position vectors \mathbf{r}_1 and \mathbf{r}_2 correspond to the pinhole locations, and t_1 and t_2 are transit times required for the optical wave to travel from the pinholes to the observation point. Irradiance of the field U_{total} on the observation screen is

$$\begin{aligned} I(\mathbf{r}_1, \mathbf{r}_2 ; t_1, t_2) &= \frac{1}{2}cn\epsilon_0 \langle |U_{\text{obs}}(\mathbf{r}_1, \mathbf{r}_2, t, t_1, t_2)|^2 \rangle_t \\ &= \frac{1}{2}cn\epsilon_0 \{ K_1 \langle |U(\mathbf{r}_1, t + t_1)|^2 \rangle_t + K_2 \langle |U(\mathbf{r}_2, t + t_2)|^2 \rangle_t \\ &\quad + 2\sqrt{K_1 K_2} \text{Re}[\langle U(\mathbf{r}_1, t + t_1)U^*(\mathbf{r}_2, t + t_2) \rangle_t] \} \end{aligned} \quad (8.4)$$

where $\langle \rangle_t$ denotes a time average, c is the speed of light in vacuum (approximately 3×10^8 ms⁻¹), n is the medium's refractive index and ϵ_0 is the permittivity of free space. The subscript t that represents a time average is suppressed throughout the remainder of the chapter and plain angle brackets $\langle \rangle$ imply a time average. Notice that the first two terms in Eq. (8.4) are simply values of irradiance at the observation point from each pinhole separately. The third term modifies the total irradiance and is called the *cross term* or *interference term*. If we assume that the constants K_1 and K_2 do not vary appreciably over the observation region, it is the cross term that, when minimum, determines I_{min} . When at maximum, the cross term determines I_{max} . Therefore, the cross term is a primary factor in determining the fringe visibility. Notice that the cross term can take on either positive or negative values.

The primary quantitative expression in the cross term of Eq. (8.4) is the *mutual coherence function*. Expressed in its most general scalar form, the mutual coherence function is given by

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2) = \langle U(\mathbf{r}_1, t + t_1)U^*(\mathbf{r}_2, t + t_2) \rangle . \quad (8.5)$$

The mutual coherence function is a statistical property that, expressed in words, is the temporal correlation of the electric field at two positions in space with respect to time.

Consider the case of *stationary fields*, which are fields that do not change their properties by a shift in the time origin.[5.5, 5.6] Stationary fields do not include any transient response that occurs when the source is turned on, and it is required that the average power of the source is not fluctuating on time scales comparable to the temporal period of the optical field or the measurement time. In this case, statistics of the optical field only depend on a time difference. That is, $t + t_1$ and

$t + t_2$ in Eq. (8.5) are replaced by $t + \tau$ and t , respectively. The mutual coherence function becomes

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle U(\mathbf{r}_1, t + \tau)U^*(\mathbf{r}_2, t) \rangle , \quad (8.6)$$

which is hereafter abbreviated by $\Gamma_{12}(\tau)$. Note that

$$\Gamma_{11}(\tau) = \langle U(\mathbf{r}_1, t + \tau)U^*(\mathbf{r}_1, t) \rangle , \quad (8.7)$$

and similarly for $\Gamma_{22}(\tau)$. The irradiance is given by

$$\begin{aligned} I(\mathbf{r}_1, \mathbf{r}_2 ; \tau) &= \frac{1}{2}cn\epsilon_0[K_1\langle |U(\mathbf{r}_1, t)|^2 \rangle + K_2\langle |U(\mathbf{r}_2, t)|^2 \rangle] \\ &\quad + cn\epsilon_0\sqrt{K_1K_2}\text{Re}[\langle U(\mathbf{r}_1, t + \tau)U^*(\mathbf{r}_2, t) \rangle] \\ &= I_1 + I_2 + cn\epsilon_0\sqrt{K_1K_2}\text{Re}[\Gamma_{12}(\tau)] , \end{aligned} \quad (8.8)$$

where I_1 and I_2 are the irradiance values at the observation point from the separate pinholes. In general, I_1 and I_2 are functions of \mathbf{r} .

The time difference $\tau = t_1 - t_2$ corresponds to the *optical path difference* $\text{OPD}_0 = v\tau$ in the observation space of the pinhole interferometer. In Fig. 8.1, OPD_0 is a function of (x_o, y_o) . The interferometer in Fig. 8.1 is called a Young's double pinhole interferometer (YDPI). It is used throughout this chapter to refine concepts of coherence.

8.1.2 The two-wavelength point source

Now consider a point source that contains two wavelengths. An implementation of the Fig. 8.1 YDPI two-pinhole screen is shown in Fig. 8.3 with an on-axis ($s_1 = s_2$) point source. The source emits spherical waves with wavelengths λ_a and λ_b . Since the point source is exactly halfway between the two pinholes, light emitted from the first pinhole is exactly in phase with light emitted from the second pinhole. Pinholes are separated by distance d with $z_0 \gg d$. The large separation between the pinhole plane and the observation plane, along with restricting the observation to a region close to the z axis, implies that the interference pattern from each wavelength contains straight-line, equally spaced fringes, as described in Chapter 4. The electric field at the observation point from each pinhole and each wavelength has the form

$$\begin{aligned} U_{\text{obs}}(s', t) &= a_0\sqrt{K}\exp[j(ks' - \omega t + \phi)] \\ &= A\exp[j(ks' - \omega t + \phi)] , \end{aligned} \quad (8.9)$$

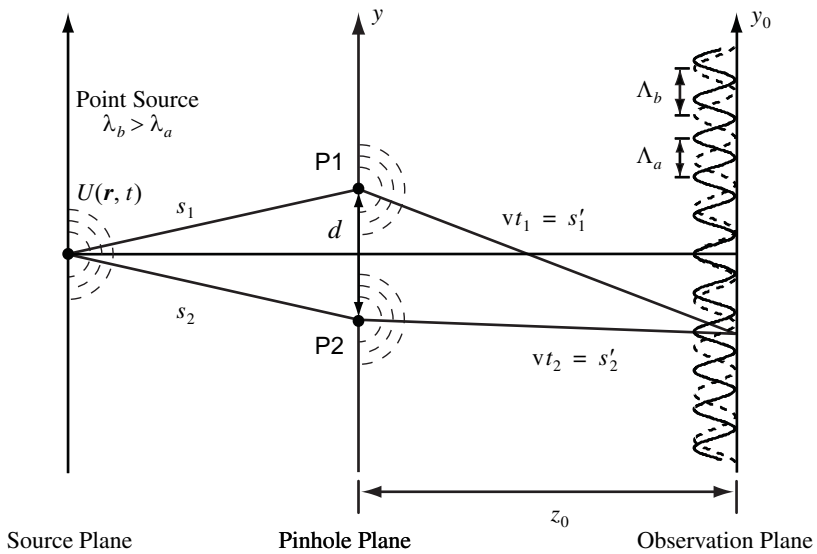


Fig. 8.3. An on-axis dual wavelength source in a YDPI, where $z_0 \gg d$ and the range of y_0 in the observation plane is small. (The drawing is not to scale.) The source produces two independent straight-line cosine fringe patterns in the observation plane with different periods Λ_a and Λ_b . Irradiance values of the independent fringe patterns are added to produce the total irradiance.

where a_0 is a constant and k , ω and ϕ depend on the wavelength. For the purposes of this discussion, assume that K is not a function of observation-plane location or wavelength.¹

Light from the source arriving at each pinhole exhibits a complicated modulation due to the combination of the two wavelengths. Media on both sides of the pinhole plane are non-dispersive, so the modulation components exhibit the same velocity v . Therefore, behaviors of waves emitted by P1 and P2 and observed at (x_0, y_0) have fixed temporal relationships. The time-varying real-valued wave components at the observation point are

1. This expression is an approximation of an expanding spherical wave that is valid in the observation region, as long as the pinholes are small and the distance s' is very large relative to the diameter of the pinholes. A more accurate description of the wave field after the pinholes is governed by diffraction, which includes a $\frac{1}{r}$ amplitude factor.

$$\begin{aligned}
U_{1a}(s_1', t) &= A \exp[j(k_a s_1' - \omega_a t + \phi_a)] , \\
U_{1b}(s_1', t) &= A \exp[j(k_b s_1' - \omega_b t + \phi_b)] , \\
U_{2a}(s_1', t, \tau) &= A \exp\{j[k_a s_1' - \omega_a(t + \tau) + \phi_a]\} \quad \text{and} \\
U_{2b}(s_1', t, \tau) &= A \exp\{j[k_b s_1' - \omega_b(t + \tau) + \phi_b]\} ,
\end{aligned} \tag{8.10}$$

where the first two terms result from P1 and the second two terms result from P2. Subscripts a and b refer to the first and second wavelengths, respectively. Equation (8.10) is written in terms of s_1' , which is a function of d , z_0 and (x_0, y_0) .¹ The square magnitude of the addition of components in Eq. (8.10) results in

$$\text{Square magnitude of addition of field components:} \tag{8.11}$$

$$\frac{1}{2} |U_{\text{total}}(s_1', t, \tau)|^2 = A^2 + A^2 \cos(\omega_a \tau) \tag{8.11a}$$

$$+ A^2 + A^2 \cos(\omega_b \tau) \tag{8.11b}$$

$$+ A^2 \cos(k_\Delta s_1' - \omega_\Delta t + \phi_\Delta) \tag{8.11c}$$

$$+ A^2 \cos(k_\Delta s_1' - \omega_\Delta t + \omega_b z + \phi_\Delta) \tag{8.11d}$$

$$+ A^2 \cos(k_\Delta s_1' - \omega_\Delta t + \omega_a \tau + \phi_\Delta) \tag{8.11e}$$

$$+ A^2 \cos[k_\Delta s_1' - \omega_\Delta(t + \tau) + \phi_\Delta] , \tag{8.11f}$$

where terms (8.11a) and (8.11b) simply result from the two wavelengths independently, and terms (8.11c) through (8.11f) are inter-modulation beat terms. The Δ subscripts in Eq. (8.11) imply modulation frequency ω_Δ or modulation wave number k_Δ , where $\omega_\Delta = \omega_a - \omega_b$ and $k_\Delta = k_a - k_b$. The constant β is the difference in the phase terms of the original optical fields, where $\phi_\Delta = \phi_a - \phi_b$. Notice that arguments of the cosines in Eqs. (8.11a) and (8.11b) do not depend on absolute time t . They depend only on wavelength, inasmuch as wavelength is related to optical frequency by $\lambda = 2\pi c/\omega$, and τ , which is a function of y_0 . Conversely, inter-modulation phase terms in Eqs. (8.11c) through (8.11f) are functions of absolute time t .

1. The variable s_2' is not used in Eq. (8.10) because it is replaced in the argument of the exponential by $k_a s_2' = k_a s_1' - \omega_a \tau$, and likewise for wavelength b terms.

In order to find irradiance $I = \frac{1}{2}cn\epsilon_0\langle|U_{\text{total}}|^2\rangle$ in the observation plane, a time average is performed over Eq. (8.11). A general relationship for averaging the sum of variables is

$$\langle C_1 + C_2 + C_3 + \dots \rangle = \langle C_1 \rangle + \langle C_2 \rangle + \langle C_3 \rangle + \dots \quad (8.12)$$

Application of Eq. (8.12) to Eq. (8.11) results in the first two terms of Eqs. (8.11a) and (8.11b) remaining unchanged, because they are not functions of absolute time t . For example, averaging Eq. (8.11a) yields

$$\langle A^2 \rangle + \langle A^2 \cos(\omega_a \tau) \rangle = A^2 + A^2 \cos(\omega_a \tau) \quad (8.13)$$

A different result is obtained when time averaging is performed on the remaining terms in Eq. (8.11). Since the cosine arguments are functions of absolute time, integration over the full temporal period [$n\lambda_a\lambda_b/(c|\lambda_a - \lambda_b|)$] of the modulation yields a zero net result. For example, averaging Eq. (8.11c) yields

$$A^2 \langle \cos(k_\Delta s'_1 - \omega_\Delta t + \phi_\Delta) \rangle = 0 \quad (8.14)$$

The result of time averaging Eq. (8.11) is

$$\begin{aligned} \frac{1}{2} \langle |U_{\text{total}}|^2 \rangle &= A^2 + A^2 \cos(\omega_a \tau) + A^2 + A^2 \cos(\omega_b \tau) \\ &= A^2 [2 + \cos(\omega_a \tau) + \cos(\omega_b \tau)] \end{aligned} \quad (8.15)$$

Equation (8.15) shows that the irradiance pattern in the observation plane is effectively the *sum of irradiance patterns from individual wavelengths*.¹

Physically, the time average represents the integration time of the detector used to measure the energy at observation point y_o . If the integration time $T \gg 2\pi/\omega_\Delta$, inter-modulation and high-frequency terms in Eq. (8.11) average to zero. For example, if $\lambda_a = 550$ nm and $\lambda_b = 570$ nm, $T \gg 5.2 \times 10^{-14}$ seconds in order to satisfy the integration-time requirement. It is usually a safe assumption to approximate the total irradiance by adding individual irradiance patterns produced from individual wavelengths, unless a detector is used for the measurement that has a short time response. For example, cones in the human eye exhibit a time response of around 10^{-2} seconds, and a fast silicon detector exhibits a time response of 10^{-10} seconds. Both of these simple detectors exhibit integration times much longer than 10^{-14} seconds.

1. Even if the integration time is not equal to the modulation period, the statistical average of the time-dependent terms in Eq. (8.11) approaches a zero limit.

In the limit where $\lambda_a \rightarrow \lambda_b$, $\omega_a \approx \omega_b = \omega$, and Eq. (8.11) reduces (before averaging) to

$$\frac{1}{2}|U_{\text{total}}|^2 = 2A^2(1 + \cos\phi_\Delta)(1 + \cos\omega\tau) \quad (8.16)$$

The leading factor $(1 + \cos\phi_\Delta)$ defines coherent addition of the wavelengths based on their phase difference. No phase shift with $\phi_\Delta = 0$ produces a coherent sum, which is equivalent to the result obtained by simply increasing the source amplitude by a factor of two.

If the wavelengths are derived from physically distinct phenomena, like the beams from two laser cavities, the phase ϕ_Δ is random in time. The time average of Eq. (8.16) with a zero-mean ϕ_Δ is

$$\begin{aligned} \frac{1}{2}\langle |U_{\text{total}}|^2 \rangle &= 2A^2(1 + \langle \cos\phi_\Delta \rangle)(1 + \cos\omega\tau) \\ &= 2A^2(1 + \cos\omega\tau) \end{aligned} \quad (8.17)$$

because the time average of $\langle \cos\phi_\Delta \rangle = 0$. Again, the total irradiance pattern of Eq. (8.11) is simply the addition of two irradiance patterns produced separately from the individual wavelengths, which both have equal amplitudes in this case. This argument can be expanded to apply to any two wavelengths in the spectrum of a polychromatic source.

In the remaining sections of this chapter, total observation plane irradiance is derived by summing irradiance patterns resulting from each wavelength of the source. Distance Λ between adjacent fringe maxima corresponds to a change of one wavelength of optical path difference $\text{OPD}_0 = v\tau$. Equation (8.15) is written in these terms as

$$\begin{aligned} \frac{1}{2}\langle |U_{\text{total}}|^2 \rangle &= A^2[2 + \cos(k_a \text{OPD}_0) + \cos(k_b \text{OPD}_0)] \\ &\approx A^2 \left[2 + 2 \cos\left(\frac{2\pi}{\bar{\lambda}} \text{OPD}_0\right) \cos\left(\frac{\pi}{\lambda_{\text{eq}}} \text{OPD}_0\right) \right] \end{aligned} \quad (8.18)$$

where $\lambda_{\text{eq}} = \lambda_a \lambda_b / |\lambda_a - \lambda_b|$ and $\bar{\lambda} = (\lambda_a + \lambda_b) / 2$. Note that the cosine terms on the top right side of Eq. (8.18) have slightly different spatial periods Λ_a and Λ_b for the same range of OPD_0 , due to the different wavelengths. The first cosine term on the bottom right side of Eq. (8.18) exhibits high-frequency fringes with period $\text{OPD}_0 = \bar{\lambda}$. The second cosine term on the bottom right side of Eq. (8.15) exhibits a slow modulation corresponding to λ_{eq} .

The central region of the observation screen at z_0 along $(0, y_0)$, which is important in the development of coherence functions in following sections, is

defined where $z_0 \gg y_0$, with the constraint that $z_0 \gg d$. As shown in Fig. 8.4, $OPD_0 = s'_1 - s'_2 \approx d \sin \theta \approx (-d)y_0/z_0$, where lengths $s'_1 = vt_1$ and $s'_2 = vt_2$ define distances from the observation point to P1 and P2, respectively. Under these conditions, OPD_0 is a linear function of position along y_0 . The period between bright fringes is $\Lambda = z_0 \lambda / d$. The cosine terms in Eq. (8.18) produce straight and equally spaced fringes in the central region for wavelengths λ_a and λ_b with periods Λ_a and Λ_b , respectively, as shown in Fig. 8.3.

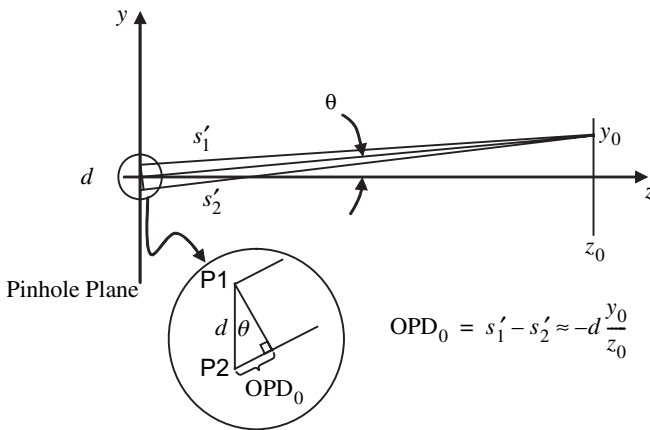


Fig. 8.4. When observation is made with $|y_0|$ small, $d \ll z_0$ and $z_0 \gg \lambda$, the optical path difference in observation space OPD_0 can be approximated by $OPD_0 \approx -dy_0/z_0$. Fringes are straight and equally spaced with period $\lambda z_0/d$.

Figure 8.5 is a gray-scale illustration of fringes in the x_0y_0 observation plane for a two-wavelength source. The observation range is restricted near $y_0 = 0$. Notice that the fringe modulation washes out periodically along y_0 . Figure 8.6 shows line profiles of the component fringes, the combined fringe pattern and the fringe visibility curve as functions of $OPD_0 = -dy_0/z_0$. In the exact center of the pattern where $y_0 = 0$, $\tau = 0$ and $OPD_0 = 0$, the cosine terms are in phase, and $V = 1$. As y_0 increases away from zero, fringes become out of phase, and visibility reduces. Fringes are π out of phase and $V = 0$ when $m\lambda_a = (n + 1/2)\lambda_b = OPD_0$, where m and n are integers. The relationship between m and n at this OPD_0 is

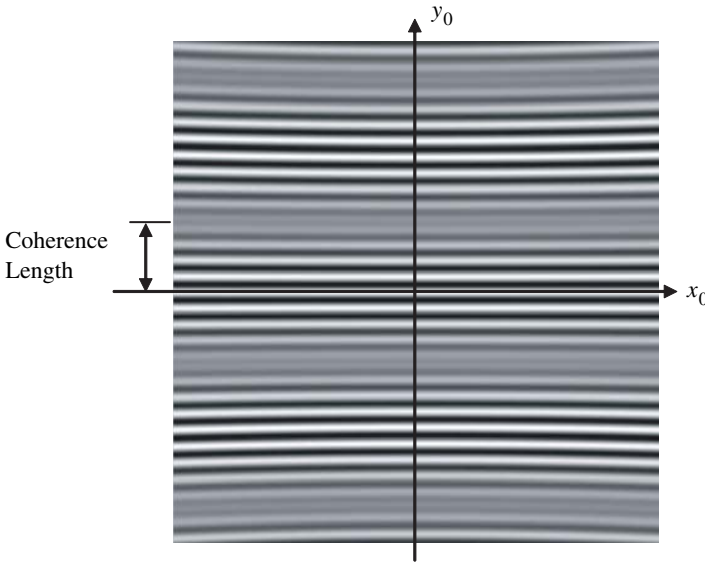


Fig. 8.5. Fringes in the YDPI observation plane for small y_0 with a dual-wavelength on-axis source. Fringes are straight and equally spaced. Regions of low V are easily observed. Notice that low V regions contain significant irradiance, but they do not contain clear modulation. The coherence length l_c is defined as the distance in OPD₀ units from the central fringe with maximum V to the first region of minimum V .

$$\frac{\Delta\lambda}{\bar{\lambda}_b} = \frac{n - m + 1/2}{m} . \tag{8.19}$$

The examples in Figs. 8.5 and 8.6 are shown for $m = 5$ and $n = 4$. Notice that the visibility is a smooth envelope that decays to zero at the point where $y_0 = z_0\lambda_{eq}/d$. The OPD₀ corresponding to the distance between the maximum fringe contrast at $y_0 = 0$ and the first zero of fringe visibility is the *coherence length* l_c . Beyond the point of minimum visibility, fringes become more in phase, until $V = 1$ again. The combined pattern in Fig. 8.6(b) has a periodic variation in visibility, between $V = 0$ and $V = 1$, but the average irradiance is constant. The high-frequency fringes occur with period $\bar{\lambda}$ in OPD₀ and change phase between modulation lobes. The visibility curve in Fig. 8.6(c) is periodic and corresponds to the modulation term in Eq. (8.18). The coherence length is $l_c = \lambda_{eq}/2$, which is approximately $\bar{\lambda}^2/2\Delta\lambda$ in OPD₀ units. As the wavelength distribution of the source widens, the range of good visibility along the observation plane decreases.

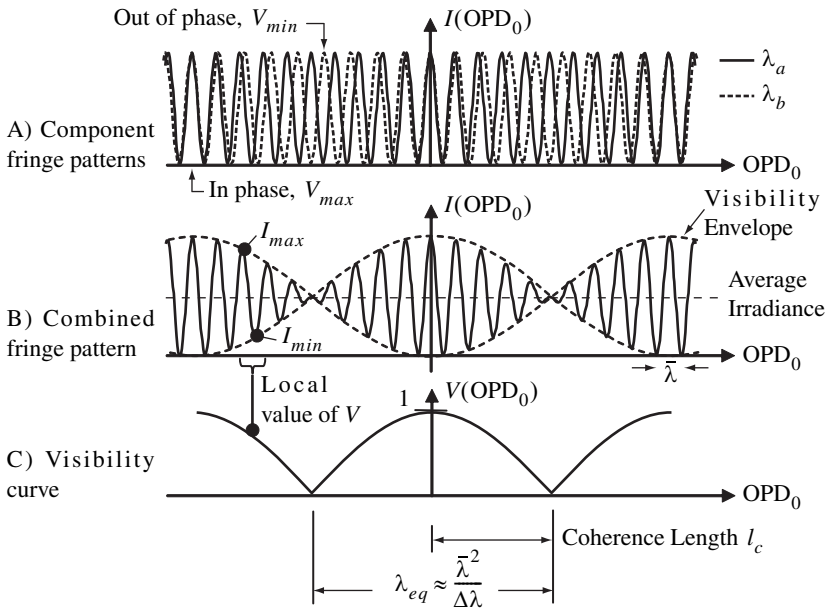


Fig. 8.6. Fringe pattern analysis for a YDPI two-wavelength source. A) Line profiles of the component fringes; B) The total irradiance, which is the linear sum of the component fringe irradiance. Notice that the modulation envelope of this pattern is the visibility function, and the high frequency fringes occur with period $\bar{\lambda}$ in units of OPD_0 . (Vertical scale is 1/2 of A.); and C) The two-wavelength fringe visibility as a function of OPD_0 shows a periodic variation. The coherence length is $l_c = \lambda_{eq}/2 = \bar{\lambda}^2 / 2\Delta\lambda$.

For example, Fig. 8.7 shows the result of adding a third wavelength, and Fig. 8.8 shows the result for five wavelengths.

8.1.3 The power spectrum of the source

The discrete wavelength character of the optical sources considered in Section 8.1.2 is not found in nature. Thermal sources, like tungsten filaments, emit a continuous distribution of wavelengths that vary as a function of the filament temperature. Gas discharge lamps can provide a narrow spectrum of radiation limited by the Lorentzian linewidth of atomic transitions, but separate molecules in the gas emit energy independently at frequencies distributed over the linewidth. Even with laser sources, a continuous description of the wavelength spectrum is required. Stimulated emission, like that found in laser cavities, can exhibit a narrow spectrum of less than one part in 10^{13} , but mechanical instabilities in the

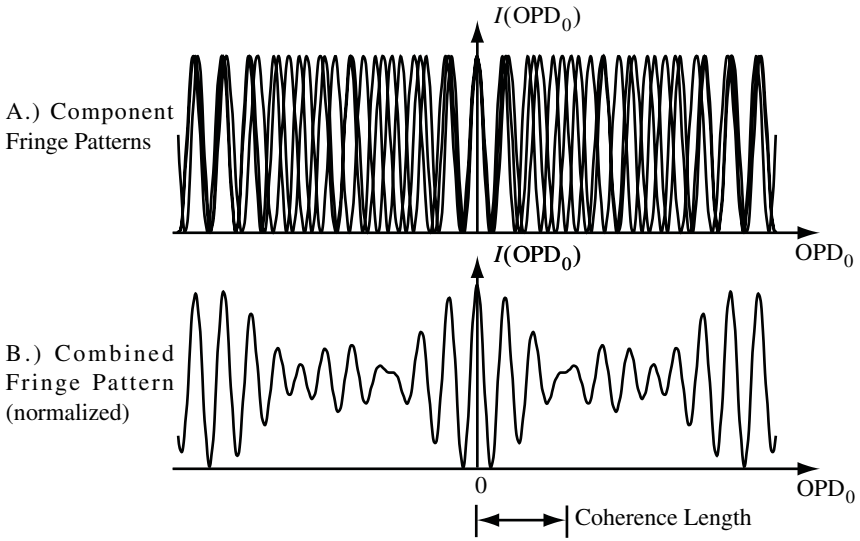


Fig. 8.7. Fringe patterns due to three equally-spaced wavelengths in the source.

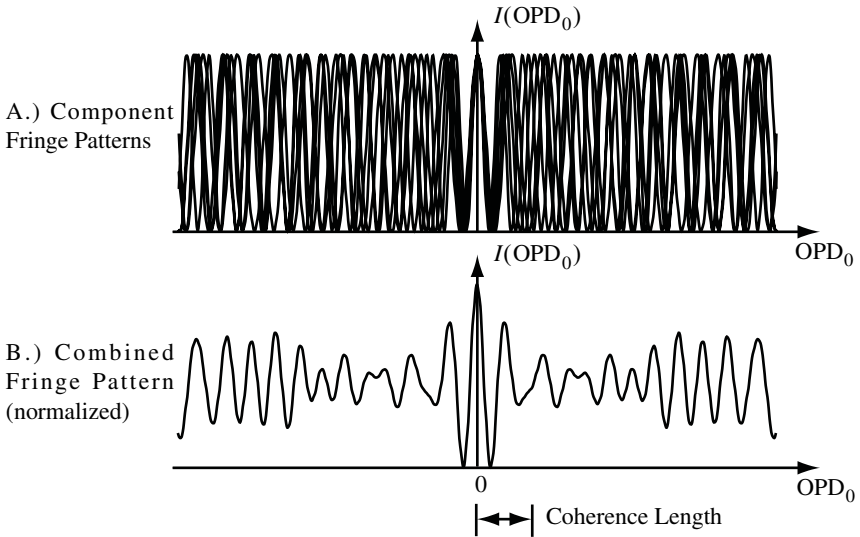


Fig. 8.8. Fringe patterns due to five equally-spaced wavelengths in the source.

laser cavity and other sources of instability result in slight variations of phase $\varphi(\mathbf{r}, t)$ of the output. The phase variations cause tiny modulations of the wavelength, which can influence interference measurements. Fortunately, these wavelength characteristics can be described in a straightforward formalism called the *power spectrum*.

Consider an optical wave that contains a continuous range of frequency components with differential amplitudes given by $\tilde{a}(\mathbf{r}, \nu)$ and phases given by $\tilde{\varphi}(\mathbf{r}, \nu)$, where ν is the temporal frequency of the vibration. The real-valued electric field is

$$E(\mathbf{r}, t) = \int_0^{\infty} \tilde{a}(\mathbf{r}, \nu) \cos[\tilde{\varphi}(\mathbf{r}, \nu) - 2\pi\nu t] d\nu . \quad (8.20)$$

Notice that Eq.(8.20) restricts the temporal frequency of the field to positive frequencies. The *analytic signal* associated with $E(\mathbf{r}, t)$ is¹

$$U(\mathbf{r}, t) = \int_0^{\infty} \tilde{a}(\mathbf{r}, \nu) \exp\{j[\tilde{\varphi}(\mathbf{r}, \nu) - 2\pi\nu t]\} d\nu , \quad (8.21)$$

where

$$E(\mathbf{r}, t) = \text{Re}[U(\mathbf{r}, t)] . \quad (8.22)$$

If real-valued $E(\mathbf{r}, t)$ is also represented by the Fourier transform

$$E(\mathbf{r}, t) = \int_{-\infty}^{\infty} \tilde{U}(\mathbf{r}, \nu) \exp(-j2\pi\nu t) d\nu , \quad (8.23)$$

its Fourier spectrum $\tilde{U}(\mathbf{r}, \nu)$ is Hermitian, so

$$\tilde{U}(\mathbf{r}, -\nu) = \tilde{U}^*(\mathbf{r}, \nu) . \quad (8.24)$$

After expanding Eq. (8.23) into the form of Eq. (8.20) and using Eq. (8.24), the spectral components are found to be related by

1. This development closely follows that of Born and Wolf, Section 10.2.

$$\begin{aligned}\tilde{U}(\mathbf{r}, \nu) &= \frac{1}{2}\tilde{a}(\mathbf{r}, \nu)\exp[j\tilde{\phi}(\mathbf{r}, \nu)] \quad \text{for } \nu \geq 0 \\ &= \frac{1}{2}\tilde{a}(\mathbf{r}, \nu)\exp[-j\tilde{\phi}(\mathbf{r}, \nu)] \quad \text{for } \nu < 0 \quad .\end{aligned}\quad (8.25)$$

The analytic signal in Eq. (8.21) can also be expressed in terms of the Fourier spectrum by

$$U(\mathbf{r}, t) = \int_0^{\infty} \tilde{U}(\mathbf{r}, \nu)\exp(-j2\pi\nu t) d\nu \quad , \quad (8.26)$$

which is the representation of the complex optical field used throughout this chapter.

If measurement of the optical field occurs around time t_0 over time span T , the sampled field amplitude can be defined as

$$\begin{aligned}E_T(\mathbf{r}, t) &= E(\mathbf{r}, t) \quad \text{when } |t - t_0| \leq \frac{T}{2} \quad , \\ &= 0 \quad \text{when } |t - t_0| > \frac{T}{2} \quad .\end{aligned}\quad (8.27)$$

The sampled Fourier spectrum is defined analogously to Eq. (8.23), where

$$E_T(\mathbf{r}, t) = \int_{-\infty}^{\infty} \tilde{U}_T(\mathbf{r}, \nu)\exp(-j2\pi\nu t) d\nu \quad , \quad (8.28)$$

and the sampled analytic signal is

$$U_T(\mathbf{r}, t) = 2\int_0^{\infty} \tilde{U}_T(\mathbf{r}, \nu)\exp(-j2\pi\nu t) d\nu \quad . \quad (8.29)$$

By Parseval's Theorem and Eqs. (8.24) and (8.28),

$$\int_{-\infty}^{\infty} E_T^2(\mathbf{r}, t) dt = 2\int_0^{\infty} |\tilde{U}_T(\mathbf{r}, \nu)|^2 d\nu \quad . \quad (8.30)$$

The irradiance is calculated from

$$\begin{aligned}
 I(\mathbf{r}) &= cn\epsilon_0 \langle E^2(\mathbf{r}, t) \rangle = cn\epsilon_0 \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} E_T^2(\mathbf{r}, t) dt \\
 &= cn\epsilon_0 \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{|\tilde{U}_T(\mathbf{r}, \nu)|^2}{2T} d\nu .
 \end{aligned} \tag{8.31}$$

Unfortunately, the limit in the integrand of Eq. (8.31) exhibits an oscillatory behavior as $T \rightarrow \infty$ with a single measurement of the electromagnetic field. However, if several measurements are acquired with significantly long T in a system that obeys ergodic statistics,¹ the average of $|\tilde{U}_T(\mathbf{r}, \nu)|^2/2T$ approaches a definite limit called the *power spectrum*, which is defined as

$$G(\mathbf{r}, \nu) = \lim_{T \rightarrow \infty} \frac{\overline{|\tilde{U}_T(\mathbf{r}, \nu)|^2}}{2T} , \tag{8.32}$$

where the bar denotes an ensemble average over many samples. Notice that $G(\mathbf{r}, \nu)$ is symmetric in ν , due to the Hermitian property of $\tilde{U}_T(\mathbf{r}, \nu)$. In terms of the power spectrum, the irradiance at point \mathbf{r} is

$$I(\mathbf{r}) = cn\epsilon_0 \int_{-\infty}^{\infty} G(\mathbf{r}, \nu) d\nu = 2cn\epsilon_0 \int_0^{\infty} G(\mathbf{r}, \nu) d\nu . \tag{8.33}$$

Note that $G(\mathbf{r}, \nu)d\nu$ is proportional to the contribution to the irradiance from the frequency range $(\nu, \nu + d\nu)$. Therefore, $G(\mathbf{r}, \nu)$ is the *spectral density* of the optical field with units of $\text{V}^2 \text{m}^{-2} \text{s}$. The positive-frequency portion of a typical power spectrum is shown in Fig. 8.9.

Just as in the monochromatic case, we can perform linear operations on the multispectral $U(\mathbf{r}, t)$ analytic signal and take the real part at the end of the calculation to get the real-valued fields. Consequently, the irradiance is found from

$$I(\mathbf{r}) = cn\epsilon_0 \langle E^2(\mathbf{r}, t) \rangle = \frac{1}{2} cn\epsilon_0 \langle |U(\mathbf{r}, t)|^2 \rangle . \tag{8.34}$$

1. Stationarity is defined in Section 8.1.1, and *ergodic statistics* imply that an ensemble average is equal to the corresponding time average involving a typical member of the ensemble.

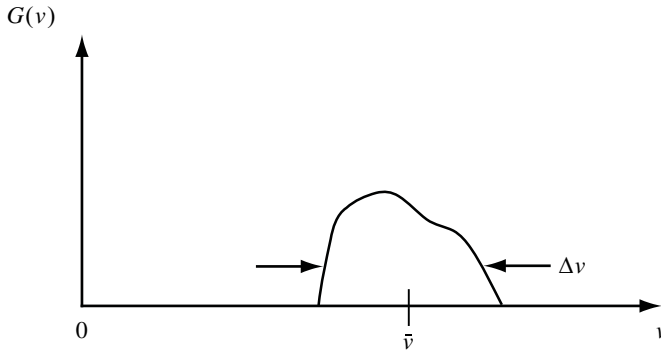


Fig. 8.9. A typical power spectrum for $v \geq 0$.

8.1.4 Basic temporal coherence

Temporal coherence is the dependence of fringe visibility on the power spectrum of the source. In Figs. 8.3 and 8.5 through 8.8, the power spectrum is a simple combination of discrete wavelengths. In this section, the functional dependence of fringe visibility for a continuous power spectrum is derived from a simple argument.

Consider the YDPI of Fig. 8.3 in air. Instead of a dual-wavelength point source, the source has a power spectrum given by $G(v)$. There is no spatial dependence of the power spectrum, because the source is a point. The form of the individual wave components arriving at the observation plane is similar to Eq. (8.9), except that the spectral density must be included.

Following a development similar to Eq. (8.9), the complex fields for one frequency arriving at the observation plane from the pinholes have the forms

$$U_1(s'_1, t) = a_0 \sqrt{K_1} \exp[j(ks'_1 - 2\pi\nu t + \phi)] \quad (8.35)$$

and

$$U_2(s'_1, t, \tau) = a_0 \sqrt{K_2} \exp\{j[ks'_1 - 2\pi\nu(t + \tau) + \phi]\}, \quad (8.36)$$

where $\omega = 2\pi\nu$ and the diffraction constant is separated explicitly from the leading constant factor a_0 in order to better illustrate the dependence on pinhole transmission. K 's are assumed constant over the observation region. Eqs. (8.35)

and (8.36) are written in terms of s'_b , τ is associated with the wave from P2. In terms of the power spectrum,

$$a_0 = 2\sqrt{G(v)dv} \quad , \quad (8.37)$$

The differential irradiance values arriving at the observation plane due to each pinhole separately are

$$\begin{aligned} I_1(v)dv &= \frac{1}{2}cn\epsilon_0|U_1(s'_b, t)|^2 \\ &= K_1I_0g(v)dv \end{aligned} \quad (8.38)$$

and

$$\begin{aligned} I_2(v)dv &= \frac{1}{2}cn\epsilon_0|U_2(s'_b, t, \tau)|^2 \\ &= K_2I_0g(v)dv \quad , \end{aligned} \quad (8.39)$$

where the fractional contribution to the irradiance from spectral components between $(v, v + dv)$ is given by $g(v)dv$, $g(v)$ is the normalized power spectrum

$$g(v) = \frac{G(v)dv}{\int_0^{\infty} G(v)dv} \quad , \quad (8.40)$$

and

$$I_0 = 2cn\epsilon_0 \int_0^{\infty} G(v)dv \quad , \quad (8.41)$$

With the effect of the spectral density, the interference term in Eq. (8.8) is

$$\begin{aligned} cn\epsilon_0\sqrt{K_1K_2}\text{Re}[\langle U_1(s'_b, t)U_2^*(s'_b, t, \tau) \rangle] \\ &= 2I_0\sqrt{K_1K_2}\text{Re}\{\exp[j(2\pi v\tau)]\}g(v)dv \quad (8.42) \\ &= 2I_0\sqrt{K_1K_2}\cos(2\pi v\tau)g(v)dv \quad . \end{aligned}$$

Combination of Eqs. (8.38), (8.39) and (8.42) lead to the observation-plane irradiance resulting from the spectral component between $(\nu, \nu + d\nu)$:

$$I(\tau; \nu) d\nu = I_0 [K_1 + K_2 + 2\sqrt{K_1 K_2} \cos(2\pi \nu \tau)] g(\nu) d\nu \quad , \quad (8.43)$$

where $\tau = \text{OPD}_0/c$ is a function of y_0 , as shown in Fig. 8.4.

Following the discussion in Section 8.1.2, addition of the individual irradiance patterns from each spectral component produces the total fringe pattern. If Eq. (8.43) is integrated over the bandwidth of the source, the description of the irradiance in the observation plane becomes

$$\begin{aligned} I(\tau) &= I_0 \left[(K_1 + K_2) \int_0^{\infty} g(\nu) d\nu + 2\sqrt{K_1 K_2} \int_0^{\infty} g(\nu) \cos(2\pi \nu \tau) d\nu \right] \\ &= I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \text{Re} \left[\int_0^{\infty} g(\nu) e^{-j2\pi \nu \tau} d\nu \right] \right\} \\ &= I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \text{Re} \left[\int_{-\infty}^{\infty} g(\nu) \text{step}(\nu) e^{-j2\pi \nu \tau} d\nu \right] \right\} \\ &= I_0 (K_1 + K_2 + 2\sqrt{K_1 K_2} \text{Re} \{ \mathbf{F}_\tau [g(\nu) \text{step}(\nu)] \}) \\ &= I_0 \{ K_1 + K_2 + 2\sqrt{K_1 K_2} m_{12}(\tau) \cos[\Phi_{12}(\tau)] \} \quad , \end{aligned} \quad (8.44)$$

where

$$\text{step}(\nu) = \begin{cases} 1 & \nu > 0 \\ \frac{1}{2} & \nu = 0 \\ 0 & \nu < 0 \end{cases} \quad (8.45)$$

The new factor multiplying the interference term is

$$m_{12}(\tau) = |\mathbf{F}_\tau [g(\nu) \text{step}(\nu)]| \quad , \quad (8.46)$$

which is the Fourier transform of the normalized positive-frequency source power spectrum, and

$$\Phi_{12}(\tau) = \arg\{\mathbf{F}_\tau[g(v)\text{step}(v)]\} \quad (8.47)$$

is the phase of the Fourier transform.¹ Fringe visibility is

$$V(\tau) = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{2\sqrt{K_1 K_2}}{K_1 + K_2} m_{12}(\tau). \quad (8.48)$$

Notice that $m_{12}(\tau)$ is the primary factor in determining visibility as a function of τ at the observation position. Usually,

$$g(v)\text{step}(v) = f(v - \bar{v}), \quad (8.49)$$

where \bar{v} is the mean frequency of the power spectrum, as shown in Fig. 8.9. In this case,

$$m_{12}(\tau) = |\mathbf{F}_\tau[f(v)]| \quad (8.50)$$

and

$$\Phi_{12}(\tau) = 2\pi\bar{v}\tau + \beta_{12}(\tau), \quad (8.51)$$

where

$$\beta_{12}(\tau) = \arg\{\mathbf{F}_\tau[f(v)]\}. \quad (8.52)$$

With these simplifications, Eq. (8.44) becomes

$$I(\tau) = I_0\{K_1 + K_2 + 2\sqrt{K_1 K_2} m_{12}(\tau) \cos[2\pi\bar{v}\tau + \beta_{12}(\tau)]\} \quad (8.53)$$

Equations (8.44) through (8.53) imply that integration of cosine fringe patterns from individual wavelengths in the power spectrum yields an aggregate fringe pattern with small-scale oscillations having period $1/\bar{v} = \bar{\lambda}/c$ and a visibility that depends on the Fourier transform of the power spectrum of the source. Fringe shift $\beta_{12}(\tau)$ is observed if the power spectrum is asymmetric about the mean frequency \bar{v} .

1. Subscripts "12" on m , Φ and β are a reference to the sampling points of the source field $U(\mathbf{r}, t)$ from which the coherence is measured. For example, the pinholes P1 and P2 define the sampling points of a YDPI.

Example 8.1: Temporal coherence with a simple rectangular power spectrum

Consider a point source with a power spectrum given by a simple rectangular function, with a mean frequency $\bar{\nu}$ and a bandwidth $\Delta\nu$. The source is used on axis in a YDPI, like the geometry of Fig. 8.3.

Mathematically, the positive portion of the normalized power spectrum is given by

$$g(\nu)\text{step}(\nu) = \frac{1}{\Delta\nu} \text{rect}\left(\frac{\nu - \bar{\nu}}{\Delta\nu}\right). \quad (8.54)$$

Calculation of the Fourier transform of Eq. (8.54) yields

$$\mathbf{F}_\tau[g(\nu)\text{step}(\nu)] = \text{sinc}(\Delta\nu\tau) \exp(-j2\pi\bar{\nu}\tau). \quad (8.55)$$

Using Eqs. (8.46) and (8.47) with the result from Eq. (8.55), $m_{12}(\tau)$ and $\Phi_{12}(\tau)$ are found to be

$$m_{12}(\tau) = |\text{sinc}(\Delta\nu\tau)| \quad (8.56)$$

and

$$\Phi_{12}(\tau) = 2\pi\bar{\nu}\tau + \arg[\text{sinc}(\Delta\nu\tau)] \quad , \quad (8.57)$$

respectively. Substitution of Eqs. (8.56) and (8.57) into Eq. (8.44) with $\tau = \text{OPD}_0/c \approx dy_0/cz_0$ in a YDPI yields the following interference pattern in the observation plane

$$I(y_0) = I_0 \left\{ 1 + \left| \text{sinc}\left(\frac{\Delta\nu dy_0}{cz_0}\right) \right| \cos \left[2\pi \frac{\bar{\nu} dy_0}{cz_0} + \beta_{12}\left(\frac{\Delta\nu dy_0}{cz_0}\right) \right] \right\} \quad , \quad (8.58)$$

or

$$I(y_0) = I_0 \left\{ 1 + \text{sinc}\left(\frac{\Delta\nu dy_0}{cz_0}\right) \cos\left(2\pi \frac{\bar{\nu} dy_0}{cz_0}\right) \right\} \quad , \quad (8.59)$$

where the cosine term creates fringes with period $\Lambda = cz_0/d\bar{\nu}$ as a function of y_0 and the sinc term is the bipolar modulation in visibility. The phase of the sinc function is formally included in $\beta_{12}\left(\frac{\Delta\nu dy_0}{cz_0}\right)$, as shown in Eq. (8.58), but it is often convenient to write the form shown in Eq. (8.59). A gray-scale illus-

tration of the fringe pattern is shown in Fig. 8.10, and plots of the irradiance and visibility along the y -axis are shown in Fig. 8.11.

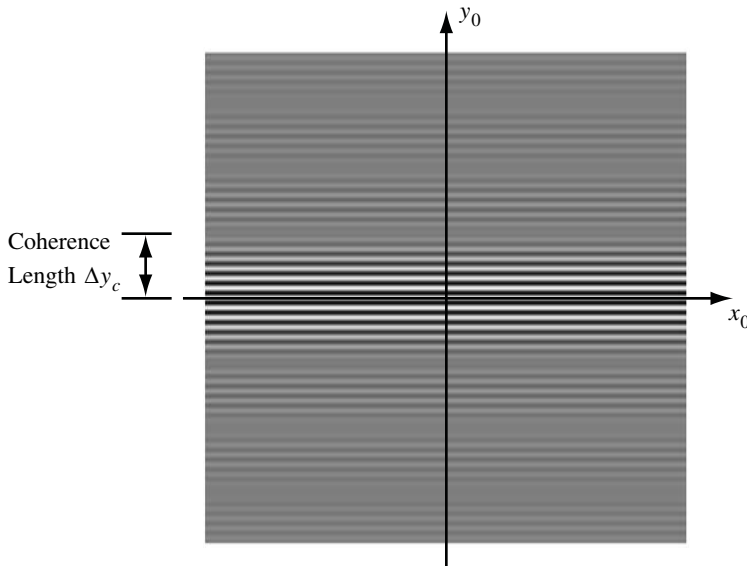


Fig. 8.10. Fringes in the YDPI observation plane for the rectangular power spectrum of Example 8.1.

In this example, visibility reduces as τ (and, thus OPD_0) increases from zero, as evidenced by the form of the sinc function in Eq. (8.59). When visibility drops to zero, there is no fringe contrast. The OPD_0 between maximum fringe visibility at $OPD_0 = 0$ and the first zero in visibility is the coherence length $l_c = c/\Delta\nu$. *Coherence time*, which corresponds to coherence length, is given by $t_c = l_c/c = 1/\Delta\nu$. In the observation space, the coherence length corresponds to the physical distance $\Delta y_c = \frac{cz_0}{d\Delta\nu}$. All two-beam interferometers exhibit similar behavior, with the general result that coherence time is approximately the inverse of the width of the power spectrum. In practice, coherence length may not be defined where visibility reaches an absolute zero, but rather where visibility reaches some minimum value, like $V = 0.2$.

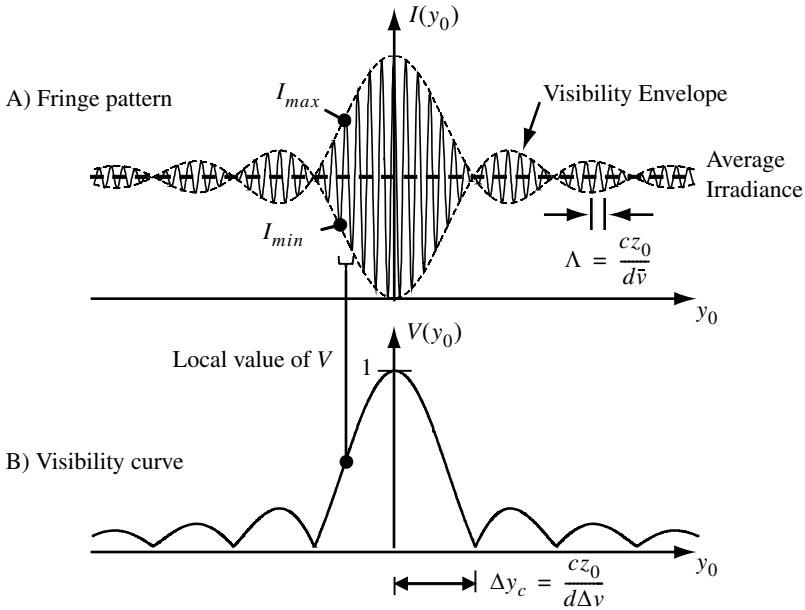


Fig. 8.11. Line profiles of the fringe pattern in Fig. 8.10, where the envelope of the fringe pattern is a sinc function.

Example 8.2: Temporal coherence properties of a Twyman-Green interferometer with a narrow Gaussian power spectrum.

Consider the power spectrum of a HeNe laser ($\bar{\lambda} = 632.8$ nm) with a Gaussian power spectrum and $\Delta\nu = 300$ kHz. The application is for a Twyman-Green interferometer, where the flatness $\epsilon(y_0)$ of a mirror is to be tested, as shown in Fig. 8.12. What is the maximum distance that can be set for the mirror separation $|d_1 - d_2|$ if the minimum resolvable visibility is $V = 0.2$?

By the peculiar property of a Gaussian function, the Fourier transformation of the Gaussian power spectrum is also in the form of a Gaussian function. If we express the normalized power spectrum $g(\nu)\text{step}(\nu) = 1/\Delta\nu \text{gaus}[(\nu - \bar{\nu})/\Delta\nu]$ where $\text{gaus}(\nu) = \exp(-\pi\nu^2)$, the calculation of its Fourier transformation yields

$$|\mathbf{F}_\tau[g(\nu)\text{step}(\nu)]| = |\text{gaus}(\Delta\nu\tau)\exp(-j2\pi\bar{\nu}\tau)| \tag{8.60}$$

From Eq. (8.46),

$$m_{12}(\tau) = |\text{gaus}(\Delta\nu\tau)| \tag{8.61}$$

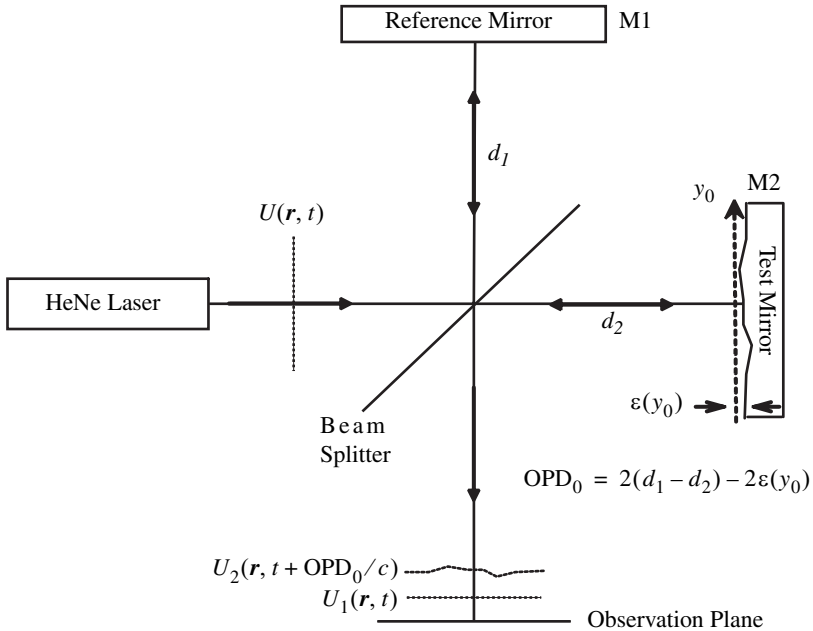


Fig. 8.12. The Twyman-Green amplitude division interferometer uses a collimated laser beam that is divided by a beam splitter. Each split beam reflects off a mirror and recombines at the observation plane. The two beams form an interference pattern in the observation plane that is determined from OPD_0 . Distances d_1 and d_2 are measured from the beamsplitter to the mirrors.

and the fringe visibility $V(\tau)$ is

$$V(\tau) = \text{gaus}(\Delta\nu\tau). \tag{8.62}$$

After replacing $\tau = OPD_0/c = \frac{2}{c}|d_1 - d_2|$ (because light in a Twyman-Green interferometer passes each arm twice) and manipulating the contrast to be larger than 0.2, we obtain the maximum mirror separation:

$$|d_1 - d_2|_{max} = \frac{c}{2\Delta\nu} \sqrt{\frac{\ln 5}{\pi}} \cong 358 \text{ m}. \tag{8.63}$$

For $V > 0.5$, where the fringes can be clearly observed, the maximum mirror separation is 235 m. Due to the long coherence length originating from the narrow spectrum width of the HeNe laser, the fringe visibility does not change significantly over the small OPD_0 changes associated with the mirror imperfections $\epsilon(y_0)$.

Example 8.3: Temporal coherence properties of a multiple-mode laser diode

Consider the power spectrum shown in Fig. 8.13 of a typical laser diode with a mean wavelength $\bar{\lambda} = 649.2 \text{ nm}$ ($\bar{\nu} = c/\bar{\lambda} = 4.621 \times 10^{14} \text{ Hz}$), where parameters of the diode laser include $n_{\text{cavity}} = 3.5$, $L = 200 \text{ }\mu\text{m}$, $\Delta\nu = 3c/2n_{\text{cavity}}L = 6.4 \times 10^{11} \text{ Hz}$ and $\Delta\nu_m = 25 \text{ MHz}$. $\Delta\nu_m$ is the bandwidth of an individual laser mode. A mathematical description of the power spectrum is

$$g(\nu)\text{step}(\nu) = A_0 \text{rect}\left(\frac{\nu - \bar{\nu}}{\Delta\nu}\right) \left[\text{comb}\left(\frac{\nu}{c/2n_{\text{cavity}}L}\right) * \text{gaus}\left(\frac{\nu}{\Delta\nu_m}\right) \right], \quad (8.64)$$

where (*) denotes one-dimensional convolution and A_0 is a normalization constant. Hence,

$$\begin{aligned} m_{12}(\tau) &= |\mathbf{F}_\tau[g(\nu)\text{step}(\nu)]| \\ &= \left| \text{sinc}(\Delta\nu\tau) * \left[\text{comb}\left(\frac{c\tau}{2n_{\text{cavity}}L}\right) \text{gaus}(\Delta\nu_m\tau) \right] \right|. \end{aligned} \quad (8.65)$$

Figure 8.14 shows the visibility curve versus $\text{OPD}_0 = c\tau$, and Fig. 8.15 shows the irradiance.

If a single-mode version of this laser is used in a Twyman-Green interferometer, the visibility only depends on $\Delta\nu_m$, which leads to a 12 m coherence length. However, the multiple-mode nature of the power spectrum in Fig. 8.13 dramatically shortens the coherence length to under 0.5 mm. The visibility returns periodically at $\text{OPD}_0 = 2mn_{\text{cavity}}L$, where m is an integer.

Example 8.4: Fourier Transform Spectroscopy

Often, it is necessary to measure the power spectrum of a source. A detailed description of the power spectrum can yield information about the chemical composition of a star, energy levels of atomic transitions, atmospheric pollution and other applications. A device often used to measure the power spectrum is a *Fourier Transform Spectrometer*. A simple Fourier Transform Spectrometer is shown in Fig. 8.16, which is basically a Twyman-Green Interferometer with the path length of one arm adjusted with a movable scan mirror. d_1 and d_2 are the lengths of each arm, respectively. Thus, the optical path difference (OPD_0) is

$$\text{OPD}_0 = 2(d_1 - d_2). \quad (8.66)$$

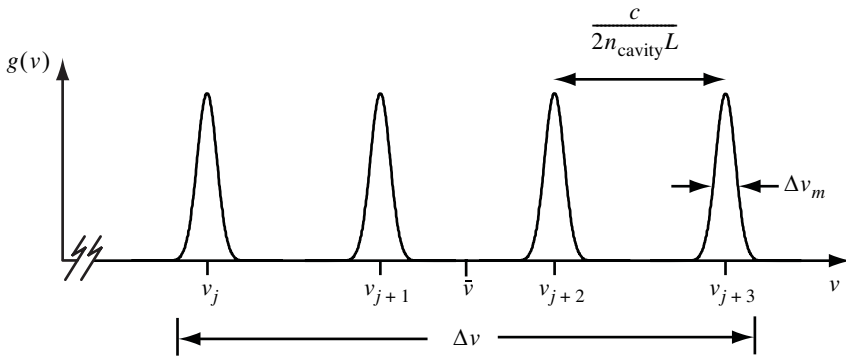


Fig. 8.13. Power spectrum of a multiple-mode laser diode. Multiple laser modes are common in laser diodes. Modes are spaced by $c/(2n_{\text{cavity}}L)$ and are centered at v_j 's determined by other laser properties. Four laser modes are illustrated with approximately equal power, each having width Δv_m . Overall width of the power spectrum is approximately Δv centered at \bar{v} .

When observing a distant point source, the fringe visibility is recorded as a function of OPD_0 . For example, Fig. 8.17 shows a hypothetical fringe pattern measured by the detector current $i(\text{OPD}_0)$ that is characteristic of astronomical observations. The fringe pattern is described mathematically by Eq. (8.44). Assuming that K_1 and K_2 do not vary over the observation space, subtraction of the constant offset from the detector current gives a function proportional to the interference term in Eq. (8.44). That is,

$$\begin{aligned}
 i(\text{OPD}_0) - i_{\text{offset}} &= A_0 m_{12}(\tau) \cos[\Phi_{12}(\tau)] \\
 &= A_0 \text{Re}\{\mathbf{F}_\tau[g(v)\text{step}(v)]\} \\
 &= \frac{1}{2}A_0 \mathbf{F}_\tau[g(v)] \quad ,
 \end{aligned}
 \tag{8.67}$$

where $\tau = \text{OPD}_0/c$ and A_0 is a constant. The power spectrum is found from the inverse Fourier transform of Eq. (8.67), which is

$$g(v) = \frac{2}{A_0} \mathbf{F}_v^{-1}[i(\tau_v) - i_{\text{offset}}] \quad .
 \tag{8.68}$$

Full width of envelope (at 0.043 amplitude) $\sim 2c/\Delta v_m = 24\text{m}$

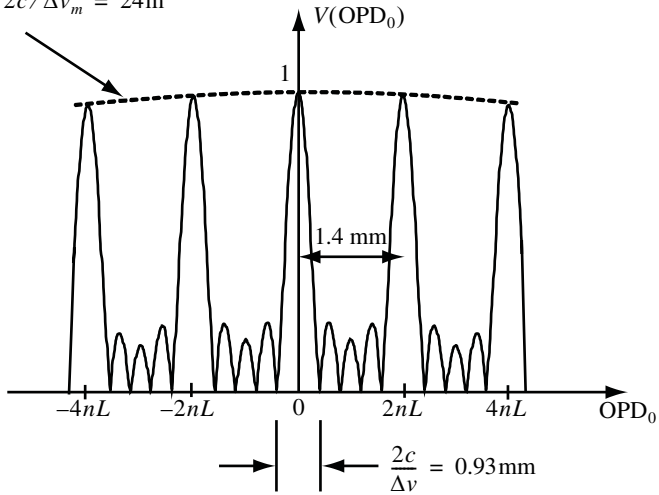


Fig. 8.14. Visibility function associated with Fig. 8.13 and the parameters of Example 8.3, where $n = n_{cavity}$.

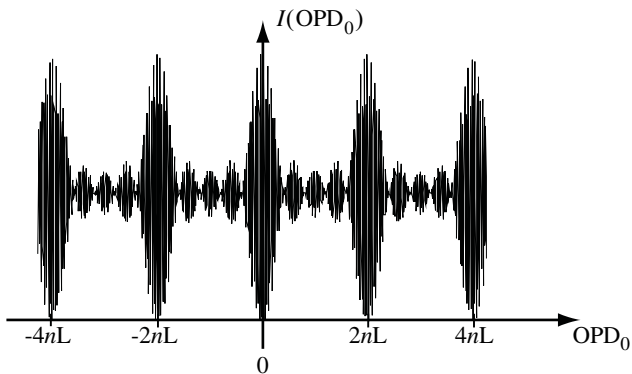


Fig. 8.15. Profile of the fringe pattern observed for Example 8.3.

The power spectrum calculated from the data in Fig. 8.17 is shown in Fig. 8.18. For more information on Fourier Transform Spectroscopy, the reader is directed to References 5.8 and 5.9.

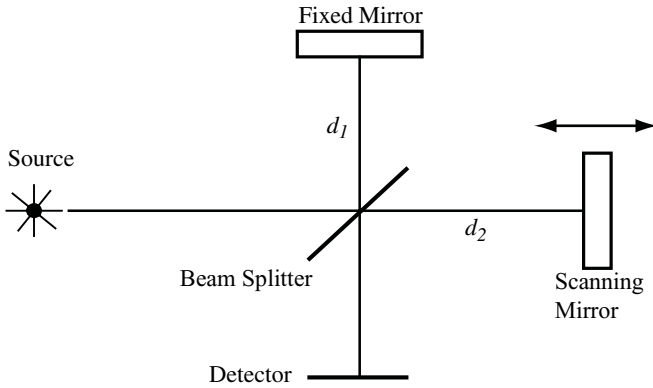


Fig. 8.16. Twyman-Green interferometer used as a Fourier Transform Spectrometer.

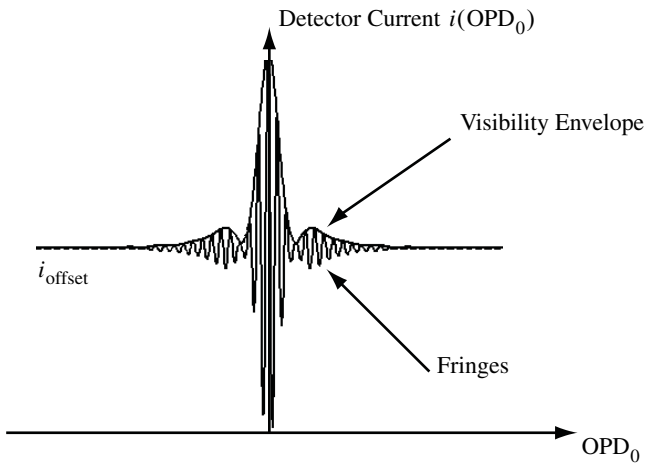


Fig. 8.17. Fringe profile used in Example 8.4, which describes Fourier Transform Spectroscopy.

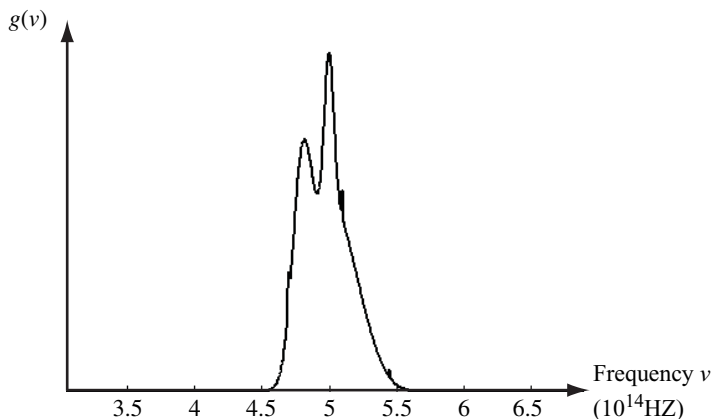


Fig. 8.18. Power spectrum calculated by the inverse transform of data displayed in Fig. 8.17.

8.1.5 Basic Spatial Coherence

Section 8.1.4 assumes that the electric field reaching the observation point from P2 is simply a delayed copy of the electric field reaching the observation point from P1. This condition occurs only if the two electric fields originate from precisely the same position in space, *i.e.* a perfect point source. However, if the electric field originates from a source of finite size, there are additional spatial fluctuations in the electric field.

Spatial coherence is the dependence of fringe visibility on the spatial extent of the source. In this section, distributed sources are treated as a collection of point-like sources that vibrate independently. The spread in wavelengths among the sources is small enough so that the coherence length is much greater than the range of OPD_0 in the measurement. Therefore, fringe visibility is not affected by the bandwidth of the power spectrum. Although the sources have nearly the same wavelength, the temporal beat frequency resulting from combination of light from the sources is beyond the temporal bandwidth of the detector used to measure the aggregate fringe pattern. These conditions serve as an informal definition of a distributed *quasimonochromatic incoherent source*. A more formal definition of a quasimonochromatic source is developed in Section 8.3.2.

For example, consider the effects of two quasimonochromatic point sources located at y_{S1} and y_{S2} , which are used as separate sources in the Young's double pinhole experiment shown in Fig. 8.19. Because the sources are independent, the

interference pattern from one source adds in irradiance with the pattern from the other source. This behavior is similar to the result found in Section 8.1.2, where the irradiance patterns from different wavelengths are added independently, except that now the irradiance patterns resulting from displaced source points are added.

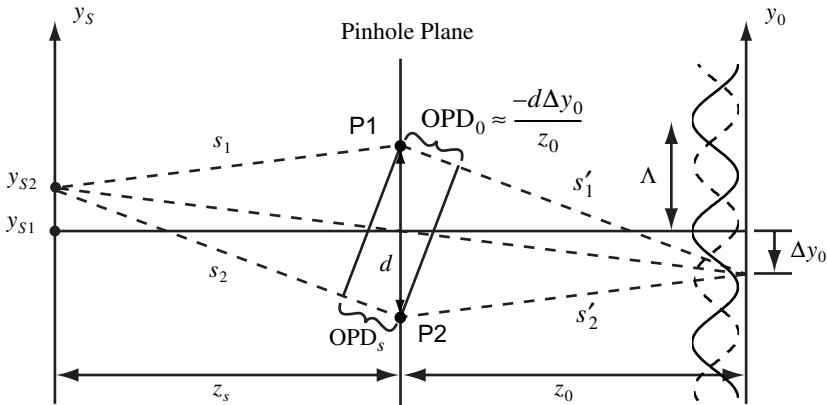


Fig. 8.19. The effect of adding a second quasimonochromatic point source at y_{S2} is to create a second fringe pattern in the observation plane that is shifted by Δy_0 . Note that a positive value of y_{S2} produces negative Δy_0 .

The on-axis source at y_{S1} generates a centered cosine fringe pattern on the observation plane. The off-axis source at y_{S2} generates a shifted fringe pattern due to the additional optical path difference in source space

$$OPD_s = s_1 - s_2 \approx -\frac{y_{S2}d}{z_s}, \tag{8.69}$$

where d is the spacing of the pinholes, $z_s \gg d$ and $z_s \gg y_{S2}$.¹ The shift Δy_0 of the fringe pattern caused by OPD_s is given explicitly by

$$\Delta y_0 = \frac{z_0}{d} OPD = -\frac{z_0 y_{S2}}{z_s}, \tag{8.70}$$

which is independent of source wavelength and pinhole separation. In fact, the shift is simply found by applying similar triangles to the geometry of Fig. 8.19. The shift of the fringe pattern is also described as the change in position of the total OPD between light emitted from the source point passing through the pinholes to the detector plane. That is,

$$\text{OPD}_{\text{total}} = \text{OPD}_s + \text{OPD}_0 = 0 \quad (8.71)$$

at the center of the fringe pattern.

The component fringe patterns and the combined fringe pattern are shown in Fig. 8.20 for two source points, where $\text{OPD}_s = \bar{\lambda}/5$ and $\bar{\lambda}$ is the mean wavelength of the quasimonochromatic source. Notice that the combined pattern has reduced visibility, and the visibility is not a function of observation-space optical path difference OPD_0 . Also, the center of the combined pattern is slightly shifted off axis. If more point sources are added to the source distribution, this trend continues. For example, in Fig. 8.21, the combined fringe pattern from three equally-spaced point sources exhibits significantly degraded visibility. Like with two point sources, the visibility is not a function of OPD_0 . If the source distribution is wide enough such that the fringe centers fill one fringe period, the result can be $V = 0$, as shown in Fig. 8.22. Note that, although $V = 0$ in Fig. 8.22, there is significant irradiance across the observation region.

Figures 8.20 through 8.22 illustrate basic properties of spatial coherence with the YDPI. Firstly, *visibility is not a function of location in the observation space*. Secondly, *visibility is inversely proportional to the source size*. These properties differentiate spatial coherence from temporal coherence in a YDPI.

The idea of adding irradiance patterns from individual source points is now extended to a continuous source distribution. Assume that the source power varies over the source plane. We define the local radiant exitance $M(x_s, y_s)$ as the power per unit area (in units of Wm^{-2}) from differential area $dx_s dy_s$. Fringes in the observation plane created by $M(x_s, y_s)$ over area $dx_s dy_s$ are derived by assuming that light emitted from $dx_s dy_s$ acts like a point source. Since light from neighboring differential areas do not interfere, the total irradiance pattern is found by integration of differential fringe patterns created by each $dx_s dy_s$.

Like the analysis of the power spectrum in Section 8.1.4, we define a normalized radiant exitance

1. In this section, it is necessary to differentiate between source-space optical path difference OPD_s and observation-space optical path difference OPD_0 . The approximation in Eq. (8.69) is also valid for source locations across the plane of the source, as can be determined by taking the first two terms of the Taylor expansions of s_1 and s_2 , as long as

$$z_s \gg \sqrt{x_s^2 + \left(y_s \pm \frac{d}{2}\right)^2}.$$

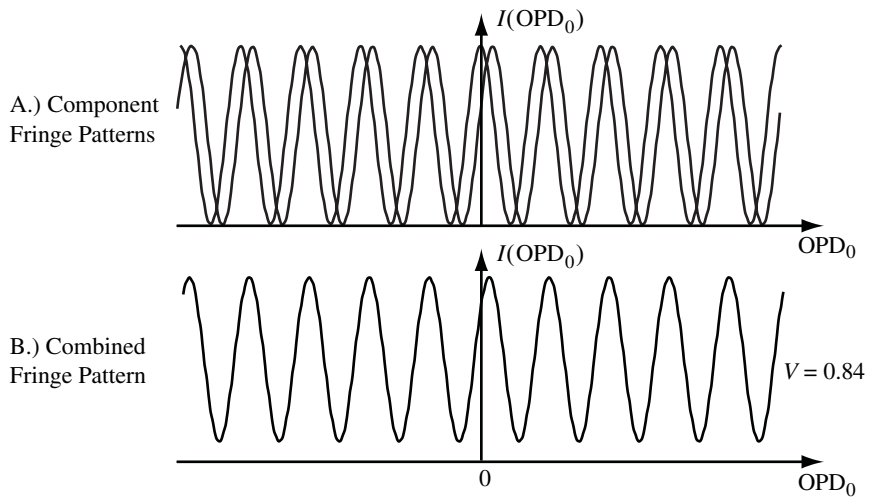


Fig. 8.20. The component fringe patterns from an on-axis source point and an off-axis source point are shown, where $OPD_s = \bar{\lambda}/5$. The total irradiance is the sum of the two fringe patterns. Notice that visibility is reduced uniformly across the observation plane, where $OPD_0 = -dy_0/z_0$. The center of the combined pattern is slightly shifted, but it retains the same period as the component fringes.

$$m_R(x_s, y_s) = \frac{M(x_s, y_s)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M(x_s, y_s) dx_s dy_s} \quad (8.72)$$

The differential irradiance pattern created by $m_R(x_s, y_s) dx_s dy_s$ has the same form as Eq. (8.43), except $g(v)dv$ is replaced by $m_R(x_s, y_s) dx_s dy_s$, with the following result

$$\begin{aligned} & I(\Delta t; x_s, y_s) dx_s dy_s \\ &= I_0 [K_1 + K_2 + 2\sqrt{K_1 K_2} \cos(2\pi v \Delta t)] m_R(x_s, y_s) dx_s dy_s \\ &= I_0 [K_1 + K_2 + 2\sqrt{K_1 K_2} \cos(kc \Delta t)] m_R(x_s, y_s) dx_s dy_s \\ &= I_0 \{ K_1 + K_2 + 2\sqrt{K_1 K_2} \cos[k(OPD_s + OPD_0)] \} m_R(x_s, y_s) dx_s dy_s, \end{aligned} \quad (8.73)$$

where $\Delta t = (OPD_s + OPD_0)/c$, $k = 2\pi/\bar{\lambda}$, and $\bar{\lambda}$ is the average wavelength of the quasimonochromatic source. The factor I_0 is the irradiance incident on the

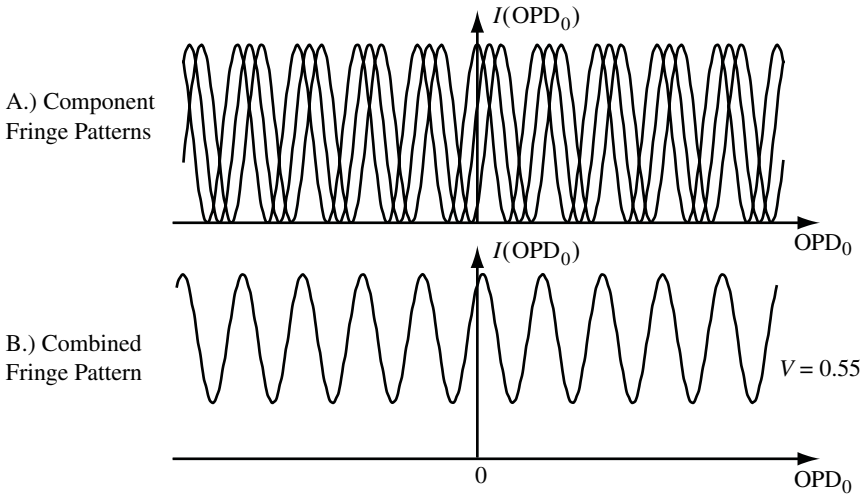


Fig. 8.21. The component fringe patterns from three distributed source points are shown, where $OPD_s = \bar{\lambda}/5$ between the source positions. The total irradiance is the sum of the three fringe patterns. Notice that visibility is reduced uniformly across the observation plane, where $OPD_0 = -dy_0/z_0$. The center of the combined pattern is slightly shifted, but it retains the same period as the component fringes.

pinholes, where we have assumed that both pinholes receive the same irradiance from the distributed source. Integration over the continuous source spatial distribution $m_R(x_s, y_s)$ removes the dependence on OPD_s and yields

$$I(OPD_0) = I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos[k(OPD_s + OPD_0)] m_R(x_s, y_s) dx_s dy_s \right\} \quad (8.74)$$

If we apply the approximation for OPD_s from Eq. (8.69), the irradiance becomes

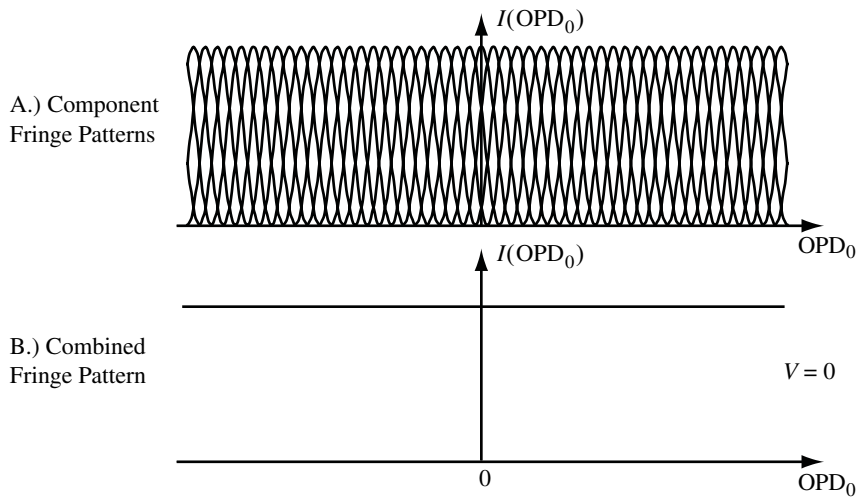


Fig. 8.22. The component fringe patterns from five distributed source points are shown, where $OPD_s = \bar{\lambda}/5$ between the source positions. The total irradiance is the sum of the five fringe patterns. Notice that visibility is reduced to zero uniformly across the observation plane, where $OPD_0 = -dy_0/z_0$. However, zero visibility does not imply zero irradiance.

$$\begin{aligned}
 &= I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos \left[k \left(\frac{-y_s d}{z_s} + OPD_0 \right) \right] m_R(x_s, y_s) dx_s dy_s \right\} \\
 &= I_0 (K_1 + K_2 \\
 &\quad + 2\sqrt{K_1 K_2} \operatorname{Re} \left\{ \exp(jkOPD_0) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m_R(x_s, y_s) \exp \left[jk \left(\frac{-y_s d}{z_s} \right) \right] dx_s dy_s \right\} \right) \quad (8.75) \\
 &= I_0 (K_1 + K_2 \\
 &\quad + 2\sqrt{K_1 K_2} \operatorname{Re} \left\{ \exp(jkOPD_0) \mathbf{F}_{d/z_s} \left[\int_{-\infty}^{\infty} m_R(x_s, y_s) dx_s \right] \right\} \right) .
 \end{aligned}$$

Notice that the Fourier transform operates on the projection of the x_s axis onto the y_s direction of the normalized source distribution. Due to the pinhole orientation along the y axis, the source distribution in the x_s direction is integrated to produce a one-dimensional function, which is then operated on with a Fourier transform to obtain the complete fringe distribution in the observation plane. The Fourier operation is a direct result of adding individual cosine distributions from each differential area of the source. Instead of individual fringe patterns with different periods, as discussed with temporal coherence in Section 8.1.4, the individual fringe patterns have the same period and are simply shifted with respect to each other.

Equation (8.75) is now simplified with several substitutions. First, the result of the Fourier transform is recognized to exhibit an amplitude and a phase, which can be combined with the exponential containing OPD_0 . The magnitude is

$$\mu_{12}\left(\frac{d}{z_s \bar{\lambda}}\right) = \left| \mathbf{F}_{d/z_s \bar{\lambda}} \left[\int_{-\infty}^{\infty} m_R(x_s, y_s) dx_s \right] \right|, \quad (8.76)$$

and the phase is

$$\beta_{12}\left(\frac{d}{\bar{\lambda} z_s}\right) = \arg \left\{ \mathbf{F}_{d/z_s \bar{\lambda}} \left[\int_{-\infty}^{\infty} m_R(x_s, y_s) dx_s \right] \right\}. \quad (8.77)$$

By applying the real part of the bracketed expression in Eq. (8.75), we obtain

$$\begin{aligned} & I(\text{OPD}_0) \\ &= I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \mu_{12}\left(\frac{d}{\bar{\lambda} z_s}\right) \cos \left[k \text{OPD}_0 + \beta_{12}\left(\frac{d}{\bar{\lambda} z_s}\right) \right] \right\}, \end{aligned} \quad (8.78)$$

or

$$I(\tau) = I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \mu_{12}\left(\frac{d}{\bar{\lambda} z_s}\right) \cos \left[2\pi \bar{\nu} \tau + \beta_{12}\left(\frac{d}{\bar{\lambda} z_s}\right) \right] \right\}, \quad (8.79)$$

as a function of $\tau = \text{OPD}_0/c$. Notice that Eq. (8.79) is very similar in form to Eq. (8.53), except the leading factor in front of the cosine and the fringe shift inside the argument of the cosine are not functions of observation-space location τ . Finally, substitution of the angular subtense of the pinholes as seen from the source $\theta_{ph} = d/z_s$ into Eq. (8.79) results in

$$I(\tau) = I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \mu_{12} \left(\frac{\theta_{ph}}{\lambda} \right) \cos \left[2\pi \bar{\nu} \tau + \beta_{12} \left(\frac{\theta_{ph}}{\lambda} \right) \right] \right\} . \quad (8.80)$$

Note that the fringe visibility,

$$V \left(\frac{\theta_{ph}}{\lambda} \right) = \frac{2\sqrt{K_1 K_2}}{K_1 + K_2} \mu_{12} \left(\frac{\theta_{ph}}{\lambda} \right) \quad (8.81)$$

is not a function of position y_0 in the observation space. Fringe spacing $\Lambda = z_0 \bar{\lambda} / d$ in the observation plane is identical for any point in the source. This development results in a similar mathematical structure to that found in Section 8.1.4. However, although Eqs. (8.48) and (8.81) are similar, Eq. (8.81) does not depend on τ .

8.1.6 Interpretation of the van Cittert-Zernike Theorem

Equations (8.76) and (8.81) are elements of the van-Cittert Zernike theorem, which can be stated for a YDPI:

If a quasimonochromatic source is a considerable distance from the aperture plane and pinhole separation is small, fringe visibility from an extended source is proportional to the inverse Fourier transform of the source's spatial distribution. The transform variable is the angular separation of the aperture-plane sampling points divided by the wavelength.

Example 8.5: Spatial coherence from a one-dimensional line source

Consider the quasimonochromatic one-dimensional line source as shown in Fig. 8.23, which can be described by

$$m_R(x_s, y_s) = \frac{1}{L} \text{rect} \left(\frac{y_s}{L} \right) \delta(x_s), \quad (8.82)$$

where, if $K_1 = K_2 = 1$,

$$V \left(\frac{\theta_{ph}}{\lambda} \right) = \mu_{12} \left(\frac{\theta_{ph}}{\lambda} \right) = \left| \text{sinc} \left(L \frac{\theta_{ph}}{\lambda} \right) \right|. \quad (8.83)$$

A plot of Eq. (8.83) is shown in Fig. 8.24 with $\theta_{ph} = d/z_s$. A reference line at $L\theta_{ph}/\lambda = 0.6$ is drawn to help illustrate variable dependencies. A complete understanding of the van Cittert-Zernike theorem is obtained by recognizing how visibility changes for each variable separately.

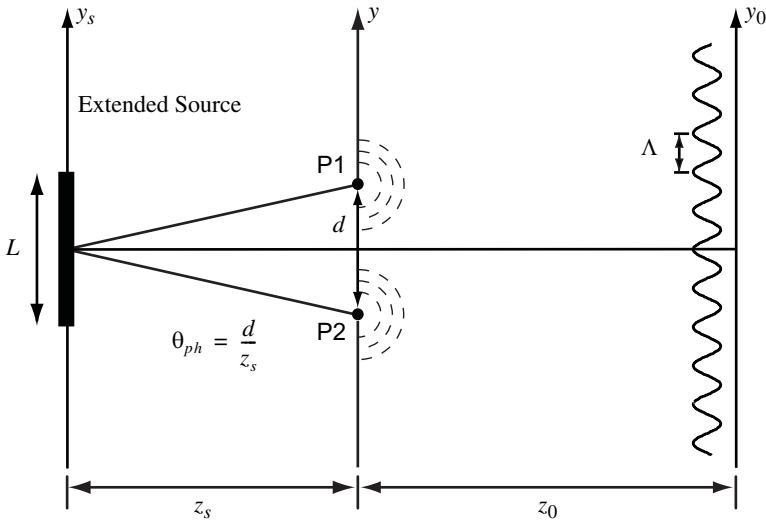


Fig. 8.23. One-dimensional source distribution with length L and angular subtense for Example 8.5.

Increasing z_s or decreasing L has the effect of decreasing the maximum fringe shift and increasing visibility. The contrast is increased because the range of shifts in the component fringe patterns is reduced compared to the fringe period. (Remember, the fringe period does not depend on z_s or L .) Thus, decreasing the angular extent of the source as seen by the pinholes increases visibility. Increasing pinhole separation d or decreasing λ has the effect of reducing the fringe spacing, which decreases visibility, because the range of shifts in the observation area does not change.

Example 8.6: Spatial coherence with a lateral shearing interferometer

Consider a lateral shearing interferometer (LSI), as shown in Fig 8.25. This type of interferometer is typically used to test collimation of laser beams and other optical systems. The LSI forms an interference pattern from two reflections off a plane parallel glass plate. The first reflection interferes with the second reflection, which is a shifted version of the same wavefront. The shift is the shear distance S , which is a function of the tilt angle, the thickness T and refractive index n of the plate. The incident wavefront must exhibit good spatial coherence over the shear distance in order to observe high visibility fringes.

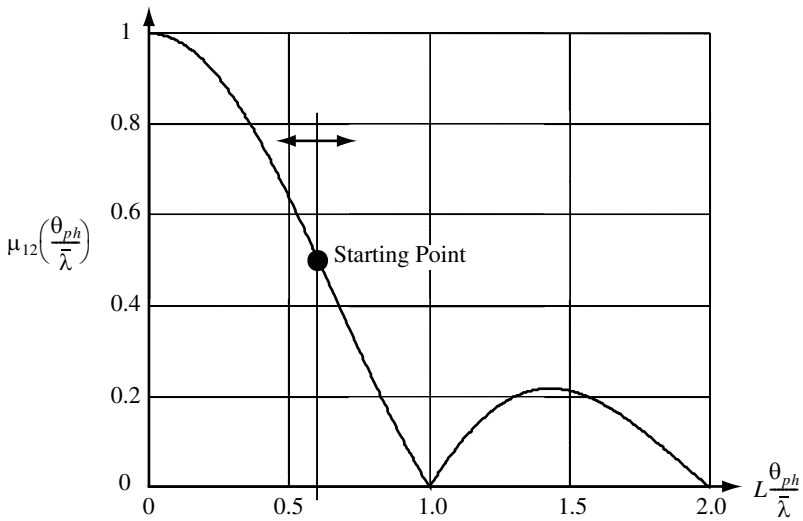


Fig. 8.24. **Visibility for an extended source.** Visibility is shown as a function of the normalized variables $L\theta_{ph}/\bar{\lambda}$ for the one-dimensional extended source in Example 8.5. In order to increase visibility from the starting point, θ_{ph} or L must be decreased, or $\bar{\lambda}$ must be increased.

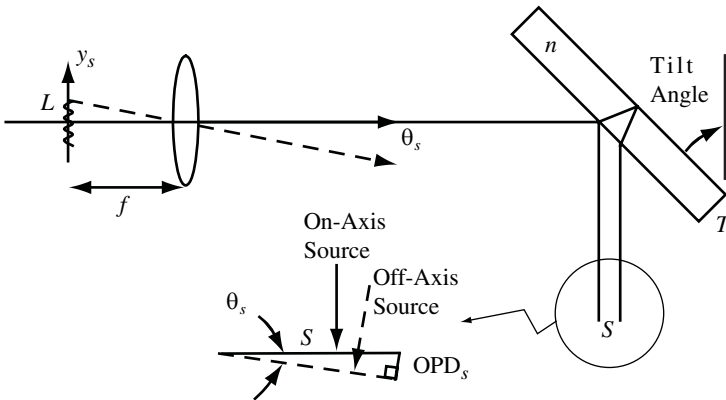


Fig. 8.25. **Spatial coherence for a lateral shearing interferometer (LSI).**

If the extended quasimonochromatic source in Fig. 8.25 is collimated by a lens of focal length f , the *angular* distribution of the light exiting the lens determines the spatial coherence properties. OPD_s between points interfering in the shifted wavefront vary, due to the angular difference between the waves from on-axis and off-axis source points. OPD_s due to source points distributed along the y_s axis is

$$\text{OPD}_s = S \sin \theta_s, \quad (8.84)$$

where

$$\sin \theta_s \approx \frac{y_s}{f}. \quad (8.85)$$

The OPD_s in Eq. (8.74) is now $\text{OPD}_s = -y_s S/f$ instead of $-y_s d/z_s$. Therefore, μ_{12} becomes

$$\mu_{12}\left(\frac{S}{\bar{\lambda}f}\right) = \left| \mathbf{F}_{S/\bar{\lambda}f} \left[\int_{-\infty}^{\infty} m_R(x_s, y_s) dx_s \right] \right|. \quad (8.86)$$

For a rectangular source distribution,

$$\mu_{12}\left(\frac{S}{\bar{\lambda}f}\right) = \left| \text{sinc}\left(L \frac{S}{\bar{\lambda}f}\right) \right|. \quad (8.87)$$

The extent L of the incoherent quasimonochromatic source must be smaller than $\bar{\lambda}f/S$ to obtain good visibility. For example, if $T = 12.5$ mm, $n = 1.5$ and the tilt angle is 45° , $S = 5.2$ mm. If $f = 100$ mm, the source size that first produces a minimum in visibility is

$$L = \frac{f\bar{\lambda}}{S} = \frac{(100 \times 10^{-3} \text{ m})(550 \times 10^{-9} \text{ m})}{5.2 \times 10^{-3} \text{ m}} = 1.1 \times 10^{-5} \text{ m} = 11 \mu\text{m}.$$

Example 8.7: Michelson Stellar Interferometer

The Michelson stellar interferometer is shown in Fig. 8.26. The fringe spacing on the observation screen depends on d , the separation of the pinholes. However, the visibility depends on h , the separation of the collection mirrors, because, like in Example 8.6, $\text{OPD}_s = h \sin \theta_s$. However, the source radiance is specified in terms of the angle as viewed from the center of the telescope, not as a physical distance in a plane. In this case, the development

of Eq. (8.80) is modified to use $m_R(\theta_{sx}, \theta_{sy})$ with $h/\bar{\lambda}$ as the transform variable. that is, μ_{12} becomes

$$\mu_{12}\left(\frac{h}{\bar{\lambda}}\right) = \left| \mathbf{F}_{\frac{h}{\bar{\lambda}}} \left[\int_{-\pi}^{\pi} m_R(\theta_{sx}, \theta_{sy}) d\theta_{sx} \right] \right| \quad (8.88)$$

Thus, the observer maps the fringe visibility by changing h . The angular radiance distribution $m_R(\theta_{sx}, \theta_{sy})$ is calculated by taking the inverse Fourier transform of measured visibility.

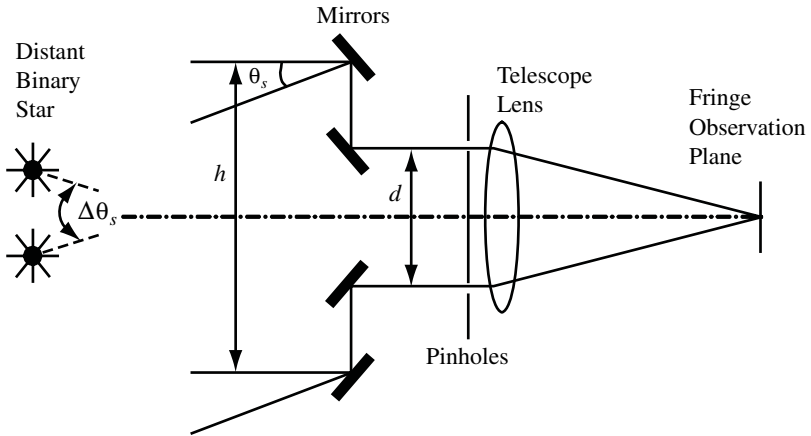


Fig. 8.26. **Michelson Stellar Interferometer** for determining properties of distant stars.

One application of the Michelson stellar interferometer is to measure the angular separation of binary stars. For this problem, the radiance distribution is

$$m_R(\theta_{sx}, \theta_{sy}) = \frac{1}{2} \left[\delta\left(\theta_{sy} + \frac{\Delta\theta_s}{2}\right) + \delta\left(\theta_{sy} - \frac{\Delta\theta_s}{2}\right) \right] \delta(\theta_{sx}) , \quad (8.89)$$

where $\Delta\theta_s$ is the total angular separation of the stars. Thus, the visibility is

$$\begin{aligned}
 V\left(\frac{h}{\bar{\lambda}}\right) &= \left| \mathbf{F}^{-1}_{h/\bar{\lambda}} \left[\int_{-\pi}^{\pi} m_R(\theta_{sx}, \theta_{sy}) d\theta_{sx} \right] \right| \\
 &= \frac{1}{2} \left| e^{j2\pi h \Delta\theta_s / 2\bar{\lambda}} + e^{-j2\pi h \Delta\theta_s / 2\bar{\lambda}} \right| \\
 &= \left| \cos(\pi h \Delta\theta_s / \bar{\lambda}) \right|
 \end{aligned} \tag{8.90}$$

Therefore, if the separation of mirrors and wavelength are determined, and the visibility is measured, the angular separation $\Delta\theta_s$ can be obtained. In the example above, if h is the mirror separation that produces the first zero in fringe visibility, $\Delta\theta_s$ is given by

$$\Delta\theta_s = \frac{\bar{\lambda}}{2h} . \tag{8.91}$$

For example, consider when the first zero of fringe visibility occurs at $h = 27$ m with $\bar{\lambda} = 500$ nm, $\Delta\theta = 9.2 \times 10^{-9}$ radians, and the angular separation of the binary stars is about 2 milli-arc-seconds (mas), where 1 mas = 4.6 nanoradians (4.6×10^{-9} radians). For comparison, the angular resolution of the human eye is about 12.5 arc-seconds or 58 microradians (58×10^{-6} radians). The Hale observatory has a resolution of about 25 mas, and the Naval Observatory Michelson Stellar Interferometer has a resolution of about 200 micro-arc-seconds.¹ In simple terms, this capability is like resolving the width of a human hair at a distance of 25 miles. The optical schematic for the Mark III Stellar Interferometer at the Naval Observatory is shown in Fig. 8.27.

8.1.7 The Concept of Coherence Area

The van Cittert-Zernike Theorem of Section 8.1.6 applied to a two-dimensional quasimonochromatic source distribution provides insight into the coherent nature of the light field. The concept of *coherence area* suggests that there is some area over which the light field behaves coherently at various distances away from the source. For example, the sketch in Fig. 8.28 shows the outline of a source distribution and the outline of its associated Fourier transform in an aperture/pinhole plane. The aperture/pinhole plane could be the pinhole plane of a YDPI, the input aperture of an interferometer or the object plane of a microscope. The distribution of the Fourier transform, appropriately scaled from Eq. (8.76), determines the visibility of interference fringes for separations d of points in the aperture plane. The coherence area is defined by *a region representing*

1. See <http://www.mtwilson.edu/Tour/NRL/mark3.html> for more information about the Naval Observatory's instrument.

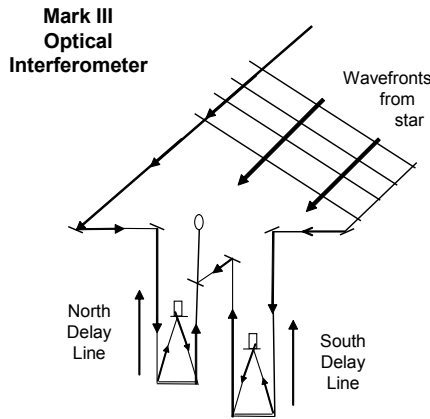


Fig. 8.27. The Mark III Stellar Interferometer at the Naval Observatory.

separations of points in the aperture plane that can interfere with high visibility. The maximum extent of this region is the boundary determined by the minimum visibility acceptable to the optical system of interest. Usually, the zeros of the visibility function are the easiest limits to consider, but a value of $V = 0.2$ is often used for experimental data.

For example, the source in Fig. 8.28 is longer in the y_s dimension than in the x_s dimension. The source distribution can be roughly approximated with a rectangle with sides of length L_x and L_y . The Fourier transform of the rectangle is separable, and the zeros of the resulting sinc visibility function are used to define the coherence area. For example, the maximum width for good coherence (zeros of V) in the y -dimension is $\Delta y = z_s \bar{\lambda} / L_y$. Note that, because of the rectangular shape in the source plane, the coherence area is also wider in one dimension than in the other. The wide dimension of the coherence area corresponds to the narrow dimension of the source. That is, *a narrow source dimension presents a wide coherence area in the pinhole/aperture plane*. In the example of Fig. 8.28, the coherence area is

$$\text{Coherence area} = \Delta x \Delta y = \frac{z_s^2 \bar{\lambda}^2}{L_x L_y} \quad (8.92)$$

Example 8.8: Coherence from a small, round distributed source.

A 1 mm diameter Gaussian incoherent extended source of average wavelength 500 nm is to be used in a diffraction experiment, for which it is desired to illuminate a distant 3 mm diameter aperture coherently. Calculate the

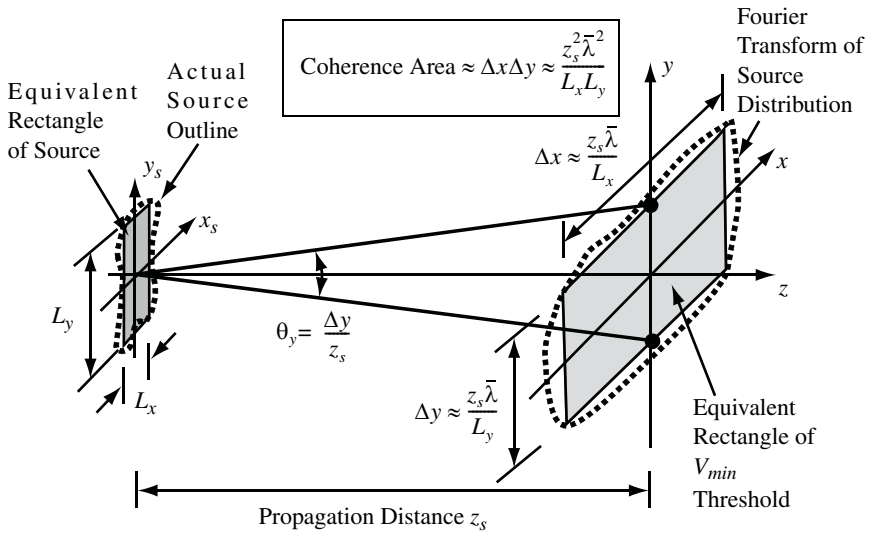


Fig. 8.28. **The concept of coherence area** is an extension of the Van Cittert Zernike Theorem.

minimum distance between the pinhole source and the diffracting aperture for good coherence. State any assumptions being made.

The spatial coherence at the aperture is obtained by applying the van Cittert-Zernike Theorem. The visibility is proportional to the Fourier transform of the projected source distribution, which can be described by

$$\int_{-\infty}^{\infty} m_R(x_s, y_s) dxs = \int_{-\infty}^{\infty} \frac{1}{r^2} \text{gaus}\left(\frac{x_s}{r}\right) \text{gaus}\left(\frac{y_s}{r}\right) dxs \quad , \quad (8.93)$$

and μ_{12} is

$$\mu_{12}\left(\frac{d}{\lambda z_s}\right) = \text{gaus}\left(\frac{rd}{\lambda z_s}\right) \quad , \quad (8.94)$$

where r is the diameter of the source at approximately the 4% radiant exitance level, d is the diameter of the distant aperture and z_s is the distance from the source to the aperture. Assume illumination is coherent for $V = \mu_{12} > 0.2$. The minimum distance for good coherence is

$$z_s = \sqrt{\frac{-\pi}{\log 0.2} \frac{rd}{\bar{\lambda}}} = \sqrt{\frac{-\pi}{\log 0.2} \frac{(0.5 \times 10^{-3})(3 \times 10^{-3})}{0.5 \times 10^{-6}}} = 4.2 \text{ m} \quad (8.95)$$

8.1.8 Combination of Temporal and Spatial Coherence

The complete description of coherence properties resulting from a multiple-wavelength, extended source is provided in Section 8.3. However, it is possible to understand the basic character of coherence from a heuristic argument. For example, consider the off-axis polychromatic point source pictured in the YDPI of Fig. 8.29. With reference to Fig. 8.19 and with the results of Section 8.1.4, maximum fringe visibility is observed at the point on the observation screen where $OPD_{\text{total}} = OPD_0 + OPD_s = 0$. The decrease in V away from this point is governed by the bandwidth $\Delta\nu$ of the source. Narrow $\Delta\nu$ implies good V over a wide distance along y_0 . The fringe pattern associated with the single source point is the point response of the YDPI system. Shifting the source point simply shifts the point response along the observation screen. In the central region of the observation screen, the relationship is linear and shift invariant in y_0 . The relationship is always linear and shift invariant in OPD_0 .

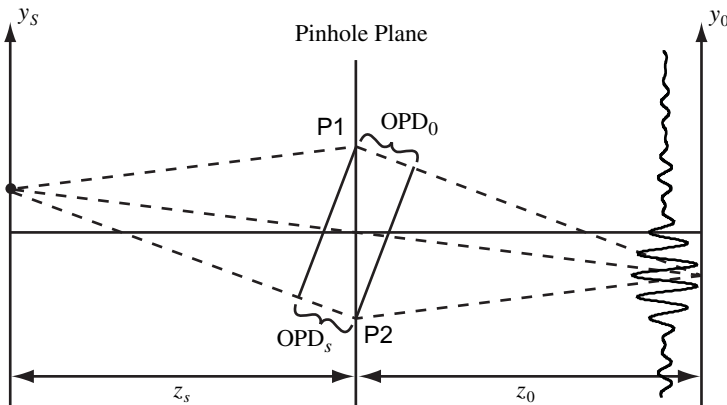


Fig. 8.29. The point response of an off-axis polychromatic source.

If more than one source point illuminates the YDPI, the point response fringe patterns are summed to produce the total fringe pattern. In the limit of a

continuous source distribution, the total fringe pattern is found by convolving the point response with the appropriately scaled source distribution.

For example, consider a line source distribution of length L along y_s , like in Example 8.5. If L is small, the shifts of the point responses are less than the width of a fringe period. Therefore, the convolution results in a uniform loss factor in visibility across observation screen, but the basic shape of the point response remains. At the point on the screen where $\text{OPD}_{\text{total}} = 0$, the loss in visibility is due only to the spatial extent of the source, and not the temporal bandwidth. As a rough approximation, the total visibility is governed by the relationship

$$V\left(\tau, \frac{\theta_{ph}}{\lambda}\right) \propto \mu_{12}\left(\frac{\theta_{ph}}{\lambda}\right) m_{12}(\tau) . \quad (8.96)$$

8.1.9 Terminology Used in Coherence Theory

Equation (8.8) can be written in the following form

$$I(\mathbf{r}_1, \mathbf{r}_2 ; \tau) = I_1 + I_2 + 2\sqrt{I_1 I_2} \text{Re}[\gamma_{12}(\tau)] , \quad (8.97)$$

where I_1 and I_2 are irradiances at the observation point from each pinhole independently, and $\gamma_{12}(\tau)$ is the *normalized mutual coherence function* (also called the *complex degree of coherence*), which is given by

$$\gamma_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} . \quad (8.98)$$

It is understood that τ is a function of the observation-space location. The normalization limits values of $\gamma_{12}(\tau)$ so that $0 \leq |\gamma_{12}(\tau)| \leq 1$. Equation (8.97) can also be written as

$$I(\mathbf{r}_1, \mathbf{r}_2 ; \tau) = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma_{12}(\tau)| \cos[\alpha_{12}(\tau)] , \quad (8.99)$$

where

$$\alpha_{12}(\tau) = \arg[\gamma_{12}(\tau)] . \quad (8.100)$$

When working with temporal coherence problems, the complex degree of coherence is related to the terms in Section 8.1.4 by

$$m_{12}(\tau) = |\gamma_{12}(\tau)| , \quad (8.101)$$

and

$$\beta_{12}(\tau) = \arg[\gamma_{12}(\tau)] . \quad (8.102)$$

When working with spatial coherence problems, the complex degree of coherence is related to the terms in Section 8.1.5 by

$$\mu_{12} = |\gamma_{12}(0)| = \left| \frac{J_{12}}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} \right| , \quad (8.103)$$

and

$$\beta_{12} = \arg[\gamma_{12}(0)] , \quad (8.104)$$

where

$$J_{12} = \Gamma_{12}(0) \quad (8.105)$$

is the *mutual intensity*. In this case, μ_{12} is the normalized mutual intensity. Notice that, if pinhole positions P1 and P2 are coincident, $J_{11} = \Gamma_{11}(0)$ is the *irradiance* at that point in the light field before the pinholes. Also, the *self coherence* of the light field is given by $\Gamma_{11}(\tau)$.

The complex degree of coherence, mutual intensity and self coherence are not functions of pinhole transmission or propagation loss from the pinholes to the observation point. They are extremely important measures of the property of a light field.

8.2 Fringe Localization

Waves traveling through space interfere and produce visible fringes if the conditions are right. In particular, the waves must have some degree of spatial and temporal coherence over a region of space. *Fringe localization* defines the region of space where interference occurs and fringes with reasonably good contrast are observed. There is a wide variety of methods for producing fringes – ranging from a simple parallel plate to complex interferometers, and each has an associated region in space where high-contrast fringes may be observed. The location of this region relative to the components of the interferometer depends on properties of the source and geometry of the interferometer. In this section, basic properties of fringe localization are illustrated for various types of sources and interferometers.

There are several different categories of fringe localization in interferometers. Some of the more common categories are:

- 1) *Localized everywhere*: The fringes have high visibility everywhere in the observation space. (Also called *non-localized* or *unlocalized*.)
- 2) *Localized*: The fringes have high visibility over some surface in the observation space. The fringe localization surface can be curved.
- 3) *Localized at infinity*: A lens is used to transform the angular distribution of fringes into a spatial distribution of fringes on an observation plane with high visibility. The observation plane is usually at the back focus of the lens. Otherwise, fringe visibility is too low for observation. (Also called *Haidinger's fringes* or *fringes of equal inclination*.)
- 4) *Fringes of equal thickness*: Localized fringes, where fringes correspond to contours of constant thickness between two surfaces.

8.2.1 Discussion of sources for illumination

Several types of sources are used for illumination in interferometers and other optical systems. This section discusses basic coherence properties of four different types of sources: monochromatic point source, polychromatic point source, quasi-monochromatic extended source and polychromatic extended source.

8.2.1.1 Monochromatic point source

Any interferometer using an ideal, monochromatic point of light as a source produces high-visibility fringes *everywhere* in observation space. In this case, two-point-source interference produces a hyperboloidal fringe field, as shown in Fig. 8.30. The fringes occur where OPD_0 is an integer number of wavelengths. These fringes are *nonlocalized* or *localized everywhere* or *unlocalized*. Of course, there is no truly ideal monochromatic point source of light. However, a good approximation is found by using high-quality laser sources, where the coherence

length due to the temporal bandwidth of the laser is much greater than the maximum OPD_0 . That is,

$$\frac{c}{n\Delta\nu} \gg (OPD_0)_{\max} , \tag{8.106}$$

where n is the refractive index in observation space.

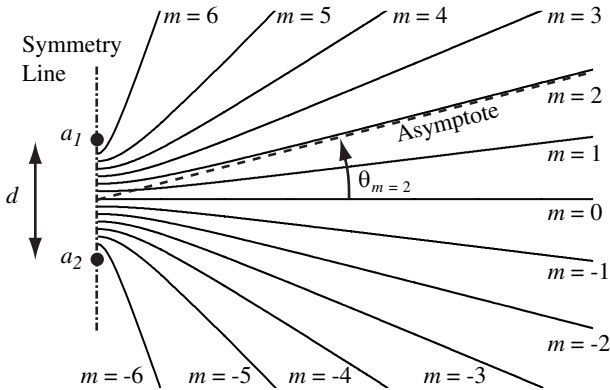


Fig. 8.30. Hyperboloidal fringe field generated from a pair of coherent point sources. Fringe surfaces are hyperboloids of revolution around the symmetry line that passes through both points. The fringe order m increases or decreases away from the plane of zero OPD_0 , which is the $m = 0$ fringe. Asymptotic lines are given by

$$\tan \theta_m = \frac{m\lambda}{\sqrt{d^2 - m^2\lambda^2}} .$$

8.2.1.2 Polychromatic point source

With a polychromatic point emitter in the source space, two-point-source interference produces hyperboloidal fringe fields in observation space for each frequency component of the source’s power spectrum. Portions of the fringe field in the observation space with high OPD_0 between the two sources exhibit reduced contrast due to wavelength-dependent fringe variations. Fringe visibility depends on where the observation is made. For example, if $\Delta\nu = c/4\bar{\lambda}$ and $n = 1$, fringe visibility near and beyond the $m = \pm 4$ fringes in Fig. 8.30 would be very poor.

8.2.1.3 Quasimonochromatic extended source

Quasimonochromatic extended sources can be treated as a collection of independent emitters oscillating with wavelength $\bar{\lambda}$.¹ This ideal case might be a good approximation for a highly filtered tungsten lamp or a spectroscopic lamp. The total fringe field in observation space is simply the addition of fringe fields from individual source points. The net effect is usually a decrease in fringe visibility, but the characteristics of the visibility reduction depend greatly on the geometry of the interferometer.

8.2.1.4 Polychromatic extended source

Analysis of systems with a polychromatic extended source must take into account both the temporal bandwidth of the source and its finite extent. In the simplest case, the polychromatic nature can be considered separately. That is, first consider the source as a collection of point emitters at $\bar{\lambda}$, then estimate the additional visibility reduction due to the temporal bandwidth. However, some complicated sources may not be well described in this way. A more complete discussion of this topic is provided in Section 8.3

8.2.2 General procedure for determining fringe localization

There are three basic steps in assessing fringe localization for most interferometers, which are:

- 1) *Determine the fringe period and position in the observation plane given an ideal, monochromatic point source in the source plane;*
- 2) *Integrate fringe patterns resulting from positioning the point source over the spatial extent of the distributed source, and/or from the separate frequencies in the power spectrum; and*
- 3) *Define regions in the observation space where V is large enough ($V > 0.2$) to detect fringes.*

Usually, results from Step (2) produce a relationship that can be expanded into a closed-form solution for fringe visibility in observation space as a function of source properties and interferometer geometry. If a closed-form solution can't be found, a computer simulation can provide the necessary information.

1. In order for the quasimonochromatic assumption to hold over the observation space, Eq. (8.106) must be satisfied.

8.2.3 Application of fringe localization techniques to specific interferometers

In this section, several specific interferometers are analyzed with respect to fringe visibility for different types of sources. The interferometers examined include Young’s double pinhole interferometer (YDPI), the plane parallel plate (PPP), a wedged plate, thin films, Fizeau interferometers, the Michelson interferometer, the Twyman-Green interferometer, Lloyd’s mirror and Fresnel’s biprism.

8.2.3.1 Young’s double pinhole interferometer (YDPI)

If a monochromatic point source is used with the YDPI displayed in Fig. 8.1, pinholes produce a coherent pair of source points. The fringes are unlocalized in the entire observation space, which is the entire region to the right of the pinhole plane.

If the YDPI is used with an on-axis polychromatic point source, it produces a region of fringe localization close to the optical axis, as shown in Fig. 8.31, which is specified by Eq. (8.48). The region of fringe localization is bounded by the fringes corresponding to $\pm OPD_0$ that produce the minimum acceptable visibility V .

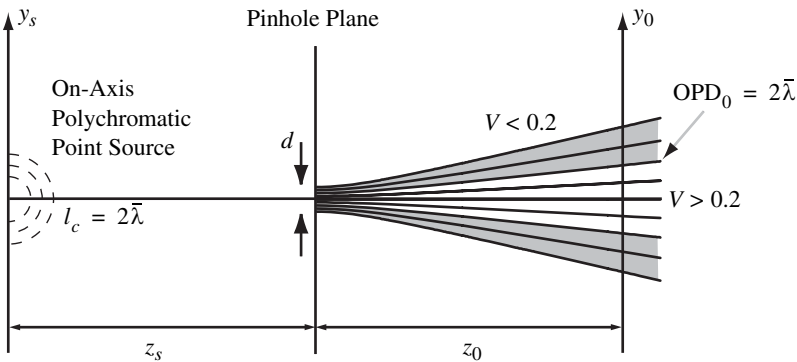


Fig. 8.31. Localization of a YDPI with an on-axis polychromatic point source. A YDPI is used with a polychromatic point source exhibiting a coherence length of $l_c = 2\lambda$. Fringes are localized near the axis where $V > 0.2$, which occurs until approximately $OPD_0 = 2\lambda$. The gray region indicates where $V < 0.2$.

Spatial coherence of a YDPI with an extended quasimonochromatic source is examined in detail in Section 8.1.5. Visibility of the fringes in observation space

is a function of the wavelength, the size of the source and its distance to the pinhole plane. Larger sources and smaller distances produce lower visibility. Unlike temporal coherence with the YDPI, visibility due to an extended source is not a function of position in the observation space. Equation (8.81) provides an explicit relationship between the source distribution, interferometer parameters, and visibility.

8.2.3.2 Plane parallel plate (PPP)

Perhaps the simplest form of interferometer is the plane parallel plate, as shown in Fig. 8.32. Light from point source a , which is a distance l in front of the plate, reflects from the two surfaces of the plate. The reflection from the front surface appears to originate from virtual source point a_1 a distance l behind the plate. Part of the wave refracts into the plate and reflects off the back surface. When this wave exits the plate from the front, it appears to originate from virtual source point a_2 . The coherent point-source pair a_1 and a_2 that are produced by the reflections are separated by the geometrical distance $d = 2t/n$. They have a symmetry line through a_1a_2 perpendicular to the surface of the plate, and they produce a hyperboloidal fringe field that can be viewed from the front of the plate.

If a is a monochromatic point source, fringe visibility is unity everywhere in the observation space to the right of the plate, and the fringes are unlocalized. (For practical experiments, the point source is isolated from the observation region with a beamsplitter or other component.) The glass causes an additional delay between the two waves that must be considered if the source is not monochromatic. Instead of the geometric separation $2t/n$ a delay caused by the maximum $OPD_0 = 2nt$ must be used in the coherence calculations. Therefore, fringes can be observed everywhere in the observation region with good visibility if $2nt \ll c/\Delta\nu$, where t is the thickness of the plate, n is its index of refraction and $\Delta\nu$ is the bandwidth of the source power spectrum. If a is a polychromatic point source with $2nt > c/\Delta\nu$, visibility reduces as the fringe order increases, because of increasing OPD_0 . Fringes of adequate visibility can be observed if the observation plane intersects fringes where $OPD_0 = m\bar{\lambda} < c/\Delta\nu$. Notice that, since the maximum OPD_0 is along the symmetry line, lowest visibility is observed in the center of the fringe pattern.

For a quasimonochromatic extended source, the PPP produces a reflection of each source point for each surface of the plate, as shown in Fig. 8.33. Each virtual source pair produces a hyperboloidal fringe pattern in observation space. The individual fringe patterns are offset in the vertical direction according to the position of the point of origin in the source. The fringe patterns add to produce the total light pattern. If the source length is greater than the fringe spacing, the hyperboloidal features wash out, and zero visibility is measured in observation region close to the plate.

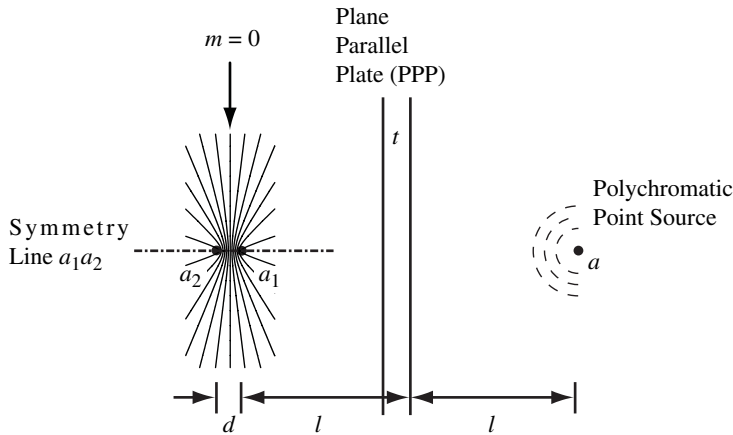


Fig. 8.32. PPP with polychromatic point-source illumination. Point source a is shown positioned a distance l in front of a plane parallel plate (PPP). A coherent pair of virtual point sources a_1a_2 is created from the two Fresnel surface reflections of the plate. The coherent point-source pair produces a hyperboloidal fringe field in the observation space to the right of the plate. If a is polychromatic, fringe visibility reduces as fringe order number m increases.

In the particular geometry of Fig. 8.33, a useful characteristic of the fringe patterns is that asymptotes for each fringe order m are parallel. This condition exists because the virtual images of the source, a_1b_1 and a_2b_2 , are parallel to each other. If the observation plane is displaced far enough to the right of the plate so that the distance between adjacent fringe orders is larger than the length of the source, fringes with acceptable visibility can be measured. A practical way of achieving a long effective distance between the source pairs and the observation plane is to use a lens positioned as shown in Fig. 8.34, where the observation plane is at the back focus of a lens. In this configuration, the image of the observation plane is at infinity with respect to the plate. Each fringe order at angle $\theta_m = \text{asin}(m\lambda/d) = \text{asin}(m\lambda/2nt)$ focuses to a circle in the observation plane, due to the symmetry of the hyperboloidal fringe field. The radius of the circle is given by $r_m = f \cot\theta_m$. The combination of all circles yields a concentric fringe pattern. Like with two-point-source interference, the maximum OPD_0 of this pattern corresponds to the innermost fringe at the center of the circular pattern. Any significant width of the source power spectrum washes out the fringe in this central region, due to temporal coherence effects. The circular fringes shown in Fig. 8.34 are called *Haidinger's fringes*. Because the fringes are only observed at a sufficiently large distance from the plate, the fringes are said to be *localized at infinity*. Since

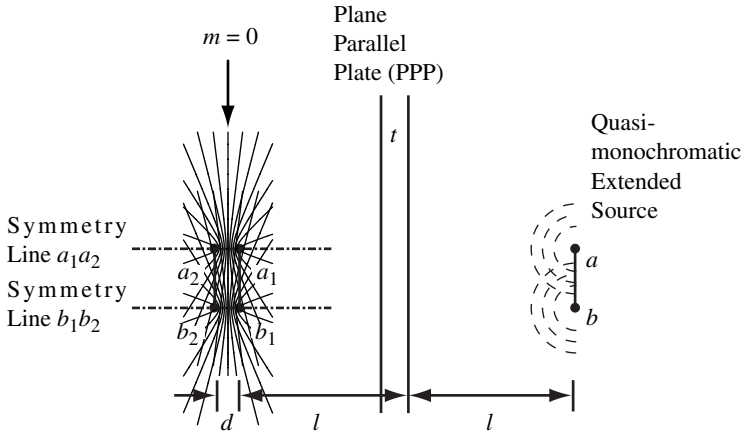


Fig. 8.33. PPP with quasimonochromatic extended-source illumination. Extended source ab is shown positioned a distance l in front of a plane parallel plate (PPP). A coherent pair of virtual point sources is created for each point in the extended source from the two Fresnel surface reflections of the plate. Each coherent point-source pair produces a hyperboloidal fringe field in the observation space. The hyperboloids are shifted with respect to each other according to the source-point position. A reduction in visibility is observed in the observation space, due to overlap of the fringes.

each fringe results from many source points, but only from a specific angular range, the fringes are also called *fringes of equal inclination*.

8.2.3.3 Wedged plate, thin films and Fizeau interferometers

Closely related to the plane parallel plate is the simple wedge, as shown in Fig. 8.35 with polychromatic point-source illumination. The virtual point sources a_1 and a_2 now exhibit a tilted symmetry line. When the wedged plate is illuminated with an extended, quasimonochromatic source, as shown in Fig. 8.36, two source images are formed. Each coherent pair in the images produces hyperbolic fringes. Unlike the cases studied so far, the images are not parallel, and the symmetry lines between coherent pairs are tilted with respect to each other. In addition, the distances d between coherent pairs are not equal, unlike with the PPP. This angled geometry complicates the discussion of fringe localization. The shifts, tilts and spacing differences produce hyperboloidal fringe fields that are also tilted, shifted and scaled. It is difficult to find regions of adequate visibility with this relatively thick-wedged-plate geometry.

More commonly, the wedge under test is very thin with respect to the distance from the source to the top surface. Effectively, the wedge is a *thin film*. In addition, the wedge angle is small. Figure 8.37 shows the geometry for a thin film illumi-

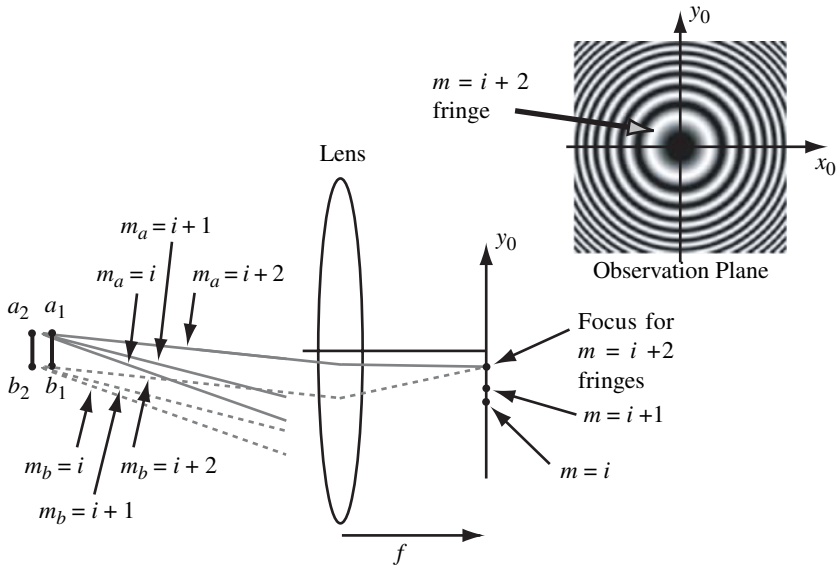


Fig. 8.34. Haidinger's fringes. *Haidinger's fringes* or *fringes of equal inclination* are formed when images of the extended source are parallel to each other in observation space. Although fringes have low visibility near the source images, a lens with focal length f can be used to form the characteristic concentric-ring fringe pattern with good visibility. The pattern forms because fringe asymptotes of corresponding fringe orders are parallel. Since the lens effectively sets the observation at an infinite conjugate with respect to the source images, the fringes are said to be *localized at infinity*.

nated by an extended quasimonochromatic source. In this case, the coherent pairs form hyperboloidal fringe fields, like those shown in the magnified view of the a_1a_2 source pair. The $m_a = 0$ fringe line is directed toward the film, where the $m_b = 0$ fringe intersects with it. A similar observation is made concerning all other fringe orders. That is, like fringe orders intersect at or near the film, which produces high-contrast fringes. Between the source images and the film, the fringes are overlapped in various ways, and the fringe visibility is low or zero. Likewise, visibility is low beyond the film. Therefore, the fringes are *localized at the thin film*.

Figure 8.38 shows the result of a simulation for the geometry of Fig. 8.36 with film thickness $t = 1 \mu\text{m}$, wedge angle = 0.1° , wavelength = 500 nm , plate refractive index $n = 1.5$, $l = 200 \text{ mm}$, and separation between source points a and b of 10 mm . The fringe fields from coherent pairs are tilted and displaced with respect to each other. However, the fringe fields converge with like orders overlapping at the wedge. That is, the m^{th} order of the fringe field from source pair a_1a_2 overlaps the m^{th} order of the fringe field from all other source pairs

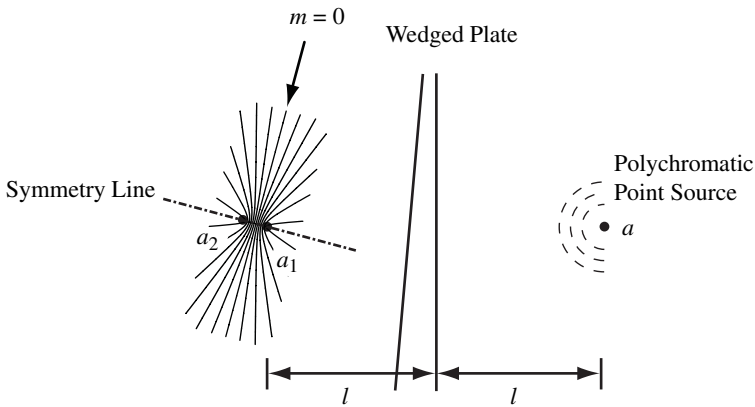


Fig. 8.35. Wedged plate with polychromatic point-source illumination. Point source a is shown positioned a distance l to the right of a wedged plate. A coherent pair of virtual point sources a_1a_2 is created from the two Fresnel surface reflections of the plate. The coherent point-source pair produces a tilted hyperboloidal fringe field in the observation space. If a is polychromatic, fringe visibility reduces as fringe order number m increases.

distributed along the images. Therefore, the visibility is maximum at the wedge and the fringes are localized at the thin film. An external optical system or the eye of an observer can be used to reimage the wedge in order to detect high visibility fringes. Notice that the $m = 0$ fringe of each coherent pair intersects the wedge. Therefore, high visibility is obtained near the intersection of the $m = 0$ fringe and the wedge, even with a broad bandwidth light source. Also, the fringe pattern consists of straight-line, equally spaced fringes near the $m = 0$ fringe, like what is observed from the basic hyperbola pattern shown in Fig. 8.30. Over a wider expanse of fringe orders, the *localization surface* is a sphere with diameter l .

If the wedge is not a simple linear function, the fringes are not distributed in straight, equally spaced lines. For example, the practical geometry shown in Fig. 8.39 shows a *Fizeau interferometer*, where an air wedge is formed between a test piece and a reference flat. Light from the extended source is blocked from directly reaching the observation plane by a beamsplitter. For most of the test surface, the wedge is linear and fringes are straight and equally spaced. In the region of the bump, fringes move to positions of equal OPD_0 , which are toward the open side of the wedge. The otherwise straight fringe lines are distorted in a shape characteristic of the bump.

The OPD_0 between real-ray reflections from wedge surfaces can be used to find locations of bright and dark fringes along the wedge. An additional factor from

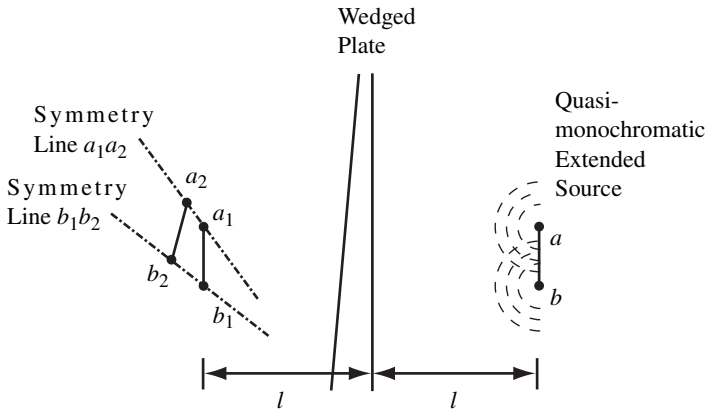


Fig. 8.36. Wedged plate with quasimonochromatic extended-source illumination. The extended source is positioned a distance l to the right of a wedged plate. A coherent pair of virtual point sources is created for each point in the source from the two Fresnel surface reflections of the plate. Each coherent point-source pair produces a hyperboloidal fringe field in the observation space.

the phase ϕ of the surface reflections must also be taken into account. When $OPD_0 = \left(m + \frac{\phi}{2\pi}\right)\lambda$, a bright fringe is observed, and when $OPD_0 = \left(m + \frac{\phi}{2} + \frac{\phi}{2\pi}\right)\lambda$, a dark fringe is observed. Since all like fringe orders overlap at the film, as shown in Fig. 8.38, it is sufficient to analyze the ray emitted from the central portion of the source. Derivation of OPD_0 is aided by the magnified view of the wedge shown in Fig. 8.40. Considering real-ray path portions \overline{AB} , \overline{CB} and \overline{AD} ,

$$\begin{aligned}
 OPD_0 &= n_2(\overline{AB} + \overline{CB}) - n_1\overline{AD} \\
 &= n_2 \frac{2t(x)}{\cos\theta_2} - n_1 \overline{AC} \sin\theta_1 \\
 &= n_2 \frac{2t(x)}{\cos\theta_2} - n_1 2t(x) \tan\theta_2 \sin\theta_1 \\
 &= n_2 \frac{2t(x)}{\cos\theta_2} - n_1 2t(x) \frac{\sin\theta_2}{\cos\theta_2} \frac{n_2 \sin\theta_2}{n_1} \quad (8.107) \\
 &= \frac{2n_2 t(x)}{\cos\theta_2} [1 - \sin^2\theta_2] \\
 &= 2n_2 t(x) \cos\theta_2
 \end{aligned}$$

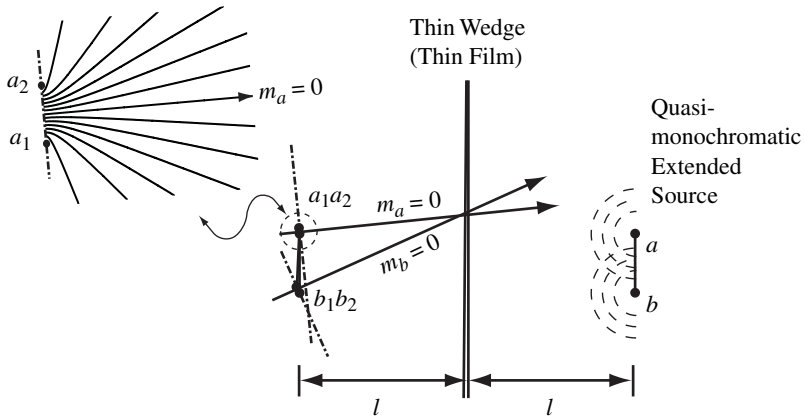


Fig. 8.37. Thin wedge (thin film) with quasimonochromatic extended-source illumination. The extended source is positioned a distance l to the right of a thin film. A coherent pair of virtual point sources is created for each point in the source from the two Fresnel surface reflections of the film. Each coherent point-source pair produces a hyperboloidal fringe field in the observation space to the right of the film. Like orders intersect at the film, so the fringes are *localized at the thin film*.

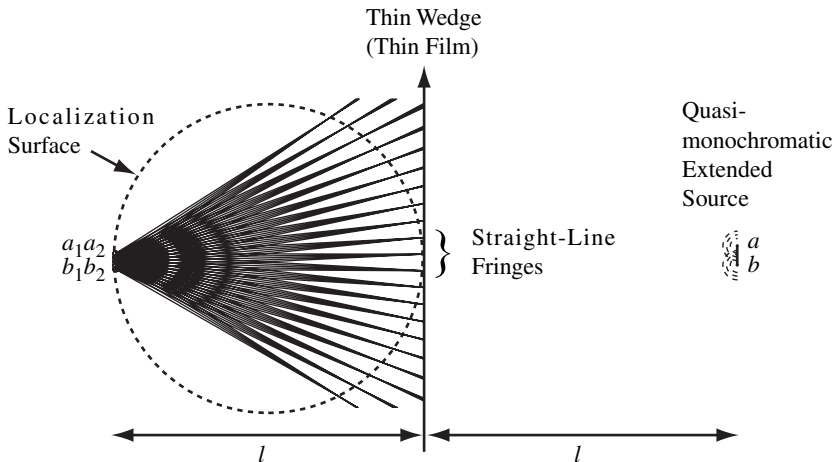


Fig. 8.38. Result of a simulation for the geometry of Fig. 8.37 with film thickness $t = 1 \mu\text{m}$, wedge angle $= 0.1^\circ$, wavelength $= 500 \text{ nm}$, plate refractive index $n = 1.5$, $l = 200 \text{ mm}$, and separation between source points a and b of 10 mm .

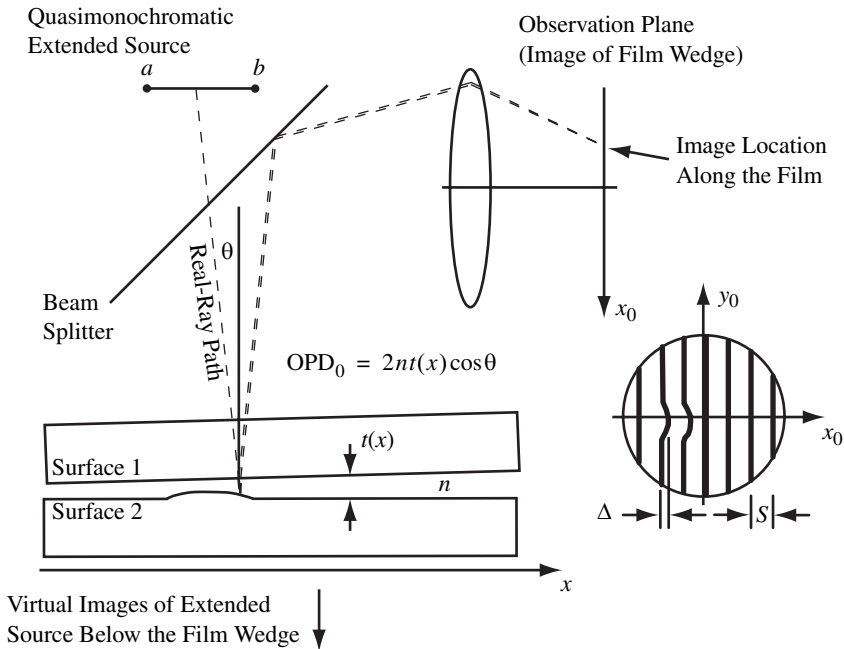


Fig. 8.39. A typical Fizeau interferometer. Fringes of equal thickness indicate contours of constant $t(x)$ and are localized at wedge. The height of the defect, as indicated by the fringe deviation, is approximately $\frac{\lambda\Delta}{2S}$.

where n_1 is the refractive index of the incident medium, n_2 is the refractive index of the plate, $t(x)$ is the thickness of the plate at the point of reflection, and θ_2 is the angle of refraction inside the wedge. For a bright fringe at the observation plane, $OPD_0 = m\lambda + \frac{\lambda\phi}{2\pi}$, where ϕ is the phase difference between Fresnel reflections from the surfaces. For example, if $n_1 = 1$, $n_2 = 1.5$ and $n_3 = 1$, $\phi = \pi$. Notice that OPD_0 in Eq. (8.107) is directly proportional to the thickness of the wedge $t(x)$. Therefore, fringes map out contours of constant wedge thickness. These fringes are called *fringes of equal thickness*.

For the case of nearly normal incidence ($\theta_1 \sim 0$) and $n = 1$, $OPD_0 = 2t(x)$, and each fringe represents one-half wavelength multiples of surface departure. For the bump shown in Fig. 8.39, the height of the defect is approximately the fractional fringe deviation multiplied by the surface departure between fringe lines. That is, $height = \frac{\lambda\Delta}{2S}$. In order to see the fringes, the wedge is imaged onto an observation plane. (Fringes can be imaged through optical systems, just like light reflected from or transmitted through real objects.) In a practical experiment, the observation plane can be a CCD array. The lens and observation plane can also be the eye and retina, respectively, of the observer if the fringes are viewed directly.

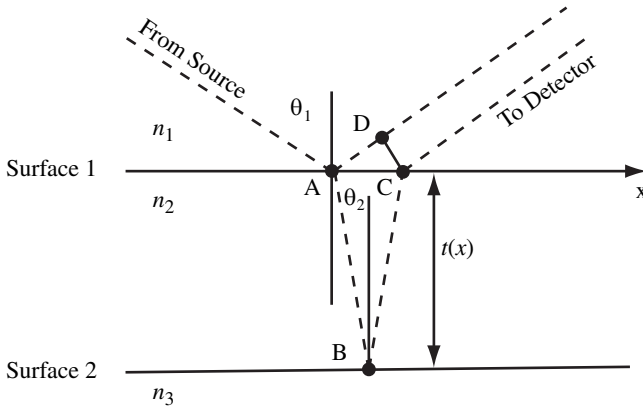


Fig. 8.40. Magnified view of real-ray intersection at the wedged plate for calculation of OPD_0 .

Notice that localization of the fringes at the wedge is an advantage to the observer. Spurious fringes caused by reflections from other surfaces typically have very low visibility and do not distract from the measurement of interest.

The analysis outlined in Fig. 8.40 can be extended to transmitted fringes, where, except for a phase shift due to the extra reflection, the OPD_0 is defined by Eq. (8.107). With a simple glass wedge, visibility of the transmitted fringes is much lower than the reflected fringes, due to the higher transmission of the first beam.

Naturally occurring thin films, like soap bubbles and oil films, produce colorful interference effects. In this case, constructive interference for different wavelengths is spread out spatially from small changes in \pm or θ_2 .

The simplified example of a thin wedge is a very useful demonstration of fringe localization, but it is also useful to develop a general technique to determine fringe localization in other types of interferometers that use quasimonochromatic extended sources. A simple construction can be used to determine fringe localization for two-beam interferometers using quasimonochromatic extended sources, as shown in Fig. 8.41. Steps in the construction are:

- 1) Determine positions of the two source images in observation space;
- 2) Locate coherent pairs a_1a_2 and b_1b_2 on ends of the source images;
- 3) Draw symmetry lines between a_1 and a_2 and between b_1 and b_2 ;
- 4) Construct the perpendicular bisectors to each symmetry line. (These bisectors are the $m = 0$ fringes for the coherent pairs.); and

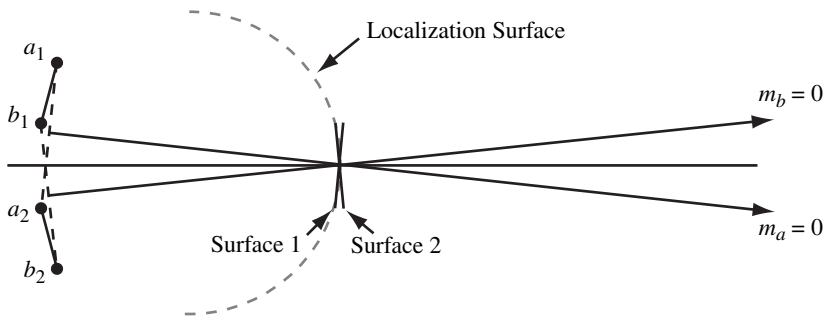


Fig. 8.41. Localization procedure for extended sources. Intersection of $m = 0$ fringes from endpoints of the source images from surfaces of the wedge determines the center of the localization region. Fringes are localized on a curved surface.

- 5) Find the intersection of the $m = 0$ fringes. This line in space is the center of the localized fringes. The localization surface is a sphere with diameter equal to the distance between the center of the source images and the intersection of the $m = 0$ fringes.

8.2.3.4 Michelson Interferometer

The basic Michelson interferometer is shown in Fig. 8.42. An extended source is used to illuminate a two-beam interferometer through a beam splitter BS. At first, we will analyze the Michelson interferometer using a quasimonochromatic extended source, with endpoints a and b . After the basic concepts of fringe localization are explained, the analysis is extended to polychromatic extended sources.

The first beam of the interferometer is defined by the light reflected off BS and mirror M1, which then transmits through BS toward the observation plane. The total path for the first beam is $d_1 + d_3 + d_3 + d_4$. The second beam of the interferometer is defined by the light that passes through BS, reflects off mirror M2 and then reflects off BS toward the observation plane. The total path for the second beam is $d_1 + d_2 + d_2 + d_4$. An imaging system is used to look into the beam splitter and form images of the fringes on an observation plane.

An unfolded diagram of the interferometer is shown in Fig. 8.43. The source images are formed by reflections off the mirrors. The images a_1b_1 and a_2b_2 are as far behind the mirrors as the source is in front. As shown, the mirrors are parallel to the source, so the source images are also parallel to the source and each other. From the perspective of the observer, as shown in Fig. 8.44, the source images are exactly analogous to the geometry shown in Figs. 8.33 and 8.34, where an extended source is used with a PPP. In this case, a simple optical system in obser-

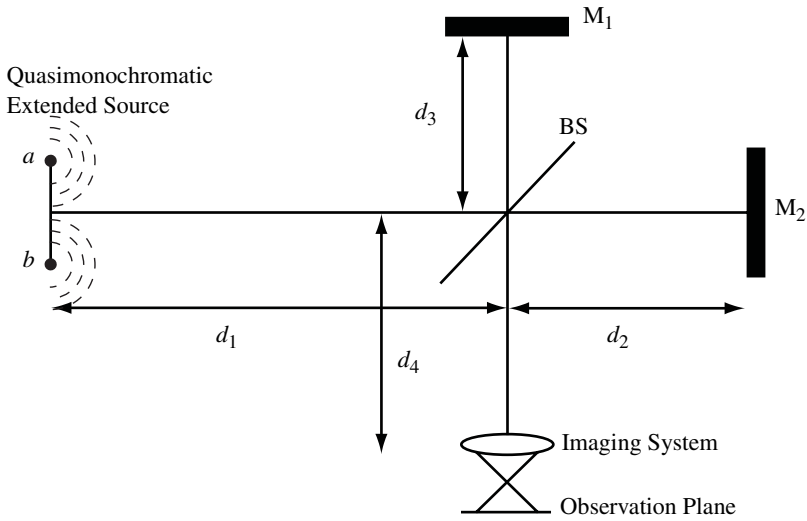


Fig. 8.42. A basic Michelson interferometer with a quasimonochromatic extended source.

vation space can be used to observe Haidinger’s fringes at its back focus plane, as shown in Fig. 8.34 for the PPP. Separation of the coherent pairs increases as the distance $2|d_3 - d_2|$ between source images increases, and more fringes are observed in the concentric pattern. Like with the plane parallel plate, maximum OPD_0 is found in the center of the pattern.

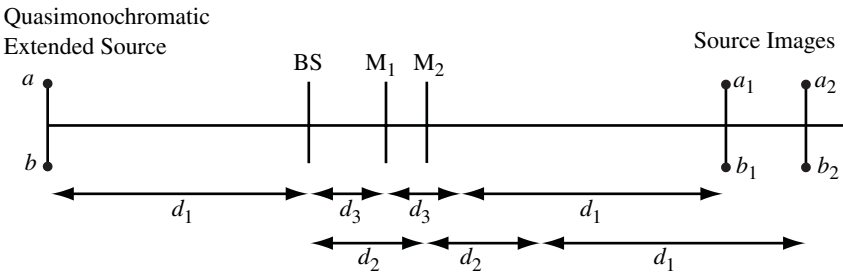


Fig. 8.43. An unfolded Michelson interferometer.

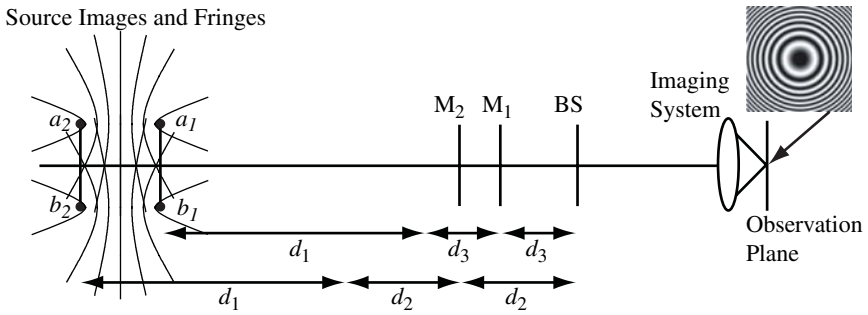


Fig. 8.44. The observation-space description of a Michelson interferometer with parallel mirrors. Haidinger fringes are observed in the back focal plane of the imaging system.

The Michelson interferometer can be configured so that the two mirrors are tilted, but not displaced, as shown in Fig. 8.45. The thin space between the mirrors in the unfolded system analogous to a Fizeau thin film. The source images are also tilted. Application of the construction described in Section 8.2.3.3 shows that the $m = 0$ fringes intersect at the mirrors, where the fringes are localized. If the mirrors are perfectly flat, straight and equally spaced fringes of equal thickness are observed when the mirrors are reimaged to the observation plane. As the tilt angle θ decreases, separations of the coherent pairs decrease, and fringe spacing increases. Fizeau-type fringes are observed if the mirrors are not flat, where the

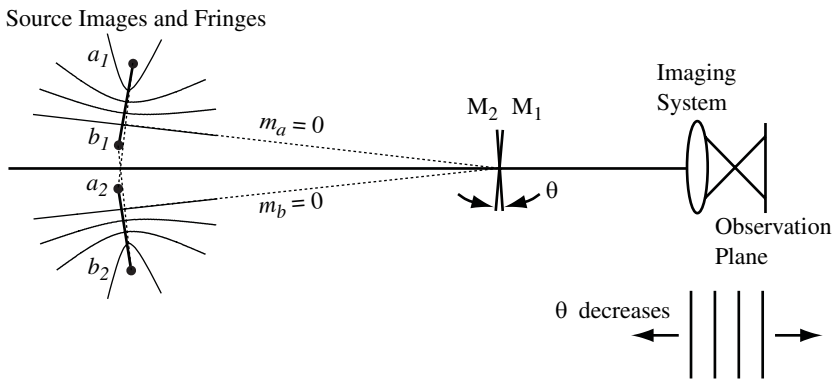


Fig. 8.45. The observation-space description of a Michelson interferometer with tilted mirrors. Fringes of equal thickness are observed when the mirrors are imaged onto the observation plane.

fringe deviation from a straight line is proportional to the surface feature. For the case of tilted mirrors with no displacement, as shown in Fig. 8.45, the $m = 0$ fringe intersects the mirrors. Therefore, if the source is polychromatic, good visibility can be observed until $OPD_0 \sim c/\Delta\nu$, where $\Delta\nu$ is the bandwidth of the source power spectrum. For example, if a white-light (400 nm to 700 nm) extended source is used, $\Delta\nu \sim 2.9 \times 10^{14}$ Hz, and $OPD_0 \sim 1 \times 10^{-6}$ m = 1 μ m before fringe visibility fades. Since the average wavelength is 550 nm, about two fringes ($m = \pm 2$) can be observed with good visibility on each side of the $m = 0$ fringe for a total of five fringes. These fringes are often called *white-light fringes*. Fringe shift due to the colored spectrum reduces fringe visibility for higher OPD_0 , as shown in Fig. 8.46.

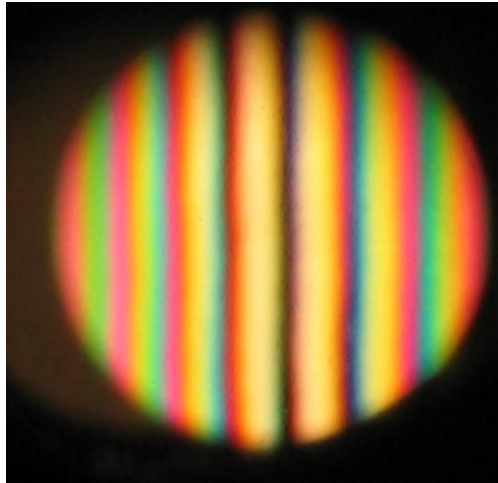


Fig. 8.46. White-light fringes.

If the two mirrors are tilted and displaced, as shown in Fig. 8.47, the center of the localization surface is now offset from the mirrors. If the offset is not large, high visibility fringes can be observed at the mirrors. Fringes intersecting the mirrors are of high order m and some curvature due to the hyperbolic shape is observed.

A practical consequence of the beam splitter in the Michelson interferometer is that one of the beams must pass through some thickness of glass, as shown in Fig. 8.48. The light from the source first transmits through the glass of BS and then reflects off its back side toward M1. Then, it passes through the glass twice more before reaching the observation plane. On the other hand, light from the source that transmits through BS and reflects off M2 only passes through the glass once. If the optical glass used to construct BS exhibits significant dispersion, OPD_0 contains additional wavelength-dependent terms that drastically reduce visibility

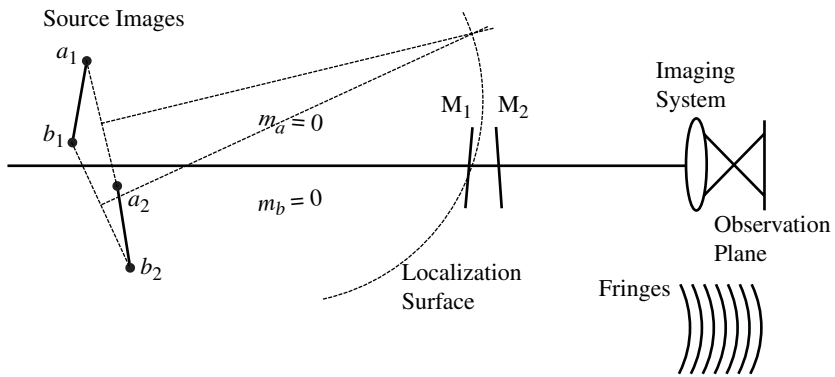


Fig. 8.47. The observation-space description of a Michelson interferometer with tilted and displaced mirrors. Fringes with hyperbolic curves are observed when the mirrors are imaged onto the observation plane.

when the interferometer is used with polychromatic sources. In order to balance OPD_0 and remove effects of the dispersion, a compensator plate C is placed in the second beam of the interferometer before M_2 . The extra glass is fabricated from the same glass as BS , and it is the same thickness. With the compensator, the two beams pass through the same type and thickness of glass before reaching the observation plane.

A summary of fringe localization effects with a Michelson interferometer is shown pictorially in Fig. 8.49, where path difference increases outwardly from the center. The top row displays the case where mirrors M_1 and M_2 are parallel to a quasimonochromatic extended source, which forms Haidinger’s fringes. As the path difference increases, fringes expand outwardly from the center. Note that Haidinger’s fringes are not observable with a white-light source. The bottom row displays the case where the mirrors are tilted and displaced. With no displacement, fringes of equal thickness are observed. If the $m = 0$ fringe can be imaged onto the observation plane, white-light fringes can be produced. As the path difference increases with increasing displacement, the center of the fringe localization surface offsets from the mirrors, and curved fringes are observed. Larger mirror displacements offset the center of the localization surface so much that fringes are no longer observable.

8.2.3.5 Twyman-Green Interferometer

In the Twyman-Green Interferometer displayed in Fig. 8.12, the coherent points are effectively located at infinity. The observation space begins as soon as the beams reflected from each mirror combine. This combination also occurs throughout any optical system that may be used to image the mirrors onto a

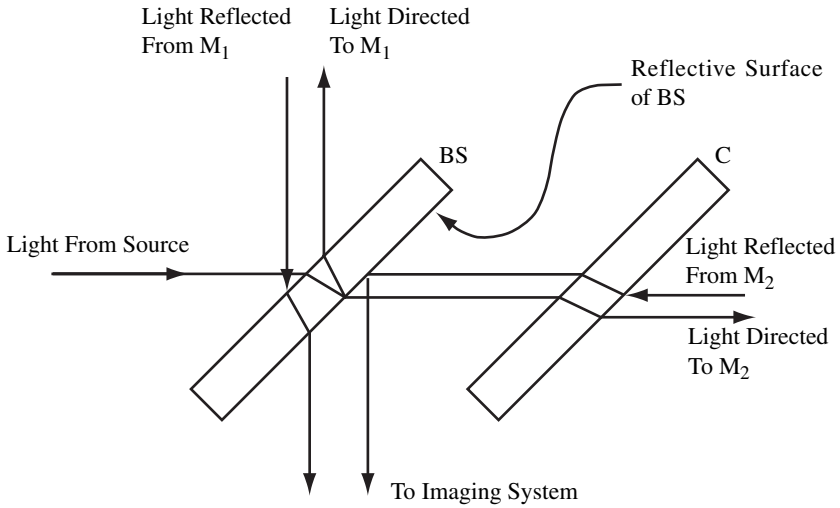


Fig. 8.48. Compensator plate C used to cancel dispersion from the glass of BS. With C in the system, each beam of the interferometer experiences three passes through the same thickness of glass.

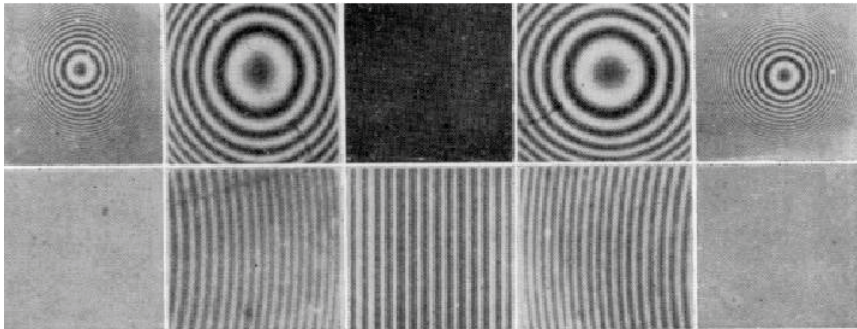


Fig. 8.49. Summary of fringe localization in a Michelson interferometer. Zero separation of the mirrors is in the center of each row. OPD_0 increases positively to the right and negatively to the left away from the center. The upper row contains fringes of equal inclination (Haidinger's fringes). The bottom row contains fringes of equal thickness.

detector plane. If a point source is used where Eq. (8.106) is satisfied, high visibility fringes are observed throughout the observation region. If a polychro-

matic point source is used where Eq. (8.106) is not satisfied, the change in visibility versus OPD_0 is given by Eq. (8.48). Since OPD_0 is a function of the mirror separation, visibility can change if the mirrors are moved. A Twyman-Green interferometer is generally not used with extended sources. In that case, the geometry must be analyzed like a Michelson interferometer.

8.2.3.6 Lloyd's mirror

A simple mirror can be used as an interferometer with a polychromatic point source. For example, the Lloyd's mirror interferometer produces a virtual image of the point source, which is located as far below the mirror surface as the real source is above it, as shown in Fig. 8.50. If the total separation is d and the distance between the source plane and the observation plane is z_0 , Eq. (8.50) can be used to predict the region of fringe localization with hyperbolic boundaries. One important difference in the fringe pattern is that the central fringe is dark, instead of bright, if there is a π phase change in reflection off the mirror. Lloyd's mirror with a single point source is similar to the YDPI, in that a hyperboloidal field is generated from two-point-source interference, except for the dark fringe at the center of the field. Straight-line cosine fringes are observed near the optical axis in the observation plane with period $\Lambda = \lambda z_0 / d$.

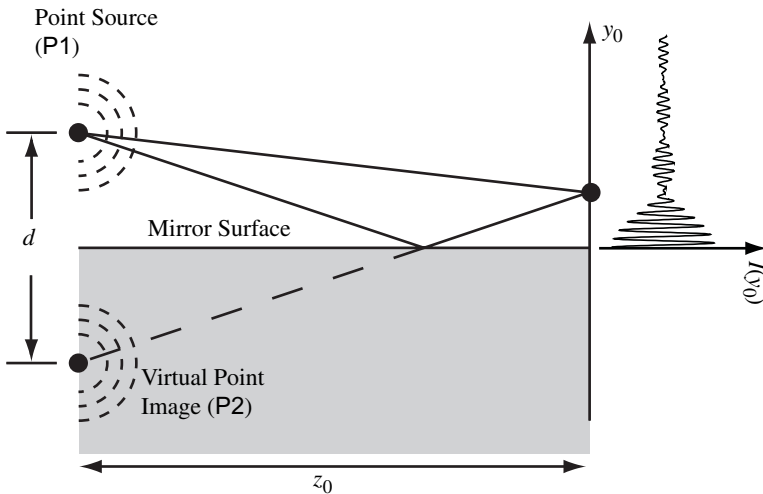


Fig. 8.50. Lloyd's mirror with a polychromatic point source. Fringe visibility is a function of y_0 that depends on the source power spectrum.

When the Lloyd's mirror is used with a quasimonochromatic extended source, the distance between individual points and the mirror is not constant, as shown in Fig. 8.51. Therefore, distances between source points and their coherent images are not constant, and the fringe spacing is a function of the source-point position y_s . For simplicity, a perfect mirror is assumed without any loss on reflection. Following a development similar to that used with Eq. (8.74), the integration over the source spatial distribution is

$$\begin{aligned}
 I(y_0) &= I_0 \left[2K + 2K \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m_R(x_s, y_s) \cos(kOPD_0) dx_s dy_s \right] \\
 &= 2KI_0 + 2KI_0 \operatorname{Re} \left\{ \mathbf{F}^{-1} \left[\frac{2y_0}{\lambda z_0} \int_{-\infty}^{\infty} m_R(x_s, y_s) dx_s \right] \right\}, \quad (8.108)
 \end{aligned}$$

where K is the fraction of the light emitted by the each source point that reaches the observation plane. If the mean distance from the source distribution to the mirror is \bar{y}_s , an argument similar to that used for Eqs. (8.49) through (8.53) results in

$$\begin{aligned}
 I(y_0) &= 2KI_0 + 2KI_0 \left| \mathbf{F}^{-1} \left[\frac{2y_0}{\lambda z_0} \int_{-\infty}^{\infty} f_R(x_s, y_s) dx_s \right] \right| \cos \left[2\pi y_0 \frac{2\bar{y}_s}{\lambda z_0} + \beta(y_0) \right], \\
 &= 2KI_L + 2KI_L \mu(y_0) \cos \left[2\pi y_0 \frac{2\bar{y}_s}{\lambda z_0} + \beta(y_0) \right], \quad (8.109)
 \end{aligned}$$

where

$$f_R(y_s) = m_R(x_s, y_s - \bar{y}_s), \quad (8.110)$$

$$\mu(y_0) = \left| \mathbf{F}^{-1} \left[\frac{2y_0}{\lambda z_0} \int_{-\infty}^{\infty} f_R(x_s, y_s) dx_s \right] \right|, \quad (8.111)$$

and

$$\beta(y_0) = \arg \left\{ \mathbf{F}^{-1} \left[\frac{2y_0}{\lambda z_0} \int_{-\infty}^{\infty} f_R(x_s, y_s) dx_s \right] \right\} \quad (8.112)$$

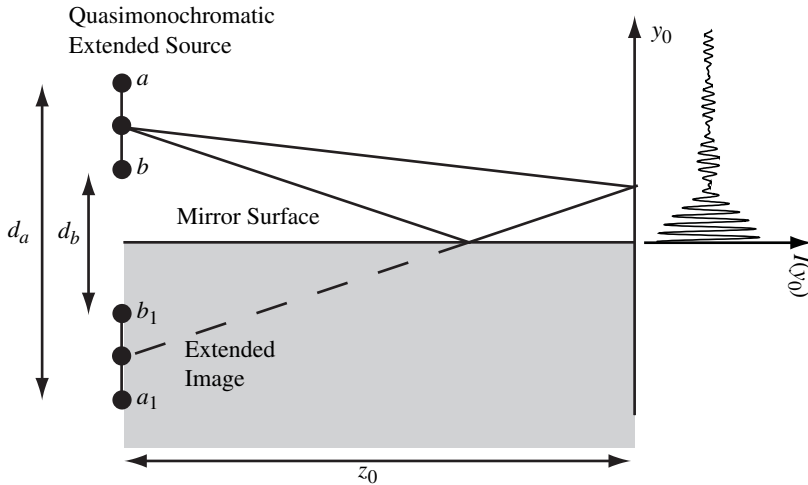


Fig. 8.51. Lloyd's mirror with an extended quasimonochromatic source. Because different positions on the source produce different separations of the source images, the fringe visibility varies as a function of y_0 .

Notice that visibility is equal to $\mu(y_0)$, which is a function of the observation-space coordinates. This behavior is dramatically different than the result obtained when an extended quasimonochromatic source is used with a YDPI, where the visibility is constant through the observation space. In fact, with a rectangular source distribution, Lloyd's mirror fringes may appear more like those shown in Fig. 8.2(C).

8.2.3.7 Fresnel's biprism

Yet another variation is Fresnel's biprism, as shown in Fig. 8.52 with a polychromatic point source, where two virtual point images a_1 and a_2 are produced by an offset pair of prisms with refractive index n . If the reflections from the prism faces are negligible, then the two virtual point images, one from each prism, produce an interference pattern in much the same way as the YDPI. The distance z_0 in Eq. (8.28) is measured from the virtual source plane, which is a distance nl behind the prisms.

Fresnel's biprism with an extended quasimonochromatic source is similar to Lloyd's mirror, except that two virtual images of the source are formed, and the parities of the images are not reversed, as shown in Fig 8.53. That is, the virtual sources are in the same orientation with respect to the y axis, where a_1b_1 and a_2b_2 are upright. (In the Lloyd's mirror, ab interferes with the reversed b_1a_1 .) In Fresnel's biprism, the distance d is a constant between coherent pairs, and the fringe spacing from coherent pairs is not a function of the source-point position.

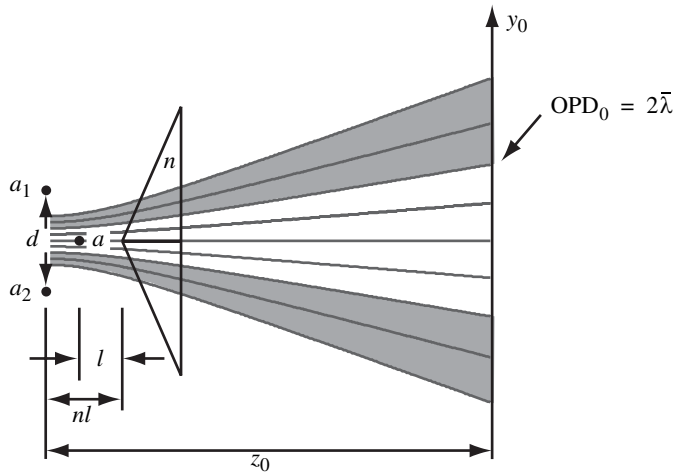


Fig. 8.52. Fresnel's biprism with a polychromatic point source.

However, the offset in source-pair position produces an offset in the fringe position, which is similar to the YDPI with an extended, quasimonochromatic source. The fringe contrast for Fresnel's biprism is not a function of the observation-space position y_0 , but it is a function of z_0 , λ and d . The net result of applying the procedure listed in Section 8.2.2 is that the visibility of the fringes is found from the coherence factor calculated from Eq. (8.76), where d is the spacing between coherent pairs and z_0 is the distance between the source images and the observation plane.

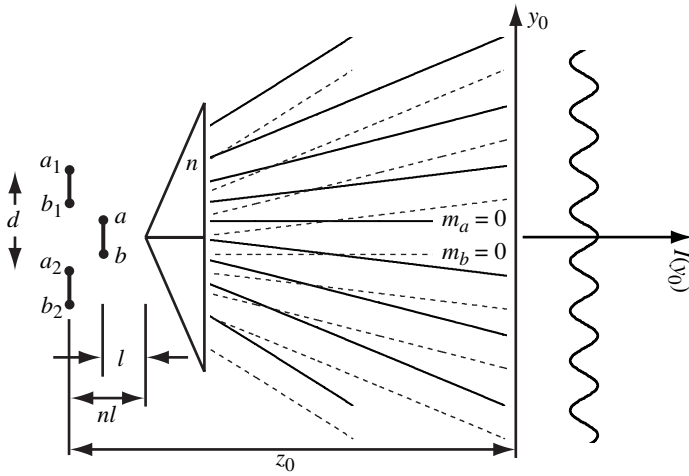


Fig. 8.53. Fresnel's biprism with an extended quasimonochromatic source. Since the parity of the source images is not reversed, fringes formed from coherent pairs have the same period, but different offsets. A uniform reduction in fringe visibility is observed in the observation space.

8.3 Advanced Topics

8.3.1 The Cross Spectral Density

In Section 8.1.5, a quasi-monochromatic distributed source provides the basis for the derivation of spatial coherence. This section explores the case where the quasimonochromatic restriction is removed.

It is reasonable to expect that a source with a finite-width power spectrum produces correlation that depends on temporal frequency. For example, results with a point source are listed in Section 8.1.4. More generally, the behavior of the mutual coherence versus frequency is given by the *cross spectral density*, which is the Fourier transform of the mutual coherence. That is,

$$\hat{\Gamma}(\mathbf{r}_1, \mathbf{r}_2; \nu) = \int_{-\infty}^{\infty} \Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau) e^{-j2\pi\nu\tau} d\tau, \quad (8.113)$$

Consequently, the mutual coherence function is the inverse Fourier transform of the cross spectral density. That is

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau) = \int_{-\infty}^{\infty} \hat{\Gamma}(\mathbf{r}_1, \mathbf{r}_2; \nu) e^{j2\pi\nu\tau} d\nu, \quad (8.114)$$

In order to determine the cross spectral density, we begin by finding an expression for the mutual coherence function.

Returning to the Young's double pinhole experiment, we consider the most general case of an extended polychromatic source. Section 8.1.4 demonstrated that, for a single polychromatic point source, the interference is a pattern of cosine fringes modulated by a properly scaled Fourier transform of the power spectral density of the source. The width of this modulation function is determined by the bandwidth of the source. This result was found by calculating the frequency dependence of the fringe pattern resulting from the point source and integrating over optical frequency, as shown in Eq. (8.44).

Section 8.1.5 demonstrated that an off-axis source point results in a shift of the fringe pattern that is related to the distance off axis of the source point. Therefore, a polychromatic point source that is off axis will produce a shifted fringe pattern that is modulated by the Fourier transform of the source power spectrum, shown in Fig. 8.29. Finally, in order to determine the fringe pattern from a collection of independent polychromatic source points, we integrate the shifted and modulated fringe pattern over the spatial extent of the source, similar to Eq. (8.75)

We begin by rewriting Eq. (8.73), except that in this case, the optical frequency is explicitly included, and not considered as just an average wavelength, where,

$$\begin{aligned}
 & I(\tau, \nu; x_s, y_s) dx_s dy_s d\nu \\
 &= \left\{ I_0 K_1 + I_0 K_2 + 2I_0 \sqrt{K_1 K_2} \cos \left[2\pi \nu \left(\tau + \frac{y_s d}{cz_s} \right) \right] \right\} m_R(x_s, y_s) g(\nu) dx_s dy_s d\nu .
 \end{aligned} \tag{8.115}$$

where $kOPD_0$ has been expressed explicitly in terms of ν as $2\pi\nu\tau$. In most cases of practical interest, it can be assumed that the spectral content of the source does not change along the source position. The function $m_R(x_s, y_s)$ describes the shape of the source, and the power spectrum $g(\nu)$ describes the spectral content of the source. Integration over source dimension and frequency yields

$$\begin{aligned}
 & I(\tau) \\
 &= I_0 K_1 + I_0 K_2 + 2I_0 \sqrt{K_1 K_2} \int \int_0^\infty \int_{-\infty}^\infty g(\nu) m_R(x_s, y_s) \cos \left[2\pi \nu \left(\tau + \frac{y_s d}{cz_s} \right) \right] dx_s dy_s d\nu .
 \end{aligned} \tag{8.116}$$

Comparing Eq. (8.116) with Eq. (8.8), we see that

$$\text{Re}[\Gamma_{12}(\mathbf{r}_1, \mathbf{r}_2; \tau)] = \frac{I_0}{\frac{1}{2}cn\epsilon_{00}} \int \int_0^\infty \int_{-\infty}^\infty g(\nu) m_R(x_s, y_s) \cos \left[2\pi \nu \left(\tau + \frac{y_s d}{cz_s} \right) \right] dx_s dy_s d\nu . \tag{8.117}$$

Playing fast and loose with complex values (i.e. because we are using analytic functions), we get

$$\Gamma_{12}(\mathbf{r}_1, \mathbf{r}_2; \tau) = \frac{I_0}{\frac{1}{2}cn\epsilon_{00}} \int_0^\infty g(\nu) m_R(x_s, y_s) \exp \left[\pm j 2\pi \nu \left(\tau + \frac{y_s d}{cz_s} \right) \right] dx_s dy_s d\nu . \tag{8.118}$$

At this stage some mention should be made in regards to notation. Firstly, Eq. (8.118) is the mutual coherence function *in the plane of the pinholes*. In physical terms, it represents the ability of the light leaving two pinholes at locations \mathbf{r}_1 and \mathbf{r}_2 to interfere. In that respect, the functional dependence of the right hand side of

Eq. (8.118) on the variables \mathbf{r}_1 and \mathbf{r}_2 is implied in the variable d , where $d = \mathbf{r}_1 - \mathbf{r}_2$ or the pinhole separation. Remember that we have assumed statistics of the light field in the plane of the pinholes are stationary. That is, the ability to interfere of the light leaving the two pinholes is not dependent on the absolute location of the two pinholes, just their separation distance.

Equation (8.118) is a complete representation of the mutual coherence function in the plane of the pinholes, if we assumed that both pinholes are illuminated equally. Simplification of this result can be done in two distinct ways. We will begin by integrating over optical frequency ν , where

$$\Gamma_{12}(\tau) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} \int \int_{-\infty}^{\infty} m_R(x_s, y_s) \left\{ \int_0^{\infty} g(\nu) \exp \left[-j2\pi\nu \left(\tau + \frac{y_s d}{cz_s} \right) \right] d\nu \right\} dx_s dy_s . \quad (8.119)$$

Recognizing the bracketed expression as simply a Fourier transform with the transform variable of $\tau + \frac{y_s d}{cz_s}$, we arrive at the following result

$$\Gamma_{12}(\tau') = \frac{I_0}{\frac{1}{2}cn\epsilon_0} \int \int_{-\infty}^{\infty} m_R(x_s, y_s) \mathbf{F}_{\tau} [g(\nu) \text{step}(\nu)] dx_s dy_s , \quad (8.120)$$

where $\tau' = \tau + \frac{y_s d}{cz_s}$. If the power spectrum has a form given by Eq.(8.49), Eq.(8.120) can be rewritten

$$\Gamma_{12}(\tau') = \frac{I_0}{\frac{1}{2}cn\epsilon_0} \int \int_{-\infty}^{\infty} m_R(x_s, y_s) m_{12}(\tau') \exp \{ g [2\pi\bar{\nu}\tau' + \beta_{12}(\tau')] \} dx_s dy_s . \quad (8.121)$$

Looking at the real part of Eq.(8.121) yields the following physically satisfying result

$$\begin{aligned} I(\tau') &= \\ &= I_0 K_1 + I_0 K_2 + 2I_0 \sqrt{K_1 K_2} \int \int_{-\infty}^{\infty} m_R(x_s, y_s) m_{12}(\tau') \cos [2\pi\bar{\nu}\tau' + \beta_{12}(\tau')] dx_s dy_s . \end{aligned} \quad (8.122)$$

We see from Eq. (8.122) that the cross term in the interference pattern is given by an integrated collection of fringe patterns that are shifted proportional to the

source point position y_s . Furthermore, each shifted fringe pattern is modulated by the function $m_{12}(\tau)$, whose width is determined by the spectral bandwidth $\Delta\nu$.

Although Eq. (8.122) is rather insightful, it can also be rather cumbersome to handle analytically. In this case, rather than integrate over the spectral bandwidth, we will integrate over the source position y_s . Equation (8.122) now becomes

$$\begin{aligned}
 I(\tau) &= I_0K_1 + I_0K_2 + 2I_0\sqrt{K_1K_2} \int_0^\infty g(\nu) \operatorname{Re} \left\{ e^{j2\pi\nu\tau} \int_{-\infty}^\infty m_R(x_s, y_s) e^{j2\pi\nu\tau \frac{y_s d}{cz_s}} dx_s dy_s \right\} d\nu \\
 &= I_0(K_1 + K_2 I_0 K_1 + I_0 K_2 + 2I_0\sqrt{K_1K_2}) \int_0^\infty g(\nu) m_{12}\left(\frac{\nu d}{cz_s}\right) \cos \left[2\pi\nu\tau + \beta_{12}\left(\frac{\nu d}{cz_s}\right) \right] d\nu
 \end{aligned} \tag{8.123}$$

Again, recognize that the bracketed expression in Eq. (8.123) as an inverse Fourier transform of the spatial function $m_R(x_s, y_s)$ with the transform variable of $\frac{\nu d}{cz_s}$. The mutual coherence function now becomes

$$\Gamma_{12}(\tau) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} \int_0^\infty g(\nu) \mathbf{F}^{-1} \left[\frac{\nu d}{cz_s} \int_{-\infty}^\infty m_R(x_s, y_s) dx \right] e^{j2\pi\nu\tau} d\nu . \tag{8.124}$$

Comparing Eqs. (8.124) and (8.114) we see that the cross spectral density of the electric field at the location of the pinholes is given by

$$\begin{aligned}
 \hat{\Gamma}_{12}(\nu) &= \frac{I_0}{\frac{1}{2}cn\epsilon_0} g(\nu) \operatorname{step}(\nu) \mathbf{F}^{-1} \left[\frac{\nu d}{cz_s} \int_{-\infty}^\infty m_R(x_s, y_s) dx \right] \\
 &= \frac{I_0}{\frac{1}{2}cn\epsilon_0} g(\nu) \operatorname{step}(\nu) m_{12}\left(\frac{\nu d}{cz_s}\right) \exp \left[j\beta_{12}\left(\frac{\nu d}{cz_s}\right) \right] .
 \end{aligned} \tag{8.125}$$

If the spatial distribution of the source is

$$m_R(x_s, y_s) = \frac{1}{L} \operatorname{rect}\left(\frac{y_s}{L}\right) \delta(x_s) , \tag{8.126}$$

the inverse Fourier transform given by

$$m_{12}\left(\frac{vd}{cz_s}\right) = \left| \operatorname{sinc}\left(\frac{Lv d}{cz_s}\right) \right|. \quad (8.127)$$

If the power spectrum of the source is given by

$$g(v) = \frac{1}{\Delta v} \operatorname{rect}\left(\frac{v - \bar{v}}{\Delta v}\right), \quad (8.128)$$

the complete mutual coherence is given by

$$\Gamma_{12}(\tau) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} \int_{-\infty}^{\infty} \frac{1}{\Delta v} \operatorname{rect}\left(\frac{v - \bar{v}}{\Delta v}\right) \operatorname{sinc}\left(\frac{Lv d}{cz_s}\right) e^{j2\pi v \tau} dv. \quad (8.129)$$

The cross spectral density, as defined by Eq. (8.125), is

$$\hat{\Gamma}_{12}(v) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} \frac{1}{\Delta v} \operatorname{rect}\left(\frac{v - \bar{v}}{\Delta v}\right) \operatorname{step}(v) \operatorname{sinc}\left(\frac{Lv d}{cz_s}\right). \quad (8.130)$$

Equation (8.130) illustrates that the bandwidth of the cross spectral density is equal to the bandwidth Δv of the source power spectrum.

8.3.2 Refinement of the quasimonochromatic assumption

If the spectral bandwidth Δv becomes very small, then Eq. (8.129) can be approximated as

$$\begin{aligned} \Gamma_{12}(\tau) &\approx \frac{I_0}{\frac{1}{2}cn\epsilon_0} \int_{-\infty}^{\infty} \delta(v - \bar{v}) \operatorname{sinc}\left(\frac{Lv d}{cz_s}\right) e^{j2\pi v \tau} dv \\ &= \frac{I_0}{\frac{1}{2}cn\epsilon_0} \operatorname{sinc}\left(\frac{L\bar{v} d}{cz_s}\right) e^{j2\pi \bar{v} \tau}. \end{aligned} \quad (8.131)$$

That is, the mutual coherence becomes the spatial coherence limit at mean frequency \bar{v} . The delta function in Eq. (8.131) is used only for the sake of mathematical convenience and can be a bit misleading. It implies that the bandwidth of the source is narrow, but it does not mean that Eq. (8.131) holds for a perfectly monochromatic light source. Implicit in the derivation of Eq. (8.131) is that the

source is incoherent (i.e. made up of a collection of independent point sources). However, for a perfectly monochromatic extended source, different points on the source cannot be treated independently. This subtlety is demonstrated quite easily by attempting to create an extended source by placing a diffuser in the path of a narrow bandwidth laser beam (a source that is close to perfectly monochromatic). The speckle pattern that is observed with this demonstration is a result of the laser light scattered from different points on the diffuser interfering. Furthermore, if a Young’s double pinhole interferometer is placed *directly after* the diffuser, the result will be fringes with a visibility of $V=1$ in the observation screen, *independent of the pinhole separation*¹. However, Eq. (8.131) implies that the interfereability (mutual coherence) decreases with pinhole separation d (proportional to $\theta_A = \frac{d}{z}$). Therefore, the source bandwidth must be narrow, but not perfectly monochromatic, for the approximation in Eq. (8.131) to hold. In order to understand how narrow the bandwidth $\Delta\nu$ needs to be for this approximation to hold, Eq. (8.129) is rewritten as

$$\Gamma_{12}(\nu) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} e^{j2\pi\bar{\nu}\tau} \int_{-1/2}^{1/2} \text{sinc}\left[\frac{Ld}{cz_s}(u\Delta\nu + \bar{\nu})\right] e^{j2\pi\Delta\nu\tau u} du \quad , \quad (8.132)$$

where

$$u = \frac{\nu - \bar{\nu}}{\Delta\nu} \quad . \quad (8.133)$$

The rect function in the integrand of Eq. (8.132) limits nonzero values over the range $u = -\left(\frac{1}{2}, \frac{1}{2}\right)$. There are no contributions to the integral for $|u| > \frac{1}{2}$. If

$$\Delta\nu\tau \ll 1 \quad , \quad (8.134)$$

the exponential term in the integrand also has no net contribution to the integral for this range of u . That is, $\exp[j2\pi(\Delta\nu\tau)u] \approx 1$ because $(\Delta\nu\tau)u \approx 0$ for all values of u over the range of integration. With these approximations, Eq. (8.132) becomes

1. The clever student might comment that a rotating ground glass plate is commonly used to reduce speckle, and thus the coherence, in a “monochromatic” laser beam. This apparent contradiction is resolved when one realizes that the effect of placing a rotating ground glass plate is the same as randomly phase modulating the wave-fronts leaving each pinhole. A phase modulated wave-front is no longer monochromatic, but has a bandwidth that is proportional to the frequency of the phase modulation, which is, of course related to the speed of rotation of the ground glass plate. If one were to use a fast enough detector (faster than the rotation speed of the ground glass plate) in the observation plane, the interference would again become observable.

$$\Gamma_{12}(\tau) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} e^{j2\pi\bar{\nu}\tau} \int_{-1/2}^{1/2} \text{sinc}\left[\frac{Ld\bar{\nu}}{cz_s}\left(u\frac{\Delta\nu}{\bar{\nu}} + 1\right)\right] du . \quad (8.135)$$

$u \ll \frac{\bar{\nu}}{\Delta\nu}$, $u\frac{\Delta\nu}{\bar{\nu}} \ll 1$, so

$$\text{sinc}\left[\frac{Ld\bar{\nu}}{cz_s}\left(u\frac{\Delta\nu}{\bar{\nu}} + 1\right)\right] \approx \text{sinc}\left(\frac{L\bar{\nu}d}{cz_s}\right) . \quad (8.136)$$

Therefore, the mutual coherence function becomes

$$\Gamma_{12}(\tau) = \frac{I_0}{\frac{1}{2}cn\epsilon_0} e^{j2\pi\bar{\nu}\tau} \text{sinc}\left(\frac{L\bar{\nu}d}{cz_s}\right) , \quad (8.137)$$

which is equivalent to the result for $\Delta\nu \rightarrow 0$ in Eq. (8.131). Therefore, a necessary condition for the assumption of quasimonochromatic illumination is that

$$\tau \ll \frac{1}{\nu} , \quad (8.138)$$

which is interpreted as requiring that $\text{OPD}_0 \ll \frac{c}{\Delta\nu}$, or that the maximum optical path difference in the optical system is much less than the coherence length of the source.

8.3.3 The Mutual Intensity Function

Notice that the interesting characteristics of the mutual coherence in Eq. (8.131) result from the sinc function, which is not a function of time difference τ . The exponential term simply defines high frequency fringes based on the mean frequency $\bar{\nu}$. Therefore, for quasi-monochromatic illumination,

$$\Gamma_{12}(\tau) \approx \Gamma_{12}(0) . \quad (8.139)$$

The function $\Gamma_{12}(0)$ is extremely useful in characterizing a quasimonochromatic optical field. This function is referred to as the *mutual intensity function*, and is usually given the variable $J(\mathbf{r}_1, \mathbf{r}_2) = J_{12}$. Expressed in terms of temporal averages, the mutual intensity, derived from Eq. (8.6) is

$$J_{12} = \langle U(\mathbf{r}_1, t)U^*(\mathbf{r}_2, t) \rangle , \quad (8.140)$$

where $U(\mathbf{r}_1, t)$ is the instantaneous electric field at the location \mathbf{r}_1 at time t .

One primary use of the mutual intensity function is that it can be used to directly calculate irradiance, while still containing information about the optical field's coherence properties. Recall that the irradiance at any point in an optical system, apart from a multiplicative constant, is given by the time average of the absolute value of the square of the total electric field at that point. That is,

$$\begin{aligned} I(\mathbf{r}_1) &= \frac{1}{2}cn\epsilon_0\langle |U(\mathbf{r}_1, t)|^2 \rangle \\ &= \frac{1}{2}cn\epsilon_0\langle U(\mathbf{r}_1, t)U^*(\mathbf{r}_1, t) \rangle . \end{aligned} \quad (8.141)$$

Comparing Eqs. (8.132) and (8.134), the relationship between mutual intensity and irradiance is given by

$$I(\mathbf{r}_1) = \frac{1}{2}cn\epsilon_0J_{11} , \quad (8.142)$$

where \mathbf{r}_1 is a location in the plane of interest (defined by the position vector \mathbf{r}_1) and J is the mutual intensity at optical plane in question. Image irradiance is just the mutual intensity evaluated at $\mathbf{r}_2=\mathbf{r}_1$.

8.3.4 Mapping of Mutual Intensity

The relationship between the mutual intensity in any two planes of an optical system can be expressed as a four-dimensional linear mapping. This mapping is demonstrated as follows. If there is a linear relationship between the electric field in one plane (say plane o) of an optical system and the electric field in another plane in the system (say plane i), this relationship is expressed as a simple two-dimensional integral transform given by

$$U_i(\mathbf{r}', t) = \int_{-\infty}^{\infty} U_o(\mathbf{r}, t)\psi_{oi}(\mathbf{r}', \mathbf{r})d^2\mathbf{r} , \quad (8.143)$$

where $\psi(\mathbf{r}', \mathbf{r})$ is the impulse response function of the linear system. This impulse response could be the point spread function of an imaging system, which is illustrated by the labeling choice of o and i for the planes in question. However, this point response function can also represent the free space propagation kernel or any

other linear relationship given by the optical system. Combining Eqs. (8.140) and (8.143) we obtain for the mutual intensity in plane i

$$\begin{aligned}
 J_i(\mathbf{r}_1', \mathbf{r}_2') &= \left\langle \int_{-\infty}^{\infty} U_o(\mathbf{r}_1, t) \psi_{oi}(\mathbf{r}_1', \mathbf{r}_1) d^2 \mathbf{r}_1 \left[\int_{-\infty}^{\infty} U_o(\mathbf{r}_2, t) \psi_{oi}(\mathbf{r}_2', \mathbf{r}_2) d^2 \mathbf{r}_2 \right]^* \right\rangle \\
 &= \left\langle \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_o(\mathbf{r}_1, t) U_o^*(\mathbf{r}_2, t) \psi_{oi}(\mathbf{r}_1', \mathbf{r}_1) \psi_{oi}^*(\mathbf{r}_2', \mathbf{r}_2) d^2 \mathbf{r}_1 d^2 \mathbf{r}_2 \right\rangle \quad (8.144) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle U_o(\mathbf{r}_1, t) U_o^*(\mathbf{r}_2, t) \rangle \psi_{oi}(\mathbf{r}_1', \mathbf{r}_1) \psi_{oi}^*(\mathbf{r}_2', \mathbf{r}_2) d^2 \mathbf{r}_1 d^2 \mathbf{r}_2
 \end{aligned}$$

The averaging brackets are brought inside the integration, because only the electric field functions $U_o(\mathbf{r}_1, t)$ and $U_o^*(\mathbf{r}_2, t)$ depend on time. This operation is equivalent to rearranging the order of integration, given that the time averaging process is simply an integration over the variable t . Finally, noticing that the averaged quantity in Eq. (8.144) is simply the mutual intensity at the plane o , the relationship between the mutual intensity at planes o and i is given by

$$J_i(\mathbf{r}_1', \mathbf{r}_2') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J_o(\mathbf{r}_1, \mathbf{r}_2) \psi_{oi}(\mathbf{r}_1', \mathbf{r}_1) \psi_{oi}^*(\mathbf{r}_2', \mathbf{r}_2) d^2 \mathbf{r}_1 d^2 \mathbf{r}_2 \quad (8.145)$$

At this stage, all that has been done is to show how the mutual intensity propagates between two planes in an optical system. In using the mutual intensity function, it is assumed that the spectral bandwidth of the electric field is narrow enough to be classified as quasimonochromatic. However, no statement has been made as to the functional form of the mutual intensity. In order to specify this functional form, the coherence must either be measured directly, or an assumption about the physical nature of the system in question must be made. The natural place to make this assumption is at the location of the source.

Considering the relationship between the plane of the light source and some other plane in the optical system, the mapping of mutual intensities between these two planes is given by

$$J_o(\mathbf{r}_1, \mathbf{r}_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J_s(s_1, s_2) \psi_{s0}(\mathbf{r}_1, s_1) \psi_{s0}^*(\mathbf{r}_2, s_2) d^2 s_1 d^2 s_2 \quad (8.146)$$

where s is a vector representing a location on the source, and $\psi_{s0}(\mathbf{r}, s)$ is the impulse response representing the linear relationship between the source plane and the plane at which knowledge of the mutual intensity is desired.

If we begin by assuming that the electric field originates from some sort of naturally radiating light source, in many cases the spatial correlations in the source are much smaller than the wavelength of light. The reason we can make this assumption is that a naturally occurring light sources, such as a mercury discharge lamp, produces light by the excited motion of atoms. It is natural to assume each of the atoms in the source is moving independently. Because the size of the atoms in question is much smaller than the wavelength of light, the spatial correlation in the atomic motion (and consequently the optical phase of the light being produce by each atom) is much smaller than the wavelength of the light being produced. Knowledge that spatial correlations of the temporal variations of the source light field phase are smaller than the wavelength of light is profound. This is because spatial features in the optical field that are smaller than the wavelength of light do not propagate, except as evanescent fields. Assuming that distance between the source and the plane of interest is greater than a few wavelengths, the mutual intensity at a source of this type can be approximated by a delta function.

$$J_s(s_1, s_2) = CI(s_1)\delta(s_1 - s_2) , \quad (8.147)$$

where $I(s)$ is the irradiance distribution of the source, and C is a constant. Whenever this approximation is valid, the source of the radiation is said to be "spatially incoherent". Substitution of Eq. (8.147) into Eq. (8.146) yields

$$J_0(\mathbf{r}_1, \mathbf{r}_2) = \int_{-\infty}^{\infty} CI(s)\psi_{s0}(\mathbf{r}_1, s)\psi_{s0}^*(\mathbf{r}_2, s)d^2s . \quad (8.148)$$

In the context of an imaging system, the function $\psi_{s0}(\mathbf{r}, s)$ will likely be the point response function of the condenser system. However, if the system in question is just a simple propagation through free space, the function $\psi_{s0}(\mathbf{r}, s)$ is the Fresnel free space propagation kernel, given by

$$\psi_{s0}(\mathbf{r}, s) = \frac{e^{jkz}}{j\lambda z} \exp\left\{\frac{jk}{2z}[(x - s_x)^2 + (y - s_y)^2]\right\} . \quad (8.149)$$

Substitution of Eq. (8.149) into Eq. (8.148) and expanding the quadratic terms in the exponential yields the explicit functional form of the well known van Cittert-Zernike Theorem, which was discussed previously in Section 8.1.6.

When the assumption about the optical source given in Eq. (8.147) is valid, the irradiance at nearly any location in an optical system can be calculated by treating the source as a collection of independent point sources that are assumed to be incoherent with respect to each other. The irradiance produced by each point source can then be added together to produce the resultant irradiance produced by the spatially incoherent source. This assumption is extremely useful for simulating many optical systems. However, it must be understood that there are many situations where this assumption is not valid. Much work has been done studying so called "quasi-homogeneous" sources. (see Section 5.6.4 of Goodman, 1985) These sources have an narrow spatial coherence, but cannot be suitably approximated by Eq. (8.147).

8.3.5 Propagation of Cross Spectral Intensity

When the light source in an optical system is assumed to be quasimonochromatic, all temporal coherence properties of the optical field are ignored, and the coherence can be described adequately using the mutual intensity function. However, if we remove this assumption, then the mutual intensity no longer provides an adequate description of the statistical properties of the optical field. The entire mutual coherence function must then be used. As shown in Section 5.3.1, the Fourier transform of the mutual coherence function is the cross spectral density of Eq.(8.147). The ability to express the mutual coherence function in terms of a Fourier transform assumes that the electric field is stationary with respect to time. As stated in Section 8.1.1, if we ignore any transient response that occurs when the source is turned on, and assume that the overall irradiance of the source is constant (over a period of time much longer than the period of the optical field's temporal fluctuations), this treatment is generally valid.

It is shown in Section 3.5 of [Goodman, 1985] that the cross spectral density can also be expressed in terms of the spectral representation of the electric field. This relationship is given by

$$\hat{\Gamma}(\mathbf{r}_1, \mathbf{r}_2; \nu) = \lim_{T \rightarrow \infty} \frac{1}{2T} \overline{U(\mathbf{r}_1, \nu)U^*(\mathbf{r}_2, \nu)} \quad , \quad (8.150)$$

where the overbar indicates an ensemble average. That is, the average is taken over an ensemble of possible results for the product $U(\mathbf{r}_1, \nu)U^*(\mathbf{r}_2, \nu)$. The fact that the cross spectral density can be expressed this way is extremely useful for determining a law of propagation.

An extrapolation of Eq. (8.143) that describes a linear relationship between the polychromatic electric field in one plane (say plane o) of an optical system and the

polychromatic electric field in another plane in the system (say plane i), is expressed as a simple two dimensional linear mapping

$$U_i(\mathbf{r}', \nu) = \int_{-\infty}^{\infty} U_o(\mathbf{r}, \nu) \psi_{oi}(\mathbf{r}', \mathbf{r}; \nu) d^2 \mathbf{r} . \quad (8.151)$$

where $\psi_{oi}(\mathbf{r}', \mathbf{r}; \nu)$ is the impulse response function of the linear system. This impulse response could be the point spread function of an imaging system illustrated by the labeling choice of o and i for the planes in question or any other linear relationship given by the optical system. Notice that the point response function in this case is also a function of optical frequency. This functionality will generally be true for any real linear relationship in an optical system. However, how strongly this function depends on optical frequency will depend on the situation in question. In some cases, the spectral dependence of the point response function can be suitably neglected, and in others it cannot. It will be assumed, for the sake of generality, that it cannot be neglected, and in situations where it can be neglected, this fact will specifically be stated. Combining Eqs. (8.150) and (8.151), we obtain for the mutual intensity in plane i

$$\begin{aligned} & \hat{\Gamma}_i(\mathbf{r}_1', \mathbf{r}_2'; \nu) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{2T} \overline{U_o(\mathbf{r}_1, \nu) U_o^*(\mathbf{r}_2, \nu)} \psi_{oi}(\mathbf{r}_1', \mathbf{r}; \nu) \psi_{oi}^*(\mathbf{r}_2', \mathbf{r}; \nu) d^2 \mathbf{r}_1 d^2 \mathbf{r}_2 , \end{aligned} \quad (8.152)$$

where the ensemble average has been brought inside the integration, because only the electric field functions are random. Even though the point response functions depend on optical frequency, they are not random and therefore not averaged. Recognizing the ensemble average as the incident mutual spectral intensity, we arrive at the following relationship

$$\hat{\Gamma}_i(\mathbf{r}_1', \mathbf{r}_2'; \nu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{\Gamma}_o(\mathbf{r}_1, \mathbf{r}_2; \nu) \psi_{oi}(\mathbf{r}_1', \mathbf{r}_1; \nu) \psi_{oi}^*(\mathbf{r}_2', \mathbf{r}_2; \nu) d^2 \mathbf{r}_1 d^2 \mathbf{r}_2 . \quad (8.153)$$

Equation (8.153) is a very powerful result, because it demonstrates that the cross spectral intensity obeys the same propagation law as the mutual intensity, except that $\bar{\nu}$ must be replaced by ν and the spectral dependence of the point response function must also be accounted for. Although the term $\bar{\nu}$ is not explicitly stated in manipulations of the mutual intensity function shown previ-

ously, it generally appears in the linear mapping kernels (such as the Fresnel propagation kernel given in Eq. (8.149) as an average wavelength. This result is proved alternatively using the Helmholtz equation in Section 5.4.3 of [Goodman, 1985]. In order to obtain the mutual coherence function for the resultant field, all that remains to be done is to calculate the inverse Fourier transform of the result of Eq. (8.153).

$$\begin{aligned}
 & I(\tau, \nu; x_s, y_s) dx_s dy_s d\nu \\
 &= I_0 \left\{ K_1 + K_2 + 2\sqrt{K_1 K_2} \cos \left[2\pi\nu \left(\frac{y_s d}{cz_s} + \tau \right) \right] \right\} g(\nu) m_R(x_s, y_s) d\nu dx_s dy_s . \quad (0.3)
 \end{aligned}$$

$$\begin{aligned}
 I(\tau) &= I_0 \left\{ K + K_2 + 2\sqrt{K_1 K_2} \int_{-\infty}^{\infty} \int_0^{\infty} g(\nu) m_R(x_s, y_s) \right. \\
 &\quad \left. \left[\cos 2\pi\nu \left(\tau + \frac{y_s d}{cz_s} \right) \right] d\nu dx_s dy_s \right\} . \quad (0.4)
 \end{aligned}$$

$$\text{Re}[\Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau)] = \text{Re} \left\{ \int_{-\infty}^{\infty} \int m_R(x_s, y_s) \left[\int_0^{\infty} g(\nu) e^{\left[-j2\pi\nu \left(\tau + \frac{y_s d}{cz_s} \right) \right]} d\nu \right] dx_s dy_s \right\} . \quad (0.7)$$

$$\text{Re}[\Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau)] = \int_{-\infty}^{\infty} \int m_R(x_s, y_s) m_{12} \left(\tau + \frac{y_s d}{cz_s} \right) \cos \left[\Phi_{12} \left(\tau + \frac{y_s d}{cz_s} \right) \right] dx_s dy_s \quad (0.8)$$

(B) if the power spectrum has a form given by Eq (8.49), Eq (0.8) can be rewritten

$$\begin{aligned}
 \text{Re}[\Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau)] &= \int_{-\infty}^{\infty} \int m_R(x_s, y_s) m_{12} \left(\tau + \frac{y_s d}{cz_s} \right) \\
 &\quad \cos \left[2\pi\bar{\nu}\tau + \beta_{12} \left(\tau + \frac{y_s d}{cz_s} \right) \right] dx_s dy_s . \quad (0.9)
 \end{aligned}$$

$$\operatorname{Re}[\Gamma(\mathbf{r}_1, \mathbf{r}_2; \tau)] = \operatorname{Re} \left(\int_0^\infty g(v) \left\{ \int_{-\infty}^\infty \int_{-\infty}^\infty m_R(x_s, y_s) e^{\left[-j2\pi v \left(\tau + \frac{y_s d}{cz_s} \right) \right]} dx_s dy_s \right\} dv \right) \quad (0.10)$$