4 ed OPTICS

Eugene Hecht

Adelphi University

Content

9 Interference

The intricate color patterns shimmering across an oil slick on a wet asphalt pavement (see photo) result from one of the more common manifestations of the phenomenon of interference.* On a macroscopic scale we might consider the related problem of the interaction of surface ripples on a pool of water. Our everyday experience with this kind of situation allows us to envision a complex distribution of disturbances (as shown, e.g., in Fig. 9.1). There might be regions where two (or more) waves have overlapped, partially or even completely canceling each other. Still other regions might exist in the pattern, where the resultant troughs and crests are even more pronounced than those of any of the constituent waves. After being superimposed, the individual waves separate and continue on, completely unaffected by their previous encounter.

Although the subject could be treated from the perspective of QED (p. 139), we'll take a much simpler approach. The wave theory of the electromagnetic nature of light provides a natural basis from which to proceed. Recall that the expression describing the optical disturbance is a second-order, homogeneous, linear, partial, differential equation [Eq. (3.22)]. As we have seen, it therefore obeys the important *Superposition Principle*. Accordingly, the resultant electric-field intensity $\vec{E}$, at a point in space where two or more lightwaves overlap, is equal to the *vector sum* of the individual constituent disturbances. Briefly then, *optical interference corresponds to the interaction of two or more lightwaves yielding a resultant irradiance that deviates from the sum of the component irradiances*.

These roughly circular interference fringes are due to an oil film on wet pavement. They are *fringes of equal thickness* (see p. 404) and so don't change when viewed at different angles. Of course, they appear in a rainbow of colors.

Out of the multitude of optical systems that produce interference, we will choose a few of the more important to examine. Interferometric devices will be divided, for the sake of discussion, into two groups: *wavefront splitting* and *amplitude splitting*. In the first instance, portions of the primary wavefront are used either directly as sources to emit secondary waves or in conjunction with optical devices to produce virtual sources of secondary waves. These secondary waves are then brought together, thereupon to interfere. In the case of amplitude splitting, the primary wave itself is divided into two segments, which travel different paths before recombining and interfering.

---

*The layer of water on the asphalt allows the oil film to assume the shape of a smooth planar surface. The black asphalt absorbs the transmitted light, preventing back reflection, which would tend to obscure the fringes.

Max   Min   Max   Min   Max   Min   Max   Min   Max   Min
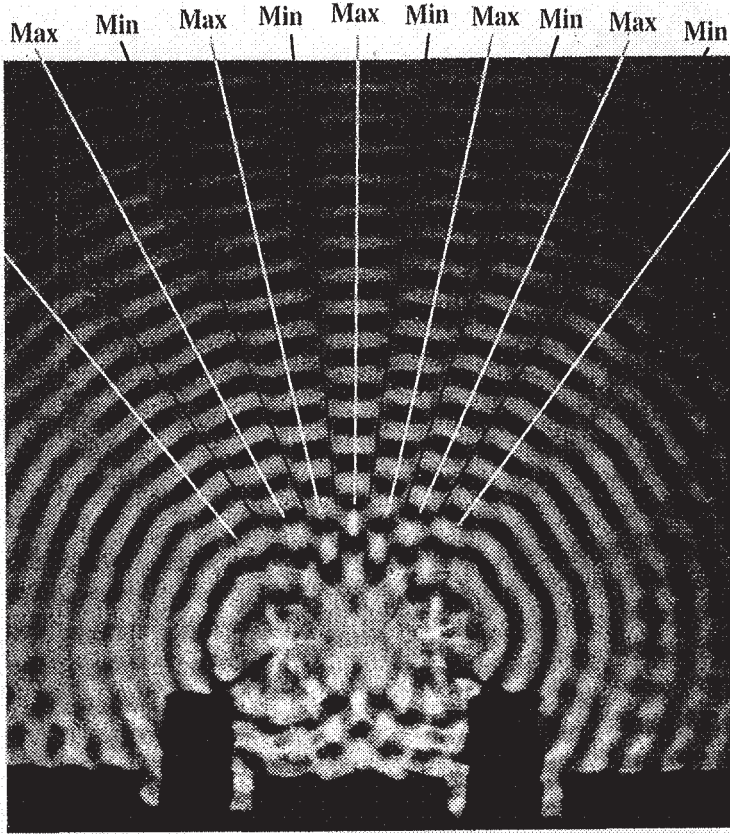
Figure 9.1 Water waves from two in-phase point sources in a ripple tank. In the middle of the pattern the wave peaks (thin bright bands), and troughs (thin black bands) lie within long wedge-shaped areas (maxima) separated by narrow dark regions of calm (minima). Although the nodal lines look straight, they're really hyperbolic. The optical equivalent is the electric field distribution depicted in Fig. 9.3c. (Photo courtesy PSSC *College Physics*, 1968, @ 1965 Educational Development Center, Inc.)

9.1 General Considerations

We have already examined the problem of the superposition of two scalar waves (Section 7.1), and in many respects those results will again be applicable. But light is, of course, a vector phenomenon; the electric and magnetic fields are vector fields. An appreciation of this fact is fundamental to any kind of intuitive understanding of interference. Still, there are many situations in which the particular optical system can be so configured that the vector nature of light is of little practical significance. We will derive the basic interference equations within the context of the vector model, thereafter delineating the conditions under which the scalar treatment is applicable.

In accordance with the Principle of Superposition, the electric field intensity $\vec{E}$, at a point in space, arising from the separate fields $\vec{E}_1$, $\vec{E}_2$,... of various contributing sources is given

by

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \cdots \qquad (9.1)$$

The optical disturbance, or light field $\vec{E}$, varies in time at an exceedingly rapid rate, roughly

$$4.3 \times 10^{14} \text{ Hz} \quad \text{to} \quad 7.5 \times 10^{14} \text{ Hz}$$

making the actual field an impractical quantity to detect. On the other hand, the irradiance I can be measured directly with a wide variety of sensors (e.g., photocells, bolometers, photographic emulsions, or eyes). The study of interference is therefore best approached by way of the irradiance.

Much of the analysis to follow can be performed without specifying the particular shape of the wavefronts, and the results are therefore quite general (Problem 9.1). For the sake of simplicity, however, consider two point sources, S_1 and S_2, emitting monochromatic waves of the same frequency in a homogeneous medium. Let their separation a be much greater than λ. Locate the point of observation P far enough away from the sources so that at P the wavefronts will be planes (Fig. 9.2). For the moment, consider only linearly polarized waves of the form

$$\vec{E}_1(\vec{r}, t) = \vec{E}_{01} \cos(\vec{k}_1 \cdot \vec{r} - \omega t + \varepsilon_1) \qquad (9.2a)$$

and

$$\vec{E}_2(\vec{r}, t) = \vec{E}_{02} \cos(\vec{k}_2 \cdot \vec{r} - \omega t + \varepsilon_2) \qquad (9.2b)$$

We saw in Chapter 3 that the irradiance at P is given by

$$I = \epsilon v \langle \vec{E}^2 \rangle_T$$

Inasmuch as we will be concerned only with relative irradiances within the same medium, we will, for the time being at least, simply neglect the constants and set

$$I = \langle \vec{E}^2 \rangle_T$$

What is meant by $\langle \vec{E}^2 \rangle_T$ is of course the time average of the magnitude of the electric-field intensity squared, or $\langle \vec{E} \cdot \vec{E} \rangle_T$. Accordingly

$$\vec{E}^2 = \vec{E} \cdot \vec{E}$$

where now

$$\vec{E}^2 = (\vec{E}_1 + \vec{E}_2) \cdot (\vec{E}_1 + \vec{E}_2)$$

and thus

$$\vec{E}^2 = \vec{E}_1^2 + \vec{E}_2^2 + 2\vec{E}_1 \cdot \vec{E}_2 \qquad (9.3)$$
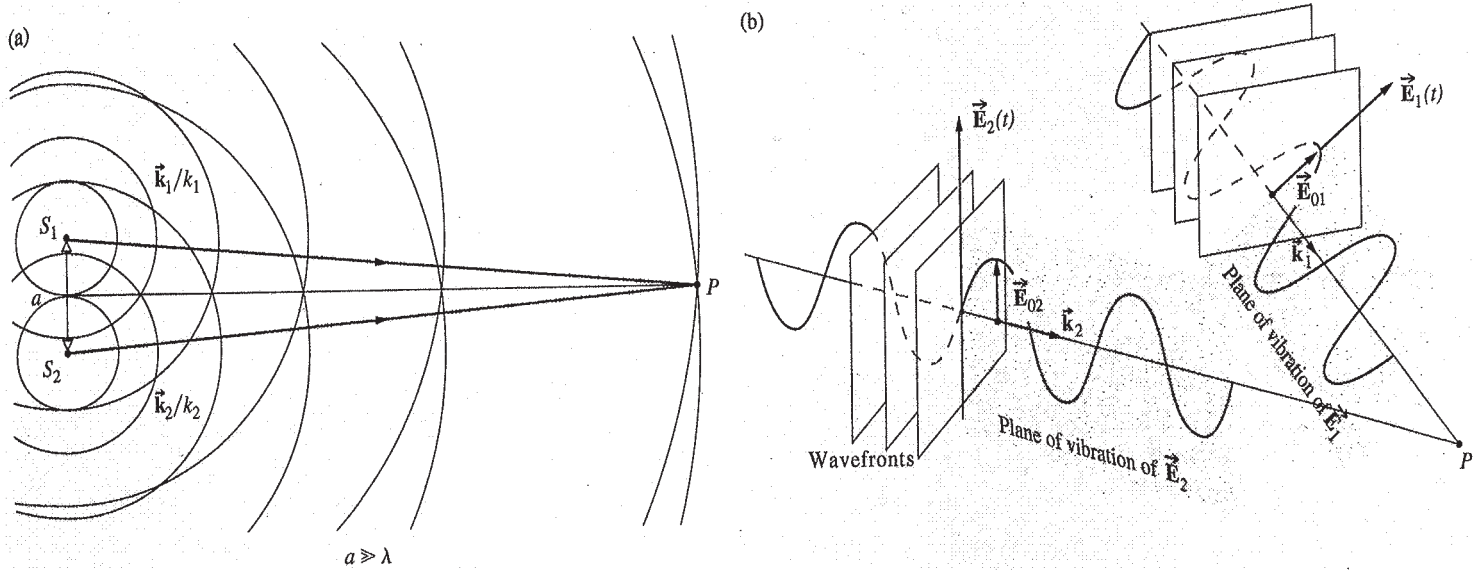
Figure 9.2 Waves from two point sources overlapping in space.

Taking the time average of both sides, we find that the irradiance becomes

$$I = I_1 + I_2 + I_{12} \tag{9.4}$$

provided that

$$I_1 = \langle \vec{E}_1^2 \rangle_T \tag{9.5}$$

$$I_2 = \langle \vec{E}_2^2 \rangle_T \tag{9.6}$$

and

$$I_{12} = 2\langle \vec{E}_1 \cdot \vec{E}_2 \rangle_T \tag{9.7}$$

The latter expression is known as the *interference term*. To evaluate it in this specific instance, we form

$$\vec{E}_1 \cdot \vec{E}_2 = \vec{E}_{01} \cdot \vec{E}_{02} \cos (\vec{k}_1 \cdot \vec{r} - \omega t + \varepsilon_1)$$
$$\times \cos (\vec{k}_2 \cdot \vec{r} - \omega t + \varepsilon_2) \tag{9.8}$$

or equivalently

$$\vec{E}_1 \cdot \vec{E}_2 =$$

$$\vec{E}_{01} \cdot \vec{E}_{02} [\cos (\vec{k}_1 \cdot \vec{r} + \varepsilon_1) \cos \omega t + \sin (\vec{k}_1 \cdot \vec{r} + \varepsilon_1) \sin \omega t]$$

$$\times [\cos (\vec{k}_2 \cdot \vec{r} + \varepsilon_2) \cos \omega t + \sin (\vec{k}_2 \cdot \vec{r} + \varepsilon_2) \sin \omega t] \tag{9.9}$$

Recall that the time average of some function $f(t)$, taken over an interval T, is

$$\langle f(t) \rangle_T = \frac{1}{T} \int_t^{t+T} f(t')\, dt' \tag{9.10}$$

The period τ of the harmonic functions is $2\pi/\omega$, and for our present concern $T \gg \tau$. In that case the $1/T$ coefficient in front of the integral has a dominant effect. After multiplying out and averaging Eq. (9.9) we have

$$\langle \vec{E}_1 \cdot \vec{E}_2 \rangle_T = \tfrac{1}{2}\vec{E}_{01} \cdot \vec{E}_{02} \cos (\vec{k}_1 \cdot \vec{r} + \varepsilon_1 - \vec{k}_2 \cdot \vec{r} - \varepsilon_2)$$

where use was made of the fact (p. 49) that $\langle \cos^2 \omega t \rangle_T = \tfrac{1}{2}$, $\langle \sin^2 \omega t \rangle_T = \tfrac{1}{2}$, and $\langle \cos \omega t \sin \omega t \rangle_T = 0$. The interference term is then

$$I_{12} = \vec{E}_{01} \cdot \vec{E}_{02} \cos \delta \tag{9.11}$$

and δ, equal to $(\vec{k}_1 \cdot \vec{r} - \vec{k}_2 \cdot \vec{r} + \varepsilon_1 - \varepsilon_2)$, is the *phase difference* arising from a combined path length and initial phase-angle difference. Notice that if $\vec{E}_{01}$ and $\vec{E}_{02}$ (and therefore $\vec{E}_1$ and $\vec{E}_2$) are perpendicular, $I_{12} = 0$ and $I = I_1 + I_2$. Two such orthogonal $\mathcal{P}$-states will combine to yield an $\mathcal{R}$-, $\mathcal{L}$-, $\mathcal{P}$-, or $\mathcal{E}$-state, but the flux-density distribution will be unaltered.

The most common situation in the work to follow corresponds to $\vec{E}_{01}$ parallel to $\vec{E}_{02}$. In that case, the irradiance reduces to the value found in the scalar treatment of Section 7.1. Under those conditions

$$I_{12} = E_{01}E_{02} \cos \delta$$

This can be written in a more convenient way by noticing that

$$I_1 = \langle \vec{E}_1^2 \rangle_T = \frac{E_{01}^2}{2} \tag{9.12}$$

and

$$I_2 = \langle \vec{\mathbf{E}}_2^2 \rangle_T = \frac{E_{02}^2}{2} \qquad (9.13)$$

The interference term becomes

$$I_{12} = 2\sqrt{I_1 I_2} \cos \delta$$

whereupon the total irradiance is

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \delta \qquad (9.14)$$

At various points in space, the resultant irradiance can be greater, less than, or equal to $I_1 + I_2$, depending on the value of I_{12}, that is, depending on δ. A maximum irradiance is obtained when $\cos \delta = 1$, so that

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \qquad (9.15)$$

when

$$\delta = 0, \pm 2\pi, \pm 4\pi,\ldots$$

In this case of **total constructive interference**, the phase difference between the two waves is an integer multiple of 2π, and the disturbances are *in-phase*. When $0 < \cos \delta < 1$ the waves are *out-of-phase*, $I_1 + I_2 < I < I_{max}$, and the result is *constructive interference*. At $\delta = \pi/2$, $\cos \delta = 0$, the optical disturbances are 90° out-of-phase, and $I = I_1 + I_2$. For $0 > \cos \delta > -1$ we have the condition of *destructive interference*, $I_1 + I_2 > I > I_{min}$. A minimum irradiance results when the waves are 180° out-of-phase, troughs overlap crests, $\cos \delta = -1$, and

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2} \qquad (9.16)$$

This occurs when $\delta = \pm \pi, \pm 3\pi, \pm 5\pi,\ldots$, and it is referred to as **total destructive interference**.

Another somewhat special yet very important case arises when the amplitudes of both waves reaching P in Fig. 9.2 are equal (i.e., $\vec{\mathbf{E}}_{01} = \vec{\mathbf{E}}_{02}$). Since the irradiance contributions from both sources are then equal, let $I_1 = I_2 = I_0$. Equation (9.14) can now be written as

$$I = 2I_0(1 + \cos \delta) = 4I_0 \cos^2 \frac{\delta}{2} \qquad (9.17)$$

from which it follows that $I_{min} = 0$ and $I_{max} = 4I_0$. For an analysis in terms of the angle between the two beams, see Problem 9.3.

Equation (9.14) holds equally well for the spherical waves emitted by S_1 and S_2. Such waves can be expressed as

$$\vec{\mathbf{E}}_1(r_1, t) = \vec{\mathbf{E}}_{01}(r_1) \exp [i(kr_1 - \omega t + \varepsilon_1)] \quad (9.18a)$$

and

$$\vec{\mathbf{E}}_2(r_2, t) = \vec{\mathbf{E}}_{02}(r_2) \exp [i(kr_2 - \omega t + \varepsilon_2)] \quad (9.18b)$$

The terms r_1 and r_2 are the radii of the spherical wavefronts overlapping at P; they specify the distances from the sources to P. In this case

$$\delta = k(r_1 - r_2) + (\varepsilon_1 - \varepsilon_2) \qquad (9.19)$$

The flux density in the region surrounding S_1 and S_2 will certainly vary from point to point as $(r_1 - r_2)$ varies. Nonetheless, from the principle of conservation of energy, we expect the spatial average of I to remain constant and equal to the average of $I_1 + I_2$. The space average of I_{12} must therefore be zero, a property verified by Eq. (9.11), since the average of the cosine term is, in fact, zero. (For further discussion of this point, see Problem 9.2.)

Equation (9.17) will be applicable when the separation between S_1 and S_2 is small in comparison with r_1 and r_2 and when the interference region is also small in the same sense. Under these circumstances, $\vec{\mathbf{E}}_{01}$ and $\vec{\mathbf{E}}_{02}$ may be considered independent of position, that is, constant over the small region examined. If the emitting sources are of equal strength, $\vec{\mathbf{E}}_{01} = \vec{\mathbf{E}}_{02}$, $I_1 = I_2 = I_0$ and we have

$$I = 4I_0 \cos^2 \tfrac{1}{2}[k(r_1 - r_2) + (\varepsilon_1 - \varepsilon_2)]$$

Irradiance maxima occur when

$$\delta = 2\pi m$$

provided that $m = 0, \pm 1, \pm 2,\ldots$ Similarly, minima, for which $I = 0$, arise when

$$\delta = \pi m'$$

where $m' = \pm 1, \pm 3, \pm 5,\ldots$, or if you like, $m' = 2m + 1$. Using Eq. (9.19) these two expressions for δ can be rewritten such that maximum irradiance occurs when

$$(r_1 - r_2) = [2\pi m + (\varepsilon_2 - \varepsilon_1)]/k \qquad (9.20a)$$

and minimum when

$$(r_1 - r_2) = [\pi m' + (\varepsilon_2 - \varepsilon_1)]/k \qquad (9.20b)$$

(a)

(b)
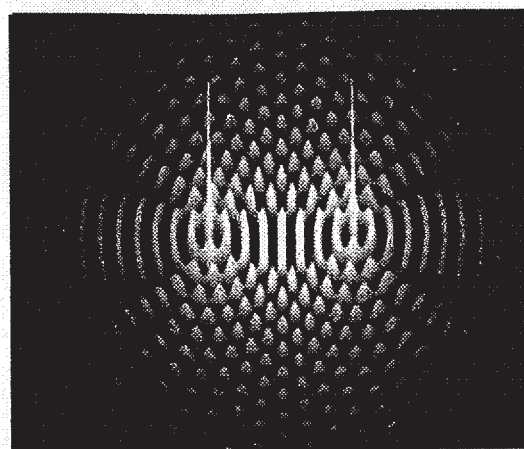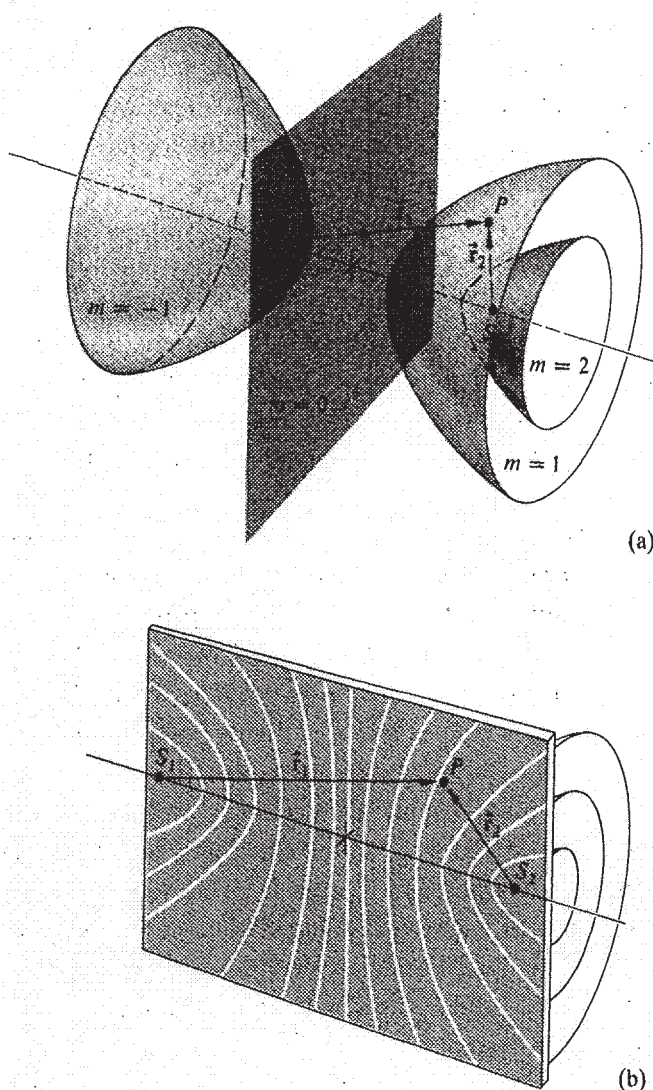
Figure 9.3 (a) Hyperboloidal surfaces of maximum irradiance for two point sources. The quantity m is positive where $r_1 > r_2$. (b) Here we see how the irradiance maxima are distributed on a plane containing S_1 and S_2. (c) The electric-field distribution in the plane shown in part (b). The tall peaks, are the point sources S_1 and S_2. Note that the spacing of the sources is different in (b) and (c). (Photo courtesy, The Optics Project, Mississippi State University.)

Either one of these equations defines a family of surfaces, each of which is a hyperboloid of revolution. The vertices of the hyperboloids are separated by distances equal to the right-hand sides of Eqs. (9.20a) and (9.20b). The foci are located at S_1 and S_2. If the waves are in-phase at the emitter, $\varepsilon_1 - \varepsilon_2 = 0$, and Eqs. (9.20a) and (9.20b) can be simplified to

[maxima]
$$(r_1 - r_2) = 2\pi m/k = m\lambda \qquad (9.21a)$$

[minima]
$$(r_1 - r_2) = \pi m'/k = \tfrac{1}{2}m'\lambda \qquad (9.21b)$$

for maximum and minimum irradiance, respectively. Figure 9.3a shows a few of the surfaces over which there are irradiance maxima. The dark and light zones that would be seen on a screen placed in the region of interference are known as **interference fringes** (Fig. 9.3b). Notice that the central bright band, equidistant from the two sources, is the so-called zeroth-order fringe ($m = 0$), which is straddled by the $m' = \pm 1$ minima, and these, in turn, are bounded by the first-order ($m = \pm 1$) maxima, which are straddled by the $m' = \pm 3$ minima, and so forth.

Since the wavelength λ for light is very small, a large number of surfaces corresponding to the lower values of m will exist close to, and on either side of, the plane $m = 0$. A number of fairly straight parallel fringes will therefore appear on a screen placed perpendicular to that ($m = 0$) plane and in the vicinity of it, and for this case the approximation $r_1 \approx r_2$ will hold. If S_1 and S_2 are then displaced normal to the $\overline{S_1 S_2}$ line, the fringes will merely be displaced parallel to themselves. *Two narrow slits will generate a large number of exactly overlapping fringes, thereby increasing the irradiance, leaving the central region of the two-point source pattern otherwise essentially unchanged.*

9.2 Conditions for Interference

If two beams are to interfere to produce a stable pattern, they must have very nearly the same frequency. A significant frequency difference would result in a rapidly varying, time-dependent phase difference, which in turn would cause I_{12} to average to zero during the detection interval (see Section 7.1). Still, if the sources both emit white light, the component reds will interfere with reds, and the blues with blues. A great many fairly similar, slightly displaced, overlapping monochromatic patterns will produce one total white-light pattern. It will not be as sharp or as extensive as a quasimonochromatic pattern, but *white light will produce observable interference.*

The clearest patterns exist when the interfering waves have equal or nearly equal amplitudes. The central regions of the dark and light fringes then correspond to complete destructive and constructive interference, respectively, yielding maximum contrast.

For a fringe pattern to be observed, the two sources need not be in-phase with each other. A somewhat shifted but otherwise identical interference pattern will occur if there is some initial phase difference between the sources, as long as it remains constant. Such sources (which may or may not be in step, but are always marching together) are **coherent**.*

9.2.1 Temporal and Spatial Coherence

Remember that because of the granular nature of the emission process, conventional quasimonochromatic sources produce light that is a mix of photon wavetrains. At each illuminated point in space there is a net field that oscillates nicely (through roughly a million cycles) for less than 10 ns or so before it randomly changes phase. This interval over which the lightwave resembles a sinusoid is a measure of its **temporal coherence**. The average time interval during which the lightwave oscillates in a predictable way we have already designated as the coherence time of the radiation. The longer the coherence time, the greater the temporal coherence of the source.

As observed from a fixed point in space, the passing lightwave appears fairly sinusoidal for some number of oscillations between abrupt changes of phase. The corresponding spatial extent over which the lightwave oscillates in a regular,

predictable way is the coherence length [Eq. (7.64)]. Once again, it will be convenient to picture the light beam as a progression of well-defined, more or less sinusoidal, wavegroups of average length Δl_c, whose phases are uncorrelated to one another. Bear in mind that *temporal coherence is a manifestation of spectral purity*. If the light were ideally monochromatic, the wave would be a perfect sinusoid with an infinite coherence length. All real sources fall short of this, and all actually emit a range of frequencies, albeit sometimes quite narrow. For instance, an ordinary laboratory discharge lamp has a coherence length of several millimeters, whereas certain kinds of lasers routinely provide coherence lengths of tens of kilometers.

Figure 9.4 summarizes some of these ideas. In (a) the wave, which arises from a point source, is monochromatic and has complete temporal coherence. What happens at P_1' will, a moment later, happen at P_2' and still later at P_3'—all totally predictably. In fact, by watching P_4' we can determine what the wave will be doing at P_1' at any time. Every point on the wave is correlated; its coherence time is unlimited. By contrast, Fig. 9.4b shows a point source that changes frequency from moment to moment. Now there's no correlation of the wave at points that are far apart like P_1' and P_4'. The waves lack the total temporal coherence displayed in (a), but they're not completely unpredictable; the behavior at points that are close together such as P_2' and P_3' are somewhat correlated. This is an instance of *partial temporal coherence*, a measure of which is the coherence length—the shortest distance over which the disturbance is sinusoidal, that is, the distance over which the phase is predictable.

Notice, in both parts of Fig. 9.4, that the behavior of the waves at points P_1, P_2, and P_3 is completely correlated. Each of the two wave streams arises from a single point source and P_1, P_2, and P_3 lie on the same wavefront in both cases; the disturbance at each of these laterally separated points is in-phase and stays in-phase. Both waves therefore exhibit complete **spatial coherence**. By contrast, suppose the source is broad, that is, composed of many widely spaced point sources (monochromatic ones of period τ), as is Fig. 9.5. If we could take a picture of the wave pattern in Fig. 9.5 every τ seconds, it would be the same; each wavefront would be replaced by an identical one, one wavelength behind it. The disturbances at P_1', P_2', and P_3' are correlated, and the wave is temporally coherent.

Now to insert a little realism; suppose each point source changes phase rapidly and randomly, emitting 10-ns long sinusoidal wavetrains. The waves in Fig. 9.5 would randomly

---

*Chapter 12 is devoted to the study of coherence, so here we'll merely touch on those aspects that are immediately pertinent.

(a)

(b)
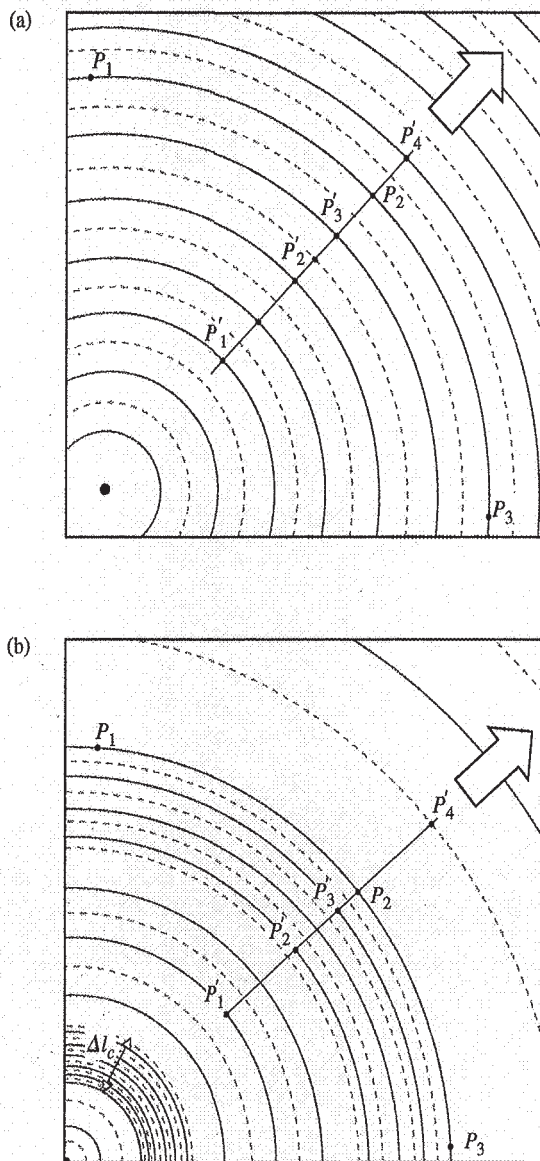
Figure 9.4 Temporal and spatial coherence. (*a*) Here the waves display both forms of coherence perfectly. (*b*) Here there is complete spatial coherence but only partial temporal coherence.

advent of the laser, it was a working principle that no two individual sources could ever produce an observable interference pattern. The coherence time of lasers, however, can be appreciable, and interference via independent lasers has been observed and photographed.* The most common means of overcoming this problem with ordinary thermal sources is to make one source serve to produce two coherent secondary sources.

9.2.2 The Fresnel–Arago Laws

In Section 9.1 it was assumed that the two overlapping optical disturbance vectors were linearly polarized and parallel. Nonetheless, the formulas apply as well to more complicated situations; indeed, the treatment is applicable regardless of the polarization state of the waves. To appreciate this, recall that
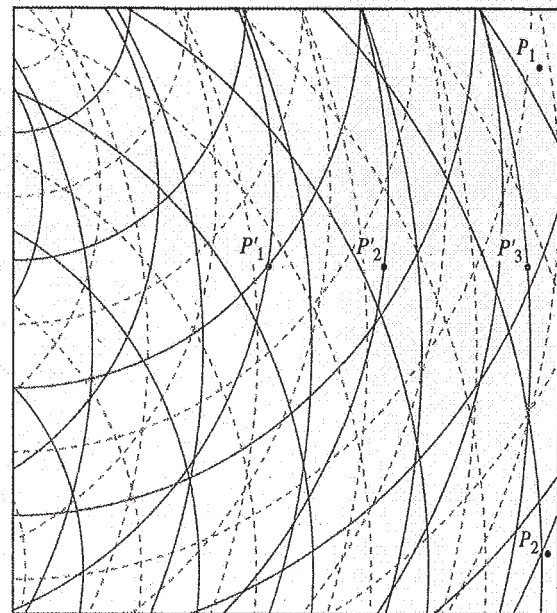
Figure 9.5 With multiple (here four) widely spaced point sources, the resultant wave is still coherent. But if those sources change phase rapidly and randomly, both the spatial and temporal coherence diminish accordingly.

change phase, shifting, combining, and recombining in a frenzied tumult. The disturbances at P_1', P_2', and P_3' would only be correlated for a time less than 10 ns. And the wave field at two, even modestly spaced, lateral points such as P_1 and P_2 would be almost completely uncorrelated depending on the size of the source. The beam from a candle flame or a shaft of sunlight is a multi-frequency mayhem much like this.

Two ordinary sources, two lightbulbs, can be expected to maintain a constant relative phase for a time no greater than Δt_c, so the interference pattern they produce will randomly shift around in space at an exceedingly rapid rate, averaging out and making it quite impractical to observe. Until the

* G. Magyar and L. Mandel, "Interference fringes produced by superposition of two independent maser light beams," *Nature* **198**, 255 (1963). F. Louradour, F. Reynaud, B. Colombeau, and C. Froehly, "Interference fringes between two separate lasers," *Am. J. Phys.* **61**, 242 (1993). L. Basano and P. Ottonello, "Interference fringes from stabilized diode lasers," *Am. J. Phys.* **68**, 245 (2000). E. C. G. Sudarshan and T. Rothman, "The two-slit interferometer reexamined," *Am. J. Phys.* **59**, 592 (1991).
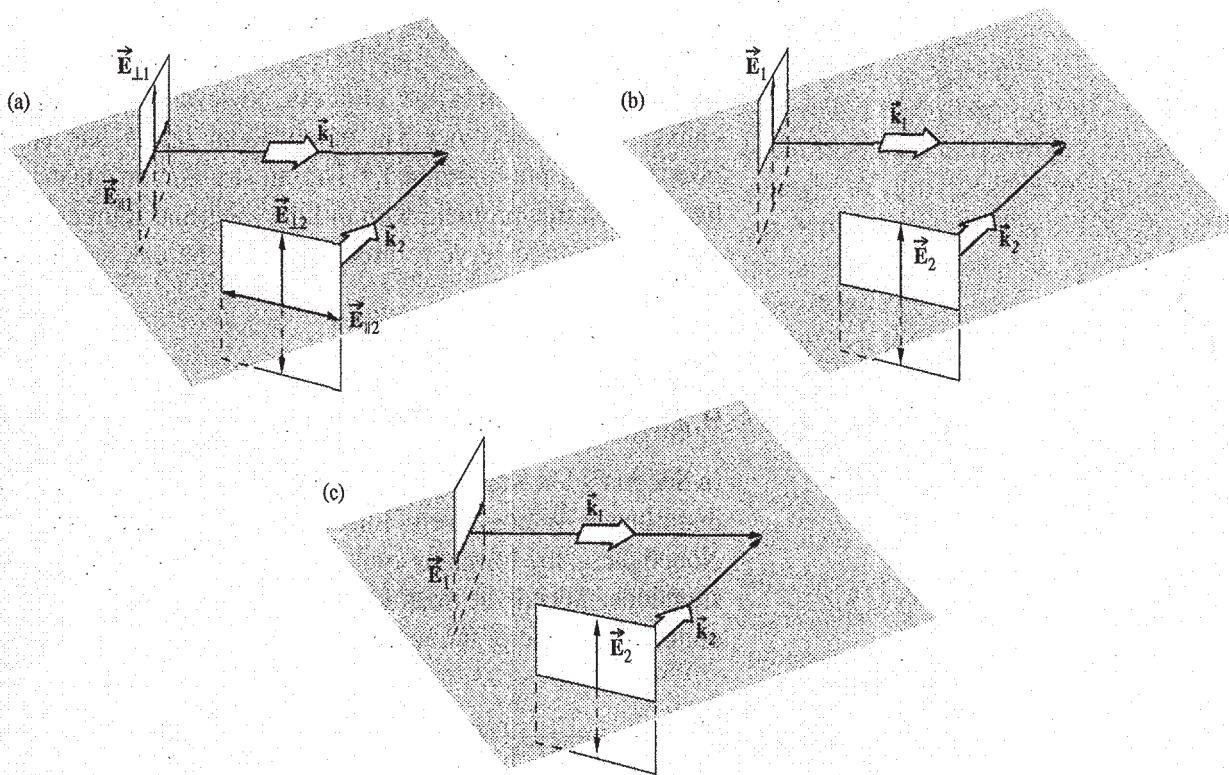
Figure 9.6 Interference of polarized light.

any polarization state can be synthesized out of two orthogonal $\mathcal{P}$-states. For natural light these $\mathcal{P}$-states are mutually incoherent, but that represents no particular difficulty.

Suppose that every wave has its propagation vector in the same plane, so that we can label the constituent orthogonal $\mathcal{P}$-states, with respect to that plane, for example, $\vec{\mathbf{E}}_\parallel$ and $\vec{\mathbf{E}}_\perp$, which are parallel and perpendicular to the plane, respectively (Fig. 9.6a). Thus any plane wave, whether polarized or not, can be written in the form $(\vec{\mathbf{E}}_\parallel + \vec{\mathbf{E}}_\perp)$. Imagine that the waves $(\vec{\mathbf{E}}_{\parallel 1} + \vec{\mathbf{E}}_{\perp 1})$ and $(\vec{\mathbf{E}}_{\parallel 2} + \vec{\mathbf{E}}_{\perp 2})$ emitted from two identical coherent sources superimpose in some region of space. The resulting flux-density distribution will consist of two independent, precisely, overlapping interference patterns $\langle(\vec{\mathbf{E}}_{\parallel 1} + \vec{\mathbf{E}}_{\parallel 2})^2\rangle_\mathrm{T}$ and $\langle(\vec{\mathbf{E}}_{\perp 1} + \vec{\mathbf{E}}_{\perp 2})^2\rangle_\mathrm{T}$. Therefore, *although we derived the equations of the previous section specifically for linear light, they are applicable to any polarization state, including natural light.*

Notice that even though $\vec{\mathbf{E}}_{\perp 1}$ and $\vec{\mathbf{E}}_{\perp 2}$ are always parallel to each other, $\vec{\mathbf{E}}_{\parallel 1}$ and $\vec{\mathbf{E}}_{\parallel 2}$, which are in the reference plane, need not be. They will be parallel only when the two beams are themselves parallel (i.e., $\vec{\mathbf{k}}_1 = \vec{\mathbf{k}}_2$). The inherent vector nature of the interference process as manifest in the dot-product representation [Eq. (9.11)] of I_{12} cannot be ignored. There are many practical situations in which the beams approach being parallel, and in these cases the scalar theory will do nicely. Even so, (b) and (c) in Fig. 9.6 are included as an urge to caution. They depict the imminent overlapping of two coherent linearly polarized waves. In Fig. 9.6b the optical vectors are parallel, even though the beams aren't, and interference would nonetheless result. In Fig. 9.6c the optical vectors are perpendicular, and $I_{12} = 0$, which would be the case here even if the beams were parallel.

Fresnel and Arago made an extensive study of the conditions under which the interference of polarized light occurs, and their conclusions summarize some of the above considerations. The **Fresnel–Arago Laws** are as follows:

1. Two orthogonal, coherent $\mathcal{P}$-states cannot interfere in the sense that $I_{12} = 0$ and no fringes result.
2. Two parallel, coherent $\mathcal{P}$-states will interfere in the same way as will natural light.
3. The two constituent orthogonal $\mathcal{P}$-states of natural light cannot interfere to form a readily observable fringe pattern even if rotated into alignment. This last point is understandable, since these $\mathcal{P}$-states are incoherent.

9.3 Wavefront-splitting Interferometers

The main problem in producing interference is the sources: they must be *coherent*. And yet separate, independent, adequately coherent sources, other than the modern laser, don't exist! That dilemma was first solved two hundred years ago by Thomas Young in his classic double-beam experiment. He brilliantly took a single wavefront, split off from it two coherent portions, and had them interfere.

9.3.1 Young's Experiment

In 1665 Grimaldi described an experiment he had performed to examine the interaction between two beams of light. He admitted sunlight into a dark room through two close-together pinholes in an opaque screen. Like a camera obscura (p. 215), each pinhole cast an image of the Sun on a distant white surface. The idea was to show that where the circles of light overlapped, darkness could result. Although at the time he couldn't possibly understand why, the experiment failed because the primary source, the Sun's disk (which subtends about 32 minutes of arc), was too large and therefore the incident light didn't have the necessary spatial coherence in order to properly simultaneously illuminate the two pinholes. To do that, the Sun would have had to subtend only a few seconds of arc.

A hundred and forty years later, Dr. Thomas Young (guided by the phenomenon of beats, which was understood to be produced by two overlapping sound waves) began his efforts to establish the wave nature of light. He redid Grimaldi's experiment, but this time the sunlight passed through an initial pinhole, which became the primary source (Fig. 9.7). This had the effect of creating a spatially coherent beam that could identically illuminate the two apertures. In this way Young succeeded in producing a system of alternating bight and dark bands—interference fringes. Today, aware of the physics involved, we generally replace the pinholes with narrow slits that let through much more light (Fig. 9.8a).

Consider a hypothetical monochromatic plane wave illuminating a long narrow slit. From that primary slit light will be diffracted out at all angles in the forward direction and a cylindrical wave will emerge. Suppose that this wave, in turn, falls on two parallel, narrow, closed spaced slits, S_1 and S_2. This is shown in a three-dimensional view in Fig. 9.8a. When symmetry exists, the segments of the primary wavefront arriving at the two slits will be exactly in-phase, and the slits will con-

stitute two coherent secondary sources. We expect that wherever the two waves coming from S_1 and S_2 overlap, interference will occur (provided that the optical path difference is less than the coherence length, $c\Delta t_c$).

Figures 9.8a, b, and c correspond to the classic arrangement of **Young's Experiment**, although there are other variations. Nowadays the first screen is usually dispensed with, and plane waves from a laser directly illuminate the aperture screen (Fig. 9.8f). In a realistic physical situation, the distance between each of the screens (Σ_a and Σ_o) in Fig. 9.8c would be very large in comparison with the distance a between the two slits, several thousand times as much, and all the fringes would be fairly close to the center O of the screen. The optical path difference between the rays along $\overline{S_1P}$ and $\overline{S_2P}$ can be determined, to a good approximation, by dropping a perpendicular from S_2 onto $\overline{S_1P}$. This path difference is given by

$$(\overline{S_1B}) = (\overline{S_1P}) - (\overline{S_2P}) \tag{9.22}$$

or

$$(\overline{S_1B}) = r_1 - r_2$$

Continuing with this approximation (see Problem 9.15), $(r_1 - r_2) = a \sin\theta$ and so

$$r_1 - r_2 \approx a\theta \tag{9.23}$$

since $\theta \approx \sin\theta$. Notice that

$$\theta \approx \frac{y}{s} \tag{9.24}$$
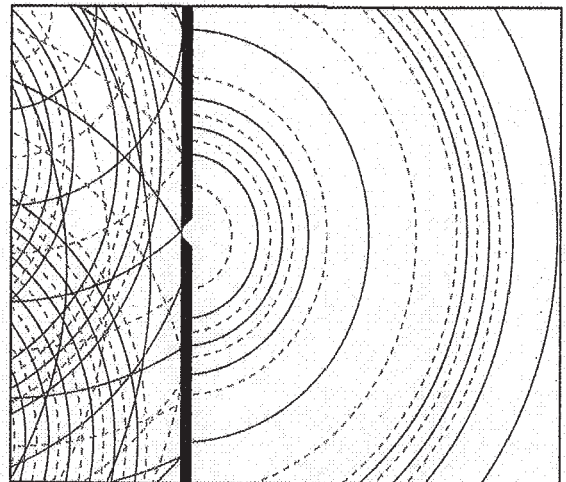
Figure 9.7 The pinhole scatters a wave that is spatially coherent, even though it's not temporally coherent.

(a)

Σ_a

Σ_o

(b)

max
min
max
min
max
min
max
min
max

Σ_a

Σ_o

(c)

P

r_2

r_1

y_m

θ_m

θ_m

O

S_2

a

S

r

S_1

B

s

Σ_a

Σ_o
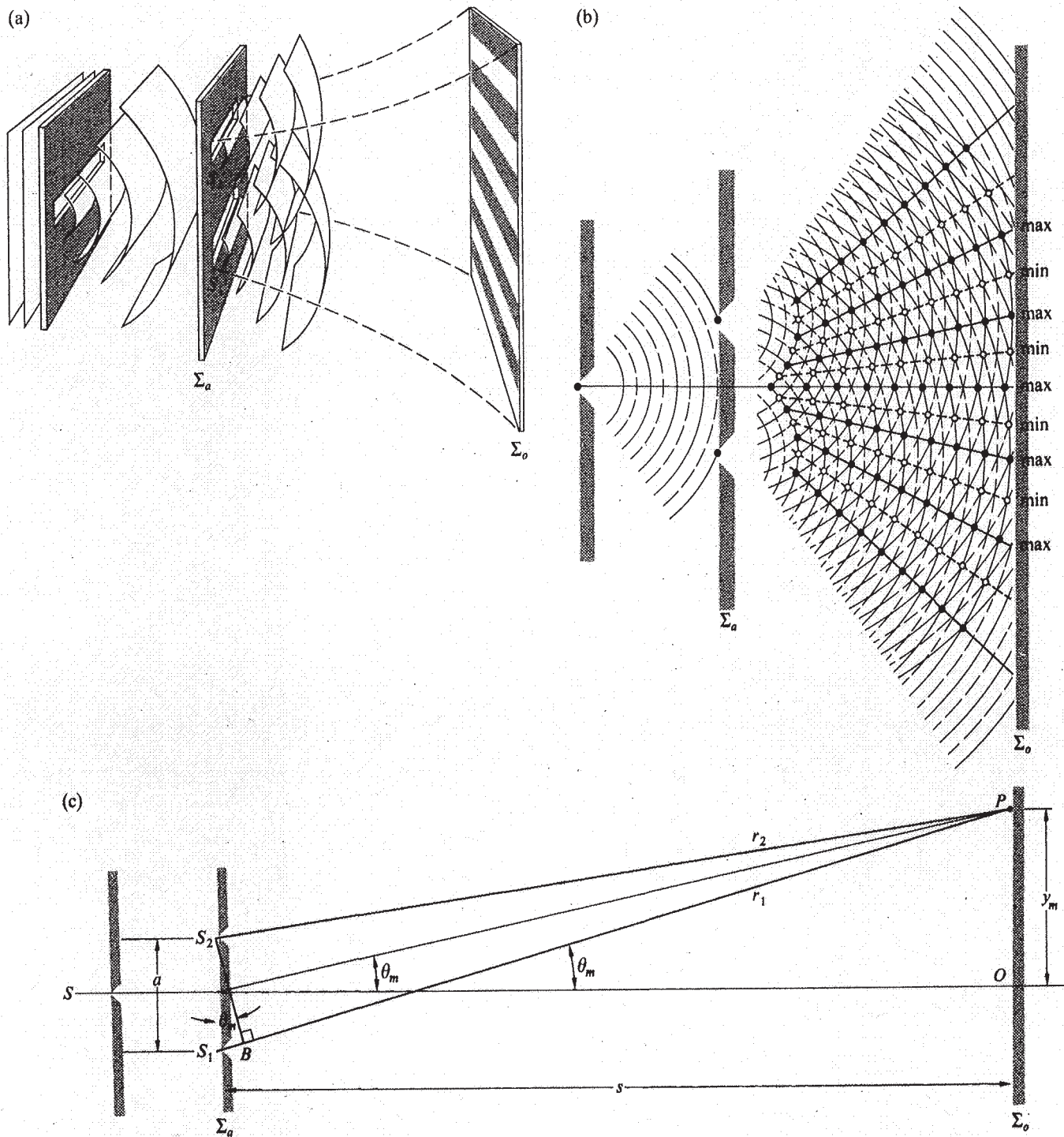
In accordance with Section 9.1, *constructive* interference

Figure 9.8 Young's Experiment. (*a*) Cylindrical waves superimposed in the region beyond the aperture screen. (*b*) Overlapping waves showing peaks and troughs. The maxima and minima lie along nearly straight hyperbolas. (*c*) The geometry of Young's Experiment. (*d*) A path length difference of one wavelength corresponds to $m = \pm 1$ and the first-order maximum. (*e*) (Photo courtesy M. Cagnet, M. Francon, and J. C. Thierr: *Atlas optischer Erscheinungen*, Berlin–Heidelberg–New York: Springer, 1962.) (*f*) A modern version of Young's Experiment using a photodetector (e.g., a photovoltaic cell or photodiode like the RS 305-462) and an $X - Y$ recorder. The detector rides on a motor-driven slide and scans the interference pattern.

and so
$$r_1 - r_2 \approx \frac{a}{s} y \qquad (9.25)$$
will occur when

(d)

(e)

(f)
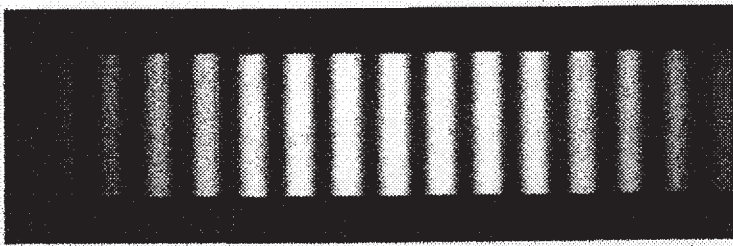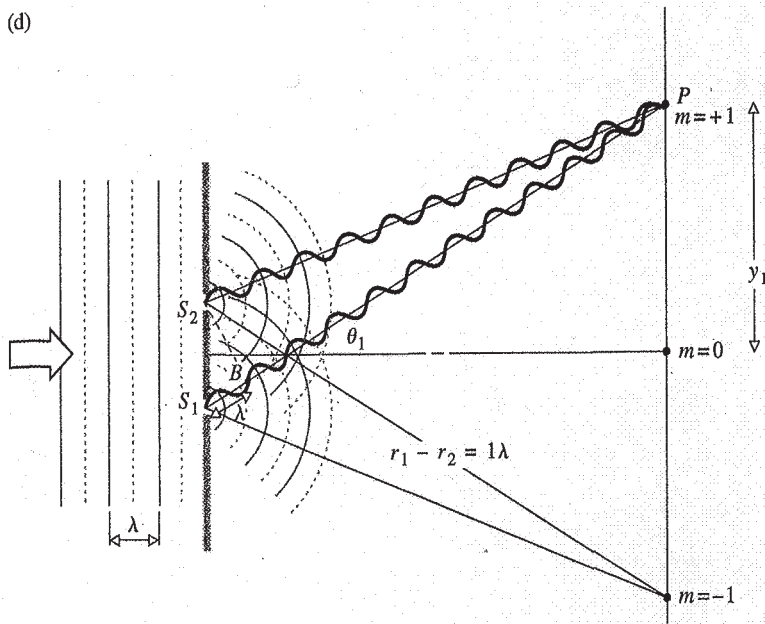
$$r_1 - r_2 = m\lambda \qquad (9.26)$$

Thus, from the last two relations we obtain

$$y_m \approx \frac{s}{a} m\lambda \qquad (9.27)$$

This gives the position of the mth bright fringe on the screen, if we count the maximum at 0 as the zeroth fringe. The angular position of the fringe is obtained by substituting the last expression into Eq. (9.24); thus

$$\theta_m = \frac{m\lambda}{a} \qquad (9.28)$$

This relationship can be obtained directly by inspecting Fig. 9.8c. For the mth-order interference maximum, m whole wavelengths should fit within the distance $r_1 - r_2$. Therefore, from the triangle $S_1 S_2 B$,

$$a \sin \theta_m = m\lambda \qquad (9.29)$$
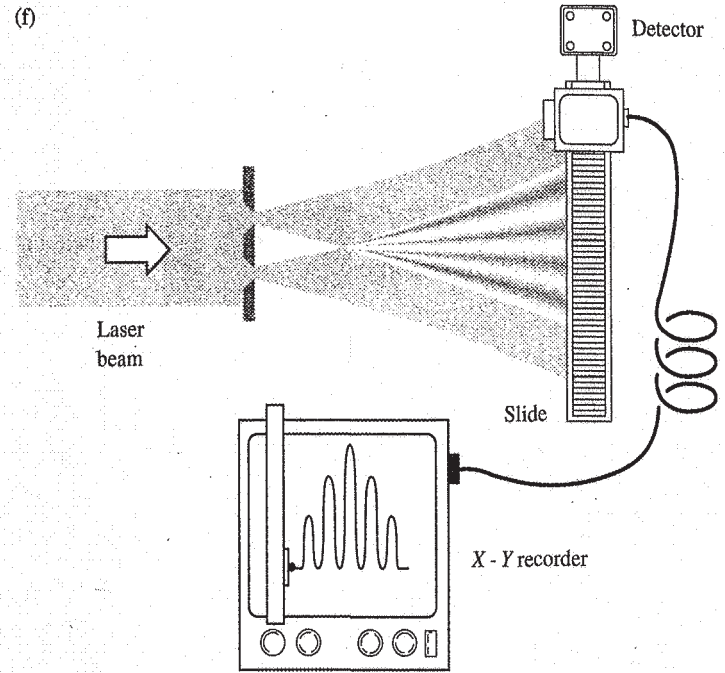
or

$$\theta_m \approx m\lambda/a$$

The spacing of the fringes on the screen can be gotten readily from Eq. (9.27). The difference in the positions of two consecutive maxima is

$$y_{m+1} - y_m \approx \frac{s}{a}(m+1)\lambda - \frac{s}{a} m\lambda$$

or

$$\Delta y \approx \frac{s}{a}\lambda \qquad (9.30)$$

Evidently, red fringes are broader than blue ones.

Since this pattern is equivalent to that obtained for two overlapping spherical waves (at least in the $r_1 \approx r_2$ region), we can apply Eq. (9.17). Using the phase difference

$$\delta = k(r_1 - r_2)$$

Eq. (9.17) can be rewritten as

$$I = 4I_0 \cos^2 \frac{k(r_1 - r_2)}{2}$$

provided, of course, that the two beams are coherent and have equal irradiances I_0. With

$$r_1 - r_2 \approx ya/s$$

the resultant irradiance becomes

$$I = 4I_0 \cos^2 \frac{ya\pi}{s\lambda} \qquad (9.31)$$

As shown in Fig. 9.9, consecutive maxima are separated by the Δy given in Eq. (9.30). Remember that we effectively assumed that each slit was infinitesimally wide, and so the cosine-squared fringes of Fig. 9.9 are really an unattainable idealization.* The actual pattern, Fig. 9.8e, drops off with distance on either side of O because of diffraction.

The Effects of Finite Coherence Length

As P in Fig. 9.8c is taken farther above or below the axis, $\overline{S_1B}$ (which is less than or equal to $\overline{S_1S_2}$) increases. If the primary source has a short coherence length, as the optical path difference increases, identically paired wavegroups will no longer be able to arrive at P exactly together. There will be an increasing amount of overlap in portions of uncorrelated wavegroups, and the contrast of the fringes will degrade. It is possible for Δl_c to be less than $\overline{S_1B}$. In that case, instead of two correlated portions of the same wavegroup arriving at P, only segments of different wavegroups will overlap, and the fringes will vanish.

As depicted in Fig. 9.10a, when the path length difference exceeds the coherence length, wavegroup-E_1 from source S_1 arrives at P with wavegroup-D_2 from S_2. There is interference, but it lasts only a short time before the pattern shifts as wavegroup-D_1 begins to overlap wavegroup-C_2, since the relative phases are different. If the coherence length was larger or the path difference smaller, wavegroup-D_1 would more or less interact with its clone wavegroup-D_2, and so on for each pair. The phases would then be correlated, and the interference pattern stable (Fig. 9.10b). Since a whitelight source will have a

*Modifications of this pattern arising as a result of increasing the width of either the primary S or secondary-source slits will be considered in later chapters (10 and 12). In the latter case, fringe contrast will be used as a measure of the degree of coherence (Section 12.1). In the latter, diffraction effects become significant.
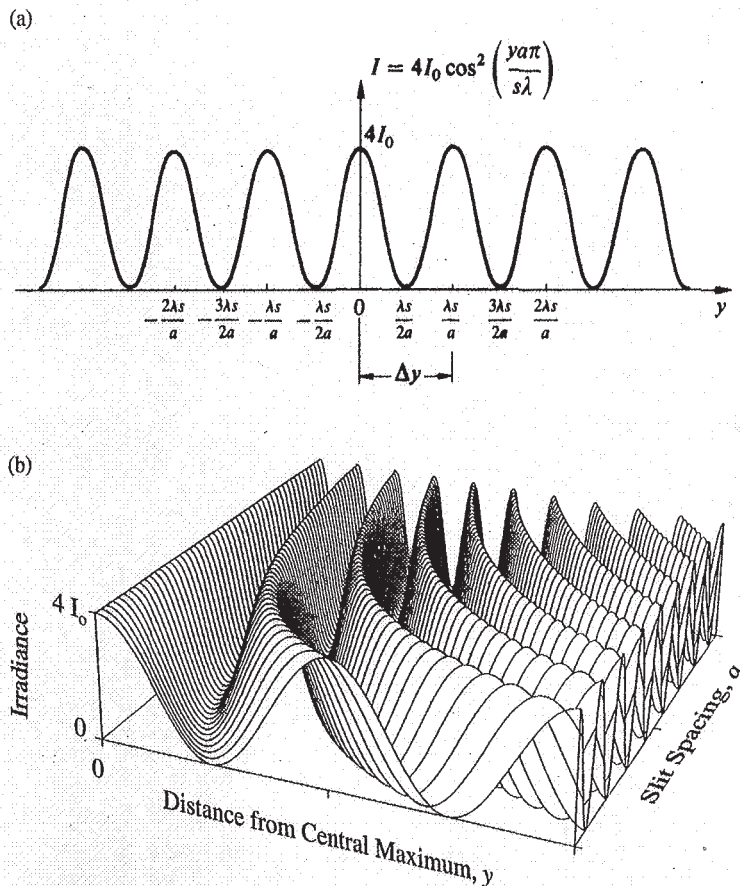
Figure 9.9 (a) Idealized irradiance versus distance curve. (b) The fringe separation Δy varies inversely with the slit separation, as one might expect from Fourier considerations; remember the inverse nature of spatial intervals and spatial frequency intervals. (Reprinted from "Graphical Representations of Fraunhofer Interference and Diffraction," *Am. J. Phys* **62**, 6 (1994), with permission of A.B. Bartlett, University of Colorado and B. Mechtly, Northeast Missouri State University and the American Association of Physics Teachers.)

coherence length of less than about three wavelengths, it follows from Eq. (9.27) that only about three fringes will be seen on either side of the central maximum.

Under white light (or with broad bandwidth illumination), all the constituent colors will arrive at $y = 0$ in-phase, having traveled equal distances from each aperture. The zeroth-order fringe will be essentially white, but all other higher order maxima will show a spread of wavelengths, since y_m is a function of λ, according to Eq. (9.27). Thus in white light we can think of the mth maximum as the mth-order band of wavelengths—a notion that will lead directly to the diffraction grating in the next chapter.

The fringe pattern can be observed by punching two small pinholes in a thin card. The holes should be approximately the size of the type symbol for a period on this page, and the separation between their centers about three radii. A street lamp,

car headlight, or traffic signal at night, located a few hundred feet away, will serve as a plane-wave source. The card should be positioned directly in front of and *very close to the eye*. The fringes will appear perpendicular to the line of centers. The pattern is much more readily seen with slits, as discussed in Section 10.2.2, but you should give the pinholes a try.

Microwaves, because of their long wavelength, also offer an easy way to observe double-slit interference. Two slits (e.g., $\lambda/2$ wide by λ long, separated by 2λ) cut in a piece of sheet metal or foil will serve quite well as secondary sources (Fig. 9.11).
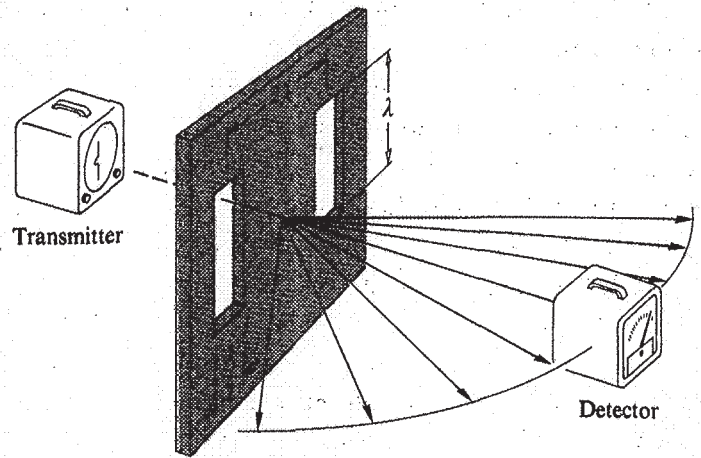
Figure 9.11 A microwave interferometer.

The Fourier Perspective

(a)

(b)
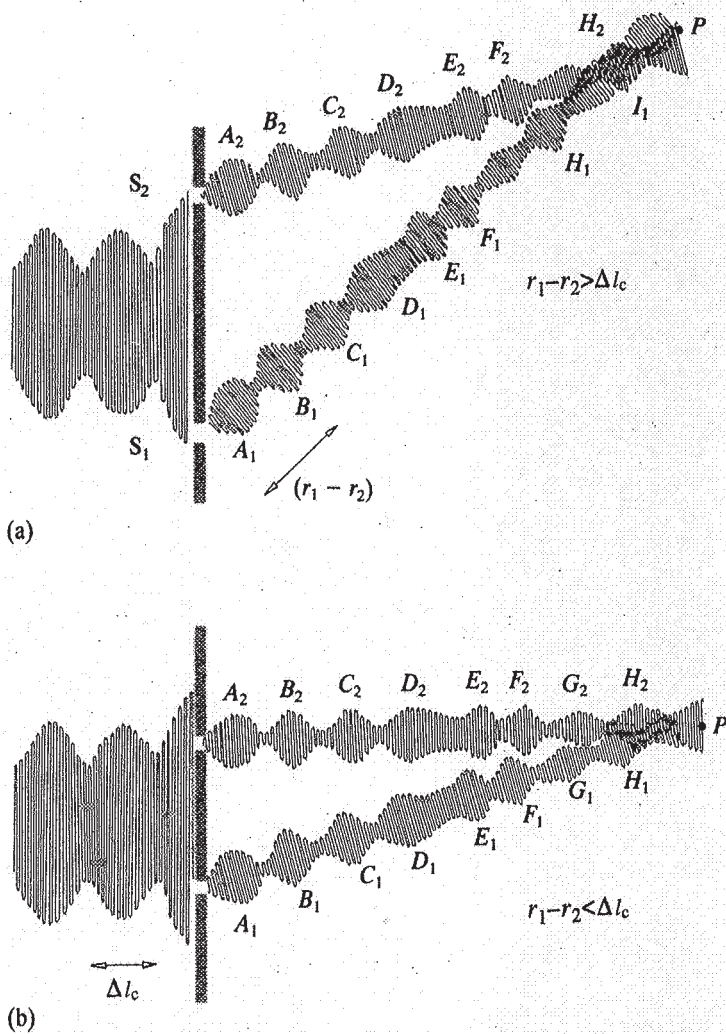
Figure 9.10 A schematic representation of how light, composed of a progression of wavegroups with a coherence length Δl_c, produces interference when (a) the path length difference exceeds Δl_c and (b) the path length difference is less than Δl_c.

When the plane waves in Fig. 9.8b illuminated the first narrow slit, light spilled out (i.e., diffracted) beyond the opaque screen in a form resembling a cylindrical wave; the narrower the slit, the more nearly cylindrical the wave. Beyond the screen the light spread over a very wide range of angles, or equivalently a wide range of spatial frequencies. From a Fourier perspective, this happened because an infinitesimally narrow source (i.e., narrow in space) generates a light field that is infinitely broad (i.e., broad in spatial frequency). The transform of a point source, an ideal one-dimensional signal spike (known as a Dirac delta function, p. 524), is a continuous constant spectrum containing all spatial frequencies, a spherical wave. In the same way, an ideal line source results in a cylindrical wave.

In practice, Young's Experiment usually consists of two in-phase slit sources arranged such that $s \gg a$. As a rule, s is so large that the resulting fringe system corresponds to a Fraunhofer diffraction pattern (p. 457). The two very thin slits resemble two line sources, two ideally narrow signal spikes, and the transform of two delta functions is a cosine function—we saw that in Fig. 7.42. To the extent that the slits can be considered infinitesimally narrow, the amplitude of the electric field in the diffraction pattern will be cosinusoidal, and the irradiance distribution will vary as the cosine squared, as in Fig. 9.9.

Several Other Interferometers

The same physical and mathematical considerations applied to Young's Experiment relate directly to a number of other wavefront-splitting interferometers. Most common among

these are Fresnel's double mirror, Fresnel's double prism, and Lloyd's mirror.

Fresnel's double mirror consists of two plane front-silvered mirrors inclined to each other at a very small angle, as shown in Fig. 9.12. One portion of the cylindrical wavefront coming from slit S is reflected from the first mirror, and another portion of the wavefront is reflected from the second mirror. An interference field exists in space in the region where the two reflected waves are superimposed. The images (S_1 and S_2) of the slit S in the two mirrors can be considered as separate coherent sources, placed at a distance a apart. It follows from the Laws of Reflection, as illustrated in Fig. 9.12a, that $\overline{SA} = \overline{S_1A}$ and $\overline{SB} = \overline{S_2B}$ so that $\overline{SA} + \overline{AP} = r_1$ and $\overline{SB} + \overline{BP} = r_2$. The optical path length difference between the two rays is then $r_1 - r_2$. The various maxima occur at $r_1 - r_2 = m\lambda$, as they do with Young's Interferometer. Again, the separation of the fringes is given by

$$\Delta y \approx \frac{s}{a}\lambda$$
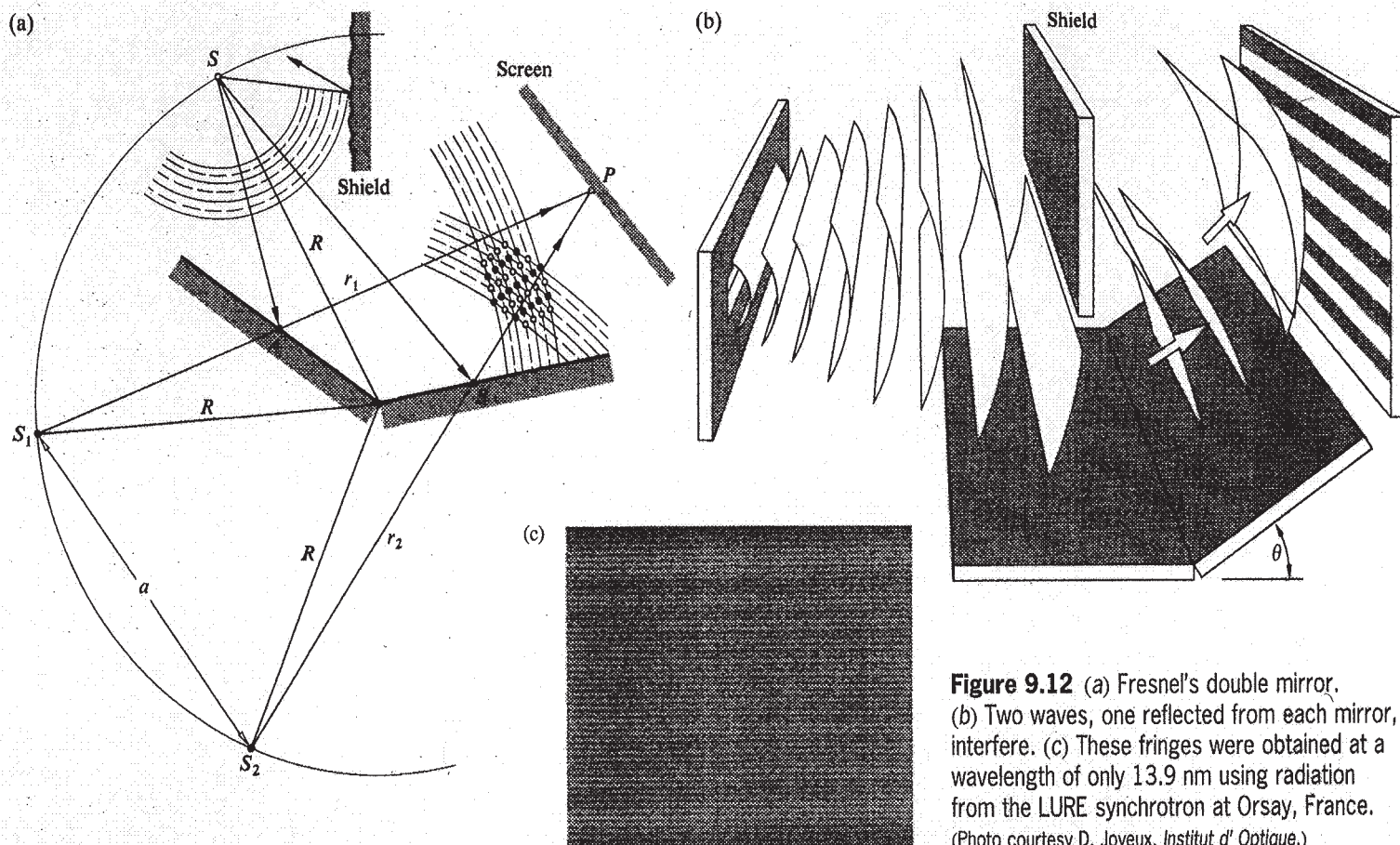
where s is the distance between the plane of the two virtual sources (S_1, S_2) and the screen. The arrangement in Fig. 9.12 has been deliberately exaggerated to make the geometry somewhat clearer. The angle θ between the mirrors must be quite small if the electric-field vectors for each of the two beams are to be parallel, or nearly so. Let $\vec{E}_1$ and $\vec{E}_2$ represent the light-waves emitted from the coherent virtual sources S_1 and S_2. At any instant in time at the point P in space, each of these vectors can be resolved into components, parallel and perpendicular to the plane of the figure. With $\vec{k}_1$ and $\vec{k}_2$ parallel to $\overline{AP}$ and $\overline{BP}$, respectively, it should be apparent that the components of $\vec{E}_1$ and $\vec{E}_2$ in the plane of the figure will approach being parallel only for small θ. As θ decreases, a decreases and the fringes broaden.

The **Fresnel double prism** or biprism consists of two thin prisms joined at their bases, as shown in Fig. 9.13. A single cylindrical wavefront impinges on both prisms. The top portion of the wavefront is refracted downward, and the lower segment is refracted upward. In the region of superposition, interference occurs. Here, again, two virtual sources S_1 and S_2

Figure 9.12 (a) Fresnel's double mirror. (b) Two waves, one reflected from each mirror, interfere. (c) These fringes were obtained at a wavelength of only 13.9 nm using radiation from the LURE synchrotron at Orsay, France. (Photo courtesy D. Joyeux, *Institut d' Optique*.)
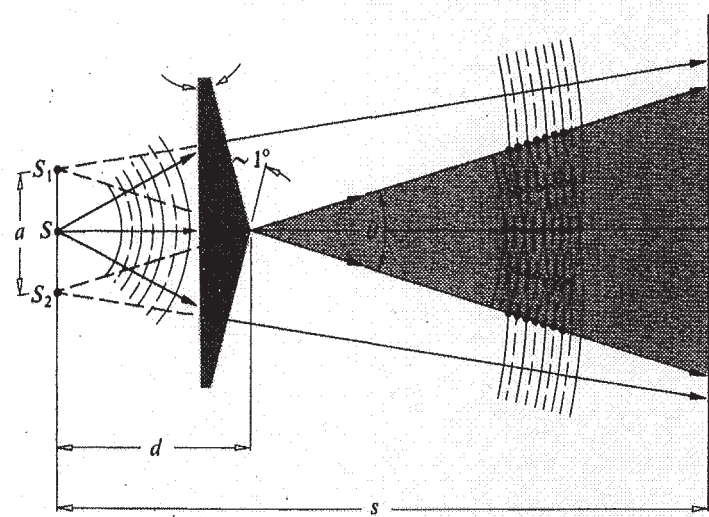
(a)
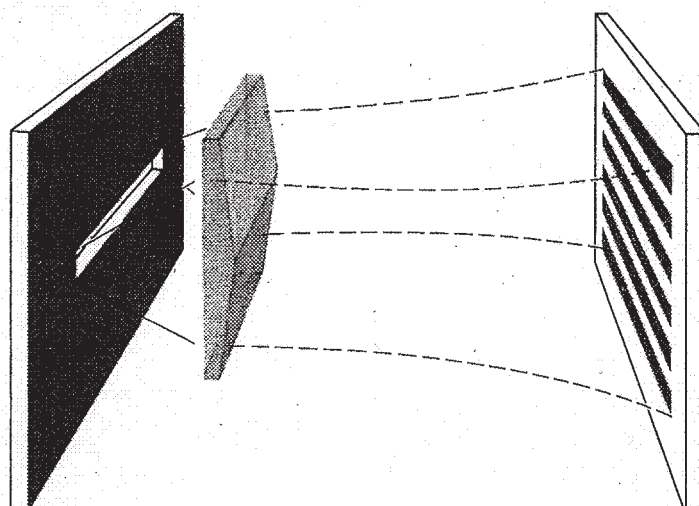
(b)
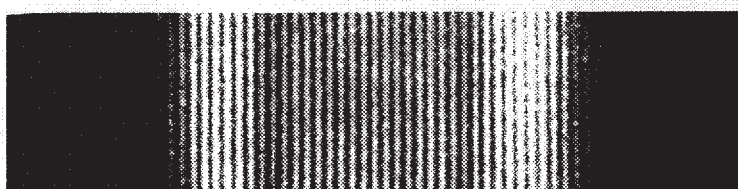
(c)

Figure 9.13 Fresnel's biprism. (*a*) The biprism creates two image sources. (*b*) With a slit source the fringes are bright bands. (*c*) Interference fringes observed with an electron biprism arrangement by G. Möllenstedt. Once again electrons behave like photons. (Photo from *Handbuch der Physik*, edited by S. Flügge, Springer-Verlag, Heidelberg.)

exist, separated by a distance *a*, which can be expressed in terms of the prism angle α (Problem 9.19), where $s \gg a$. The expression for the separation of the fringes is the same as before.

The last wavefront-splitting interferometer that we will consider is **Lloyd's mirror**, shown in Fig. 9.14. It consists of a flat piece of either dielectric or metal that serves as a mirror, from which is reflected a portion of the cylindrical wavefront coming from slit *S*. Another portion of the wavefront proceeds directly from the slit to the screen. For the separation *a*, between the two coherent sources, we take the distance between the actual slit and its image S_1 in the mirror. The spacing of the fringes is once again given by $(s/a)\lambda$. The distinguishing feature of this device is that at glancing incidence $(\theta_i = \pi/2)$ the reflected beam undergoes a 180° phase shift. (Recall that the amplitude-reflection coefficients are then both equal to −1.) With an additional phase shift of $\pm\pi$,

$$\delta = k(r_1 - r_2) \pm \pi$$

and the irradiance becomes

$$I = 4I_0 \sin^2\left(\frac{\pi a y}{s\lambda}\right)$$

The fringe pattern for Lloyd's mirror is complementary to that of Young's Interferometer; the maxima of one pattern exist at values of *y* that correspond to minima in the other pattern. The top edge of the mirror is equivalent to $y = 0$ and will be the center of a dark fringe rather than a bright one, as in Young's device. The lower half of the pattern will be obstructed by the presence of the mirror itself. Consider what
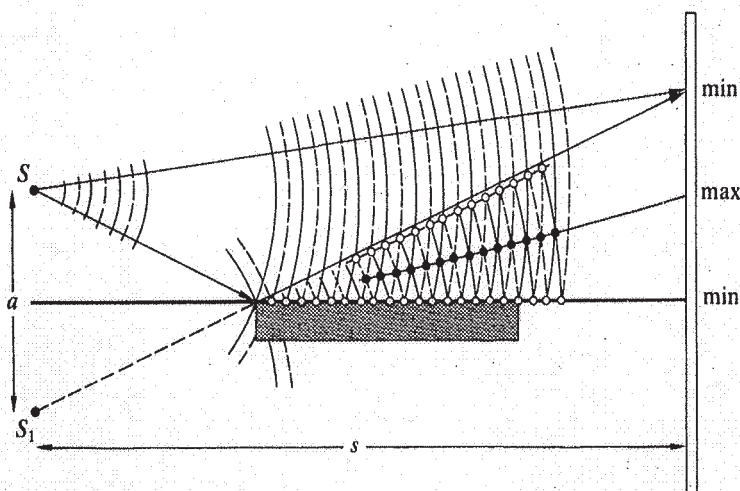
Figure 9.14 Lloyd's mirror.

would happen if a thin sheet of transparent material were placed in the path of the rays traveling directly to the screen. The transparent sheet would have the effect of increasing the number of wavelengths in each direct ray. The entire pattern would accordingly move upward, where the reflected rays would travel a bit farther before interfering. Because of the obvious inherent simplicity of this device, it has been used over a very wide region of the electromagnetic spectrum. The actual reflecting surfaces have ranged from crystals for X-rays, ordinary glass for light, and wire screening for microwaves to a lake or even the Earth's ionosphere for radiowaves.*

All the above interferometers can be demonstrated quite readily either using a laser or, for white light, something a bit more old-fashioned like a discharge lamp or a carbon arc (Fig. 9.15).

9.4 Amplitude-Splitting Interferometers

Suppose that a lightwave is incident on a half-silvered mirror,† or simply on a sheet of glass. Part of the wave is transmitted and part reflected. Both the transmitted and reflected waves have lower amplitudes than the original one. It can be said figuratively that the amplitude has been "split."

If the two separate waves could be brought together again at a detector, interference would result, as long as the original coherence between the two had not been destroyed. If the path lengths differed by a distance greater than that of the wavegroup (i.e., the coherence length), the portions reunited at the detector would correspond to different wavegroups. No unique phase relationship would exist between them in that case, and the fringe pattern would be unstable to the point of being unobservable. We will get back to these ideas when we consider coherence theory in more detail. For the

Figure 9.15 Bench setup to study wavefront-splitting arrangements with a carbon arc source. The water cell is needed to keep things cool.

moment the discussion is restricted, for the most part, to those cases for which the path difference is less than the coherence length.

9.4.1 Dielectric Films—Double-Beam Interference

Interference effects are observable in sheet transparent materials, the thicknesses of which vary over a very broad range, from films less than the length of a lightwave (e.g., for green light λ_0 equals about $\frac{1}{150}$ the thickness of this printed page) to plates several centimeters thick. A layer of material is referred to as a *thin film* for a given wavelength of electromagnetic radiation when its thickness is of the order of that wavelength. Before the early 1940s, interference phenomena associated with thin dielectric films, although well known, had fairly limited practical value. The rather spectacular color displays arising from oil slicks and soap films, however pleasing aesthetically and theoretically, were mainly curiosities.

With the advent of suitable vacuum deposition techniques in the 1930s, precisely controlled coatings could be produced on a commercial scale, and that, in turn, led to a rebirth of interest in dielectric films. During the Second World War, both sides were finding the enemy with a variety of coated optical devices, and by the 1960s multilayered coatings were in widespread use.

Fringes of Equal Inclination

---

*For a discussion of the effects of a finite slit width and a finite frequency bandwidth, see R. N. Wolfe and F. C. Eisen, "Irradiance Distribution in a Lloyd Mirror Interference Pattern," *J. Opt. Soc. Am.* **38**, 706 (1948).

---

†A *half-silvered mirror* is one that is semitransparent, because the metallic coating is too thin to be opaque. You can look through it, and at the same time you can see your reflection in it. *Beamsplitters*, as devices of this kind are called, can also be made of thin stretched plastic films, known as *pellicles*, or even uncoated glass plate.

Consider the simple case of a transparent parallel plate of dielectric material having a thickness d (Fig. 9.16). Suppose that the film is nonabsorbing and that the amplitude-reflection coefficients at the interfaces are so low that only the first two reflected beams E_{1r} and E_{2r} (both having undergone only one reflection) need be considered (Fig. 9.17). In practice, the amplitudes of the higher-order reflected beams (E_{3r}, etc.) gen-

Figure 9.17 Fringes of equal inclination.

erally decrease very rapidly, as can be shown for the air–water and air–glass interfaces (Problem 9.25). For the moment, consider S to be a monochromatic point source.

The film serves as an amplitude-splitting device, so that E_{1r} and E_{2r} may be considered as arising from two coherent virtual sources lying behind the film; that is, the two images of S formed by reflection at the first and second interfaces. The reflected rays are parallel on leaving the film and can be brought together at a point P on the focal plane of a telescope objective or on the retina of the eye when focused at infinity. From Fig. 9.17, the optical path length difference for the first two reflected beams is given by

$$\Lambda = n_f[(\overline{AB}) + (\overline{BC})] - n_1(\overline{AD})$$

and since $(\overline{AB}) = (\overline{BC}) = d/\cos \theta_t$,

$$\Lambda = \frac{2n_f d}{\cos \theta_t} - n_1(\overline{AD})$$

Now, to find an expression for $(\overline{AD})$, write

Figure 9.16 The wave and ray representations of thin-film interference. Light reflected from the top and bottom of the film interferes to create a fringe pattern.

$$(\overline{AD}) = (\overline{AC}) \sin \theta_i$$

Using Snell's Law, this becomes

$$(\overline{AD}) = (\overline{AC}) \frac{n_f}{n_1} \sin \theta_t$$

where

$$(\overline{AC}) = 2d \tan \theta_t \qquad (9.32)$$

The expression for Λ now becomes

$$\Lambda = \frac{2n_f d}{\cos \theta_t}(1 - \sin^2 \theta_t)$$

or finally

$$\Lambda = 2n_f d \cos \theta_t \qquad (9.33)$$

The corresponding phase difference associated with the optical path length difference is then just the product of the free-space propagation number and Λ, that is, $k_0 \Lambda$. If the film is immersed in a single medium, the index of refraction can simply be written as $n_1 = n_2 = n$. Realize that n may be less than n_f, as in the case of a soap film in air, or greater than n_f, as with an air film between two sheets of glass. In either case *there will be an additional phase shift arising from the reflections themselves*. Recall that for incident angles up to about 30°, regardless of the polarization of the incoming light, the two beams, one internally and one externally reflected, will experience a *relative phase shift* of π radians (Fig. 4.44 and Section 4.3). Accordingly,

$$\delta = k_0 \Lambda \pm \pi$$

and more explicitly

$$\delta = \frac{4\pi n_f}{\lambda_0} d \cos \theta_t \pm \pi \qquad (9.34)$$

or

$$\delta = \frac{4\pi d}{\lambda_0}(n_f^2 - n^2 \sin^2 \theta_i)^{1/2} \pm \pi \qquad (9.35)$$

The sign of the phase shift is immaterial, so we will choose the negative sign to make the equations a bit simpler. In reflected light an interference maximum, a bright spot, appears at P when $\delta = 2m\pi$—in other words, an even multiple of π. In that case Eq. (9.34) can be rearranged to yield

[maxima] $$d \cos \theta_t = (2m + 1)\frac{\lambda_f}{4} \qquad (9.36)$$

where $m = 0, 1, 2, \ldots$ and use has been made of the fact that $\lambda_f = \lambda_0/n_f$. This also corresponds to minima in the transmitted

light. Interference minima in reflected light (maxima in transmitted light) result when $\delta = (2m \pm 1)\pi$, that is, odd multiples of π. For such cases Eq. (9.34) yields

[minima] $$d \cos \theta_t = 2m\frac{\lambda_f}{4} \qquad (9.37)$$

The appearance of odd and even multiples of $\lambda_f/4$ in Eqs. (9.36) and (9.37) is significant, as we will see presently. We could, of course, have a situation in which $n_1 > n_f > n_2$ or $n_1 < n_f < n_2$, as with a fluoride film deposited on an optical element of glass immersed in air. The π phase shift would then not be present, and the above equations would simply be modified appropriately.

If the lens used to focus the rays has a small aperture, interference fringes will appear on a small portion of the film. Only the rays leaving the point source that are reflected directly into
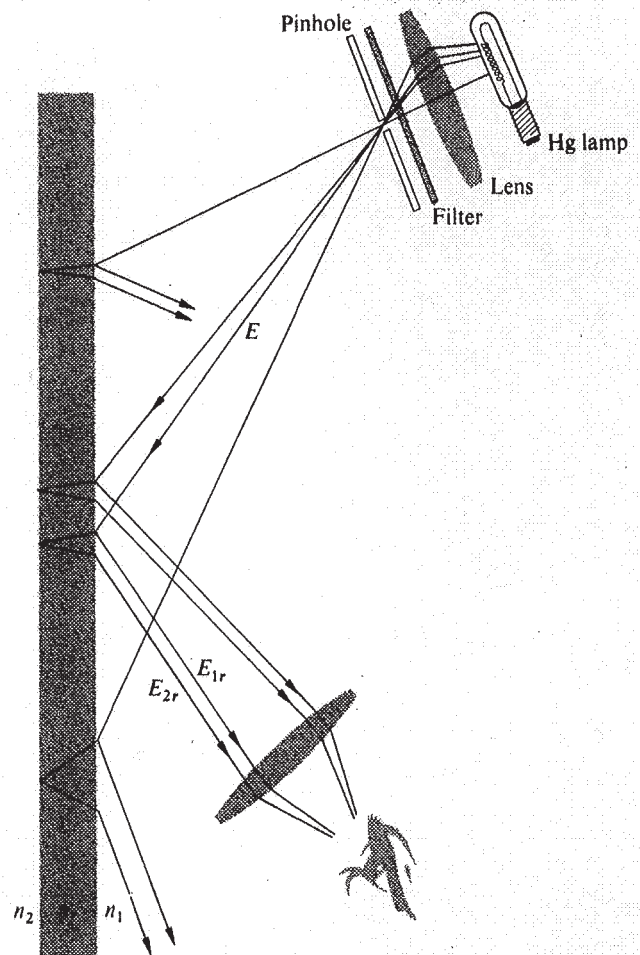
Figure 9.18 Fringes seen on a small portion of the film.
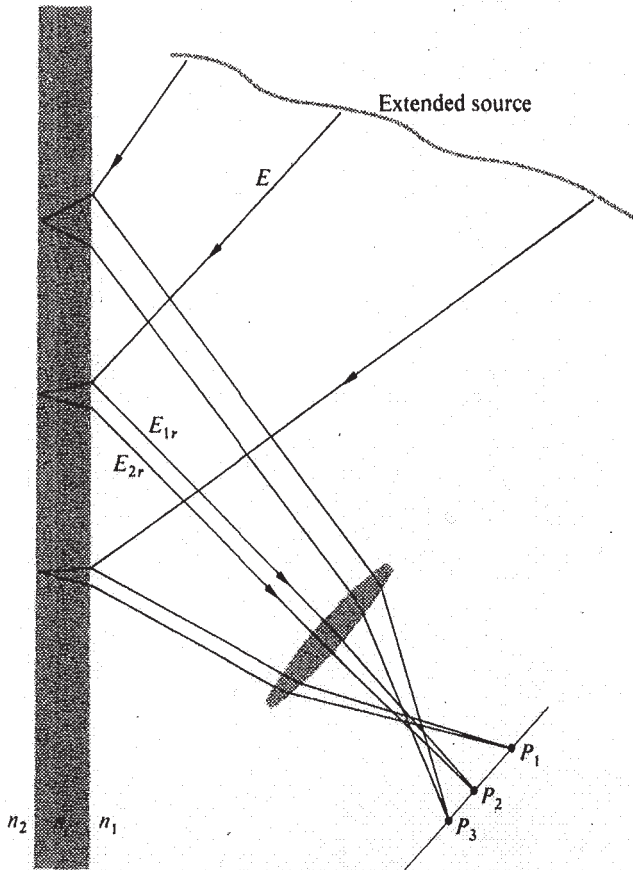
Figure 9.19 Fringes seen on a large region of the film.

the lens will be seen (Fig. 9.18). For an extended source, light will reach the lens from various directions, and the fringe pattern will spread out over a large area of the film (Fig. 9.19).

The angle θ_i or equivalently θ_t, determined by the position of P, will in turn control δ. The fringes appearing at points P_1 and P_2 in Fig. 9.20 are known as **fringes of equal inclination**. (Problem 9.30 discusses some easy ways to see these fringes.) Keep in mind that each source point on the extended source is incoherent with respect to the others. When the image of the extended source is reflected in the surface, it will be seen to be banded with bright and dark fringes. Each of these is an arc of a circle centered on the intersection of a perpendicular dropped from the eye to the film.

As the film becomes thicker, the separation $\overline{AC}$ between E_{1r} and E_{2r} also increases, since

$$\overline{AC} = 2d \tan \theta_t \qquad [9.32]$$

When only one of the two rays is able to enter the pupil of the

eye, the interference pattern will disappear. The larger lens of a telescope can then be used to gather in both rays, once again making the pattern visible. The separation can also be reduced by reducing θ_t and therefore θ_i, that is, by viewing the film at nearly normal incidence. The equal-inclination fringes that are seen in this manner for thick plates are known as **Haidinger fringes**, after the Austrian physicist Wilhelm Karl Haidinger (1795–1871). With an extended source, the symmetry of the setup requires that the interference pattern consists of a series of concentric circular bands centered on the perpendicular drawn from the eye to the film (Fig. 9.21). As the observer moves, the interference pattern follows along.

Haidinger fringes can be seen in the ordinary window glass of a store front. Find one with a neon sign in the window and look out at the street, at night, very close to the glowing tube. You'll see circular fringes centered on your eye floating off in the distance.
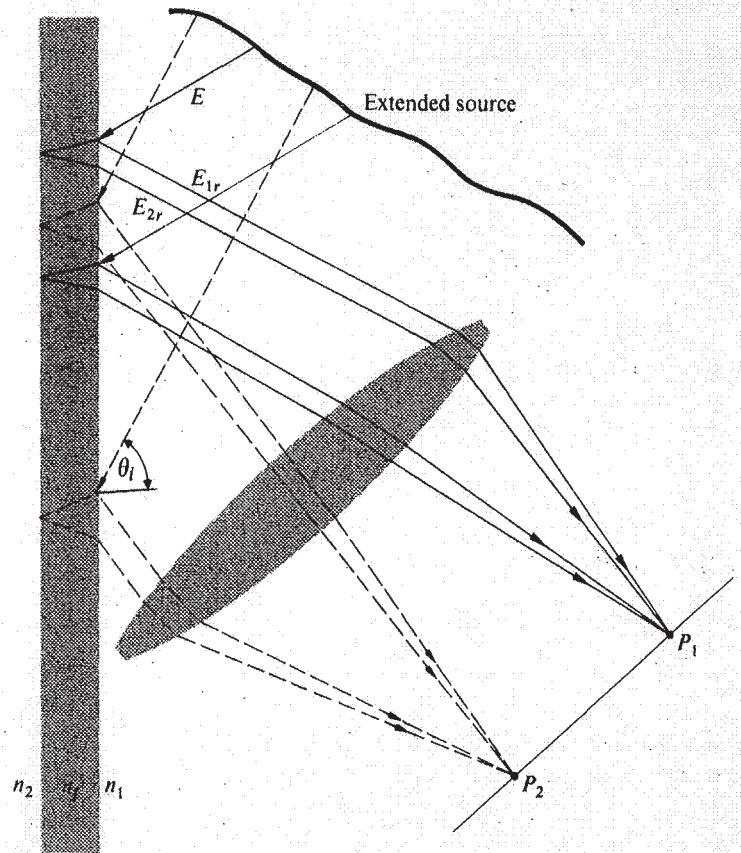
Figure 9.20 All rays inclined at the same angle arrive at the same point.
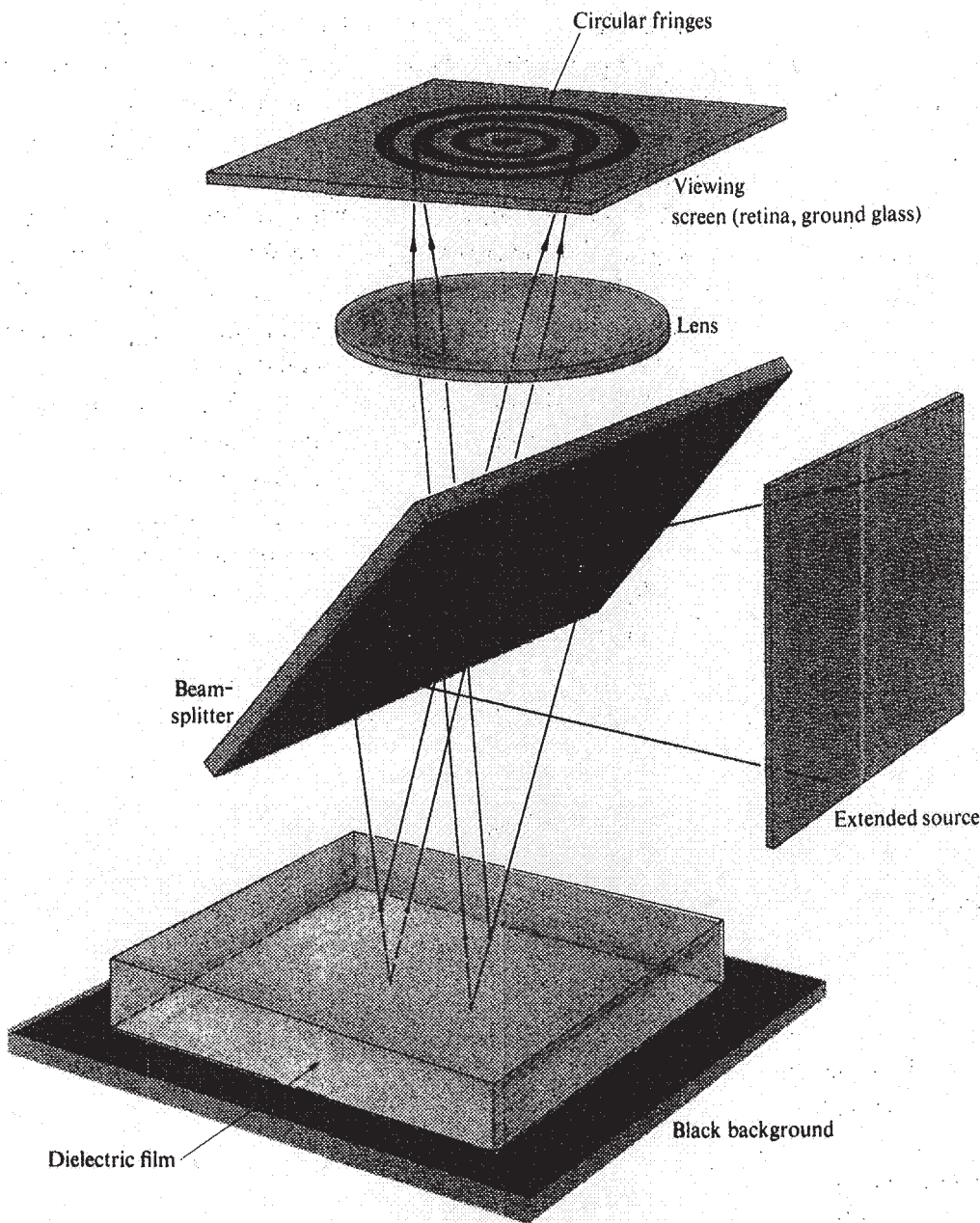
Figure 9.21 Circular Haidinger fringes centered on the lens axis.

Fringes of Equal Thickness

A whole class of interference fringes exists for which the optical thickness, $n_f d$, is the dominant parameter rather than θ_i. These are referred to as **fringes of equal thickness**. Under white-light illumination the iridescence of soap bubbles, oil slicks (a few wavelengths thick), and even oxidized metal surfaces is the result of variations in film thickness. Interference bands of this kind are analogous to the constant-height contour lines of a topographical map. Each fringe is the locus of all points in the film for which the optical thickness is a constant. In general, n_f does not vary, so that the fringes correspond to regions of constant film thickness. As such, they can be quite useful in determining the surface features of optical elements (lenses, prisms, etc.). For example, a surface to be examined may be put into contact with an *optical flat*.* The air in the space between the two generates a thin-film interference pat-

---

*A surface is said to be optically flat when it deviates by not more than about $\lambda/4$ from a perfect plane. In the past, the best flats were made of clear fused quartz. Now glass-ceramic materials (e.g., CERVIT) having extremely small thermal coefficients of expansion (about one-sixth that of quartz) are available. Individual flats of $\lambda/200$ or a bit better can be made.

tern. If the test surface is flat, a series of straight, equally spaced bands indicates a wedge-shaped air film, usually resulting from dust between the flats. Two pieces of plate glass separated at one end by a strip of paper will form a satisfactory wedge with which to observe these bands.

When viewed at nearly normal incidence in the manner illustrated in Fig. 9.22, the contours arising from a nonuniform film are called **Fizeau fringes**. For a thin wedge of small angle α, the optical path length difference between two reflected rays may be approximated by Eq. (9.33), where d is the thickness at a particular point, that is,

$$d = x\alpha \qquad (9.38)$$

For small values of θ_i the condition for an interference maximum becomes

$$(m + \tfrac{1}{2})\lambda_0 = 2n_f d_m$$

or

$$(m + \tfrac{1}{2})\lambda_0 = 2\alpha x_m n_f$$

Since $n_f = \lambda_0/\lambda_f$, x_m may be written as

$$x_m = \left(\frac{m + 1/2}{2\alpha}\right)\lambda_f \qquad (9.39)$$

Maxima occur at distances from the apex given by $\lambda_f/4\alpha$, $3\lambda_f/4\alpha$, and so on, and consecutive fringes are separated by a distance Δx, given by

$$\Delta x = \lambda_f/2\alpha \qquad (9.40)$$

Notice that the difference in film thickness between adjacent maxima is simply $\lambda_f/2$. Since the beam reflected from the lower surface traverses the film twice ($\theta_i \approx \theta_t \approx 0$), adjacent maxima differ in optical path length by λ_f. Note, too, that the film thickness at the various maxima is given by

$$d_m = (m + \tfrac{1}{2})\frac{\lambda_f}{2} \qquad (9.41)$$

which is an odd multiple of a quarter wavelength. Traversing the film twice yields a phase shift of π, which, when added to the shift of π resulting from reflection, puts the two rays back in-phase.

The accompanying photo shows a soap film held vertically so that it settles into a wedge shape under the influence of gravity. When illuminated with white light, the bands are various colors. The black region at the top is a portion where the film is less than $\lambda_f/4$ thick. Twice this, plus an additional shift of $\lambda_f/2$ due to the reflection, is less than a whole wavelength. The reflected rays are therefore out-of-phase. As the thickness decreases still further, the total phase difference approaches π. The irradiance at the observer goes to a mini-

Figure 9.22 Fringes from a wedge-shaped film.
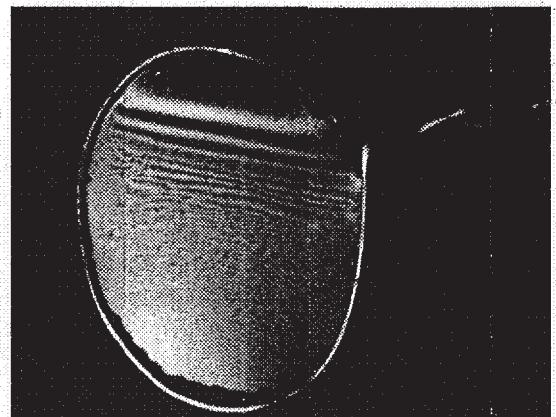
A wedge-shaped film made of liquid dishwashing soap. (Photo by E. H.)
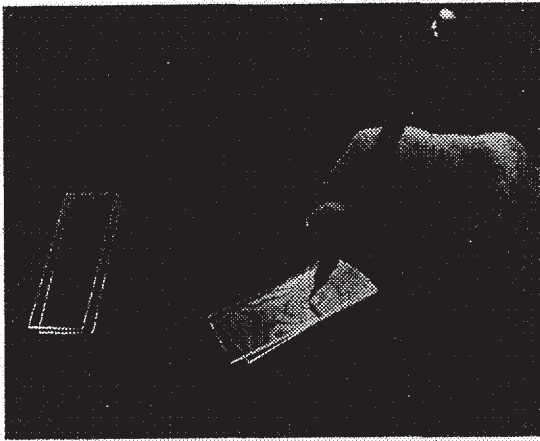
Fringes created by an air film between two microscope slides.
(Photo by E. H.)

mum (Eq. 9.16), and the film appears black in reflected light.*

Press two well-cleaned microscope slides together. The enclosed air film will usually not be uniform. In ordinary room light a series of irregular, colored bands (fringes of equal thickness) will be clearly visible across the surface. The thin glass slides distort under pressure, and the fringes move and change accordingly. Tape two slides together with transparent (matt-surfaced) tape. It will scatter light and make the reflected fringes more easily seen.

If the two pieces of glass are forced together at a point, as might be done by pressing on them with a sharp pencil, a
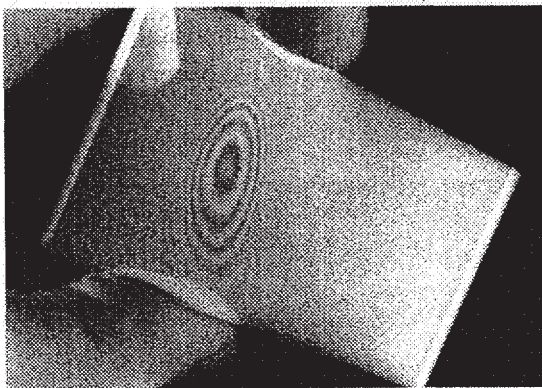
Newton's rings with two microscope slides. The thin film of air between the slides creates the interference pattern. (Photo by E. H.)

*The relative phase shift of π between internal and external reflection is required if the reflected flux density is to go to zero smoothly, as the film gets thinner and finally disappears.

series of concentric, nearly circular, fringes is formed about that point. Known as **Newton's rings***, this pattern is more precisely examined with the arrangement of Fig. 9.23. Here a lens is placed on an optical flat and illuminated at normal incidence with quasimonochromatic light. The amount of uniformity in the concentric circular pattern is a measure of the degree of perfection in the shape of the lens. With R as the radius of curvature of the convex lens, the relation between the distance x and the film thickness d is given by

$$x^2 = R^2 - (R - d)^2$$
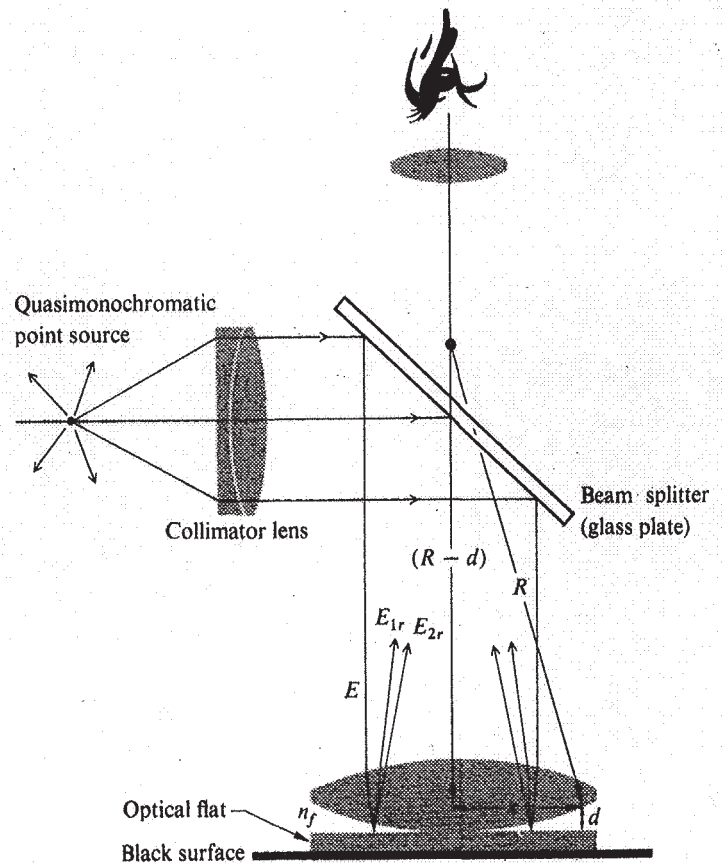
Figure 9.23 A standard setup to observe Newton's rings

*Robert Hooke (1635–1703) and Isaac Newton independently studied a whole range of thin-film phenomena, from soap bubbles to the air film between lenses. Quoting from Newton's *Opticks*:

> I took two Object-glasses, the one a Planoconvex for a fourteen Foot Telescope, and the other a large double Convex for one of about fifty Foot; and upon this, laying the other with its plane side downwards, I pressed them slowly together to make the Colours successfully emerge in the middle of the Circles.
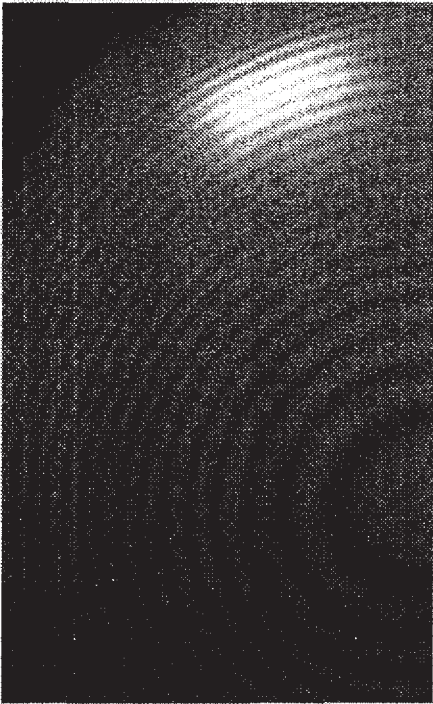
Interference from the thin air film between a convex lens and the flat sheet of glass it rests on. The illumination was quasimonochromatic. These fringes were first studied in depth by Newton and are known as Newton's rings. (Photo by E.H.)

or more simply by

$$x^2 = 2Rd - d^2$$

Since $R >> d$, this becomes

$$x^2 = 2Rd$$

Assume that we need only examine the first two reflected beams E_{1r} and E_{2r}. The mth-order interference maximum will occur in the thin film when its thickness is in accord with the relationship

$$2n_f d_m = (m + \tfrac{1}{2})\lambda_0$$

The radius of the mth bright ring is therefore found by combining the last two expressions to yield

[bright ring] $$x_m = [(m + \tfrac{1}{2})\lambda_f R]^{1/2} \qquad (9.42)$$

Similarly, the radius of the mth dark ring is

[dark ring] $$x_m = (m\lambda_f R)^{1/2} \qquad (9.43)$$

If the two pieces of glass are in good contact (no dust), the central fringe at that point ($x_0 = 0$) will clearly be a minimum in irradiance, an understandable result since d goes to zero at that point. In transmitted light, the observed pattern will be the complement of the reflected one discussed above, so that the center will now appear bright.

Newton's rings, which are Fizeau fringes, can be distinguished from the circular pattern of Haidinger's fringes by the manner in which the diameters of the rings vary with the order m. The central region in the Haidinger pattern corresponds to the maximum value of m (Problem 9.29), whereas just the opposite applies to Newton's rings.

An optical shop, in the business of making lenses, will have a set of precision spherical test plates or gauges. A designer can specify the surface accuracy of a new lens in terms of the number and regularity of the Newton rings that will be seen with a particular test gauge. The use of test plates in the manufacture of high-quality lenses, however, is giving way to far more sophisticated techniques involving laser interferometers (Section 9.8.2).

9.4.2 Mirrored Interferometers

There are a good number of amplitude-splitting interferometers that utilize arrangements of mirrors and beamsplitters. By far the best known and historically the most important of these is the **Michelson Interferometer**. Its configuration is illustrated in Fig. 9.24. An extended source (e.g., a diffusing ground-glass plate illuminated by a discharge lamp) emits a wave, part of which travels to the right. The beamsplitter at O divides the wave into two, one segment traveling to the right and one up into the background. The two waves are reflected by mirrors M_1 and M_2 and return to the beamsplitter. Part of the wave coming from M_2 passes through the beamsplitter going downward, and part of the wave coming from M_1 is deflected by the beamsplitter toward the detector. The two waves are united, and interference can be expected.

Notice that one beam passes through O three times, whereas the other traverses it only once. Consequently, each beam will pass through equal thicknesses of glass only when a *compensator plate C* is inserted in the arm OM_1. The compensator is an exact duplicate of the beamsplitter, with the exception of any possible silvering or thin film coating on the beamsplitter. It is positioned at an angle of 45°, so that O and C are parallel to each other. With the compensator in place, any optical path difference arises from the actual path difference. In addition,

because of the dispersion of the beamsplitter, the optical path is a function of λ. Accordingly, for quantitative work, the interferometer without the compensator plate can be used only with a quasimonochromatic source. The inclusion of a compensator negates the effect of dispersion, so that even a source with a very broad bandwidth will generate discernible fringes.

To understand how fringes are formed, refer to the construction shown in Fig. 9.25, where the physical components are represented more as mathematical surfaces. An observer at the position of the detector will simultaneously see both mirrors M_1 and M_2 along with the source Σ in the beamsplitter. We can redraw the interferometer as if all the elements were in a straight line. Here M_1' corresponds to the image of mirror M_1 in the beamsplitter, and Σ has been swung over in line with O and M_2. The positions of these elements in the diagram depend on their relative distances from O (e.g., M_1' can be in front of, behind, or coincident with M_2 and can even pass through it). The surfaces Σ_1 and Σ_2 are the images of the source Σ in mirrors M_1 and M_2, respectively. Now consider a single point S on the source emitting light in all directions; let's follow the course of one emerging ray. In actuality a wave from S will be split at O, and its segments will thereafter be reflected by M_1 and M_2. In our schematic diagram we represent this by reflecting the ray off both M_2 and M_1'. To an observer at D, the two reflected rays will appear to have come from the image points S_1 and S_2. [Note that all rays shown in (a) and (b) of Fig. 9.25 share a common plane-of-incidence.] For all practical purposes, S_1 and S_2 are coherent point sources, and we can anticipate a flux-density distribution obeying Eq. (9.14).

As the figure shows, the optical path difference for these rays is nearly $2d \cos \theta$, which represents a phase difference of $k_0 2d \cos \theta$. There is an additional phase term arising from the fact that the wave traversing the arm OM_2 is internally reflected in the beamsplitter, whereas the OM_1-wave is externally reflected at O. If the beamsplitter is simply an uncoated glass plate, the relative phase shift resulting from the two reflections will be π radians. *Destructive*, rather than constructive, inter-
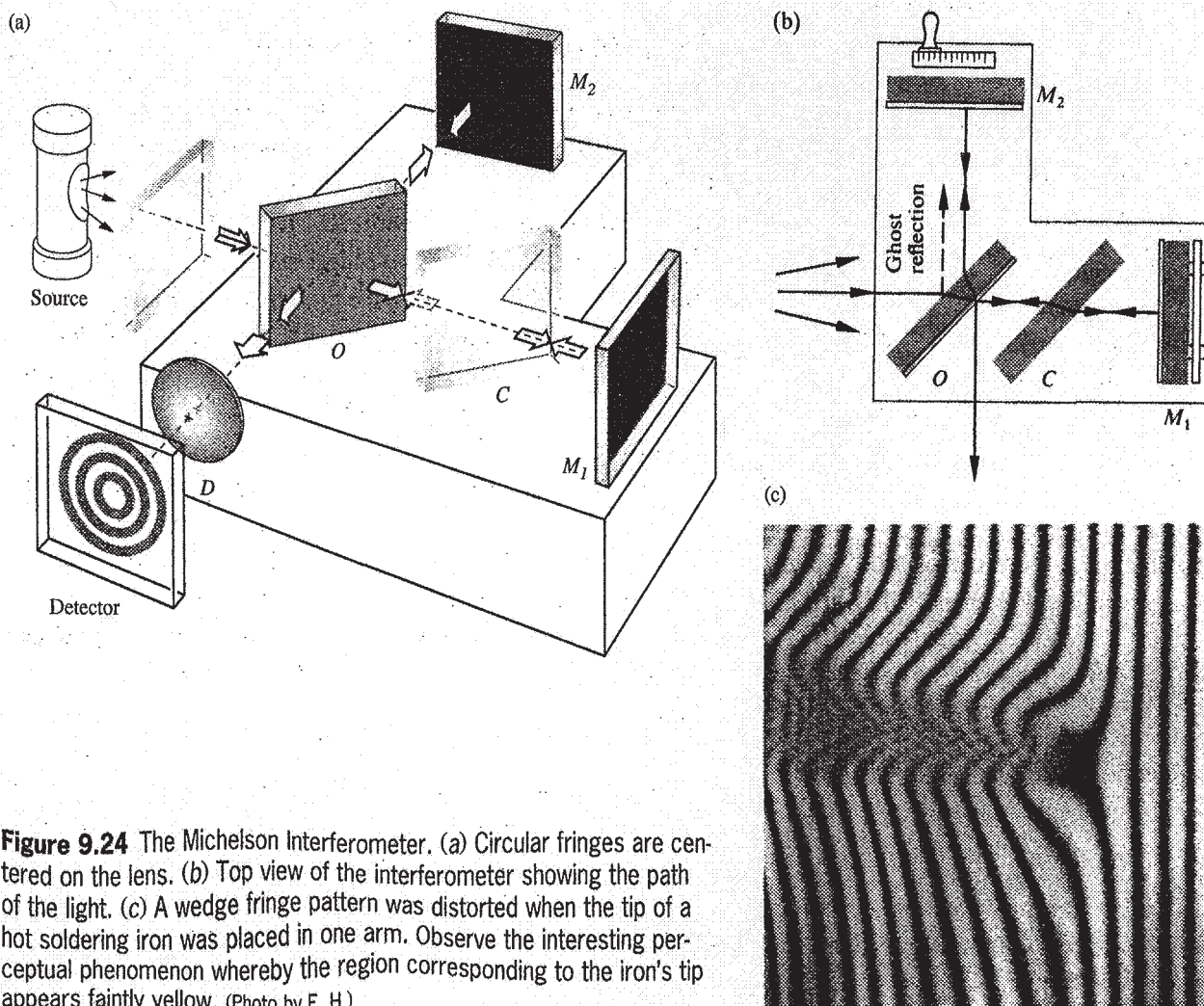
Figure 9.24 The Michelson Interferometer. (a) Circular fringes are centered on the lens. (b) Top view of the interferometer showing the path of the light. (c) A wedge fringe pattern was distorted when the tip of a hot soldering iron was placed in one arm. Observe the interesting perceptual phenomenon whereby the region corresponding to the iron's tip appears faintly yellow. (Photo by E. H.)

(a)

(b)
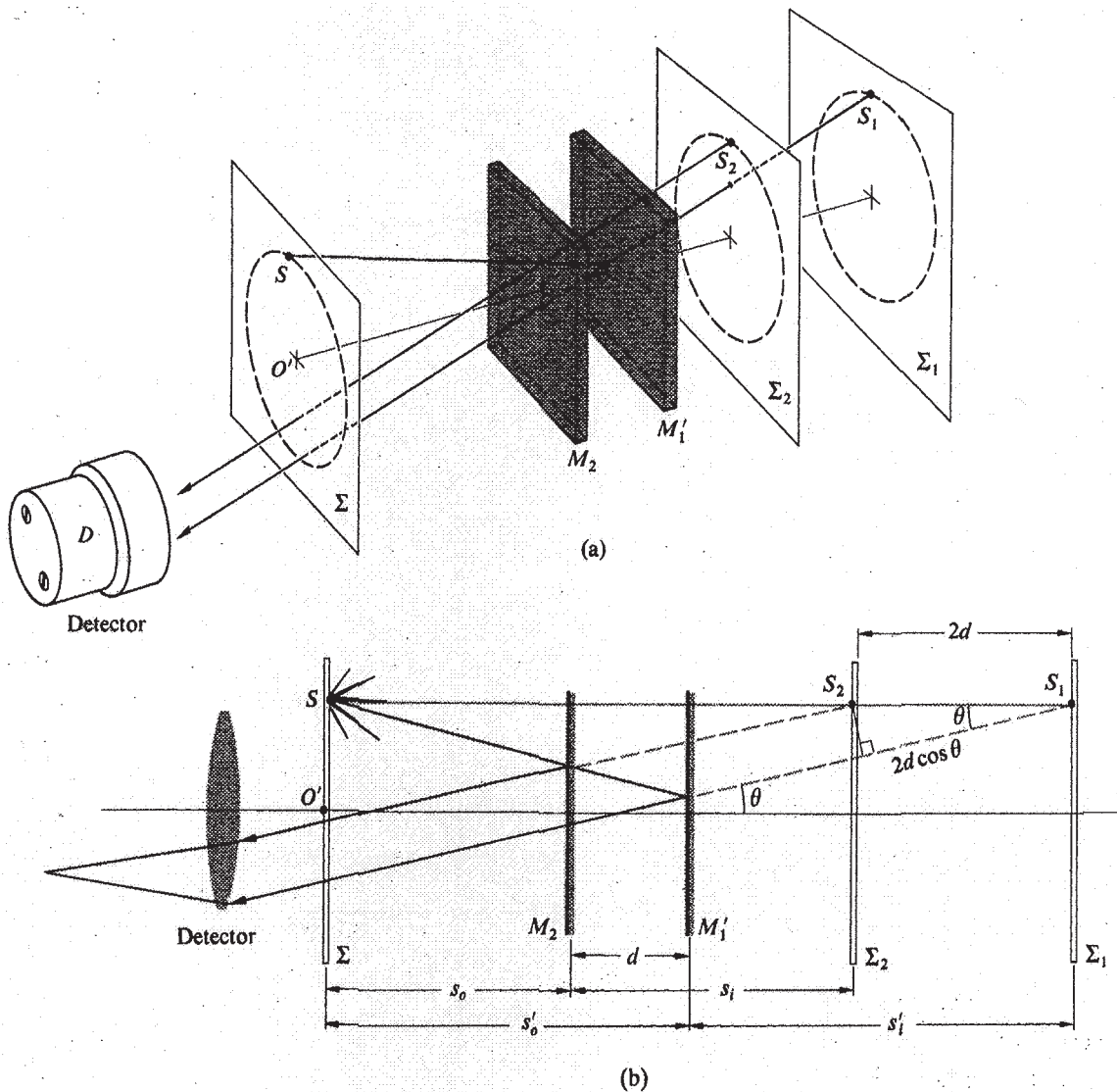
Figure 9.25 A conceptual rearrangement of the Michelson Interferometer.

ference will then exist when

$$2d \cos \theta_m = m\lambda_0 \qquad (9.44)$$

where m is an integer. If this condition is fulfilled for the point S, then it will be equally well fulfilled for any point on Σ that lies on the circle of radius $O'S$, where O' is located on the axis of the detector. As illustrated in Fig. 9.26, an observer will see a circular fringe system concentric with the central axis of her eye's lens. Because of the small aperture of the eye, the observer will not be able to see the entire pattern without the use of a large lens near the beamsplitter to collect most of the emergent light.

If we use a source containing a number of frequency components (e.g., a mercury discharge lamp), the dependence of θ_m on λ_0 in Eq. (9.44) requires that each such component generate a fringe system of its own. Note, too, that since $2d \cos \theta_m$ must be less than the coherence length of the source,

it follows that laser light will be particularly easy to use in demonstrating the interferometer (see Section 9.5). This point would be made strikingly evident were we to compare the fringes produced by laser light with those generated by "white" light from an ordinary tungsten bulb or a candle. In the latter case, the path difference must be very nearly zero, if we are to see any fringes at all, whereas in the former instance a difference of 10 cm has little noticeable effect.

An interference pattern in quasimonochromatic light typically consists of a large number of alternatively bright and dark rings. A particular ring corresponds to a fixed *order m*. As M_2 is moved toward M_1', d decreases, and according to Eq. (9.44), $\cos \theta_m$ increases while θ_m therefore decreases. The rings shrink toward the center, with the highest-order one disappearing whenever d decreases by $\lambda_0/2$. Each remaining ring broadens as more and more fringes vanish at the center, until only a few fill the whole screen. By the time $d = 0$ has been reached, the central fringe will have spread out, filling the
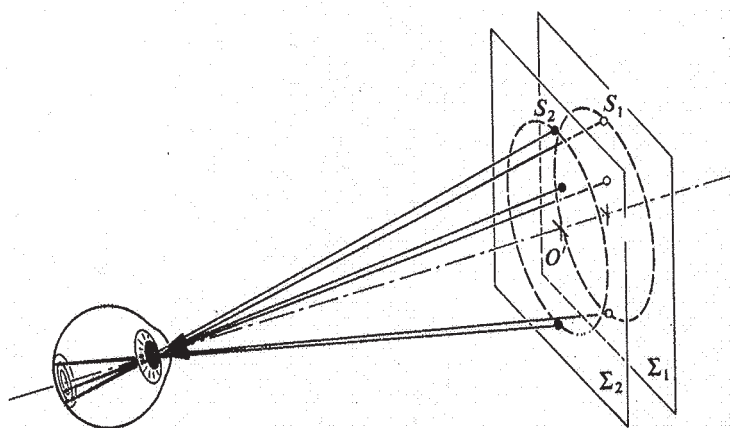
Figure 9.26 Formation of circular fringes.

entire field of view. With a phase shift of π resulting from reflection off the beamsplitter, the whole screen will then be an interference minimum. (Lack of perfection in the optical elements can render this unobservable.) Moving M_2 still farther causes the fringes to reappear at the center and move outward.

Notice that a central dark fringe for which $\theta_m = 0$ in Eq. (9.44) can be represented by

$$2d = m_0\lambda_0 \tag{9.45}$$

(Keep in mind that this is a special case. The central region might correspond to neither a maximum nor a minimum.) Even if d is 10 cm, which is fairly modest in laser light, and $\lambda_0 = 500$ nm, m_0 will be quite large, namely, 400 000. At a fixed value of d, successive dark rings will satisfy the expressions

$$2d \cos \theta_1 = (m_0 - 1)\lambda_0$$
$$2d \cos \theta_2 = (m_0 - 2)\lambda_0$$
$$\vdots$$
$$2d \cos \theta_p = (m_0 - p)\lambda_0 \tag{9.46}$$

The angular position of any ring, for example, the pth ring, is determined by combining Eqs. (9.45) and (9.46) to yield

$$2d(1 - \cos \theta_p) = p\lambda_0 \tag{9.47}$$

Since $\theta_m \equiv \theta_p$, both are just the half-angle subtended at the detector by the particular ring, and since $m = m_0 - p$, Eq. (9.47) is equivalent to Eq. (9.44). The new form is somewhat

more convenient, since (using the same example as above) with $d = 10$ cm, the sixth dark ring can be specified by stating that $p = 6$, or in terms of the *order* of the pth ring, that $m = 399\,994$. If θ_p is small,

$$\cos \theta_p = 1 - \frac{\theta_p^2}{2}$$

and Eq. (9.47) yields

$$\theta_p = \left(\frac{p\lambda_0}{d}\right)^{1/2} \tag{9.48}$$

for the angular radius of the pth fringe.

The construction of Fig. 9.25 represents one possible configuration, the one in which we consider only pairs of parallel emerging rays. Since these rays do not actually meet, they cannot form an image without a condensing lens of some sort. Indeed, that lens is most often provided by the observer's eye focused at infinity. The resulting *fringes of equal inclination* ($\theta_m = $ constant) located at infinity are also *Haidinger fringes*. A comparison of Figs. 9.25b and 9.3a, both showing two coherent point sources, suggests that in addition to these (virtual) fringes at infinity, there might also be (real) fringes formed by converging rays. These fringes do in fact exist. Hence, if you illuminate the interferometer with a *broad source* and shield out all extraneous light, you can easily see the projected pattern on a screen in a darkened room (see Section 9.5). The fringes will appear in the space in front of the interferometer (i.e., where the detector is shown), and their size will increase with increasing distance from the beamsplitter. We will consider the (real) fringes arising from point-source illumination a little later on.

When the mirrors of the interferometer are inclined with respect to each other, making a small angle (i.e., when M_1 and M_2 are not quite perpendicular), *Fizeau fringes* are observed. The resultant wedge-shaped air film between M_2 and M_1' creates a pattern of straight parallel fringes. The interfering rays appear to diverge from a point behind the mirrors. The eye would have to focus on this point in order to make these *localized fringes* observable. It can be shown analytically* that by appropriate adjustment of the orientation of the mirrors M_1 and M_2, fringes can be produced that are straight, circular, elliptical, parabolic, or hyperbolic—this holds as well for the real and virtual fringes.

....

*See, for example, Valasek, *Optics*, p.135.

The Michelson Interferometer can be used to make extremely accurate length measurements. As the moveable mirror is displaced by $\lambda_0/2$, each fringe will move to the position previously occupied by an adjacent fringe. Using a microscope arrangement, one need only count the number of fringes N, or portions thereof, that have moved past a reference point to determine the distance traveled by the mirror Δd, that is,

$$\Delta d = N(\lambda_0/2)$$

Nowadays this can be done fairly easily by electronic means. Michelson used the method to measure the number of wavelengths of the red cadmium line corresponding to the standard meter in Sévres near Paris.*

The Michelson Interferometer can be used along with a few polaroid filters to verify the Fresnel–Arago Laws. A polarizer inserted in each arm will allow the optical path length difference to remain fairly constant, while the vector field directions of the two beams are easily changed.

A microwave Michelson Interferometer can be constructed with sheet-metal mirrors and a chicken-wire beamsplitter. With the detector located at the central fringe, it can easily measure shifts from maxima to minima as one of the mirrors is moved, thereby determining λ. A few sheets of plywood, plastic, or glass inserted in one arm will change the central fringe. Counting the number of fringe shifts yields a value for the index of refraction, and from that we can compute the dielectric constant of the material.

The **Mach–Zehnder Interferometer** is another amplitude-splitting device. As shown in Fig. 9.27, it consists of two beamsplitters and two totally reflecting mirrors. The two waves within the apparatus travel along separate paths. A difference between the optical paths can be introduced by a slight tilt of one of the beamsplitters. Since the two paths are separated, the interferometer is relatively difficult to align. For the same reason, however, the interferometer finds myriad applications. It has even been used, in a somewhat altered yet conceptually similar form, to obtain electron interference fringes.†

Figure 9.27 The Mach–Zehnder Interferometer.

An object interposed in one beam will alter the optical path length difference, thereby changing the fringe pattern. A common application of the device is to observe the density variations in gas-flow patterns within research chambers (wind tunnels, shock tubes, etc.). One beam passes through the optically flat windows of the test chamber, while the other beam traverses appropriate compensator plates. The beam within the chamber will propagate through regions having a spatially
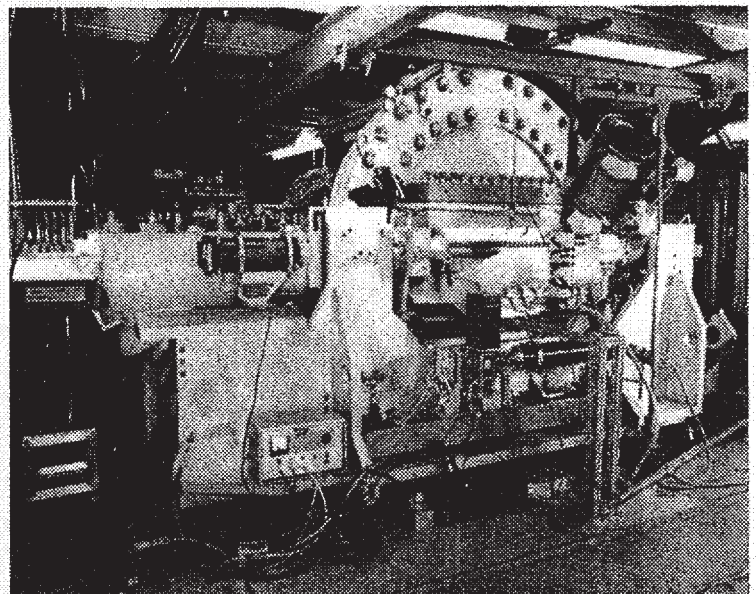
Scylla IV, an early setup for studying plasma. (Courtesy of University of California, Lawrence Livermore National Laboratory, and the Department of Energy.)

*A discussion of the procedure he used to avoid counting the 3 106 327 fringes directly can be found in Strong, *Concepts of Classical Optics*, p. 238, or Williams, *Applications of Interferometry*, p. 51.

†L. Marton, J. Arol Simpson, and J. A. Suddeth, *Rev. Sci. Instr.* **25**, 1099 (1954), and *Phys. Rev.* **90**, 490 (1953).
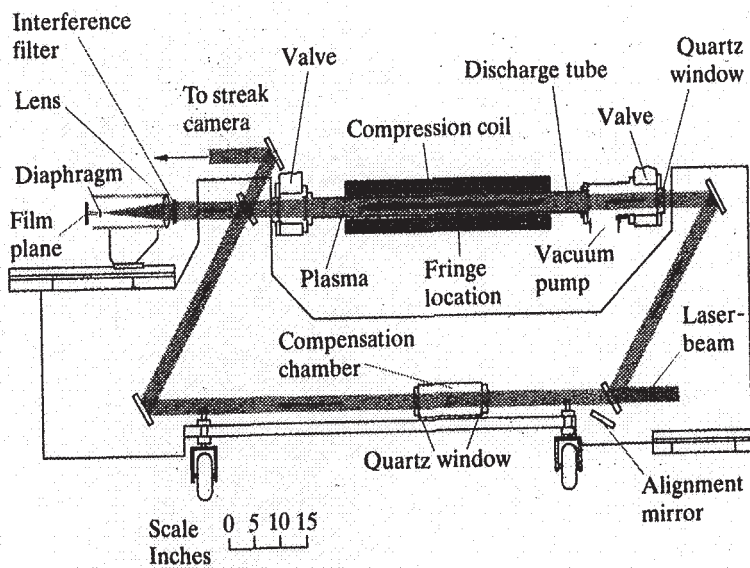
Figure 9.28 Schematic of Scylla IV.

Interferogram with plasma. (Photo courtesy Los Alamos National Laboratory.)

varying index of refraction. The resulting distortions in the wavefront generate the fringe contours. A particularly nice application is shown in Fig. 9.28, which depicts the magnetic compression device known as Scylla IV. It was used to study controlled thermonuclear reactions at the Los Alamos Scientific Laboratory. In this case, the Mach–Zehnder Interferometer appears in the form of a parallelogram. The two ruby laser *interferograms*, as these photographs are called, show the background pattern without a plasma in the tube and the density contours within the plasma during a reaction.

Another amplitude-splitting device, which differs from the previous instrument in many respects, is the **Sagnac Interferometer**. It is very easy to align and quite stable. An interesting application of the device is discussed in the last section of this chapter, where we consider its use as a gyroscope. One form of the Sagnac Interferometer is shown in Fig. 9.29*a* and another in Fig. 9.29*b*; still others are possible. Notice that the main feature of the device is that there are two identical but oppositely directed paths taken by the beams and that both form closed loops before they are united to produce interference. A deliberate slight shift in the orientation of one of the mirrors will produce a path length difference and a resulting fringe pattern. Since the beams are superimposed and therefore inseparable, the interferometer cannot be put to any of the conventional uses. These in general depend on the possibility of imposing variations on only one of the constituent beams.

Real Fringes

Before we examine the creation of real, as opposed to virtual, fringes, let's first consider another amplitude-splitting interferometric device, the **Pohl fringe-producing system**, illustrated in Fig. 9.30. It is simply a thin transparent film illuminated by the light coming from a point source. In this case, the fringes are real and can accordingly be intercepted on a screen placed anywhere in the vicinity of the interferometer without a condensing-lens system. A convenient light source to use is a mercury lamp covered with a shield having a small hole ($\approx \frac{1}{4}$ inch diameter) in it. As a thin film, use a piece of

Interferogram without plasma. (Photo courtesy Los Alamos National Laboratory.)
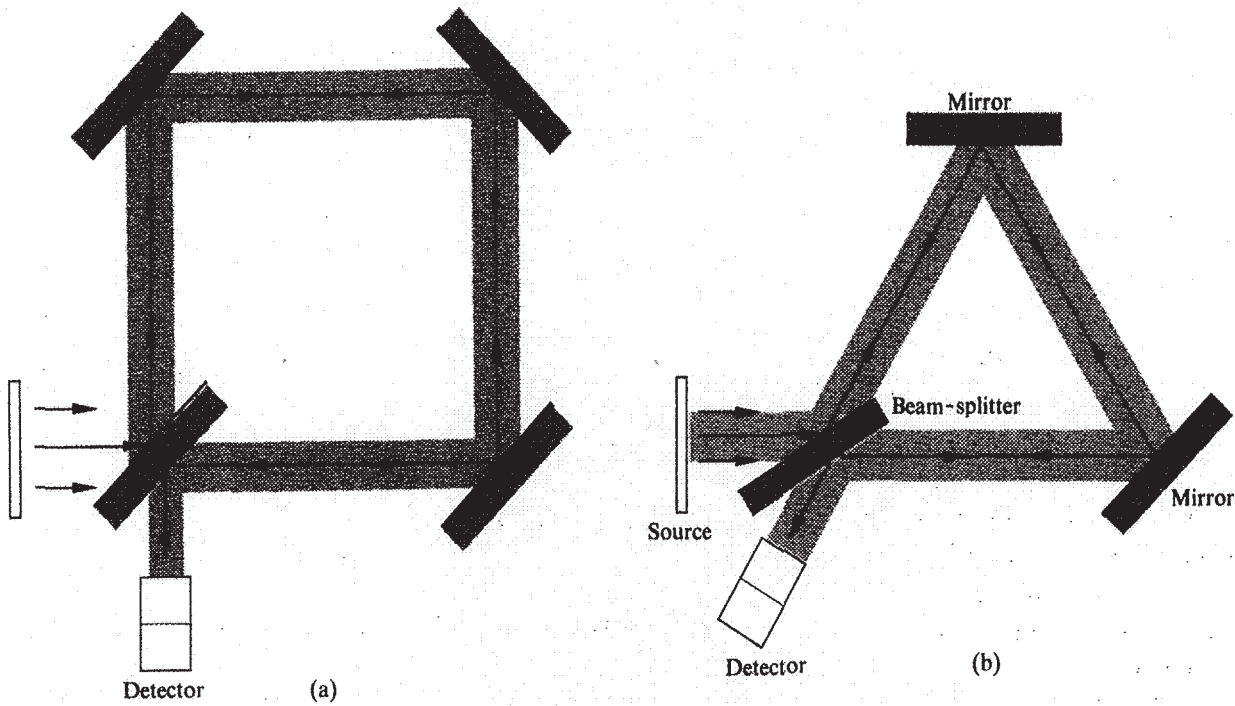
Figure 9.29 (*a*) A Sagnac Interferometer. (*b*) Another variation of the Sagnac Interferometer.
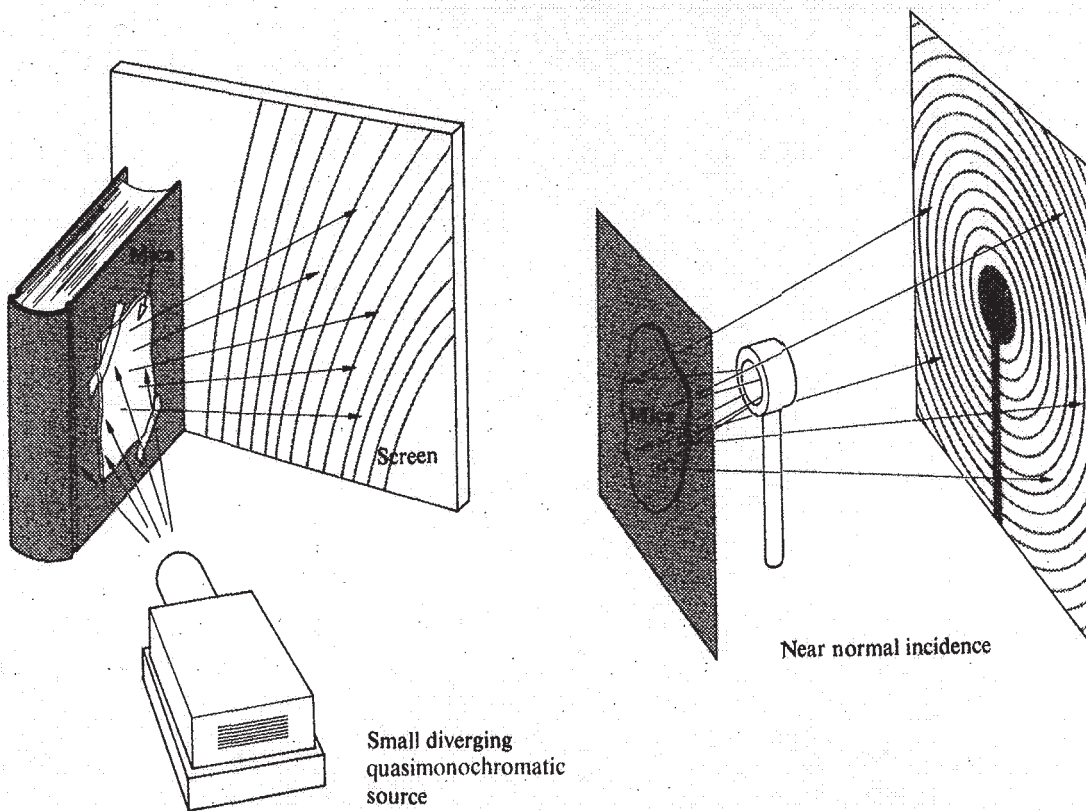
Figure 9.30 The Pohl Interferometer.
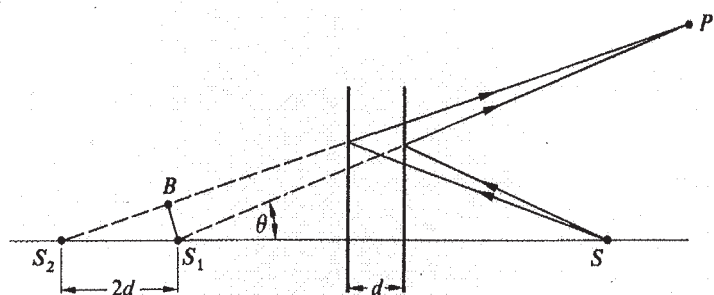
Figure 9.31 Point-source illumination of parallel surfaces.

ordinary mica taped to a dark-colored book cover, which serves as an opaque backing. If you have a laser, its remarkable coherence length and high flux density will allow you to perform this same experiment with almost anything smooth and transparent. Expand the beam to about an inch or two in diameter by passing it through a lens (a focal length of 50 to 100 mm will do). Then just reflect the beam off the surface of a glass plate (e.g., a microscope slide), and the fringes will be evident within the illuminated disk wherever it strikes a screen.

The underlying physical principle involved with point-source illumination for all four of the interferometric devices considered above can be appreciated with the help of a construction, variations of which are shown in Figs. 9.31 and 9.32.* The two vertical lines in Fig. 9.31, or the inclined ones in Fig. 9.32, represent either the positions of the mirrors or the two sides of the thin sheet in the Pohl Interferometer. Let's assume that point P in the surrounding medium is a point at

Real Michelson fringes using He–Ne laser light. (Photo by E. H.)
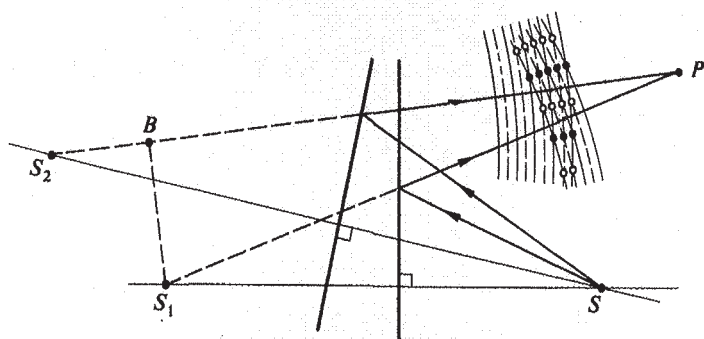
Figure 9.32 Point-source illumination of inclined surfaces.

*A. Zajac, H. Sadowski, and S. Licht, "The Real Fringes in the Sagnac and the Michelson Interferometers," *Am. J. Phys.* **29**, 669 (1961).

which there is constructive interference. A screen placed at that point would intercept this maximum, as well as a whole fringe pattern, without any condensing system. The coherent virtual sources emitting the interfering beams are mirror images S_1 and S_2 of the actual point source S. It should be noted that this kind of real fringe pattern can be observed with both the Michelson and Sagnac Interferometers. If either device is illuminated with an expanded laserbeam, a real fringe pattern will be generated directly by the emerging waves. This is an extremely simple and beautiful demonstration.

9.5 Types and Localization of Interference Fringes

Often it is important to know where the fringes produced in a given interferometric system will be located, since that is the region where we need to focus our detector (eye, camera, telescope). In general, the problem of locating fringes is characteristic of a given interferometer; that is, it has to be solved for each individual device.

Fringes can be classified, first, as either *real* or *virtual* and, second, as either *nonlocalized* or *localized*. Real fringes are those that can be seen on a screen without the use of an additional focusing system. The rays forming these fringes converge to the point of observation, all by themselves. Virtual

fringes cannot be projected onto a screen without a focusing system. In this case the rays obviously do not converge.

Nonlocalized fringes are real and exist everywhere within an extended (three-dimensional) region of space. The pattern is literally nonlocalized, in that it is not restricted to some small region. Young's Experiment, as illustrated in Fig. 9.8, fills the space beyond the secondary sources with a whole array of real fringes. Nonlocalized fringes of this sort are generally produced by small sources, that is, point or line sources, be they real or virtual. In contrast, localized fringes are clearly observable only over a particular surface. The pattern is literally localized, whether near a thin film or at infinity. This type of fringe will always result from the use of extended sources but can be generated with a point source as well.

The Pohl Interferometer (Fig. 9.30) is particularly useful in illustrating these principles, since with a point source it will produce both real nonlocalized and virtual localized fringes. The real nonlocalized fringes (Fig. 9.33, upper half) can be intercepted on a screen almost anywhere in front of the mica film.

For the nonconverging rays, realize that since the aperture of the eye is quite small, it will intercept only those rays that are directed almost exactly at it. For this small pencil of rays, the eye, at a particular position, sees either a bright or dark spot but not much more. To perceive an extended fringe pattern formed by parallel rays of the type shown in the bottom half of Fig. 9.33, a large lens will have to be used to gather in light entering at other orientations. In practice, however, the source is usually somewhat extended, and fringes can generally be seen by looking into the film with the eye focused at infinity. These virtual fringes are localized at infinity and are equivalent to the *equal-inclination fringes* of Section 9.4. Sim-
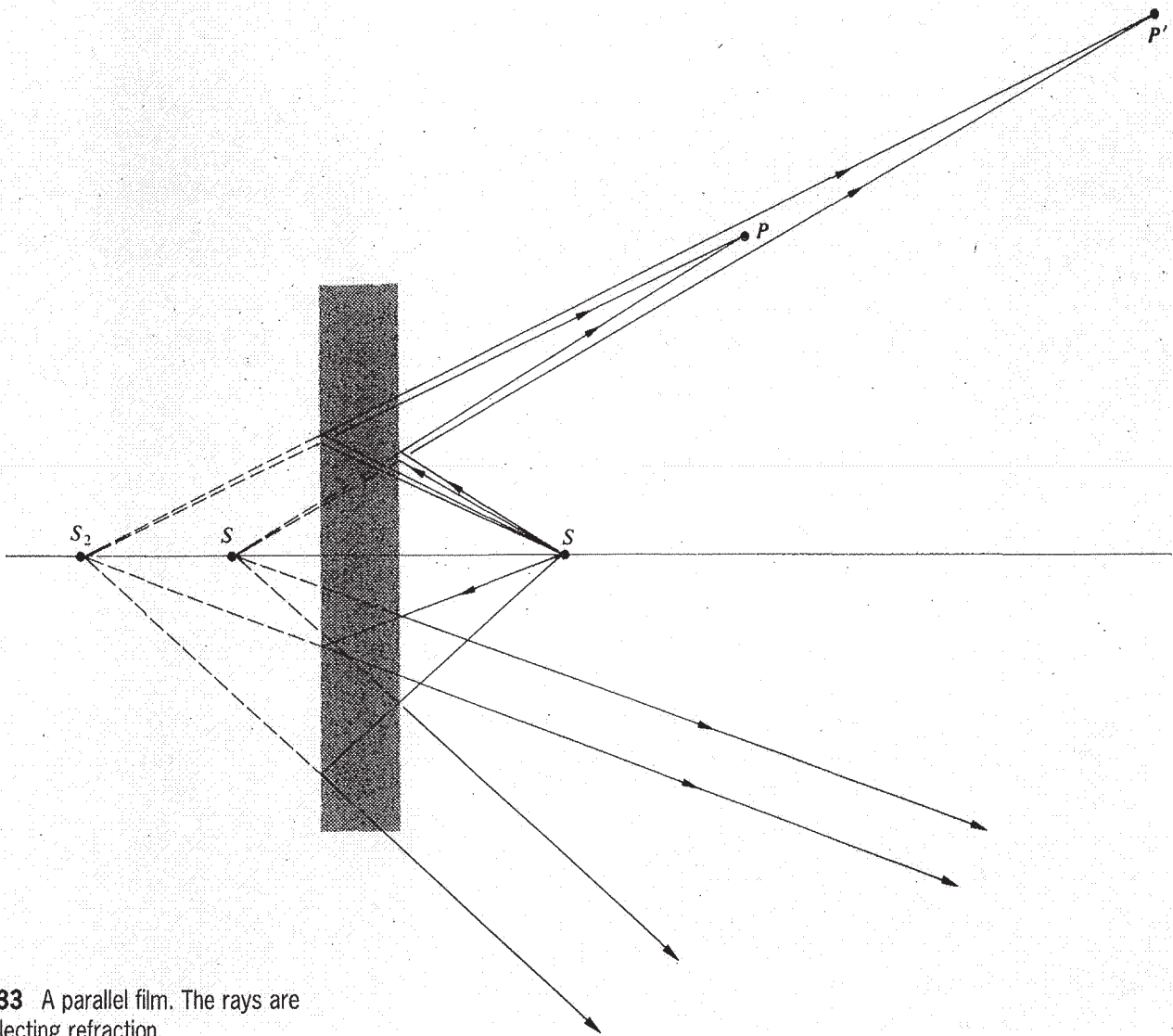
Figure 9.33 A parallel film. The rays are drawn neglecting refraction.
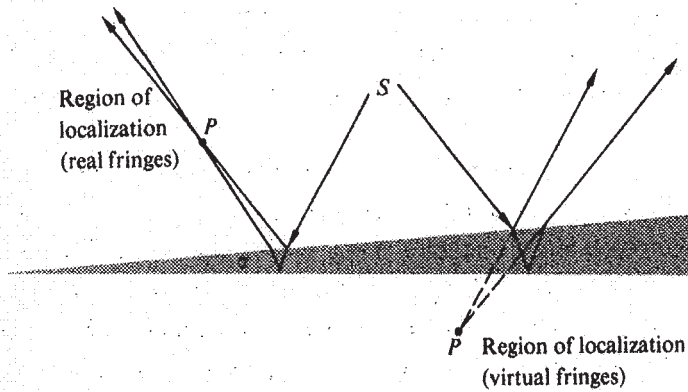
Figure 9.34 Fringes formed by a wedge-shaped film.

ilarly, if the mirrors M_1 and M_2 in the Michelson Interferometer are parallel, the usual circular, virtual, equal-inclination fringes localized at infinity will be seen. We can imagine a thin air film between the surfaces of the mirrors M_2 and M_1' acting to generate these fringes. As with the configuration of Fig. 9.30 for the Pohl device, real nonlocalized fringes will also be present.

The geometry of the fringe pattern seen in reflected light from a transparent wedge of small angle α is shown in Fig. 9.34. The fringe location P will be determined by the direction of incidence of the incoming light. Newton's rings have this same kind of localization, as do the Michelson, Sagnac, and other interferometers for which the equivalent interference system consists of two reflecting planes inclined slightly to
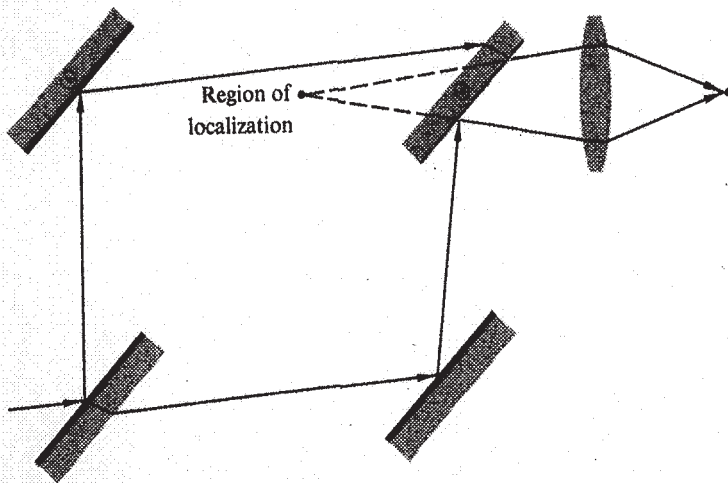
Figure 9.35 Fringes in the Mach–Zehnder Interferometer.

each other. The wedge setup of the Mach–Zehnder Interferometer is distinctive in that by rotating the mirrors, one can localize the resulting virtual fringes on any plane within the region generally occupied by the test chamber (Fig. 9.35).

9.6 Multiple Beam Interference

Thus far we have examined a number of situations in which two coherent beams are combined under diverse conditions to produce interference patterns. There are, however, circumstances under which a much larger number of mutually coherent waves are made to interfere. In fact, whenever the amplitude-reflection coefficients, the r's, for the parallel plate illustrated in Fig. 9.17 are not small, as was previously the case, the higher-order reflected waves $\vec{E}_{3r}$, $\vec{E}_{4r}$,... become quite significant. A glass plate, slightly silvered on both sides so that the r's approach unity, will generate a large number of multiply internally reflected rays. For the moment, we will consider only situations in which the film, substrate, and surrounding medium are transparent dielectrics. This avoids the more complicated phase changes resulting from metal-coated surfaces.

To begin the analysis as simply as possible, let the film be nonabsorbing and let $n_1 = n_2$. The notation will be in accord with that of Section 4.10; the amplitude-transmission coefficients are represented by t, the fraction of the amplitude of a wave transmitted on entering into the film, and t', the fraction transmitted when a wave leaves the film. The rays are actually lines drawn perpendicular to the wavefronts and therefore are also perpendicular to the optical fields $\vec{E}_{1r}$, $\vec{E}_{2r}$, and so forth. Since the rays will remain nearly parallel, the scalar theory will suffice as long as we are careful to account for any possible phase shifts.

As shown in Fig. 9.36, the scalar amplitudes of the reflected waves $\vec{E}_{1r}$, $\vec{E}_{2r}$, $\vec{E}_{3r}$,..., are respectively $E_0 r$, $E_0 t r' t'$, $E_0 t r'^3 t'$,..., where E_0 is the amplitude of the initial incoming wave and $r = -r'$ via Eq. (4.89). The minus sign indicates a phase shift, which we will consider later. Similarly, the transmitted waves $\vec{E}_{1t}$, $\vec{E}_{2t}$, $\vec{E}_{3t}$,... will have amplitudes $E_0 t t'$, $E_0 t r'^2 t'$, $E_0 t r'^4 t'$,.... Consider the set of parallel reflected rays. Each ray bears a fixed phase relationship to all the other reflected rays. The phase differences arise from a combination of optical path length differences and phase shifts occurring at the various reflections. Nonetheless, the waves are mutually
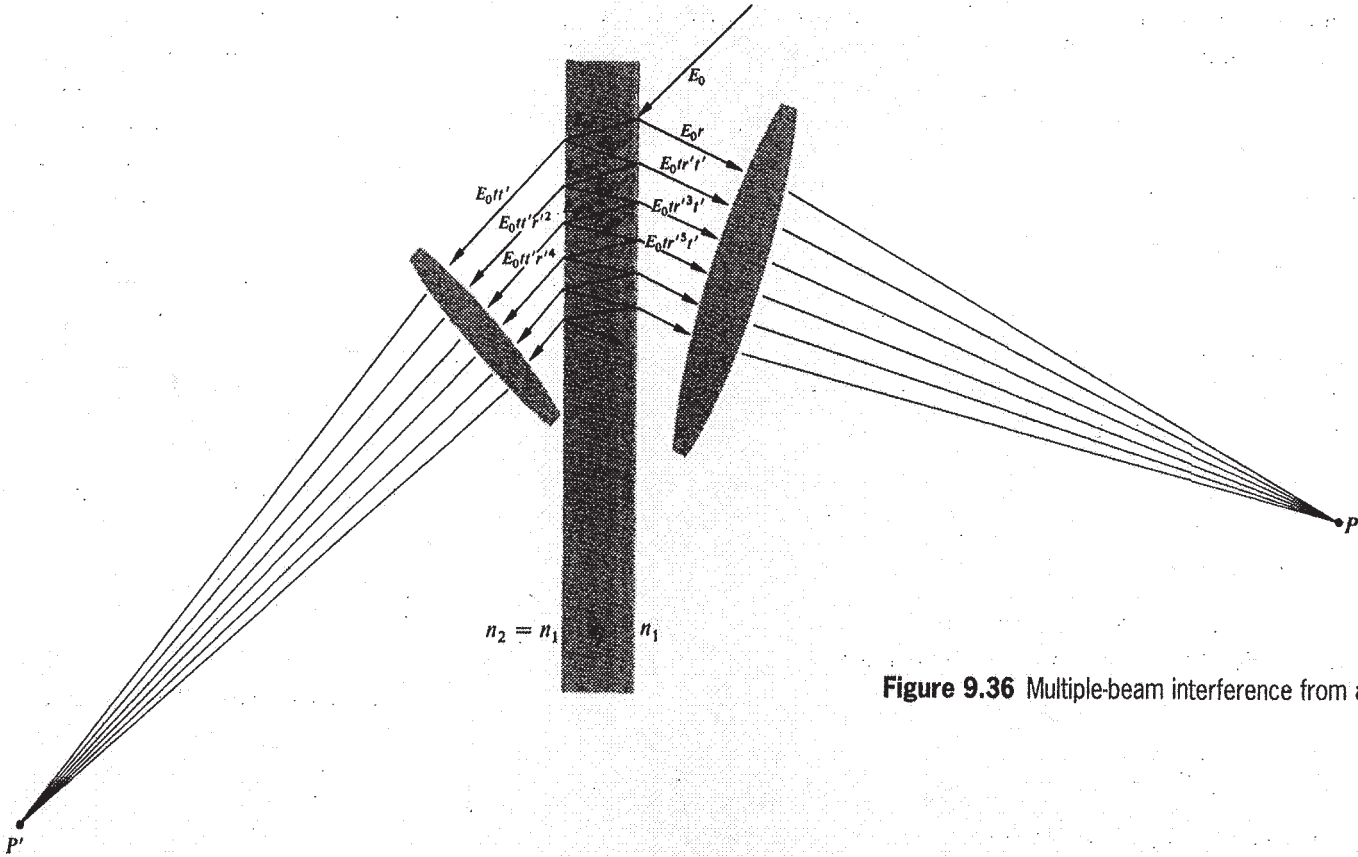
Figure 9.36 Multiple-beam interference from a parallel film.

coherent, and if they are collected and brought to focus at a point P by a lens, they will all interfere. The resultant irradiance expression has a particularly simple form for two special cases.

The difference in optical path length between adjacent rays is given by

$$\Lambda = 2n_f d \cos \theta_t \qquad [9.33]$$

All the waves except for the first, $\vec{E}_{1r}$, undergo an odd number of reflections *within* the film. It follows from Fig. 4.44 that at each internal reflection the component of the field parallel to the plane-of-incidence changes phase by either 0 or π, depending on the internal incident angle $\theta_i < \theta_c$. The component of the field perpendicular to the plane-of-incidence suffers no change in-phase on internal reflection when $\theta_i < \theta_c$. Clearly then, no relative change in-phase among these waves results from an odd number of such reflections (Fig. 9.37). As the *first special case*, if $\Lambda = m\lambda$, the second, third, fourth, and successive waves will all be in-phase at P. The wave $\vec{E}_{1r}$, however, because of its reflection at the top surface of the film, will be out-of-phase by 180° with respect to all the other waves. The
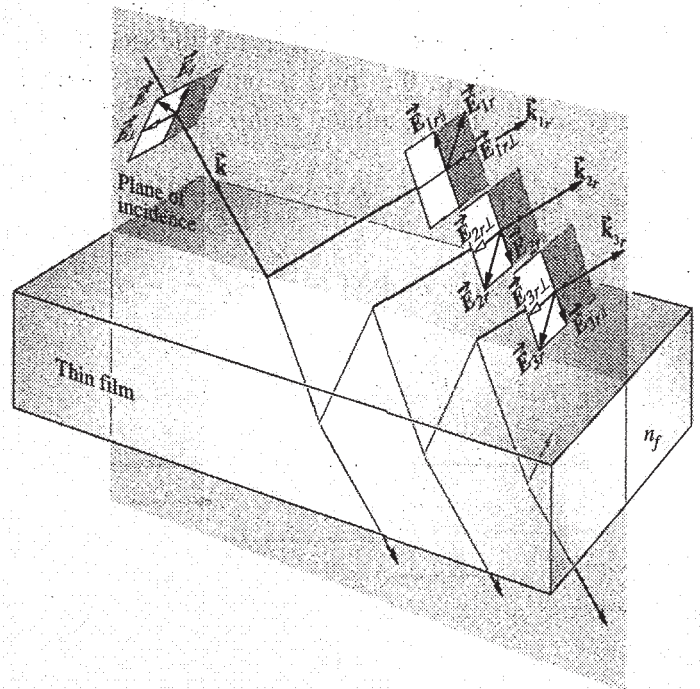
Figure 9.37 Phase shifts arising purely from the reflections (internal $\theta_i < \theta'_p$).

phase shift is embodied in the fact that $r = -r'$ and r' occurs only in odd powers. The sum of the scalar amplitudes, that is, the total *reflected amplitude* at point P, is then

$$E_{0r} = E_0 r - (E_0 trt' + E_0 tr^3 t' + E_0 tr^5 t' + \cdots)$$

or

$$E_{0r} = E_0 r - E_0 trt'(1 + r^2 + r^4 + \cdots)$$

where since $\Lambda = m\lambda$, we've just replaced r' by $-r$. The geometric series in parentheses converges to the finite sum $1/(1 - r^2)$ as long as $r^2 < 1$, so that

$$E_{0r} = E_0 r - \frac{E_0 trt'}{(1 - r^2)} \qquad (9.49)$$

It was shown in Section 4.10, when we considered Stokes's treatment of the principle of reversibility (Eq. 4.86), that $tt' = 1 - r^2$, and it follows that

$$E_{0r} = 0$$

Thus when $\Lambda = m\lambda$ the second, third, fourth, and successive waves exactly cancel the first reflected wave, as shown in Fig. 9.38. In this case no light is reflected; all the incoming energy is transmitted. The *second special case* arises when $\Lambda = (m + \frac{1}{2})\lambda$. Now the first and second rays are in-phase, and all other adjacent waves are $\lambda/2$ out-of-phase; that is, the second is out-of-phase with the third, the third is out-of-phase with the fourth, and so on. The resultant *scalar amplitude* is then

$$E_{0r} = E_0 r + E_0 trt' - E_0 tr^3 t' + E_0 tr^5 t' - \cdots$$

or

$$E_{0r} = E_0 r + E_0 rtt'(1 - r^2 + r^4 - \cdots)$$

$E_{0r} = 0$ (Resultant amplitude)

Figure 9.38 Phasor diagram.

(Resultant amplitude) E_{0r}
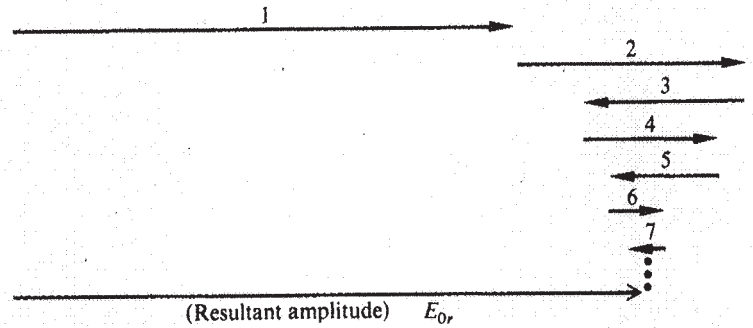
Figure 9.39 Phasor diagram.

The series in parentheses is equal to $1/(1 + r^2)$, in which case

$$E_{0r} = E_0 r \left[1 + \frac{tt'}{(1 + r^2)} \right]$$

Again, $tt' = 1 - r^2$; therefore, as illustrated in Fig. 9.39,

$$E_{0r} = \frac{2r}{(1 + r^2)} E_0$$

Since this particular arrangement results in the addition of the first and second waves, which have relatively large amplitudes, it should yield a large reflected flux density. The irradiance is proportional to $E_{0r}^2/2$, so from Eq. (3.44)

$$I_r = \frac{4r^2}{(1 + r^2)^2} \left(\frac{E_0^2}{2} \right) \qquad (9.50)$$

That this is in fact the maximum, $(I_r)_{max}$, will be shown later.

We will now consider the problem of multiple-beam interference in a more general fashion, making use of the complex representation. Again let $n_1 = n_2$, thereby avoiding the need to introduce different reflection and transmission coefficients at each interface. The optical fields at point P are given by

$$\tilde{E}_{1r} = E_0 r e^{i\omega t}$$
$$\tilde{E}_{2r} = E_0 tr' t' e^{i(\omega t - \delta)}$$
$$\tilde{E}_{3r} = E_0 tr'^3 t' e^{i(\omega t - 2\delta)}$$
$$\vdots$$
$$\tilde{E}_{Nr} = E_0 tr'^{(2N-3)} t' e^{i[\omega t - (N-1)\delta]}$$

where $E_0 e^{i\omega t}$ is the incident wave.

The terms $\delta, 2\delta, \ldots, (N-1)\delta$ are the contributions to the phase arising from an optical path length difference between adjacent rays ($\delta = k_0\Lambda$). There is an additional phase contribution arising from the optical distance traversed in reaching point P, but this is common to each ray and has been omitted. The relative phase shift undergone by the first ray as a result of the reflection is embodied in the quantity r'. The resultant *reflected scalar wave* is then

$$\tilde{E}_r = \tilde{E}_{1r} + \tilde{E}_{2r} + \tilde{E}_{3r} + \cdots + \tilde{E}_{Nr}$$

or upon substitution (Fig. 9.40)

$$\tilde{E}_r = E_0 r e^{i\omega t} + E_0 t r' t' e^{i(\omega t - \delta)} + \cdots + E_0 t r'^{(2N-3)} t'$$
$$\times e^{i[\omega t - (N-1)\delta]}$$

This can be rewritten as

$$\tilde{E}_r = E_0 e^{i\omega t}\{r + r't t' e^{-i\delta}[1 + (r'^2 e^{-i\delta})$$
$$+ (r'^2 e^{-i\delta})^2 + \cdots + (r'^2 e^{-i\delta})^{N-2}]\}$$

If $|r'^2 e^{-i\delta}| < 1$, and if the number of terms in the series approaches infinity, the series converges. The resultant wave becomes

$$\tilde{E}_r = E_0 e^{i\omega t}\left[r + \frac{r't t' e^{-i\delta}}{1 - r'^2 e^{-i\delta}}\right] \qquad (9.51)$$

In the case of zero absorption, no energy being taken out of the waves, we can use the relations $r = -r'$ and $tt' = 1 - r^2$ to rewrite Eq. (9.51) as

$$\tilde{E}_r = E_0 e^{i\omega t}\left[\frac{r(1 - e^{-i\delta})}{1 - r^2 e^{-i\delta}}\right]$$
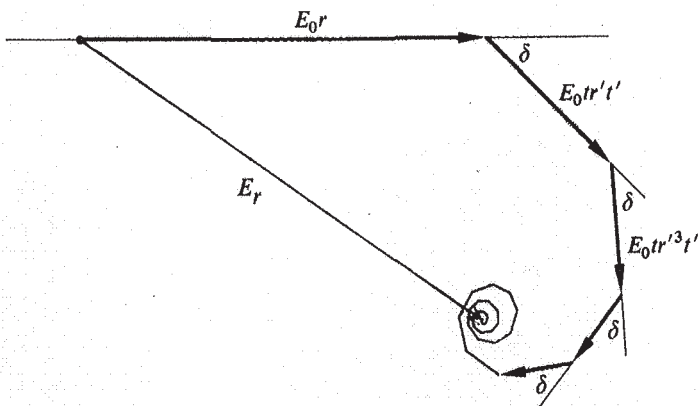
Figure 9.40 Phasor diagram.

The reflected flux density at P is then $I_r = \tilde{E}_r \tilde{E}_r^*/2$, that is,

$$I_r = \frac{E_0^2 r^2 (1 - e^{-i\delta})(1 - e^{+i\delta})}{2(1 - r^2 e^{-i\delta})(1 - r^2 e^{+i\delta})}$$

which can be transformed into

$$I_r = I_i \frac{2r^2(1 - \cos\delta)}{(1 + r^4) - 2r^2\cos\delta} \qquad (9.52)$$

The symbol $I_i = E_0^2/2$ represents the incident flux density, since, of course, E_0 was the amplitude of the incident wave. Similarly, the amplitudes of the transmitted waves given by

$$\tilde{E}_{1t} = E_0 tt' e^{i\omega t}$$
$$\tilde{E}_{2t} = E_0 tt' r'^2 e^{i(\omega t - \delta)}$$
$$\tilde{E}_{3t} = E_0 tt' r'^4 e^{i(\omega t - 2\delta)}$$
$$\vdots$$
$$\tilde{E}_{Nt} = E_0 tt' r'^{2(N-1)} e^{i[\omega - (N-1)\delta]}$$

can be added to yield

$$\tilde{E}_t = E_0 e^{i\omega t}\left[\frac{tt'}{1 - r^2 e^{-i\delta}}\right] \qquad (9.53)$$

(Because we are interested in the irradiance, a common factor of $e^{-i\delta/2}$, arising from the transmission through the film, was omitted. It contributes to the fact that there is a phase difference of $\pi/2$ between the reflected and transmitted waves, but that is of no concern here.)

Multiplying Eq. (9.53) by its complex conjugate yields (Problem 9.39) the irradiance of the transmitted beam

$$I_t = \frac{I_i (tt')^2}{(1 + r^4) - 2r^2\cos\delta} \qquad (9.54)$$

Using the trigonometric identity $\cos\delta = 1 - 2\sin^2(\delta/2)$, Eqs. (9.52) and (9.54) become

$$I_r = I_i \frac{[2r/(1 - r^2)]^2 \sin^2(\delta/2)}{1 + [2r/(1 - r^2)]^2 \sin^2(\delta/2)} \qquad (9.55)$$

and

$$I_t = I_i \frac{1}{1 + [2r/(1 - r^2)]^2 \sin^2(\delta/2)} \qquad (9.56)$$

where energy is not absorbed, that is, $tt' + r^2 = 1$. If indeed

none of the incident energy is absorbed, the flux density of the incoming wave should exactly equal the sum of the flux density reflected off the film and the total transmitted flux density emerging from the film. It follows from Eqs. (9.55) and (9.56) that this is indeed the case, namely,

$$I_i = I_r + I_t \qquad (9.57)$$

This will not be true, however, if the dielectric film is coated with a thin layer of semitransparent metal. Surface currents induced in the metal will dissipate a portion of the incident electromagnetic energy.

Consider the transmitted waves as described by Eq. (9.54). A maximum will exist when the denominator is as small as possible, that is, when $\cos \delta = 1$, in which case $\delta = 2\pi m$ and

$$(I_t)_{max} = I_i$$

Under these conditions, Eq. (9.52) indicates that

$$(I_r)_{min} = 0$$

as we would expect from Eq. (9.57). Again, from Eq. (9.54) it is clear that a minimum transmitted flux density will exist when the denominator is a maximum, that is, when $\cos \delta = -1$. In that case $\delta = (2m + 1)\pi$ and

$$(I_t)_{min} = I_i \frac{(1 - r^2)^2}{(1 + r^2)^2} \qquad (9.58)$$

The corresponding maximum in the reflected flux density is

$$(I_r)_{max} = I_i \frac{4r^2}{(1 + r^2)^2} \qquad (9.59)$$

Notice that the constant-inclination fringe pattern has its maxima when $\delta = (2m + 1)\pi$ or

$$\frac{4\pi n_f}{\lambda_0} d \cos \theta_t = (2m + 1)\pi$$

which is the same as the result we arrived at previously, in Eq. (9.36), by using only the first two reflected waves. Note, too, that Eq. (9.59) verifies that Eq. (9.50) was indeed a maximum.

The form of Eqs. (9.55) and (9.56) suggests that we introduce a new quantity, the *coefficient of finesse F*, such that

$$F \equiv \left(\frac{2r}{1 - r^2} \right)^2 \qquad (9.60)$$

whereupon these equations can be written as
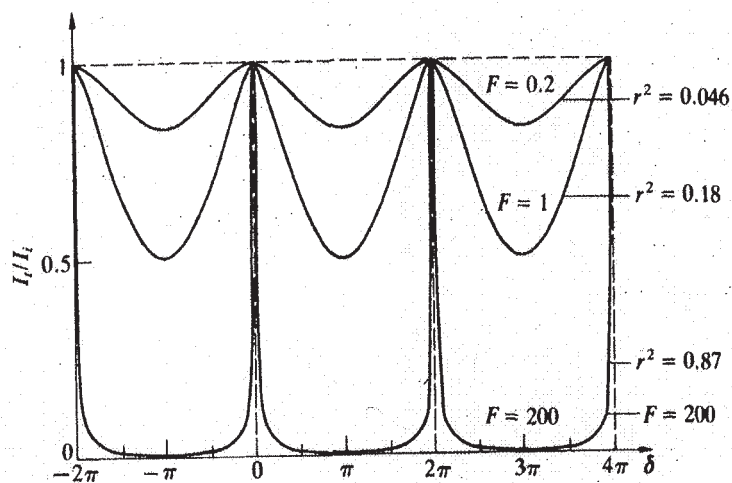
Figure 9.41 Airy function.

$$\frac{I_r}{I_i} = \frac{F \sin^2 (\delta/2)}{1 + F \sin^2(\delta/2)} \qquad (9.61)$$

and

$$\frac{I_t}{I_i} = \frac{1}{1 + F \sin^2(\delta/2)} \qquad (9.62)$$

The term $[1 + F \sin^2 (\delta/2)]^{-1} \equiv \mathscr{A}(\theta)$ is known as the **Airy function**. It represents the transmitted flux-density distribution and is plotted in Fig. 9.41. The complementary function $[1 - \mathscr{A}(\theta)]$, that is, Eq. (9.61), is plotted as well, in Fig. 9.42. When $\delta/2 = m\pi$, the Airy function is equal to unity for all values of F and therefore r. When r approaches 1, the transmitted flux density is very small, except within the sharp spikes centered about the points $\delta/2 = m\pi$. Multiple-beam interference has resulted in a redistribution of the energy density in comparison to the sinusoidal two-beam pattern (of
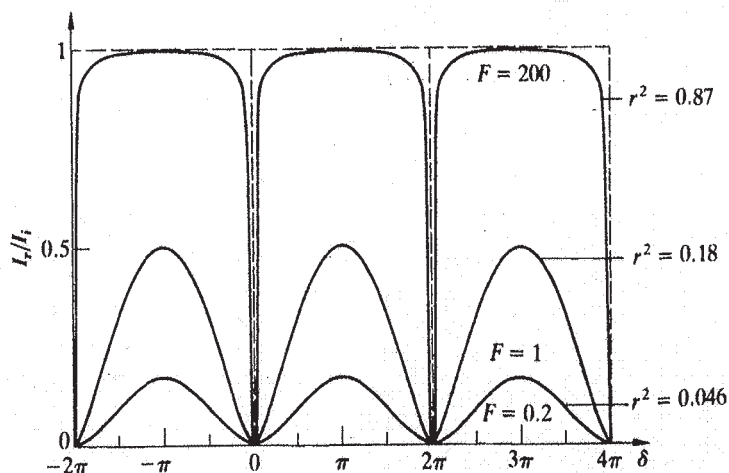
Figure 9.42 One minus the Airy function.

which the curves corresponding to a small reflectance are reminiscent). This effect will be further demonstrated when we consider the diffraction grating. At that time we will see this same peaking effect, resulting from an increased number of coherent sources contributing to the interference pattern. Remember that the Airy function is, in fact, a function of θ_t or θ_i by way of its dependence on δ, which follows from Eqs. (9.34) and (9.35), ergo the notation $\mathscr{A}(\theta)$. Each spike in the flux-density curve corresponds to a particular δ and therefore a particular θ_i. For a plane-parallel plate, the fringes, in transmitted light, will consist of a series of narrow bright rings on an almost completely dark background. In reflected light, the fringes will be narrow and dark on an almost uniformly bright background.

Constant-thickness fringes can also be made sharp and narrow by applying a light silver coating to the relevant reflecting surfaces to produce multiple-beam interference.

9.6.1 The Fabry–Perot Interferometer

The multiple-beam interferometer, first constructed by Charles Fabry and Alfred Perot in the late 1800s, is of considerable contemporary interest. Besides being a spectroscopic device of extremely high resolving power, it serves as the basic laser resonant cavity. In principle, the device consists of two plane, parallel, highly reflecting surfaces separated by some distance d. This is the simplest configuration, and as we shall see, other forms are also widely in use. In practice, two semisilvered or aluminized glass optical flats form the reflecting boundary surfaces. The enclosed air gap generally ranges from several millimeters to several centimeters when the apparatus is used interferometrically, and often to considerably greater lengths when it serves as a laser resonant cavity. If the gap can be mechanically varied by moving one of the mirrors, it's referred to as an interferometer. When the mirrors are held fixed and adjusted for parallelism by screwing down on some sort of spacer (invar or quartz is commonly used), it's said to be an **etalon** (although it is still an interferometer in the broad sense). If the two surfaces of a single quartz plate are appropriately polished and silvered, it too will serve as an etalon; the gap need not be air. The unsilvered sides of the plates are often made to have a slight wedge shape (a few minutes of arc) to reduce the interference pattern arising from reflections off these sides.

The etalon in Fig. 9.43 is shown illuminated by a broad source, which might be a mercury arc or a He–Ne laserbeam spread out in diameter to several centimeters. This can be done rather nicely by sending the beam into the back end of a telescope focused at infinity. The light can then be made diffuse by passing it through a sheet of ground glass. Only one ray emitted from some point S_1 on the source is traced through the etalon. Entering by way of the partially silvered plate, it is multiply reflected within the gap. The transmitted rays are collected by a lens and brought to a focus on a screen, where they interfere to form either a bright or dark spot. Consider this particular plane-of-incidence, which contains all the reflected rays. Any other ray emitted from a different point S_2, parallel to the original ray and in that plane-of-incidence, will form a spot at the same point P on the screen. As we shall see, the discussion of the previous section is again applicable, so that Eq. (9.54) determines the transmitted flux density I_t.

The multiple waves generated in the cavity, arriving at P from either S_1 or S_2, are coherent among themselves. But the rays arising from S_1 are completely incoherent with respect to those from S_2, so that there is no sustained mutual interference. The contribution to the irradiance I_t at P is just the sum of the two irradiance contributions.

All the rays incident on the gap at a given angle will result in a single circular fringe of uniform irradiance (Fig. 9.44). With a broad diffuse source, the interference bands will be narrow concentric rings, corresponding to the multiple-beam transmission pattern.

The fringe system can be observed visually by looking directly into the etalon, while focusing at infinity. The job of the focusing lens, which is no longer needed, is done by the eye. At large values of d, the rings will be close together, and a telescope might be needed to magnify the pattern. A relatively inexpensive monocular will serve the same purpose and will allow for photographing the fringes localized at infinity.
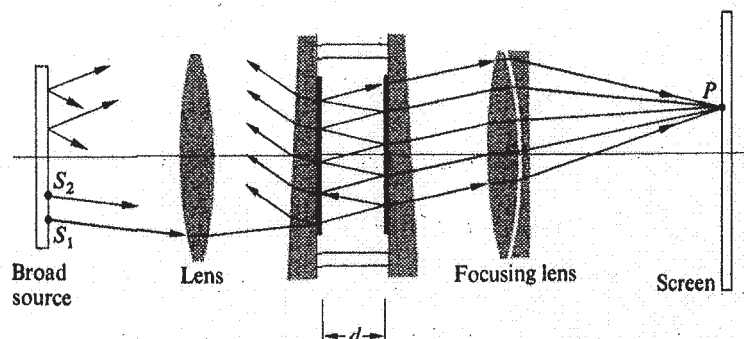
Figure 9.43 Fabry–Perot etalon.

(a)

Source

Lens

Etalon

Lens

Screen

(b)
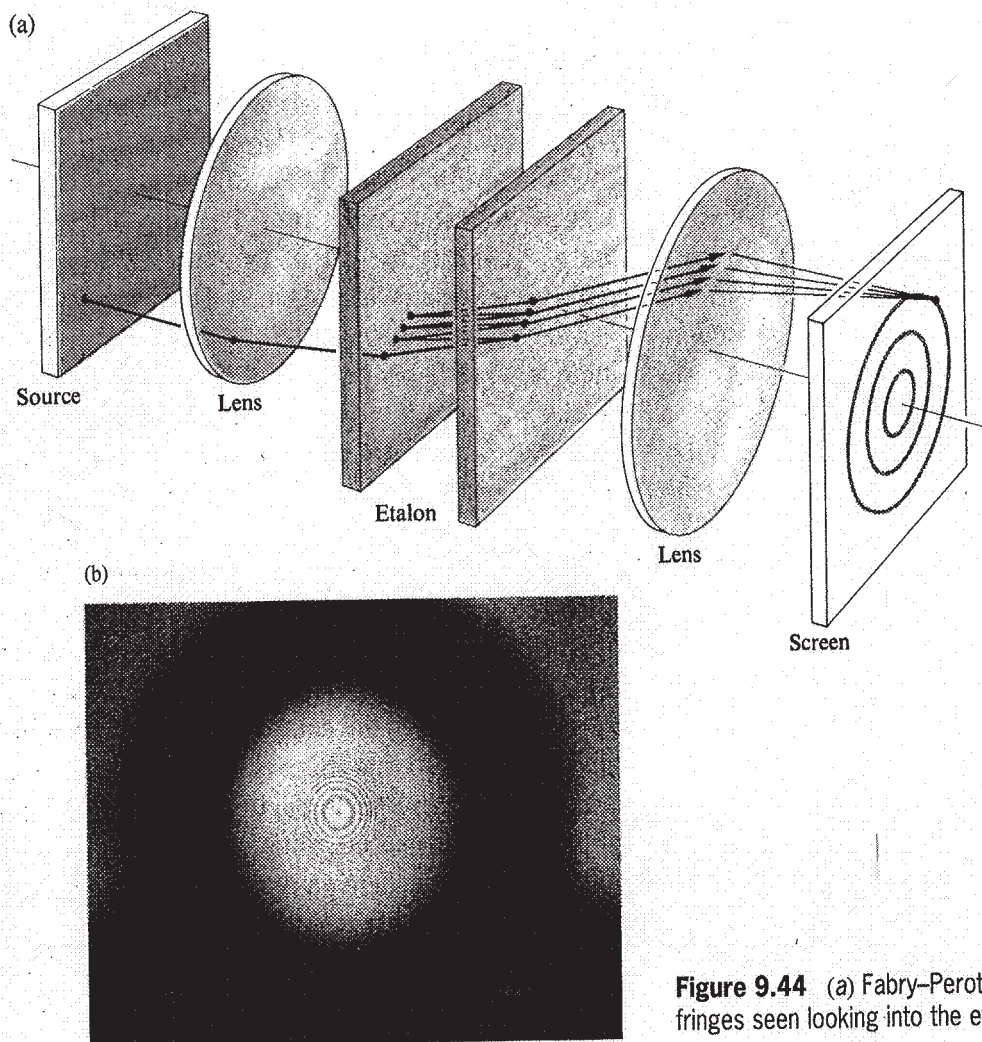
Figure 9.44 (a) Fabry–Perot etalon. (b) The axially symmetric fringes seen looking into the etalon. (Photo by E. H.)

As might be expected from the considerations of Section 9.5, it is possible to produce real nonlocalized fringes using a bright point source.

The partially transparent metal films that are often used to increase the reflectance ($R = r^2$) will absorb a fraction A of the flux density; this fraction is referred to as the **absorptance**.

The expression

$$tt' + r^2 = 1$$

or

$$T + R = 1 \qquad [4.60]$$

where T is the transmittance, must now be rewritten as

$$T + R + A = 1 \qquad (9.63)$$

One further complication introduced by the metallic films is an additional phase shift $\phi(\theta_i)$, which can differ from either zero or π. The phase difference between two successively transmitted waves is then

$$\delta = \frac{4\pi n_f}{\lambda_0} d \cos \theta_t + 2\phi \qquad (9.64)$$

For the present conditions, θ_i is small and ϕ may be considered to be constant. In general, d is so large, and λ_0 so small, that ϕ can be neglected. We can now express Eq. (9.54) as

$$\frac{I_t}{I_i} = \frac{T^2}{1 + R^2 - 2R \cos \delta}$$

or equivalently

$$\frac{I_t}{I_i} = \left(\frac{T}{1 - R} \right)^2 \frac{1}{1 + [4R/(1 - R)^2] \sin^2(\delta/2)} \qquad (9.65)$$

Making use of Eq. (9.63) and the definition of the Airy function, we obtain

$$\frac{I_t}{I_i} = \left[1 - \frac{A}{(1 - R)} \right]^2 \mathcal{A}(\theta) \qquad (9.66)$$

as compared with the equation for zero absorption

$$\frac{I_t}{I_i} = \mathscr{A}(\theta) \qquad [9.62]$$

Inasmuch as the absorbed portion A is never zero, the transmitted flux-density maxima $(I_t)_{max}$ will always be somewhat less than I_i. [Recall that for $(I_t)_{max}$, $\mathscr{A}(\theta) = 1$.]

Accordingly, the *peak transmission* is defined as $(I_t/I_i)_{max}$:

$$\frac{(I_t)_{max}}{I_i} = \left[1 - \frac{A}{(1-R)}\right]^2 \qquad (9.67)$$

A silver film 50 nm thick would be approaching its maximum value of R (e.g., about 0.94), while T and A might be, respectively, 0.01 and 0.05. In this case, the peak transmission will be down to $\frac{1}{36}$. The relative irradiance of the fringe pattern will still be determined by the Airy function, since

$$\frac{I_t}{(I_t)_{max}} = \mathscr{A}(\theta) \qquad (9.68)$$

A measure of the sharpness of the fringes, that is, how rapidly the irradiance drops off on either side of the maximum, is given by the half-width γ. Shown in Fig. 9.45, γ is the width of the peak, in radians, when $I_t = (I_t)_{max}/2$.

Peaks in the transmission occur at specific values of the phase difference $\delta_{max} = 2\pi m$. Accordingly, the irradiance will drop to half its maximum value [i.e., $\mathscr{A}(\theta) = \frac{1}{2}$] whenever $\delta = \delta_{max} \pm \delta_{1/2}$. Inasmuch as

$$\mathscr{A}(\theta) = [1 + F \sin^2(\delta/2)]^{-1}$$
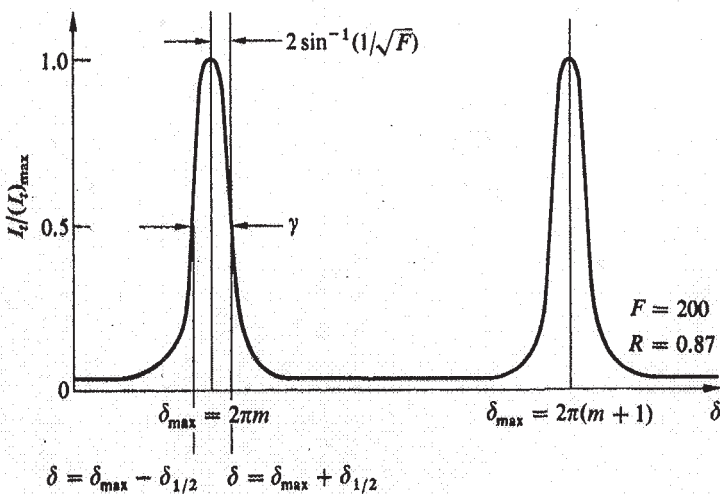
Figure 9.45 Fabry–Perot fringes.

then when

$$[1 + F \sin^2(\delta_{1/2}/2)]^{-1} = \frac{1}{2}$$

it follows that

$$\delta_{1/2} = 2 \sin^{-1}(1/\sqrt{F})$$

Since F is generally rather large, $\sin^{-1}(1/\sqrt{F}) \approx 1/\sqrt{F}$, and therefore the half-width, $\gamma = 2\delta_{1/2}$, becomes

$$\gamma = 4/\sqrt{F} \qquad (9.69)$$

Recall that $F = 4R/(1-R)^2$, so that the larger R is, the sharper the transmission peaks will be.

Another quantity of particular interest is the ratio of the separation of adjacent maxima to the half-width. Known as the **finesse**, $\mathscr{F} \equiv 2\pi/\gamma$ or, from Eq. (9.69),

$$\mathscr{F} = \frac{\pi\sqrt{F}}{2} \qquad (9.70)$$

Over the visible spectrum, the finesse of most ordinary Fabry–Perot instruments is about 30. The physical limitation on $\mathscr{F}$ is set by deviations in the mirrors from plane parallelism. Keep in mind that as the finesse increases, the half-width decreases, but so too does the peak transmission. Incidentally, a finesse of about 1000 is attainable with curved-mirror systems using dielectric thin-film coatings.*

Fabry–Perot Spectroscopy

The Fabry–Perot Interferometer is frequently used to examine the detailed structure of spectral lines. We will not attempt a complete treatment of interference spectroscopy, but rather will define the relevant terminology, briefly outlining appropriate derivations.†

....

*The paper "Multiple Beam Interferometry," by H. D. Polster, *Appl. Opt.* **8**, 522 (1969), should be of interest. Also look at "The Optical Computer," E. Abraham, C. Seaton, and S. Smith, *Sci. Am.* (Feb. 1983), p. 85, for a discussion of the use of the Fabry–Perot Interferometer as an optical transistor.

....

†A more complete treatment can be found in Born and Wolf, *Principles of Optics*, and in W. E. Williams, *Applications of Interferometry*, to name only two.

As we have seen, a hypothetical, purely monochromatic lightwave generates a particular circular fringe system. But δ is a function of λ_0, so that if the source were made up of two such monochromatic components, two superimposed ring systems would result. When the individual fringes partially overlap, a certain amount of ambiguity exists in deciding when the two systems are individually discernible, that is, when they are said to be *resolved*. Lord Rayleigh's* criterion for resolving two equal-irradiance overlapping slit images is well accepted, even if somewhat arbitrarily in the present application. Its use, however, will allow a comparison with prism or grating instruments. The essential feature of this criterion is that the fringes are *just resolvable* when the combined irradiance of both fringes at the center, or saddle point, of the resultant broad fringe is $8/\pi^2$ times the maximum irradiance. This simply means that one would see a broad bright fringe with a grey central region. To be a bit more analytic about it, examine Fig. 9.46, keeping in mind the previous derivation of the half-width. Consider the case in which the two constituent fringes have equal irradiances, $(I_a)_{max} = (I_b)_{max}$. The peaks in the resultant, occurring at $\delta = \delta_a$ and $\delta = \delta_b$, will have equal irradiances,

$$(I_t)_{max} = (I_a)_{max} + I' \qquad (9.71)$$

At the saddle point, the irradiance $(8/\pi^2)(I_t)_{max}$ is the sum of

Figure 9.46 Overlapping fringes.

....

*The criterion will be reconsidered with respect to diffraction in the next chapter (see Fig. 10.26).

the two constituent irradiances, so that, recalling Eq. (9.68),

$$(8/\pi^2)\frac{(I_t)_{max}}{(I_a)_{max}} = [\mathscr{A}(\theta)]_{\delta=\delta_a+\Delta\delta/2} + [\mathscr{A}(\theta)]_{\delta=\delta_b+\Delta\delta/2}$$
$$(9.72)$$

Using $(I_t)_{max}$ given by Eq. (9.71), along with the fact that

$$\frac{I'}{(I_a)_{max}} = [\mathscr{A}(\theta)]_{\delta=\delta_a+\Delta\delta}$$

we can solve Eq. (9.72) for $\Delta\delta$. For large values of F,

$$(\Delta\delta) \approx \frac{4.2}{\sqrt{F}} \qquad (9.73)$$

This then represents the smallest phase increment, $(\Delta\delta)_{min}$, separating two resolvable fringes. It can be related to equivalent minimum increments in wavelength $(\Delta\lambda_0)_{min}$, frequency $(\Delta\nu)_{min}$, and wave number $(\Delta k)_{min}$. From Eq. (9.64), for $\delta = 2\pi m$, we have

$$m\lambda_0 = 2n_f d \cos\theta_t + \frac{\phi\lambda_0}{\pi} \qquad (9.74)$$

Dropping the term $\phi\lambda_0/\pi$, which is clearly negligible, and then differentiating, yields

$$m(\Delta\lambda_0) + \lambda_0(\Delta m) = 0$$

or

$$\frac{\lambda_0}{(\Delta\lambda_0)} = -\frac{m}{(\Delta m)}$$

The minus will be omitted, since it means only that the order increases when λ_0 decreases. When δ changes by 2π, m changes by 1,

$$\frac{2\pi}{(\Delta\delta)} = \frac{1}{(\Delta m)}$$

and thus

$$\frac{\lambda_0}{(\Delta\lambda_0)} = \frac{2\pi m}{(\Delta\delta)} \qquad (9.75)$$

The ratio of λ_0 to the least resolvable wavelength difference, $(\Delta\lambda_0)_{min}$, is known as the **chromatic resolving power** $\mathscr{R}$ of any spectroscope. At nearly normal incidence

$$\mathscr{R} \equiv \frac{\lambda_0}{(\Delta\lambda_0)_{min}} \approx \mathscr{F}\frac{2n_f d}{\lambda_0} \qquad (9.76)$$

or

$$\mathscr{R} \approx \mathscr{F}m$$

For a wavelength of 500 nm, $n_f d = 10$ mm, and $R = 90\%$, the

resolving power is well over a million, a range achieved by the finest diffraction gratings. It follows as well, in this example, that $(\Delta\lambda_0)_{min}$ is less than a millionth of λ_0. In terms of frequency, the *minimum resolvable bandwidth* is

$$(\Delta\nu)_{min} = \frac{c}{\mathscr{F}2n_f d} \qquad (9.77)$$

inasmuch as $|\Delta\nu| = |c\Delta\lambda_0/\lambda_0^2|$.

As the two components present in the source become increasingly different in wavelength, the peaks shown overlapping in Fig. 9.46 separate. As the wavelength difference increases, the mth-order fringe for one wavelength λ_0 will approach the $(m + 1)$th-order for the other wavelength $(\lambda_0 - \Delta\lambda_0)$. The particular wavelength difference at which overlapping takes place, $(\Delta\lambda_0)_{fsr}$, is known as the *free spectral range*. From Eq. (9.75), a change in δ of 2π corresponds to $(\Delta\lambda_0)_{fsr} = \lambda_0/m$, or at near normal incidence,

$$(\Delta\lambda_0)_{fsr} \approx \lambda_0^2/2n_f d \qquad (9.78)$$

and similarly

$$(\Delta\nu)_{fsr} \approx c/2n_f d \qquad (9.79)$$

Continuing with the above example (i.e., $\lambda_0 = 500$ nm and $n_f d = 10$ mm), $(\Delta\lambda_0)_{fsr} = 0.0125$ nm. If we attempt to increase the resolving power by merely increasing d, the free spectral range will decrease, bringing with it the resulting confusion from the overlapping of orders. What is needed is that $(\Delta\lambda_0)_{min}$ *be as small as possible and* $(\Delta\lambda_0)_{fsr}$ *be as large as possible*. But lo and behold,

$$\frac{(\Delta\lambda_0)_{fsr}}{(\Delta\lambda_0)_{min}} = \mathscr{F} \qquad (9.80)$$

This result should not be too surprising in view of the original definition of $\mathscr{F}$.

Both the applications and configurations of the Fabry–Perot Interferometer are numerous indeed. Etalons have been arranged in series with other etalons, as well as with grating and prism spectroscopes, and multilayer dielectric films have been used to replace the metallic mirror coatings.

Scanning techniques are now widely in use. These take advantage of the superior linearity of photoelectric detectors over photographic plates, to obtain more reliable flux-density measurements. The basic setup for *central-spot scanning* is illustrated in Fig. 9.47. Scanning is accomplished by varying δ, by changing n_f or d rather than cos θ_t. In some arrangements, n_f is smoothly varied by altering the air pressure with-
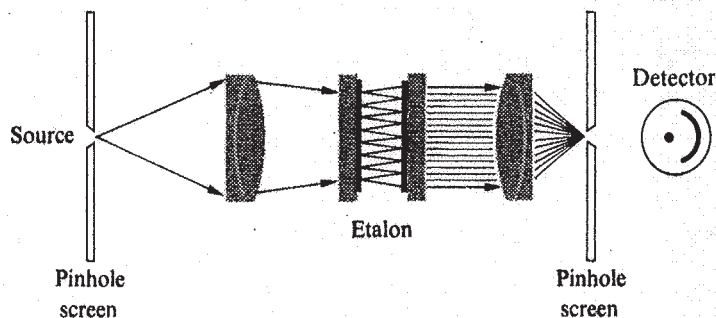

Figure 9.47 Central spot scanning.

in the etalon. Alternatively, mechanical vibration of one mirror with a displacement of $\lambda_0/2$ will be enough to scan the free spectral range, corresponding as it does to $\Delta\delta = 2\pi$. A popular technique for accomplishing this utilizes a piezoelectric mirror mount. This kind of material will change its length, and therefore d, as a voltage is applied to it. The voltage profile determines the mirror motion.

Instead of photographically recording irradiance over a large region in space, at a single point in time, this method records irradiance over a large region in time, at a single point in space.

The actual configuration of the etalon itself has also undergone some significant variations. Pierre Connes in 1956 first described the *spherical-mirror Fabry–Perot Interferometer*. Since then, curved-mirror systems have become prominent as laser cavities and are also finding increasing use as spectrum analyzers.

9.7 Applications of Single and Multilayer Films

The optical uses to which coatings of thin dielectric films have been put in recent times are many indeed. Coatings to eliminate unwanted reflections off a diversity of surfaces, from showcase glass to high-quality camera lenses, are now commonplace (see photo on page 426). Multilayer, nonabsorbing beamsplitters and *dichroic* mirrors (color-selective beamsplitters that transmit and reflect particular wavelengths) can be purchased commercially.

Figure 9.48 is a segmented diagram illustrating the use of a *cold mirror* in combination with a *heat reflector* to channel infrared radiation to the rear of a motion-picture projector. The intense unwanted infrared radiation emitted by the source is
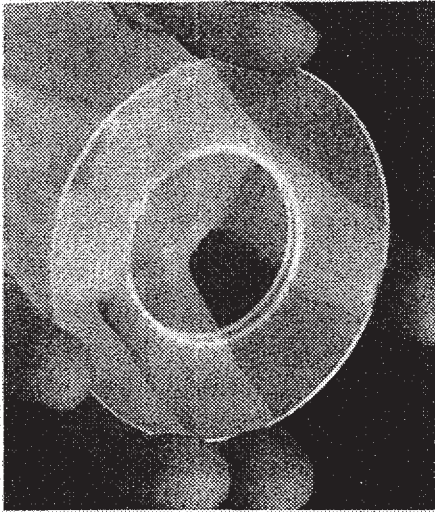
This glass disk has an antireflection coating in the shape of a circle applied to the central region of both its sides. (Photo by E. H.)

removed from the beam to avoid heating problems at the photographic film. The top half of Fig. 9.48 is an ordinary back-silvered mirror shown for comparison. Solar cells, which are one of the prime power-supply systems for space vehicles, and even the astronauts' helmets and visors, are shielded with similar heat control coverings.

Multilayer broad and narrow band-pass filters that transmit only over a specific spectral range can be made to span the region from infrared to ultraviolet. In the visible, for example, they play an important part in splitting up the image in color television cameras, and in the infrared they're used in missile guidance systems, CO_2 lasers, and satellite horizon sensors. The applications of thin-film devices are manifold, as are their
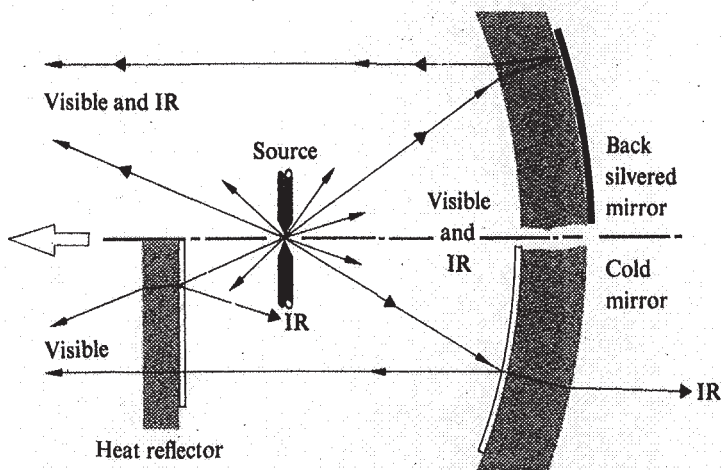

Figure 9.48 A composite drawing showing an ordinary system in the top half and a coated one in the bottom.

structures, which extend from the simplest single coatings to intricate arrangements of one hundred or more layers.

The treatment of multilayer film theory used here will deal with the *total* electric and magnetic fields and their boundary conditions in the various regions. This is a far more practical approach for many-layered systems than is the multiple-wave technique used earlier.*

9.7.1 Mathematical Treatment

Consider the linearly polarized wave shown in Fig. 9.49, impinging on a thin dielectric film between two semi-infinite transparent media. In practice, this might correspond to a dielectric layer a fraction of a wavelength thick, deposited on
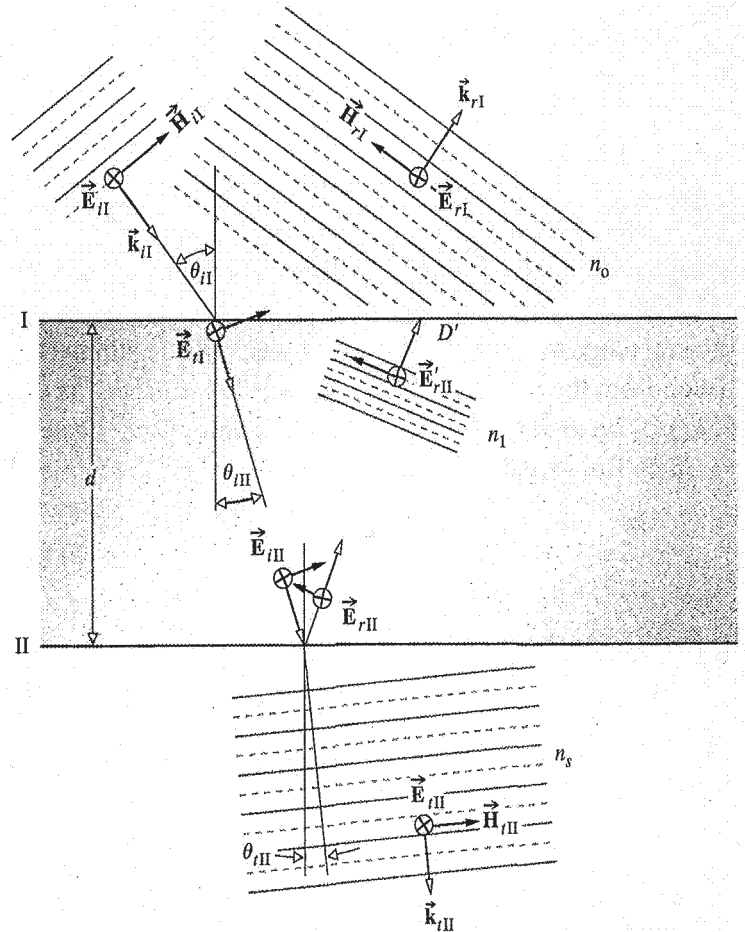

Figure 9.49 Fields at the boundaries.

*For a very readable nonmathematical discussion, see P. Baumeister and G. Pincus, "Optical Interference Coatings," *Sci. Amer.* **223**, 59 (December 1970).

the surface of a lens, a mirror, or a prism. One point must be made clear at the outset: each wave E_{rI}, E'_{rII}, E_{tII}, and so forth, represents the resultant of all possible waves traveling in that direction, at that point in the medium. The summation process is therefore built in. As discussed in Section 4.6.2, the boundary conditions require that the tangential components of both the electric ($\vec{E}$) and magnetic ($\vec{H} = \vec{B}/\mu$) fields be continuous across the boundaries (i.e., equal on both sides). At boundary I

$$E_I = E_{iI} + E_{rI} = E_{tI} + E'_{rII} \tag{9.81}$$

and
$$H_I = \sqrt{\frac{\epsilon_0}{\mu_0}}(E_{iI} - E_{rI})n_0 \cos \theta_{iI}$$

$$H_I = \sqrt{\frac{\epsilon_0}{\mu_0}}(E_{tI} - E'_{rII})n_1 \cos \theta_{iII} \tag{9.82}$$

where use is made of the fact that $\vec{E}$ and $\vec{H}$ in nonmagnetic media are related through the index of refraction and the unit propagation vector:

$$\vec{H} = \sqrt{\frac{\epsilon_0}{\mu_0}} n\hat{k} \times \vec{E}$$

At boundary II
$$E_{II} = E_{iII} + E_{rII} = E_{tII} \tag{9.83}$$

and
$$H_{II} = \sqrt{\frac{\epsilon_0}{\mu_0}}(E_{iII} - E_{rII})n_1 \cos \theta_{iII}$$

$$H_{II} = \sqrt{\frac{\epsilon_0}{\mu_0}} E_{tII} n_s \cos \theta_{tII} \tag{9.84}$$

the substrate having an index n_s. In accord with Eq. (9.33), a wave that traverses the film once undergoes a shift in-phase of $k_0(2n_1 d \cos \theta_{iII})/2$, which will be denoted by $k_0 h$, so that

$$E_{iII} = E_{tI} e^{-ik_0 h} \tag{9.85}$$

and
$$E_{rII} = E'_{rII} e^{+ik_0 h} \tag{9.86}$$

Equations (9.83) and (9.84) can now be written as

$$E_{II} = E_{tI} e^{-ik_0 h} + E'_{rII} e^{+ik_0 h} \tag{9.87}$$

and $H_{II} = (E_{tI} e^{-ik_0 h} - E'_{rII} e^{+ik_0 h})\sqrt{\frac{\epsilon_0}{\mu_0}} n_1 \cos \theta_{iII}$ (9.88)

These last two equations can be solved for E_{tI} and E'_{rII}, which when substituted into Eqs. (9.81) and (9.82) yield

$$E_I = E_{II} \cos k_0 h + H_{II}(i \sin k_0 h)/Y_1 \tag{9.89}$$

and $\quad H_I = E_{II} Y_1 i \sin k_0 h + H_{II} \cos k_0 h \tag{9.90}$

where
$$Y_1 \equiv \sqrt{\frac{\epsilon_0}{\mu_0}} n_1 \cos \theta_{iII}$$

When $\vec{E}$ is in the plane-of-incidence, the above calculations result in similar equations, provided that now

$$Y_1 \equiv \sqrt{\frac{\epsilon_0}{\mu_0}} n_1/\cos \theta_{iII}$$

In matrix notation, the above linear relations take the form

$$\begin{bmatrix} E_I \\ H_I \end{bmatrix} = \begin{bmatrix} \cos k_0 h & (i \sin k_0 h)/Y_1 \\ Y_1 i \sin k_0 h & \cos k_0 h \end{bmatrix} \begin{bmatrix} E_{II} \\ H_{II} \end{bmatrix} \tag{9.91}$$

or
$$\begin{bmatrix} E_I \\ H_I \end{bmatrix} = \mathcal{M}_I \begin{bmatrix} E_{II} \\ H_{II} \end{bmatrix} \tag{9.92}$$

The *characteristic matrix* $\mathcal{M}_I$ relates the fields at the two adjacent boundaries. It follows, therefore, that if two overlaying films are deposited on the substrate, there will be three boundaries or interfaces, and now

$$\begin{bmatrix} E_{II} \\ H_{II} \end{bmatrix} = \mathcal{M}_{II} \begin{bmatrix} E_{III} \\ H_{III} \end{bmatrix} \tag{9.93}$$

Multiplying both sides of this expression by $\mathcal{M}_I$, we obtain

$$\begin{bmatrix} E_I \\ H_I \end{bmatrix} = \mathcal{M}_I \mathcal{M}_{II} \begin{bmatrix} E_{III} \\ H_{III} \end{bmatrix} \tag{9.94}$$

In general, if p is the number of layers, each with a particular value of n and h, then the first and the last boundaries are related by

$$\begin{bmatrix} E_I \\ H_I \end{bmatrix} = \mathcal{M}_I \mathcal{M}_{II} \cdots \mathcal{M}_p \begin{bmatrix} E_{(p+1)} \\ H_{(p+1)} \end{bmatrix} \tag{9.95}$$

The characteristic matrix of the entire system is the resultant of the product (in the proper sequence) of the individual 2×2 matrices, that is,

$$\mathcal{M} = \mathcal{M}_I \mathcal{M}_{II} \cdots \mathcal{M}_p = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \tag{9.96}$$

To see how all this fits together, we will derive expressions for the amplitude coefficients of reflection and transmission using the above scheme. By reformulating Eq. (9.92) in terms of the boundary conditions [(9.81), (9.82), and (9.84)] and

setting

$$Y_0 = \sqrt{\frac{\epsilon_0}{\mu_0}}\, n_0 \cos\theta_{iI}$$

and

$$Y_s = \sqrt{\frac{\epsilon_0}{\mu_0}}\, n_s \cos\theta_{tII}$$

we obtain

$$\begin{bmatrix} (E_{iI} + E_{rI}) \\ (E_{iI} - E_{rI})Y_0 \end{bmatrix} = \mathcal{M}_1 \begin{bmatrix} E_{tII} \\ E_{tII}Y_s \end{bmatrix}$$

When the matrices are expanded, the last relation becomes

$$1 + r = m_{11}t + m_{12}Y_s t$$

and

$$(1 - r)Y_0 = m_{21}t + m_{22}Y_s t$$

inasmuch as

$$r = E_{rI}/E_{iI} \quad \text{and} \quad t = E_{tII}/E_{iI}$$

Consequently,

$$r = \frac{Y_0 m_{11} + Y_0 Y_s m_{12} - m_{21} - Y_s m_{22}}{Y_0 m_{11} + Y_0 Y_s m_{12} + m_{21} + Y_s m_{22}} \quad (9.97)$$

and

$$t = \frac{2Y_0}{Y_0 m_{11} + Y_0 Y_s m_{12} + m_{21} + Y_s m_{22}} \quad (9.98)$$

To find either r or t for any configuration of films, we need only compute the characteristic matrices for each film, multiply them, and then substitute the resulting matrix elements into the above equations.

9.7.2 Antireflection Coatings

Now consider the extremely important case of normal incidence, that is,

$$\theta_{iI} = \theta_{tII} = \theta_{tII} = 0$$

which in addition to being the simplest, is also quite frequently approximated in practical situations. If we put a subscript on r to indicate the number of layers present, the reflection coefficient for a single film becomes

$$r_1 = \frac{n_1(n_0 - n_s)\cos k_0 h + i(n_0 n_s - n_1^2)\sin k_0 h}{n_1(n_0 + n_s)\cos k_0 h + i(n_0 n_s + n_1^2)\sin k_0 h} \quad (9.99)$$

Multiplying r_1 by its complex conjugate leads to the reflectance

$$R_1 = \frac{n_1^2(n_0 - n_s)^2 \cos^2 k_0 h + (n_0 n_s - n_1^2)^2 \sin^2 k_0 h}{n_1^2(n_0 + n_s)^2 \cos^2 k_0 h + (n_0 n_s + n_1^2)^2 \sin^2 k_0 h} \quad (9.100)$$

This formula becomes particularly simple when $k_0 h = \frac{1}{2}\pi$, which is equivalent to saying that the optical thickness h of the film is an odd multiple of $\frac{1}{4}\lambda_0$. In this case $d = \frac{1}{4}\lambda_f$, and

$$R_1 = \frac{(n_0 n_s - n_1^2)^2}{(n_0 n_s + n_1^2)^2} \quad (9.101)$$

which, quite remarkably, will equal zero when

$$n_1^2 = n_0 n_s \quad (9.102)$$

Generally, d is chosen so that h equals $\frac{1}{4}\lambda_0$ in the yellow-green portion of the visible spectrum, where the eye is most sensitive. Cryolite ($n = 1.35$), a sodium aluminum fluoride compound, and magnesium fluoride ($n = 1.38$) are common low-index films. Since MgF$_2$ is by far the more durable, it is used more frequently. On a glass substrate, ($n_s \approx 1.5$), both these films have indices that are still somewhat too large to satisfy Eq. (9.102). Nonetheless, a single $\frac{1}{4}\lambda_0$ layer of MgF$_2$ will reduce the reflectance of glass from about 4% to a bit more than 1%, over the visible spectrum. It is now common practice to apply antireflection coatings to the elements of optical instruments. On camera lenses, such coatings produce a decrease in the haziness caused by stray internally scattered light, as well as a marked increase in image brightness. At wavelengths on either side of the central yellow-green region, R increases and the lens surface will appear blue-red in reflected light.

For a double-layer, quarter-wavelength antireflection coating,

$$\mathcal{M} = \mathcal{M}_I \mathcal{M}_{II}$$

or more specifically

$$\mathcal{M} = \begin{bmatrix} 0 & i/Y_1 \\ iY_1 & 0 \end{bmatrix} \begin{bmatrix} 0 & i/Y_2 \\ iY_2 & 0 \end{bmatrix} \quad (9.103)$$

At normal incidence this becomes

$$\mathcal{M} = \begin{bmatrix} -n_2/n_1 & 0 \\ 0 & -n_1/n_2 \end{bmatrix} \quad (9.104)$$

Substituting the appropriate matrix elements into Eq. (9.97) yields r_2, which, when squared, leads to the reflectance
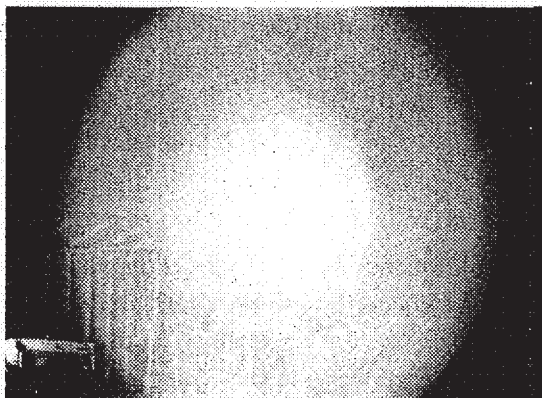
$$R_2 = \left[\frac{n_2^2 n_0 - n_s n_1^2}{n_2^2 n_0 + n_s n_1^2} \right]^2 \qquad (9.105)$$

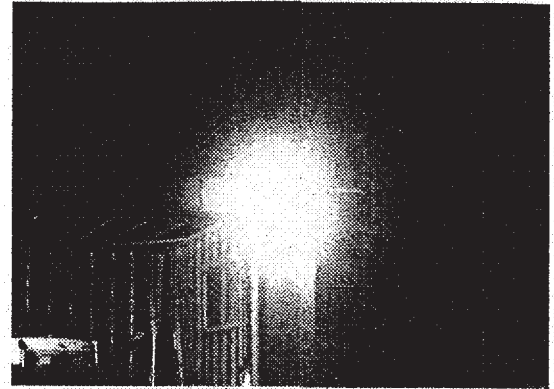For R_2 to be exactly zero at a particular wavelength, we need

$$\left(\frac{n_2}{n_1} \right)^2 = \frac{n_s}{n_0} \qquad (9.106)$$

This kind of film is referred to as a *double-quarter, single-minimum* coating. When n_1 and n_2 are as small as possible, the reflectance will have its single broadest minimum equal to zero at the chosen frequency. It should be clear from Eq. (9.106) that $n_2 > n_1$; accordingly, it is now common practice to designate a (glass)–(high index)–(low index)–(air) system as $gHLa$. Zirconium dioxide ($n = 2.1$), titanium dioxide ($n = 2.40$), and zinc sulfide ($n = 2.32$) are commonly used for H-layers, and magnesium fluoride ($n = 1.38$) and cerium fluoride ($n = 1.63$) often serve as L-layers.

Other double- and triple-layer schemes can be designed to satisfy specific requirements for spectral response, incident angle, cost, and so on. The accompanying photo is a scene photographed through a 15-element zoom lens, with a 150-W lamp pointing directly into the camera. The lens elements were covered with a single layer of MgF_2. When a triple-layer antireflection coating is used (see photo), the improved contrast and glare reduction are apparent.

Lens elements coated with a multilayer film structure. (Photo courtesy Optical Coating Laboratory, Inc., Santa Rosa, CA.)

9.7.3 Multilayer Periodic Systems

The simplest kind of periodic system is the *quarter-wave stack*, which is made up of a number of quarter-wave layers. The periodic structure of alternately high- and low-index materials, illustrated in Fig. 9.50, is designated by

$$g(HL)^3 a$$

Figure 9.51 illustrates the general form of a portion of the spectral reflectance for a few multilayer filters. The width of
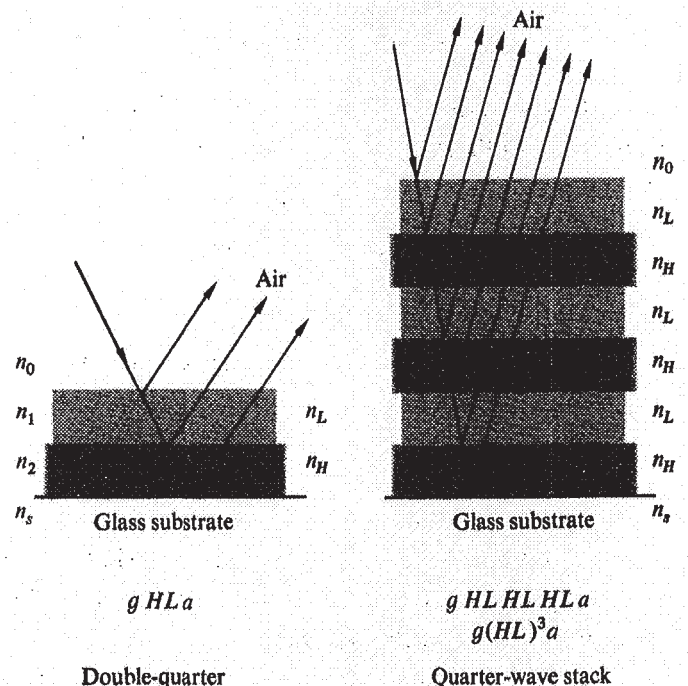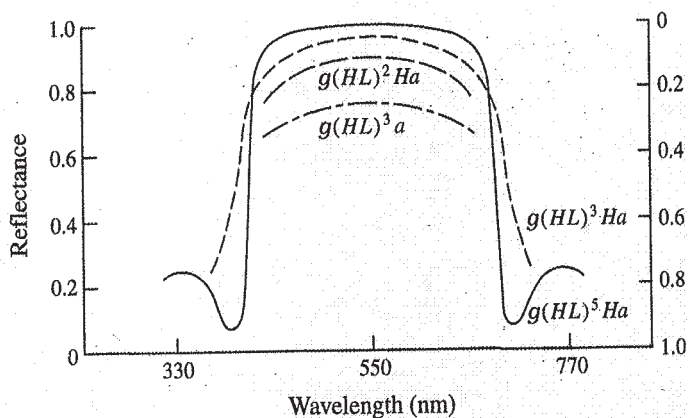

Lens elements coated with a single layer of MgF_2. (Photo courtesy Optical Coating Laboratory, Inc. Santa Rosa, CA.)

Figure 9.50 A periodic structure. Refraction has been omitted for simplicity.

Figure 9.51 Reflectance and transmittance for several periodic structures.

the high-reflectance central zone increases with increasing values of the index ratio n_H/n_L, and its height increases with the number of layers. Note that the maximum reflectance of a periodic structure such as $g(HL)^m a$ can be increased further by adding another H-layer, so that it has the form $g(HL)^m Ha$. Mirror surfaces with very high reflectance can be produced using this arrangement.

The small peak on the short-wavelength side of the central zone can be decreased by adding an eighth-wave low-index film to both ends of the stack, in which case the whole arrangement will be denoted by

$$g(0.5L)(HL)^m H(0.5L)a$$

This has the effect of increasing the short-wavelength high-frequency transmittance and is therefore known as a *high-pass filter*. Similarly, the structure

$$g(0.5H)L(HL)^m(0.5H)a$$

merely corresponds to the case in which the end H-layers are $\lambda_0/8$ thick. It has a higher transmittance at the long-wavelength, low-frequency range and serves as a *low-pass filter*.

At nonnormal incidence, up to about 30°, there is quite frequently little degradation in the response of thin-film coatings. In general, the effect of increasing the incident angle is a shift in the whole reflectance curve down to slightly shorter wavelengths. This kind of behavior is evidenced by several naturally occurring periodic structures, for example, peacock and hummingbird feathers, butterfly wings, and the backs of several varieties of beetles.

The last multilayer system to be considered is the *interference*, or more precisely the *Fabry–Perot, filter*. If the separa-

tion between the plates of an etalon is of the order of λ, the transmission peaks will be widely separated in wavelength. It will then be possible to block all the peaks but one by using absorbing filters of colored glass or gelatin. The transmitted light corresponds to a single sharp peak, and the etalon serves as a narrow band-pass filter. Such devices can be fabricated by depositing a semitransparent metal film onto a glass support, followed by a MgF_2 spacer and another metal coating.

All-dielectric, essentially nonabsorbing Fabry–Perot filters have an analogous structure, two possible examples of which are

$$g \; HLH \; LL \; HLH \; a$$

and

$$g \; HLHL \; HH \; LHLH \; a$$

The characteristic matrix for the first of these is

$$\mathcal{M} = \mathcal{M}_H \mathcal{M}_L \mathcal{M}_H \mathcal{M}_L \mathcal{M}_L \mathcal{M}_H \mathcal{M}_L \mathcal{M}_H$$

but from Eq. (9.104)

$$\mathcal{M}_L \mathcal{M}_L = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

or

$$\mathcal{M}_L \mathcal{M}_L = -\mathcal{I}$$

where $\mathcal{I}$ is the unity or identity matrix. The central double layer, corresponding to the Fabry–Perot cavity, is a half-wavelength thick ($d = \frac{1}{2}\lambda_f$). It therefore has no effect on the reflectance *at the particular wavelength under consideration*. Thus, it is said to be an absentee layer, and as a consequence,

$$\mathcal{M} = -\mathcal{M}_H \mathcal{M}_L \mathcal{M}_H \mathcal{M}_H \mathcal{M}_L \mathcal{M}_H$$

The same conditions prevail over and over again at the center and will finally result in
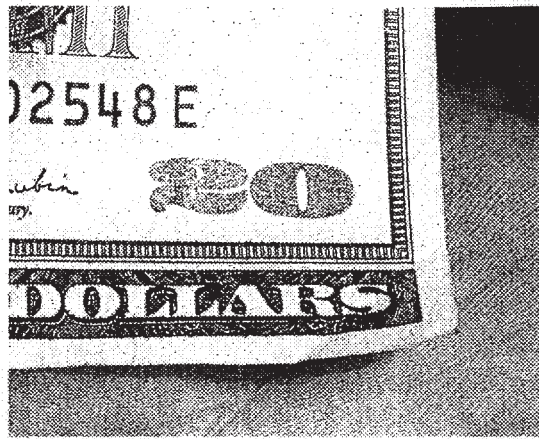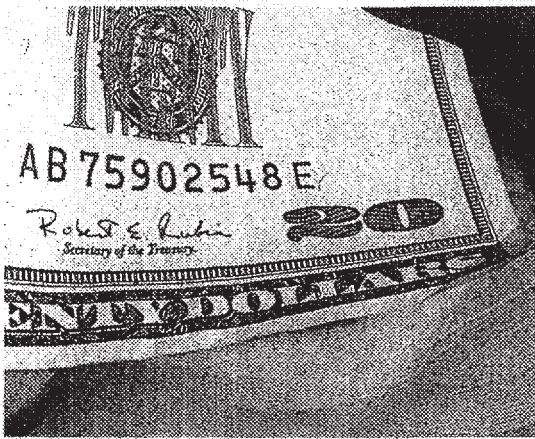
$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

At the special frequency for which the filter was designed, r at normal incidence, according to Eq. (9.97), reduces to

$$r = \frac{n_0 - n_s}{n_0 + n_s}$$

the value for the uncoated substrate. In particular, for glass ($n_s = 1.5$), in air ($n_0 = 1$) the theoretical peak transmission is

In Fig. 9.16 the optical path length difference depends on λ (i.e., the color of the light) and on the viewing angle. In a similar way the ink used to print the denominations on U.S. currency now contains structured particles that produce interference colors. The ink is infused with tiny flakes all oriented in the same direction. Each flake is a multilayered interference filter. Here the number 20 changes from black to green as the viewing angle changes.

96% (neglecting reflections from the back surface of the substrate, as well as losses in both the blocking filter and the films themselves).

9.8 Applications of Interferometry

There have been many physical applications of the principles of interferometry. Some of these are only of historical or pedagogical significance, whereas others are now being used extensively. The advent of the laser and the resultant availability of highly coherent quasimonochromatic light have made it particularly easy to create new interferometer configurations.

9.8.1 Scattered-Light Interference

Probably the earliest recorded study of interference fringes arising from scattered light is to be found in Sir Isaac Newton's *Optiks* (1704, Book Two, Part IV). Our present interest in this phenomenon is twofold. First, it provides an extremely easy way to see some rather beautiful colored interference fringes. Second, it is the basis for a remarkably simple and highly useful interferometer.

To see the fringes, lightly rub a thin layer of ordinary talcum powder onto the surface of any common back-silvered mirror (dew will do as well). Neither the thickness nor the uniformity of the coating is particularly important. The use of a bright point source, however, is crucial. A satisfactory source can be made by taping a heavy piece of cardboard having a hole about $\frac{1}{4}$ inch in diameter over a good flashlight. Initially, stand back from the mirror about 3 or 4 feet; the fringes will be too fine and closely spaced to see if you stand much nearer.

Hold the flashlight alongside your cheek and illuminate the mirror so that you can see the brightest reflection of the bulb in it. The fringes will then be clearly seen as a number of alternately bright and dark bands.

In Fig. 9.52 two coherent rays leaving the point source are shown arriving at point P after traveling different routes. One ray is reflected from the mirror and then scattered by a single transparent talcum grain toward P. The second ray is first scattered downward by the grain, after which it crosses the mirror
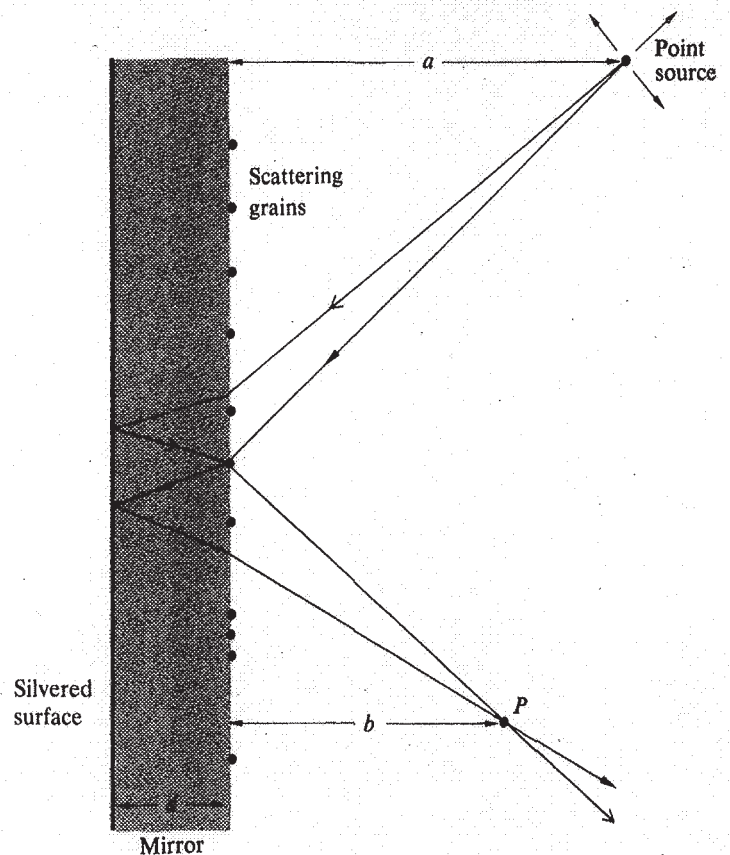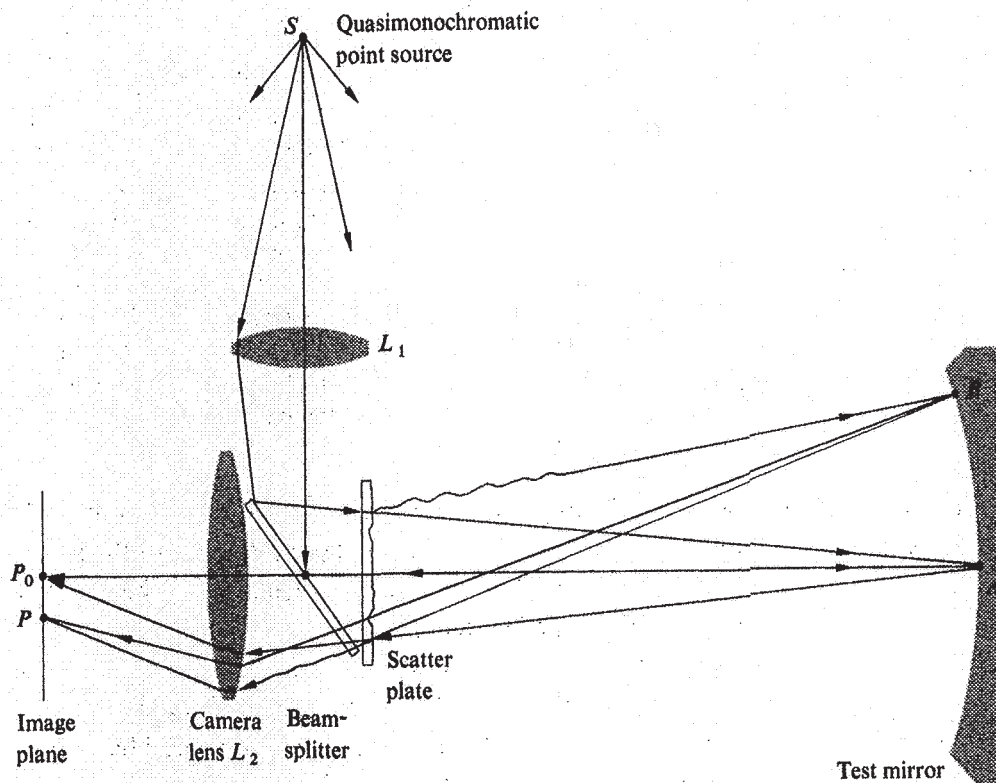

Figure 9.52 Interference of scattered light.

Figure 9.53 Scatter plate setup. Adapted from R. M. Scott, *Appl. Opt.* **8**, 531 (1969).

and is reflected back toward P. The resulting optical path length difference determines the interference at P. At normal incidence, the pattern is a series of concentric rings of radius*

$$\rho \approx \left[\frac{nm\lambda a^2 b^2}{d(a^2 - b^2)} \right]^{1/2}$$

Now consider a related device, which is very useful in testing optical systems. Known as a **scatter plate**, it generally consists of a slightly rough-surfaced, transparent sheet. In an arrangement such as the one shown in Fig. 9.53, it serves as an amplitude-splitting element. In this application it must have a center of symmetry; that is, each scattering site is required to have a duplicate, symmetrically located about a central point.

In the system under consideration, a point source of quasi-monochromatic light S is imaged, by means of lens L_1 on the surface, at point A of the mirror being tested. A portion of the light coming from the source is scattered by the scatter plate and thereafter illuminates the entire surface of the mirror. The mirror, in turn, reflects light back to the scatter plate. This

wave, as well as the light forming the image of the pinhole at point A, passes through the scatter plate again and finally reaches the image plane (either on a screen or in a camera). Fringes are formed on this latter plane. The interference process, which is manifest in the formation of these fringes, occurs because each point in the final image plane is illuminated by light arriving via two dissimilar routes, one originating at A and the other at some point B, which reflects scattered light. Indeed, as strange as they may look at first sight, well-defined fringes do result (see photo).

Examining the passage of light through the system in a bit more detail, consider the light initially incident on the scatter plate and assume that the wave is planar, as shown in Fig. 9.54. After it passes through the scatter plate, the incident plane wavefront $\vec{E}_i$ will be distorted into a transmitted wavefront $\vec{E}_T$. We envision this wave, in turn, split into a series of Fourier components consisting of plane waves, that is,

$$\vec{E}_T = \vec{E}_1 + \vec{E}_2 + \cdots \tag{9.107}$$

Two of these constituents are shown in Fig. 9.54a. Now suppose we attach a specific meaning to these components; namely, $\vec{E}_1$ is taken to represent the light traveling to the point A in Fig. 9.53, and $\vec{E}_2$ that traveling toward B. The analysis of the

*For more of the details, see A. J. deWitte, "Interference in Scattered Light," *Am. J. Phys.* **35**, 301 (1967).
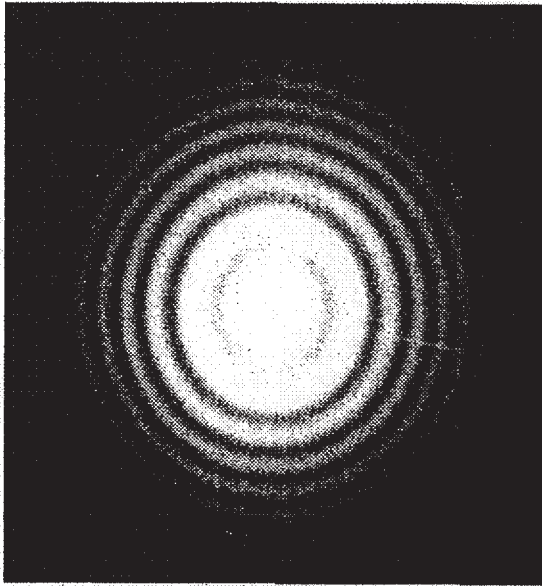
Fringes in scattered light.

The wavefront returning from B to the scatter plate is denoted by $\vec{\mathbf{E}}_B$ in Fig. 9.54c. Upon traversing the scatter plate, it will be reshaped into the wave $\vec{\mathbf{E}}_{BT}$. One of the Fourier components of this wavefront, denoted by $\vec{\mathbf{E}}_{B\theta}$, is inclined at the angle θ and will therefore be focused at the same point P on the screen.

Some of the waves arriving at P will be coherent in the sense that interference occurs. To obtain the resultant irradiance I_P, first add the amplitudes of all the waves arriving at P, that is, $\vec{\mathbf{E}}_P$, and then square and time average $\vec{\mathbf{E}}_P$.

In the discussion above, only two point sources at the mirror were considered. Actually, of course, the whole surface of the mirror is illuminated by the ongoing light, and every point of it will serve as a secondary source of returning waves. All the waves will be deformed by the scatter plate, and these, in turn, can be split into plane-wave components. In each series of component waves, there will be one inclined at an angle θ, and all of these will be focused at the same point P on the screen. The resultant amplitude will then have the form

$$\vec{\mathbf{E}}_P = \vec{\mathbf{E}}_{A\theta} + \vec{\mathbf{E}}_{B\theta} + \cdots$$

The light reaching the image plane can be envisioned as made up in part of two optical fields of special interest. One of these results from light that was scattered only on its passage through the plate toward the mirror, and the other results from light that was scattered only on the way toward the image plane. The former broadly illuminates the test mirror and ultimately results in an image of it on the screen. The latter, which was initially focused to the region about A, scatters a diffuse

stages that follow could be continued in the same way. Let the portion of the wavefront returning from A be represented by the wavefront $\vec{\mathbf{E}}_A$ in Fig. 9.54b. The scatter plate will transform it into an irregular transmitted wave, denoted by $\vec{\mathbf{E}}_{AT}$ in the same figure. This again corresponds to a complicated configuration, but it can be split into Fourier components consisting of plane waves, as in the above case. In Fig. 9.54b, two of these component wavefronts have been drawn, one traveling to the left, and the other inclined at an angle θ. The latter wavefront, which is denoted by $\vec{\mathbf{E}}_{A\theta}$, is focused by lens L_2 at the point P on the screen (Fig. 9.53).

$$\mathbf{E}_i \qquad \mathbf{E}_T \quad \mathbf{E}_1 \quad \mathbf{E}_2$$
(a)

$$\mathbf{E}_{A\theta} \; \mathbf{E}_{AT} \qquad \mathbf{E}_A$$
(b)

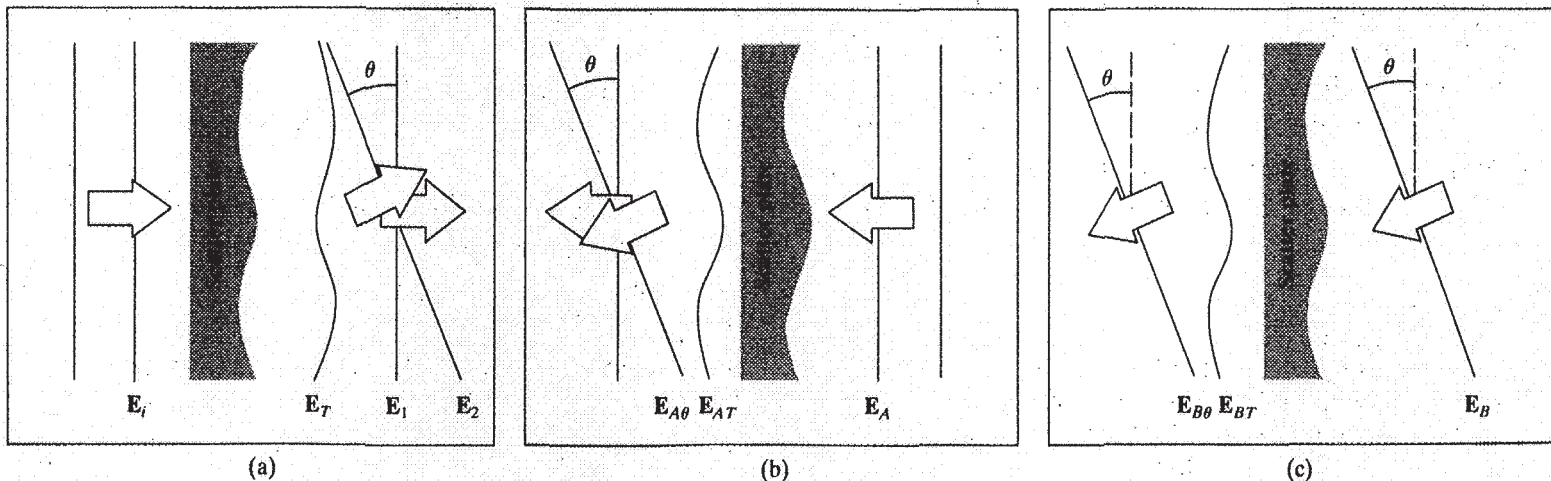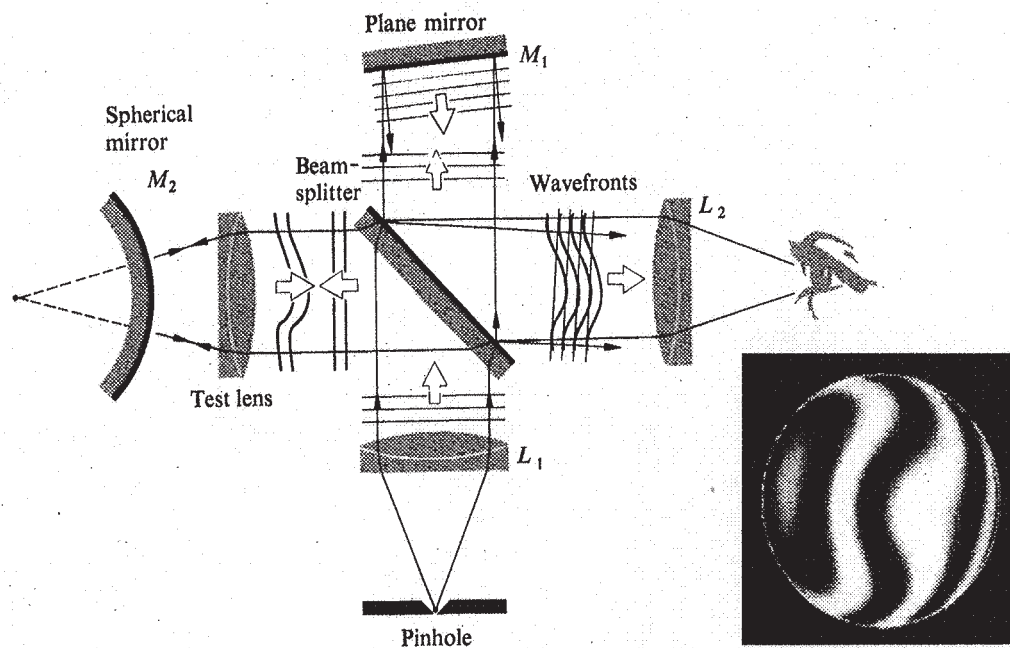$$\mathbf{E}_{B\theta} \; \mathbf{E}_{BT} \qquad \mathbf{E}_B$$
(c)

Figure 9.54 Wavefronts passing through the scatter plate.

Figure 9.55 The Twyman–Green interferometer. (Photo by E.H.)

blur across the screen. The point A is chosen so that the small area in the vicinity of it is free of aberrations. In that case, the wave reflected from it serves as a reference with which to compare the wavefront corresponding to the entire mirror surface. The interference pattern will show, as a series of contour fringes, any deviations from perfection in the mirror surface.*

9.8.2 The Twyman-Green Interferometer

The Twyman–Green is essentially a variation of the Michelson Interferometer. It's an instrument of great importance in the domain of modern optical testing. Among its distinguishing physical characteristics (illustrated in Fig. 9.55) are a quasimonochromatic point source and lens L_1, to provide a source of incoming *plane waves*, and a lens L_2, which permits all the light from the aperture to enter the eye so that the entire field can be seen, that is, any portion of M_1 and M_2. A continuous laser serves as a superior source in that it provides the

convenience of long path length differences and, in addition, short photographic exposure times. These tend to minimize unwanted vibration effects. Laser versions of the Twyman–Green are among the most effective testing tools in Optics. As shown in the figure, the device is set up to examine a lens. The spherical mirror M_2 has its center of curvature coincident with the focal point of the lens. If the lens being tested is free of aberrations, the emerging reflected light returning to the beamsplitter will again be a plane wave. If, however, astigmatism, coma, or spherical aberration deforms the wavefront, a fringe pattern clearly manifesting these distortions can be seen and photographed. When M_2 is replaced by a plane mirror, a number of other elements (prisms, optical flats, etc.) can be tested equally well. The optician interpreting the fringe pattern can then mark the surface for further polishing to correct high or low spots. In the fabrication of the finest optical systems, telescopes, high-altitude cameras, and so forth, the interferograms may even be scanned electronically, and the resulting data analyzed by computer. Computer-controlled plotters can then automatically produce surface contour maps or perspective "three-dimensional" drawings of the distorted wavefront generated by the element being tested. These procedures can be used throughout the fabrication process to ensure the highest quality optical instruments. Complex systems with wavefront aberrations in the fractional-wavelength range are the result of what might be called the new technology.

*For further discussion of the scatter plate, the reader might consult the rather succinct papers by J. M. Burch, *Nature* **171**, 889 (1953), and *J. Opt. Soc. Am.* **52**, 600 (1962). Reference should be made to J. Strong, *Concepts of Classical Optics*, p. 383. Also see R. M. Scott, "Scatter Plate Interferometry," *Appl. Opt.* **8**, 531 (1969), and J. B. Houston, Jr., "How to Make and Use a Scatterplate Interferometer," *Optical Spectra* (June 1970), p. 32.
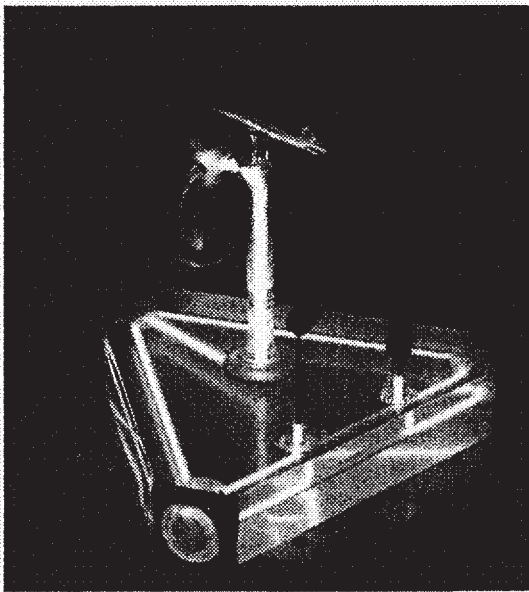
9.8.3 The Rotating Sagnac Interferometer

The Sagnac Interferometer is widely used to measure rotational speed. In particular, the ***ring laser***, which is essentially a Sagnac Interferometer containing a laser in one or more of its arms, was designed specifically for that purpose. The first ring laser gyroscope was introduced in 1963, and work is continuing on various devices of this sort (see photo). The initial experiments that gave impetus to these efforts were performed by Sagnac in 1911. At that time he rotated the entire interferometer, mirrors, source, and detector, about a perpendicular axis passing through its center (Fig. 9.56). Recall, from Section 9.4.2, that two overlapping beams traverse the interferometer, one clockwise, the other counterclockwise. The rotation effectively shortens the path taken by one beam in comparison to that of the other. In the interferometer, the result is a fringe shift proportional to the angular speed of rotation ω. In the ring laser, it is a frequency difference between the two beams that is proportional to ω.

Consider the arrangement depicted in Fig. 9.56. The corner A (and every other corner) moves with a linear speed $v = R\omega$, where R is half the diagonal of the square. Using classical reasoning, we find that the time of travel of light along AB is

$$t_{AB} = \frac{R\sqrt{2}}{c - v/\sqrt{2}}$$

An early ring laser gyro. (Photo courtesy Autonetics, a Division of Boeing North America, Inc.)
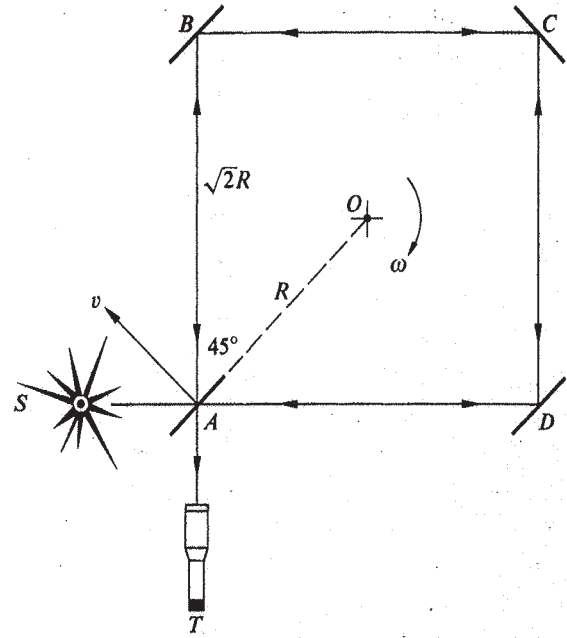

Figure 9.56 The rotating Sagnac Interferometer. Originally it was 1 m × 1 m with $\omega = 120$ rev/min.

or

$$t_{AB} = \frac{2R}{\sqrt{2}c - \omega R}$$

The time of travel of the light from A to D is

$$t_{AD} = \frac{2R}{\sqrt{2}c + \omega R}$$

The total time for counterclockwise and clockwise travel is given respectively by

$$t_\circlearrowleft = \frac{8R}{\sqrt{2}c + \omega R}$$

and

$$t_\circlearrowright = \frac{8R}{\sqrt{2}c - \omega R}$$

For $\omega R << c$ the difference between these two intervals is

$$\Delta t = t_\circlearrowright - t_\circlearrowleft$$

or, using the Binomial Series,

$$\Delta t = \frac{8R^2\omega}{c^2}$$

This can be expressed in terms of area $A = 2R^2$ of the square formed by the beams of light as

$$\Delta t = \frac{4A\omega}{c^2}$$

Let the period of the monochromatic light used be $\tau = \lambda/c$; then the fractional displacement of the fringes, given by $\Delta N = \Delta t/\tau$, is

$$\Delta N = \frac{4A\omega}{c\lambda}$$

a result that has been verified experimentally. In particular, Michelson and Gale* used this method to determine the angular velocity of the Earth.

The preceding classical treatment is obviously lacking, inasmuch as it assumes speeds in excess of c, an assumption that is contrary to the dictates of Special Relativity. Furthermore, it would appear that since the system is accelerating, General Relativity would prevail. In fact, these formalisms yield the same results.

9.8.4 Radar Interferometry

In February 2000 the Space Shuttle Endeavour completed a mission to create a "three-dimensional" map of the Earth covering 119 million square kilometers. The feat was accomplished using synthetic aperture radar (SAR). In general, the larger the aperture of a viewing system, the greater the resolution (p. 471) and the more details one can see. SAR is a technique for using the motion of an airplane or spacecraft along with signal processing methods to simulate a large antenna.

Using a phased array antenna (p. 98), the Shuttle swept a radar beam back and forth perpendicular to its line of motion painting a 225-km wide swath over the Earth's surface (Fig. 9.57). Orbiting upside-down, Endeavour extended a 60-m mast with two receiving antennas at its end (Fig. 9.58). The SAR then sent out a stream of about 1700 high-powered electromagnetic pulses per second from its main antenna in the cargo bay, which was both a transmitter and receiver. Actually, the mission utilized two different radars: a C-band system operating at a wavelength of 5.6 cm that provided most of the coverage, and a higher resolution X-band 3-cm system that gave a detailed view of a narrow 50-km swath (Fig. 9.57). A radar image is made up of countless tiny uniform dots known as pixels (p. 473). The pixel is the smallest bit of information in the picture—nothing can be seen that's smaller than a single pixel. For the main C-band system, each pixel is about 12.5 m in diameter, and the smallest object that can be resolved is about 30 m across.

*Michelson and Gale, *Astrophys. J.* **61**, 140 (1925).
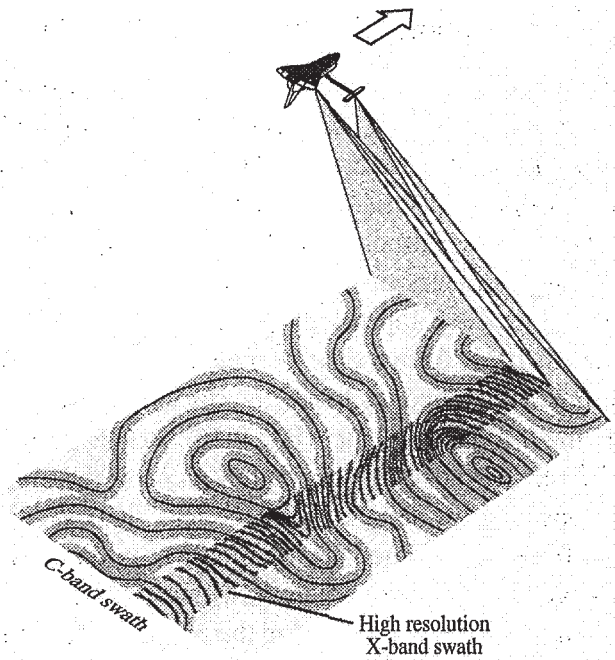
Figure 9.57 As the Shuttle orbited, its two radar systems swept out a swath across the surface of the Earth.

Ordinarily, a radar system sends out a pulse (with a pulsewidth of 10-50 μs), and then, picking up the backscattered wave, it records both the amplitude and round-trip time. That gives a rough idea of the size and location of the target. However, in order to gather data about the elevation of surface features on the Earth, the Shuttle Radar Topography Mission (SRTM) utilized interferometry—in a way that suggests
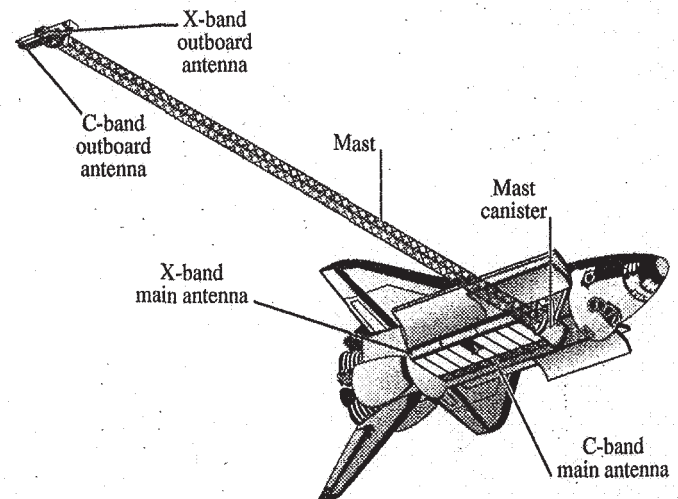
Figure 9.58 The Shuttle Endeavour carried the main C-band transmitter-receiver antenna in its cargo bay and a second receiver at the end of a 60-m long mast.

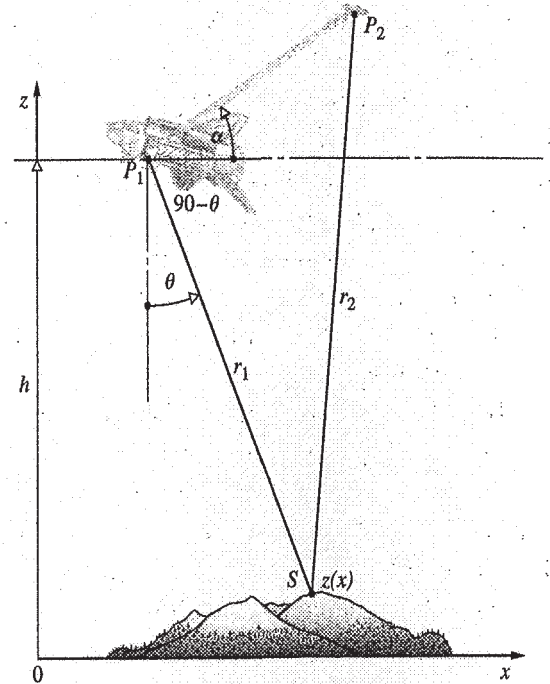**Figure 9.59** A radar pulse emitted from the Shuttle strikes the ground and reflects back. The echo is picked up by both the outboard and inboard antennas.

Figure 9.60 The basic geometry of the SAR interferometer. The source point S is the spot on the ground that reflects the radar pulse back to the Shuttle. Points P_1 and P_2 correspond to the two receivers, one on the mast and the other in the cargo bay of the Shuttle.
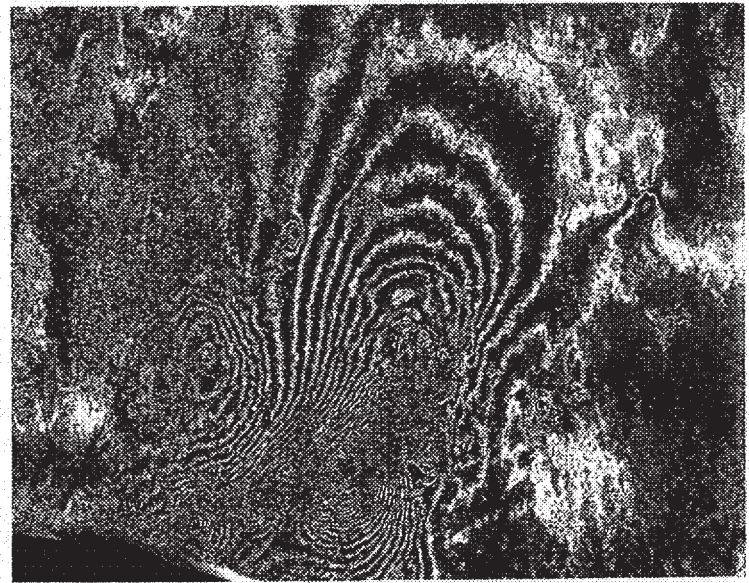
Young's Experiment run backwards (p. 393). In any event, similar interferometric techniques are of growing importance in radio and optical astronomy.

The SAR is a coherent imaging system, and it retains information about both the amplitude and phase of the radar echo during data acquisition and processing. A signal is emitted from the Shuttle (much like the flash from an ordinary camera but spectrally more controlled); it strikes the ground (Fig. 9.59) and returns to the two antennas, one in the cargo bay (P_1), the other on the boom (P_2). These are separated by a 60-m baseline a. The two radar echoes are converted into digital data which are recorded for later processing and display as an image. It's left for Problem 48 (Fig. 9.60) to show that the topography in the form of the function $z(x)$ can be expressed in terms of the altitude h, the *look angle* of the radar θ, and measured phase angle difference, or *interferometric phase* ϕ between the two signals;

$$z(x) = h - \frac{(\lambda\phi/2\pi)^2 - a^2}{2a\sin(\alpha - \theta) - (\lambda\phi/2\pi)}\cos\theta \quad (9.108)$$

Synthetic aperture radar was used to produce this interferogram arising from the June 1992 earthquake in Landers, California. Images taken by the ERS-1 satellite before and after the quake were combined to generate this fringe pattern which reveals the shift in the ground that took place. The picture covers an area of about 125 by 175 km. (Photo courtesy Didier Massonnet, Centre National d'Etudes Spatiales, France.)

An interferometer of this sort measures ϕ, the difference in phase between the signals arriving at the ends of its baseline. It does this by analytically interfering those signals using a process called *cross correlation* (p. 545). When the two separate data sets, one from each antenna, are combined on the ground the first thing produced is an interferogram or fringe map (see photo) which encodes the topography. The interferogram corresponds to a collection of "fringes of equal height," or if you will, contours of equal height. But the information needs further refining; the elevations of the contours are unknown. Based on accurate knowledge of the mast length and orientation, the height of each contour, $z(x)$, is determined, essentially via triangulation. Data collected over the oceans provides a sea-level reference for all elevations. After a considerable amount of computation, pixel by pixel, a 3-D topographical map is finally created (see photo).

These are radar images of San Andreas, California taken by the Space Shuttle Endeavour in 2000. The picture on the left (which looks a lot better in color) shows an interferogram overlaying the terrain; the picture on the right is the corresponding "three-dimensional" map that results from the analysis of all of the data.

PROBLEMS

Complete solutions to all problems—except those with an asterisk—can be found in the back of the book.

9.1 Returning to Section 9.1, let

$$\tilde{\mathbf{E}}_1(\vec{r}, t) = \tilde{\mathbf{E}}_1(\vec{r})e^{-i\omega t}$$

and

$$\tilde{\mathbf{E}}_2(\vec{r}, t) = \tilde{\mathbf{E}}_2(\vec{r})e^{-i\omega t}$$

where the wavefront shapes are not explicitly specified, and $\tilde{\mathbf{E}}_1$ and $\tilde{\mathbf{E}}_2$ are complex vectors depending on space and initial phase angle. Show that the interference term is then given by

$$I_{12} = \tfrac{1}{2}(\tilde{\mathbf{E}}_1 \cdot \tilde{\mathbf{E}}_2^* + \tilde{\mathbf{E}}_1^* \cdot \tilde{\mathbf{E}}_2) \qquad (9.109)$$

You will have to evaluate terms of the form

$$\langle \tilde{\mathbf{E}}_1 \cdot \tilde{\mathbf{E}}_2 e^{-2i\omega t} \rangle_T = (\tilde{\mathbf{E}}_1 \cdot \tilde{\mathbf{E}}_2/T) \int_t^{t+T} e^{-2i\omega t'} dt'$$

for $T \gg \tau$ (take another look at Problem 3.10). Show that Eq. (9.109) leads to Eq. (9.11) for plane waves.

9.2 In Section 9.1 we considered the spatial distribution of energy for two point sources. We mentioned that for the case in which the separation $a \gg \lambda$, I_{12} spatially averages to zero. Why is this true? What happens when a is much less than λ?

9.3* Return to Fig. 2.22 and prove that if two electromagnetic plane waves making an angle θ have the same amplitude, E_θ, the resulting interference pattern on the yx-plane is a cosine-squared irradiance distribution given by

$$I(y) = 4E_0^2 \cos^2\left(\frac{\pi}{\lambda} y \sin\theta\right)$$

Locate the zeros of irradiance. What is the value of the fringe separation? What happens to the separation as θ increases? Compare your analysis with that leading to Eq. (9.17). [*Hint*: Begin with the wave expressions given in Section 2.7, which have the proper phases already worked out, and write them as exponentials.]

9.4 Will we get an interference pattern in Young's Experiment (Fig. 9.8) if we replace the source slit S by a single long-filament lightbulb? What would occur if we replaced the slits S_1 and S_2 by these same bulbs?

9.5* Figure P.9.5 shows an output pattern that was measured by a tiny microphone when two small piezo-loudspeakers separated by 15 cm were pointed toward the microphone at a distance of 1.5 m away. Given that the speed of sound at 20°C is 343 m/s, determine the approximate frequency at which the speakers were driven. Discuss the nature of the pattern and explain why it has a central minimum.

Figure P.9.5 (Data courtesy of CENCO.)

Position of Microphone (cm, zero-point arbitrary)
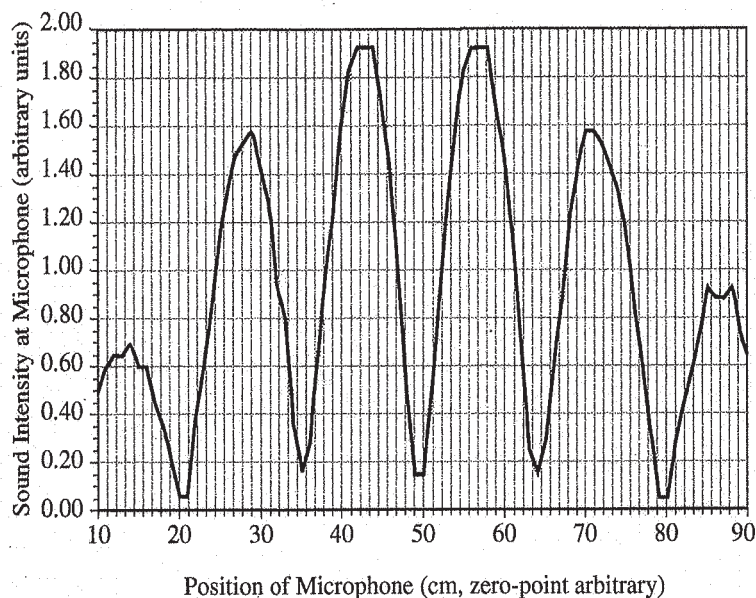
9.6* Two 1.0-MHz radio antennas emitting in-phase are separated by 600 m along a north–south line. A radio receiver placed 2.0 km east is equidistant from both transmitting antennas and picks up a fairly strong signal. How far north should that receiver be moved if it is again to detect a signal nearly as strong?

9.7 An expanded beam of red light from a He–Ne laser ($\lambda_0 = 632.8$ nm) is incident on a screen containing two very narrow horizontal slits separated by 0.200 mm. A fringe pattern appears on a white screen held 1.00 m away.

(a) How far (in radians and millimeters) above and below the central axis are the first zeros of irradiance?

(b) How far (in mm) from the axis is the fifth bright band?

(c) Compare these two results.

9.8* Red plane waves from a ruby laser ($\lambda_0 = 694.3$ nm) in air impinge on two parallel slits in an opaque screen. A fringe pattern forms on a distant wall, and we see the fourth bright band 1.0° above the central axis. Kindly calculate the separation between the slits.

9.9* A 3 × 5 card containing two pinholes, 0.08 mm in diameter and separated center to center by 0.10 mm, is illuminated by parallel rays of blue light from an argon ion laser ($\lambda_0 = 487.99$ nm). If the fringes on an observing screen are to be 10 mm apart, how far away should the screen be?

9.10* White light falling on two long narrow slits emerges and is observed on a distant screen. If red light ($\lambda_0 = 780$ nm) in the first-order fringe overlaps violet in the second-order fringe, what is the latter's wavelength?

9.11* Considering the double-slit experiment, derive an equation for the distance $y_{m'}$ from the central axis to the m'th irradiance *minimum*, such that the first dark bands on either side of the central maximum correspond to $m' = \pm 1$. Identify and justify all your approximations.

9.12* With regard to Young's Experiment, derive a general expression for the shift in the vertical position of the mth *maximum* as a result of placing a thin parallel sheet of glass of index n and thickness d directly over one of the slits. Identify your assumptions.

9.13* Plane waves of monochromatic light impinge at an angle θ_i on a screen containing two narrow slits separated by a distance a. Derive an equation for the angle measured from the central axis which locates the mth maximum.

9.14* Sunlight incident on a screen containing two long narrow slits 0.20 mm apart casts a pattern on a white sheet of paper 2.0 m beyond. What is the distance separating the violet ($\lambda_0 = 400$ nm) in the first-order band from the red ($\lambda_0 = 600$ nm) in the second-order band?

9.15 To examine the conditions under which the approximations of Eq. (9.23) are valid:

(a) Apply the law of cosines to triangle S_1S_2P in Fig. 9.8c to get

$$\frac{r_2}{r_1} = \left[1 - 2\left(\frac{a}{r_1}\right)\sin\theta + \left(\frac{a}{r_1}\right)^2\right]^{1/2}$$

(b) Expand this in a Maclaurin series yielding

$$r_2 = r_1 - a\sin\theta + \frac{a^2}{2r_1}\cos^2\theta + \cdots$$

(c) In light of Eq. (9.17), show that if $(r_1 - r_2)$ is to equal $a\sin\theta$, it is required that $r_1 >> a^2/\lambda$.

9.16 A stream of electrons, each having an energy of 0.5 eV, impinges on a pair of extremely thin slits separated by 10^{-2} mm. What is the distance between adjacent minima on a screen 20 m behind the slits? ($m_e = 9.108 \times 10^{-31}$ kg, 1 eV $= 1.602 \times 10^{-19}$ J.)

9.17* It is our intention to produce interference fringes by illuminating some sort of arrangement (Young's Experiment, a thin film, the Michelson Interferometer, etc.) with light at a mean wavelength of 500 nm, having a linewidth of 2.5×10^{-3} nm. At approximately what optical path length difference can you expect the fringes to vanish? [*Hint:* Think about the coherence length and revisit Problem 7.39.]

9.18* Imagine that you have an opaque screen with three horizontal very narrow parallel slits in it. The second slit is a center-to-center

distance a beneath the first, and the third is a distance $5a/2$ beneath the first. (a) Write a complex exponential expression in terms of δ for the amplitude of the electric field at some point P at an elevation θ on a distant screen where $\delta = ka \sin \theta$. Prove that

$$I(\theta) = \frac{I(0)}{3} + \frac{2I(0)}{9}(\cos \delta + \cos 3\delta/2 + \cos 5\delta/2)$$

Verify that at $\theta = 0$, $I(\theta) = I(0)$.

9.19* Show that a for the Fresnel biprism of Fig. 9.13 is given by $a = 2d(n - 1)\alpha$.

9.20* In the Fresnel double mirror $s = 2$ m, $\lambda_0 = 589$ nm, and the separation of the fringes was found to be 0.5 mm. What is the angle of inclination of the mirrors, if the perpendicular distance of the actual point source to the intersection of the two mirrors is 1 m?

9.21* The Fresnel biprism is used to obtain fringes from a point source that is placed 2 m from the screen, and the prism is midway between the source and the screen. Let the wavelength of the light be $\lambda_0 = 500$ nm and the index of refraction of the glass be $n = 1.5$. What is the prism angle, if the separation of the fringes is 0.5 mm?

9.22 What is the general expression for the separation of the fringes of a Fresnel biprism of index n immersed in a medium having an index of refraction n'?

9.23 Using Lloyd's mirror, X-ray fringes were observed, the spacing of which was found to be 0.0025 cm. The wavelength used was 8.33 Å. If the source–screen distance was 3 m, how high above the mirror plane was the point source of X-rays placed?

9.24 Imagine that we have an antenna at the edge of a lake picking up a signal from a distant radio star (Fig. P.9.24), which is just coming up above the horizon. Write expressions for δ and for the angular position of the star when the antenna detects its first maximum.

Figure P.9.24
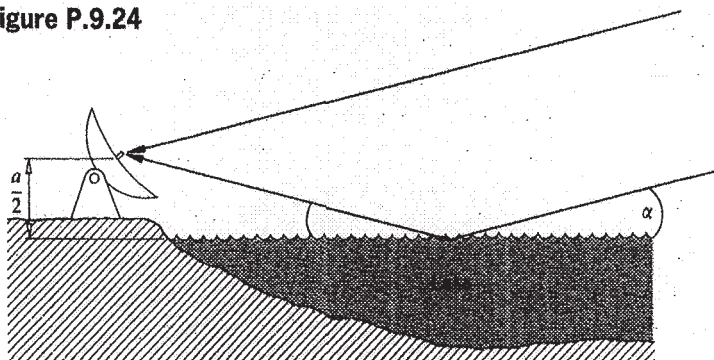
9.25* If the plate in Fig. 9.17 is glass in air, show that the amplitudes of E_{1r}, E_{2r}, and E_{3r} are respectively $0.2E_{0i}$, $0.192E_{0i}$, and $0.008E_{0i}$, where E_{0i} is the incident amplitude. Make use of the Fres-

nel coefficients at normal incidence, assuming no absorption. You might repeat the calculation for a water film in air.

9.26 A soap film surrounded by air has an index of refraction of 1.34. If a region of the film appears bright red ($\lambda_0 = 633$ nm) in normally reflected light, what is its minimum thickness there?

9.27* A thin film of ethyl alcohol ($n = 1.36$) spread on a flat glass plate and illuminated with white light shows a color pattern in reflection. If a region of the film reflects only green light (500 nm) strongly, how thick is it?

9.28* A soap film of index 1.34 has a region where it is 550.0 nm thick. Determine the vacuum wavelengths of the radiation that is not reflected when the film is illuminated from above with sunlight.

9.29 Consider the circular pattern of Haidinger's fringes resulting from a film with a thickness of 2 mm and an index of refraction of 1.5. For monochromatic illumination of $\lambda_0 = 600$ nm, find the value of m for the central fringe ($\theta_t = 0$). Will it be bright or dark?

9.30 Illuminate a microscope slide (or even better, a thin cover-glass slide). Colored fringes can easily be seen with an ordinary fluorescent lamp (although some of the newer versions don't work well at all) serving as a broad source or a mercury street light as a point source. Describe the fringes. Now rotate the glass. Does the pattern change? Duplicate the conditions shown in Figs. 9.18 and 9.19. Try it again with a sheet of plastic food wrap stretched across the top of a cup.

9.31 Figure P.9.31 illustrates a setup used for testing lenses. Show that

$$d = x^2(R_2 - R_1)/2R_1R_2$$

when d_1 and d_2 are negligible in comparison with $2R_1$ and $2R_2$, respectively. (Recall the theorem from plane geometry that relates the products of the segments of intersecting chords.) Prove that the radius of the mth dark fringe is then

$$x_m = [R_1R_2m\lambda_f/(R_2 - R_1)]^{1/2}$$

How does this relate to Eq. (9.43)?

9.32* Newton rings are observed on a film with quasimonochromatic light that has a wavelength of 500 nm. If the 20th bright ring has a radius of 1 cm, what is the radius of curvature of the lens forming one part of the interfering system?

9.33 Fringes are observed when a parallel beam of light of wavelength 500 nm is incident perpendicularly onto a wedge-shaped film with an index of refraction of 1.5. What is the angle of the wedge if the fringe separation is $\frac{1}{3}$ cm?

Figure P.9.31

Test plate

R_1

R_2

Figure P.9.36
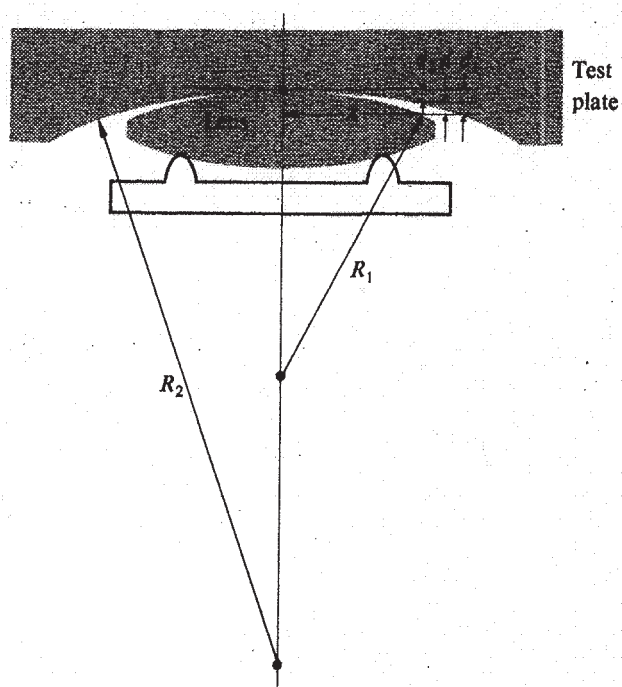
9.34* Suppose a wedge-shaped air film is made between two sheets of glass, with a piece of paper 7.618×10^{-5} m thick used as the spacer at their very ends. If light of wavelength 500 nm comes down from directly above, determine the number of bright fringes that will be seen across the wedge.

9.35 A Michelson Interferometer is illuminated with monochromatic light. One of its mirrors is then moved 2.53×10^{-5} m, and it is observed that 92 fringe-pairs, bright and dark, pass by in the process. Determine the wavelength of the incident beam.

9.36* One of the mirrors of a Michelson Interferometer is moved, and 1000 fringe-pairs shift past the hairline in a viewing telescope during the process. If the device is illuminated with 500-nm light, how far was the mirror moved?

9.37* Suppose we place a chamber 10.0 cm long with flat parallel windows in one arm of a Michelson Interferometer that is being illuminated by 600-nm light. If the refractive index of air is 1.000 29 and all the air is pumped out of the cell, how many fringe-pairs will shift by in the process?

9.38* A form of the Jamin Interferometer is illustrated in Fig. P.9.38. How does it work? To what use might it be put?

9.39 Starting with Eq. (9.53) for the transmitted wave, compute the flux density, that is, Eq. (9.54).

9.40 Given that the mirrors of a Fabry–Perot Interferometer have an
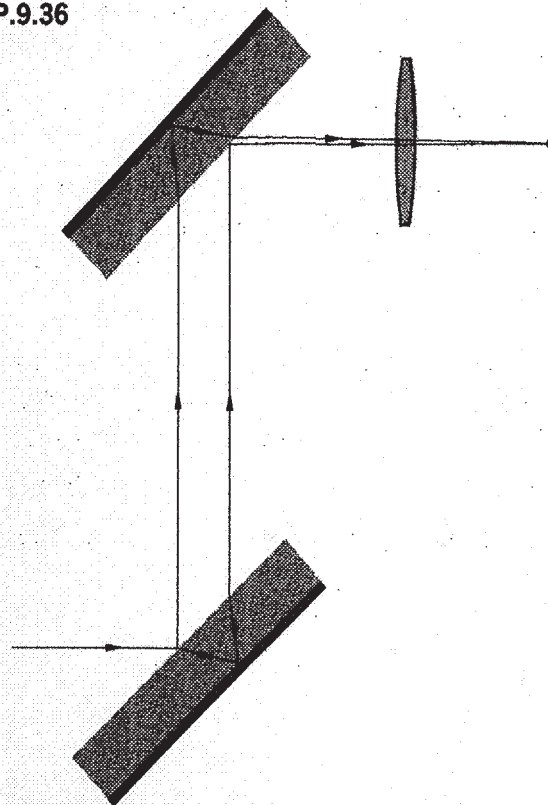
amplitude reflection coefficient of $r = 0.8944$, find

(a) the coefficient of finesse,

(b) the half-width,

(c) the finesse, and,

(d) the *contrast factor* defined by

$$C \equiv \frac{(I_t/I_i)_{max}}{(I_t/I_i)_{min}}$$

9.41 To fill in some of the details in the derivation of the smallest phase increment separating two resolvable Fabry–Perot fringes, that is,

$$(\Delta\delta) \approx 4.2/\sqrt{F} \qquad [9.73]$$

satisfy yourself that

$$[\mathscr{A}(\theta)]_{\delta \,=\, \delta_a \pm \Delta\delta/2} = [\mathscr{A}(\theta)]_{\delta \,=\, \Delta\delta/2}$$

Show that Eq. (9.72) can be rewritten as

$$2[\mathscr{A}(\theta)]_{\delta \,=\, \Delta\delta/2} = 0.81\{1 + [\mathscr{A}(\theta)]_{\delta \,=\, \Delta\delta}\}$$

When F is large γ is small, and $\sin(\Delta\delta) = \Delta\delta$. Prove that Eq. (9.73) then follows.

9.42 Consider the interference pattern of the Michelson Interferometer as arising from two beams of equal flux density. Using Eq.

(9.17), compute the half-width. What is the separation, in δ, between adjacent maxima? What then is the finesse?

9.43* Satisfy yourself of the fact that a film of thickness $\lambda_f/4$ and index n_1 will always reduce the reflectance of the substrate on which it is deposited, as long as $n_s > n_1 > n_0$. Consider the simplest case of normal incidence and $n_0 = 1$. Show that this is equivalent to saying that the waves reflected back from the two interfaces cancel one another.

9.44 Verify that the reflectance of a substrate can be increased by coating it with a $\lambda_f/4$, high-index layer, that is, $n_1 > n_s$. Show that the reflected waves interfere constructively. The quarter-wave stack $g(HL)^m Ha$ can be thought of as a series of such structures.

9.45 Determine the refractive index and thickness of a film to be deposited on a glass surface ($n_g = 1.54$) such that no normally incident light of wavelength 540 nm is reflected.

9.46 A glass microscope lens having an index of 1.55 is to be coated with a magnesium fluoride film to increase the transmission of normally incident yellow light ($\lambda_0 = 550$ nm). What minimum thickness should be deposited on the lens?

9.47* A glass camera lens with an index of 1.55 is to be coated with a cryolite film (n $\approx$ 1.30) to decrease the reflection of normally incident green light ($\lambda_0 = 500$ nm). What thickness should be deposited on the lens?

9.48* Using Fig. 9.60, which depicts the geometry of the Shuttle radar interferometer, show that

$$z(x) = h - r_1 \cos \theta$$

Then use the Law of Cosines to establish that Eq. (9.108) is correct.