Chapter 3
Interference

Light wave interference is observed as a modulation of irradiance, usually bright fringes and dark fringes on an observation screen. The geometry of the fringes depends on the shape of the wavefronts and the difference in the optical path traveled by the waves. Differences in the order of the wavelength of light cause changes in irradiance from a bright fringe to a dark fringe, making interference a highly accurate tool for measuring refractive indices, wavefronts, forms of optical surfaces, thicknesses, etc. The physical parameter that determines the quality of the interference (the possibility of generating fringes) is the coherence between the waves. The coherence has its origin in the fluctuations of the optical field emitted by the sources. Natural sources, like the sun, emit spontaneously (randomly), but in artificial sources, like lasers, the emission has a high degree of correlation.

Two interference patterns generated with a He-Ne laser are shown in Fig. 3.1. The laser beam is focused with a positive lens into a small hole in an opaque screen, which is seen as a point of light in the figure (point source). The lens is behind the screen and cannot be seen in the image. The divergent (spherical) wavefront passes through several optical elements. First, it passes through a 1 mm thick microscope slide (flat piece of glass). There, light is reflected from each slide face and interference occurs between the two reflected signals, which is seen on the opaque screen (two-source interference pattern at bottom left). This interference pattern consists of roughly circular fringes, where the thickness of the bright fringes is similar to the thickness of the dark fringes. This is the typical result of the interference of two sources that emit spherical waves. The beam transmitted by the microscope slide is then allowed to enter a Fabry–Pérot interferometer, which consists of two thick plates of highly reflective glass parallel to each other. The separation between the plates is less than a millimeter, and the facing faces have a thin aluminum film that increases reflectance. This generates multiple reflections, with similar amplitude coefficients, so there is now interference from more than two waves. The effect on the reflected interference pattern is a thinning of

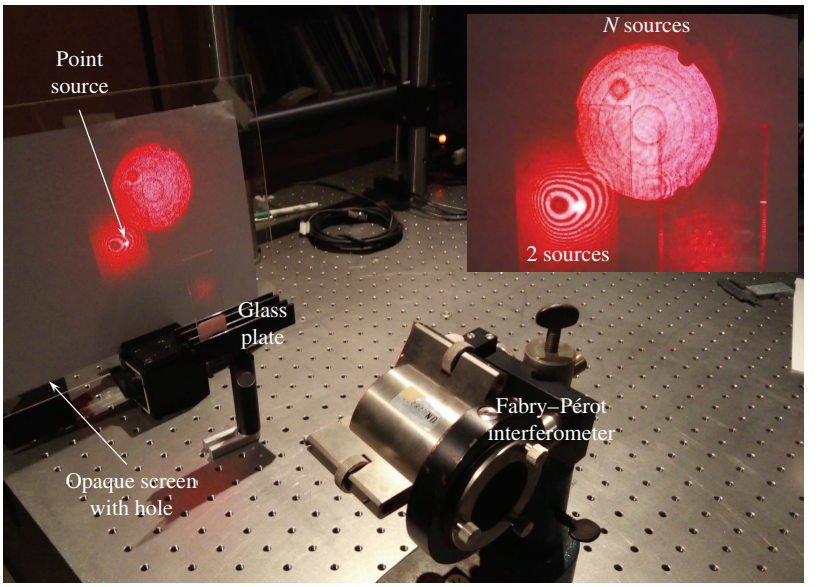


Figure 3.1 Two-wave and multi-wave interference.

the dark fringes, as can be seen on the opaque screen (N sources interference pattern at top right).

In this chapter the most common interferometers are presented, starting with the Michelson interferometer, since it is very illustrative to study the interference of plane waves and spherical waves. The interference of multiple beams in a plate with parallel faces is then discussed, which also explains the operation of the Fabry–Pérot interferometer. In all cases, it is assumed that the surfaces of the optical elements that compose the interferometers are ideal, i.e., they coincide with their mathematical description. In practice, the manufacturing process of these elements limits the optical quality of the surfaces. Finally, some practical aspects of the Michelson interferometer, which may also be present in other interferometers, are discussed in Section 3.4.

3.1 Interference and Coherence

Consider the sum of two harmonic plane waves given by $\mathbf{E}_1(\mathbf{r}, t) = \mathbf{E}_{01}e^{i(\mathbf{k}_1 \cdot \mathbf{r} - \omega t + \phi_1)}$ and $\mathbf{E}_2(\mathbf{r}, t) = \mathbf{E}_{02}e^{i(\mathbf{k}_2 \cdot \mathbf{r} - \omega t + \phi_2)}$ at a point $\mathbf{r}$ of empty space (air). The phases ϕ_1 and ϕ_2 are functions of time that depend on the light emission process and account for the fluctuations of the fields in $\mathbf{r}$. The amplitudes $\mathbf{E}_{01}$ and $\mathbf{E}_{02}$ are assumed to be constant in time. The resulting wave is

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t). \tag{3.1}$$

The irradiance at $\mathbf{r}$, according to Eq. (2.29), is given by

$$I(\mathbf{r}) = \frac{\epsilon_0 c}{2}\langle \mathbf{E}_1 \cdot \mathbf{E}_1^* + \mathbf{E}_2 \cdot \mathbf{E}_2^* + 2\mathrm{Re}\{\mathbf{E}_1 \cdot \mathbf{E}_2^*\}\rangle, \tag{3.2}$$

i.e.,

$$I(\mathbf{r}) = \frac{\epsilon_0 c}{2}\langle E_1^2\rangle + \frac{\epsilon_0 c}{2}\langle E_2^2\rangle + 2\frac{\epsilon_0 c}{2}\mathrm{Re}\langle \mathbf{E}_1 \cdot \mathbf{E}_2^*\rangle. \tag{3.3}$$

Let us set the plane that contains the wave vectors $\mathbf{k}_1 = 2\pi\hat{\mathbf{s}}_1/\lambda$ and $\mathbf{k}_2 = 2\pi\hat{\mathbf{s}}_2/\lambda$ as a reference plane. With respect to this plane, the parallel and orthogonal components of $\mathbf{E}_1$ and $\mathbf{E}_2$ can be defined, so that $\mathbf{E}_{01} = \mathbf{E}_{01}^{\perp} + \mathbf{E}_{01}^{\parallel}$ and $\mathbf{E}_{02} = \mathbf{E}_{02}^{\perp} + \mathbf{E}_{02}^{\parallel}$. Then,

$$\mathbf{E}_1 \cdot \mathbf{E}_2^* = \mathbf{E}_{01}^{\perp} \cdot (\mathbf{E}_{02}^{\perp})^* e^{-i(\Delta\mathbf{k}\cdot\mathbf{r}+\Delta\phi)} + \mathbf{E}_{01}^{\parallel} \cdot (\mathbf{E}_{02}^{\parallel})^* e^{-i(\Delta\mathbf{k}\cdot\mathbf{r}+\Delta\phi)}, \tag{3.4}$$

where $\Delta\mathbf{k} = \mathbf{k}_2 - \mathbf{k}_1$ and $\Delta\phi = \phi_2 - \phi_1$. In Eq. (3.4), the terms $\mathbf{E}_{01}^{\parallel} \cdot (\mathbf{E}_{02}^{\perp})^*$ and $\mathbf{E}_{01}^{\perp} \cdot (\mathbf{E}_{02}^{\parallel})^*$ are not included because the components are orthogonal to each other, resulting in 0. In other words, waves with polarizations orthogonal to each other do not interfere.

The vectors $\mathbf{E}_{01}^{\perp}$ and $\mathbf{E}_{02}^{\perp}$ are parallel, while the vectors $\mathbf{E}_{01}^{\parallel}$ and $\mathbf{E}_{02}^{\parallel}$ form an angle equal to the angle between $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$; thus,

$$I(\mathbf{r}) = \frac{\epsilon_0 c}{2}\langle E_1^2\rangle + \frac{\epsilon_0 c}{2}\langle E_2^2\rangle + 2\frac{\epsilon_0 c}{2}\mathrm{Re}\langle [E_{02}^{\perp}E_{01}^{\perp} + E_{02}^{\parallel}E_{01}^{\parallel}\cos(2\alpha)]e^{-i(\Delta\mathbf{k}\cdot\mathbf{r}+\Delta\phi)}\rangle, \tag{3.5}$$

where 2α is the angle between $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$. From Eq. (3.5), the interference due to waves whose polarization states are parallel to the reference plane depends on the angle 2α. Thus, if $2\alpha = \pi/2$, the waves do not interfere and the irradiance will simply be the sum of the irradiances due to the parallel components, $I_{\parallel} = (I_1 + I_2)_{\parallel}$. On the other hand, the interference due to waves whose polarization states are orthogonal to the reference plane does not depend on the angle formed between $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$.

In what follows, the interfering waves will be assumed to be in the polarization state that is orthogonal to the reference plane; thus, the expression for two-wave interference is

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}\mathrm{Re}\langle e^{-i(\Delta\mathbf{k}\cdot\mathbf{r}+\Delta\phi)}\rangle, \tag{3.6}$$

where $I_1 = (\epsilon_0 c/2)\langle E_1^2\rangle = (\epsilon_0 c/2)(E_{01}^{\perp})^2$ and $I_2 = (\epsilon_0 c/2)\langle E_2^2\rangle = (\epsilon_0 c/2)(E_{02}^{\perp})^2$ are the irradiances in $\mathbf{r}$ generated by each of the waves. In the term $\langle e^{-i(\Delta\mathbf{k}\cdot\mathbf{r}+\Delta\phi)}\rangle$, the phase difference $\Delta\phi$ depends on time. So, to get the

average value, one needs to explicitly know the variation of ϕ_1 and ϕ_2 over time.

3.1.1 Degree of coherence

To get to Eq. (3.6), it is assumed that there is no time delay between the waves, which can occur if the origin of one of the sources is displaced or if one of the waves travels through a medium that produces a change in the speed of propagation. If there is a time lag τ, then the sum of the waves in general would be

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t - \tau). \tag{3.7}$$

Now, the irradiance is given by

$$\begin{aligned} I(\mathbf{r}) &= I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}\langle e^{-i(\Delta\mathbf{k}\cdot\mathbf{r}+\omega\tau+\Delta\phi(\tau))}\rangle \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}\{e^{-i\Delta\mathbf{k}\cdot\mathbf{r}}\langle e^{-i\omega\tau}e^{-i(\Delta\phi(\tau))}\rangle\}, \end{aligned} \tag{3.8}$$

with $\Delta\phi(\tau) = \phi_2(t - \tau) - \phi_1(t)$. The complex degree of coherence is defined as

$$\begin{aligned} \gamma(\tau) &= \langle e^{-i\omega\tau}e^{-i(\Delta\phi(\tau))}\rangle \\ &= |\gamma(\tau)|e^{-i(\alpha(\tau)+\omega\tau)}, \end{aligned} \tag{3.9}$$

where $\alpha(\tau) + \omega\tau$ is the phase of the degree of coherence. Then,

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}|\gamma(\tau)|\cos(\Delta\mathbf{k}\cdot\mathbf{r} + \omega\tau + \alpha(\tau)). \tag{3.10}$$

The modulus of the degree of coherence satisfies $0 \le |\gamma(\tau)| \le 1$. In its limits, it takes the value 0 if $\Delta\phi$ is a random function and the value 1 if $\Delta\phi$ is constant in time. The superposition of the waves is considered to be incoherent, partially coherent, or coherent if the following is satisfied for the magnitude of $\gamma(\tau)$:

$$|\gamma(\tau)| = \begin{cases} 0, & \text{incoherent} \\ \{0, 1\}, & \text{partially coherent} \\ 1, & \text{coherent.} \end{cases} \tag{3.11}$$

Although the interference expression given by Eq. (3.10) has been obtained specifically for homogeneous plane waves (i.e., when the amplitude of each wave is constant over a given wavefront), the result is valid for the sum of two waves in general. In the latter case, the degree of coherence is the normalized version of the correlation of the fields at point $\mathbf{r}$. *Principles of Optics* by Born and Wolf contains the general formalism for interference with partially coherent waves [1].

3.1.2 Interference and coherence

To see the effect that $\gamma(\tau)$ has on interference, let us consider a simple but illustrative example. Following Fowles [2], let us suppose there is a light source consisting of a two-level electronic system, which spontaneously emits a pulse given by

$$E(t) = E_0 \text{rect}(t/\tau_0) e^{-i(2\pi\nu_0 t - \phi(t))}, \tag{3.12}$$

where the function $\text{rect}(t/\tau_0)$ describes a rectangular signal as follows:

$$\text{rect}\left(\frac{t}{\tau_0}\right) = \begin{cases} 1, \text{ if } \quad |t| \leq \tau_0/2 \\ 0, \text{ otherwise.} \end{cases}. \tag{3.13}$$

Hence, the pulse has cosine form with frequency ν_0 and duration τ_0, such that $\tau_0 > T$, where the period $T = 1/\nu_0$. The initial phase $\phi(t)$ is random within the range $-\pi < \phi < \pi$. The spontaneous character of the emission is described by the initial random phase of the pulse. Imagining a continuous emission of pulses but with random phases, the graphical representation of the phase in the emission process can be described as shown in Fig. 3.2.

Now suppose that the light thus generated is used in an optical system with which we can produce interference of two waves; this is an interferometer. For example, in Fig. 3.3, the diagram of a Michelson interferometer is shown. The (plane) wave coming out of the source S is called the primary wave (0). The first task of the interferometer is to generate two (secondary) waves from the primary wave. For this, a flat semi-mirror can be used, which reflects 50% of the amplitude of the primary wave and lets the other 50% of the amplitude pass. The element that does this task is called a beamsplitter (BS). In this way, the secondary waves (1) are obtained. The second task of the interferometer is to

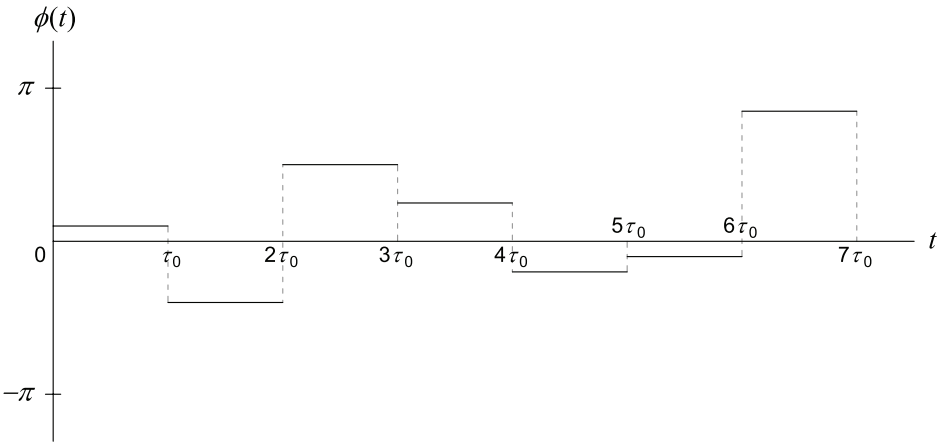


Figure 3.2 Initial phases of pulses emitted randomly one after another by a two-level electronic transition source.

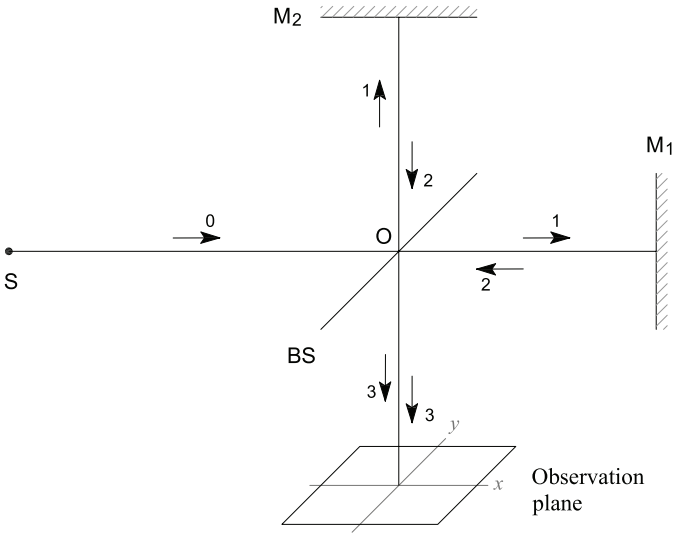


Figure 3.3 Diagram of a Michelson interferometer. A light beam (0) coming out of the source S is divided by the beamsplitter (BS) into two separate beams (1). One of the rays passes through the BS and goes to the flat mirror M_1 and the other beam is reflected by the BS and goes to the flat mirror M_2. The reflected beams (2) in each mirror reach the BS, resulting in two new beams (3) that overlap to produce interference.

add the two secondary waves. This is achieved by means of the two mirrors, M_1 and M_2, with which the direction of propagation of the secondary waves (2) is changed, directing them again toward the beamsplitter so that the reflected and transmitted waves (3) in the beamsplitter overlap. In particular, if the mirrors are orthogonal to the wave vectors of the secondary waves (with the beamsplitter at 45°) in the interference region, the unit propagation vectors $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ will be parallel, i.e., $\mathbf{k}_2 - \mathbf{k}_1 = 0$.

The final result is the sum of two plane waves (3) with $\Delta\mathbf{k} = 0$ and with a time delay τ resulting from the difference in the optical path traveled by the two secondary waves (1, 2, 3) from point O at the beamsplitter. If the distance between O and M_1 is d_1 and the distance between O and M_2 is d_2, the optical path difference (in air $n = 1$) between the waves (3) in the interference region is $2(d_2 - d_1)$. Consequently, the time lag will be $\tau = 2(d_2 - d_1)/c$. For $d_2 > d_1$, $\tau > 0$. Then, displacing M_2 (or M_1) axially, the desired time delay is achieved.

Let us assume that the time delay between the two waves is less than the duration of the pulse generated by the light source, $\tau < \tau_0$. The degree of coherence between the two waves depends on the difference in the initial phases of the two waves. Because a copy of a wave is made with the interferometer, the phase difference given in Fig. 3.2 would look like the one shown in Fig. 3.4. At the top of Fig. 3.4, the phase $\phi(t)$ of one of the waves is shown along with the phase of the other wave, including the delay time. The result of the subtraction is displayed at the bottom of Fig. 3.4. In the intervals of duration $(\tau_0 - \tau)$, the

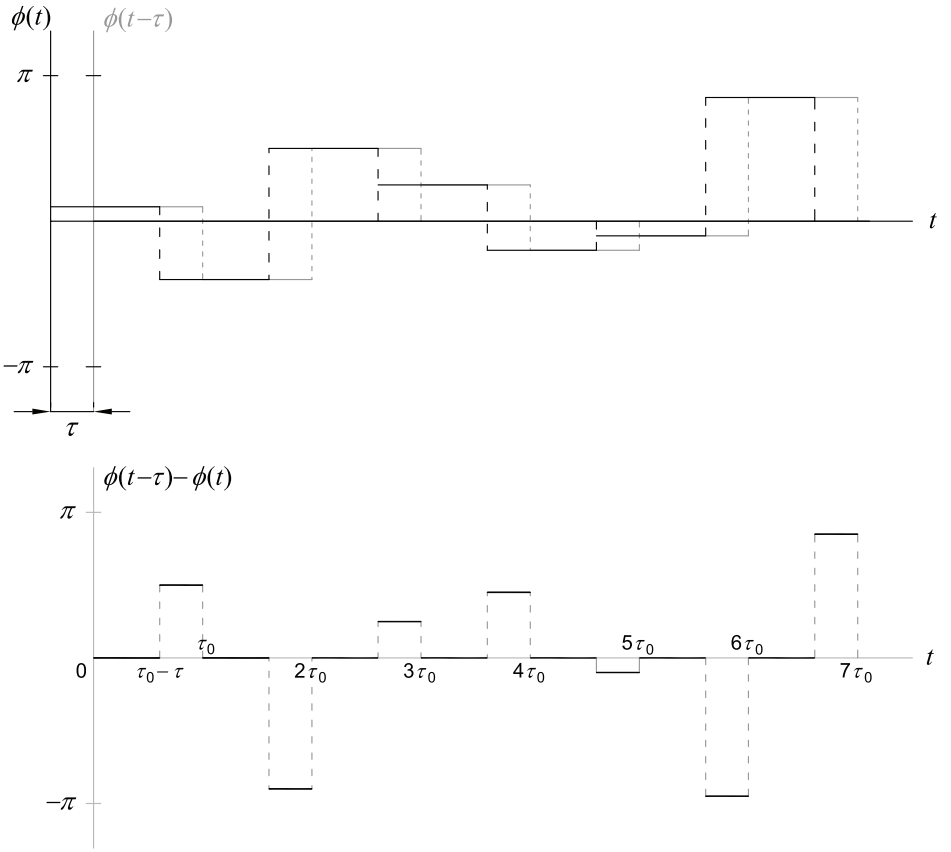


Figure 3.4 Difference of the phases $\phi(t-\tau)$ and $\phi(t)$ for the time delay τ.

phases coincide and, therefore, the result is 0. In the other intervals, the result is not null, but it still has a random distribution of phases.

Thus, from Eq. (3.9),

$$\gamma(\tau) = \frac{e^{-i\omega\tau}}{\check{T}} \int_0^{\check{T}} e^{-i(\phi(t-\tau)-\phi(t))} dt. \tag{3.14}$$

This integral can be solved by adding the M integrals over each interval of duration τ_0, assuming that $\check{T} = M\tau_0$. Thus,

$$\gamma(\tau) = \frac{e^{-i\omega\tau}}{\check{T}} \left[\int_0^{\tau_0} e^{-i(\phi(t-\tau)-\phi(t))} dt + \ldots + \int_{(M-1)\tau_0}^{M\tau_0} e^{-i(\phi(t-\tau)-\phi(t))} dt \right], \tag{3.15}$$

and each of these integrals, in turn, can be decomposed into two intervals: the duration interval ($\tau_0 - \tau$), where the phase difference is 0, and the interval of

duration τ, where the phase difference takes any value, say Δ_j for the jth interval. Therefore, grouping the integrals for which the phase shift is 0 and putting in another group of integrals with phase shifts Δ_j,

$$\begin{aligned}\gamma(\tau) = &\frac{e^{-i\omega\tau}}{\breve{T}}\left[\int_0^{\tau_0-\tau} dt + \int_{\tau_0}^{2\tau_0-\tau} dt + \ldots + \int_{(M-1)\tau_0}^{M\tau_0-\tau} dt\right] \\ &+ \frac{e^{-i\omega\tau}}{\breve{T}}\left[\int_{\tau_0-\tau}^{\tau_0} e^{i\Delta_1} dt + \int_{2\tau_0-\tau}^{2\tau_0} e^{i\Delta_2} dt + \ldots + \int_{M\tau_0-\tau}^{M\tau_0} e^{i\Delta_M} dt\right],\end{aligned} \tag{3.16}$$

which is equal to

$$\gamma(\tau) = \frac{e^{-i\omega\tau}}{\breve{T}}[M(\tau_0 - \tau)] + \frac{e^{-i\omega\tau}}{\breve{T}}[0]. \tag{3.17}$$

Because the result of the second group of integrals is 0, since the phases Δ_j are random, the sum (for $\breve{T} > \tau_0$) is 0. Thus, the final result is

$$\begin{aligned}\gamma(\tau) &= e^{-i\omega\tau}\frac{\tau_0 - \tau}{\tau_0}; \qquad \tau < \tau_0 \\ &= 0; \qquad\qquad\quad \tau \geq \tau_0.\end{aligned} \tag{3.18}$$

The modulus of the degree of coherence is

$$|\gamma(\tau)| = 1 - \frac{\tau}{\tau_0}, \tag{3.19}$$

and the phase $\alpha(\tau) = 0$. The graphical representation is shown in Fig. 3.5. When $\tau = 0$, the correlation of the waves is maximum and the degree of coherence is equal to 1. For values of $\tau > \tau_0$, the correlation between the waves

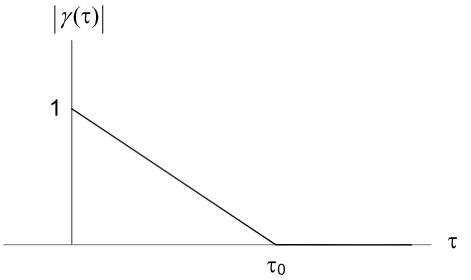


Figure 3.5 Modulus of the degree of coherence of a source that emits random and consecutive pulses of duration τ_0.

is null; therefore, there is no interference and the irradiance is reduced to the sum of the individual irradiances of the waves, i.e., $I = I_1 + I_2$.

To see in detail the effect of $\gamma(\tau)$ on the irradiance in the observation screen of the interferometer represented in Fig. 3.3, let us consider the irradiance at the point $\{x = 0,\ y = 0\}$. Because the beamsplitter divides the amplitude of the wave into two equal parts, $I_1 = I_2 = I_0$. Therefore, the irradiance at $\mathbf{r} = \{0, 0\}$ is

$$I(0, 0) = 2I_0\left[1 + \left(1 - \frac{\tau}{\tau_0}\right)\cos\left(\frac{2\pi\tau}{T}\right)\right], \quad \tau > 0. \tag{3.20}$$

The condition for $\tau > 0$ holds assuming $d_2 > d_1$. However, mirror 2 can also be displaced axially so that $d_2 < d_1$, leading to $\tau < 0$. With this in mind,

$$I(0, 0) = 2I_0\left[1 + \left(1 \pm \frac{\tau}{\tau_0}\right)\cos\left(\frac{2\pi\tau}{T}\right)\right], \tag{3.21}$$

where the + sign is used for $-\tau_0 < \tau < 0$ and the – sign is used for $0 < \tau < \tau_0$.

In Fig. 3.6, the irradiance is shown as a function of the time delay [Eq. (3.21)] in the range $-12T < \tau < 12T$, assuming that the source emits pulses of duration $\tau_0 = 10T$. When the mirrors are at the same distance from the beamsplitter (point O, Fig. 3.3), $d_2 = d_1$, then $\tau = 0$ and the maximum value of irradiance is obtained. By moving one of the mirrors axially, the irradiance oscillates, progressively decreasing (increasing) the maximum (minimum) of the irradiance. This irradiance variation occurs according to the modulus of the degree of coherence shown in Fig. 3.5. When the time delay reaches the value of τ_0, the irradiance oscillations disappear, and for $|\tau| > \tau_0$ the irradiance remains constant, i.e., $I = 2I_0$. Note that the envelope of the irradiance plot corresponds to the modulus of the degree of coherence. Thus, we have an experimental way (with a Michelson interferometer) to measure the degree of coherence of a light source.

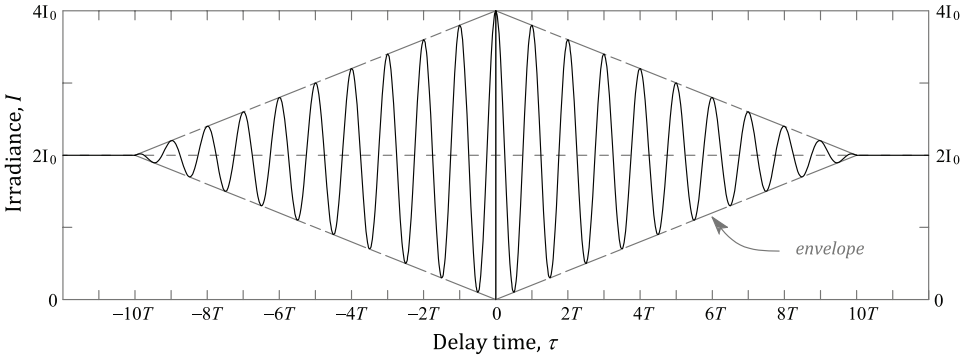


Figure 3.6 Irradiance as a function of delay time τ.

A complete oscillation of the irradiance occurs every time the optical path difference between the waves changes by λ. In other words, every time the argument of the cosine function changes by 2π, a maximum (minimum) is obtained. Setting the argument of the cosine function equal to 2π, i.e., $2\pi\Delta\tau/T = 2\pi(2\Delta d)/(cT) = 2\pi$,

$$\Delta d = \frac{\lambda}{2} \tag{3.22}$$

is the spatial increase in the separation of the mirrors corresponding to two consecutive maxima or minima.

3.1.3 Coherence length

From the above, there would be interference (radiation modulation) if the delay time were less than τ_0. This is also equivalent to saying that there would be interference if the path difference between the waves in the interference region were less than $l_c = c\tau_0$. This length is called the *coherence length* and is often used to characterize the (temporal) coherence of a light source. With the Michelson interferometer, we have a way to measure the coherence length. It can also be done in another way, which depends on the spectral width (bandwidth) of the light source. To see this, let us return to the pulse example from Eq. (3.12). Although the cosine vibration within the rect function [Eq. (3.13)] has a specific frequency, ν_0, the pulse spectrum is not a Dirac delta located at ν_0, but rather a distribution of frequencies centered at ν_0 due to the finite duration of the pulse. A signal is strictly monochromatic if it is described by a cosine function of infinite duration.

To evaluate the spectral content of the pulse, Fourier analysis can be used. For a particular pulse, the function $\phi(t)$, which measures the initial phase, will have a fixed value. Omitting this term, the Fourier transform $\tilde{E}(\nu) = F[E(t)]$ is

$$\begin{aligned}\tilde{E}(\nu) &= E_0 \int_{-\infty}^{\infty} \text{rect}(t/\tau_0) e^{-i(2\pi\nu_0 t)} e^{i2\pi\nu t} dt \\ &= E_0 \left[\int_{-\infty}^{\infty} \text{rect}(t/\tau_0) e^{i2\pi\nu t} dt\right] * \left[\int_{-\infty}^{\infty} e^{-i(2\pi\nu_0 t)} e^{i2\pi\nu t} dt\right],\end{aligned} \tag{3.23}$$

where the symbol $*$ denotes the convolution operation. Solving the two integrals, then

$$\tilde{E}(\nu) = E_0[\tau_0 \text{sinc}(\pi\nu\tau_0)] * [\delta(\nu - \nu_0)], \tag{3.24}$$

where the function $\text{sinc}(x) = \sin(x)/x$. The delta function centers the function $\text{sinc}(\pi\nu\tau_0)$ at $\nu = \nu_0$. The graph of Eq. (3.24) is shown in Fig. 3.7. The spectral width of the function $\text{sinc}(\pi\nu\tau_0)$ can be defined as half the spectral separation

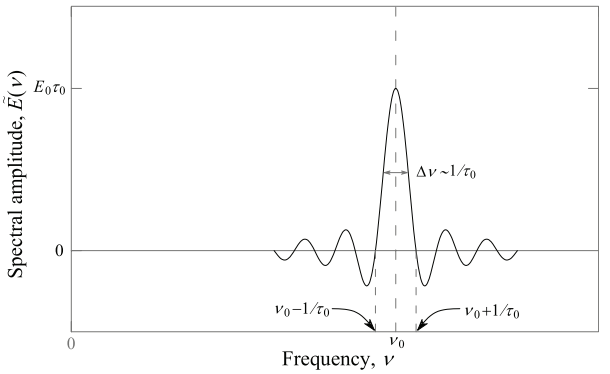


Figure 3.7 Spectrum of the pulse given by Eq. (3.12). The spectral width $\Delta\nu$ approaches half the spectral separation of the leading zeros of the function $\text{sinc}(\pi\nu\tau_0)$.

between its leading zeros closest to ν_0. Let $\Delta\nu$ be the spectral width of the initial zeros closest to ν_0; the spectral positions of the leading zeros are given by $\pi(\nu_0 + \Delta\nu)\tau_0 = \pi$ and $\pi(\nu_0 - \Delta\nu)\tau_0 = -\pi$. Subtracting them leads to

$$\Delta\nu = \frac{1}{\tau_0}. \tag{3.25}$$

Thus, the inverse of the spectral width is a measure of the pulse duration time, which, in turn, determines the coherence length. Therefore, the coherence length of a light source having spectral width $\Delta\nu$ is given by

$$l_c = \frac{c}{\Delta\nu}. \tag{3.26}$$

Because $\nu = c/\lambda$, taking the differentials of ν and λ, $\Delta\nu = -\nu\Delta\lambda/\lambda$, leads to an alternative form for the coherence length given by

$$l_c = \frac{\overline{\lambda}^2}{|\Delta\lambda|}, \tag{3.27}$$

where $\overline{\lambda}$ is the central value of the wavelength in the spectral range of width $\Delta\lambda$.

Considering the classical model for an illumination source, the spectral width of the line (for electronic oscillators) is a universal constant [3],

$$\Delta\lambda = 1.2 \times 10^{-5}\,\text{nm}. \tag{3.28}$$

With this value, the coherence length for a line of the visible spectrum varies as 13 m $< l_c <$ 41 m (for 400 nm $< \lambda_0 <$ 700 nm). In contrast, for white light, between 400 nm (violet) and 700 nm (red), the coherence length would be approximately $l_c = 1$ μm, with $\lambda = 550$ nm and $\Delta\lambda = 300$ nm. Other sources, such as mercury arc lamps, have coherence lengths of the order of 3 cm, and

Kr discharge lamps have coherence lengths of the order of 30 cm. Lasers have long coherence lengths; e.g., a stabilized He-Ne laser can have a coherence length of 300 m.

3.2 Interference of Two Plane Waves

In the previous section, it is shown that the expression for the interference of two plane waves of angular frequency $\omega = 2\pi\nu_0$ is given by

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}|\gamma(\tau)|\cos[\Delta\mathbf{k}\cdot\mathbf{r} + \omega\tau + \alpha(\tau)],$$

where $|\gamma(\tau)|$ is the modulus of the degree of coherence, $\Delta\mathbf{k} = \mathbf{k}_2 - \mathbf{k}_1$ is the difference of the wave vectors at the observation point $\mathbf{r}$, τ is the delay time of the waves with respect to a reference point or plane (point O in the beamsplitter), and $\alpha(\tau)$ is the phase (together with $\omega\tau$) of the degree of coherence. Taking the two waves from the same source, $\alpha(\tau) = 0$. In practice, to observe the interference of two plane waves, it is usual to expand the beam (increase its cross section) that reaches the beamsplitter. This can be done by focusing light at a point S, which is made to coincide with the primary focal point of an aberration-corrected lens. This lens is called a *collimating lens* (CL), as shown in Fig. 3.8. After the lens, there will be planar wavefronts orthogonal to the optical axis of the lens, which will be taken as the optical axis of the interferometer. Any of these wavefronts can be taken as the reference plane; here, let us take the wavefront arriving at point O of the

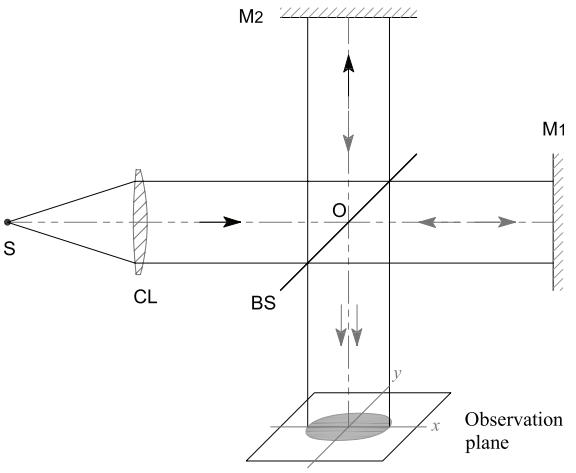


Figure 3.8 Plane waves interference with a Michelson interferometer. A collimating lens (CL) converts the divergent spherical wave from the point light source S into a plane wave. When the mirrors are orthogonal to the wave vectors, there is a uniformly illuminated region on the screen that changes in irradiance level as one of the mirrors moves axially.

beamsplitter as the reference plane. Once the beams are split, the path differences from O can be measured.

Let us suppose that the illumination source has a degree of coherence that remains constant for the displacements of the mirrors considered, $|\gamma(\tau)| = |\gamma|$. By setting the plane mirrors M_1 and M_2 orthogonal to the wave vectors of the beams transmitted and reflected by the beamsplitter, the interference in the observation plane would be given by

$$I(x, y) = I_1 + I_2 + 2|\gamma|\sqrt{I_1 I_2}\cos(\omega\tau). \tag{3.29}$$

For τ fixed, $I(x, y)$ does not depend on x or y, so in the observation plane there would be a uniform irradiance distribution, as shown in Fig. 3.8. By axially displacing one of the mirrors, the irradiance level on the observation screen changes, obtaining maximum values at $(d_2 - d_1) = m\lambda/2$, where m is an integer. Figure 3.9 shows three examples of the variation of the irradiance level (with $I_1 = I_2 = I_0$) in the observation plane as a function of the relative axial displacement between the mirrors M_1 and M_2, for three degrees of coherence: $|\gamma| = 1$, 1/2, and 0.

By observing the variation of irradiance as a function of $(d_2 - d_1)$, $|\gamma|$ can be measured directly. In particular, the maximum and minimum irradiances can be measured, which in turn are given by

$$I_{\max} = I_1 + I_2 + 2|\gamma|\sqrt{I_1 I_2}, \tag{3.30}$$

$$I_{\min} = I_1 + I_2 - 2|\gamma|\sqrt{I_1 I_2}, \tag{3.31}$$

according to Eq. (3.29). By defining the visibility or contrast of irradiance modulation as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{3.32}$$

the modulus of the degree of coherence is obtained by substituting Eqs. (3.30) and (3.31) in Eq. (3.32); i.e.,

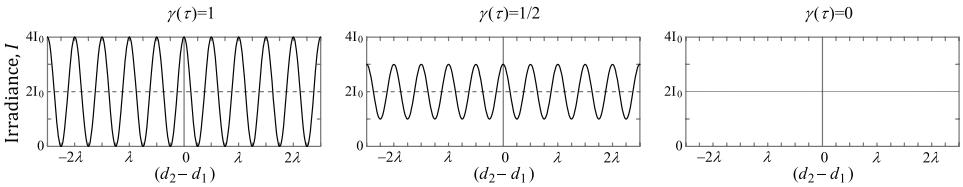


Figure 3.9 Modulation of the irradiance for different degrees of coherence in the interferometer shown in Fig. 3.8 as a function of the difference in position of the mirrors with respect to the point O of the beamsplitter.

$$|\gamma| = \frac{(I_1 + I_2)}{2\sqrt{I_1 I_2}} C. \tag{3.33}$$

Thus, from the irradiance measurements, the modulus of the degree of coherence is determined. Irradiances of each wave and the contrast are measured individualy. In particular, if the beamsplitter reflects 50% and transmits 50%, then $I_1 = I_2$ and $|\gamma| = C$. In this case, the contrast is the modulus measure of the degree of coherence.

3.2.1 Interference with inclined plane waves

In what follows, let us assume that $|\gamma(\tau)| = 1$, e.g., by illuminating with a laser whose coherence length is much greater than the axial displacements of the mirrors. If the mirrors M_1 and M_2 rotate through an angle $\alpha/2$, as shown in Fig. 3.10, each beam in the interference region would be inclined by an angle α with respect to the optical (vertical) axis. In each mirror, the rotation is performed with respect to the point of intersection of the mirror with the optical axis, i.e, O_1 in mirror M_1 and O_2 in mirror M_2. Therefore, the angle between the two beams in the interference region is 2α. At the place of overlap there will be a pattern of straight fringes (interference pattern) whose separation will arise from the angle of inclination.

To determine the geometry of the fringes, let us first assume that the mirror distances $d_1 = \overline{OO_1}$ and $d_2 = \overline{OO_2}$ are equal. This implies that there will not be a time delay between waves along the optical axis. The intersection

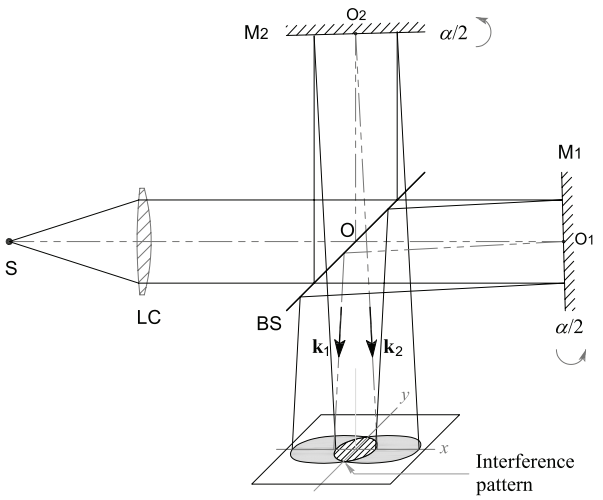


Figure 3.10 Interference with inclined plane waves. On the observation screen there will be a pattern of fringes. The separation of the fringes depends on the angle of inclination of the mirrors.

of the optical axis with the viewing screen is taken as the coordinate origin for the interferogram ($x = 0$, $y = 0$). Now the expression for the interference is

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\Delta\mathbf{k}\cdot\mathbf{r}). \tag{3.34}$$

Assuming that the observation screen is in the $z = 0$ plane, according to the geometry of Fig. 3.10, $\mathbf{k}_1 = (2\pi/\lambda)\{s_{1x}, 0, s_{1z}\}$ and $\mathbf{k}_2 = (2\pi/\lambda)\{s_{2x}, 0, s_{2z}\}$, with $s_{1z} = s_{2z}$ and $s_{1x} = -s_{2x}$. On the other hand, $\mathbf{r} = \{x, y, 0\}$. Thus, $\Delta\mathbf{k}\cdot\mathbf{r} = (2\pi/\lambda)(s_{2x} - s_{1x})x$ and because $s_{2x} = \sin\alpha$,

$$\Delta\mathbf{k}\cdot\mathbf{r} = \frac{2\pi}{\lambda}(2x\sin\alpha). \tag{3.35}$$

This indicates that on the observation screen there will be a modulation of the irradiance in the x direction, according to

$$I(x) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left(\frac{4\pi}{\lambda}x\sin\alpha\right). \tag{3.36}$$

A simulated interference pattern for two plane waves is shown in Fig. 3.11 when $\alpha = 0.01°$, $\lambda = 632.8$ nm, and $I_1 = I_2$. The region where the interference pattern is observed corresponds to $(-5 < x < 5)$ mm and $(-5 < y < 5)$ mm.

The separation Δx between two consecutive maxima or minima of irradiance occurs every time $(4\pi/\lambda)x\sin\alpha$ changes by 2π; i.e.,

$$\Delta x = \frac{\lambda}{2\sin\alpha}. \tag{3.37}$$

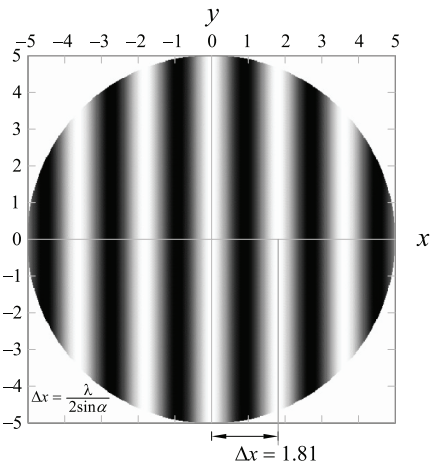


Figure 3.11 Simulation of an interference pattern generated by two inclined plane waves when $|\gamma(\tau)| = 1$ ($\alpha = 0.01°$ and $\lambda = 632.8$ nm). The scale of the axes is in millimeters.

Note that to have a separation between fringes equal to 1.81 mm, in the example of Fig. 3.11, the angle of rotation of the mirrors is very small: $\alpha/2 = 0.005°$.

3.2.2 Displacement of interference fringes

When one of the mirrors in the interferometer shown in Fig. 3.8 is moved axially, there is a change in the level of irradiance on the observation screen. There is no spatial modulation of irradiance (along the screen); the modulation is axial, i.e., temporal. If in the interferometer shown in Fig. 3.10 one of the mirrors moves axially ($d_2 - d_1 \neq 0$), at each point of the observation screen there is also a change in the level of irradiance, but the net effect observed is a transversal displacement of the fringes in the x direction. The axial displacement of one of the mirrors introduces a time delay $\tau = 2(d_2 - d_1)/c$ between the waves (with respect to point O), and the irradiance is given by

$$I(x) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left[\frac{4\pi}{\lambda}x\sin\alpha + \frac{4\pi}{\lambda}(d_2 - d_1)\right]. \tag{3.38}$$

By bringing the mirror M_1 closer to point O, so that $d_2 - d_1 > 0$, the value of the argument of the cosine function increases and, to maintain the initial value, x must take a negative value; thus, the interference fringes shift to the left. In Fig. 3.12, three interferograms with $\alpha = 0.01°$, $\lambda = 632.8$ nm, and $I_1 = I_2$ are simulated when the axial displacement of the mirror M_1 leads to the optical path difference $d_2 - d_1 = 0$, $\lambda/8$, and $\lambda/4$. Moving mirror M_1 away from point O, so that $d_2 - d_1 < 0$, the interference fringes will shift to the right.

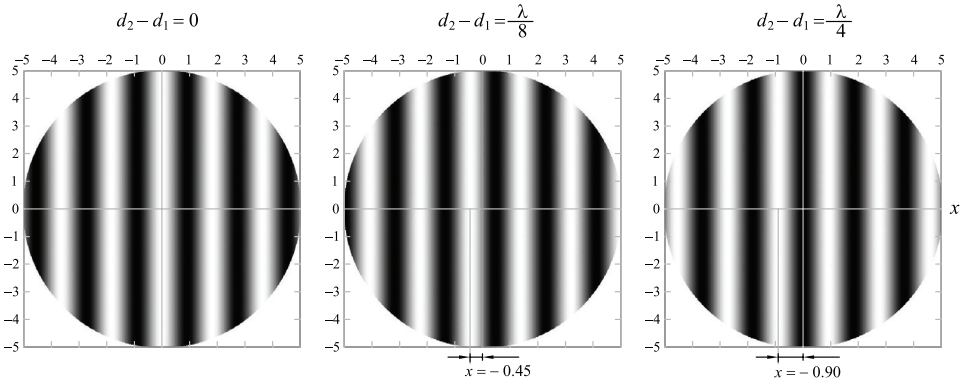


Figure 3.12 Transversal displacement of interference fringes as a function of the axial displacement of the mirror M_1 (approaching the beamsplitter). The optical path difference in the images is $d_2 - d_1 = 0$, $\lambda/8$, and $\lambda/4$. The scale of the axes is in millimeters.

3.2.3 Interferogram visibility

The modulation contrast of irradiance in the axial direction was defined by Eq. (3.32). This quantity, based on the irradiances, allows the degree of coherence to be measured. The visibility of the interference fringe pattern can also be calculated from Eq. (3.32), where I_{max} and I_{min} are the maximum and minimum irradiance values of the fringe pattern.

The change in visibility depends on the degree of coherence and the relationship between the intensities of the two waves; e.g., if $|\gamma| = 1$, then the visibility

$$C = \frac{2\sqrt{I_1 I_2}}{(I_1 + I_2)} \tag{3.39}$$

only depends on the ratio between I_1 and I_2. In Fig. 3.13, three interferograms are shown along with their profiles in the x direction when the wave amplitudes are $E_1 = E_0$ and $E_2 = E_0$, $E_1 = E_0$ and $E_2 = 0.4E_0$, and $E_1 = E_0$ and $E_2 = 0.1E_0$. In the first case, $I_{max} = 4I_0$, $I_{min} = 0$, and $C = 1$; in the second case, $I_{max} = 1.96I_0$, $I_{min} = 0.36I_0$, and $C = 0.69$; and in the third case, $I_{max} = 1.21I_0$, $I_{min} = 0.81I_0$, and $C = 0.20$. Note that the irradiance oscillates spatially around the mean value $I_1 + I_2$, which in the first case is $2I_0$, in the second case is $1.16I_0$, and in the third case is $1.01I_0$.

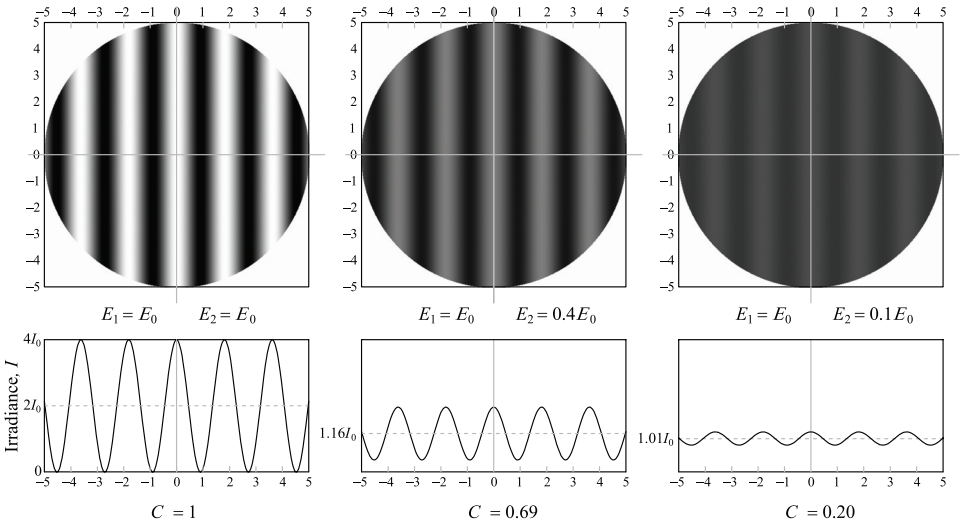


Figure 3.13 Interferograms when the visibility of the fringes depends on the ratio of the wave amplitudes: $C = 1$ ($E_1 = E_0$ and $E_2 = E_0$), $C = 0.69$ ($E_1 = E_0$ and $E_2 = 0.4E_0$), and $C = 0.20$ ($E_1 = E_0$ and $E_2 = 0.1E_0$). The modulus of the degree of coherence is set to $|\gamma| = 1$. The scale of the axes is in millimeters.

Thus, when $I_1 = I_2 = I_0$ and $|\gamma| = 1$,

$$I = 4I_0 \cos^2\left(\frac{4\pi}{\lambda} x \sin\alpha\right), \tag{3.40}$$

from where $I_{max} = 4I_0$ and $I_{min} = 0$, and the visibility is $C = 1$. If $I_1 \neq I_2$, the visibility decreases, and it becomes zero if $I_1 = 0$ or if $I_2 = 0$.

If in addition $|\gamma| < 1$, then there is a greater decrease in the amplitude of the spatial oscillation of the fringes (maintaining the mean value). In other words, the width of the profiles shown in Fig. 3.13 decreases.

In practice, one of the interferometer mirrors is fixed and aligned, while the other can be moved and tilted using precision screws. The analysis carried out for the formation of the interference fringes is still valid, obtaining the same results.

3.3 Interference of Two Spherical Waves

A light source is considered to be point-like if its apparent size is insignificant relative to the distance at which the signal is detected. The detected wavefront will be seen as a spherical wavefront whose radius is equal to the distance between the source and the point of observation. In the previous section, with the Michelson interferometer shown in Fig. 3.8, two plane waves are generated. This is because the point source is located at the primary focal point of the CL. If the point source is displaced axially from the primary focal point of the CL, the wavefront refracted by the lens would be spherical and the radius of curvature will be equal to the distance between the conjugate of the point source and the point at which the wavefront is measured. Once the spherical wave is divided in the beamsplitter, the mirrors M_1 and M_2 generate two virtual images (secondary point sources) of the conjugate of the point source, and in the interference region we would have the superposition of two spherical waves. As in Section 3.2.1, here we also study the interference phenomena when the mirrors are orthogonal to the optical axis of the interferometer and when the mirrors are tilted at a small angle.

Before dealing with the two interference cases of greatest interest, let us describe the geometry of the problem according to Fig. 3.14. Suppose there are two point sources in an isotropic and homogeneous medium (air, $n = 1$) separated by a distance a, located at $z = -a/2$ and $z = a/2$, and the irradiance at a point P located at $\mathbf{r} = \{x, y, z\}$ is wanted. If the two sources are actually images of a primary source, which emits spherical waves of angular frequency $\omega = 2\pi\nu_0$, the initial phases will be equal and the fields of the spherical waves emitted by S_1 and S_2 at point P can be written as*

*$E_0^{\dagger}$ is the amplitude of the field multiplied by the unit length. Thus, the amplitude of the optical field in a spherical wavefront of radius r would be $E_0 = E_0^{\dagger}/r$.

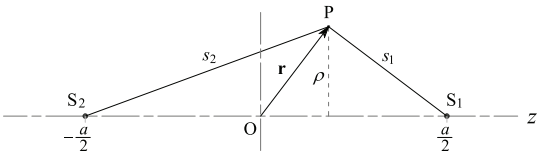


Figure 3.14 Geometry to describe the interference of two spherical waves emitted by point sources S_1 and S_2.

$$E_1(s_1,\, t) = \frac{E_{01}^{\dagger}}{s_1} e^{i(ks_1 - \omega t)} \tag{3.41}$$

and

$$E_2(s_2,\, t) = \frac{E_{02}^{\dagger}}{s_2} e^{i(ks_2 - \omega t)}, \tag{3.42}$$

where $s_1 = [\rho^2 + (a/2 - z)^2]^{1/2}$ and $s_2 = [\rho^2 + (a/2 + z)^2]^{1/2}$, where in turn, $\rho = [x^2 + y^2]^{1/2}$ is the radial coordinate of the projection of point P on the xy plane. The scalar form of these equations implies that at P, the fields have the same polarization state.

If $|\gamma(\tau)| = 1$, the irradiance at P due to the superposition of the two waves would be

$$I(\mathbf{r}) = \frac{\epsilon_0 c}{2}(E_1 + E_2)(E_1^* + E_2^*). \tag{3.43}$$

The individual irradiances at P will be $I_1 = \epsilon_0 c(E_{01}^{\dagger}/s_1)^2/2$ and $I_2 = \epsilon_0 c(E_{02}^{\dagger}/s_2)^2/2$. With this in mind, Eq. (3.43) becomes

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left[\frac{2\pi}{\lambda}(s_2 - s_1)\right]. \tag{3.44}$$

In this situation, $(s_2 - s_1)/c$ measures the delay time of the waves arriving at P. When the argument of the cosine function takes a constant value q, then $2\pi(s_2 - s_1)/\lambda = q$ describes surfaces where the irradiance is constant. Specifically, if $2\pi(s_2 - s_1)/\lambda = 2\pi m$ ($m = 0, \pm 1, \pm 2, \ldots$), the condition of surfaces of maximum irradiance is fulfilled. These surfaces are hyperboloids of revolution defined by

$$(s_2 - s_1) = m\lambda, \tag{3.45}$$

where the point sources are the focal points of the hyperbolas.

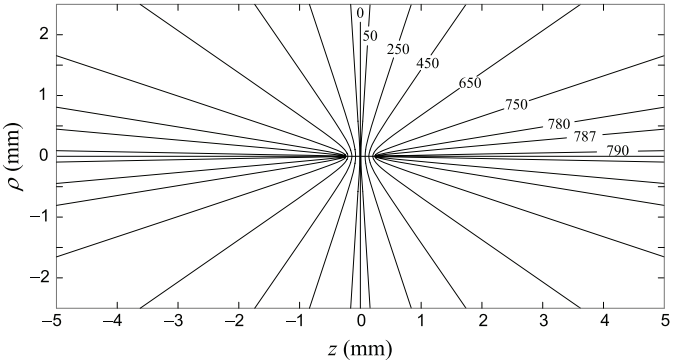


Figure 3.15 Some maximum irradiance curves produced by point sources located at $z = -0.25$ mm and $z = 0.25$ mm for $m = 0$, 50, 250, 450, 650, 750, 780, 787, and 790, with $\lambda = 632.8$ nm.

In Fig. 3.15, some of the hyperbolas resulting from the intersection of a meridional plane with the z axis, when the two point sources are separated by $a = 0.5$ mm and $\lambda = 632.8$ nm, are shown for $m = 0$, 50, 250, 450, 650, 750, 780, 787, and 790. Along these curves, the irradiance has a maximum value. The number of curves along which the maximum irradiance is obtained is determined by the nearest integer to the quotient a/λ, i.e., $m_{\max} \equiv a/\lambda$, which in our example is 790. The surface for $m = 0$ is the plane $z = 0$ (which is in the middle of the two sources). In this case, the optical path difference is zero for any point of coordinates $\{x, y, 0\}$.

Let us see in detail the interference in planes parallel to the $z = 0$ plane and in planes parallel to the $y = 0$ plane. In the first case, due to the symmetry of revolution around the z axis, the maximum irradiance curves are circles. To determine the radii of these circles, let us explicitly write Eq. (3.45) and solve for s_2:

$$\sqrt{\rho^2 + (a/2 + z)^2} = m\lambda + \sqrt{\rho^2 + (a/2 - z)^2}. \tag{3.46}$$

Squaring and simplifying, the radius of the circle for a given m in the $z = z_0$ plane is given by

$$\rho_m = \sqrt{\left(\frac{2az_0 - m^2\lambda^2}{2m\lambda}\right)^2 - \left(\frac{a}{2} - z_0\right)^2}. \tag{3.47}$$

In Fig. 3.16, the circular interference fringes and the fringe radius variation are shown for the example of two point sources separated by $a = 0.5$ mm (located at $z = -a/2$ and $z = a/2$, with $\lambda = 632.8$ nm) in the $z = 100$ mm plane. Moving away from the center, the fringes get closer to each other.

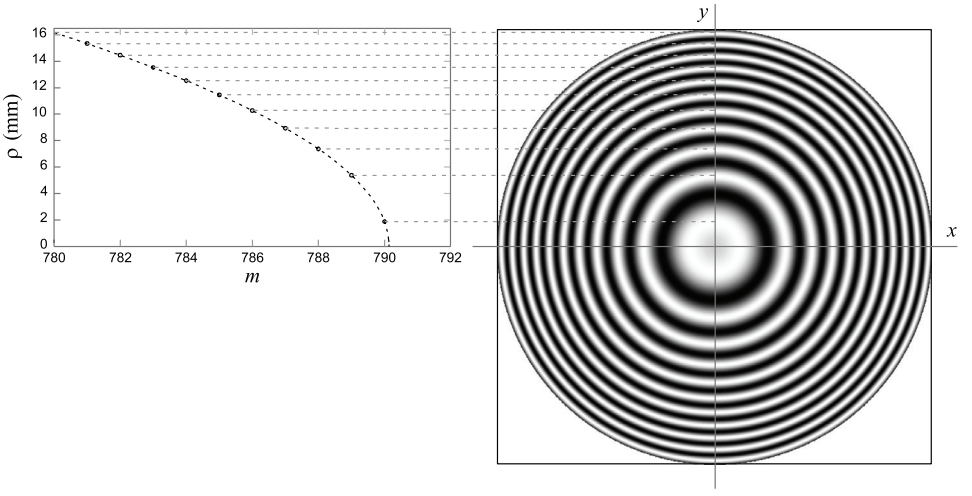


Figure 3.16 Circular interference fringes in the $z = 100$ mm plane with the center at $\{x, y\} = \{0, 0\}$, generated by two point sources separated by $a = 0.5$ mm and located at $z = -a/2$ and $z = a/2$, when the wavelength is $\lambda = 632.8$ nm.

In the second case, i.e., for a plane parallel to the plane $y = 0$, the intersections of the hyperboloids are open curves (hyperbolas). The position of the curves of maximum irradiance along the z direction as a function of m is obtained by making $x = 0$ and $y = y_0$ in Eq. (3.46), i.e.,

$$z_m = \frac{m\lambda}{2}\sqrt{\frac{4y_0^2}{a^2 - m^2\lambda^2} + 1}. \tag{3.48}$$

In Fig. 3.17 some interferograms obtained in different planes parallel to the plane $y = 0$ are shown for different separations between the sources, in an observation region of 80 mm × 80 mm centered at $z = 0$ and $x = 0$. The first interferogram, in the plane $y = 10$ mm, has $m_{max} = 1$, so $z_1 \to \infty$. In other words, there is only one interference fringe, the central fringe ($m = 0$); the rest is an irradiance distribution that extends to infinity on each side. The second

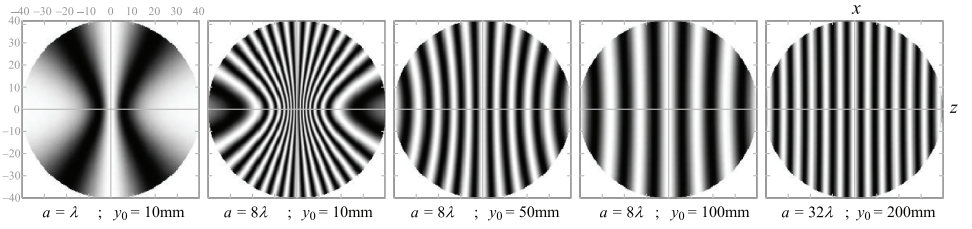


Figure 3.17 Interference fringes in planes parallel to the $y = 0$ plane located at $y_0 = 10$, 50, 100, and 200 mm, generated by two separate point sources $a = \lambda$, 8λ, and 32λ, with $\lambda = 632.8$ nm. The scale of the axes is in millimeters.

interferogram, also in the $y = 10$ mm plane, has $m_{max} = 8$ and all the interference fringes that can be observed in this situation are shown: 7 interference fringes on each side of the central fringe ($m = \pm 1, \pm 2, \pm 3, \pm 4, \pm 5, \pm 6$, and ± 7, as they move away from the central band). The shape of the fringes is hyperbolic. In the third interferogram, the viewing plane moves away to $y = 50$ mm, so there are fewer interference fringes. The fourth interferogram zooms out even farther, to $y = 100$ mm, again decreasing the number of fringes; now the fringes appear straight and equally spaced. The last interferogram, collected at $y = 200$ mm, has $m_{max} = 32$. The interference fringes in the observation region are straight and equally spaced. This last statement can be supported by the fact that when the maximum number of fringes m in the observation region satisfies the $m^2 \ll m_{max}^2$ relation, $a^2 \gg m^2\lambda^2$ and Eq. (3.48) can be approximated by $z_m = m\lambda y_0/a$. Therefore, the fringes in that region are evenly spaced with a separation equal to

$$\Delta z = \frac{\lambda y_0}{a}. \tag{3.49}$$

Going back to the last interferogram, the maximum number m of fringes in the observation screen is 5 and, in fact, $25 \ll 1024$ holds.

The two situations corresponding to Figs. 3.16 and 3.17 are obtained in the Michelson interferometer when illuminated by a spherical wave: the first with the configuration shown in Fig. 3.8 and the second with the configuration shown in Fig. 3.10. These two cases are discussed below.

3.3.1 Circular fringes with the Michelson interferometer

If in the interferometer shown in Fig. 3.8 the CL is moved axially or the lens is simply removed, a spherical wavefront arrives at the beamsplitter, as shown in Fig. 3.18. Mirrors M_1 and M_2 will form virtual images of the point source from which the wavefront diverges. These virtual images constitute two secondary point sources S_1 and S_2. Looking from the observation screen toward the beamsplitter, two secondary point sources can be seen, one after the other, along the optical axis of the interferometer,* separated by a distance $a = 2(d_2 - d_1)$, where $d_2 = \overline{OO_2}$ and $d_1 = \overline{OO_1}$. The radii of curvature of the wavefronts emitted by S_2 and S_1, on the observation screen, are $R_2 = 2d_2 + \overline{OS} + \overline{OO_P}$ and $R_1 = 2d_1 + \overline{OS} + \overline{OO_P}$. O_P is the intersection of the optical axis of the interferometer with the observation screen ($x = 0$ and $y = 0$). Keeping the origin of coordinates in the middle of the two secondary sources (as in Fig. 3.14),

*Note that S_2 is the virtual image of S generated by mirror M_2, but S_1 is the reflection in the beamsplitter of the virtual image generated by mirror M_1 of S. This virtual image is to the right of mirror M_1, but to an observer on the observation screen, the two secondary sources are aligned.

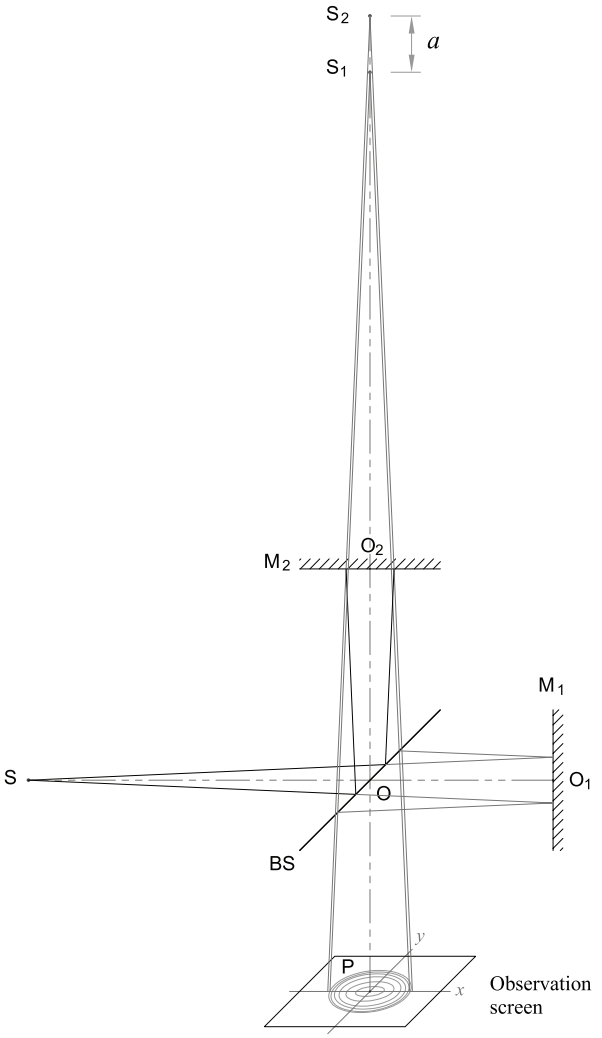


Figure 3.18 A Michelson interferometer illuminated by a spherical wave. On the observation screen, along the optical axis of the interferometer, there is a pattern of circular fringes when the separation between the secondary sources S_1 and S_2 is $a \neq 0$.

$z_0 = (R_1 + R_2)/2$. Consequently, $R_1 = z_0 - a/2$ and $R_2 = z_0 + a/2$, and $\Delta R = R_2 - R_1 = a$. In what follows, let us assume that $R_2 > R_1$.

The interference of the two spherical waves on the observation screen would look as shown in Fig. 3.16. Let us now consider what happens to the interference fringes if one of the mirrors is displaced axially, say mirror M_2. To see this, suppose we shift mirror M_2 so that the separation between the secondary sources is $a = 500\lambda$, 400λ, 300λ, 200λ, 100λ, and 0 (with $\lambda = 632.8$ nm). Let us keep the observation screen at $z_0 = 100$ mm and the observation region centered on the z axis with dimensions 33 mm $\times$ 33 mm. The radii of

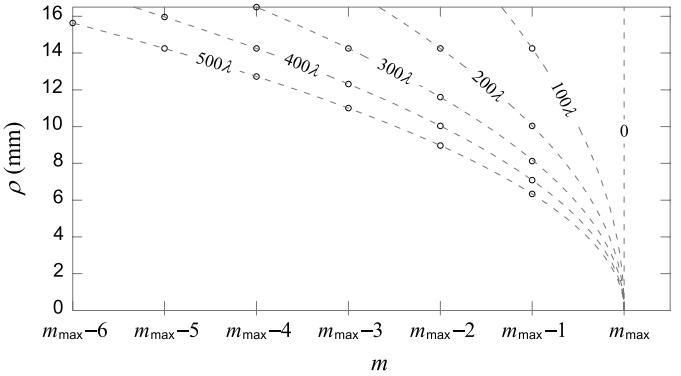


Figure 3.19 Radii of the circles of maximum irradiance as a function of the number m with respect to $m_{\max}$ for different values of the separation (a = 500λ, 400λ, 300λ, 200λ, 100λ, and 0) of the secondary point sources in the interferometer shown in Fig. 3.18.

the circles of maximum irradiance are shown in Fig. 3.19. The maximum value of m, $m_{\max} = a/\lambda$, is taken as a reference point, and in each case it tells us that in the interferogram we have, there is a maximum irradiance at $\{0, 0, z_0\}$. The first ring of irradiance (circular fringe) is given by $m_{\max} - 1$, and each curve tells us what the radius of the circle of maximum irradiance that corresponds to it is (which from now on is taken as the radius of the ring). The second irradiance ring occurs for $m_{\max} - 2$, and so on. The interference patterns for $a = 500\lambda$, 400λ, 300λ, 200λ, and 100λ are shown in Fig. 3.20. For $a = 0$, the region will be completely illuminated, i.e., there are no interference rings, which is illustrated by Fig. 3.19 with the vertical line labeled with 0.

Let us consider the position of the first ring in each interferogram. If initially the position of mirror M_2 with respect to mirror M_1 gives a separation $a = 300\lambda$, the radius of the first ring is $\rho_{m_{\max}-1} = 8.18$ mm. By displacing the mirror M_2, approaching the beamsplitter, so that $a = 200\lambda$, the radius of the first ring increases to $\rho_{m_{\max}-1} = 10.04$ mm. And when the mirror M_2 moves away from the beamsplitter such that $a = 400\lambda$, the radius of the first ring decreases to $\rho_{m_{\max}-1} = 7.08$ mm. Thus, if $R_2 > R_1$, by displacing the mirror M_2,

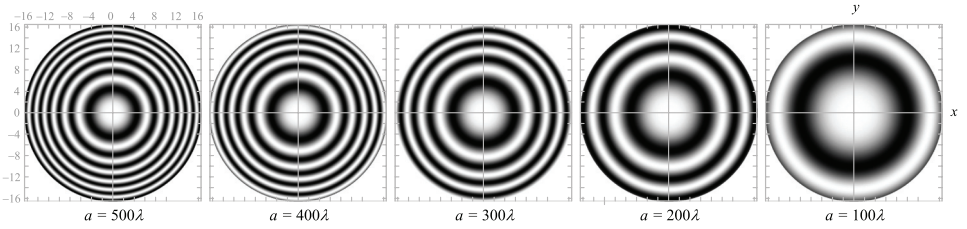


Figure 3.20 Interference patterns for different separations ($a = 500\lambda$, 400λ, 300λ, 200λ, and 100λ; $\lambda = 632.8$ nm) in the interferometer shown in Fig. 3.18 when the observation screen is at $z_0 = 100$ mm. The scale of the axes is in millimeters.

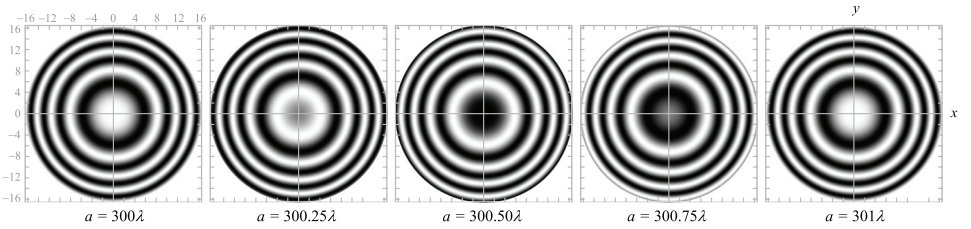


Figure 3.21 Interference patterns for different separations ($a = 300\lambda$, 300.25λ, 300.50λ, 300.75λ, and 301λ; $\lambda = 632.8$ nm) in the interferometer shown in Fig. 3.18 when the observation screen is at $z_0 = 100$ mm. The scale of the axes is in millimeters.

the separation a decreases, leading to an increase in the radius of the rings and therefore fewer rings in the region of observation. And on the contrary, moving the mirror M_2 to increase the separation a led to a decrease in the radius of the rings and therefore more rings in the observation region. When the displacement is followed slowly, in fractions of a wavelength, the effect when a increases is that the rings emerge from the center and the effect when a decreases is that the rings converge toward the center, as shown in Fig. 3.21 for $a = 300\lambda$, $a = (300 + 1/4)\lambda$, $a = (300 + 1/2)\lambda$, $a = (300 + 3/4)\lambda$, and $a = 301\lambda$. An analogous situation holds if $R_2 < R_1$.

Approximation to calculate the radius of the rings

The radii of the fringes that have the maximum irradiance are given by Eq. (3.47). In many practical situations in interferometry, the separation of the secondary sources is found to be much smaller than the radii of curvature R_1 and R_2, i.e., $\Delta R \ll \{R_1, R_2\}$, and the size of the region in which the interference pattern is observed is also much smaller than the radii of curvature R_1 and R_2. In this case, the distance of the virtual sources S_1 and S_2 to a point on the observation screen can be approximated as

$$s_1 = R_1 + \Delta_1 \tag{3.50}$$

and

$$s_2 = R_2 + \Delta_2, \tag{3.51}$$

with

$$\Delta_1 = \frac{\rho^2}{2R_1} \tag{3.52}$$

and

$$\Delta_2 = \frac{\rho^2}{2R_2}, \tag{3.53}$$

where ρ is the radial distance of the observation screen point with respect to $x = 0$ and $y = 0$. Δ_1 and Δ_2 are the distances of the wavefronts of radii R_1 and R_2 to the observation screen point in the second order of approximation.

Thus, the difference $(s_2 - s_1)$ in Eq. (3.44) for a point on the observation screen is

$$s_2 - s_1 = R_2 - R_1 + \frac{\rho^2}{2}\left(\frac{1}{R_2} - \frac{1}{R_1}\right). \tag{3.54}$$

Therefore, the circles of maximum irradiance on the observation screen are obtained when

$$\Delta R\left(1 - \frac{\rho^2}{2R_1R_2}\right) = m\lambda. \tag{3.55}$$

And the radius of the circles as a function of m is

$$\rho_m = \sqrt{2(\Delta R - m\lambda)\frac{R_1R_2}{\Delta R}}. \tag{3.56}$$

Again, the maximum value of m (circular fringe closest to the center) is given by $m = \Delta R/\lambda = a/\lambda$.

Whereas the exact value of the radii of the circles of maximum irradiance can be obtained from Eq. (3.47), Eq. (3.56) serves to obtain the values of the radii of the circumferences in the second-order approximation. To compare the results given by the two equations, recall that $R_1 = z_0 - a/2$ and $R_2 = z_0 + a/2$. The results obtained with Eqs. (3.47) and (3.56) are compared for the first circular fringes when $a = 0.5$ mm and $z_0 = 100$ mm, with $\lambda = 632.8$ nm, in Fig. 3.22.

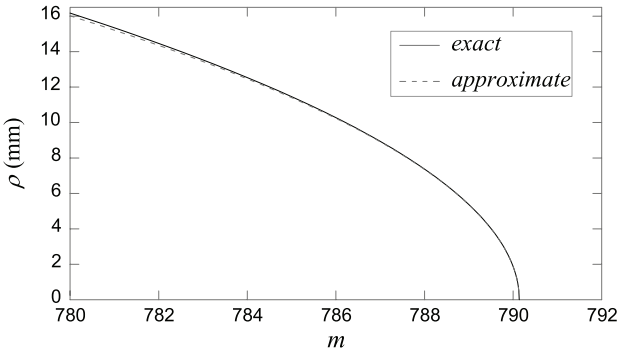


Figure 3.22 Comparison of the radii of circles of constant irradiance calculated with the exact form [Eq. (3.47)] and the approximate form [Eq. (3.56)] for the first 11 circular fringes, when $a = 0.5$ mm and $z_0 = 100$ mm, with $\lambda = 632.8$ nm.

The separation between two consecutive fringes, taken as the difference of the radii of the circles of maximum irradiance, can be evaluated from computing $\rho_m - \rho_{m+1}$. Let us first perform the subtraction of the squares:

$$\rho_m^2 - \rho_{m+1}^2 = 2\lambda \frac{R_1 R_2}{\Delta R}. \tag{3.57}$$

Rewriting the left side as $(\rho_m - \rho_{m+1})(\rho_m + \rho_{m+1})$ and defining the separation between two consecutive fringes as $\Delta\rho = (\rho_m - \rho_{m+1})$ and the mean value of the radii as $\overline{\rho} = (\rho_m + \rho_{m+1})/2$,

$$\Delta\rho = \frac{\lambda}{\overline{\rho}} \frac{R_1 R_2}{\Delta R}. \tag{3.58}$$

Note that we have omitted the subscript m in $\Delta\rho$ and $\overline{\rho}$, because the right-hand side of Eq. (3.57) does not depend on m. Equation (3.58) shows how the separation between fringes decreases inversely proportional to the position (radius) of the fringes as we move away from the center of the interferogram. Finally, if $\Delta R \ll \{R_1, R_2\}$, then $R_1 R_2 \approx z_0^2$, and Eq. (3.58) can also be written as

$$\Delta\rho = \frac{\lambda}{\overline{\rho}} \frac{z_0^2}{a}. \tag{3.59}$$

In practice, $\overline{\rho}$ can be taken as the radius of the dark ring between the two bright fringes for which the separation is to be measured, where a is twice the difference in the separation of the mirrors (with respect to the beamsplitter) and z_0 is the geometrical mean of the radii of curvature of the interfering wavefronts.

3.3.2 Parallel fringe approximation with the Michelson interferometer

If the interferometer shown in Fig. 3.10 is illuminated by spherical waves, keeping the mirror separation the same, i.e., $(d_2 - d_1) = 0$, the secondary point sources S_1 and S_2 would be seen as illustrated in Fig. 3.23. The sources S_1 and S_2 are at the same axial distance from the observation screen but have a transverse separation a', which depends on the angle of tilt $\alpha/2$ of the mirrors.

The interference pattern on the observation screen consists of fringes whose maximum irradiance follows curves resulting from the intersection of the hyperboloids of revolution with the plane on the observation screen, similar to those shown in the example in Fig. 3.17 (except for the names of the axes). In Fig. 3.23, if z_0 is the axial distance between the sources $\{S_1, S_2\}$ and the observation screen, and a' is the separation between S_1 and S_2, the separation of the fringes along the x direction on the observation screen will be given by Eq. (3.48), but the name of the parameters will be changed to

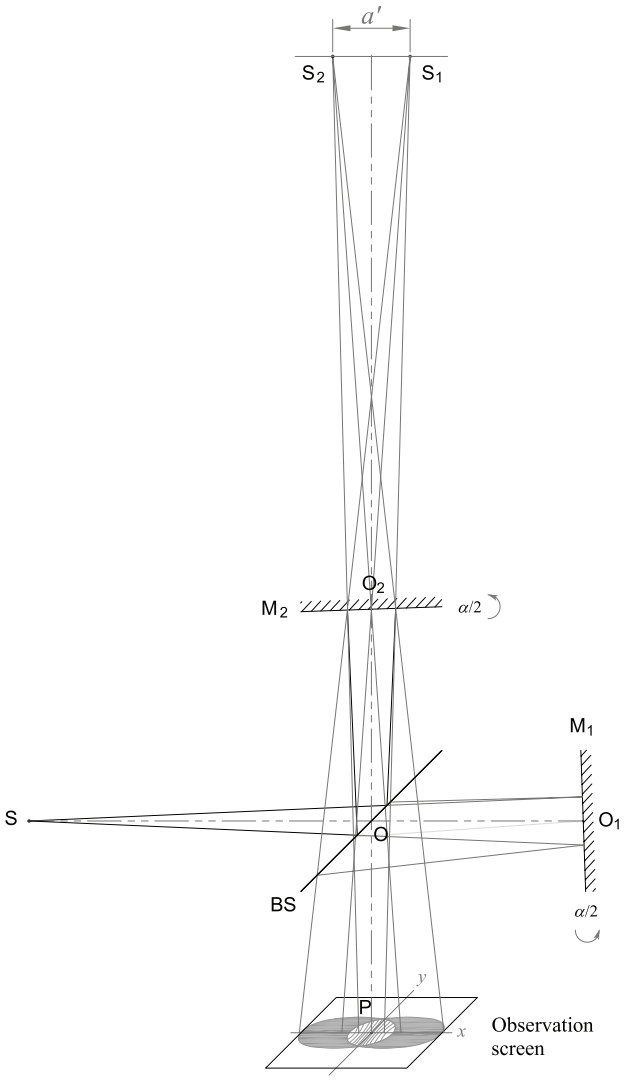


Figure 3.23 A Michelson interferometer with mirrors M_1 and M_2 tilted at a small angle. The two mirrors are at the same distance from the beamsplitter.

$y_0 \to z_0$ and $a \to a'$, and the name of the variable will be changed to $z_m \to x_m$. Therefore, for the geometry of the interferometer shown in Fig. 3.23, the position of the interference fringes in the x direction is

$$x_m = \frac{m\lambda}{2}\sqrt{\frac{4z_0^2}{a'^2 - m^2\lambda^2} + 1}, \tag{3.60}$$

where a' depends on the angle of tilt according to

$$a' = 2(\overline{O_2O} + \overline{OS}) \sin \alpha. \tag{3.61}$$

The distance between point O_2 and the midpoint of sources S_1 and S_2 is $(a'/2)/\tan \alpha$; hence, z_0 is

$$z_0 = \frac{a'}{2 \tan \alpha} + (\overline{O_2O} + \overline{OO_P}). \tag{3.62}$$

In the case where $a' \ll (\overline{O_2O} + \overline{OS})$ and in turn $a' \ll \{R_1, R_2\}$, $\tan \alpha$ can be approximated as $\sin \alpha$. In the observation region where the square of the maximum number of lateral fringes is much smaller than the square of $m_{\max} \equiv a'/\lambda$, Eq. 3.60 can be approximated by $x_m = m\lambda z_0/a'$, which corresponds to a pattern of evenly spaced parallel straight fringes. According to Eq. (3.62), the separation between fringes would be

$$\Delta x = \frac{\lambda}{2 \sin \alpha} + \frac{\lambda}{a'}(\overline{O_2O} + \overline{OO_P}). \tag{3.63}$$

This equation is similar to Eq. (3.37), which was obtained for plane wave interference. The difference is that if in the interferometer shown in Fig. 3.10 the observation screen moves axially, the separation of the fringes does not change. But if in the interferometer shown in Fig. 3.23 the viewing screen is moved axially (changing $\overline{OO_P}$), the fringe spacing changes. The addend $\lambda(\overline{O_2O} + \overline{OO_P})/a'$ acts as a scale factor in the separation of the fringes. The first addend of Eq. (3.63) is identical to Eq. (3.37).

In conclusion, the wavefronts that reach the observation screen in the interferometer shown in Fig. 3.23 resemble plane wavefronts when

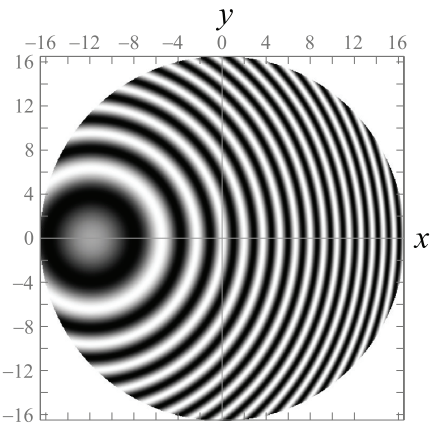


Figure 3.24 An interference pattern generated by spherical waves in a Michelson interferometer when one of the mirrors is tilted. The axial separation of the sources S_1 and S_2 is $a = 400\lambda$, and the lateral displacement of the source S_2 is x_0 $(= a'/2) = 0.03$ mm. The scale of the axes is in millimeters.

$a' \ll \{R_1, R_2\}$ and the observation region has dimensions such that $m^2 \ll m_{\max}^2$, where m is the number of lateral fringes.

Finally, if only one of the mirrors in the interferometer shown in Fig. 3.18 is rotated, a lateral displacement of the center of the rings is observed, as shown in Fig. 3.24, where the axial separation of the sources S_1 and S_2 is $a = 400\lambda$ and the mirror M_2 rotates by an angle that laterally displaces the source S_2 by $x_0 (= a'/2) = 0.03$ mm. If the angle of tilt is increased, the center of the rings moves out of the observation region and the fringe patterns approach straight parallel fringe patterns.

3.4 Practical Aspects in the Michelson Interferometer

In the previous sections of this chapter, the Michelson interferometer was used as an optical tool to study, in some detail, the interference patterns generated by the superposition of plane waves and spherical waves. The Michelson interferometer is one of many interferometers that can be used for this purpose. In any case, the goal is to generate two secondary waves from one primary wave, so that interference is guaranteed (provided that the optical path difference between the beams is not greater than the coherence length). The use of an interferometer goes beyond the explanation of the interference patterns of plane or spherical waves. Its power is applied in the evaluation of surfaces or wavefronts, either to characterize the same surfaces or to measure some other parameter such as the refractive index or optical aberrations. For this type of application, one of the beams is taken as the reference beam, while the other beam is used to analyze the optical element to be measured. Most commonly, the reference beam is flat or spherical. In practice, some drawbacks arise with the generation of the reference beam due to the optical quality of the interferometer's components and its configuration. In this section, some practical elements of the Michelson interferometer are considered: the point source, the collimating lens, the mirrors, the peak–valley error (as a measure of the quality of the optical surfaces), and an example of a Michelson interferometer.

Point source

Let us suppose that we want to illuminate the interferometer with a plane wave, as shown in Fig. 3.8. To do this, we should have a point source S. This is already a first precaution because physically we can build small sources, but not point sources (mathematically speaking), and with a very narrow spectral band (laser), but not a Dirac delta function. Now let us not dwell on that and assume we have a quasi-monochromatic source* small enough to be considered a point source.

*A source is said to be nearly monochromatic if its bandwidth $\Delta\lambda$ is much smaller than the value of the wavelength centered on the bandwidth, i.e., $\Delta\lambda/\overline{\lambda} \ll 1$.

Collimating lens

The next thing is to select a collimating lens to generate a plane wave. A first option could be a simple lens (positive or negative) with spherical faces, aligned and with its primary focus coinciding with the point source. Because the point source is on the optical axis, the only aberration present will be spherical. Although the aberration can be reduced by optimally shaping the lens (correctly choosing the radii of curvature of the lens faces), it is not possible to eliminate this aberration. A better option is to use an achromatic doublet. In addition to correcting chromatic aberration for two colors (spectral lines F and C, Appendix C), this type of lens, greatly reduces spherical aberration. This may already be a good solution. But if full correction of spherical aberration is desired, an aspherical lens for finite (primary focal point) and infinite conjugates can be designed. This solution is not always possible due to its high cost. Finally, there is an aspect that has not been mentioned, which has to do with the finite extension of the wavefront. In the previous sections, it was assumed that the wavefront has a circular edge (as is usually the case, either by the edge of the collimating lens or by a diaphragm that is placed before or after the lens to determine the size of the cross section of the beams). This physical limitation on the wavefront extension causes edge diffraction, so the wavefront is not strictly a plane in its entirety.

Beamsplitter

This element has been represented with a diagonal line, which of course is another idealization. In practice, this element is usually a plate with parallel faces made of glass or another material. When an oblique ray hits the plate, the ray refracted to the other side of the plate exits at an angle equal to that of the incident ray, but with a lateral shift, as shown in Fig. 3.25. This is not a problem when the interferometer is illuminated with collimated light (Fig. 3.8), because any ray associated with the wavefront is shifted by the same amount, so the output remains a plane wavefront. However, when illuminated with a spherical wave (Fig. 3.18), the rays are refracted according to the angle of incidence, giving rise to an aberrated wavefront.

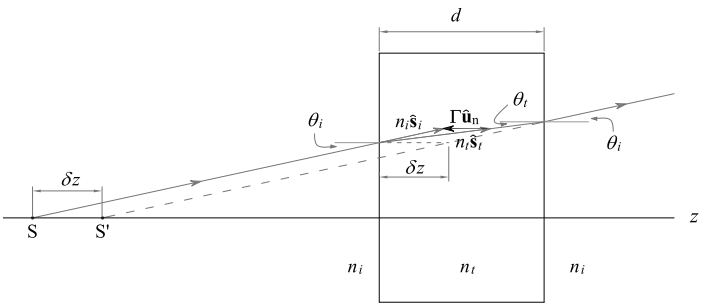


Figure 3.25 Deviation of the ray in a plate of parallel faces.

To calculate the lateral shift of the refracted ray, suppose that the plate of parallel faces is separated by a distance d and is made of a material with refractive index n_t. Let us place a point source S at a certain distance from the plate. The line orthogonal to the faces, which passes through S, defines the z axis. Now consider a ray diverging from S with a tilt angle θ_i, as shown in Fig. 3.25. The back projection of the refracted ray leaving the second face passes through the point S′. Therefore, an observer behind the plate will see that the refracted ray comes from point S′ and not from S. The separation between S and S′ is the measure of the z-axis deviation of the refracted ray, $\delta z = \overline{\mathrm{S'S}}$. At the point of incidence of the first face of the plate, the vectors $n_i\hat{\mathbf{s}}_i$ and $n_t\hat{\mathbf{s}}_t$ associated with the incident and refracted rays satisfy the vector form of Snell's law [Eq. (2.75)], so $\Gamma = n_t \cos\theta_t - n_i \cos\theta_i$. Considering similar triangles,

$$\frac{\Gamma}{n_t} = \frac{\delta z}{d/\cos\theta_t}, \tag{3.64}$$

then

$$\delta z = d\left(1 - \frac{n_i \cos\theta_i}{n_t \cos\theta_t}\right). \tag{3.65}$$

The deviation from $\delta z|_{\theta i=0} = d(1 - n_i/n_t)$ is shown as a function of the angle of incidence θ_i for $d = 1$ mm, $n_i = 1$, and $n_t = 1.5168$ (BK7 glass), with $\lambda = 587.56$ nm, in Fig. 3.26. The curve shows that as the angle of incidence increases, point S′ moves away from S and toward the plate.

In interferometers, the beamsplitter plates can be nitrocellulose membranes (pellicles), thin plates (a few millimeters thick), and cubes formed by two right prisms (Fig. 3.27). In all three options, one side will have a thin film coating to increase reflection and create beamsplitters of, e.g., 50% reflection

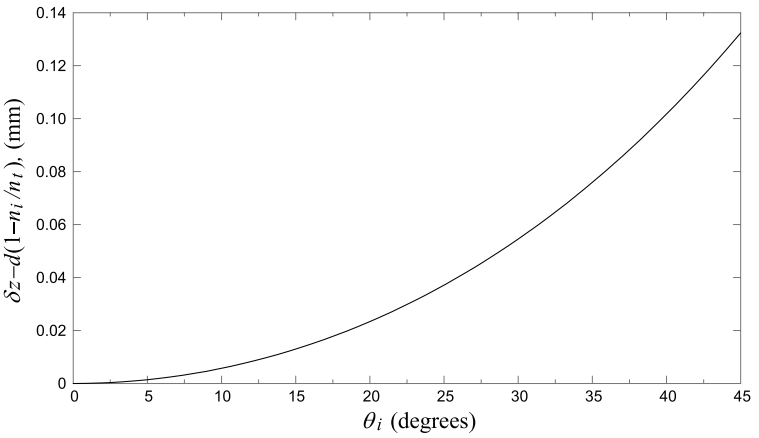


Figure 3.26 Deviation of a refracted ray in a plate of parallel faces with a thickness equal to 1 mm.

Figure 3.27 Beamsplitters from Thorlabs, Inc (www.thorlabs.com): (a) pellicle, (b) plate, and (c) cube.

and 50% transmission. The pellicles are very thin (~2 μm), so the deflection of the refracted rays can be considered negligible. This element is very close to the idealization that we have made of the beamsplitter. The drawback is that they are elements that must be handled carefully so as not to break the membrane and must be protected from dust. They are a good solution in closed systems free of mechanical vibrations. Thin plates are very common in interferometers. One of the beams will pass through the plate once, while the other must pass through it three times, which must be taken into account when using low-coherence sources. Cube beamsplitters are made up of two right prisms joined at their diagonal faces (Appendix E). One of these faces is coated with a thin film to increase reflection. Unlike thin plates, the two rays pass through the cube the same number of times.

In short, by using collimated light to illuminate the interferometer, any of these beamsplitters can be used. But if the light used diverges from a point source, the best option would be the pellicle. However, the usual option is a thin plate. The cube beamsplitter is not a good choice if the radius of curvature of the wavefront reaching the cube is comparable to the side of the cube. Because the deflection of rays in a plate is a function of the angle of incidence, this effect is similar to the spherical aberration that affects rays converging to form an image on a spherical refracting surface, as shown in Fig. 1.88. So, in a thick plate, such as the dividing cube, the end result is that an incident spherical wavefront upon refraction takes a shape other than spherical, i.e., an aberrated wavefront (aspherical wavefront). For the point source in the optical axis of the interferometer shown in Fig. 3.18, the aspherical wavefront will have symmetry of revolution, so the interference patterns generated on the observation screen will be rings of interference, but their spatial distribution will be different from that of spherical wave interference.

Mirrors

After the beamsplitter, mirrors M_1 and M_2 follow. They are flat mirrors, i.e., sheets of glass with the flat face covered with a metallic film, usually aluminum, but silver and gold are also found. To protect the metallized face

from oxidation, dust, or dirt (fingerprints), a thin (half-wavelength) film of dielectric material, such as silicon oxide, is usually deposited. This is a layer that allows the surface to be cleaned.

Peak–valley error

Now let us consider an aspect that affects all the optical elements previously mentioned. In practice, what is meant by a spherical or flat optical surface? In the manufacturing process, polished surfaces (like a mirror) very close to the mathematical surface design can be obtained. The deviation between the real surface and the mathematical surface gives a measure of the optical quality of the surface. For example, an optical plane of precision λ/4 refers to a polished surface whose deviations from a reference plane do not exceed one-fourth of the nominal wavelength. Although λ/4 may seem like a small quantity, it implies a considerable change in the irradiance distribution. A change in λ/2 implies going from a bright area to a dark area. Let us suppose that the interferometer shown in Fig. 3.8 is wanted, with optical elements of quality λ/4. By adjusting the distance of the mirrors from the beamsplitter to have an optical path difference equal to $2(d_2 - d_1) = m\lambda$, a fully illuminated region is expected on the observation screen. However, this is not the case, but rather an illuminated region with some less luminous and even dark areas will be observed. For best results, the quality of the optical elements should be changed, e.g., to a precision of λ/10, λ/20, or λ/100.

3.4.1 Laboratory interferometer

A Michelson interferometer built in a laboratory with optical elements of precision λ/4 is shown in Fig. 3.28. The point source is a laser beam (He-Ne, $\lambda = 632.8$ nm) focused on a small hole (*pinhole*) of about 15 μm. The objective of focusing it on the hole is to eliminate the high spatial frequencies present in the beam, thus obtaining an approximately homogeneous illumination. This is discussed in Chapter 4. A circular stop is then placed to set the extent of the interferogram to approximately 25 mm in diameter when the light is collimated. Then there is the collimating lens, which in this case is an achromatic doublet with a focal length of 300 mm. This lens is mounted on an axial slider to facilitate beam collimation, which is achieved when the primary focus of the lens coincides with the point source. The axial slider also makes it easy to change the beam collimation to obtain spherical wavefronts. At 225 ±1 mm from the lens is the center of a beamsplitter cube (point O in Fig. 3.8), with a 50 mm side made of BK7 glass. Next, the flat aluminum mirrors are placed, as shown in the figure, 125 ± 1 mm from the center of the cube. Mirror 2 is mounted on another axial slider with which the optical path difference between the rays in the interference region can be changed. Each of the mirrors is supported by a mount that has two fine-thread screws with which the mirror can be tilted. Finally, there is the observation screen,

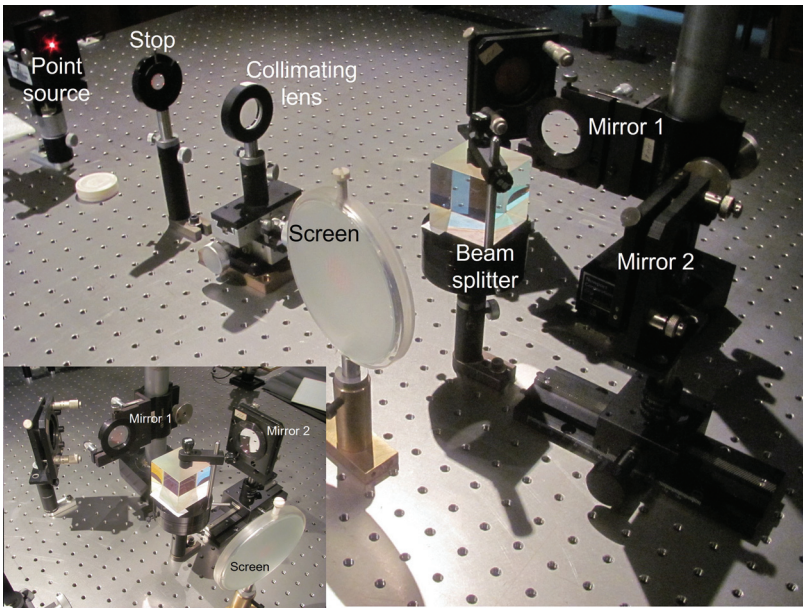


Figure 3.28 A Michelson interferometer. Mirror 2 is mounted on an axial slider.

125 ± 1 mm from the center of the cube. The observation screen is ground glass (translucent). Interferograms on the observation screen are recorded with a photographic camera (not shown in the figure).

With this interferometer, the interference patterns shown in Fig. 3.29 are generated. In (a), the two mirrors are approximately the same distance from the center of the beamsplitter cube and, furthermore, the light beams are collimated and aligned (the path difference between the collimated beams is

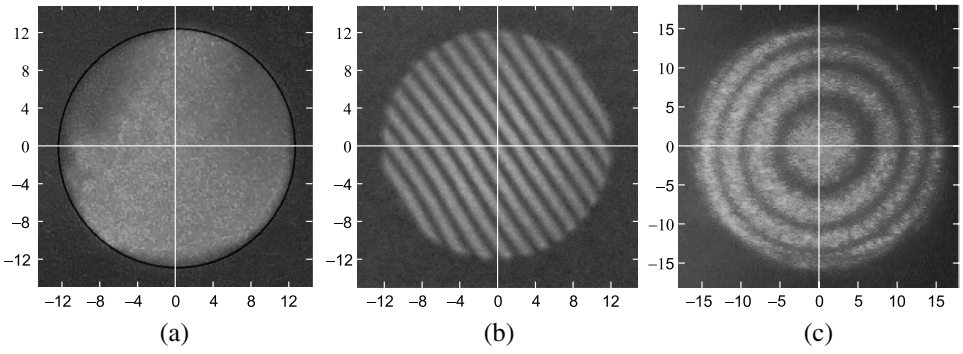


Figure 3.29 Experimental interference patterns. In (a) and (b) the beams are collimated, and in (c) the beams are divergent spherical. (a) The optical path difference is approximately zero, and the mirrors are aligned; (b) the optical path difference is approximately zero, and one of the mirrors is tilted at a small angle; (c) keeping the mirrors aligned, the collimating lens moves away from the beamsplitter cube (40 mm) and the mirror M_1 also moves away about 60 mm. The scale of the axes is in millimeters.

approximately zero). However, the observation region does not have a homogeneous or symmetric illumination. This is an example of how the precision of the surfaces of the optical elements used to generate the interferogram (of λ/4) affects its quality. To obtain this interferogram, mirrors were adjusted until the illuminated region without interference fringes was obtained. In (b), a pattern of parallel straight fringes is obtained by tilting the mirror 2 at a small angle. In reality, the fringes are somewhat distorted, which is more noticeable on the lower left side. This is also a result of the quality of the optical surfaces of the interferometer and shows that the two interfering wavefronts are not strictly flat. In (c), the mirror 2 moves axially about 60 mm away from the beamsplitter cube and the collimating lens moves about 40 mm away from the cube. By moving the collimating lens, the wavefront exiting the lens is divergent spherical and its center of curvature will be 1950 mm from the lens. Now the lens is about 265 mm from the center of the cube.

Taking into account these distances, the distance of the observation screen, and the additional optical path due to the beamsplitter cube (with refraction index ~1.51), $R_1 = 2641$ mm and $R_2 = 2761$ mm. Therefore, $\Delta R = 120$ mm, where $m_{\max} = \Delta R/\lambda = 189633$ is the value of m corresponding to the central zone shown in Fig. 3.29(c). In this figure, three circular fringes are counted and the radius of the third one is estimated to be about 15 mm. This can be verified from Eq. (3.56), by substituting $\Delta R = m_{\max}\lambda$ and $m = m_{\max} - 3$ for the third fringe in

$$\rho_{m_3} = \sqrt{\frac{6\lambda R_1 R_2}{\Delta R}} = 15.19 \text{ mm},$$

which agrees very well with the experimental estimate.

In conclusion, a real planar reference beam will be a wavefront distorted by an amount similar to the deformations of the optical surfaces of the beamsplitter and the planar mirror that sends the reference beam to the observation region. For a spherical reference beam, the effect of the thickness of the beamsplitter must also be taken into account. The other beam of the interferometer, also initially affected by the beamsplitter, can be used to assess the optical quality of an element, e.g., a flat or spherical mirror. The optical quality of the reference beam will determine the accuracy with which the surfaces under test can be measured. In the example shown in Fig. 3.29(c), the cube effect is negligible because the angle of the marginal ray turns out to be 0.37°, which gives an axial deviation of the ray of 0.34 μm.

3.5 Interference in a Plate of Parallel Faces

In the interferometer shown in Fig. 3.8, what an observer sees from the observation screen is a pair of parallel reflecting flat surfaces separated by a distance $2(d_2 - d_1)$. The light reflected by the two surfaces generates an

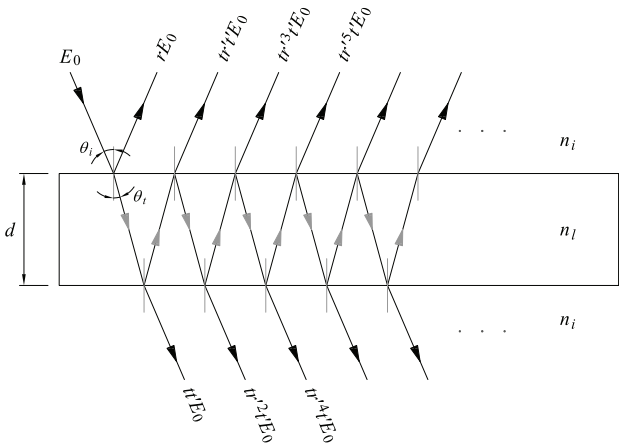


Figure 3.30 Reflection and transmission in a plate of parallel faces.

interference pattern in the plane of observation. A system that emulates the previous one is a plate with parallel faces. Now the reflected and transmitted light will be on the faces of the plate, and it is possible to observe interference on both sides of the plate: if it is on the same side as the light source, it is said to be reflection interference, but if it is on the side where there is no light source, it is said to be transmission interference. To see this, let us consider Fig. 3.30, which shows the incidence of a ray with an angle θ_i on a plate of refractive index n_l and thickness d, immersed in a medium of index n_i.

At each of the interfaces, there will be multiple reflected and transmitted rays. The amplitudes of the waves associated with light rays are given by the Fresnel equations, Eqs. (2.96–2.99). In Fig. 3.30, r and t denote the reflection ($r_\perp$ or $r_\parallel$) and transmission ($t_\perp$ or $t_\parallel$) coefficients when light passes from the medium of refractive index n_i to the medium of refractive index n_l, and r' and t' denote reflection and transmission coefficients when the light passes from the medium of refractive index n_l to the medium of refractive index n_i.

3.5.1 Stokes relations

The relations between r and r' and between t and t' are obtained from the Stokes relations, which are deduced from the illustrations in Fig. 3.31. In (a), the reflection and transmission of a ray of amplitude E_0 incident at an angle θ_i at an interface separating two media of refractive indices n_i and n_t are shown. The amplitude of the reflected wave would be rE_0, and the amplitude of the transmitted wave would be tE_0. Taking into account the principle of reversibility (or reciprocity), if two rays are sent in opposite directions to the reflected and transmitted rays in (a), with amplitudes rE_0 and tE_0, respectively, the incident beam with amplitude E_0 must again be obtained but going in the opposite direction. In (b), the incident ray is opposite the reflected ray in (a), but has the same amplitude. This ray will have a reflection and a

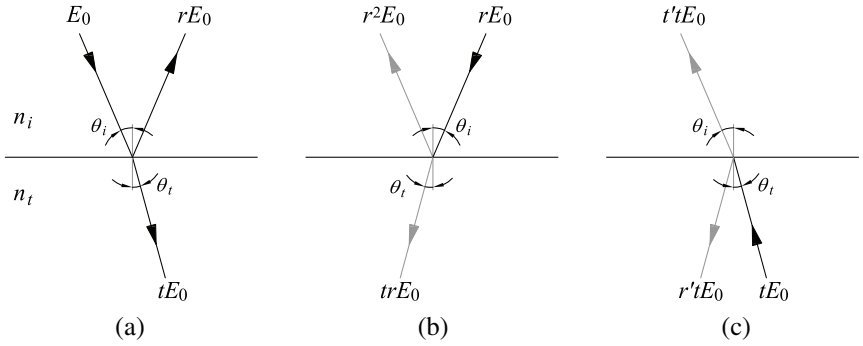


Figure 3.31 Stokes coefficients. Reflection and transmission of (a) a wave of amplitude E_0, (b) a wave of amplitude rE_0, and (c) a wave of amplitude tE_0.

transmission with amplitudes r^2E_0 and trE_0, respectively. On the other hand, in (c), the incident ray is opposite the transmitted ray in (a), but has the same amplitude. This ray will also have a reflection and a transmission, but with amplitudes $r'tE_0$ and $tt'E_0$, respectively. If the principle of reversibility holds, then

$$r^2 + tt' = 1 \tag{3.66}$$

and

$$tr + r't = 0. \tag{3.67}$$

Therefore,

$$tt' = 1 - r^2 \tag{3.68}$$

and

$$r' = -r. \tag{3.69}$$

These last two equations are called Stokes relations and are especially useful for adding the amplitudes of the multiple reflected and transmitted waves shown in Fig. 3.30. The reflected/transmitted irradiance will be the square modulus of the sum of the reflected/transmitted waves, including the optical path difference between consecutive waves.

3.5.2 Multiple-wave interference

This section will consider the superposition of the multiple reflected and transmitted waves (rays) shown in Fig. 3.30.

Interference by reflection

For the reflected irradiance,

$$I_r = \frac{\epsilon_0 c}{2}(E_r E_r^*), \tag{3.70}$$

with

$$E_r = rE_0 + tr't'E_0 e^{i\delta} + tr'^3 t' E_0 e^{i2\delta} + tr'^5 t' E_0 e^{i3\delta} + \cdots +, \tag{3.71}$$

and E_r^* the conjugate of E_r. The phase δ depends on the optical path difference Λ between two consecutive reflected rays, i.e., $\delta = 2\pi\Lambda/\lambda$. Now, Eq. (3.71) can be rewritten as

$$E_r = rE_0 + tt'r'E_0 e^{i\delta}[1 + r^2 e^{i\delta} + (r^2)^2 e^{i2\delta} + (r^2)^3 e^{i3\delta} + \cdots +]. \tag{3.72}$$

And in a more compact form,

$$E_r = rE_0 + tt'r'E_0 e^{i\delta} \sum_{q=0}^{Q} (r^2 e^{i\delta})^q, \tag{3.73}$$

where $q = 0, 1, 2, \ldots$ defines the reflection $q+2$, and $Q+2$ is the total number of reflections. The sum in Eq. (3.73) represents a geometrical series whose ratio is $r^2 e^{i\delta}$ and with a modulus ≤ 1. The sum of such series is given by

$$\frac{1 - (r^2 e^{i\delta})^{Q+1}}{1 - r^2 e^{i\delta}}. \tag{3.74}$$

If the number of reflections is much greater than 1 (there will be infinite reflections if $\theta_i = 0$ or if the plate has infinite extension), the sum reduces to $1/(1 - r^2 e^{i\delta})$. With this in mind and taking the Stokes relations into account, the reflected field will be

$$E_r = rE_0 \left[\frac{1 - e^{i\delta}}{1 - r^2 e^{i\delta}}\right]. \tag{3.75}$$

Finally, the irradiance of the multiple reflected waves [Eq. (3.70)] is given by

$$I_r = I_0 \left[\frac{4R\sin^2(\delta/2)}{(1-R)^2 + 4R\sin^2(\delta/2)}\right], \tag{3.76}$$

where $I_0 = \epsilon_0 c (E_0)^2/2$ and the reflectance $R = r^2$.

To calculate the phase δ, let us consider Fig. 3.32. The optical path difference between the two reflected beams would be

$$\Lambda = n_l(\overline{AB} + \overline{BC}) - n_i \overline{AD}. \tag{3.77}$$

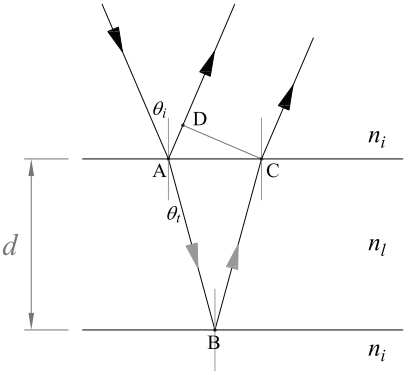


Figure 3.32 Geometry for calculating the optical path difference between two consecutive reflections.

From similar triangles, $\overline{AB} = \overline{BC} = d/\cos\theta_t$ and $\overline{AD} = (2d\tan\theta_t) \times \sin\theta_i$. Then the phase $\delta = 2\pi\Lambda/\lambda$ is given by

$$\delta = \frac{2\pi}{\lambda}\left(\frac{2n_l d}{\cos\theta_t} - (2d\tan\theta_t)n_i\sin\theta_i\right). \tag{3.78}$$

Using Snell's law and simplifying,

$$\delta = \frac{2\pi}{\lambda}(2n_l d\cos\theta_t). \tag{3.79}$$

Interference by transmission

For the case of transmitted waves (Fig. 3.30), the resulting sum is

$$E_t = tt'E_0 + tr'^2t'E_0e^{i\delta} + tr'^4t'E_0e^{i2\delta} + \cdots +, \tag{3.80}$$

i.e.,

$$E_t = tt'E_0[1 + r^2e^{i\delta} + (r^2)^2e^{i2\delta} + \cdots +]. \tag{3.81}$$

Analogous to the derivation of the reflected multi-wave irradiance, the transmitted multi-wave irradiance is given by

$$I_t = I_0\left[\frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\delta/2)}\right]. \tag{3.82}$$

Equations (3.76) and (3.82) are complementary when there is no absorption.

Case 1. Plane wave interference.

A first result that can be observed is when the plate is illuminated by (coherent) monochromatic plane waves in a direction orthogonal to the plate,

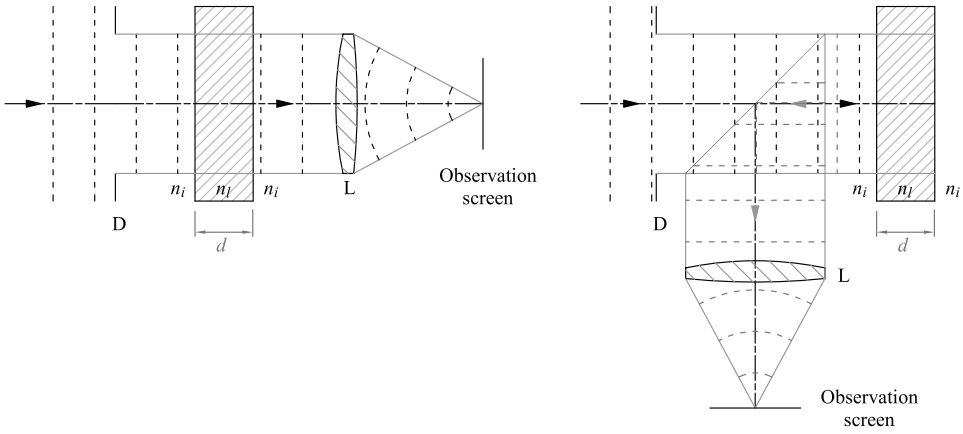


Figure 3.33 Axial interference of multiple plane waves on a plate with parallel faces: (a) transmitted and (b) reflected.

i.e., when $\theta_i = 0$ and therefore $\theta_t = 0$. In Fig. 3.33(a), a situation is shown in which a plane wave bounded by a diaphragm D is incident orthogonally to a plate of thickness d. If the interference behind the plate is observed in one plane (parallel to the plate), the situation is similar to that of the Michelson interferometer shown in Fig. 3.8 and a homogeneously illuminated region will be seen. But if a lens L is added and the observation screen is moved to the secondary focal plane of the lens, a bright spot will be observed. In both cases, the irradiance would be given by Eq. (3.82) and depends on the thickness of the plate. In Fig. 3.33(b), the interferometer is designed to observe the interference of the multiple waves reflected by the faces of the plate. With the help of the beamsplitter, the reflected waves are deflected toward the lens L, which focuses the light on the observation screen. The irradiance would be given by Eq. (3.76).

For $\gamma(\tau) = 1$, Fig. 3.9 shows how the transmitted irradiance in the observation plane changes as the separation between the mirrors changes. In this case, the irradiance contrast is 1. By changing the thickness of the plate in Fig. 3.33(a), a modulation of the irradiance in the observation plane (focal plane) is also obtained. Figure 3.34 shows the value of the irradiance in the focal plane as a function of the thickness of the plate (on the order of the wavelength) for three values of the reflection coefficient r: 0.2, 0.56, and 0.96.

In particular, the reflection coefficient $r = 0.2$ ($r_{\parallel} = -r_{\perp}$) is obtained for a glass plate ($n_l = 1.5$, $\lambda = 632.8$ nm) immersed in air, when $\theta_i = 0$. The modulation of the irradiance behaves similarly to that of the irradiance shown in Fig. 3.9, for $\gamma(\tau) = 1$, in the Michelson interferometer. The difference is in the contrast of the modulation, given that in the case of the plate, it is low, equal to 0.08. This resemblance is no accident. When $r = 0.2$, then

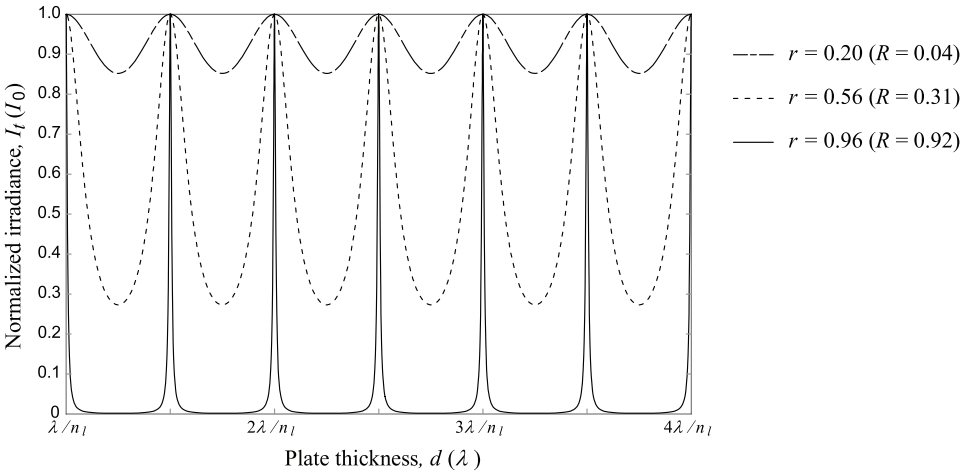


Figure 3.34 Modulation of the irradiance of multiple transmitted waves in the focal plane [Fig. 3.33(a)] when the reflection coefficient r of the faces of the plate is 0.2, 0.56, and 0.96.

$r^2 = 0.04$ and $r^4 = 0.0016$, so $(r^2)^2 e^{i2\delta}$, and the other higher order addends do not contribute significantly to the sum of Eq. (3.81). Thus, in this low reflectance case ($R = 0.04$), the interference is determined only by the first two transmitted beams, i.e.,

$$E_t = tt' E_0[1 + r^2 e^{i\delta}], \tag{3.83}$$

and the irradiance will be (omitting the term r^4)

$$I_t = I_0(1 - R)^2 \left[1 + 2R\cos\left(\frac{4\pi}{\lambda} n_l d\right)\right]. \tag{3.84}$$

This expression is analogous to Eq. (3.29) of the interference of two plane waves when $|\gamma| = 1$ and $\tau = 2(d_2 - d_1)/c$.

Thus, when the reflectance of the faces of the plate is low (as in a glass plate), the effective interference of the multiple transmissions resembles the interference of two waves. If the reflectance increases, other addends will be added and the result will move away from that corresponding to two waves. To increase reflectance, one option is to increase the refractive index of the plate. In practice, however, not much can be improved, as high refractive indices are around 2.5 (e.g., in diamonds), giving reflectances of around 0.18. To achieve reflectances such as those shown in Fig. 3.34, of 0.31 ($r = 0.56$) or 0.92 ($r = 0.96$), thin metallic or dielectric films are deposited on the faces of the plate.

In the case where the reflection coefficient is $r = 0.96$, the interference result changes remarkably compared with the case where $r = 0.2$. The maxima irradiance (I_0) is obtained when the thickness of the plate is a multiple of half a

wavelength divided by the refractive index of the plate (same as for two waves), but the irradiance values decay rapidly when a little change in the thickness of the plate is produced. In other words, the most prominent irradiance values are found only in very narrow bands.

In the case of multiple reflected waves [Fig. 3.33(b)], an analogous situation is obtained, as shown in Fig. 3.35, for the same values of the reflection coefficient. Irradiance minima (zero) are obtained when the thickness of the plate is a multiple of half a wavelength divided by the refractive index of the plate (same as for two waves). Now when $r = 0.2$, the irradiance contrast is equal to 1, as in Fig. 3.9, for $\gamma(\tau) = 1$. And when $r = 0.96$, there are narrow bands where the irradiance is around zero, but then irradiance values increase rapidly when the thickness of the plate changes a bit. This is the working principle of antireflection thin films.

Note that the transmission (Fig. 3.34) and reflection (Fig. 3.35) graphs are complementary, resulting in conservation of energy if there is no absorption in the plate.

Case 2. Interference of spherical waves I

Let us now consider a spherical wave incident on the plate with parallel faces. As we have already seen, there is a pattern of circular fringes when the Michelson interferometer is illuminated with a spherical wave. Something similar happens in the case of the plate, but the structure of the fringes depends on the reflectance of the faces of the plate. Consider the configuration shown in Fig. 3.36 in which the transmission interference fringes are observed.

In (a), the interference at a point on the screen is caused by the superposition of rays arriving at different angles. In (b), a lens is inserted into the setup and the observation screen is placed in the secondary focal

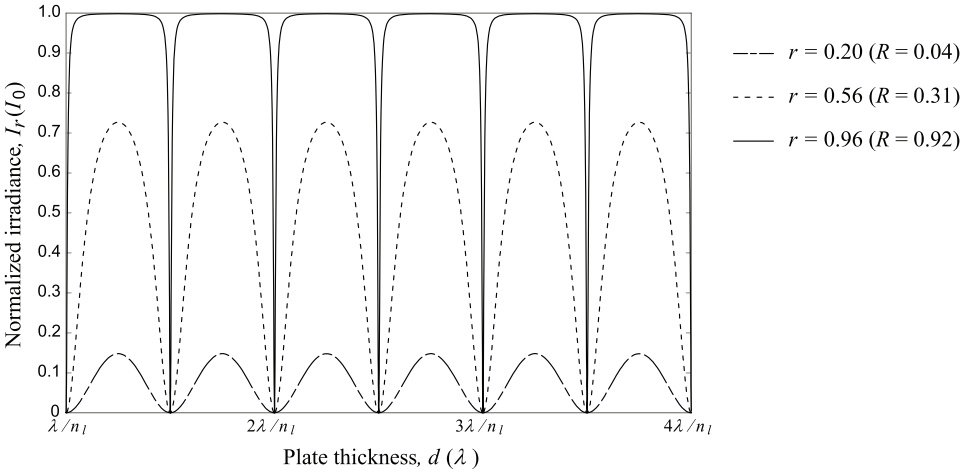


Figure 3.35 Modulation of the irradiance of multiple reflected waves in the focal plane [Fig. 3.33(b)] when the reflection coefficient r of the faces of the plate is 0.2, 0.56, and 0.96.

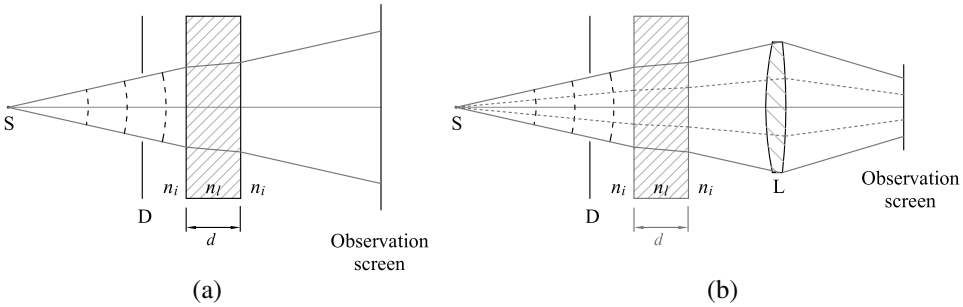


Figure 3.36 Interference of multiple transmitted waves on a plate with parallel faces when illuminated by a spherical wave. (a) Nonlocalized fringe formation. (b) Localized fringe formation.

plane of such a lens. The interference now occurs at one point on the observation screen due to overlapping rays hitting the lens at the same angle. Whereas in (a) the circular interference fringes are observed for any position of the observation screen along the optical axis (nonlocalized fringes), in (b) the circular interference fringes are observed focused in the focal plane of the lens (localized fringes).

First, let us see the formation of the interference fringes in configuration (b), since the required algebra has already been developed in Eq. (3.82). To obtain this equation, the incidence of a ray with a certain angle was considered in Fig. 3.30. Then, from Fig. 3.33(a), the specific case of $\theta_i = 0$ was examined: the irradiance given by Eq. (3.82) is observed at the secondary focal point. If now in the configuration shown in Fig. 3.33(a) we assume that $\theta_i \neq 0$, the interference of the multiple waves of the inclined plane θ_i will be seen at an off-axis bright point at the radial position $f \tan \theta_i$, where f is the lens focal distance. This is what happens with the multiple transmitted rays corresponding to a divergent ray from S in Fig. 3.36(b), with $\theta_i \neq 0$: the transmitted rays will be focused at a point located at $f \tan \theta_i$. Taking into account the symmetry of revolution around the optical axis, what we would have is a pattern of rings. The radii of the circles of maximum irradiance will be given by

$$\rho_m = f \tan \theta_m, \tag{3.85}$$

where

$$\sin \theta_m = \frac{\sqrt{n_l^2 - (m\lambda/2d)^2}}{n_i}, \tag{3.86}$$

according to Eq. (3.79), when $\delta = 2m\pi$. The maximum value of m, $m_{\max}$, will be the nearest integer to $2n_l d/\lambda$. Moving away from the center, the value of m decreases. Thus, with $m = m_{\max} - 1, m_{\max} - 2, \ldots$, the interference rings of the center are identified.

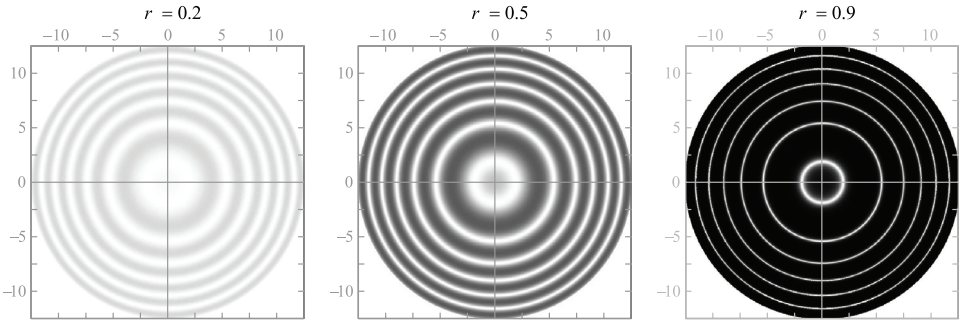


Figure 3.37 Transmission interferograms generated by a plate with an optical thickness of 0.5 mm for $r = 0.2$, $r = 0.5$, and $r = 0.9$, using a lens with a focal length of 65.65 mm. The scale of the axes is in millimeters.

To compare the interferograms generated by spherical waves with the Michelson interferometer and with the parallel plate interferometer, let us assume that in the Michelson interferometer shown in Fig. 3.18 the separation of the mirrors is $a/2 = 0.25$ mm and the glass plate has a thickness $d = (a/2)/n_l$, with $n_l = 1.5$ (for $\lambda = 632.8$ nm). The interferogram obtained with the Michelson interferometer, when the distance between the midpoint of the virtual sources and the observation screen is $z_0 = 100$ mm, is shown in Fig. 3.16. And the interferograms that are obtained when using a lens of focal length $f = 65.65$ mm for $r = 0.2$ (glass plate without reflective coatings), $r = 0.5$, and $r = 0.9$ are shown in Fig. 3.37. The focal length was set to this value taking into account the angle subtended by the first interference ring in the pattern shown in Fig. 3.16. The radius of this ring is $\rho_{790} = 1.88$ mm. On the other hand, for the plate, the corresponding angle with the first ring is $\theta_{790} = 1.61°$ [Eq. (3.86)], which gives a focal length equal to 65.65 mm [Eq. (3.85)].

In the interferograms shown in Fig. 3.37, the position of the rings (the radius of the circles of maximum irradiance) is the same. This is because they do not depend on the reflection coefficient [Eqs. (3.85) and (3.86)]. However, the reflection coefficient determines the width of the circular fringes. The greater the effective number of interfering beams (which occurs as the reflection coefficient increases), the smaller the width of the fringes. When $r = 0.2$, the pattern looks like the two-spherical wave pattern shown in Fig. 3.16, but the contrast is low, $C = 0.08$. In the plate, as the reflectance of the faces increases, the contrast of the fringes also increases. Thus, with $r = 0.5$, $C = 0.47$; and with $r = 0.9$, $C = 0.98$.

For reflection interference, the configuration shown in Fig. 3.33(b) can be used, but illuminating with a spherical wave. In this case, the irradiance maxima occur when $\delta = 2m\pi \pm \pi$. Reflection interference patterns generated with the same parameters as those used in the transmission interference example are shown in Fig. 3.38. As expected, these interferograms are the

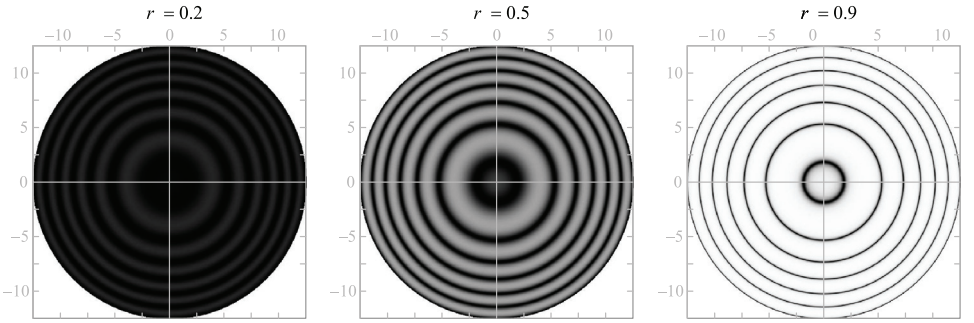


Figure 3.38 Reflection interferograms generated by a plate with an optical thickness of 0.5 mm for $r = 0.2$, $r = 0.5$, and $r = 0.9$. The scale of the axes is in millimeters.

complement of the transmission interferograms. Unlike the transmission interferograms, the contrast of the interference fringes in the reflection is $C = 1$ because the dark areas have zero irradiance value.

The irradiance of the interferogram shown in Fig. 3.37(a) was calculated using $r = 0.2$ regardless of the angle. Strictly speaking, this should not be the case because the reflection and transmission coefficients depend on the angle, according to the Fresnel equations [Eqs. (2.96–2.99)]. However, in the example under consideration, the range of the angle of incidence with which the interference pattern is generated goes from 0 to 10°, and for that range the value of r changes very little, around 2% (Fig. 2.13). Even for larger angles, the variation remains small, so for the first interference rings it suffices to take the reflection coefficient equal to that of $\theta_i = 0$. This also applies to plates with reflective coatings.

At first glance, the interferograms shown in Figs. 3.16 and 3.37(a) are similar. This can be verified if we compare the radial positions of the interference rings [Eqs. (3.47) and (3.85)]. A comparison of the radial position of the first 12 interference rings in the Michelson interferometer, when $a = 0.1$ and 0.5 mm, and the first 12 interference rings on the plate of parallel faces, when $2dn_l = 0.1$ and 0.5 mm, is shown in Fig. 3.39. In fact, when $a = 0.5$ mm, the radial position of the rings of the two interferograms is approximately the same. However, when decreasing to, e.g., $a = 0.1$ mm, the interference rings generated with the Michelson interferometer have a smaller radius compared with the rings generated with the plate.

Case 3. Interference of spherical waves II

Let us now consider the configuration shown in Fig. 3.36(a). Suppose that three rays (labeled 1, 2, and 3) diverge from the point source S, as shown in Fig. 3.40. In transmission, there will be multiple rays parallel to the incident rays. In Fig. 3.40, only three transmitted rays are drawn for each incident ray.

In each case, the first of the transmitted rays suffers a lateral deviation produced by the plate of thickness d. The backward extensions of the first

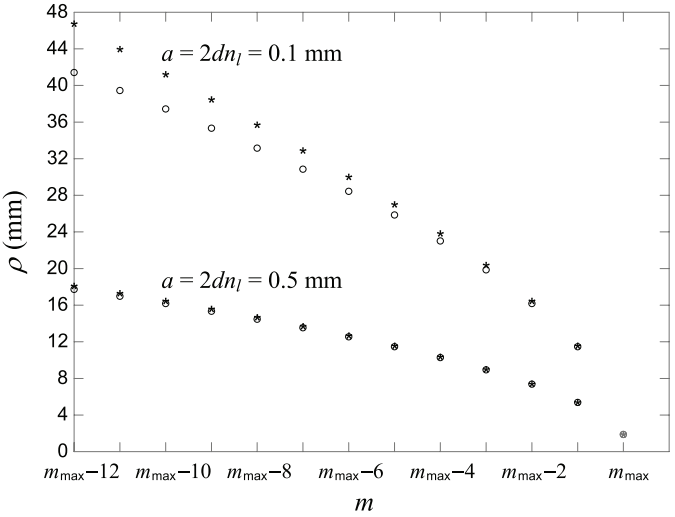


Figure 3.39 Radius (ρ) of the first 12 interference rings of spherical waves with the Michelson interferometer (o) when $a = 0.5$ and 0.1 mm and with a plate with parallel faces ($\star$) when $2dn_l = 0.1$ and 0.5 mm.

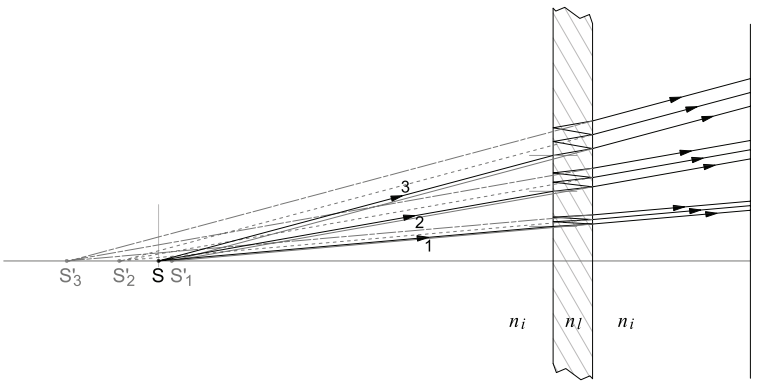


Figure 3.40 Refraction of three rays diverging from the source S together with some transmitted rays.

transmitted rays (solid gray lines) converge approximately at the point S'_1. Also in each case, the second of the transmitted rays undergoes a lateral deviation produced by an equivalent plate of thickness $3d$. The backward extensions of the second transmitted rays (straight gray lines with equal segments) converge approximately at the point S'_2. The third of the transmitted rays experiences, in each case, a lateral deviation produced by an equivalent plate of thickness $5d$. The backward projections of the third transmitted rays (long and short gray straight segments) converge approximately at the point S'_3. And so on it continues for the other transmitted rays. The extensions of the jth ($j = 1, 2, 3, \ldots$) transmitted rays do not converge

exactly at a point (S'_j) because the axial deviation experienced by the refracted rays depends on the angle of incidence, according to Eq. (3.65).

For the analysis that follows, let us assume that the axial prolongations of the jth transmitted rays converge at the point S'_j. This implies that the angle of incidence of the rays diverging from S and reaching the plate is small. With this in mind, the rays arriving at a point P on the screen appear to come from multiple virtual sources located at S'_1, S'_2, $S'_3, \ldots$, as shown in Fig. 3.41(a). Therefore, the interference at P is produced by the superposition of rays with different angles. Thus, the superimposed rays at P are rays emerging from the source S at different angles. So how is it possible that the refracted rays meet at point P? This is possible if the different rays experience different internal reflections in the plate, as shown in Fig. 3.41(b). For example, three rays arrive at point P_2 [Fig. 3.41(a)]: the ray that appears to come out of S'_1 arrives

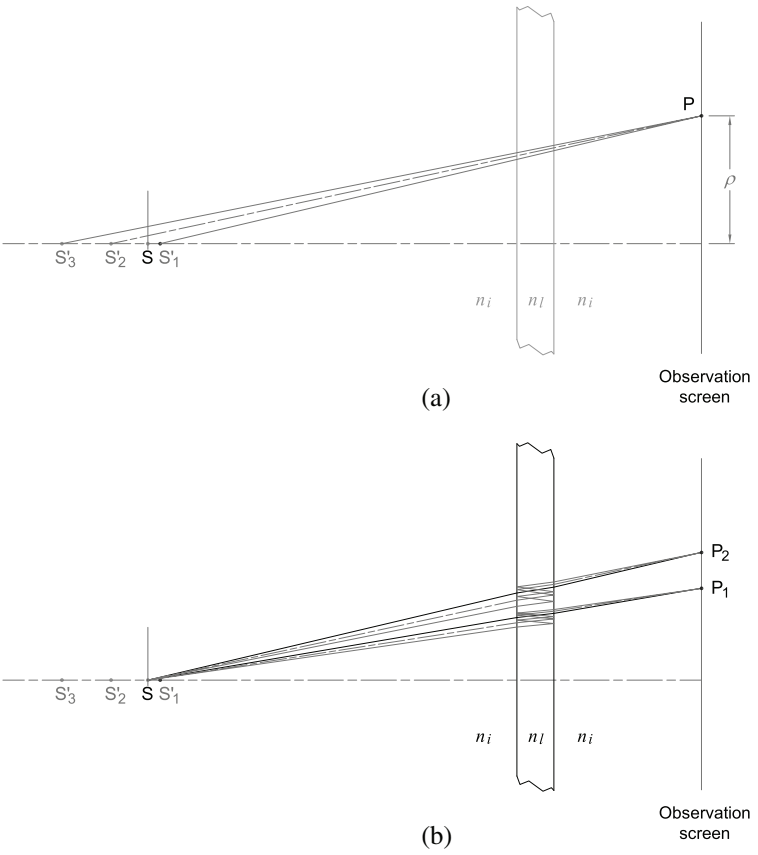


Figure 3.41 (a) Virtual sources. Rays arriving at point P appear to come from multiple virtual sources located at S'_1, S'_2, and S'_3. (b) Real rays. The interference at a point on the observation screen results from the superposition of rays with different angles and with different numbers of internal reflections in the plate.

after having been transmitted by both sides of the plate without reflecting internally (ray in black); the ray that appears to leave S'_2 arrives after being transmitted in the first face, then internally reflected from the second face, plus internally reflected from the first face, and transmitted from the second face (segmented gray ray); and the ray that appears to leave S'_3 arrives after being transmitted from the first face, then internally reflected from the second face twice, plus internally reflected from the first face twice, and transmitted from the second face (ray on solid gray). This will happen progressively for the rays that emerge from the other virtual sources.

The separation between two consecutive virtual sources can be calculated with the help of Fig. 3.40. Let us consider extensions of the corresponding transmitted rays with one of the rays diverging from S, as shown in Fig. 3.42. From the geometry of the figure, the separation between any pair of consecutive sources is constant. By defining this separation as a, $a = \overline{S'_1S'_2} = \overline{S'_2S'_3} = \ldots = \overline{S'_jS'_{j+1}}$. It is also true that $\overline{B_1B_2} = \overline{A_1A_2} = \overline{A_2A_3} = \ldots = \overline{A_jA_{j+1}}$. Therefore, for any pair of consecutive sources (j and $j+1$), $\tan\theta_i = \overline{B_1B_2}/a$. Because $\overline{B_1B_2} = 2d\tan\theta_t$,

$$a = \frac{2d\tan\theta_t}{\tan\theta_i}. \tag{3.87}$$

Equation (3.87) says that the separation between the virtual sources varies with the angle of incidence. However, in a practical situation, limiting the angles of incidence to small values (around 10°) so that $\cos\theta_i \approx 1$ and $\cos\theta_t \approx 1$ allows the change from $\tan\theta_t/\tan\theta_i$ to $\sin\theta_t/\sin\theta_i$. Using Snell's law,

$$a = 2d\frac{n_i}{n_l}, \tag{3.88}$$

which is independent of the angle. Thus, in this situation the interference of spherical waves in the configuration shown in Fig. 3.36(a) can be seen as the superposition of spherical waves that diverge from multiple virtual sources

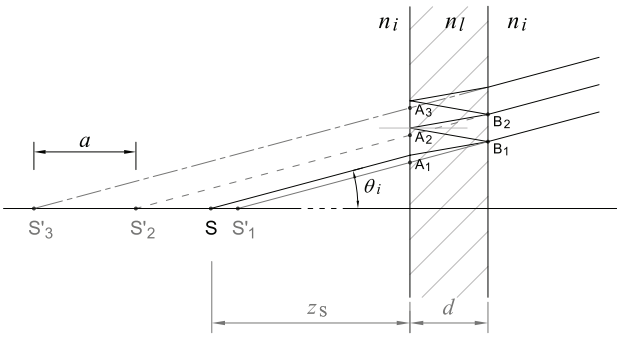


Figure 3.42 Geometry to calculate the distance between two consecutive virtual sources.

uniformly separated by a distance of $2dn_i/n_l$ along the optical axis. The distance from the virtual source S'_j to the plate (front face) turns out to be

$$z_j = z_S - \delta z_0 + (j-1)a, \tag{3.89}$$

where z_S is the distance between the point source S and the plate, and δz_0 is the axial deviation given by Eq. (3.65), with $n_t = n_l$ for $\theta_t = 0$.

Based on this, the optical field for a point P on the observation screen at radial distance $\rho = \sqrt{x^2 + y^2}$ (omitting the time phase term ωt and the initial phase term) will be

$$E_t = tt' \frac{E_0^\dagger}{s_1} e^{iks_1} + tr^2 t' \frac{E_0^\dagger}{s_2} e^{iks_2} + tr^4 t' \frac{E_0^\dagger}{s_3} e^{iks_3} + \cdots +, \tag{3.90}$$

or, more compactly,

$$E_t = tt' E_0^\dagger \sum_{j=1}^{J} r^{2(j-1)} \frac{e^{iks_j}}{s_j}, \tag{3.91}$$

where s_j is the distance between the virtual source S'_j and the point P, and J is the total number of reflections ($J \to \infty$). Assuming that the distance between the plate and the observation screen is $d + z_P$, then

$$s_j = \sqrt{R_j^2 + \rho^2}, \tag{3.92}$$

with

$$R_j = z_j + d + z_P. \tag{3.93}$$

As in cases 1 and 2, the reflection coefficient determines the weight with which each addend intervenes in the sum. For example, for the glass plate without reflective coating, by just taking the first two addends, the transmitted irradiance would be

$$I_t = (1-R)^2 \left\{ I_1 + I_2 R^2 + 2R\sqrt{I_1 I_2} \cos\left[\frac{2\pi}{\lambda} (s_2 - s_1) \right] \right\}, \tag{3.94}$$

where $I_1 = \epsilon_0 c (E_0^\dagger / s_1)^2 / 2$ and $I_2 = \epsilon_0 c (E_0^\dagger / s_2)^2 / 2$. This irradiance is analogous to Eq. (3.44), except for the contrast of the fringes, which is low for the plate case. By increasing the reflectance of the plate, there will be substantial contributions from more addends and the irradiance resulting from the superposition of J transmitted beams ($J \to \infty$) can no longer be written in a simple way, as is done in Eq. (3.82) in the case of the superposition of parallel beams.

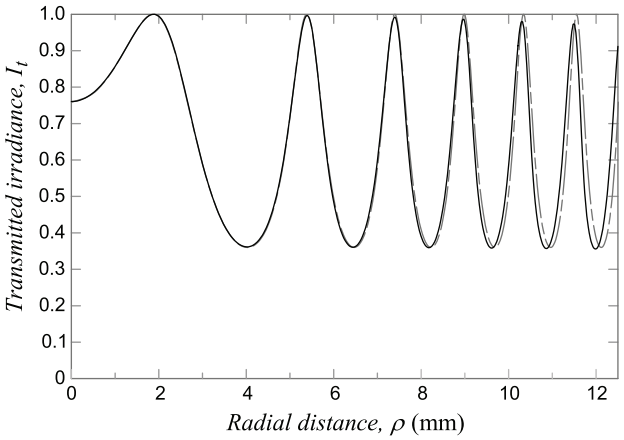


Figure 3.43 Profiles of the interferograms obtained with the interferometers shown in Figs. 3.36(a) (in black) and (b) (in gray dashed).

To see the difference between the interferogram obtained with the configuration shown in Fig. 3.36(a) and the interferogram obtained with the configuration shown in Fig. 3.36(b), let us compare the profile of the interferogram shown in Fig. 3.37(b), which is obtained when $r = 0.5$, and the profile of the interferogram expressed in Eq. (3.91), with the same value of r. For the calculation of the second profile, the following parameters were set: $d = 0.25n_l$ (in mm), $n_l = 1.5$, and $n_i = 1.0$. With this, the virtual sources are separated by $a = 0.5$ mm. The distances $z_S = 50 - \delta z_0$ and $z_P = 50 - d$, both in mm, were also fixed. In Fig. 3.43, the two profiles for the first six rings are shown as a function of the radial coordinate: in black, the one obtained with the interferometer shown in Fig. 3.36(a), and in gray dashed, the one obtained with the interferometer shown in Fig. 3.36(b). Two things stand out: the first, moving away from the center, the irradiance maxima separate; second, the rings in the black interferogram fade (going away from the center). The separation between the maxima corresponds to that shown in Fig. 3.39 for $a = 0.5$. Although the calculation of the black profile must be done with $J \to \infty$, using the first 12 addends of the sum of Eq. (3.91) is more than sufficient because the 12th addend turns out to be $\approx 0.0001\%$ of the second addend.*

Equation (3.94) is completely analogous to Eq. (3.44), which implies that the radii of the circles of maximum irradiance for the interferometer shown in Fig. 3.36(a), with $r = 0.2$, coincide with the radii of the interference rings generated by two waves in the Michelson interferometer. On the other hand, if the effect of increasing the reflectance of the plate translates into a thinning of the interference rings while maintaining the radius of the circles of maximum irradiance, then it can be anticipated that the circles of maximum irradiance in

*Another example: if the reflection coefficient is $r = 0.9$, the first 68 addends are required so that the last addend is approximately 0.0001% of the second addend.

the interferometer of Fig. 3.36(a) are also calculated with Eq. (3.47). In terms of the interference by reflection, a result similar to that shown in Fig. 3.38 will also be obtained because, of course, it will be the complement of the interferograms generated by transmission.

3.5.3 Two-wave interference

It has already been mentioned that in a plate with parallel flat faces, when the reflectance of the faces is low, in practice, only the first two reflected or transmitting waves interfere. In both cases, the analysis of the formation of interference fringes can be carried out by considering the two virtual sources that are generated by reflection or by transmission. This section deals with a situation of special interest: the interference of the first two spherical waves reflected by a plate of low reflectance ($r \sim 0.2$). This case will allow us to analyze the interference produced by extended polychromatic light sources.

According to the previous sections, there are two possible configurations in which to observe the interference, which are illustrated in Figs. 3.44(a) and (b). So there are two virtual sources (on the opposite side to the source S) generated by the reflection on each of the faces of the plate. In (a), two rays arrive at P at different angles from the virtual sources. For any point P on the source side the same will happen, so there will be interference fringes anywhere on the source side (nonlocalized fringes). In (b), two parallel rays are focused by the lens at P; therefore, there will only be interference fringes in the focal plane (localized fringes) that depend on the angle of the rays (fringes of equal inclination).

In this section, we consider fringe formation for the configuration shown in Fig. 3.44(a). With the help of Fig. 3.45, it is possible to see that, due to the symmetry around the optical axis (the line through S and the virtual sources), the interference fringes in a plane orthogonal to the optical axis containing P are circular. The separation between the virtual sources $a = \overline{S_1' S_2'}$ is given by

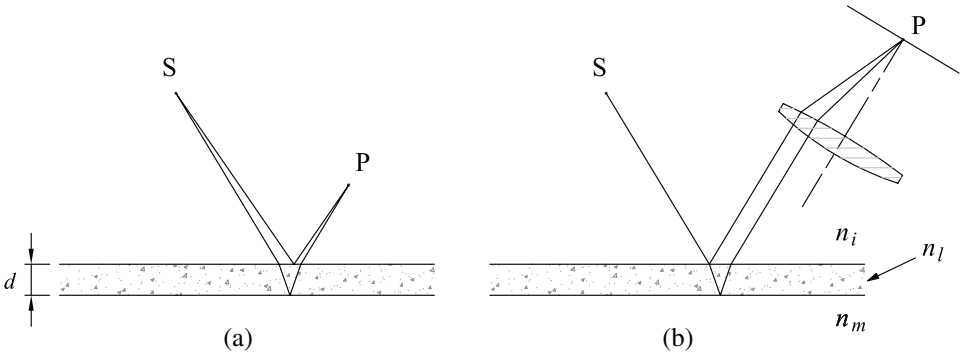


Figure 3.44 Interference by reflection: (a) nonlocalized fringes for any point P and (b) localized fringes in the focal plane of the positive lens.

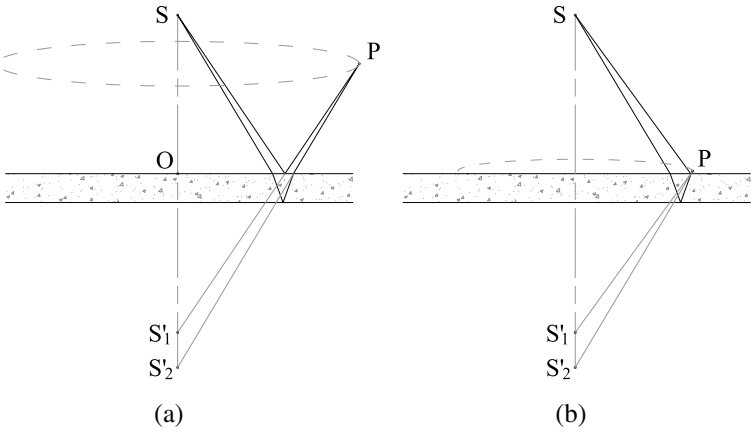
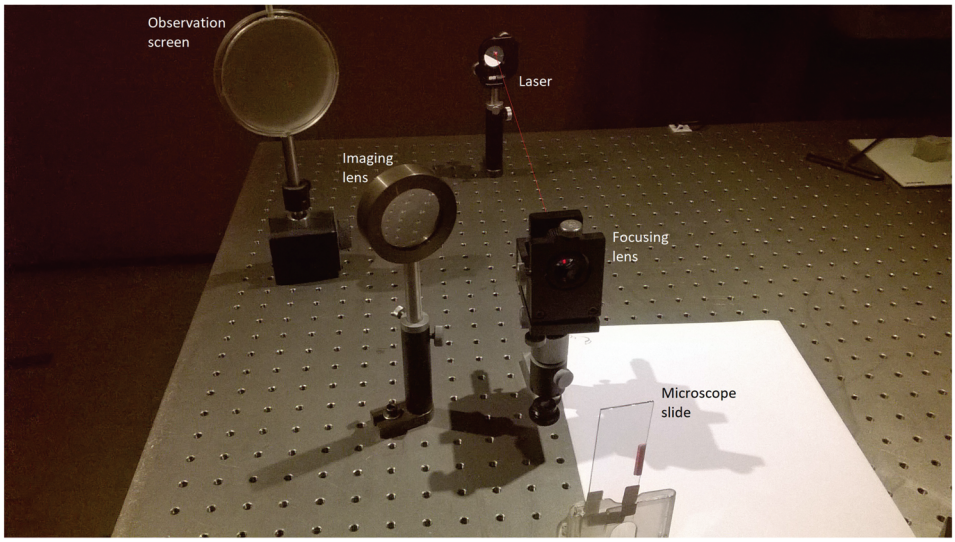


Figure 3.45 Fringe formation by reflection. (a) Circular fringe in an observation plane at a certain distance from the plate. (b) Circular fringe in an observation plane close to the first face of the plate.

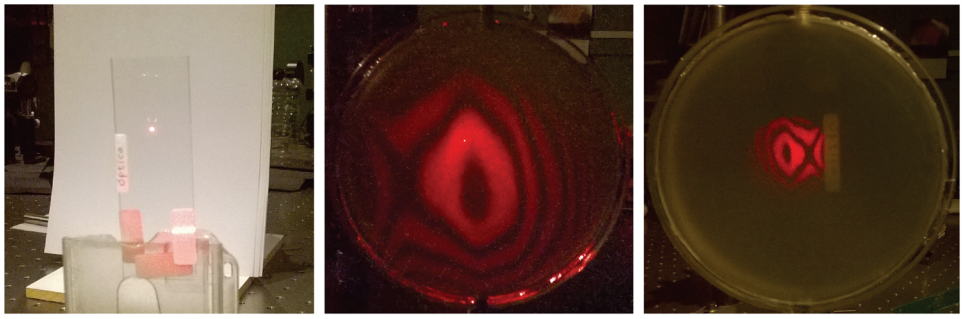
$a = 2d \tan\theta_t / \tan\theta_i$ [Eq. (3.87)], and the radius of the interference rings is given by Eq. (3.47), with $z_0 \geq (\overline{SO} + a/2)$.

The formation of an interference ring in an observation plane at a certain distance $z_0 > (\overline{SO} + a/2)$ is shown in Fig. 3.45(a). Changing the distance z_0 changes the scale of the rings (Fig. 3.15). In the limit, when the observation plane coincides with the first face of the plate [$z_0 \approx (\overline{SO} + a/2)$], the fringes will have the smallest possible size [Fig. 3.45(b)]. If the plate has irregularities in its optical thickness (refractive index variations and geometrical thickness variations), the interference fringes are distorted. In particular, if the observation plane coincides with the first face of the plate, then the interference fringes in that plane will allow the irregularities of the sheet to be measured with respect to the pattern of regular rings that would be obtained for an ideal plate with parallel flat faces. Furthermore, if the distance from the source S to the plate is such that the wavefronts are practically flat, then the interference pattern will be a topographic map of optical thickness (similar to contour lines on geographic maps). This allows direct measurement of the optical quality of the plate.

In addition to the above, an experiment to observe reflection interference from a glass plate ($n = 1.51$) is shown in Fig. 3.46. In (a), the complete setup is shown: a laser beam (He-Ne, 632.8 nm), a focusing lens (focal length, 8 mm), a microscope slide (1 mm thick), an imaging lens (150 mm focal distance and 60 mm in diameter), and an observation screen (frosted glass). The laser beam is focused with the positive lens to generate the point source. The diverging light (spherical wave) reaches the plate where the two relevant reflections for the interference take place. In (b), a detail of the plate is shown to which a label (with the word "óptica") has been placed that will serve to form the



(a)



(b) (c) (d)

Figure 3.46 (a) Experimental setup to view reflection interference in a plate. (b) Detail of the plate with a label (the word "óptica"). (c) Interferogram on the observation screen about 60 cm from the plate, without the imaging lens. (d) Image (formed with the imaging lens) of the interferogram on the first side of the plate.

image of the plate on the observation screen. In (c), the interference pattern on the display screen is shown when the imaging lens has been removed; i.e., when light travels freely from the plate to the display screen. In (d), the image of the interference pattern is observed just on the first side of the plate. For this, the imaging lens is used. Sure enough, the label image next to the interferogram can be seen. Because the fringes are not localized, in (d) we have a scaled version of (c). This interferogram shows the defects of the plate that can be caused by variations in the refractive index or by irregularities in the flatness of the faces.

It is worth noting that the aperture of the imaging lens must be large enough to collect all (or nearly all) of the reflected beams on the slide if the full

interferogram is to be observed. If instead of this lens one of our eyes is placed to look at the plate, the interferogram will not be seen because the beam of rays that enters is very small, limited by our pupil.

3.6 Interference from *N* Point Sources

In the previous sections, the phenomenon of interference by two point sources through the Michelson interferometer and by J point sources with $J \to \infty$ through a plate with parallel faces has been considered. Increasing the number of sources, with a number that tends to infinity, and making the contribution of each wave to the interference relevant lead to a thinning of the interference rings. If the reflectance approaches 1, the fringe profile tends to a distribution of very narrow bands (like Dirac deltas).

This section will study the intermediate case: the interference of coherent waves with each other generated by N point sources ($2 < N < \infty$). In particular, there will be two situations to deal with: a set of sources evenly spaced in the axial direction, as shown in Fig. 3.47(a), and an array of evenly spaced sources in the transverse direction, as shown in Fig. 3.47(b). In both cases, let us assume that the point sources have the same amplitude and initial phase, so the field at a point P on the observation screen becomes (omitting the time term ωt)

$$E = E_0^{\dagger} \sum_{j=1}^{N} \frac{e^{iks_j}}{s_j}. \tag{3.95}$$

In a way, interference with the first source array has already been discussed in the previous section. The distance s_j is determined by Eq. (3.92),

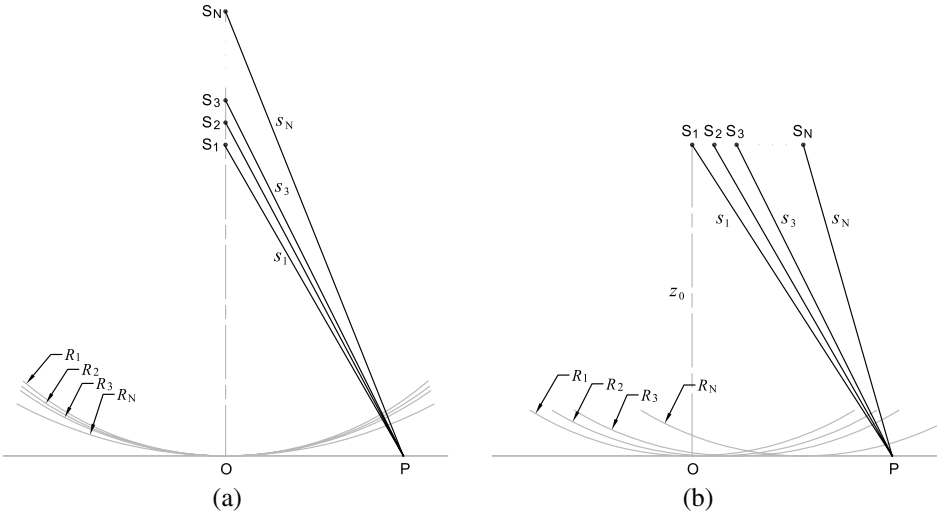


Figure 3.47 An array of N point sources: (a) axial and (b) transversal.

$$s_j = \sqrt{R_j^2 + \rho^2},$$

where R_j is the radius of curvature of the wavefront with the center at S_j and the vertex at point O on the observation screen. The radial distance $\rho = (x^2 + y^2)^{1/2}$ measures the separation between O and P. If a is the separation between consecutive sources, then

$$R_j = R_1 + (j - 1)a. \tag{3.96}$$

Normalized interference patterns (and their profiles) are shown in Fig. 3.48 when $a = 1000\lambda$, $\lambda = 632.8$ nm, $R_1 = 100$ mm, and $N = 2$, 3, and 5. As expected, increasing the number of sources decreases the width of the fringes. But now there is a remarkable fact: between the maxima of the brightest fringes, fringes of much lower intensity appear. The first fringes are called *primary maxima*, and the second fringes are called *secondary maxima*. In fact, there are $(N - 2)$ secondary maxima between two consecutive principal maxima. As the number of sources increases, the intensity of the secondary maxima decreases, and if $N \to \infty$, the secondary maxima disappear and there will be very narrow (diffraction-limited) fringes, as shown in Fig. 3.37(c).

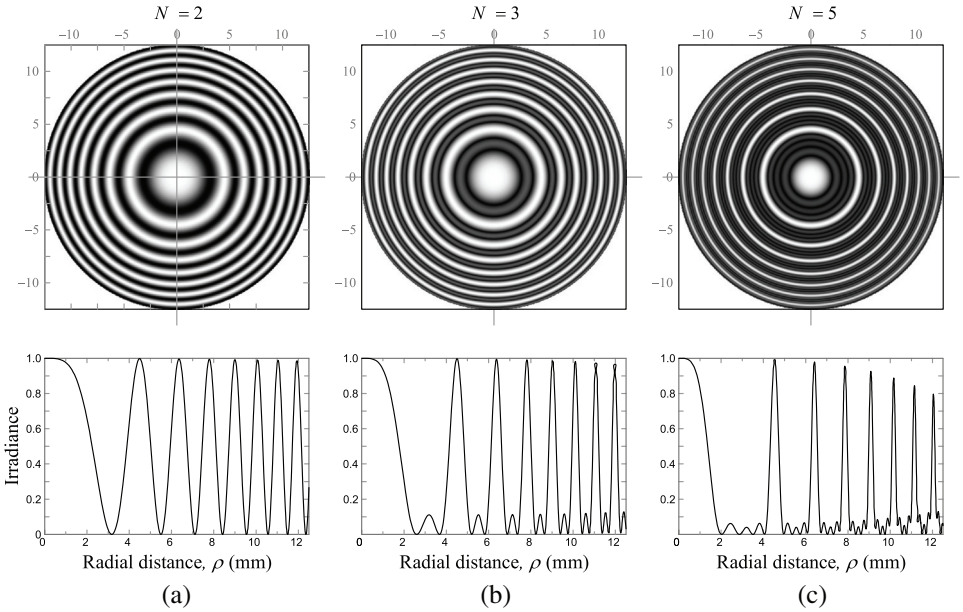


Figure 3.48 Interference of spherical waves generated by 2-, 3-, and 5-point sources located axially and uniformly separated by a distance $a = 1000\lambda$. The distance from the first source to the observation screen is $R_1 = 100$ mm.

For the second array of point sources [Fig. 3.47(b)], the radii of curvature of the wavefronts with the vertices in the observation screen are equal to z_0. Assuming that the linear array of sources is set along the x direction, the distance from the source S_j to the point P will be

$$s_j = \sqrt{z_0^2 + (x - (j-1)a)^2 + y^2}. \tag{3.97}$$

Normalized interference patterns (and their profiles) for 2- and 4-point source arrays are shown in Fig. 3.49. In (a) and (b), the sources are $a = 8\lambda$ apart and at a distance $z_0 = 50$ mm from the observation screen. Again, going from two to four sources, the major fringes become thinner and two minor fringes emerge, but the position of the principal fringes is maintained. In (c), the quantities a and z_0 for the set of four sources have been multiplied by 10, but the observation region is the same as in (a) and (b). With this change, the interference pattern changes the shape of the fringes from hyperbolic to straight lines. Now the principal fringes are the same distance apart, and of course, the secondary maxima are still visible.

3.6.1 Plane wave approximation

In the array shown in Fig. 3.47(b), when the observation region is much smaller than z_0, the waves from the N sources when they reach the observation

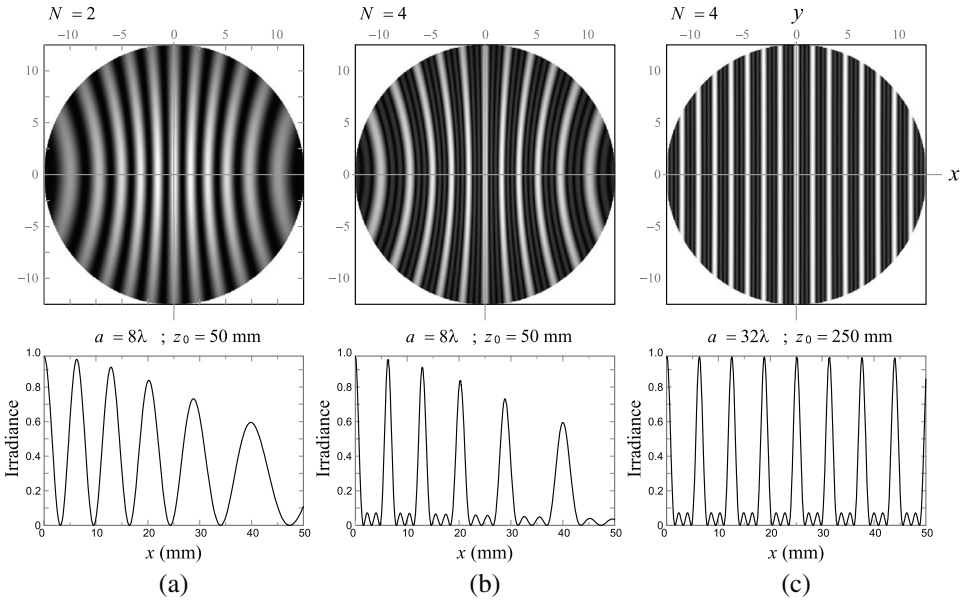


Figure 3.49 Interference of spherical waves generated by 2- and 4-point source arrays located laterally. (a and b) The distance of the sources from the observation screen is $z_0 = 50$ mm, and the sources are uniformly separated by a distance $a = 8\lambda$. (c) The distance of the sources from the observation screen is $z_0 = 500$ mm and the sources are uniformly separated by a distance $a = 80\lambda$, but the same observation region of (a) and (b) is maintained.

screen can be approximated by plane waves. This explains why in Fig. 3.49(c) the fringes resemble a pattern of equally spaced straight fringes.

If, in addition to the condition mentioned above, the extension of the line array of sources satisfies $(N-1)a \ll z_0$, then the amplitudes of the N waves at point P will be approximately equal and s_j in the denominator of Eq. (3.95) can be changed to z_0. By limiting the analysis to the x direction,

$$s_j = z_0\sqrt{1+\frac{x^2-2(j-1)xa+(j-1)^2a^2}{z_0^2}}, \tag{3.98}$$

which in first approximation is

$$s_j = z_0 + \frac{x^2}{2z_0} - \frac{(j-1)xa}{z_0}. \tag{3.99}$$

With this, Eq. (3.95) can be written as

$$E = \frac{E_0^\dagger}{z_0} e^{ikz_0} e^{ikx^2/2z_0} \sum_{j=1}^{N} e^{-ikx(j-1)a/z_0}. \tag{3.100}$$

Defining $I_0 = (\epsilon_0 c/2)E_0^{\dagger 2}/z_0^2$, the irradiance $I = (\epsilon_0 c/2)(E^* E)$ is

$$I = I_0 \left| \sum_{j=1}^{N} e^{-ikx(j-1)a/z_0} \right|^2. \tag{3.101}$$

The sum turns out to be

$$\sum_{j=1}^{N} e^{-ikx(j-1)a/z_0} = \frac{1-e^{-ikxNa/z_0}}{1-e^{-ikxa/z_0}}, \tag{3.102}$$

which can be rewritten as

$$e^{-k(N-1)a/2z_0} \frac{i2\sin(kxNa/2z_0)}{i2\sin(kxa/2z_0)}. \tag{3.103}$$

In the end, the irradiance is

$$I = I_0 \left[\frac{\sin(kxNa/2z_0)}{\sin(kxa/2z_0)}\right]^2. \tag{3.104}$$

Principal maxima

The zeros in the denominator of Eq. (3.104) determine the position of the principal maxima. These are obtained by

$$\frac{ka}{2z_0}x = m\pi, \tag{3.105}$$

where m is an integer. Then, the irradiance at these points would be

$$I = I_0\left[\frac{\sin(Nm\pi)}{\sin(m\pi)}\right]^2. \tag{3.106}$$

Because both the denominator and the numerator become zero for m integer, from L'Hôpital rule,

$$\lim_{m\to integer}\left[\frac{\sin(Nm\pi)}{\sin(m\pi)}\right] = \frac{N\cos(Nm\pi)}{\cos(m\pi)} = \pm N. \tag{3.107}$$

Therefore,

$$I = N^2 I_0 \tag{3.108}$$

is the value of the principal maxima found in

$$x_m = m\frac{\lambda z_0}{a}. \tag{3.109}$$

That is, the separation between consecutive maxima turns out to be

$$\Delta x = \frac{\lambda z_0}{a}. \tag{3.110}$$

There are other zeros in the numerator, between the zeros in the denominator, which are given by

$$N\frac{ka}{2z_0}x = m'\pi, \tag{3.111}$$

with m' (integer) < N, located at

$$x_{m'} = \frac{m'}{N}\frac{\lambda z_0}{a}. \tag{3.112}$$

Thus, there will be $(N-1)$ minima ($I=0$) between two consecutive principal maxima. Consequently, there should be $(N-2)$ secondary maxima interspersed with the minima.

The profile of the interference patterns for $N=2$-, 3-, and 5-point sources is shown in Fig. 3.50 when $a=80\lambda$, $z_0=1000$ mm, and $\lambda=632.8$ nm. In all cases, because it is assumed that the amplitude of the sources is the same, the irradiance produced by a source at a point on the observation screen is I_0.

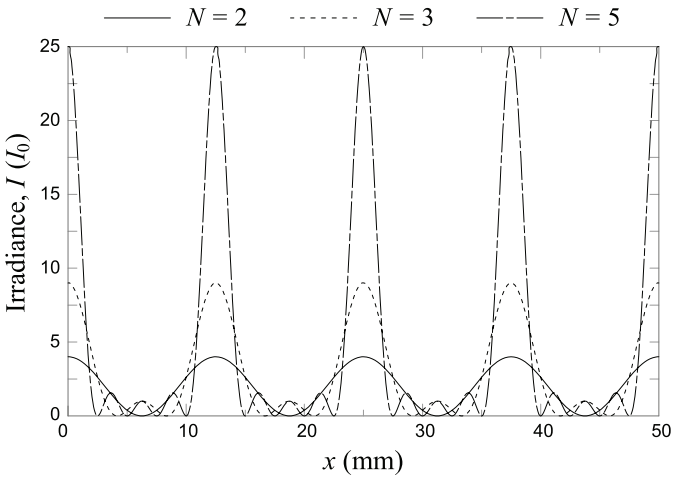


Figure 3.50 Interference pattern profile of $N = 2$, 3, and 5 equally spaced point sources of equal amplitude.

The results shown in Fig. 3.50 are the basis for diffraction gratings, a topic that is covered in Chapter 4.

3.7 Interference with Extended Light Sources

Until now, our focus in the previous sections has been limited to the interference of waves emitted by a monochromatic point source. This situation is very common in the laboratory, where we have laser sources and optical elements to form a point source. Outside the laboratory, white light (sunlight) interference can be seen in soap bubbles or oil films in water. In both cases, there should be a thin film with a thickness of about half the coherence length of light. Direct or diffuse sunlight illuminating the thin film arrives from different directions and can be modeled as an extended incoherent source, i.e., a set of spatially distributed incoherent point sources. The color interference pattern recorded by a camera focusing on an oil film on wet asphalt on a rainy day is shown in Fig. 3.51.

To explain the formation of the fringes shown in Fig. 3.51, consider Fig. 3.52, which shows a plate with parallel plane faces, of index n_l, resting on a block of index n_m. An extended source, represented by the curve where S and S′ are, illuminates the plate. Let us analyze this figure in parts. First, suppose that there is only the point source S, which generates a pattern of nonlocalized fringes. Pay attention to the interference pattern near the front face of the plate, as in Fig. 3.45(b). Changing S for another point source S′, analogously to the case of the source S, there would be another pattern of nonlocalized fringes. If the two sources are left active, assuming that the two sources are incoherent with each other, there will be a superposition in P of the

Figure 3.51 Interference generated by an oil stain on wet asphalt (water) with white light (sun).

interference patterns generated individually by each source. By adding other incoherent sources between each other, until completing the extended source, the effect of the superposition of the different interference patterns will be an illuminated region without modulation of the irradiance by interference; i.e., with the extended source, the nonlocalized interference fringes that might be observed with a point source disappear.

Returning to the situation with only the point source S, the nonlocalized interference fringes near the front face of the plate can be observed by imaging the plate through a lens L, as shown in Fig. 3.52. If the aperture diaphragm is small (as in the eye or in a photographic camera), the region illuminated by S that can be observed is limited to point P and its neighborhood. This neighborhood is defined by the rays leaving S, striking the plate, and reflecting through the aperture diaphragm. The angular size of the neighborhood is defined by the angle subtended by the virtual image of the aperture diaphragm (the image generated by the first face of the plate) with respect to the source S.

Let us assume that the neighborhood of point P is very small. The interference at P can be explained as the superposition of the chief ray (axis-like line) reflected from the first face of the plate and the neighboring ray reflected from the second face, with both rays leaving S. Rays leaving S at other angles do not fall within the neighborhood of P. To see the interference in other regions of the plate, rays from other sources are needed, e.g., with the source S′, interference in P′ can be observed. By considering all the point sources that make up the extended source, the full interferogram would be

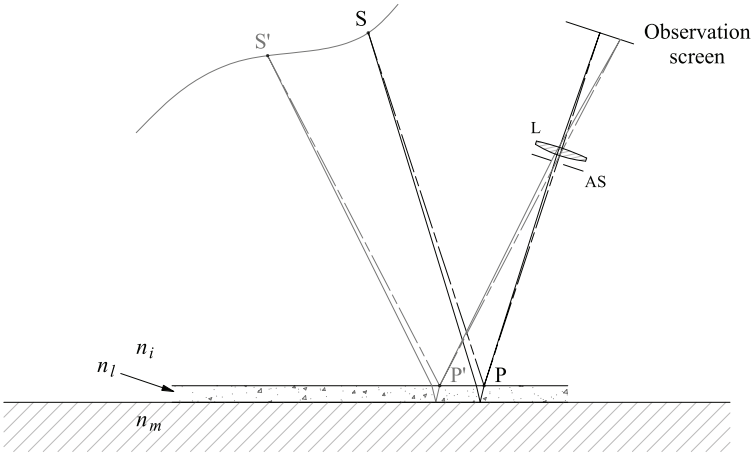
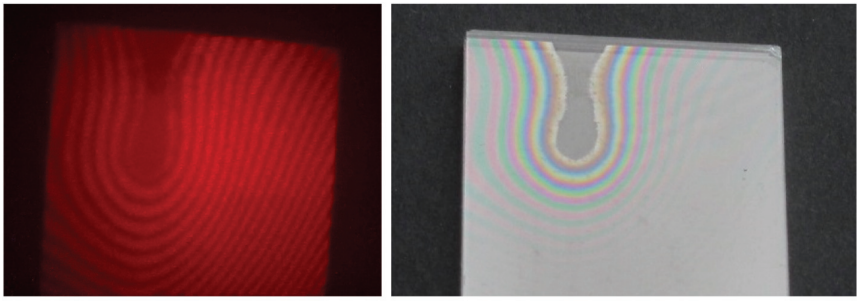


Figure 3.52 Interference fringes with an extended source.

observed, similar to what happens with a single point source (and a large aperture lens). Now the size of the observation screen (field stop) defines the region in which the interferogram would be seen.

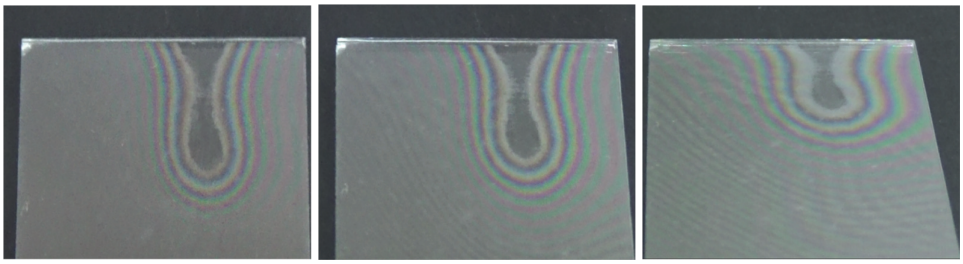
Finally, if the point sources are polychromatic, each color satisfies the interference conditions, and because the position of the fringes depends on the wavelength, the final result is color fringes. As the wavelength increases, the size of the interference fringes decreases [Eq. (3.56)].

Reflection interference fringes generated by the film of air between two microscope slides are shown in Fig. 3.53. In (a), fringes on an observation screen separated from the slides by about 25 cm when illuminated with a point source ($\lambda_0 = 632.8$ nm) are shown. In (b), fringes on the slides recorded with a photographic camera when illuminated with white light are shown. To generate the air wedge, a very small drop of water was placed near the edge of



(a) (b)

Figure 3.53 Reflection interference patterns generated by the film of air between two microscope slides put in contact. (a) Fringes on an observation screen (25 cm from the slides) when illuminating with a point source (wavelength, 632.8 nm). (b) Fringes recorded by a camera when the microscope slides are illuminated with white (solar) light.



(a) $\sim 10°$ (b) $\sim 45°$ (c) $\sim 70°$

Figure 3.54 Interference fringes with white light (sun) when viewed from different angles with respect to the normal of the slides.

one of the slides, and then the other slide was placed on top of the first and mechanical pressure was exerted on the slides, thus spreading the drop of water. Beyond the boundary of the water stain there is air, and it is there that the interference fringes are observed. The two patterns in (a) and (b) are observed in a direction close to the normal of the sliders. When the color fringes are viewed from another angle, the position of the fringes changes,* as shown in Fig. 3.54. In (a), the fringes are observed at an approximate angle of 10° with respect to the normal of the slides, which can be considered close to the normal. In this case, the fringes are called *fringes of equal thickness* and can be used to make a topographic map of the air film. In (b) and (c), the photographic camera was tilted approximately 45° and 70°, respectively. This change in fringe size depends not only on the change in the optical path in the angled film, but also on the thickness of the layer above it. In the case of the oil film shown in Fig. 3.51, there is no other medium (other than air), so changing the viewing angle changes the size of the fringes less noticeably.

Thus, with light from an extended polychromatic source, localized color fringes can be seen on a film (of air, soap, oil, etc.). Viewed in the direction normal to the film, we essentially have fringes of equal thickness.

3.7.1 Artificial extended sources

Direct and diffuse white light (sun) are an example of an extended source, in which the point sources are practically at infinity. One way to obtain an extended artificial source is through the use of a polished glass (diffusing plate), i.e., a glass plate to which one or both smooth surfaces have been modified through a sanding process (using some abrasive material, such as silicon oxide). Therefore, the plate is seen as a translucent surface (like the viewing screen used in Fig. 3.46 to see interferograms). By illuminating the polished glass, the microroughnesses of the sandblasted face act as secondary light sources that

*The phase difference changes in a similar way to that expressed by Eq. (3.79) for parallel rays. As the angle of incidence increases, the phase difference decreases, so the fringes increase in size.

emit roughly spherical waves. Illumination can be done with an extended source (white light, incandescent lamp, discharge lamp, etc.) or with a spherical wave generated by a laser. The first case is usually used to homogenize the lighting. The second case is used to get an extended monochromatic source, but now the different secondary point sources are coherent with each other. The superposition of the fields generated by the multitude of these secondary coherent sources generates an interference pattern with a grainy (random) structure known as a *speckle* pattern. The average size of the speckles depends inversely on the size of the illuminated region on the frosted glass. In general, to obtain an extended monochromatic source, the illumination region has to be of a size such that the average size of the speckles is very small, and what is observed is a homogeneous illumination (of very small grains).

Michelson interferometer with an extended source.
An example of artificial extended source interference common in teaching laboratories employs a Michelson interferometer and a diffuser plate illuminated with a discharge lamp (usually sodium or mercury vapor lamps). In these lamps, most of the emitted radiation occurs in a few spectral lines; e.g., for the sodium lamp, most of the radiation is around the 589.0 and 589.6 nm lines (hence its intense yellow color).

A schematic of the Michelson interferometer illuminated by the extended source Σ (lamp-illuminated diffusing screen) is shown in Fig. 3.55(a). The source Σ is a collection of mutually incoherent point sources with random initial phases. Mirrors M_1 and M_2 generate the virtual images Σ'_1 and Σ'_2 of

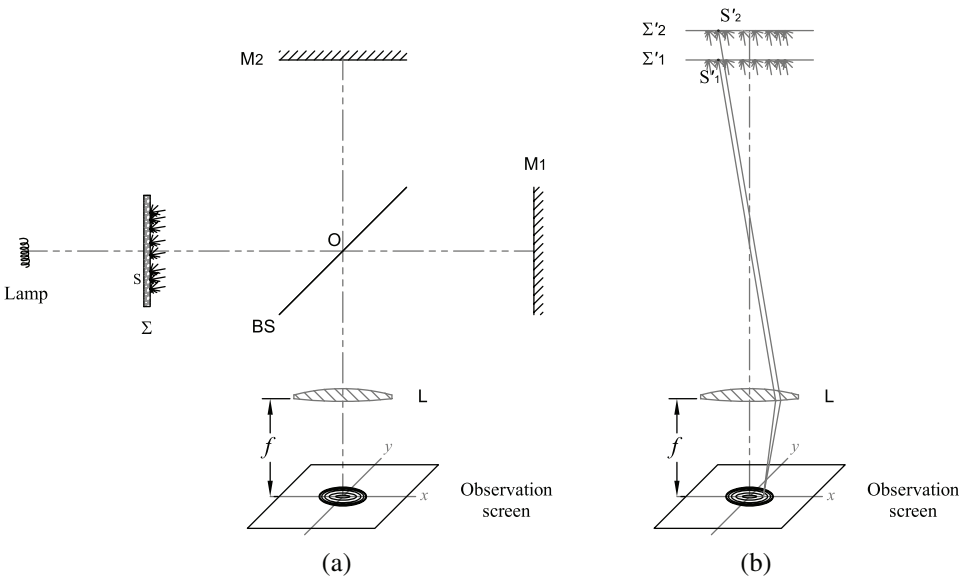


Figure 3.55 (a) A Michelson interferometer illuminated with an extended source Σ. (b) Illustration of the virtual images Σ'_1 and Σ'_2 of the extended source Σ generated by the mirrors M_1 and M_2, seen from the lens L.

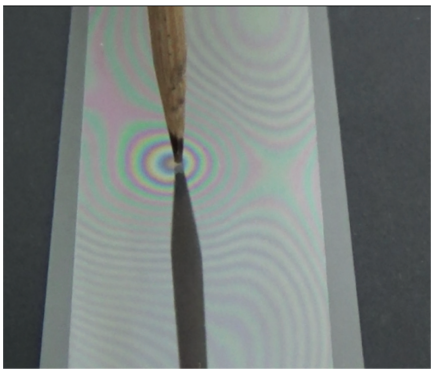


Figure 3.56 An interference pattern between two microscope slides when pressed with a pencil and using white light (sun).

the source Σ, which are seen one after the other from the observation screen, as shown in Fig. 3.55(b). Assuming that the coherence length of the lamp is greater than the separation between Σ_1' and Σ_2', the waves emitted by the virtual point sources S_1' and S_2' (images of the point source S in Σ) interfere. This occurs for all virtual point sources. But there would not be interference between waves emitted by the different point sources in Σ. Consequently, without the lens L, no interference fringes will be seen on the observation screen. With the help of lens L, by placing the observation plane in the focal plane of the lens, interference fringes of equal inclination can be observed. These fringes are the result of the superposition in intensity of the patterns generated by each point source in Σ. Equations (3.85) and (3.86), with $n_l = n_i = 1$, describe these fringes.

The interference between two microscope slides, when pressed with a pencil tip, is shown in Fig. 3.56. Based on what is covered in this section, the reader is invited to make a qualitative description of the formation process of the color fringes and the thickness of the air film between the slides.

3.8 Young Interferometer I

In the eighteenth century, the corpuscular theory of light developed by Newton prevailed. Although a wavelike behavior of light was already observed, the difficulty of observing diffraction (as it was already known in sound or in surface waves of water) motivated Newton to develop a corpuscular theory of light, which was accepted thanks to Newton's scientific reputation. Around the same time, other scientists such as Christiaan Huygens (1629–1695) and Robert Hooke (1635–1703) advocated wave theories of light. A little earlier, Francesco Grimaldi (1618–1663) discovered the diffraction of light through small openings, which suggested a wavelike behavior of light. Despite this background, there was no strong evidence that Newton's

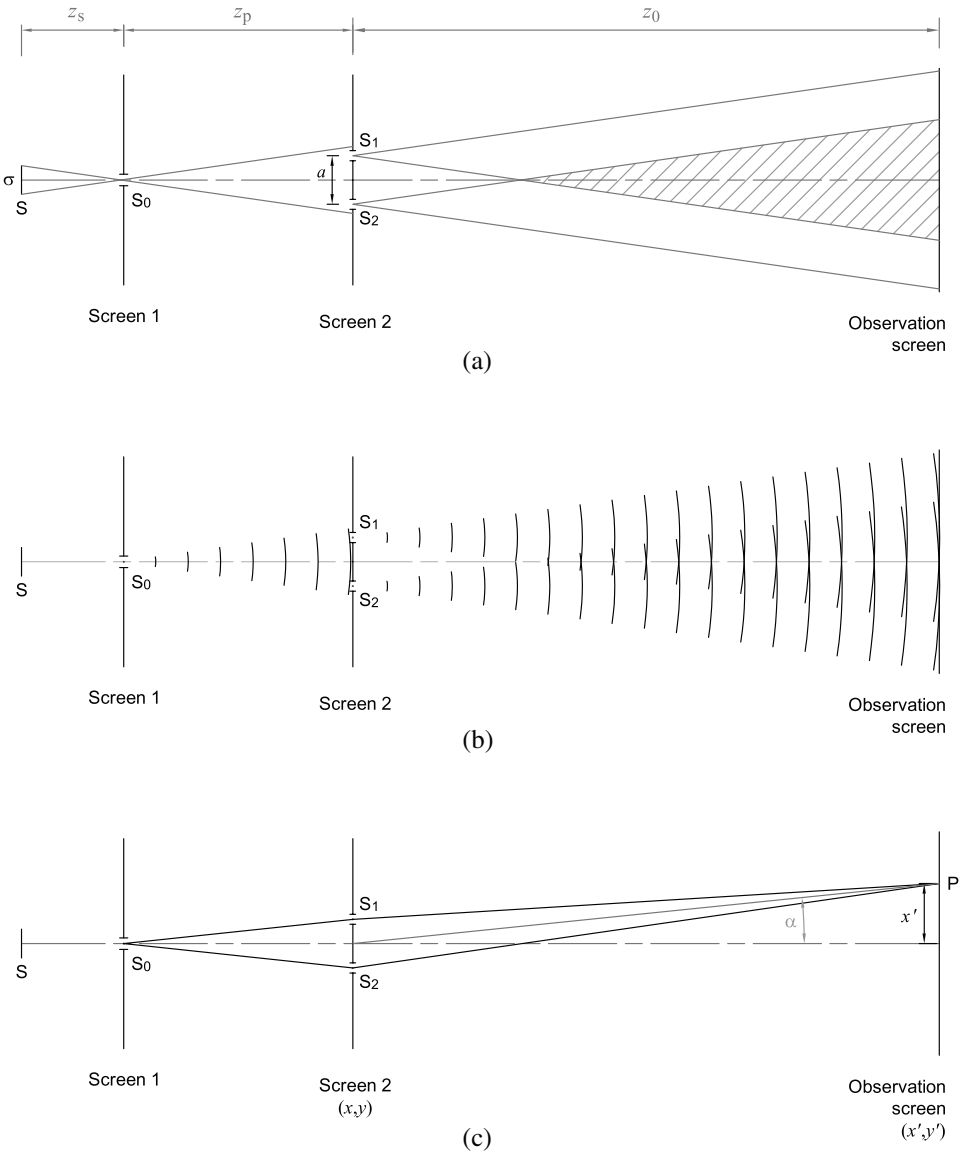


Figure 3.57 Young's experiment. (a) Geometry of the experiment. In screen 1, there is a small hole to let through some of the light emitted by the extended source S of size σ. On screen 2, there are two other holes, S_1 and S_2, also small, which in turn allow part of the light that diverges from S_0 to pass through. Light diverging from holes S_1 and S_2 is superimposed on the display screen. (b) Simplified scheme with point sources at S_0, S_1, and S_2. (c) Geometry for the calculation of the optical path difference.

corpuscular theory was wrong. In 1803, Thomas Young (1773–1829) performed an experiment that seriously challenged the corpuscular theory [4].

The modern version of Young's interferometer is illustrated in Fig. 3.57. The general scheme is shown in Fig. 3.57(a). The illumination source S is an

extended source of size σ. At distance z_s is screen 1, which has a small hole (S_0). Then, at a distance z_p from screen 1 there is screen 2, which contains two other identical holes (S_1 and S_2) separated from each other by the distance a. Finally, at distance z_0 from screen 2 is the observation screen. The size of S_0 determines the spatial correlation of the fields in S_1 and S_2. If S_0 is a point source, there will be the maximum spatial correlation between S_1 and S_2. On the other hand, the sizes of S_1 and S_2 determine the diffraction modulation of the interference pattern on the observation screen. The analysis of the interference pattern, taking into account the geometry of the holes and the size of the light source, is covered in Chapter 4.

In what follows, let us consider the ideal situation where the holes are so small that it can be assumed that there is one primary point source at S_0 and two secondary point sources at S_1 and S_2, as shown in Fig. 3.57(b). In Young's experiment, it is typical to have symmetry about the optical axis, so that S_1 and S_2 are the same distance from the optical axis and therefore $\overline{S_0S_1} = \overline{S_0S_2}$. Thus, on the observation screen, there will be a superposition of the spherical wavefronts that diverge from S_1 and S_2. This situation is discussed in Section 3.3. Some interference patterns are shown in Fig. 3.17 by changing the distance between the sources S_1 and S_2, and the distance between the sources and the observation screen.

Assuming that the source is monochromatic with wavelength λ and greater coherence length than the optical path difference at P, i.e., $l_c > |(\overline{S_0S_2} + \overline{S_2P}) - (\overline{S_0S_1} + \overline{S_1P})| = |\overline{S_2P} - \overline{S_1P}|$ [Fig. 3.57(c)], the irradiance maxima on the screen along the x' direction are given by

$$x'_m = \frac{m\lambda}{2}\sqrt{\frac{4z_0^2}{a^2 - m^2\lambda^2} + 1}, \tag{3.113}$$

according to Eq. (3.48), changing $y_0 \to z_0$ and $z_m \to x'_m$. The parameter m labels the interference fringes. Thus, $m = 0$ corresponds to the central fringe, $m = 1, 2, 3, \ldots$ corresponds to the fringes on the right (left) side of the central fringe according to their order, and $m = -1, -2, -3, \ldots$ corresponds to the fringes on the left (right) side of the central fringe according to their order. In Young's experiment, $z_0 \gg a \gg \lambda$ holds. If the region in which the interferogram is observed has an extent much smaller than z_0, the observed fringes are evenly spaced parallel fringes (as in the right image in Fig. 3.17 when the separation between the sources and the observation screen is 200 mm and $a = 32\lambda$). In this case, the position of the mth fringe can be obtained by making the corresponding approximations in Eq. (3.113), i.e.,

$$x'_m = m\frac{z_0\lambda}{a}, \tag{3.114}$$

and the separation between two consecutive fringes would be

$$\Delta x' = \frac{z_0 \lambda}{a}. \tag{3.115}$$

These results [Eqs. (3.114) and (3.115)] were also obtained in Section 3.6 [Eqs. (3.109) and (3.110)], in which spherical wavefronts were approximated by planar wavefronts, an approximation also found in Young's experiment. With this in mind, the irradiance on the observation screen of Young's experiment is given by Eq. (3.104) with $N = 2$, i.e.,

$$I = I_0 \left[\frac{\sin(kx'a/z_0)}{\sin(kx'a/2z_0)} \right]^2, \tag{3.116}$$

with $I_0 = (\epsilon_0 c/2) E_0^{\dagger 2}/z_0^2$, where $E_0^{\dagger}$ is the field amplitude (per unit length) of sources S_1 and S_2. Using the trigonometric identity $\sin(2\alpha) = 2 \sin \alpha \cos \alpha$ in the numerator of Eq. (3.116) leads to

$$I = 4I_0 \left[\cos\left(\frac{\pi a}{z_0 \lambda} x' \right) \right]^2. \tag{3.117}$$

Another approximate way to determine the irradiance is obtained directly from Fig. 3.57(c). Taking into account the condition $z_0 \gg a$, the field at P due to S_1 is $E_1(P) = (E_0^{\dagger}/z_0) e^{iks_1}$ and the field at P due to S_2 is $E_2(P) = (E_0^{\dagger}/z_0) e^{iks_2}$. Therefore, the irradiance at P, given by $I = (\epsilon_0 c/2)|E_1(P) + E_2(P)|^2$, would be

$$I = 2I_0[1 + \cos(k(s_2 - s_1))]. \tag{3.118}$$

Approximating $s_2 - s_1 \approx a\alpha \approx ax'/z_0$ leads to

$$I = 2I_0 \left[1 + \cos\left(\frac{2\pi a}{\lambda z_0} x' \right) \right], \tag{3.119}$$

which is equivalent to Eq. (3.117).

3.8.1 Division of wavefront and division of amplitude

In Young's interferometer there is a remarkable fact as to how to generate the two secondary sources S_1 and S_2 compared with the Michelson interferometer. In the Michelson interferometer (as in the plate with parallel faces), the secondary sources are virtual images of a primary source obtained by the beamsplitter and the mirrors M_1 and M_2 (Fig. 3.18). The beamsplitter divides the amplitude of the incident wave (into a reflected wave and a transmitted wave). Interferometers based on this principle are also called *amplitude-splitting interferometers*. In contrast, in Young's interferometer, the secondary sources S_1 and S_2 are obtained by isolating regions of the wavefront emitted

by the primary source S_0. Secondary sources are then said to be obtained by dividing the wavefront. Interferometers based on this principle are also called *wavefront-splitting interferometers*.

In amplitude-splitting interferometers, the secondary sources are copies of the primary source, so the interference between the secondary waves depends on the temporal correlation of the superimposed fields. If the delay time between the waves is greater than the coherence time (Fig. 3.6), no interference will be observed. In the case of wavefront-splitting interferometers, the interference also depends on the spatial similarity of the optical fields in S_1 and S_2. The spatial correlation of the fields in S_1 and S_2 measures the *spatial coherence* of the waves. In particular, in Young's interferometer, the size of the observation region of the interferogram is usually such that for any point P within the region, the coherence length is greater than the optical path difference, so the coherence of interference waves is basically determined by the spatial correlation of the fields in S_1 and S_2. Thus, coherence has two aspects, one temporal and one spatial, which can be measured with the help of Michelson's interferometer and Young's interferometer, respectively.

3.9 Other Interferometers

An interferometer can be said to be an optical system that splits a primary wave in amplitude or wavefront into two or more secondary waves, and then, after the secondary waves have traveled different optical paths, they are superimposed to observe a pattern of interference. The geometry of the fringes accounts for some parameters of the interferometric system, e.g., quality of the optical surfaces, homogeneity of the optical glass components (variation of the refractive index), curvatures of the surfaces, etc.

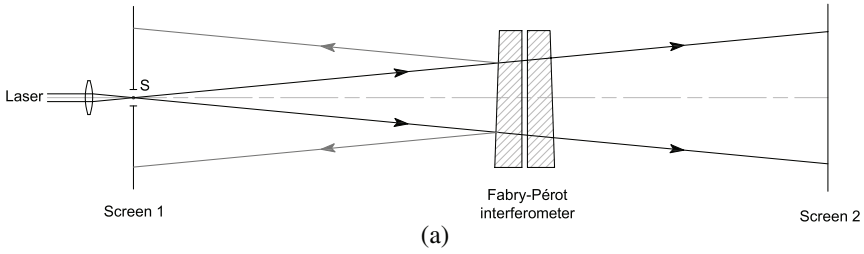
In this section, some types of interferometers other than the Michelson and Young interferometers are described.

3.9.1 Fabry–Pérot interferometer

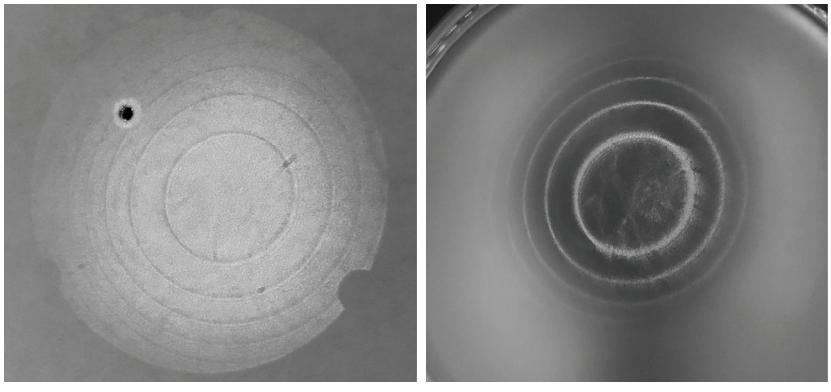
Figure 3.58 shows a system consisting of two flat mirrors, a spacer ring about 0.5 mm thick, and a cylindrical support. Mirrors are thick (wedge-shaped) plates of glass with a metallic (aluminum) coating on one side to produce high reflectance ($0.9 < r < 1$). The spacer ring is placed in the middle of the two mirrors (in contact with their metallized faces) guaranteeing a fixed distance between the mirrors. The mirror and ring assembly is placed inside the cylindrical support and adjusted. This device is known as a Fabry–Pérot interferometer, in its simplest form. The operation of the interferometer is explained with the concepts developed for the plate with parallel faces, but with $n_l = 1$ (air between the mirrors).

A schematic of an experimental setup is shown in Fig. 3.59(a) for simultaneously observing interference fringes (nonlocalized) by reflection and

Figure 3.58 A basic Fabry–Pérot interferometer. It consists of two flat mirrors of semi-silvered glass on one of the faces, a spacer ring (∼0.5 mm thick), and a cylindrical support that adjusts the mirror system with the spacer in the middle.



(a)



(b) (c)

Figure 3.59 (a) Optical system for simultaneously observing reflected and transmitted interference. (b) Reflection interferogram on screen 1. (c) Transmission interferogram on screen 2.

transmission with the Fabry–Pérot interferometer shown in Fig. 3.58. A laser beam is focused at the center of a hole about 3 mm in diameter located in an opaque screen (screen 1). The point source S will be there. The divergent spherical wavefront is transmitted and reflected several times at the interferometer plates. The interferograms seen by reflection on screen 1 and by transmission on screen 2 are shown in Figs. 3.59(b) and (c), respectively. In (b), the dark dot corresponds to the hole in which the point source is located. In effect, the two interferograms complement each other, as in Figs. 3.37(c) and 3.38(c).

Fabry–Pérot interferometer as a resonant cavity

A very important application of the Fabry–Pérot interferometer is in lasers, where the interferometer is the resonant cavity. To see this, let us consider Fig. 3.34, which shows that for $r = 0.96$, a very narrow band of transmitted irradiance is obtained. This figure is based on the thickness of the plate. However, if the thickness of the plate is fixed and the irradiance is plotted as a function of the frequency of the wave, a similar plot is obtained, since the phase given in Eq. (3.79) also depends linearly on the frequency $\nu = c/\lambda$. That is, for very high reflectances of the mirrors, the transmitted irradiance only occurs in very narrow frequency bands, which explains the high degree of monochromaticity of lasers.

This book does not deal with the details of the Fabry–Pérot interferometer as a resonant cavity, which can be found in specialized works on lasers [5]. An introduction to the subject can also be obtained by consulting elementary optics books [6].

3.9.2 Antireflective thin film

One application of parallel face plate interference is in antireflective coatings. The simplest system consists of a plate of parallel faces whose thickness is of the order of a wavelength, called a *thin film*, placed on top of another plate of parallel faces, called a *substrate*, as illustrated in Fig. 3.60.

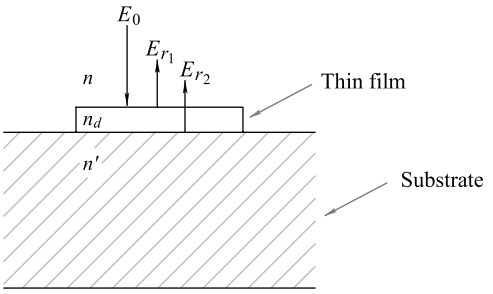


Figure 3.60 A thin film on a substrate. When the film thickness is $d = \lambda/4n_d$, Er_1 and Er_2 interfere destructively.

The reflected field is $E_{r_1} = r_1 E_0$ on the first face and $E_{r_2} = t_1 r_2 t_1' E_0$ on the second face. Thus, the total reflected field (omitting the reflection at the base of the substrate) would be

$$E_r = r_1 E_0 + t_1 r_2 t_1' E_0 e^{i\delta},$$

where

$$\delta = \frac{4\pi n_d}{\lambda} d.$$

The reflected irradiance would be

$$I_r = I_0 R,$$

with $I_0 = (\epsilon_0 c/2) E_0^2$ and the reflectance

$$R = r_1^2 + (t_1 t_1' r_2)^2 + 2 r_1 t_1 t_1' r_2 \cos\delta.$$

At normal incidence, the reflection and transmission coefficients (for the parallel polarization state) are

$$r_1 = \frac{n_d - n}{n_d + n}; \quad r_2 = \frac{n' - n_d}{n' + n_d}; \quad t_1 = \frac{2n}{n_d + n}; \quad t_1' = \frac{2n_d}{n_d + n}.$$

In practice, it is common to use MgF_2 (magnesium fluoride), whose index is $n_d = 1.38$, for the thin film. If this film is deposited on a glass substrate ($n' = 1.5$) and $n = 1$ (air), the relation $n < n_d < n'$ holds, so the reflectance takes a minimum value when $\delta = \pi$; i.e., the thickness of the film must be

$$d = \frac{\lambda}{4 n_d}.$$

With a little more work, we find that the minimum takes the value of zero if

$$n_d^2 = n' n.$$

For example, for the glass substrate, the thin film should have a refractive index of 1.22, somewhat less than the refractive index of MgF_2.

3.9.3 Newton and Fizeau interferometers

In general, any arrangement of two contacting optical surfaces illuminated with monochromatic light is called a Newton interferometer [7]. The name given to this type of interferometer comes from the first reported experiments by Newton [8] of bringing a pair of telescope lenses into contact. In this experiment, Newton observes colored circular fringes centered on the contact region. Something similar to this is shown in Fig. 3.56, where the center of the

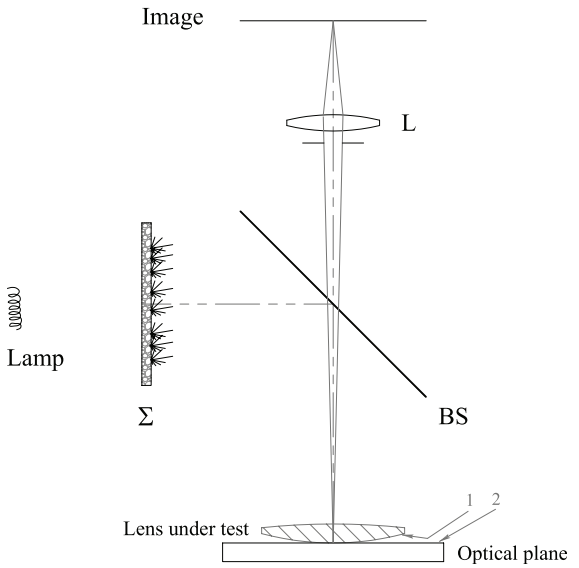


Figure 3.61 A mount to observe Newton's rings generated by the curved face of the lens in contact with an optical reference plane.

colored fringes is at the tip of the pencil. These circular fringes are also called *Newton's rings*.

The basic setup for observing Newton's rings is shown in Fig. 3.61. A lens with spherical faces is placed in an optical reference plane (peak–valley error, λ/100). The surfaces in contact, spherical surface 1 and flat surface 2, form a film of air. An extended monochromatic source (mercury discharge lamp) is used to illuminate the surfaces in contact. The interference pattern formed by the superposition of externally reflected waves in the optical plane (surface 2) and internally reflected waves in the lens (surface 1) is observed with the help of a lens L. With this configuration, the quality of the curved surface of the lens that is in contact with an optical reference plane is examined. If the face of the lens is spherical, there will be an interference pattern made up of circular rings. The radius of the mth bright ring is given by

$$\rho_m = \sqrt{(m + 1/2)\lambda R}, \tag{3.120}$$

and the radius of the mth dark ring will be

$$\rho_m = \sqrt{m\lambda R}, \tag{3.121}$$

where R is the radius of curvature of the spherical surface 1 and $m = 0, 1, 2, \ldots$ It is left as a task for the reader to derive Eqs. (3.120) and (3.121). Note that now the central fringe is labeled $m = 0$ (where the contact region is), whereas for fringes generated by two virtual sources, as in Fig. 3.16, the value of m for

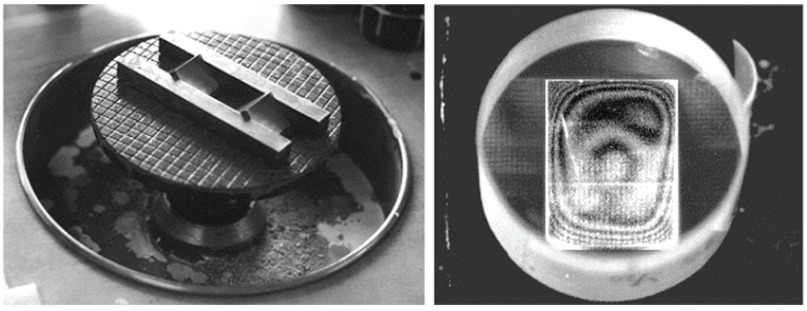


Figure 3.62 (a) Polishing of the diagonal surfaces of two right prisms. (b) Interference fringes generated by the diagonal surface of one of the prisms in (a) when it comes into contact with the optical reference plane of the Newton interferometer.

the central fringe depends on the relationship between the separation of the virtual sources and the wavelength.

The configuration in Fig. 3.61 can also be used to assess the quality of the flat surfaces of other optical elements (lenses, planes, prisms); e.g., in the process of polishing the flat surfaces of a prism [Fig. 3.62(a)], the quality of the surfaces can be checked by putting the surface to be evaluated in contact with the optical plane of the Newton interferometer shown in Fig. 3.61. In Fig. 3.62(b), the fringe pattern obtained for the diagonal surface of one of the prisms in Fig. 3.62(a) is shown. Because the fringes have the same thickness, a map of the topography of the surface under analysis can be obtained.

In general, the reference surface should have a similar shape to the surface to be evaluated by the interference fringes. If the shape of the two surfaces

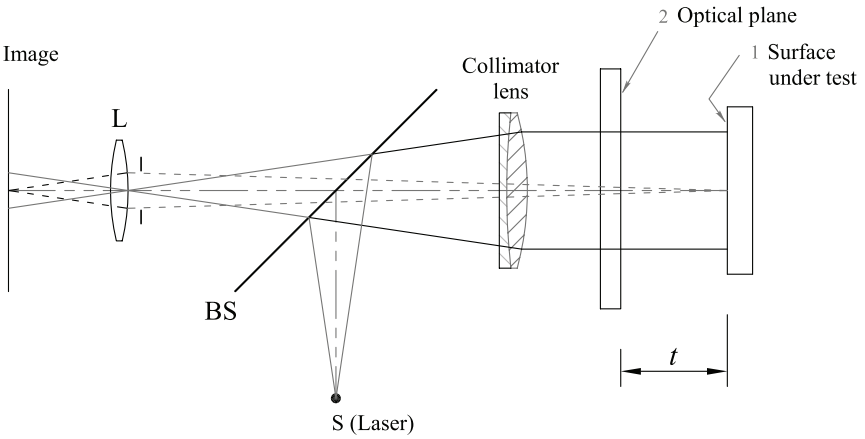


Figure 3.63 Scheme of the Fizeau interferometer to evaluate a test surface that is separated by a distance t from the optical reference plane.

differs by a large number of wavelengths, there will be many fringes and the imaging system will not be able to resolve them.

If in practice it is not possible to place the two surfaces in contact, a variation of the optical system shown in Fig. 3.61 can be made, as shown in Fig. 3.63 (Fizeau interferometer). The light source must have a coherence length greater than twice the separation t between the optical reference plane and the test surface. Therefore, it is common to use a laser beam focused in S. The interference fringes will be of equal thickness.

References

[1] M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, Oxford, England (1993).
[2] G. R. Fowles, *Introduction to Modern Optics*, 2nd ed., Dover Publications, Mineola, New York (1989).
[3] J. D. Jackson, *Classical Electrodynamics*, 3rd ed., Wiley, New York (1999).
[4] T. Young, "The Interference of light," Chap. 7 in *Great Experiments in Physics: Firsthand Accounts from Galileo to Einstein*, 2nd ed., M. H. Shamos, Ed., Dover Publications, Mineola, New York, 93–107 (1987).
[5] M. Csele, *Fundamentals of Light Sources and Lasers*, Wiley, New York (2004).
[6] E. Hecht, *Optics*, Global Edition, 5th ed., Pearson, Harlow, England (2017).
[7] M. V. Mantravadi and D. Malacara, "Newton, Fizeau, and Haidinger interferometers," Chap. 1 in *Optical Shop Testing*, 3rd ed., D. Malacara, Ed., Wiley, New York, 1–45 (2007).
[8] I. Newton, *Opticks: or, a Treatise of the Reflections, Refractions, Inflections and Colours of Light*, Sam. Smith and Benj. Walford, London (1704).