16 APPLICATIONS OF PHYSICAL OPTICS

16.1 INTRODUCTION

16.1.1.1 Restatement of principles. In instruments for purposes of interferometry, the problems of geometrical optical designs are usually simple. However, since such instruments depend upon the interference of light waves for their proper functioning, a knowledge of the principles of interference is necessary for proper design. These principles have been already presented. A brief recapitulation of these principles is now presented, followed by detailed examples of their application to the design of several typical instruments of this class.

16.1.1.2 As stated in Section 3, the instantaneous magnitude of a plane-polarized light wave will be equivalent to the instantaneous magnitude of the electric vector and can be specified by the trigonometric function

$$E(z, t) = a \cos (knz + \phi - \omega t) \tag{1}$$

where

z = distance measured along Z
t = time
$k = 2\pi/\lambda$
$\omega = 2\pi/T$
λ = wavelength
T = period for one complete vibration
ϕ = phase angle
n = refractive index. It can be a function of z for variable media.
a = amplitude of the wave. It is an exponential decreasing function of z for absorbing media.

It is also shown that the time-averaged energy density for a single wave over a single period T of oscillation will be proportional to the square of the amplitude, that is,

$$W = a^2/2. \tag{2}$$

16.1.1.3 If interference phenomena of two or more waves are considered, the time-averaged energy density, W, will be the sum of the instantaneous energies of the electric vectors, the average over a single period of T of the square of the sum of the instantaneous magnitudes of the electric vectors. Thus, for two collinear waves,

$$W = \frac{1}{2}\left[a_1^2 + 2 a_1 a_2 \cos (\phi_1 - \phi_2) + a_2^2\right] \tag{3}$$

where

ϕ_1, ϕ_2 = phase angles of each electric vector
$\phi_1 - \phi_2$ = fixed phase difference, δ.

16.1.1.4 Referring again to Section 3, the conditions of Equation (3) depend on the direction of propagation and the source of radiation. Collinear, coherent waves will reinforce each other when the phase difference is zero or an even multiple of π and oppose each other when the phase difference is an odd multiple of π. For collinear, non-coherent waves, this reinforcement or opposition does not apply, but the time-averaged energy densities will add according to

$$W = \frac{1}{2}\left[a_1^2 + a_2^2\right]. \tag{4}$$

16.1.1.5 If the waves are non-collinear and coherent, as if Figure 16.1, their phases Φ_1 and Φ_2 will be given by

$$\Phi_1 = knz + \phi_1$$
$$\Phi_2 = kn(x \sin\theta + z \cos\theta) + \phi_2 \tag{5}$$

where

ϕ_1 = phase angle of the wave propagated along OZ,
ϕ_2 = phase angle of the wave propagated along OP.

The difference in phase angles will then be

$$\Phi_1 - \Phi_2 = \phi_1 - \phi_2 - knx \sin\theta + knz(1 - \cos\theta). \tag{6}$$

Letting $\phi_1 - \phi_2 = \delta$ and using Equation (3), the time-averaged energy density will be

$$2W = a_1^2 + a_2^2 + 2\, a_1\, a_2 \cos\left[\, \delta - knx \sin\theta + knz\, (1 - \cos\theta)\, \right] \qquad (7)$$

where a_1 and a_2 are the amplitudes of the interfering waves at the point $(0,\ y,\ 0)$. By choosing θ to be suitably small, we can set $\sin\theta = \theta$ and $1 - \cos\theta = \theta^2/2$. If observation is to be made in the xy plane near $z = 0$, the z term can be neglected and Equation (7) becomes

$$2W = a_1^2 + a_2^2 + 2\, a_1\, a_2 \cos\left(\delta - \frac{2\pi nx\theta}{\lambda}\right) \qquad (8)$$

which is the usual interference formula.

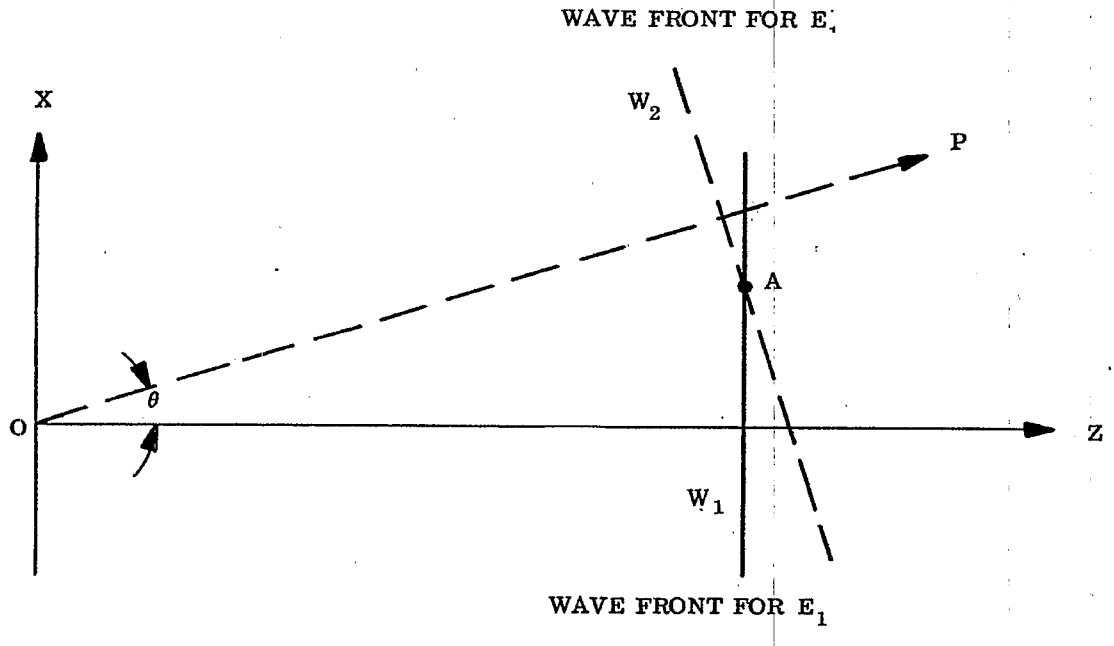

Figure 16.1 - Interference between two plane wavefronts W_1 and W_2 that are propagated along different directions.

16.2 THE FIZEAU INTERFEROSCOPE

16.2.1 Principles of operation.

16.2.1.1 A group of interferometers known as Fizeau interferoscopes or Fizeau double beam interferometers have been devised around afore mentioned principles for the purpose of testing the flatness and parallelism of the surfaces S_1 and S_2, Figure 16.2, of a plane parallel plate or for testing the flatness of a surface against an optical flat. The essential characteristics of these interferometers are illustrated in Figure 16.2 in which either of the surfaces S_1 or S_2 may be the optically flat surface of reference. Monochromatic light issues from a pinhole H and emerges from the collimator L_1 as plane waves. By a slight angular adjustment (not shown) of the upper plate surface, S_2 can be made to reflect ray HA approximately back upon itself. We take this direction as the OZ-direction. Surface S_1 now reflects ray HAB along a direction BR such that angle $\theta = 2\alpha$ where α is the indicated angle between surfaces S_1 and S_2. We choose the direction OP parallel to BR. The coordinate line OX falls in the wave front reflected by S_2. Lines OP, OX and OZ together with the angle are now corresponding elements in Figures 16.1 and 16.2. Equation (7) or (8) may therefore be applied to determine the energy densities at any point (x, z). If both surfaces S_1 and S_2 are flat, σ is constant; but we must take $\sigma = \sigma(x)$ when both surfaces are not flat. In the interests of simplicity, we shall suppose at first that surfaces S_1 and S_2 are flat.

16.2.1.2 When S_1 and S_2 are uncoated surfaces of glass, the two waves formed by reflection at S_1 and S_2 will have amplitudes a_1 and a_2 so nearly alike that one may set $a_1 = a_2 = a$ and write Equation (7) as

$$W = a^2 \left\{ 1 + \cos \left[\delta - knx \sin \theta + knz \left(1 - \cos \theta \right) \right] \right\}. \tag{9}$$

Furthermore, $\delta = \pi$ for optically flat surfaces of glass since the phase changes due to reflection at A and B differ by 180^0. Thus,

$$W = a^2 \left\{ 1 - \cos kn \left[x \sin \theta - z \left(1 - \cos \theta \right) \right] \right\} \tag{10}$$

if the surfaces S_1 and S_2 are optically flat surfaces of glass.

16.2.1.3 Lens L_1 and L_2 are invariably arranged so that the plane $z = 0$ or a neighboring plane is focused upon the retina or upon the photographic emulsion, i.e., one arranges to observe the energy density W in the interference fringes in a plane for which z is either zero or small. Also, the angle θ is very small in actual practice. Thus both z and $1 - \cos \theta$ become so small that one will ordinarily be justified in neglecting the term $z(1 - \cos \theta)$ in Equation (10), and in writing

$$W = a^2 \left[1 - \cos kx\theta \right] = 2a^2 \sin^2 \left(\frac{\pi x \theta}{\lambda} \right)$$

$$= 2a^2 \sin^2 \left(\frac{2\pi x \alpha}{\lambda} \right) \tag{11}$$

when the space between S_1 and S_2 is air.

16.2.1.4 The actual energy density W is of little interest in practical interferometry. Interest centers, rather, upon the fringe-width, the distance from one fringe to the next similar interference fringe. The fringe system is repeated, according to Equation (11), whenever x is altered by the amount Δx such that $2\pi\alpha \, \Delta x / \lambda = \pi$, i.e., whenever

$$h = |\Delta x| = \frac{\lambda}{2\alpha} \tag{12}$$

where h denotes fringe-width and α is the angle in radians between surfaces S_1 and S_2. Equation (12) can be used to measure α. If $\alpha = 0$, the fringe-width is infinite, and conversely.

16.2.1.5 It will be seen from Figure 16.2 that

$$d = x\alpha \tag{13}$$

where d is the thickness of the air gap at point x. Equation (11) can therefore be written in the highly instructive form

$$W = 2a^2 \sin^2 \left(\frac{2\pi d}{\lambda} \right) \tag{14}$$

Hence W is constant for those loci along which the separation d of the surfaces is constant. W is, of course, constant along an interference fringe. Therefore, each interference fringe is the locus of points for which the separation d of the surfaces is constant. This statement holds throughout interferometry with very few exceptions or qualifications. With respect to Equation (14), we note that W has the period $d = \lambda/2$. This means that in the Fizeau interferoscope the separation d changes by $\lambda/2$ in going, say, from one bright fringe to the

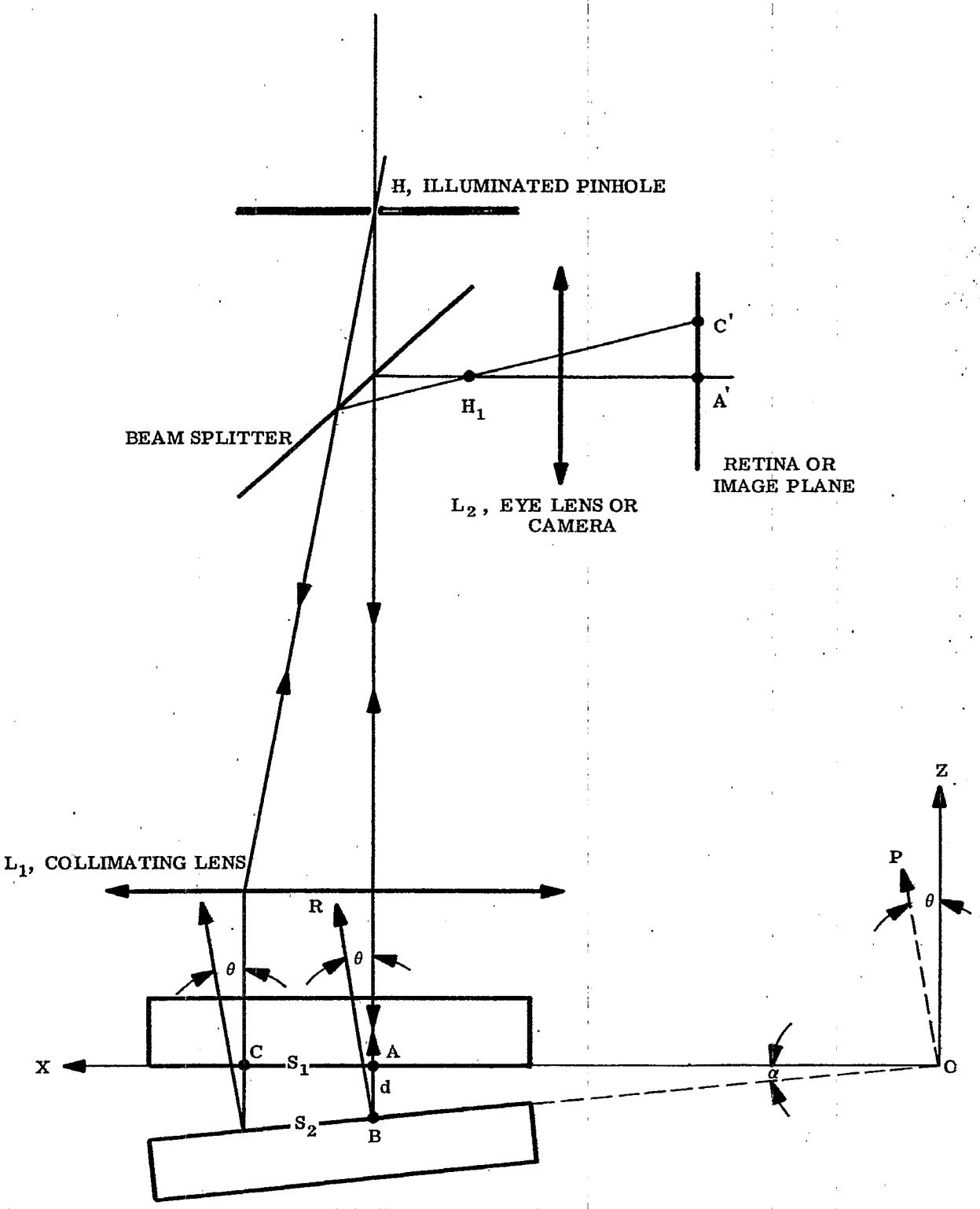

FIGURE 16.2 -Notation with respect to the Fizeau Interferoscope or Interferometer.

next. More generally, it is the optical path nd that changes by $\lambda/2$.

16.2.1.6 The use of the Fizeau interferoscope for examining parallelism of plates amounts to considering the gap between surfaces S_1 and S_2, Figure 16.2, as the plate. The refractive index of the gap is now that of the plate.

16.2.2 The Fizeau interferoscope in testing for optical flatness.

16.2.2.1 The separation d of surfaces S_1 and S_2, Figure 16.2 can be large. Consequently the risk of scratching a surface during testing for flatness is avoided. Moreover, the reference flat is not subjected to wear by frequent rubbing, etc.

16.2.2.2 As can be expected, the interference fringes will not be straight unless the test surface is also an optical flat.

16.2.2.3 In paragraph 16.2.1.2 we saw that $\delta = \pi$ for optically flat, uncoated surfaces S_1 and S_2 of glass. The main effect of a small departure of the test surface from a plane is to introduce a local irregularity in the separation d, Figure 16.2, over the range of x at which the departure occurs. It is natural, then, to consider δ in the form

$$\delta = \pi - 2kn\ D(x) \tag{15}$$

where D(x) shall express the phase difference introduced between the two interfering waves on account of the departure of the reflecting surfaces from a plane. We suppose, as in the argument leading to Equation (11), that the term knz (1 - cos θ) is negligible in Equation (9) and introduce δ from Equation (15). The result is

$$W = a^2 \left[1 - \cos k (2\,D(x) + x\theta) \right]; \quad n = 1; \tag{16}$$

where θ is so small that one can set $\sin \theta = \theta$. Since $\theta = 2\alpha$,

$$W = a^2 \left[1 - \cos 2k (D(x) + x\alpha) \right] = 2a^2 \sin^2 \left[\frac{2\pi}{\lambda} (D(x) + x\alpha) \right] \tag{17}$$

The exact physical significance of D(x) is now clear. Since $x\alpha = d$, the separation between surfaces S_1 and S_2 (see Figure 16.2) at point x, D(x) must be the increase in separation due to a local bulge in one of the reflecting surfaces. $D(x) > 0$ when the bulge increases the separation between the two surfaces.

16.2.2.4 It should be observed from Equation (17) that W = constant whenever

$$D(x) + x\alpha = D(x) + d = \text{constant}. \tag{18}$$

Since D(x) + d is the actual separation of the surfaces at point x, it follows that an interference fringe is the locus of those points x for which the separation of the interferometer surfaces is a constant. If the surfaces are plane, D(x) = 0 and the fringes are straight.

16.2.2.5 Suppose one of the interferometer flats is pressed or moved so as to decrease d by a small amount. Since each fringe is the locus of equal separations D(x) + d, the whole family of fringes will move in the positive x - direction of Figure 16.2 wherever D(x) = 0. In localities where D(x) = 0, each fringe will move in a slightly more complex manner so as to find the location where D(x) + d remains constant.

16.3 THE TWYMAN-GREEN INTERFEROMETER

16.3.1 Principles of operation.

16.3.1.1 The essential characteristics of the Twyman Green interferometer are shown in Figure 16.3. The physical principles utilized in the Twyman-Green interferometer and in the Fizeau interferoscope are so similar that the corresponding elements of Figures 16.2 and 16.3 are recognized easily. These corresponding elements are denoted by the same symbols. A small pinhole H, illuminated by monochromatic light, is located at the first focal plane of the collimator L_1 so that a plane wave front is reflected by surfaces S_1 and S_2 of the end-mirrors. A telescope is added to produce an image of the pinhole H at H_1. The surface S_1 appears to be located at S_1'. If S_1' makes the angle α with S_2, the ray reflected from S_1 will appear to be a ray BR such that BR makes the angle $\theta = 2\alpha$ with the ray AQ reflected from S_2. We take OZ parallel to AQ and OP parallel to BR. The coordinate OX falls in the wave front reflected by S_2. This time, to complement Figure 16.2, we show the passage of ray BR to the vicinity of the eye lens where a second image H_2 of the pinhole H is formed. The width of the interference fringes is increased by decreasing the separation $H_1\ H_2$ of the images of the pinhole by tilting mirror S_2 in the direction for decreasing angles θ and α.

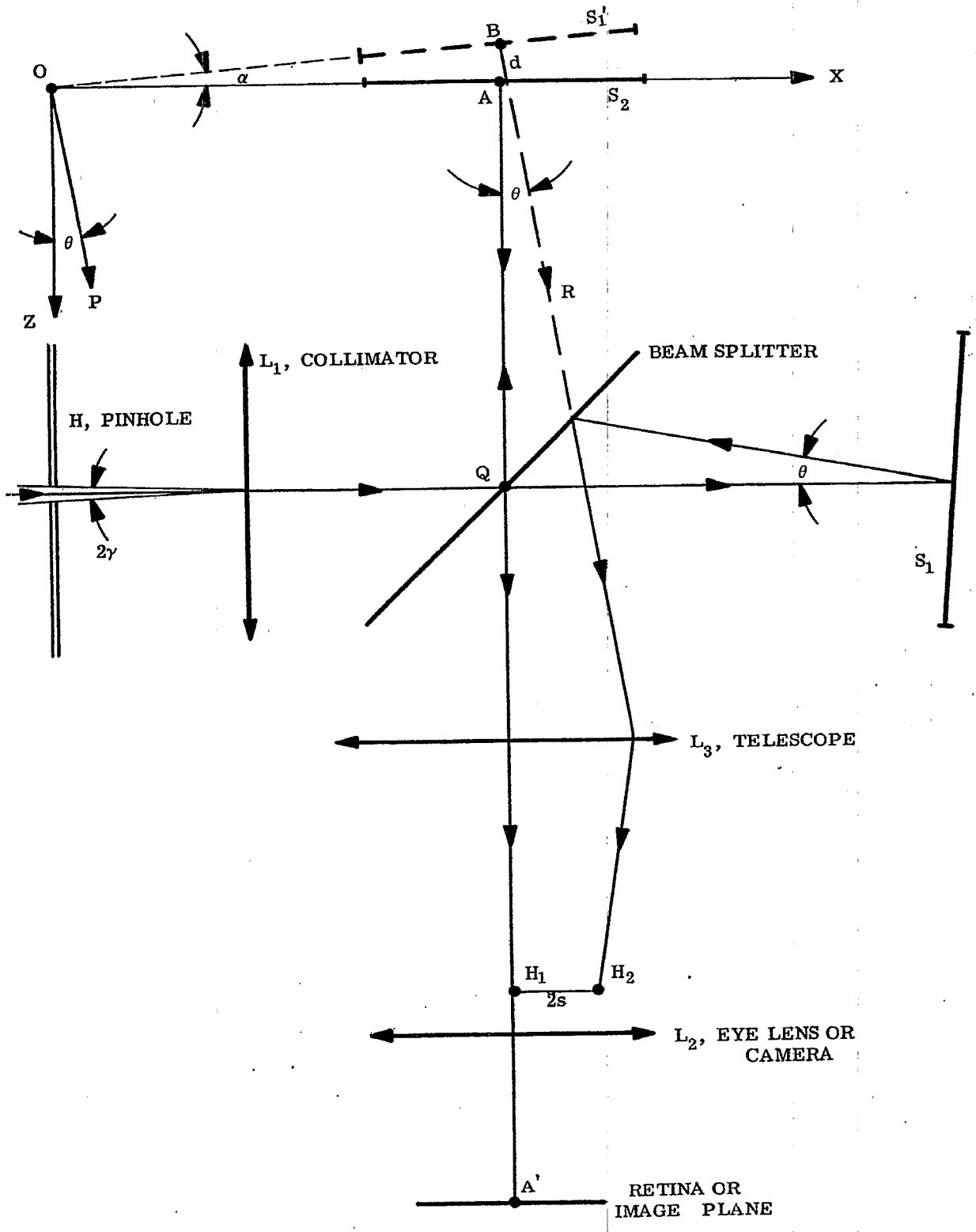

FIGURE 16.3 -Notation with respect to the Twyman Green Interferometer.

16.3.1.2 One major use of the Twyman Green interferometer is to examine the optical quality of glass plates, prisms, etc. The sample to be examined is placed in one arm of the interferometer and the effect of the sample upon the fringes noted. It is usually necessary to readjust the angular setting of at least one of the mirrors S_1 and S_2 and to alter the length arm AQ so as to obtain best contrast in the fringes. We shall not be concerned here with the details of the many applications of the Twyman Green interferometer but, rather, with the principles involved.

16.3.1.3 The beam splitter is usually an optically parallel plate, one of whose surfaces is coated with a uniform film of silver or aluminum. Whereas it is not necessary that the transmittance and reflectance of the filmed surface shall be alike, they should not be markedly dissimilar. Where utmost contrast in the fringes is desired, the second surface of the beam splitter should be rendered low reflecting. Henceforth, it will be supposed that the beam splitter consists, in effect, of a single surface as in Figure 16.3.

16.3.1.4 The amplitudes a_1 and a_2 of the two, plane, interfering waves that enter the telescope L_3 are not likely to be as nearly equal as in the Fizeau interferoscope. However, these amplitudes will be nearly alike provided that the end mirrors S_1 and S_2 have practically equal reflectances and provided that the test sample transmits well. It is, of course, possible to compensate the effects of the sample in one arm by placing a suitable absorbing plate in the second arm.

16.3.1.5 We saw that the phase difference δ between the two interfering waves will be π in the Fizeau interferoscope. The phase changes on reflection at the surfaces S_1 and S_2 of the Twyman Green interferometer are likely to be nearly alike to that δ can be sensibly zero. However, one cannot always be certain that δ is sensibly zero or that a_1 and a_2 are sensibly alike.

16.3.1.6 The product knz $(1 - \cos\theta)$ of Equation (7) will usually be negligible in the Twyman Green interferometer for the same reason that applies to the Fizeau interferoscope. We introduce this approximation into Equation (7). Instead of writing δ as in Equation (15), we set $\delta = \delta_o - 2kn\, D(x)$. The result is

$$2W = a_1^2 + a_2^2 + 2a_1 a_2 \cos\left[\delta_o - 2k\,(x\alpha + D(x))\right] \qquad (19)$$

since $\sin\theta - \theta$ and $\theta = 2\alpha$.

Equations (17) and (19) differ mainly in that the fringe system is shifted slightly with respect to x and in that the fringe contrast obtained from (19) will be inferior to the constrast obtained from (17) except when $a_1 = a_2 = a$, i.e., except when the amplitudes of the two interfering waves are made substantially alike in the Twyman Green interferometer.

16.4 EFFECT OF MONOCHROMATICITY ON FRINGE CONTRAST

16.4.1 Discussion of problem.

16.4.1.1 Fringes obtained with Fizeau interferoscopes or with Twyman Green interferometers "wash out" when the path difference d, Figures 16.2 and 16.3, becomes too great relative to the spectral purity of the monochromatic source. It can be shown that the effect of the presence of many different wavelengths $(k = 2\pi/\lambda)$ in Equation (19) is to reduce the average value of the cosine term to zero as the spread of wavelengths is increased. The physical circumstances become similar to those which cause Equation (3) to degenerate to Equation (4). We may say that the source of light becomes incoherent. An insight into the nature and magnitude of the difficulty can be obtained from the following simplified considerations.

16.4.1.2 With respect to Equation (19), suppose for convenience of argument that $D(x) = 0$ and suppose that the source contains wavelengths from $\lambda_o - |\Delta\lambda|$ to $\lambda_o + |\Delta\lambda|$ where $|\Delta\lambda|$ is small. If $|\Delta\lambda|$ is too large, an interference maximum for $\lambda = \lambda_o$ will fall at the same point $x\alpha = d$ as the first interference minimum due to the wavelength $\lambda = \lambda_o - |\Delta\lambda|$. Thus, if

$$\frac{4\pi x\alpha}{\lambda_o} - \delta_o = \nu\, 2\pi \; ; \; \nu \text{ an integer;} \tag{20}$$

and

$$\frac{4\pi x\alpha}{\lambda_o - |\Delta\lambda|} - \delta_o = \nu\, 2\pi + \pi \; ; \tag{21}$$

then since $1/(\lambda_o - |\Delta\lambda|) - 1/\lambda_o + |\Delta\lambda|/\lambda_o^2$, it follows by subtraction of Equation (20) from (21) that $4x\alpha\,|\Delta\lambda|/\lambda_o^2 = 1$. When, therefore,

$$|\Delta\lambda| = \lambda_o^2/4x\alpha = \frac{\lambda_o^2}{4d} \; , \tag{22}$$

a bright fringe due to λ_o will fall upon a dark fringe due to $\lambda = \lambda_o - |\Delta\lambda|$. If the radiant fluxes of the two wavelengths are approximately equal and if the wavelengths λ_o and $\lambda_o - |\Delta\lambda|$ do not differ appreciably in color, the interference fringes will be practically obliterated when $|\Delta\lambda|$ and the separation d are related as in Equation (22).

16.4.1.3 When $|\Delta\lambda|$ is less than that given by Equation (22), one can expect that the fringes will be visible. In fact, we must expect from Equation (22) that the condition for the appearance of interference fringes is

$$|\Delta\lambda|\; d < \frac{\lambda_o^2}{4} \; . \tag{23}$$

Contrast in the fringes is improved by choosing $|\Delta\lambda|$ and the path difference d so that their product is small.

16.5 EFFECT OF PINHOLE SIZE ON CONTRAST

16.5.1 Discussion of problem.

16.5.1.1 As can be expected intuitively, the effect of opening the pinhole H too far is to reduce contrast in the fringes even though the light is so monochromatic that one can set $|\Delta\lambda| = 0$. It can be shown that when the pinhole size cannot be neglected, one obtains, instead of monochromatic law of Equation (19), the result

$$2W_T = a_1^2 + a_2^2 + 2a_1\, a_2 \left[2\, \frac{J_1(k2\gamma s)}{k2\gamma s} \right] \cos\left[\delta_o - 2k\,(x\alpha + D(x)) \right] \tag{24}$$

in which W_T is the total energy density due to all of the points in the illuminated pinhole H (see Figure 16.3), 2γ is the angle subtended by the pinhole at the collimating lens L_1, 2s is the indicated separation of the images H_1 and H_2 of the pinhole and J_1 is a Bessel function of first order and first kind.

16.5.1.2 The function $2J_1(t)/t$ assumes its maximum value of unity at $t = 0$. Therefore, Equation (24) is identical to Equation (19) whenever the angle 2γ subtended by the pinhole at the collimator is so small that one can accept the approximation $2J_1(k2\gamma s)/k2\gamma s = 1$. Contrast in the fringes is excellent provided that the amplitudes a_1 and a_2 of the interfering beams are not too unlike. For a given value of 2γ, the fringes should show better contrast as they are broadened, i.e., as the separation s of the two pinhole images is decreased.

16.5.1.3 The function $J_1(t)/t$ has an infinite number of zeros the first of which occurs at $t = 3.8317$. Whenever the product $2\gamma s$ becomes so large that $2\pi\, 2\gamma s/\lambda = 3.8317$, $J_1(k2\gamma s)/k2\gamma s = 0$. Hence W_T becomes constant and should be independent of x and the fringes should vanish when

$$2\gamma s = \frac{3.8317\lambda}{2\pi} = 0.61\lambda \; . \tag{25}$$

Since $J_1(t)/t$ changes sign as t passes through any of the roots of $J_1(t)/t = 0$, the fringes should shift abruptly by one half fringe width as $2\gamma s$ passes through the value given by Equation (25).

16.5.1.4 Whereas it is the writer's experience that Equation (25) does not agree in an excellent quantitative manner with observations in, say, the Twyman-Green interferometer, it does serve as semi-quantitative basis for predicting the degree of contrast in the fringes.

16.6 YOUNG'S PINHOLE INTERFEROMETER

16.6.1 Introduction.

16.6.1.1 The Fizeau and Twyman Green interferometers belong to a broad class of doubled pinhole interferometers in which two actual pinholes are illuminated or in which the image of one illuminated pinhole is doubled by any one of a variety of beam splitting devices. The rudimentary theory of formation of the interference fringes is essentially the same for this group of interferometers. If the plane of observation is sufficiently far from the location of the pinholes, the two corresponding waves that arrive at the plane of observation are essentially plane so that the theory of the foregoing paragraphs applies.

16.6.1.2 The following argument presents a second, very useful point of view that encroaches to some extent upon Huygens' principle. Let us consider the simplest of all double pinhole interferometers, namely Young's famous interferometer of Figure 16.4. Monochromatic light is focused upon a small pinhole H. Coherent, spherical waves emanate from H and illuminate the small pinholes H_1 and H_2. If H falls upon the Z-axis, the light reaching H_1 and H_2 will be in phase. Otherwise, a phase difference δ_o will be introduced. Pairs of coherent, spherical waves emerge from pinholes H_1 and H_2 and reach point (x, y) of the observation plane after traversing paths r_1 and r_2. If distance D is large relative to the separation 2s of the pinholes H_1 and H_2 and if point (x, y) is not too far from the Z-axis, the distances r_1 and r_2 will be so nearly alike that the two waves from H_1 and H_2 will arrive at point (x, y) with substantially equal amplitude provided that they leave H_1 and H_2 with substantially equal amplitude. We shall suppose for sake of generality that the interfering waves from H_1 and H_2 reach point (x, y) with the amplitudes a_1 and a_2, respectively. (The amplitude of one of the waves might be reduced, for example, by placing an absorbing glass plate over one of the pinholes or by making the pinholes small but unlike in area).

From Figure 16.4,

$$r_1^2 = D^2 + (x - s)^2 + y^2 ; \tag{26}$$

$$r_2^2 = D^2 + (x + s)^2 + y^2 . \tag{27}$$

therefore

$$r_2^2 - r_1^2 = (r_2 - r_1)(r_2 + r_1) = 4xs \tag{28}$$

or

$$r_2 - r_1 = 2xs \Big/ \left(\frac{r_2 + r_1}{2}\right). \tag{29}$$

It matters to a considerable extent which approximation one wishes to accept for $(r_2 + r_1)/2$, the average value of r_1 and r_2. In case the point of observation (x, y), Figure 16.4, falls near the Z-axis, both r_1 and r_2 differ only slightly from $R = \sqrt{D^2 + s^2}$, and the average value of r_1 and r_2 will fall nearer R than either r_1 or r_2. Accordingly, we suppose that the point of observation (x, y), falls near the Z-axis and accept the approximation

$$r_2 - r_1 = \frac{2xs}{\sqrt{D^2 + s^2}} \tag{30}$$

Then from Figure 16.4, $\dfrac{s}{\sqrt{D^2 + s^2}} = \sin\dfrac{\theta}{2}$. Since D is great relative to s,

$$\sin\frac{\theta}{2} \approx \frac{\theta}{2} = \frac{s}{\sqrt{D^2 + s^2}} ; \tag{31}$$

therefore,

$$r_2 - r_1 = x\theta \tag{32}$$

in which θ is very nearly equal to the actual angle between the direction of propagation of the two waves that reach point (x, y) from the pinholes H_1 and H_2.

16.6.1.3 We find that the two coherent waves which interfere at point (x, y) have amplitudes a_1 and a_2 and the phase difference $\phi_1 - \phi_2$ such that

$$\phi_1 - \phi_2 = \delta_o + k(r_1 - r_2) = \delta - kx\theta \tag{33}$$

wherein the portion $kx\theta$ is due to the path difference $r_2 - r_1$ and wherein δ_o specifies the phase difference

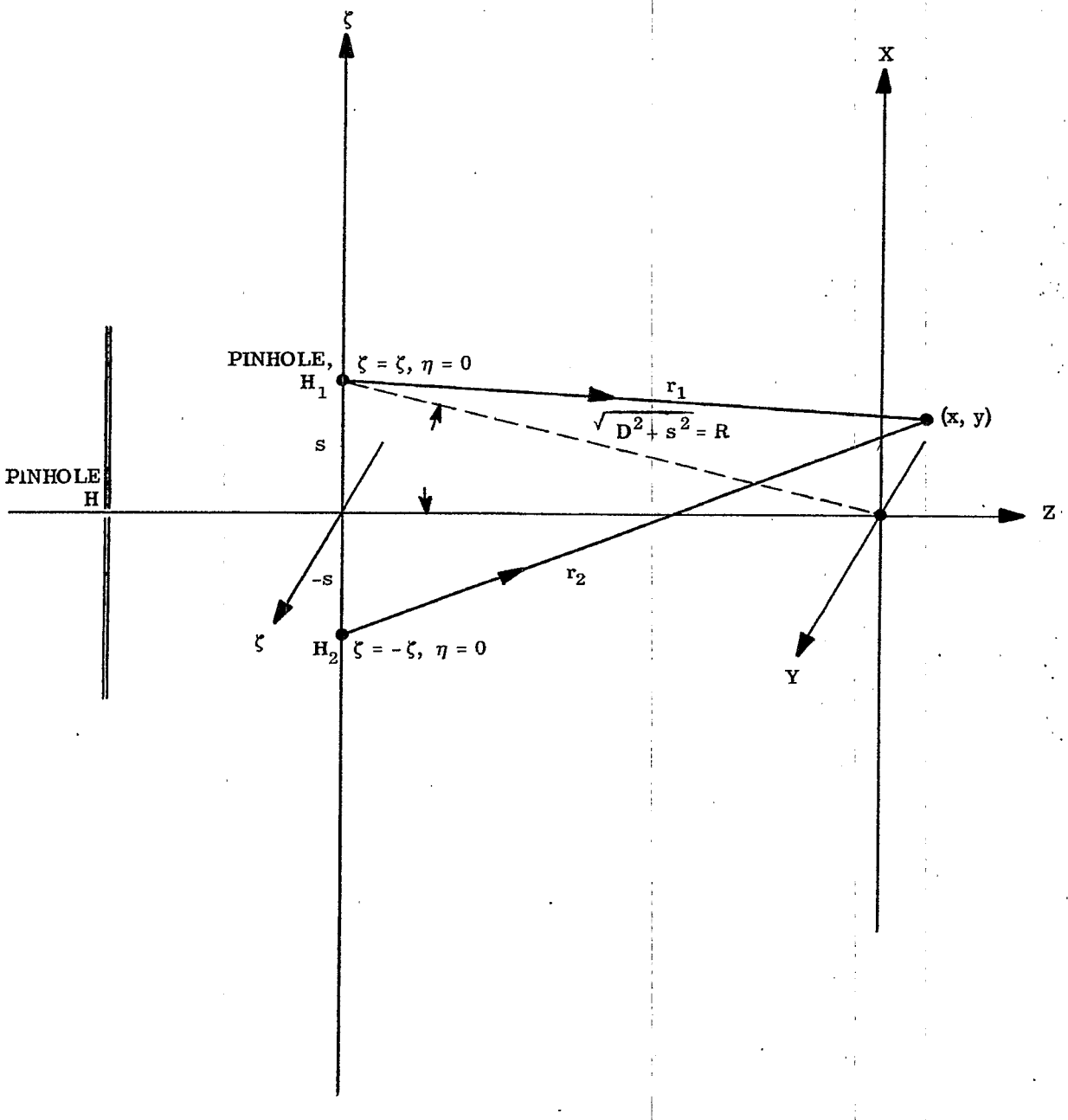

FIGURE 16.4 - Young's Pinhole Interferometer

between the two, interfering, non-collinear waves as they leave the pinholes H_1 and H_2. The time-averaged energy density W produced by interfering waves is given again by Equation (3). Thus, from Equations (3) and (33)

$$2W = a_1^2 + a_2^2 + 2\, a_1\, a_2 \cos\,(\,\delta_o - kx\theta\,) \tag{34}$$

in which

$$k = 2\pi/\lambda \quad \text{and} \quad \theta = 2s\,/\sqrt{D^2 + s^2}\ . \tag{35}$$

where 2s is the separation of the pinholes H_1 and H_2, Figure 16.4.

16.6.1.4 Comparision of Equations (19) and (34) shows that they agree since $2\alpha = \theta$ and since D(x), as defined in Paragraph 16.3, is zero as applied to Young's pinhole interferometer.

16.6.1.5 The fringes formed in Young's pinhole interferometer will not be straight, as predicted by the approximation of Equation (34), unless the point of observation is near the Z-axis of Figure 16.4. As the point of observation is moved out to distances $\sqrt{x^2 + y^2}$ that become appreciable with respect to D, the average value of r_1 and r_2 becomes a function of both x and y. It follows from Equations (29) and (33) that $\phi_1 - \phi_2$ will not vary in a simple linear manner with x. The fringes become curved in a manner that is not difficult to ascertain from a further study of Equations (26), (27), and (29).

16.6.1.6 Young's slit interferometer is obtained by replacing pinholes H, H_1 and H_2 by very narrow slits perpendicular to the plane of the paper. With this arrangement, the interference fringes seen at plane x y will remain straight over a greatly increased portion of the xy plane provided that the slits are sufficiently long.

16.6.1.7 Young's interferometer is both useful and simple to construct. The difference in optical path, for example, of two similar glass plates of nearly the same thickness can be ascertained by applying the following principles. We observe that if pinhole H is on the Z axis, Figure 16.4, a bright white fringe will be formed at O, where x = 0, when H is illuminated with white light because the optical paths from H to O are equal. Constructive interference occurs at O for all wavelengths. If the pinhole H is not on the Z axis, the bright white fringe will be found at a location $x \neq 0$. This location is called the white light position and determines a point of reference at which the optical paths from H to O are equal. When monochromatic light is substituted for white light, the fringes appear in best contrast about the white light position. Suppose that the optical path H H_1 O is increased by a slight amount δ_o relative to the optical path H H_2 O by the insertion at H_1 and H_2 of glass plates that differ slightly in optical path. We see from Figure 16.4 that the ray H_2 x must be inclined toward larger x-values in order to equalize the optical path difference between the paths $H_2 H_1$ x and H H_2 x. Therefore, the white light fringe or any monochromatic fringe must move outward from the axis Z in the direction of that pinhole H_1 or H_2 over which has been placed the plate having the greater optical path. The magnitude of δ_o can be found as follows from the measurement of the fringe shift produced by δ_o.

16.6.1.8 First, the fringe width h is the increase in x for which $k(x + h)\theta$ exceeds $kx\theta$ by 2π in Equation (34). Thus

$$kh\theta = 2\pi \qquad \text{or} \qquad h = \frac{\lambda}{\theta}. \tag{36}$$

Secondly, a given interference fringe occupies that position x for which

$$\delta_o - kx\theta = \text{constant} = C. \tag{37}$$

Suppose, for generality, that δ_o has successively the values δ_1 and δ_2. Denote the corresponding position of a given fringe by x_1 and x_2. Then from Equation (37)

$$\begin{aligned} \delta_1 - kx_1\theta &= C \\ \delta_2 - kx_2\theta &= C \end{aligned} \tag{38}$$

By subtraction of Equations (38) one finds that

$$\delta_2 - \delta_1 = k\theta\,(x_2 - x_1) = \frac{2\pi}{\lambda}\,\theta\,(x_2 - x_1). \tag{39}$$

From Equations (37) and (39) we obtain the extremely useful result

$$\delta_2 - \delta_1 = 2\pi\,\frac{x_2 - x_1}{h}, \text{ radians.} \tag{40}$$

In other words, the phase change in the two arms H H_1 x and H H_2 x is given by the ratio of the fringe shift, $(x_2 - x_1)$, to the fringe width, h.

16.6.1.9 Difficulties can appear when $\delta_2 - \delta_1$ exceeds 2π; for then the fringe shift, $x_2 - x_1$, exceeds the fringe width, h, by a number of fringes that may not be obvious. This ambiguity about the "fringe jump" can be settled by considering the shift of the white light position or by making measurements of the fringe locations at more than one wavelength.

16.7 LLOYD'S INTERFEROMETER

16.7.1 Description. Lloyd's double pinhole or double slit arrangement for obtaining interference fringes is illustrated in Figure 16.5. Corresponding elements are denoted by the same symbols in Figures 16.4 and 16.5 to emphasize their similarity. The interpretations of the interferometers due to Lloyd, Young, Fizeau, and Twyman Green are alike provided that the pinholes are small and provided that the distance D is great. It should be observed that the virtual image H_2 is a mirror image of H_1. The relative locations of the corresponding coherent points in the "images" H_1 and H_2 will therefore be significantly different in Lloyd's interferometer as compared with the Fizeau and Twyman Green interferometers. This mirror image relation between H_1 and H_2 is avoided by Fresnel's double mirror interferometer which is illustrated in Figure 16.6. Both H_1 and H_2 are now virtual images whose separation 2s is governed by the angle α between the interferometer mirrors M_1 and M_2.

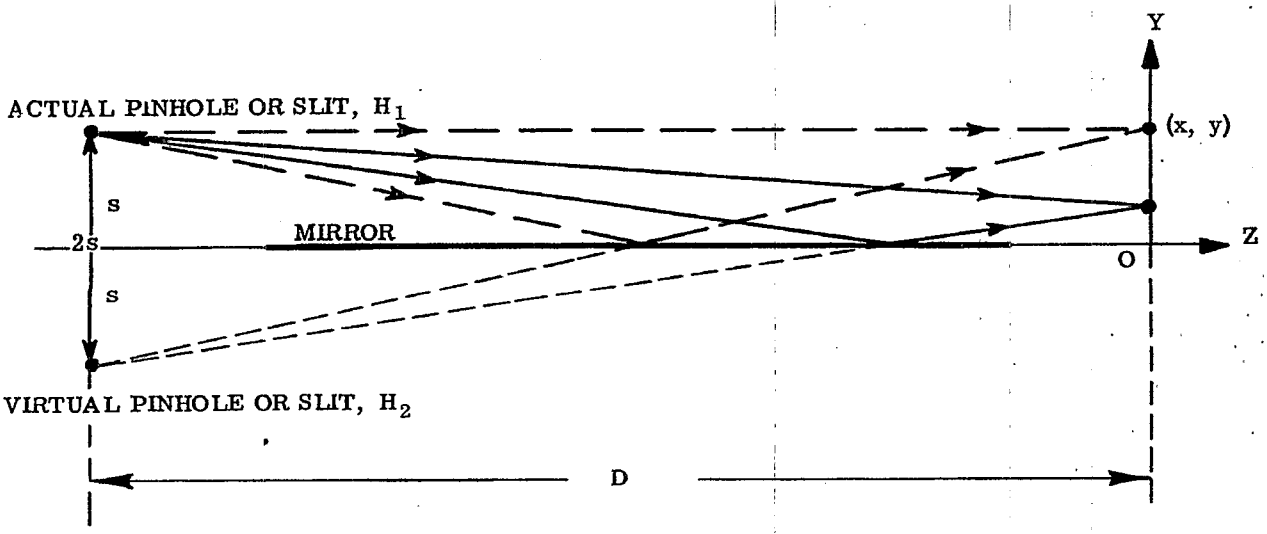

FIGURE 16.5 - Lloyd's Interferometer

16.8 FRESNEL COEFFICIENTS FOR NORMAL INCIDENCE

16.8.1 Computing amplitude reflectance and transmittance.

16.8.1.1 Let n and nK denote the optical constants of two media that are in contact across a plane interface as in Figure 16.7. Then for normal incidence upon the interface along the indicated direction, the amplitude reflectance ρ is given by

$$\rho = \frac{M_o - M_1}{M_o + M_1} \tag{41}$$

and the amplitude transmittance τ across the interface is given by

$$\tau = \frac{2\,M_o}{M_o + M_1} \tag{42}$$

where

$$M_\nu \equiv n_\nu\,(1 + i\,K_\nu); \quad \nu = 0,\ 1. \tag{43}$$

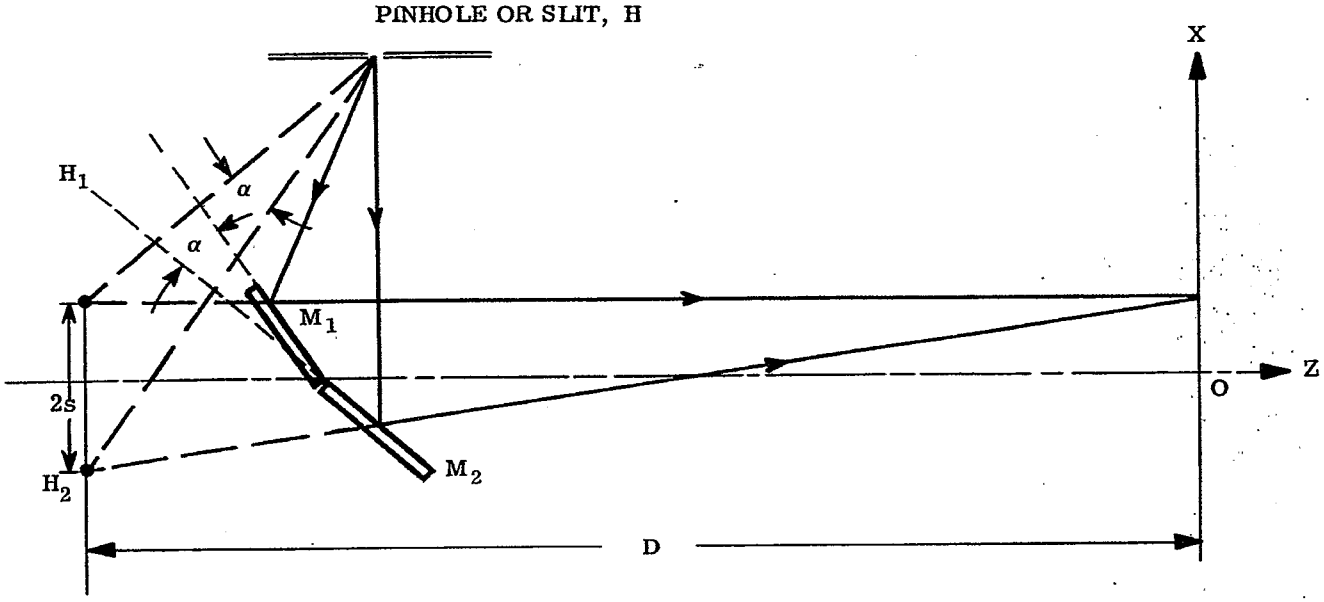

FIGURE 16.6 - Fresnel's Mirror Interferometer.

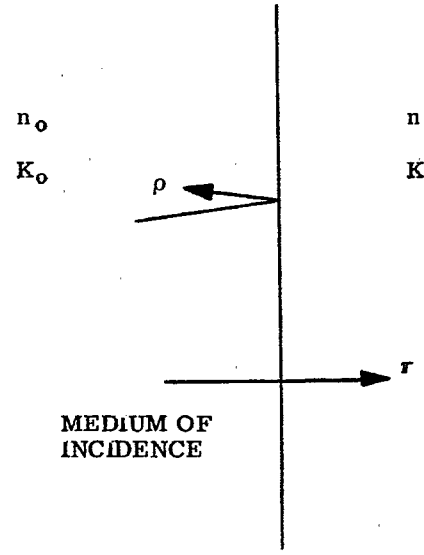

FIGURE 16.7 -Transmittance and reflectance at Translucent Interface

16.8.1.2 Suppose that neither medium absorbs so that $K_o = K_1 = 0$. Then

$$\rho = \frac{n_o - n_1}{n_o + n_1}; \qquad \tau = \frac{2n_o}{n_o + n_1}. \tag{44}$$

We see that $\tau > 0$, but the amplitude reflectance ρ is greater or less than zero according as n_o is greater than or less than n_1. If we write ρ in the form

$$\rho = \left| \frac{n_o - n_1}{n_o + n_1} \right| \cos \sigma, \tag{45}$$

we see that σ, the phase change on reflection, is zero when $n_o > n_1$ but is π when $n_o < n_1$. Furthermore, the phase change on transmission across an interface between two non-absorbing media is always zero.

16.8.1.3 Interferometers usually involve the splitting of a light beam at one or more interfaces between two media. In order to compute or to estimate the amplitudes a_1 and a_2 of the interfering waves thus produced, knowledge of the Fresnel coefficients is essential. The Fresnel coefficients at normal incidence will suffice for the purposes of the present text. The application of Fresnel's coefficients for normal incidence to cases involving oblique incidence can, however, be misleading. The reader who needs to compute the amplitudes a_1 and a_2 for oblique incidence should consult paragraph 24.1.

16.9 INTERFERENCE WITH PLANE PARALLEL PLATES AND DISTANT LIGHT SOURCES

16.9.1 Discussion of Problem.

16.9.1.1 A ray AB from one point in a distant source of light is incident upon a plate of thickness d with refractive index n_1. Reflected rays R_1, R_2, R_3, etc., and transmitted rays T_1, T_2, etc., are formed in the manner indicated in Figure 16.8. We suppose that the plate is non-absorbing and that the reflectance of its surfaces is so low that only rays R_1 and R_2 need be considered in the reflected beam of rays. The problem is to find the optical path difference δ between rays R_2 and R_1 under the assumption that the surfaces of

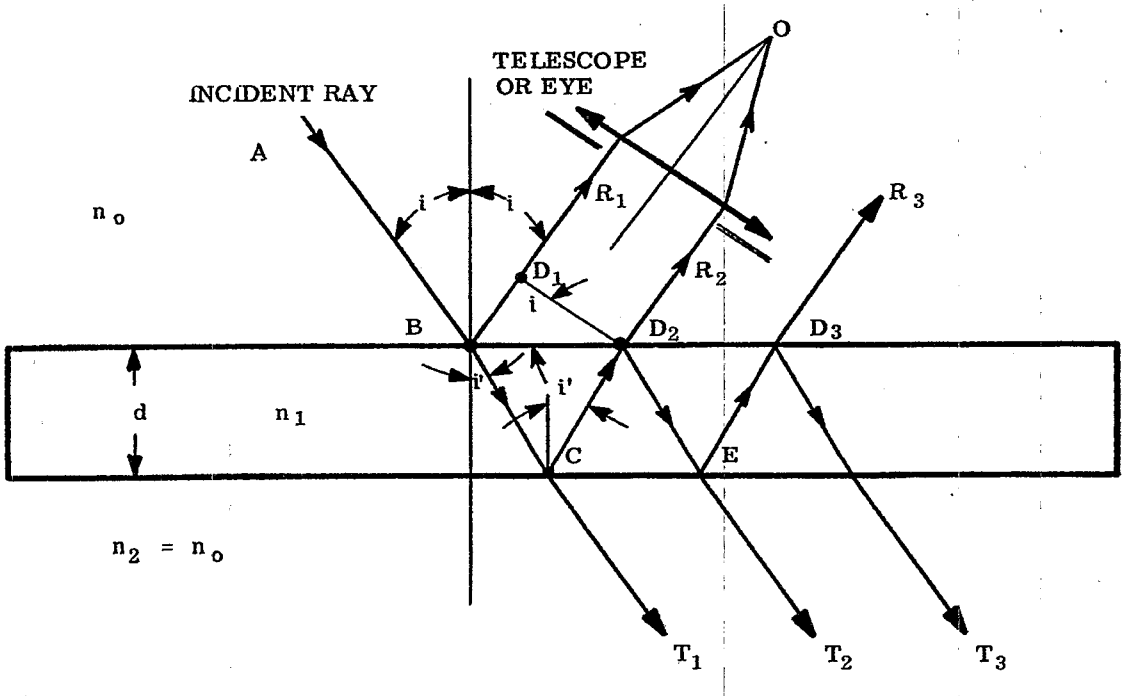

FIGURE 16.8 - The Dielectric Plate as an Interferometer.

the plate are parallel. Let line segment D_1 D_2 be drawn perpendicular to rays R_1 and R_2. Then

$$\delta = n_1 (BC + CD_2) - n_o BD_1 \quad (46)$$

$$BC = CD_2 = \frac{d}{\cos i'}$$

$$BD_2 = 2d \tan i'$$

$$BD_1 = BD_2 \sin i = 2d \tan i' \sin i. \quad (47)$$

Substitution of relations (47) into Equation (46) yields

$$\delta = \frac{2n_1 d}{\cos i'} (1 - \frac{n_o}{n_1} \sin i \sin i') . \quad (48)$$

Since $n_o \sin i = n_1 \sin i'$, it follows that the optical path difference δ between rays R_1 and R_2, Figure 16.8, is given by

$$\delta = 2n_1 d \cos i' \quad (49)$$

where n_1 and d are, respectively, the refractive index and thickness of the plate, and i' is the indicated angle of refraction. Equation (49) is of great importance to the interpretation of interferometry with films and plates.

16.9.1.2 Let us suppose that the plate is immersed in a single medium. Then $n_2 = n_o$. It follows from the principles of the preceding section that the phase changes on reflection at B and C, Figure 16.8, differ by π radians. Thus,

$$\Delta = \frac{2\pi}{\lambda} 2n_1 d \cos i' + \pi \text{ radians} \quad (50)$$

where Δ is the total phase difference introduced between rays R_1 and R_2 due to the optical path difference δ and the phase changes on reflection. We have supposed tacitly that the angle i' is not so large that it is essential to distinguish sharply between normal and oblique incidence.

16.9.1.3 The optical path difference between the transmitted rays T_1 and T_2 is also given by δ as in Equation (49). More generally, the optical path difference between any two, consecutive reflected or transmitted rays, such as R_2 and R_3, is given by δ.

16.9.1.4 According to paragraph 16.8, the amplitude a of the wave reflected at points B in the first surface, Figure 16.8 will be

$$a_1 = \frac{|n_o - n_1|}{n_o + n_1} \quad (51)$$

The wave corresponding to rays R_2 is transmitted twice through the first surface in opposite directions and is reflected at points C. Hence, from Equations (44),

$$a_2 = \frac{2n_o}{n_o + n_1} \frac{|n_1 - n_o|}{n_1 + n_o} \frac{2n_1}{n_o + n_1} = \frac{4n_o n_1 |n_1 - n_o|}{(n_o + n_1)^3} . \quad (52)$$

If, for example, $n_o = 1$ and $n_1 = 1.5$, then $a_1 = 0.2$ and $a_2 = 0.192$ so that a_1 and a_2 are substantially alike. These two collinear waves interfere to produce the time-averaged energy density or illumination at O, Figure 16.8, proportional to W as given by Equation (3) with δ. From Equations (3) and (50)

$$2W = a_1^2 - 2 a_1 a_2 \cos\left(\frac{4\pi n_1 d}{\lambda} \cos i'\right) + a_2^2 . \quad (53)$$

The illumination produced by interference in the reflected beam can therefore be varied by changing any one of the following parameters:

(a) The optical thickness end of the plate
(b) The angle of refraction, i'
(c) The wavelength, λ.

The illumination at point O, Figure 16.8, is minimum when

$$4\pi n_1 d \frac{\cos i'}{\lambda} = \nu\, 2\pi; \quad \nu = 0, 1, 2, 3, \text{ etc.} \tag{54}$$

On the other hand, this illumination is maximum when

$$4\pi n_1 d \frac{\cos i'}{\lambda} = \mu\, \pi; \quad \mu \text{ an odd integer.} \tag{55}$$

The minima will be quite dark since a_1 and a_2 are substantially alike.

16.9.1.5 It is emphasized that with distant sources of light, the eye or telescope is focused for infinity, as illustrated in Figure 16.8, in order to observe the phenomena discussed in this section.

16.10 INTERFERENCE WITH PLANE PARALLEL PLATES AND NEARBY LIGHT SOURCES

16.10.1 Discussion of Problem.

16.10.1.1 The manner in which interference phenomena can be observed with nearby light sources is illustrated in Figure 16.9. Consider the coherent spherical wave that enamates from point S in the source. Suppose that the eye or camera is focused upon the upper surface of the plate and that the distances SD_2 and SB are large compared to the thickness d of the plate. A pair of rays SBR_2 and SD_2CBR_1 leaves point S and reaches point O in the manner indicated.

16.10.1.2 With point S as center and SD_2 as radius, draw arc D_2D_1. If the distance SD_2 is large and if the thickness d is relatively small, the arc D_2D_1 will be practically straight and perpendicular to SB. Moreover, the difference between angles i_1 and i_2 will be so small that either i_1 or i_2 or an intermediate angle, such as i, can be regarded as the angle of incidence together with i' as the angle of refraction. The optical path difference δ between rays SD_2CBO and SBO is

$$\delta = n_1 (BC + CD_2) - n_o BD_1. \tag{56}$$

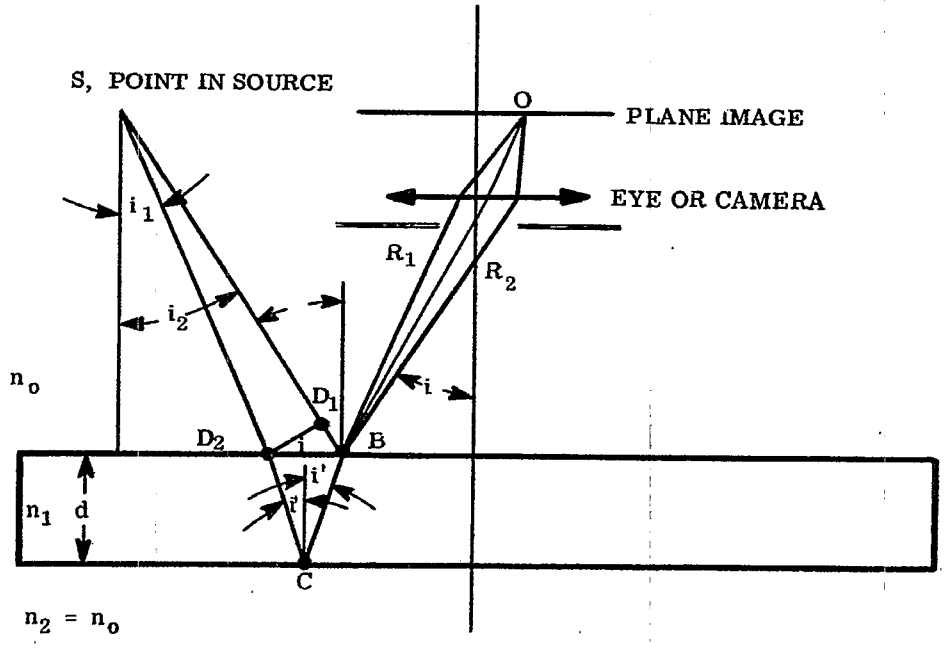

FIGURE 16.9 - The Parallel Plate Interferometer with nearby light sources.

Comparison of Equations (56) and (46) shows that they are alike. Moreover, comparison of points B, C, D_2, and D_1 in Figures 16.9 and 16.8 shows that they play similar roles. Hence

$$\delta = 2n_1 d \cos i' \tag{57}$$

as in Equation (49), and what has been said in the preceding section applies with excellent approximation to illumination with nearby sources provided that the thickness d of the plate is small as compared to the distance from the plate to the source.

16.11 HAIDINGER'S INTERFERENCE FRINGES

16.11.1 Interpretation of Haidinger's Fringes.

16.11.1.1 A simple arrangement for observing Haidinger's fringes is shown in Figure 16.10. The eye is preferably focused at infinity, where fringe contrast is best, but can be focused when desired on any suitable plane B.

16.11.1.2 The discussions in paragraphs 16.9 and 16.10 apply directly to the interpretation of Haidinger's fringes. In the interests of simplicity, let us accept the approximation $a_1 = a_2 = a$ in writing the energy density W of Equation (53) so that

$$2W = a^2 \left[2 - 2 \cos \left(\frac{4\pi n_1 d \cos i'}{\lambda} \right) \right].$$

Therefore, the energy density in the observed fringes is proportional to

$$W = a^2 \left[1 - \cos \left(\frac{4\pi n_1 d \cos i'}{\lambda} \right) \right], \tag{58}$$

with

$$\sin i = n_1 \sin i'. \tag{59}$$

Dark fringes or bright fringes are seen at angles i, Figure 16.10, for which cos i' obeys Equations (54) or (55), respectively. Since the angles i or i' are constant on circles about the axis AO, Haidinger's fringes are observed as circular fringes about an axis AO that moves with the observer's eye.

16.11.1.3 Suppose, for example, that $n_1 = 1.5$, $d = 1.8$mm, and $\lambda = 0.54 \times 10^{-3}$ mm. Then $2n_1 d/\lambda = 10^4$. Therefore, from Equation (54), $\nu = 10^4$ when $i' = 0$. The order number $\nu = 10^4$ is the highest possible order - and for it the central fringe is black. The next black fringe occurs when $\nu = 9999$, i.e., when

$$\cos i' = \frac{\frac{9999}{2n_1 d}}{\lambda} = 0.9999 \quad \text{or} \quad i' = 0.81^\circ.$$

Since $\sin i = n_1 \sin i'$, the angle i subtended at the observer by the radius of the first dark ring is 1.21°. Because the angular resolving power of the eye is approximately one minute of arc, plates much thicker than 1.8mm can be inspected for parallelism with the unaided eye by moving the plate along the arrow direction Q of Figure 16.10.

16.11.1.4 In applying Haidinger's fringes to the inspection of parallelism of plates, the distance from the eye to the plate should be made three feet, or so. The point O, Figure 16.10, can then be in the plate itself, i.e., one can focus his eye approximately upon the plate. Even though the plate may not be plane parallel, substantially circular Haidinger's fringes will be seen. The central Haidinger fringe will oscillate in brightness a number of times that depend upon the departure of the surfaces of the plate from parallelism as the plate is moved across the field of view in the Q-direction of Figure 16.10. At the central fringe, $\cos i' = 1$. Equation (54) now shows that when ν changes by unity (that is, as the central fringe changes from one state of blackness to the next), the corresponding change Δd in thickness is given by $2n_1 \Delta d/\lambda = 1$ or by

$$\frac{n_1 \Delta d}{\lambda} = \frac{1}{2}. \tag{60}$$

This means that each time the central fringe passes through one cycle, the optical path through the plate has changed (as could be expected without the aid of the theory) by one-half wavelength. Counting the number of blinks of the central Haidinger fringe forms a sensitive and simple method for measuring the amount of departure from parallelism of a plate.

16.11.1.5 It is worth noting that Haidinger's fringes are essentially fringes of equal inclination. Each fringe corresponds to a definite angle i' of inclination. Changes in i' (rather than changes in λ or $n_1 d$) govern the observed changes in the fringes.

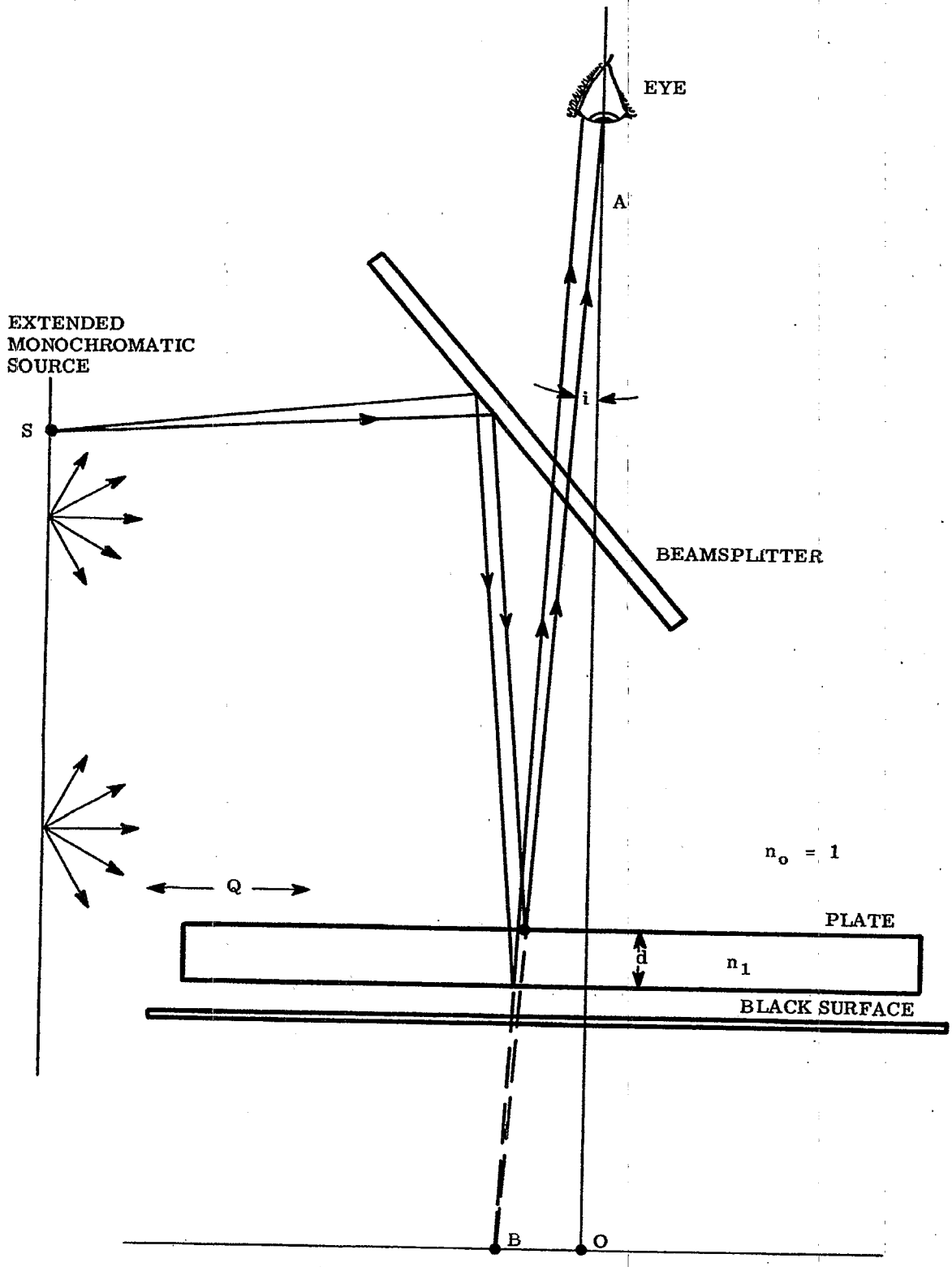

FIGURE 16. 10 -Arrangement for observing Haidinger's Fringes.

16.12 FIZEAU FRINGES

16.12.1 Introduction.

16.12.1.1 The fringes seen with the arrangement shown in Figure 16.11 under illumination from an extended and fairly monochromatic source are often called Fizeau fringes or Fizeau bands. These fringes are similar in formation to those obtained in Michelson's interferometer of Figure 16.12. The method of Figure 16.11 is used widely for testing one polished surface against another for flatness or for sphericity. The reference surface, S_2, may be flat or spherical.

16.12.1.2 Owing to the presence of dust, surfaces S_1 and S_2 will ordinarily be inclined so that the space between them is approximated by an air wedge whose angle θ is constant only when both surfaces are plane. Figure 16.13 illustrates how the Fizeau fringes can appear localized in a chosen plane containing point P. Note that each point P receives coherent light from a corresponding point S in the source. Each point P is, in effect, illuminated by a different point in the source. An extended source becomes necessary for viewing fringes over an extended surface S_1.

16.12.1.3 It will be observed that Figures 16.13 and 16.9 are so similar that they become identical when $\theta = 0$. The argument leading to Equation (57) for the optical path difference δ between rays SPP' and SQPP' applies again with excellent approximation provided that one takes for d the thickness of the wedge at point P as indicated in Figure 16.13. Equations (54) and (55) govern the location of the fringes. Minima occur where

$$2n_1 \, d \cos i' = \nu \lambda ; \quad \nu = 0, 1, 2, 3, \text{ etc.}, \tag{61}$$

and maxima occur where

$$4n_1 \, d \cos i' = \mu\lambda ; \quad \mu \text{ an odd integer.} \tag{62}$$

Since the space between S_1 and S_2 is usually air,

$$n_1 = 1$$
$$i' = i \tag{63}$$

where i is the angle of incidence.

16.12.1.4 The advantage of simplicity obtained through the use of Fizeau fringes rests upon the fact that variations in the angle of incidence i, Figure 16.11, have negligibly small effects upon the location of the fringes because the separations d between S_1 and S_2 are small. Suppose, for example, that $d = 10\lambda$. The maximum value of ν occurs at $i' = 0$, and here $\nu = 20$ from Equation (61) since $n_1 = 1$. If d and λ remain constant as point P moves away from point O, the next dark fringe occurs at $\nu = 19$ so that $\cos i' = \cos i = 19/20$. Correspondingly, $i = 18.19^{\circ}$. Let the distance AO from the eye to the test plate be made so large relative to the lateral dimension of the test plate that the maximum value of i cannot exceed 4°. The variation of 4° is obviously small compared to the amount 18.19° required for decreasing $d \cos i'$ by the amount $\lambda/2$. In fact when $0 \leq i' \leq 4^{\circ}$, $0.9976 \leq \cos i' \leq 1$. Hence, $10\lambda \geq d \cos i' \geq 9.976\lambda$. This means that $d \cos i'$ cannot change by more than 0.034 wavelengths due to any variation of the angle of incidence when i_{max} is constrained to 4° by the choice of the distance AO. If, therefore, one arranges to observe the Fizeau fringes at normal incidence, he is justified in setting $\cos i' = 1$ in Equations (61) and (62) and accepting the well known approximation that the separation d changes by the amount $\lambda/2$ in passing, for example, from one bright fringe to the next. Each fringe may be regarded as the locus of points for which the separation of the surfaces S_1 and S_2, Figure 16.11, is constant.

16.13 NEWTON'S RINGS AND NEWTON'S FRINGES

16.13.1 Interpretation of Newton's Fringes.

16.13.1.1 An experimental arrangement for observing Newton's rings or fringes is illustrated in Figure 16.14 In honor of Sir Isaac Newton, the colored circular fringes seen around the point O with white light sources are called Newton's rings. The central fringe is black at O when S_1 and S_2 are substantially in contact because there is a phase difference of one-half vibration between the reflections at S_1 and S_2. It is preferable for most purposes to view the interference bands with monochromatic sources. These circular bands are often called Newton's Fringes. Comparison of Figures 16.11 and 16.14 shows that Fizeau's and Newton's fringes can become practically identical.

16.13.1.2 It was seen in the previous section that a Fizeau fringe can be regarded with good approximation as the locus of points for which the separation of the surfaces S_1 and S_2 is constant. We may take the view that

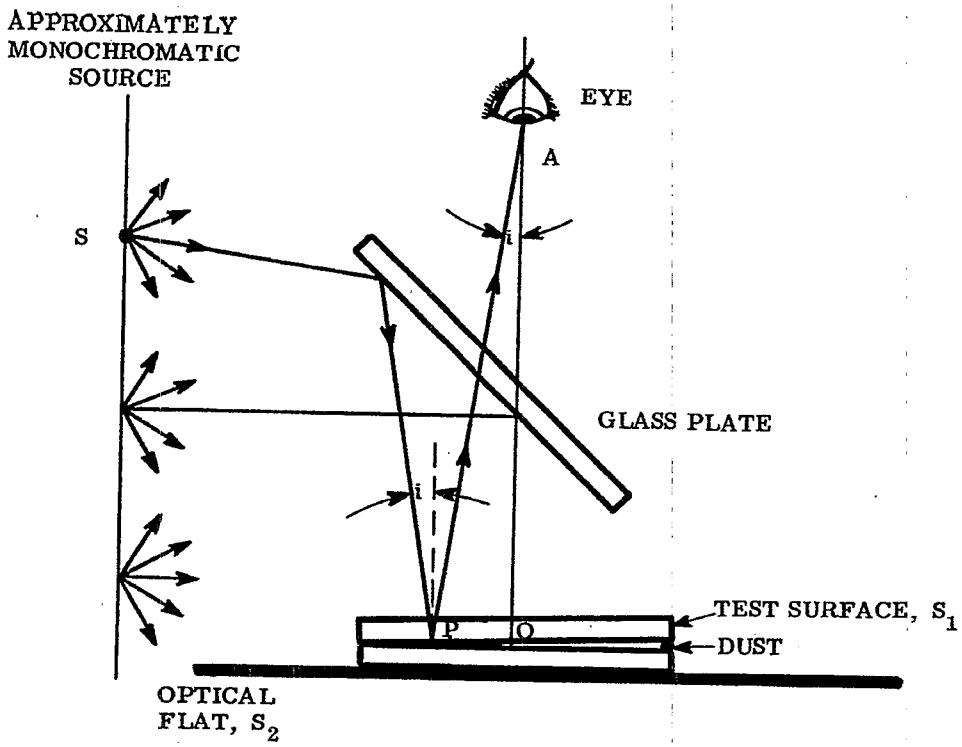

FIGURE 16. 11- Method for obtaining Fizeau Fringes.

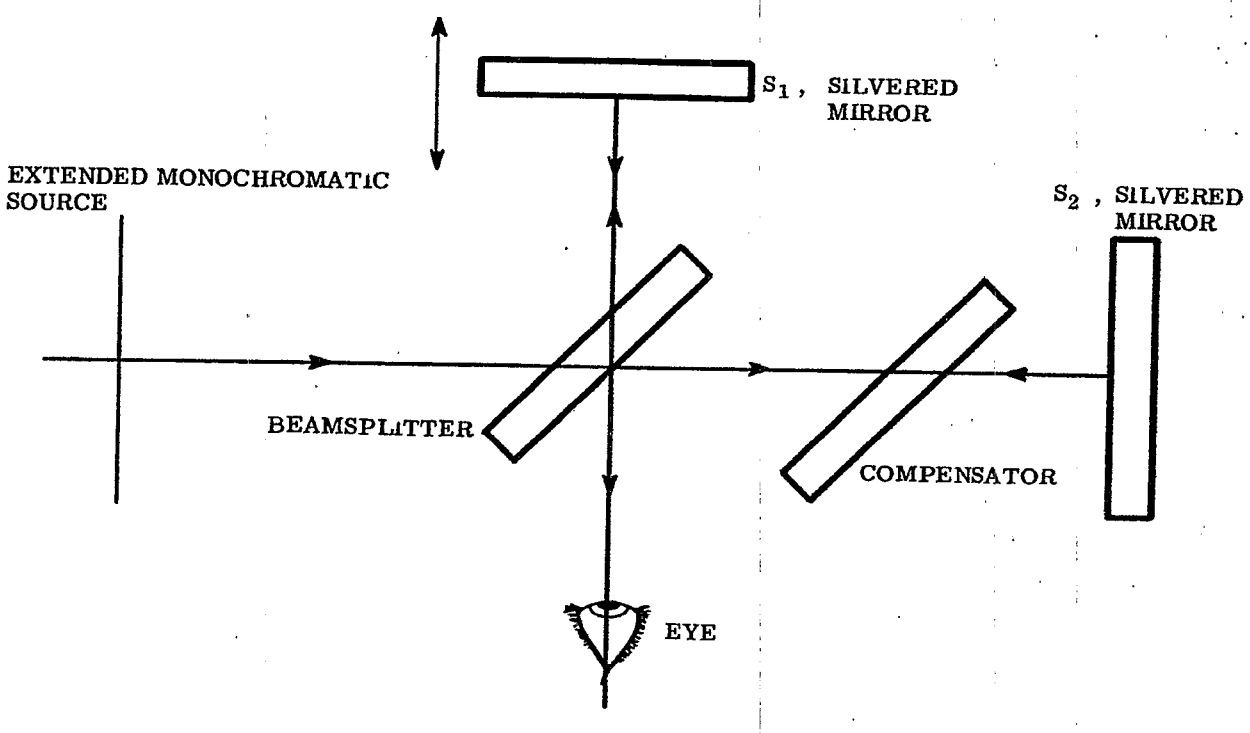

FIGURE 16. 12-Michelson's Interferometer.

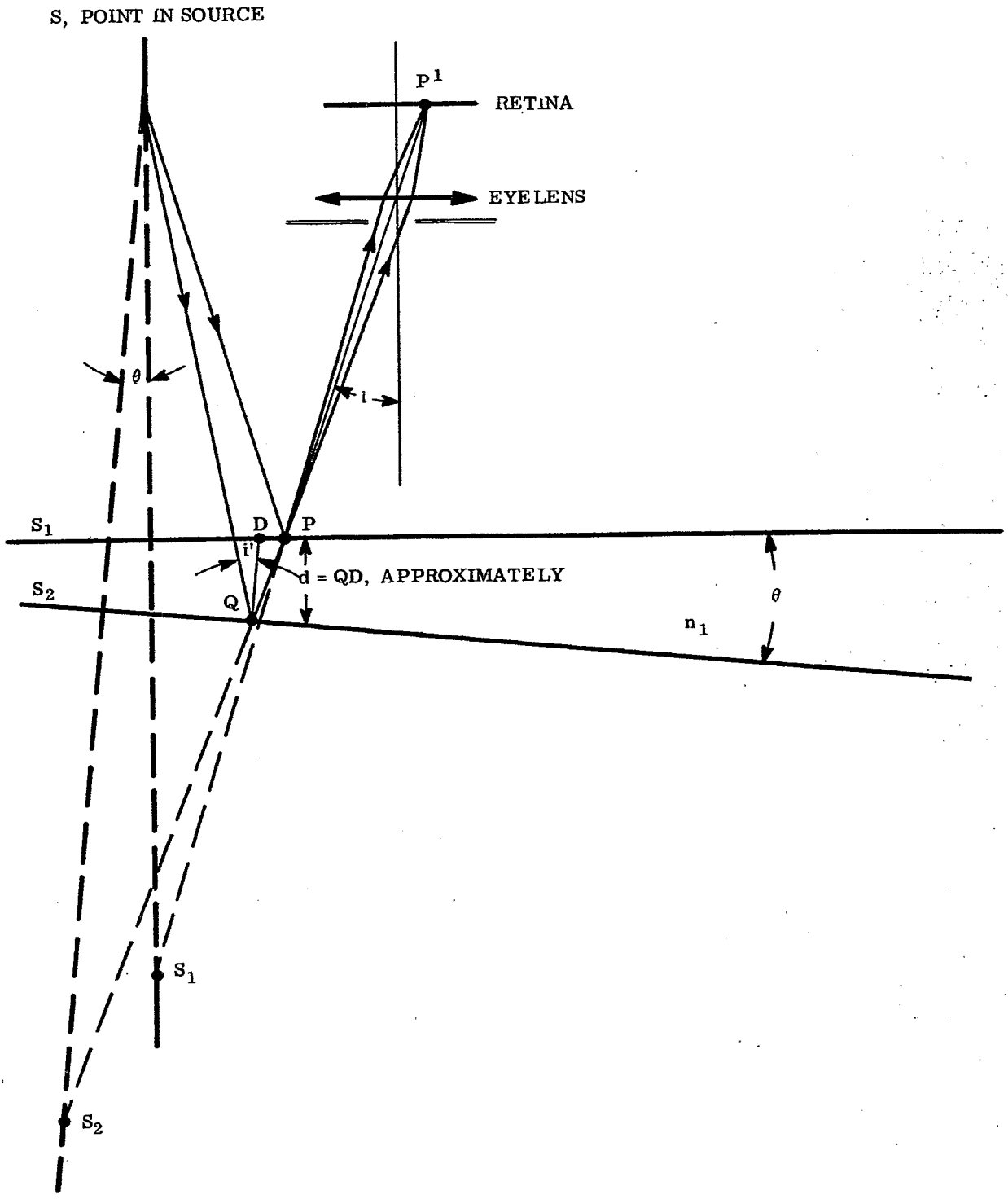

FIGURE 16. 13 - Construction showing how the Fizeau Fringes can appear localized at points, P, near the reflecting surfaces S_1 and S_2

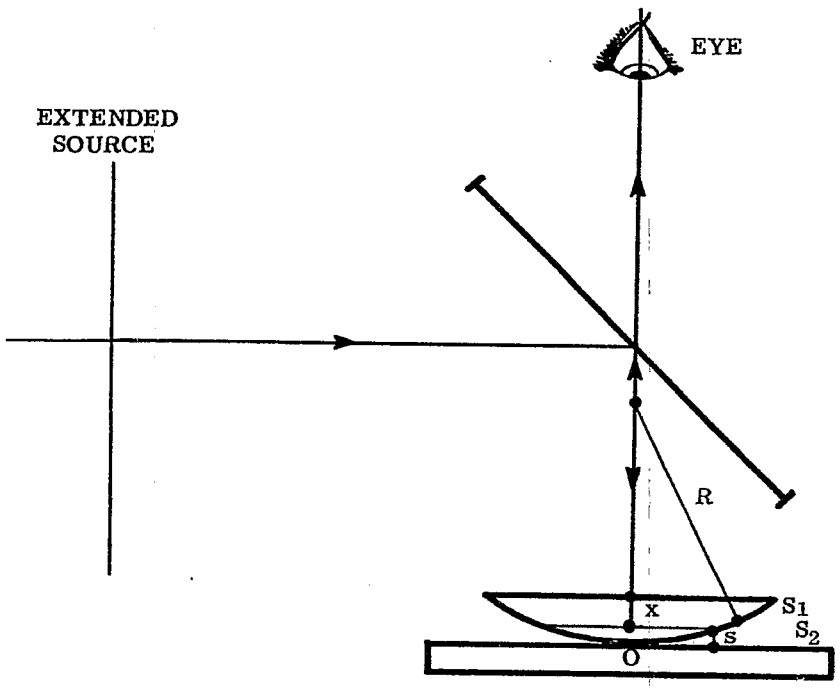

FIGURE 16. 14-Arrangement for obtaining Newton's Rings or Newton's Fringes.

Newton's fringes are Fizeau fringes along which the sagitta s of Figure 16. 14 is constant such that dark fringes occur when

$$s = \nu \frac{\lambda}{2} ; \quad \nu = 0, 1, 2, 3, \text{ etc.}, \tag{64}$$

and such that bright fringes occur when

$$s = \mu \frac{\lambda}{4} ; \quad \mu = 1, 3, 5, \text{ etc.} \tag{65}$$

The sagitta s obeys the relation $x^2 = 2Rs - s^2$, where R is the radius of the surface. By neglecting s^2 in comparison with 2Rs, one obtains the approximation

$$x = \sqrt{2Rs} \ . \tag{66}$$

Thus, from Equations (64), (65), and (66)

$$x_\nu = \sqrt{\nu R \lambda} ; \quad \nu = 0, 1, 2, 3, \text{ etc.}, \tag{67}$$

where x_ν are the radii of the dark fringes and

$$x_\mu = \sqrt{\mu R \lambda / 2} ; \quad \mu = 1, 3, 5, \text{ etc.}$$

where the x_μ are the radii of the bright fringes. The radius R of the surface can be computed from the measured values of the radii x_ν or x_μ.

16.13.1.3 The adoption of the theory of Fizeau fringes to Newton's fringes is, in itself, an approximation. The method of the sagitta should be regarded merely as a first approximation to the interpretation of Newton's fringes with extended sources of light. More critical investigations reveal that the choice of observation plane matters, as does also the location of the eye with respect to the points x_ν or x_μ.

16.13.1.4 In viewing both Fizeau and Newton's fringes, the tendency and practice is to focus upon the thin film between the interferometer surfaces S_1 and S_2.

16.13.1.5 For increased accuracy in using the sagitta method for determining the radius R_1, it is preferable to choose as the reference surface S_2 a spherical surface of known radius R_2 that departs only slightly from R_1. The "effective sagitta" s, Figure 16.15, is now given by $s = s_1 - s_2$ in which $x^2 = 2R_1 s_1 - s_1^2 = 2R_2 s_2 - s_2^2$. By neglecting s_1^2 and s_2^2 in comparison with $2R_1 s_1$ and $2R_2 s_2$, respectively, one obtains

$$s = \frac{x^2}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \frac{x^2}{2}\left(\frac{R_2 - R_1}{R_1 R_2}\right). \tag{68}$$

Thus, from Equations (68) and (64), dark fringes occur at radii x_ν for which

$$x_\nu = \sqrt{\left(\frac{R_1 R_2}{R_1 - R_2}\right)\nu\lambda}, \tag{69}$$

a result that reduces to Equation (67) when $R_2 = \infty$. If R_1 and R_2 are nearly alike, one may set $R_1 R_2 = R_2^2$. Within the validity of this approximation,

$$R_2 - R_1 = \left(\frac{R_2}{x_\nu}\right)^2 \nu\lambda. \tag{70}$$

16.13.1.6 Systematic error of interpretation of Newton's fringes due to inadequacies of the sagitta method can be avoided or minimized, as will now be shown, by replacing one of the end-mirrors of the Tyman Green interferometer by the spherical surface S_1 as illustrated in Figure 16.16.

16.13.1.7 We suppose that the end-mirror S_1 has a large radius R and seek to compute R from the radii of the circular fringes seen about point O when the eye lens and telescope focus the plane z = 0 upon the retina. We may suppose for simplicity that pinhole H and the center C of spherical surface S_1 fall upon the axis of the instrument. We take plane z = 0 through point O as the plane of reference. The plane wave reflected from S_2 appears to return to the observer as a plane wave along the direction OZ. The wave reflected from S_1 appears (apart from spherical aberration produced on reflection) as a spherical wave that expands from point F located at distance R/2 behind point O. We suppose that the distances x are small enough that spherical aberration on reflection can be ignored.

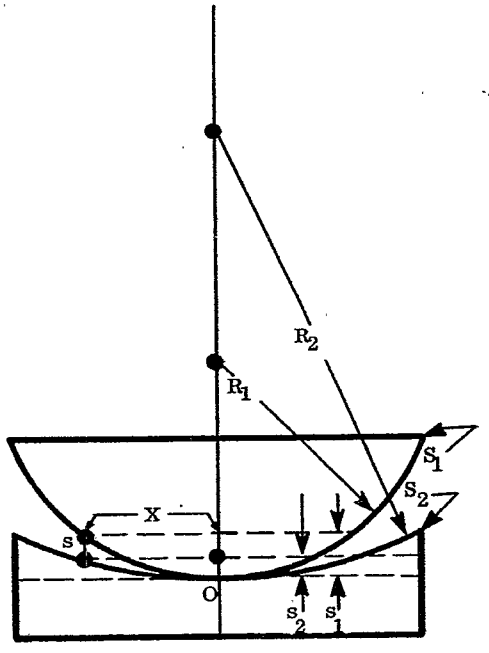

FIGURE 16.15--The Sagitta Method when the Reference Surface S_2 is a sphere.

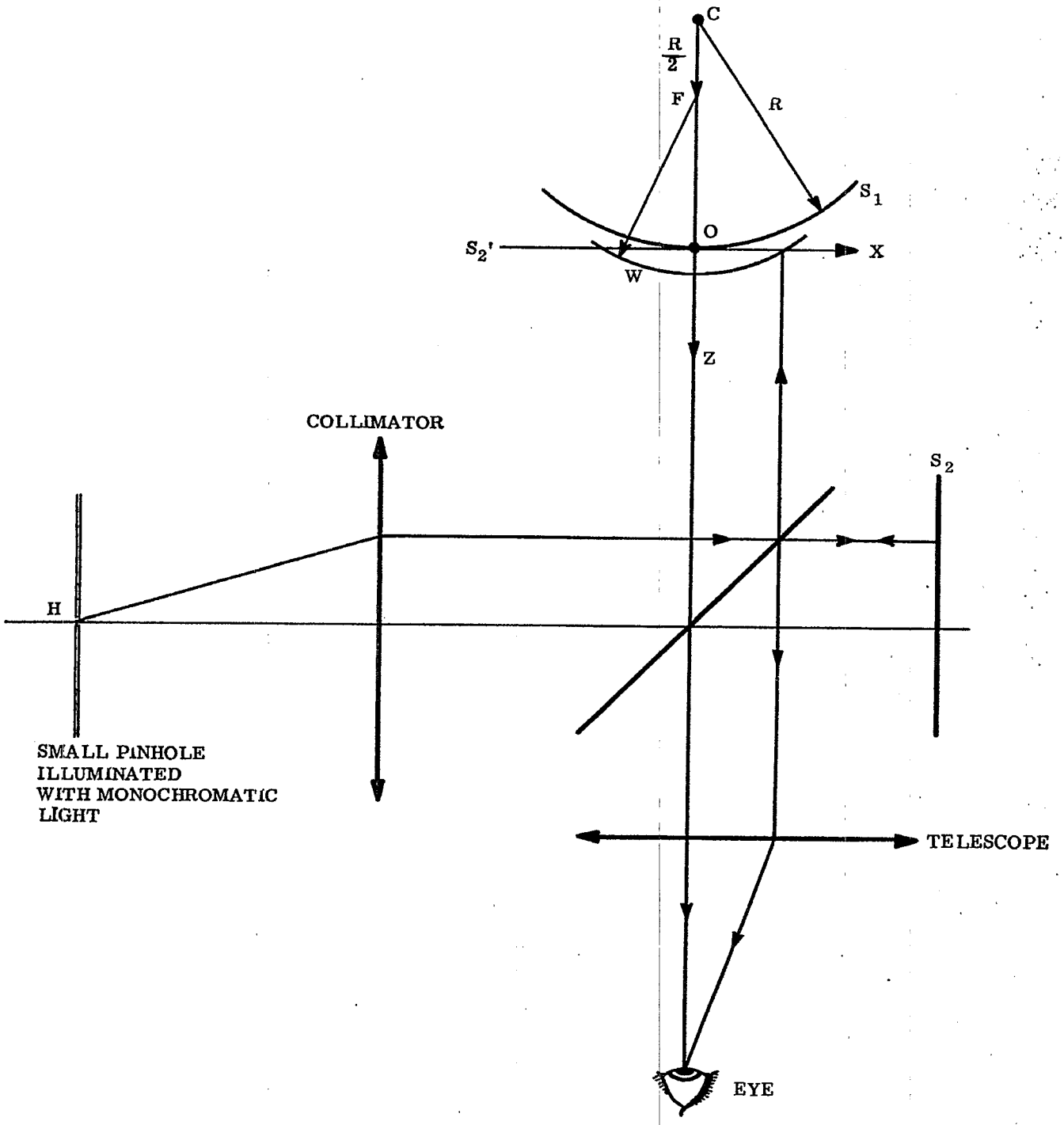

FIGURE 16. 16-Formation of Newton's Fringes with a Tyman Green Interferometer.

16.13.1.8 The plane wave returned to the observer has the form

$$E_2 = a_2 \cos (\delta_o + kz - \omega t) \tag{71}$$

where a_2 denotes the amplitude and δ_o has been added in order to account for the difference in phase change on reflection. The spherical wave returned to the observer has the form

$$E_1 = a_1 \cos \left\{ \delta_1 + k \left[r^2 + (z - \frac{R}{2})^2 \right]^{1/2} - \omega t \right\} \tag{72}$$

in which

$$r \equiv (x^2 + y^2)^{1/2} . \tag{73}$$

On the circle $x^2 + y^2 = r^2$ in the plane of observation $z = 0$ the phase difference $\phi_1 - \phi_2$ between E_1 and E_2 is given by

$$\phi_1 - \phi_2 = \delta_1 - \delta_o + k \left(r^2 + \frac{R^2}{4} \right)^{1/2} . \tag{74}$$

But at point O, where $x = y = z = 0$, $\phi_1 - \phi_2$ must equal $-\delta_o$ because the separation of S_1 and S_2' is zero. Hence, with respect to the undetermined value of δ_1 ,

$$\delta_1 = - k \frac{R}{2} = - \pi \frac{R}{\lambda}$$

so that

$$\phi_1 - \phi_2 = - \delta_o - \frac{\pi R}{\lambda} + \frac{2\pi}{\lambda} \left[r^2 + \frac{R^2}{4} \right]^{1/2} . \tag{75}$$

The time-averaged energy density on circles of radii r the plane $z = 0$ is given by Equation (3) wherein $\phi_1 - \phi_2$ obeys Equation (75). It follows from Equations (3) and (75) that the fringes display maximum brightness at r-values for which

$$- \delta_o - \frac{\pi R}{\lambda} + \frac{2\pi}{\lambda} \left[r_\nu^2 + \frac{R^2}{4} \right]^{1/2} = \nu 2\pi ; \quad \nu = 0, 1, 2, 3, \text{ etc.}, \tag{76}$$

and minimum brightness at r-values for which

$$- \delta_o - \frac{\pi R}{\lambda} + \frac{2\pi}{\lambda} \left[r_\mu^2 + \frac{R^2}{4} \right]^{1/2} = \mu \pi ; \quad \mu = 1, 3, 5, \text{ etc.} \tag{77}$$

Equations (76) and (77) enable one to compute both δ_o and R from measured values of r_ν and r_μ in cases where δ_o is not known.

16.13.1.9 Either the Twyman Green interferometer or the Fizeau interferoscope of Figure 16.2 may be used. With Fizeau interferoscopes, $\delta_o = \pi$ in Equations (76) and (77). With Twyman-Green interferometers, $\delta_o = 0$ when the end-mirrors are unsilvered or equally silvered surfaces of glass.

16.13.1.10 The exact form of Equations (76) and (77) will rarely be required. The following excellent approximation leads to a much simpler pair of working relations. We write

$$\left[r_\nu^2 + \frac{R^2}{4} \right]^{1/2} = \frac{R}{2} \left[1 + \frac{4 r_\nu^2}{R^2} \right]^{\frac{1}{2}}$$

It will be impractical to utilize either the Twyman-Green or Fizeau interferometers unless the radius R of the test surface is so great that $1 >> 4r_\nu^2 / R^2$ and that

$$\frac{R}{2} \left[1 + \frac{4 r_\nu^2}{R^2} \right]^{\frac{1}{2}} \approx \frac{R}{2} \left[1 + \frac{2 r_\nu^2}{R^2} \right] = \frac{R}{2} + \frac{r_\nu^2}{R} . \tag{78}$$

By combining Equations (76) and (77) with Equation (78), one obtains the simplified results

$$- \delta_o + \frac{2\pi r_\nu^2}{\lambda R} = \nu 2\pi ; \quad \text{(bright fringes)} \tag{79}$$

$$- \delta_o + \frac{2\pi r_\mu^2}{\lambda R} = \mu \pi ; \quad \text{(dark fringes).} \tag{80}$$

If, for example, $\delta = \pi$ as in the Fizeau interferoscope,

$$2\pi r_\mu^2 / \lambda R = (\mu + 1)\pi \quad \text{so that} \quad r_\mu^2 = \lambda R (\mu + 1)/2.$$

Consequently, for circular dark fringes

$$r_\mu = \sqrt{\lambda R (\mu + 1)/2}\ . \tag{81}$$

Since μ is an odd integer, $(\mu + 1)$ is an even integer, and $(\mu + 1)/2$ generates the integers 0, 1, 2, 3, etc. (To obtain the zero-value, one takes $\mu = -1$.)

16.13.1.11 Comparision of Equations (81) and (67) shows that they agree. This means that the sagitta method is more reliable as applied to measuring R in the Twyman-Green or the Fizeau interferometers than it is likely to be as applied to methods based upon Fizeau fringes or Newton's fringes. This conclusion is not surprising because the Fizeau and Twyman-Green interferometers utilize small sources of light and are constructed so that the observer is forced to view the fringes under conditions of normal incidence.

16.14 COMPLEX NUMBERS

16.14.1 Introduction.

16.14.1.1 Many of the following discussions are both shorter and more readily understood by employing complex numbers instead of the trigonometric functions. Only the most elementary properties of complex numbers will be needed.

16.14.1.2 One well known method of expressing a complex number Z is illustrated by the equation

$$Z = a + i\,b \tag{82}$$

wherein a and b are real numbers and $i = \sqrt{-1}$. The real numbers a and b are often called the real and imaginary parts, respectively. The so-called complex conjugate $\overline{Z}$ of Z is defined by the relation

$$\overline{Z} = a - i\,b. \tag{83}$$

It follows at once that

$$a = \frac{Z + \overline{Z}}{2} \equiv R_e(Z), \quad \text{the real part of } Z \tag{84}$$

and that

$$b = \frac{Z - \overline{Z}}{2i} \equiv I_m(Z)\ , \quad \text{the imaginary part of } Z. \tag{85}$$

Furthermore,

$$|Z|^2 = Z\,\overline{Z} = a^2 + b^2 \ ; \quad (i^2 = -1) \tag{86}$$

where $|Z|$ is the absolute value or amplitude of Z, i. e., the length of Z as illustrated in Figure 16. 17

16.14.1.3 For our purposes the exponential form of Z is much to be preferred. Thus,

$$Z = |Z|\, e^{i\theta} \tag{87}$$

16.14.1.4 By definition,

$$|Z|\, e^{i\theta} = |Z|\ (\cos\theta + i \sin\theta) \tag{88}$$

where the angle θ, illustrated in Figure 16. 17, is called the argument of Z and written arg(Z). It follows by comparison of Equations (82) and (83) that

$$a = |Z|\ \cos\theta\ ; \quad b = |Z|\ \sin\theta\ ; \tag{89}$$

consequently,

$$\tan\theta = \frac{b}{a}\ . \tag{90}$$

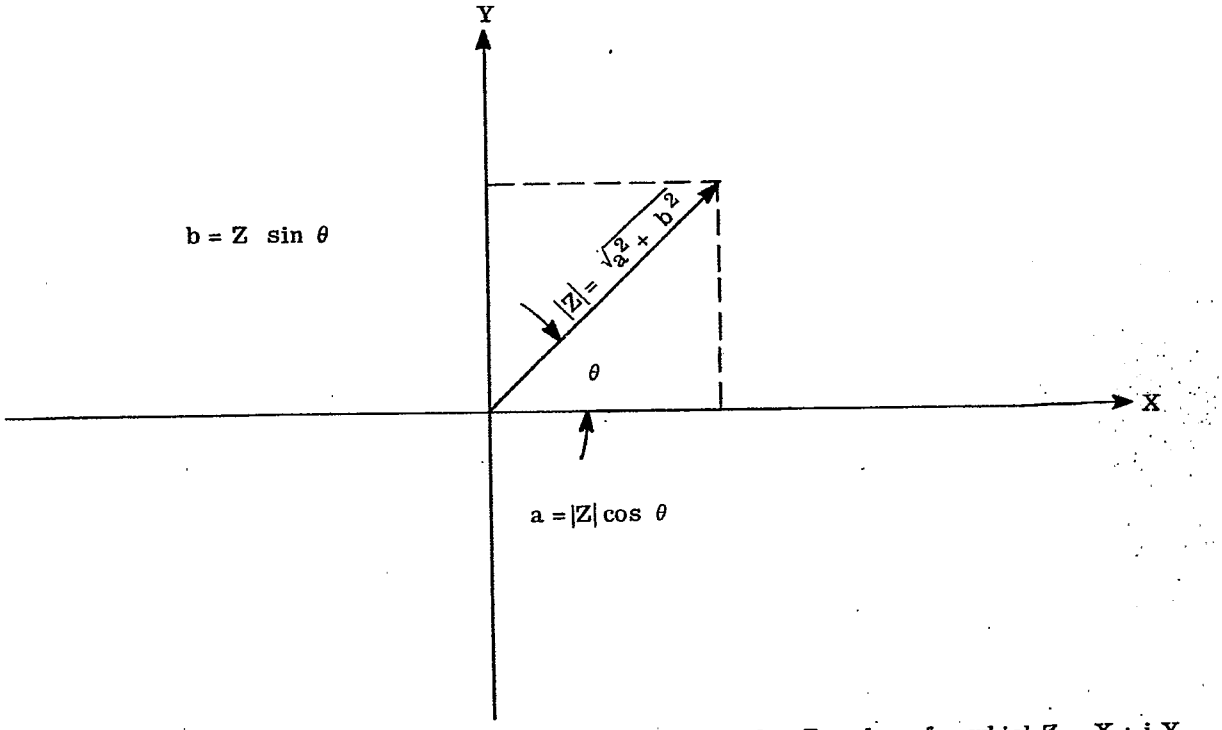

FIGURE 16. 17-Representation of complex numbers in the complex Z - plane for which Z = X + i Y.

16.14.1.5 That Equation (88) is a reasonable definition can be seen from the following considerations. From the series

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$$

and

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

we obtain

$$\cos\theta + i \sin\theta = 1 + i\theta + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \ldots$$

which is in the form of

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

wherein $x = i\theta$, and $e^x = e^{i\theta}$.

16.14.1.6 Given two complex numbers Z_1 and Z_2 in exponential form, their product Z is given by

$$Z = |Z_1| e^{i\theta_1} |Z_2| e^{i\theta_2} = |Z_1| |Z_2| e^{i(\theta_1 + \theta_2)} \quad (91)$$

The rule for multiplying two complex numbers is to multiply their amplitudes and to add their arguments. Similarly with respect to division,

$$Z = |Z_1| e^{i\theta_1} \Big/ |Z_2| e^{i\theta_2} = \frac{|Z_1|}{|Z_2|} e^{i(\theta_1 - \theta_2)} \quad (92)$$

16.14.1.7 Finally, if $Z = |Z| e^{i\theta}$,

$$\bar{Z} = |Z| e^{-i\theta}. \tag{93}$$

Consider, for example, the statement

$$E = ae^{i(\phi - \omega t)} = a\cos(\phi - \omega t) + i \sin(\phi - \omega t). \tag{94}$$

We see that the wave form of Equation (1) is the real part of E as expressed by the complex form of Equation (94). This means that, when desired, the instantaneous value of E can be computed as the real part, R (E), of E as given by Equation (94). However, one's chief interests center finally upon the time-averaged energy density. From Equation (94)

$$|E|^2 = E\bar{E} = a^2. \tag{95}$$

By comparing Equations (95) and (2), we find that

$$2W = |E|^2 = E\bar{E} \tag{96}$$

where W is the time-averaged density. This property of the complex wave form is of great convenience.

16.14.1.8 Suppose that the complex wave traverses a medium whose amplitude transmittance is τ and whose phase transmittance (optical path) is nd. We can write the transmittance of this medium in the complex form

$$T = \tau e^{ind} \tag{97}$$

If E is given by Equation (94) upon entry into the medium, then if E' denotes the value of E as the wave leaves the medium

$$E' = TE = \tau a e^{i(\phi + nd - \omega t)} \tag{98}$$

Similarly, if the wave corresponding to Equation (94) is reflected from an interface between two media

$$E' = \rho E = |\rho| e^{i(\phi + \psi - \omega t)} \tag{99}$$

in which $\rho = |\rho| e^{i\psi}$ wherein ρ denotes amplitude reflectance and ψ denotes the phase change upon reflection.

16.15 TRANSMITTANCE OF PLANE PARALLEL PLATES

16.15.1 Introduction.

16.15.1.1 The simplified treatment of paragraph 16.9 applies with excellent approximation to plates whose surfaces have low reflectance. As the reflectance of the surfaces increases, the effects of the inter-reflected beams ultimately dominate and exert, as we shall see, profound effects upon the distribution of energy density in the observed fringes. The most conspicuous of these effects is a pronounced sharpening of the fringes to the point where they can appear as narrow bright lines on a dark background in transmitted light. These narrow fringes can be utilized to obtain more accurate measurements of surface irregularities, etc., than is possible with the sinusoidal fringes that are produced by double beam interferometers, such as the Michelson interferometer or the Fizeau interferoscope.

16.15.1.2 The theory of this paragraph applies directly to the Fabry-Perot and related multiple beam interferometers.

16.15.1.3 With respect to Figure 16.18:

n_1 = refractive index of the plate
n_o = refractive index of medium of incidence
n_2 = refractive index of last medium
d = thickness of the plate
i = angle of incidence
i' = angle of refraction
τ_1 = internal amplitude transmittance of the plate
$\tau_{0,1}$ = amplitude transmittance from the 0^{th} into the 1^{st} medium
$r_{1,0}$ = amplitude reflectance from the 1^{st} upon the 0^{th} medium
$\delta_{1,0}$ = phase change on reflection associated with $r_{1,0}$
$\tau_{1,2}$ = amplitude transmittance from 1^{st} into the 2^{nd} medium
$r_{1,2}$ = amplitude reflectance from the 1^{st} upon the 2^{nd} medium
$\delta_{1,2}$ = phase change on reflection associated with $r_{1,2}$

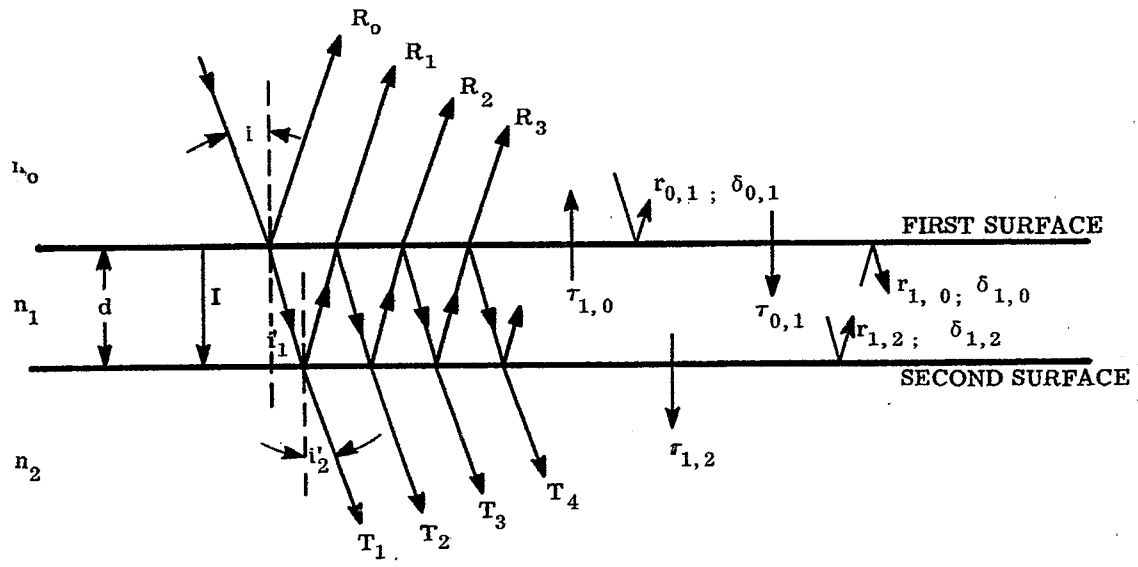

FIGURE 16. 18-Convention with respect to the transmitted beam in a plate or Fabry-Perot interferometer.

16. 15.1.4 We bear in mind that $\tau_{0,1}$, $\rho_{1,0} = r_{1,0}\, e^{i\delta_{1,0}}$, $\tau_{1,2}$ and $\rho_{1,2} = r_{1,2}\, e^{i\delta_{1,2}}$ are Fresnel coefficients that depend in general upon i and upon whether the incident E-vector vibrates in, or perpendicular to the plane of the paper.

$$n_o \sin i = n_1 \sin i'_1 = n_2 \sin i'_2 \,. \tag{100}$$

The optical path difference between any two rays T_j and T_{j+1} will be $2n_1 d \cos i'_1$ (see paragraph 5.10). Let

$$\alpha \equiv \frac{4\pi n_1}{\lambda} \, d \cos i'_1 + \delta_{1,0} + \delta_{1,2} \tag{101}$$

and let β be the optical path for the directly transmitted beam T_1. Then, under the supposition that the incident beam has the amplitude unity,

$$T_1 = \tau_{0,1}\, \tau_1\, \tau_{1,2}\, e^{i\beta}$$

$$T_2 = \tau_{0,1}\, \tau_1^3\, \tau_{1,2}\, e^{i\beta}\, r_{1,2}\, r_{1,0}\, e^{i\alpha}$$

$$T_3 = \tau_{0,1}\, \tau_1^5\, \tau_{1,2} e^{i\beta}\, (r_{1,2}\, r_{1,0})^2\, e^{i\,2\alpha}\,; \quad \text{etc.}$$

If we consider N inter-reflections so that there are N emergent rays T_j, the emergent wave is now determined from the scalar quantity

$$E = e^{-i\omega t} \sum_{\nu=1}^{N} T_\nu = \tau_{0,1}\ \tau_1\ ,\ \tau_{1,2}\ e^{i\beta} e^{-i\omega t} \sum_{\nu=0}^{N} A^\nu\ e^{i\nu\alpha} \qquad (102)$$

where

$$A = \tau_1^2\ \ r_{1,0}\ ,\ r_{1,2} \leq 1. \qquad (103)$$

But

$$\sum_{\nu=0}^{N} A^\nu\ e^{i\nu\alpha} = \frac{1 - A^{N+1}\, e^{i\alpha(N+1)}}{1 - A\ e^{i\alpha}} \qquad (104)$$

Therefore,

$$E = \tau_{0,1}\ ,\ \tau_1\ ,\ \tau_{1,2}\ e^{i(\beta - \omega t)} \frac{1 - A^{N+1}\ e^{i\alpha(N+1)}}{1 - A\ e^{i\alpha}}. \qquad (105)$$

16.15.1.5 The time-averaged energy density 2W is proportional to $|E|^2 = E\ \bar{E}$. It is obtained in a straightforward manner from Equation (105). The result is

$$2W = (\tau_{0,1}\ ,\ \tau_1\ ,\ \tau_{1,2})^2 \left\{ \frac{1 - 2\,A^{N+1} \cos\ [(N+1)]\alpha\ + A^{2(N+1)}}{1 - 2\,A \cos\,\alpha + A^2} \right\}. \qquad (106)$$

wherein α and A are given by Equations (101) and (103), respectively.

16.15.1.6 In a thick plate the number N of inter-reflections is restricted by the length of the incident wave train or by the tendency of each successive reflection to "walk" the beam out through the ends of the plate. However, with thin films, such as soap films, or with evaporated films one is usually justified in setting $N = \infty$. Whenever one can accept the approximation $N = \infty$, the time-averaged energy density W in the transmitted beam is given by the simpler expression

$$2W = \frac{(\tau_{0,1}\ \tau_1,\ \tau_{1,2})^2}{1 - 2\,A \cos\,\alpha + A^2}. \qquad (107)$$

16.15.1.7 With respect to both Equations (106) and (107), major maxima occur in the transmitted fringes when

$$\alpha = \nu\, 2\pi\ ;\quad \nu = 0,\ 1,\ 2,\ 3,\ \text{etc.} \qquad (108)$$

This result can be expected intuitively; for it requires that all rays T_j of Figure 16.18 shall emerge in phase.

16.15.1.8 The integers ν are often called spectral orders.

16.15.1.9 Equation (106) for N transmitted rays T_j differs from Equation (107) in that it predicts the existence of N + 1 subsidiary maxima between any two consecutive spectral orders ν and $\nu + 1$.

16.15.1.10 When $A = \tau_1^2\ \ r_{1,0}\ \ r_{1,2}$ becomes small in Equations (106) and (107),

$$2W \rightarrow \frac{(\tau_{0,1}\ \tau_1\ \tau_{1,2})^2}{1 - 2\,A \cos\,\alpha} \rightarrow (\tau_{0,1}\ \tau_1\ \tau_{1,2})^2\ (1 + 2\,A \cos\,\alpha)\ . \qquad (109)$$

This means that the transmitted fringes assume the sinusoidal distributions typical of double beam interferometers when A becomes small due to reduction of the internal transmittance τ_1 of the plate or of the amplitude reflectances $r_{1,0}$ and $r_{1,2}$ of its surfaces. Contrast in the transmitted fringes will be poor when A is so small that Equation (109) is an acceptable approximation to Equation (107).

16.15.1.11 It is more difficult to demonstrate that Equations (107) and (106) predict the appearance of sharp fringes

as A approaches unity. Let the energy density W be plotted against α as in Figure 16.19. At $\alpha = \nu\, 2\pi$,

$$W = W_{max} = \frac{B^2}{(1 - A)^2} \tag{110}$$

in which $B = \tau_{0,1}\, \tau_1\, \tau_{1,2}$. We wish to find the neighboring value of α for which $W = W_{max}/2$. Set

$$\alpha = \nu\, 2\pi + \Delta\alpha \tag{111}$$

and suppose that $\Delta\alpha$ is so small that $\cos\alpha = \cos \nu\, 2\pi \cos \Delta\alpha = 1 - (\Delta\alpha)^2/2$. Then for $W = W_{max}/2$ from Equation (107), $B^2/\left[1 - 2A + A^2 + A(\Delta\alpha)^2\right] = B^2/\left[2(1 - A)^2\right]$ so that $A(\Delta\alpha)^2 = (1 - A)^2$. Hence,

$$\Delta\alpha = \frac{1 - A}{\sqrt{A}} \tag{112}$$

where $\Delta\alpha$ is the increment that must be added to $\alpha = \nu\, 2\pi$ in order to drop W from W_{max} to $W_{max}/2$. If α is increased by 2π, the next fringe for which $W = W_{max}$ is obtained. In other words, the fringe-width is 2π in terms of α. We define

$$w = \frac{\Delta\alpha}{2\pi} = \frac{1}{2\pi}\quad\frac{1 - A}{\sqrt{A}} \tag{113}$$

and call it the optical half-width of the Fabry-Perot fringes. We see that this optical half-width decreases rapidly as A approaches unity. If, for example, $A = 0.9$, $2w = 0.032$. This means that the width 2w, Figure 16.19, is approximately 0.03 times the width from one bright fringe to the next. The fringes become exceedingly sharp as A approaches unity. A-values of 0.9 are obtained easily by silvering the two surfaces

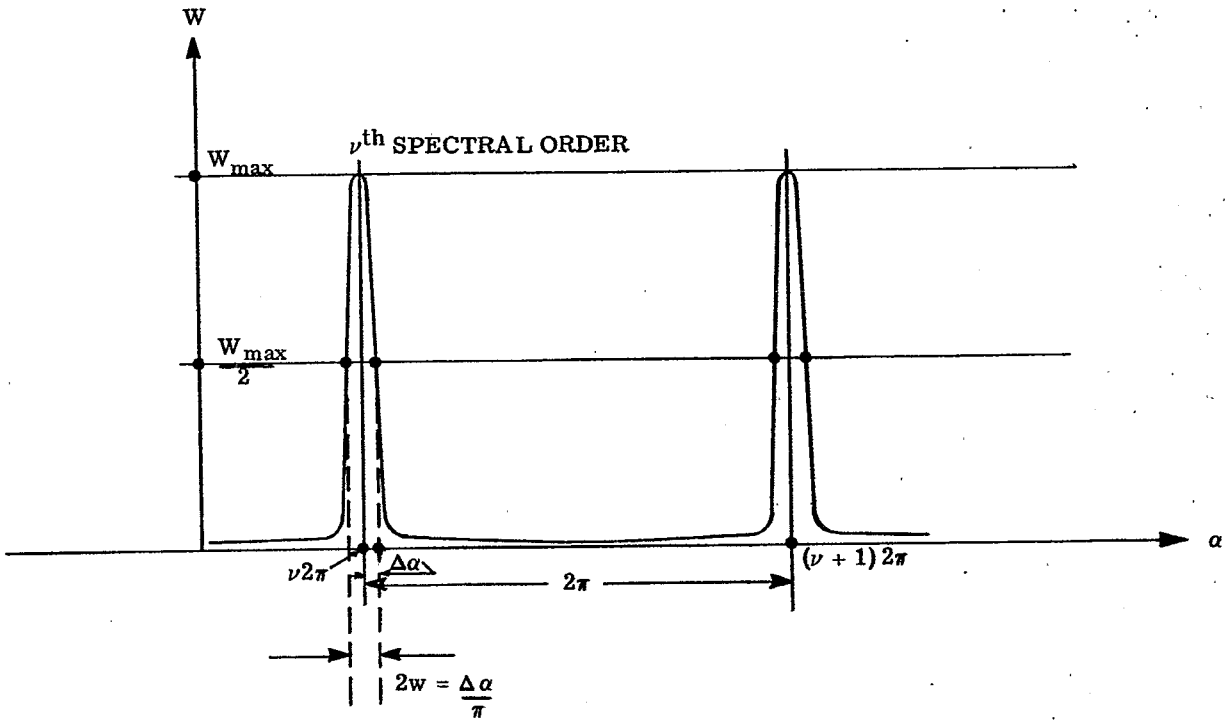

FIGURE 16.19- The Sharpness Quality of Multiple Beam Interference Fringes.

of the plate.

16.15.1.12 The limiting sharpness of the multiple beam fringes depends ultimately upon freedom from absorption. As a high reflecting coating, silver has remarkably low absorption. It is not difficult to obtain evaporated films of silver that have absorptions less than 5 per cent even when the film is practically opaque. Whereas much lower absorptions are possible with silver, the use of high reflecting, multi-layered films is becoming more common when the narrowest half-widths are required.

16.15.1.13 Two methods for viewing the multiple beam interference fringes that are transmitted by a plate are illustrated in Figures 16.20 and 16.21. Sharp, circular fringes will be seen provided that the surfaces of the plate are sufficiently parallel and silvered. Since n_1, d, and λ are fixed, it follows at once from Equation (101) that the sharp bright fringes are fringes of equal inclination, i.e., the angle of refraction i_1' is constant along each fringe. When the thickness d of the plate or film is large, the number of circular fringes becomes so great that the determination of their spectral order ν is difficult.

16.16 REFLECTANCE FROM PLANE PARALLEL PLATES

16.16.1 Introduction.

16.16.1.1 The dark fringes usually appear sharp in the reflected family. However, it is not necessarily true that a dark fringe must appear in the reflected family of fringes at values of α for which a bright fringe occurs in the transmitted family.

16.16.1.2 With respect to Figure 16.18, let $r_{0,1}$ and $\delta_{0,1}$ denote amplitude reflectance and phase change on reflection for a beam incident from the 0^{th} medium. Then,

$$
\begin{aligned}
R_0 &= r_{0,1}\, e^{i\delta_{0,1}} \\
R_1 &= \tau_{0,1}\, \tau_{1,0}\, \tau_1^2\, r_{1,2}\, e^{i(\alpha - \delta_{1,0})} \\
R_2 &= \tau_{0,1}\, \tau_{1,0}\, \tau_1^4\, r_{1,2}^2\, r_{1,0}\, e^{i(2\alpha - \delta_{1,0})} \\
R_3 &= \tau_{0,1}\, \tau_{1,0}\, \tau_1^6\, r_{1,2}^3\, r_{1,0}^2 \cdot e^{i(3\alpha - \delta_{1,0})}, \text{ etc.}
\end{aligned} \tag{114}
$$

Therefore,

$$
R = \sum_{\nu=0}^{N} R^{\nu} = r_{0,1}\, e^{i\delta_{0,1}} + C\, e^{(\alpha - \delta_{1,0})} \sum_{\nu=0}^{N-1} A^{\nu}\, e^{i\nu\alpha} \tag{115}
$$

in which α and A are defined by Equations (101) and (103), R is a complex number that determines the amplitude and phase of the reflected beam and

$$
C \equiv \tau_{0,1}\, \tau_{1,0}\, \tau_1^2\, r_{1,2}. \tag{116}
$$

Comparison of Equations (102) and (115) shows that Equation (115) contains the additional term $r_{0,1}\, e^{i\delta_{0,1}}$ due to the first reflection R_o of Figure 16.18. It is the presence of this extra term that complicates the nature and the interpretation of the reflected fringes.

16.16.1.3 Suppose that α has any one of the values $\nu 2\pi$ of Equation (108), the condition for bright fringes in the transmitted beam. Then from Equation (115)

$$
\begin{aligned}
R &= r_{0,1}\, e^{i\delta_{0,1}} + C\, e^{-i\delta_{1,0}} \sum_{\nu=0}^{N-1} A^{\nu} \\
&= r_{0,1}\, e^{i\delta_{0,1}} \left[1 + C\, e^{-i(\delta_{1,0} + \delta_{0,1})} \sum_{\nu=0}^{N-1} A^{\nu} \right]
\end{aligned} \tag{117}
$$

We see that $|R|$ will be minimum when $\alpha = \nu\, 2\pi$, provided that with respect to the phase changes $\delta_{0,1}$ and $\delta_{1,0}$ on reflection at the first surface of the plate

$$
\delta_{1,0} + \delta_{0,1} = \mu\pi \tag{118}
$$

where μ is an odd integer. In other words, dark reflected fringes will occur at the same α - values as bright transmitted fringes, provided that the sum of the phase changes on reflection for incidence from opposite directions upon the first surface, Figure 16.18, is an odd number of half-wavelengths. This condition is rarely fulfilled. Consequently, one has to expect that the reflected fringes will be darkest at α - values that differ suitably from $\alpha = \nu\, 2\pi$ where ν is an integer. However, this complication does not detract from the utility of

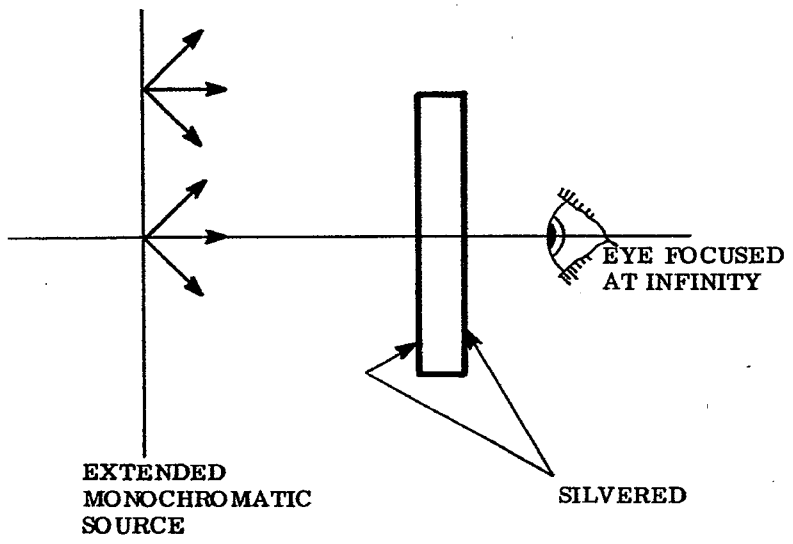

FIGURE 16. 20-Simple Parallel Plate Interferometer.

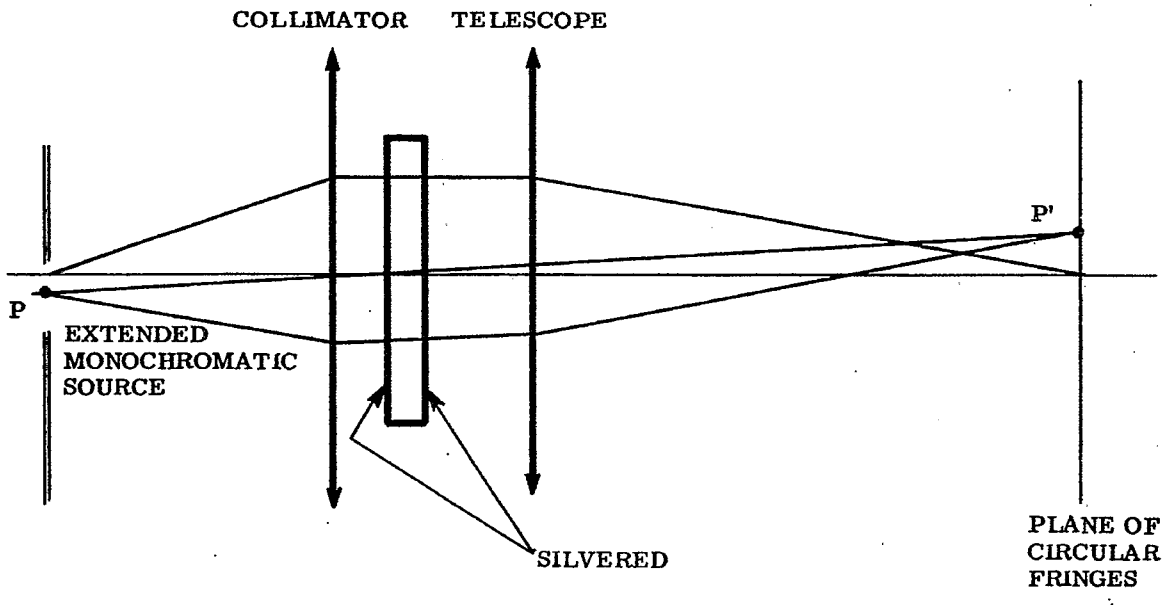

FIGURE 16. 21-The Fabry-Perot Interferometer.

the reflected fringes, except in those cases in which it leads to fringes that are only slightly darker than the background.

16.16.1.4 Sharp reflected fringes can be observed, for example, by replacing the elements bearing surfaces S_1 and S_2 of Figure 16. 11 by a plane parallel plate whose major surfaces are suitably silvered, aluminized, etc. The eye is preferably focused for infinity.

16.17 MULTIPLE BEAM INTERFERENCE FRINGES FROM SLIGHTLY INCLINFD SURFACES

16.17.1 General.

16.17.1.1 Let a wavefront V be incident upon the wedge formed between two reflecting surfaces that have the small included angle α as illustrated in Figure 16. 22. Wavefronts V_0, V_1, V_2, etc., inclined at the angles 0, 2α, 4α, etc., will emerge from the wedge after an appropriate number of inter-reflections within the wedge. The corresponding emergent rays are indicated by T_0, T_1, T_2, etc. A series of coherent, plane waves are formed in this manner by inter-reflections within the wedge.

16.17.1.2 Let

$t_1 \equiv$ amplitude transmittance of surface S_1;
$t_2 \equiv$ amplitude transmittance of surface S_2;
$r_1 \equiv$ amplitude reflectance of surface S_1;
$r_2 \equiv$ amplitude reflectance of surface S_2;
$\delta_1 \equiv$ phase change on reflection at surface S_1;
$\delta_2 \equiv$ phase change on reflection at surface S_2.

16.17.1.3 We choose the X-axis along OP and the Z-axis parallel to PT_0 and suppose that the amplitude of the incident wavefront is unity. We note that such phase changes as may occur upon transmission through surfaces S_1 and S_2 can be ignored since they alter all of the emergent waves equally. The space between S_1 and S_2 is assumed to be nonabsorbing.

16.17.1.4 The emergent wave propagated along PT_0, i.e., along Z, has the complex form

$$T_o = t_1 t_2 e^{iknz} e^{-i\omega t}$$

The wave emergent along PT_1 has the form

$$T_1 = t_1 t_2 r_1 r_2 e^{i(\delta_1+\delta_2)} e^{ikn[x \sin 2\alpha + z \cos 2\alpha]} e^{-i\omega t}$$

Similarly,

$$T_2 = t_1 t_2 (r_1 r_2)^2 e^{i2(\delta_1+\delta_2)} e^{ikn[x \sin 4\alpha + z \cos 4\alpha]} e^{-i\omega t};$$

$$T_3 = t_1 t_2 (r_1 r_2)^3 e^{i3(\delta_1+\delta_2)} e^{ikn[x \sin 6\alpha + z \cos 6\alpha]} e^{-i\omega t};$$

etc.,

16.17.1.5 Introduce

$$R \equiv r_1 r_2;$$

$$\tau \equiv t_1 t_2;$$

$$\phi \equiv \delta_1 + \delta_2. \qquad (119)$$

Then

$$T = \sum_{\nu=0}^{N} T^{\nu} = \tau e^{-i\omega t} \sum_{\nu=0}^{N} R^{\nu} e^{i\nu\phi} e^{ikn[x \sin(2\nu\alpha) + z \cos(2\nu\alpha)]} \qquad (120)$$

where T specifies the amplitude and phase determined by the interference of the emergent waves $T_0, T_1, \ldots, T_N$. The fringes described by Equation (120) are of a type far more general than those ordinarily used. We obtain the conventional type multiple beam fringes formed by a wedge by supposing that the angle α of the wedge is so small that

$$\sin(2\nu\alpha) \rightarrow 2\nu\alpha; \quad 0 \leqq \nu \leqq N. \qquad (121)$$

$$\cos(2\nu\alpha) \rightarrow 1; \quad 0 \leqq \nu \leqq N.$$

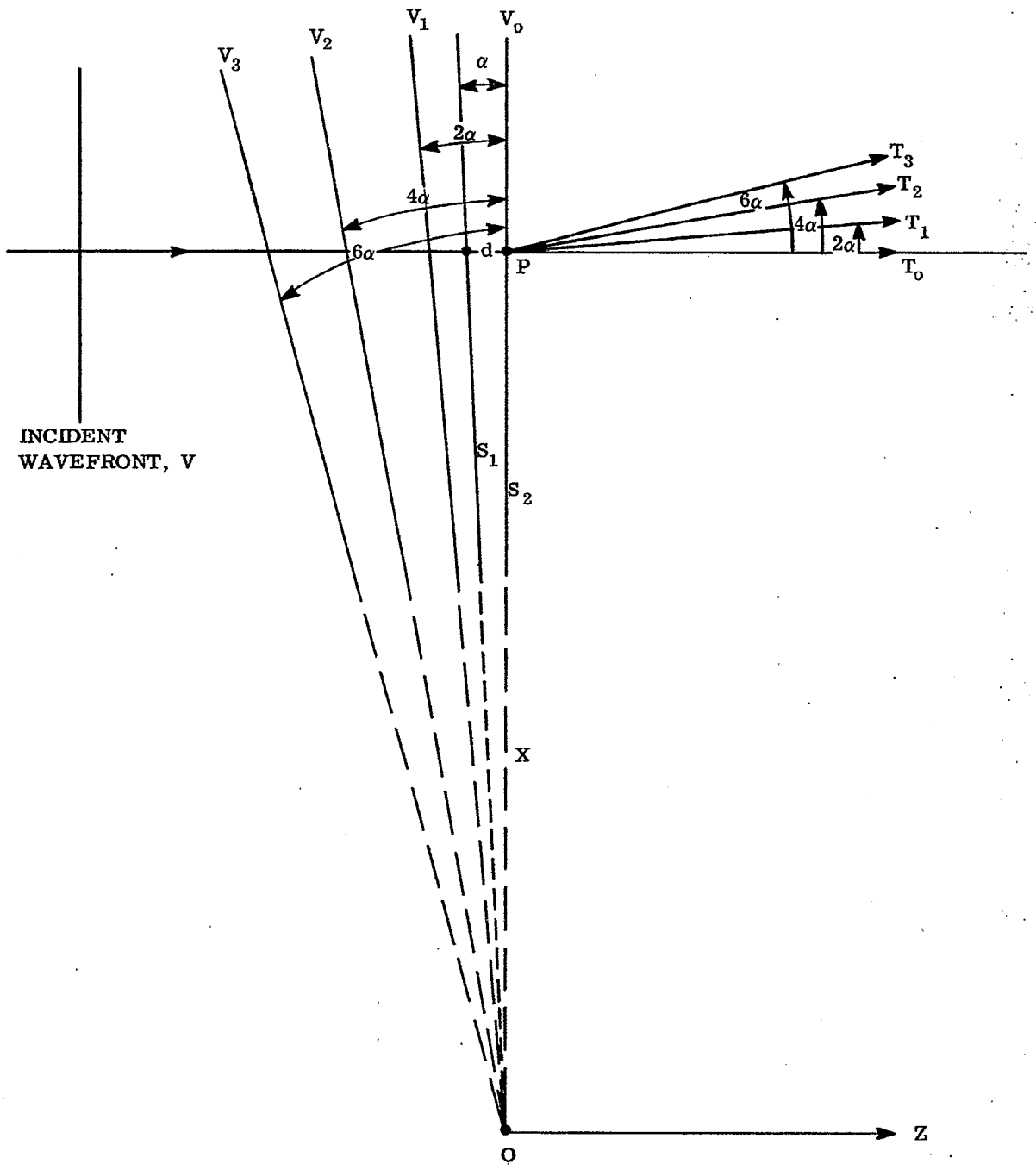

FIGURE 16. 22- Multiple Reflections in two reflecting surfaces S_1 and S_2 .

From Equations (120) and (121)

$$T = \tau\ e^{-i\omega t}\ e^{iknz} \sum_{\nu=0}^{N} R^{\nu}\ e^{i\nu\ (\phi + 2knx\alpha)} \tag{122}$$

Applying Equation (104) to Equation (122), we obtain

$$T = \tau\ e^{-i\omega t}\ e^{iknz}\ \frac{1 - R^{N+1}\ e^{i(N+1)(\phi + 2knx\alpha)}}{1 - R\ e^{i(\phi + 2knx\alpha)}} \tag{123}$$

16.17.1.6 The time-averaged energy density W_T in the fringes seen on transmission is given by

$$2W_T = |T|^2 = \tau^2\ \frac{1 - 2R^{N+1}\cos\left[(N+1)(\phi + 2knx\alpha)\right] + R^{2(N+1)}}{1 - 2R\cos(\phi + 2knx\alpha) + R^2} \tag{124}$$

in which

$$\tau \equiv t_1 t_2\ ;\quad R \equiv r_1 r_2\ ;\quad \phi \equiv \delta_1 + \delta_2\ ; \tag{119}$$

$k \equiv 2\pi/\lambda$; and n is the refractive index of the medium within the wedge. ϕ is the sum of the phase changes on reflection at the surfaces S_1 and S_2 of the wedge. α is the angle of the wedge. The result of Equation (124) is independent of z (which suggests most strongly that the fringes are not necessarily localized within the wedge). However, it should be remembered that the requirement of Equation (121) is unlikely to be met in actual practice when the included number of inter-reflections N is high. Dependence of the fringe system upon the plane z of observation must be expected from Equation (120) when one is not entitled to set $\cos(2\nu\theta) = 1$.

16.17.1.7 A common method for obtaining and viewing transmitted multiple beam fringes in a wedge is illustrated in Figure 16. 23. The rays PT_o, PT_1, PT_2, etc., of Figure 16. 22 form images H_o, H_1, H_2, H_3, etc., of the pinhole H at the second focal plane of the objective. The number N of inter-reflections is frequently restricted by the diaphragm stop D of the objective, i.e., by the numerical aperture of the objective. In Figure 16. 23 rays from the zero order ($\nu = 0$) pass through H_o; rays from the 1st order ($\nu = 1$) pass through

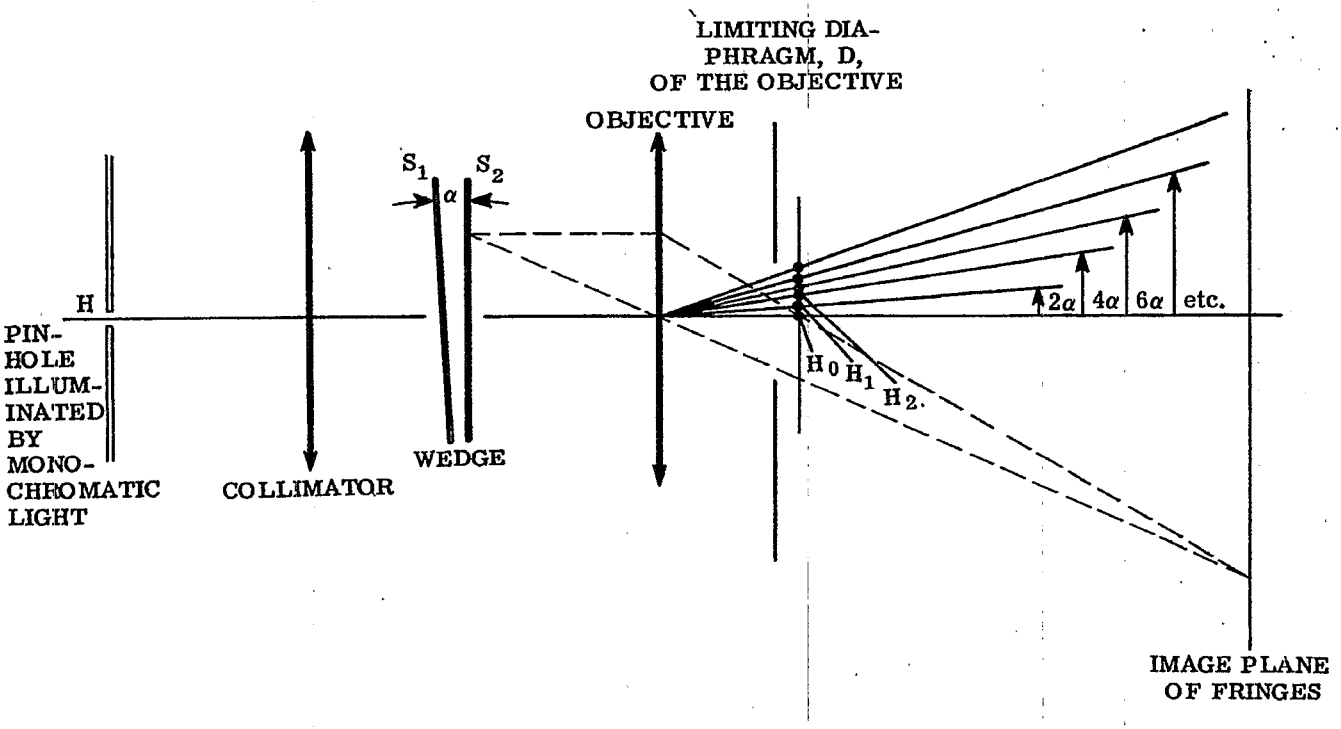

FIGURE 16. 23- Method of producing and viewing Transmitted, Multiple Beam Fringes in a wedge.

H_1, etc. Rays belonging to the seventh order ($\nu = 7$) are interrupted by the diaphragm. Thus, with respect to Equation (124), one would have $\nu_{max} = N = 6$. Apart from restricting the possible number of spectral orders N that get to the image plane, the objective may be regarded as a means of observing the object plane which is usually selected at, or within, the wedge where z approaches zero. The pinhole images H_ν, that correspond to the spectral orders ν, are easily seen by viewing the back of the objective, provided the system is in proper adjustment and that α is neither too small nor too large. The image plane is frequently viewed with an eyepiece. A microscope forms an excellent means for viewing the transmitted fringes. One has only to replace the conventional substage condensor by a more suitable lens to act as collimator. The selected pinhole H should be small enough so that it does not reduce the sharpness of the multiple beam fringes as determined experimentally.

16.17.1.8 We now return to complete our interpretation of Equation (124). Comparison of Equations (106) and (107) for multiple beam fringes with plane parallel plates shows that they are very similar to Equation (124). Most of the conclusions drawn in paragraph 16.15 apply again with minor modifications or qualifications. For example, we may conclude at once that bright fringes will occur when

$$\phi + \frac{4\pi}{\lambda} mx\alpha = \nu\, 2\pi . \tag{125}$$

It will be seen from Figure 16.22 that

$$x\alpha = d \tag{126}$$

where d is the thickness of the wedge at the point P under observation. Hence, we may rewrite Equation (125) in the well known form

$$2nd = \nu\lambda - \frac{\lambda\phi}{2\pi} \tag{127}$$

in which ϕ, expressed in radians, is the sum of the phase changes on reflection at the surfaces S_1 and S_2 of the wedge. ϕ is in general a function of the wavelength. Again we observe that each fringe is the locus of points x for which the optical path nd is constant. The fringe width $|\Delta x| = h$ must be, according to Equation (125), that value of $|\Delta x|$ for which $4\pi n\alpha\, |\Delta x| /\lambda = 2\pi$. Therefore, the fringe width h is given by

$$h = |\Delta x| = \frac{\lambda}{2n\alpha} . \tag{128}$$

Comparison of Equation (128) with Equation (12) shows that when the refractive indices n of the space between the reflecting surfaces are alike, the fringe widths are the same, whether one is using a Fizeau type interferometer or the multiple beam interferometer.

16.17.1.9 With respect to Figure 16.22, reflected plane waves emerge from the wedge and are propagated along the negative Z-direction. Corresponding to Equation (120), a series R for the reflected fringes is obtained. As in Equation (115) for parallel plates (case $\alpha = 0$), the series for R is complicated by the term R_o that corresponds to direct reflection from the first surface of the wedge. In general, the remarks and conclusions of paragraph 16.16 also apply to the multiple beam fringes formed by reflection from a wedge for which $\alpha \neq 0$. The narrow reflected fringes are likely to be dark. A useful method for observing reflected multiple beam fringes is illustrated in Figure 16.24. The pinhole is placed at the first focal plane of the objective. The images H_o, H_1, . . . , H_N of the pinhole H formed by the light belonging to the spectral orders ν fall along a straight line. When an undue amount of parasitic light is present at the plane of H_o, H_1, . . . , H_N, contrast in the fringes can be improved markedly by inserting at this plane a diaphragm with a slit which is oriented to pass the spectral orders. It is possible also to block the spectral order $\nu = 0$ by blocking the light in the image H_o. When this is done, the reflected fringes have the appearance of the transmitted fringes — in fact, these narrow, bright, reflected fringes now obey Equation (125).

16.18 MEASUREMENTS WITH MONOCHROMATIC LIGHT

16.18.1 Introduction.

16.18.1.1 The effects of thin films upon the phase change introduced into a wave that traverses the optical system are being considered by some designers as an integral portion of the optical design of complex, high quality systems that contain many coated elements. The multiple beam interferometer is used frequently for measuring the thickness of thin films. The following principles belong to a method that has been applied to many different types of thickness measurements notably by S. Tolansky.

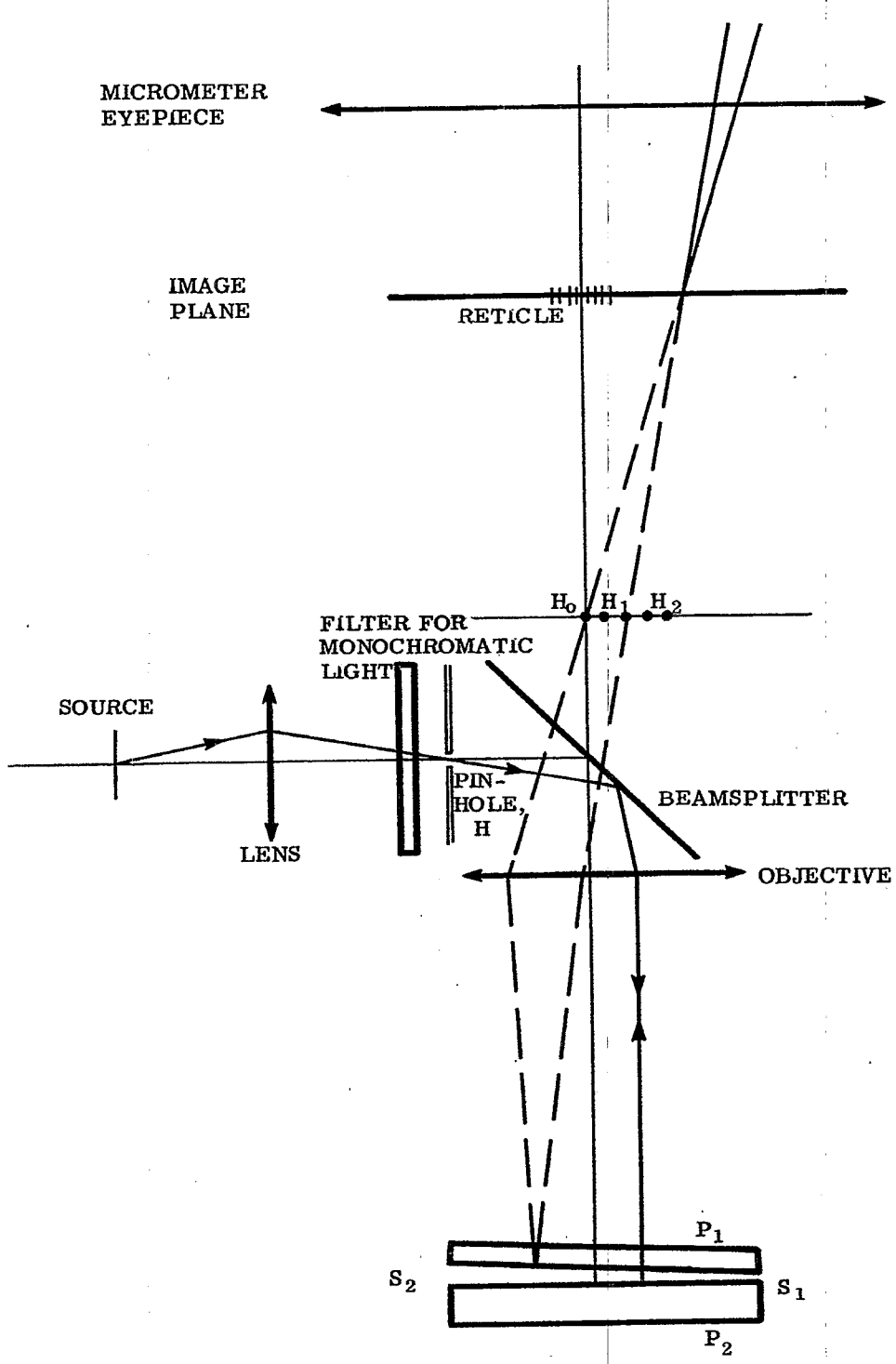

FIGURE 16. 24-Microscope for viewing reflected fringes under vertical illumination.

16.18.1.1 The preferred arrangement for measuring thicknesses of thin films utilizes multiple beam fringes that are formed by reflection as illustrated in Figure 16. 24. A micrometer eyepiece, containing any suitable reticule, is needed for measuring fringe widths and the fringe shifts that occur at the edge of a film that has been deposited upon surface S_2 and covered with a uniform coating of, say, silver as illustrated in Figure 16. 25. Evaporated coatings of silver and other metals produce a sharp step, whose height is equal to that of the film. The evaporated overcoating must be sufficiently opaque so that the phase changes on reflection at S_2 are not changed by the presence of the substrate or the film. The optically flat surface S_1 must be placed in close contact with surface S_2 in order to obtain reliable measurements of the thickness of the film. The usual practice is to lay plate P_1 directly upon plate P_2, Figure 16. 24, after making certain that no large dust particles are present to increase the separation between the silvered surfaces. It is good practice to make the fringes approximately perpendicular to the edge AB as in Figure 16. 26.

16.18.1.2 Let t denote the thickness of the film. We shall now show that

$$t = \frac{1}{2} \frac{\Delta x}{n h} \lambda \qquad (129)$$

where Δx and h are respectively, the fringe shift and fringe width determined with the aid of the micrometer eyepiece (see Figure 16. 26). It is presumed that t is so small that the fringe shift is less than one fringe width. (This method is not well suited to measure thicknesses for which the fringe shifts Δx exceed the fringe width.) We have seen that a fringe is the locus of points Δx for which the separation d of the reflecting surfaces is constant. If then, a fringe is located at the point x in the absence of the film, it will move to a point $x + \Delta x$ on the film so as to keep d constant in the manner illustrated in Figure 16. 27. Since the angle α between S_1 and S_2 is to be small,

$$t = \alpha \, \Delta x . \qquad (130)$$

But from Equation (128), $\alpha = \lambda/2 \, n \, h$. Substitution of this value of α into Equation (130) gives Equation (129) directly. The wedge between surfaces S_1 and S_2 is ordinarily air so that n = 1. This simple argument leading to Equation (129) applies to both the reflected and the transmitted fringes.

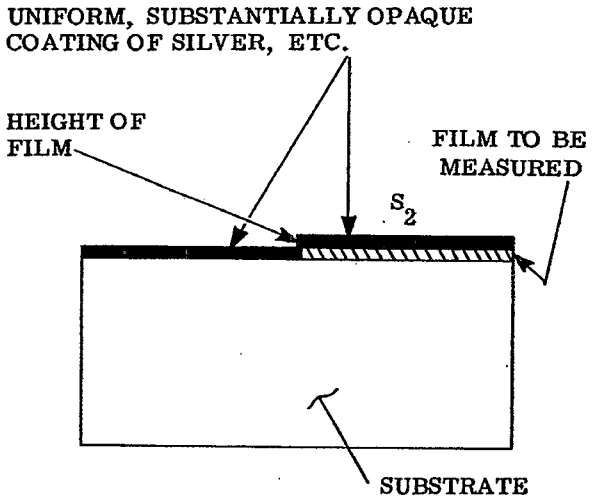

Figure 16. 25 - The usual method of preparing the sample film for thickness measurement in the Multiple Beam Interferometer.

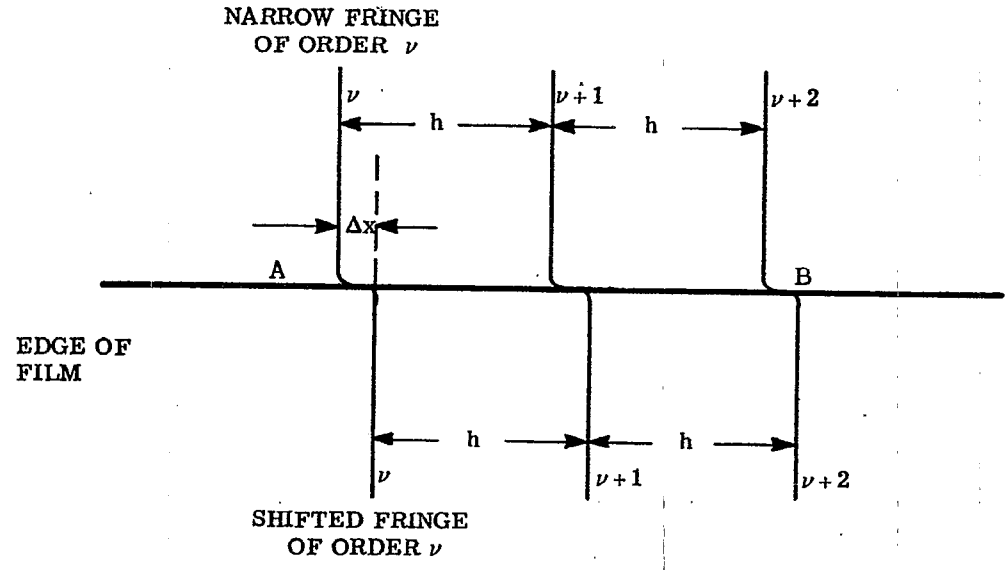

FIGURE 16. 26- Appearance of the narrow fringes when the thickness of the film is a small fraction of a wavelength and the film occupies the portion below the edge AB. If the surfaces S_1 and S_2 are optically flat, the indicated fringe widths, h, will be alike within the expiramental error of measuring h.

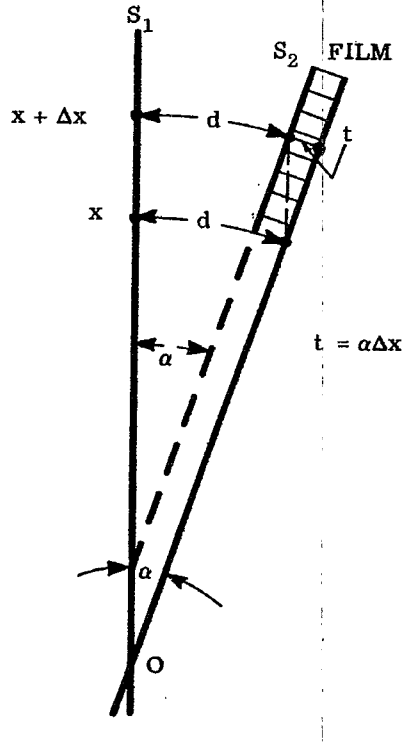

FIGURE 16. 27- Movement of an interference frings, x, to the position, x + Δ x, by the introduction of a film at thickness, t.

16.18.1.3 One soon finds that the attainable precision is restricted by the roughness of the polished glass surfaces that ordinarily serve as the reflectors. These surfaces present, so to speak, a mountainous terrain whose peaks and valleys range between 10 and 60 Angstroms in height and depth. Correspondingly, the sharp fringes will not remain straight under increasing magnification but become so wiggly that one has difficulty in estimating their "center of gravity." These wiggly fringes are valuable for comparing different methods of polishing and molding the surfaces of optical elements. The method is so sensitive that the height of a molecule of mica has been determined with an accuracy that compares favorably with the result obtained from x-rays.

16.19 THE METHOD OF CHANNELED SPECTRA

16.19.1 General.

16.19.1.1 The conventional method for observing channeled spectra (also called the FECO bands, i.e., fringes of equal chromatic order) is illustrated in Figure 16.28 for the case in which the FECO bands are formed by transmission at the interferometer. Collimated white light passes through the interferometer. The image of the wedge is focused upon the entrance slit of a wavelength monochromator. One may view or photograph the FECO bands that appear at the exit pupil of the wavelength monochromator. If the surfaces S_1 and S_2 of the interferometer are plane, the interference bands seen at the eyepiece will be straight, of different wavelength, and of consecutive spectral order ν as indicated. In this method, the surfaces S_1 and S_2 are preferably parallel. Since it is an accidental matter to achieve parallelism by pressing surface S_1 against surface S_2, the practical compromise is to alter the relative inclination of surfaces S_1 and S_2 to the point at which the interference bands formed at the eyepiece of the wavelength monochromator are parallel to the image of the entrance slit.

16.19.1.2 The arrangement illustrated in Figure 16.26 allows white light to pass through the interferometer plates. Consequently, a relatively large amount of light flux is available to disturb the thermal equilibrium of the interferometer plates. The observed wavelengths of the interference bands can drift for hours before reliable readings can be taken. A more satisfactory arrangement that minimizes drifts due to thermal causes has been described by H. Osterberg and D. LaMarre.* Their arrangement, as applied to obtaining multiple beam fringes by reflection, is illustrated in Figure 16.29. Monochromatic light of measured, variable wavelength illuminates the interferometer. The interference fringes seen at the eyepiece of the microscope are of the same wavelength for a given setting of the wavelength drum and differ consecutively, as indicated, in order number. Indeed, the fringes resemble those of Figure 16.26 and could be measured as discussed in paragraph 16.18 with the aid of an eyepiece micrometer for determining the thickness of a film. To do so would defeat several advantages of this arrangement. Instead, advantage is taken of the fact that the fringes move as the wavelength drum is turned. In this way, consecutive fringes from each side of the step can be brought into coincidence with a fixed pointer or marker on the reticule and the corresponding wavelength recorded. With this arrangement the surfaces S_1 and S_2 should not be parallel but should be preferably (although not necessarily) inclined so that the multiple beam fringes are approximately perpendicular to the image of the step that marks the edge of the film whose thickness is to be measured. This step is imaged sharply upon the plane of the reticule. Consequently, each wavelength determination is made across a definite, localized, and selected area at the edge of the film. This area is that portion of the surface S_2 which is projected upon the pointer at the plane of the reticule. It follows that slight or even marked departures of the test surfaces from flatness have secondary effects upon the accuracy of this method of channeled spectra. One looks for a spot at which the fringe runs quite straight across the edge of the film and makes his measurements here.

16.19.1.3 The main advantage of the method of channeled spectra over the direct method of multiple beam fringes discussed in paragraph 16.18 is that the flatness of the surfaces S_1 and S_2 is much less critical for the purpose of making thickness measurements. A second advantage consists of the fact that channeled spectra enable one to measure either thin or thick films without ambiguity relative to whether the fringe shift exceeds or does not exceed a suspected number of fringe widths.

16.20 INTERPRETATION OF MEASUREMENTS WITH CHANNELED SPECTRA

16.20.1 Introduction.

16.20.1.1 Examination of the theory of multiple beam interferometry stated in paragraphs 16.15 through 16.17 shows that whether one is dealing with fringes obtained in either reflection or transmission from parallel plates or from wedges, the analytic condition for the appearance of the sharp fringes is of the form

$$\nu \lambda = 2d + \lambda f \; ; \quad n = 1 \; ; \tag{131}$$

where ν is an integer, d is the separation of the interferometer surfaces, λ is the wavelength, and f is a function related to the phase changes that take place on reflection at the coated surfaces of the reflecting surfaces. The function f can vary with wavelength and will be different for the transmitted and reflected fringes.

*H. Osterberg and D. LaMarre, J. Opt. Soc. Amer., 46, 777-778 (1956).

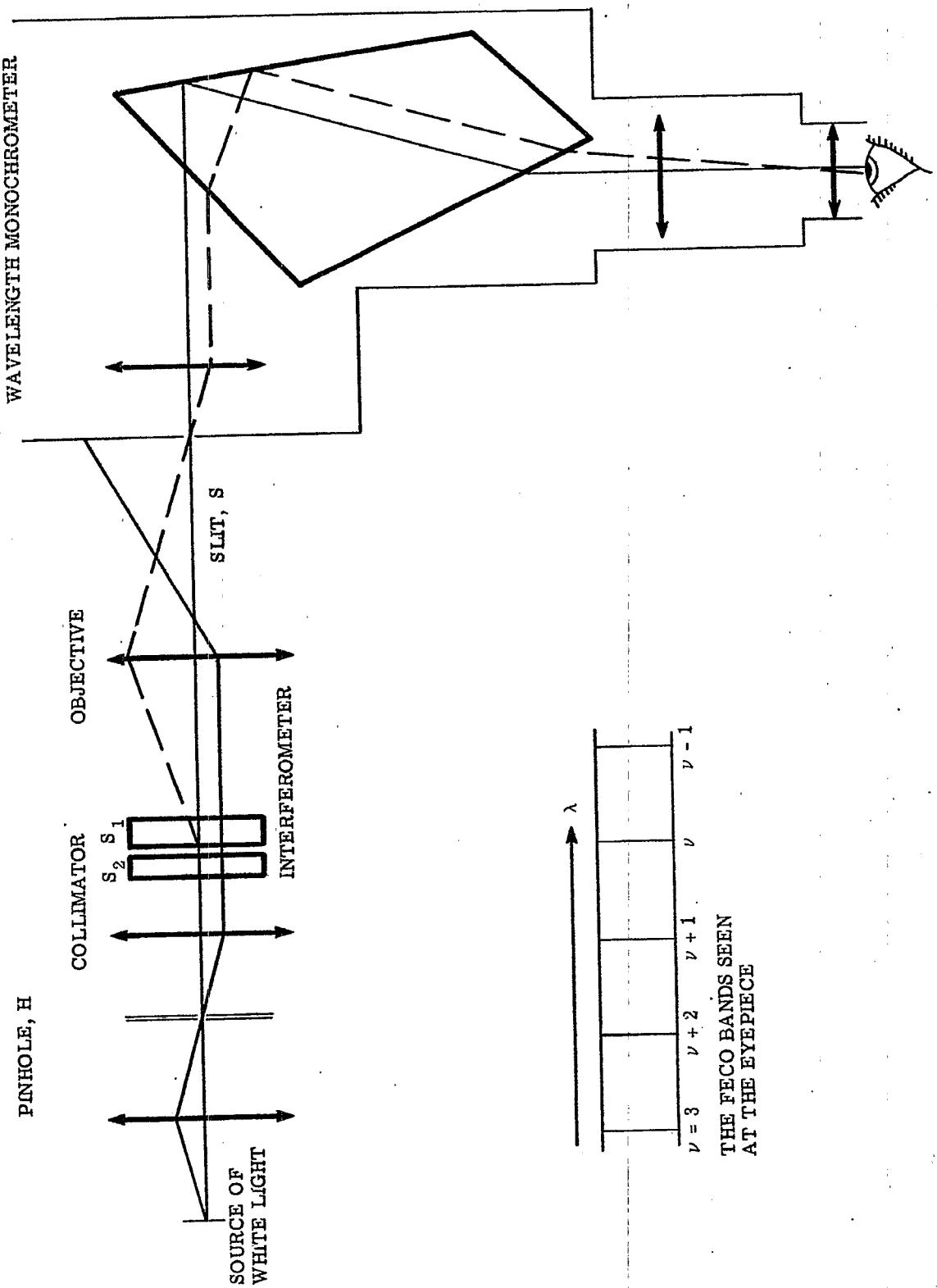

FIGURE 16. 28- Conventional method for obtaining channeled spectra from a source of white light.

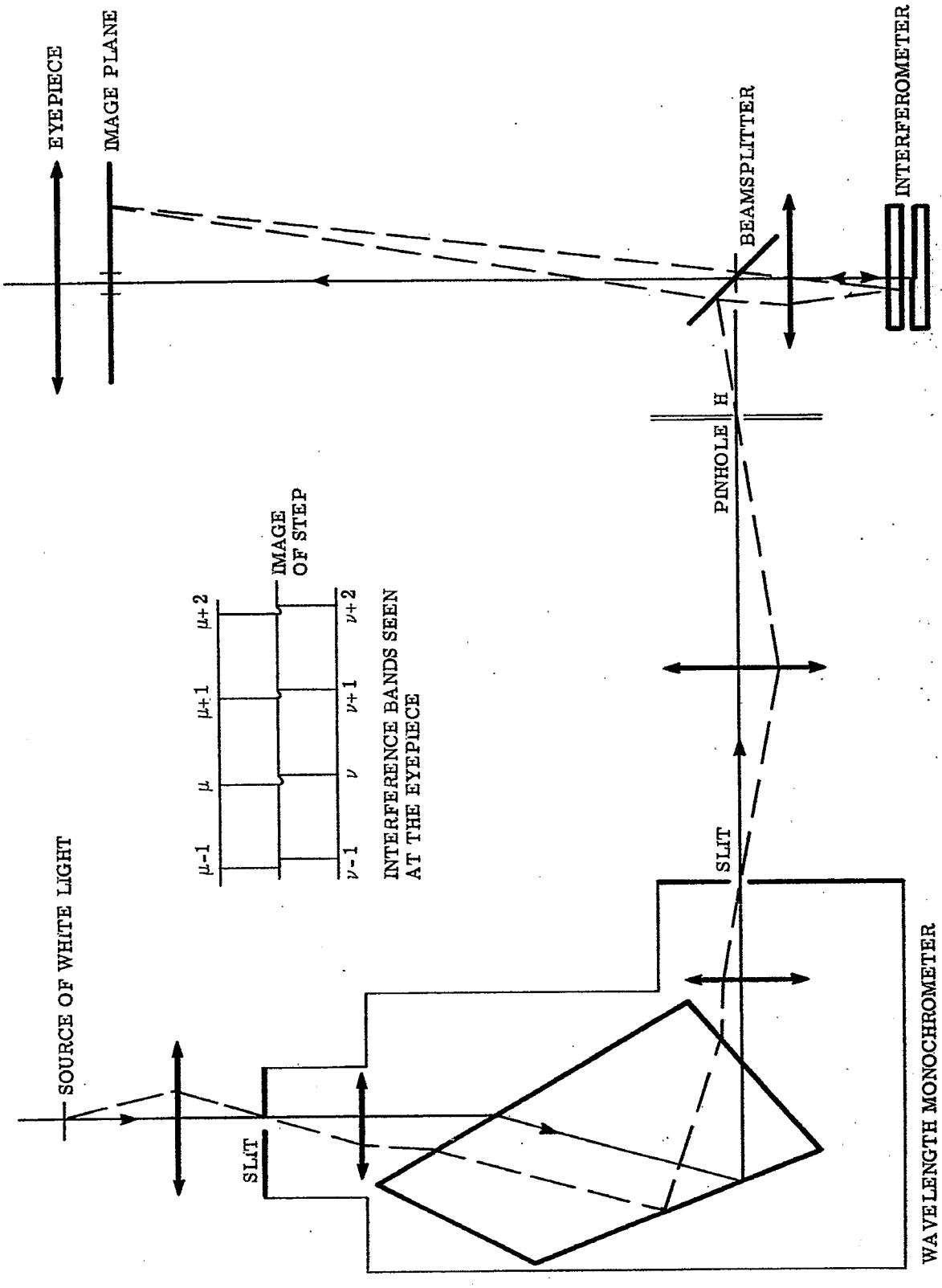

FIGURE 16. 29-A second method for producing channeled spectra.

16.20.1.2 With regard to the transmitted fringes, it has been customary to take $f = 0$ and to state that each bright fringe occurs at those wavelengths $\lambda = \lambda_\nu$ for which $d = \nu(\lambda_\nu/2)$. Interpretations based upon this simplified view are, however, inadequate.

16.20.1.3 Let f be expanded as a function of wavelength about the wavelength λ_o such that

$$f = f_o + b(\lambda - \lambda_o) + c(\lambda - \lambda_o)^2 + \ldots . \tag{132}$$

Experience has shown that with silver coatings or with high reflecting multilayers, there will exist an extended range for which $\nu\lambda$ is, with good approximation, a linear function of λ about an appropriately chosen λ_o in the visible region. For this range of wavelengths the first two terms of Taylor's expansion of $\nu\lambda$ about the point $\lambda = \lambda_o$ from Equations (131) and (132) yield the approximation

$$\nu\lambda = 2d - b\lambda_o^2 + s_o\lambda \tag{133}$$

in which

$$s_o = f_o + b\lambda_o . \tag{134}$$

16.20.1.4 With respect to Figure 16.28, consider two nearby spectral orders ν and $\nu + p$ where $p = 0, \pm 1, \pm 2, \pm 3$, etc. Let λ_ν and $\lambda_{\nu+p}$ be the central wavelengths of these spectral bands. We have seen that channeled spectra are obtained from a single localized area for which the separation d of the interferometer mirrors is constant. Since $b\lambda_o^2$ is constant, it follows from Equation (133) that

$$\lambda_\nu(\nu - s_o) = 2d - b\lambda_o^2 = \text{constant} = \lambda_{\nu+p}(\nu + p - s_o). \tag{135}$$

Hence,

$$\nu - s_o = p\,\frac{\lambda_{\nu+p}}{\lambda_\nu - \lambda_{\nu+p}} . \tag{136}$$

As will be seen from Figure 16.28, determining p is simply a matter of counting bands from the band whose order number is labelled ν. Since p, λ_ν, and $\lambda_{\nu+p}$ are known, one can compute $\nu - s_o$ from Equation (136). It is good practice to compute $\nu - s_o$ for at least three values of $|p|$ when enough bands are available. If the values $\nu - s_o$ thus obtained are not alike within a range corresponding to one's experimental error in reading the wavelengths, the separation d of the interferometer mirrors is changing or $|p|$ has been chosen so large that $\lambda_{\nu+p}$ falls outside of the range for which $\nu\lambda$ is adequately linear in λ.

16.20.1.5 The values $\nu - s_o$ will fall in the range 10 to 70 when the interferometer mirrors are laid one upon the other except when great precaution is taken to avoid dust particles. The corresponding separation of the interferometer mirrors falls in the range 5 to 35 wavelengths. This explains why the separation can vary with temperature, etc. When $\nu - s_o$ has been determined, the separation d is given by

$$2d = (\nu - s_o)\lambda_\nu + b\lambda_o^2 . \tag{137}$$

Unfortunately, one needs to know $b\lambda_o^2$ in order to compute d accurately. One may, of course, accept $2d = (\nu - s_o)\lambda_\nu$ as his approximation and expect that this will be a better approximation than obtained by asserting that $2d = \nu\lambda_\nu$.

16.20.1.6 On the other hand, a knowledge of $b\lambda_o^2$ is not required in order to determine accurately the thickness t of a film. With respect to the interference bands $\nu + p$ and $\mu + p$ seen on each side of the step formed at the edge of the film (see Figure 16.29), one determines $\nu - s_o$ and $\mu - s_o$ from the wavelengths λ_ν, $\lambda_{\nu+p}$ and λ_μ, $\lambda_{\mu+p}$ for which the interference bands are brought into coincidence with the marker on the reticule by turning the wavelength drum. The "non-integral spectral orders" $\nu - s_o$ and $\mu - s_o$ become known on each side of the step at the film. Then, from Equation (137)

$$2t = 2(d_1 - d_2) = (\nu - s_o)\lambda_\nu - (\mu - s_o)\lambda_\mu . \tag{138}$$

If the film is thin enough, one finds automatically that $\nu - s_o = \mu - s_o$ or that $\nu = \mu$. In such cases Equation (138) reduces to

$$t = \frac{1}{2}(\nu - s_o)(\lambda_\nu - \lambda) . \tag{139}$$

16.20.1.7 Let us consider the sensitivity and accuracy of the method of channeled spectra in, for example, the measurement of the thickness t of the thin films to which Equation (139) applies. If the error in reading the wavelengths λ_ν and λ_μ is $\delta\lambda$ and if δt is the corresponding error in t, then for estimating δt, we observe from Equation (139) that

$$\delta t \leqq \left(\frac{\nu - s_o}{2}\right) 2\ |\delta\lambda| \leqq (\nu - s_o)\ |\delta\lambda| \ . \qquad (140)$$

It becomes clear that the error $|\delta t|$ is reduced by making measurements at low values of $\nu - s_o$, i.e., at low separations d of the interferometer mirrors. Reducing $\nu - s_o$ to values in the neighborhood of 1 or 2 causes the spectral bands to broaden and to become excessively wiggly when polished surfaces are employed. The added difficulty of setting upon the center of gravity of the interference bands now appears. With the use of diffraction gratings, such as monochromators, and of photographic methods involving microdensitometry, errors $|\delta t|$ of 0.1 Angstrom or less may become possible. To carry the method to such extremes is however costly, cumbersome, and tedious. A typical example of the actual error obtained by making routine visual settings with a prism monochromator has been cited by Osterberg and LaMarre. They found that the visual settings with a Hilger Barfit monochromator are reproducible to about one Angstrom. With $\nu - s_o = 35$, the corresponding maximum error δt in the thickness t of the film is 35 Angstroms. The actual computed values of t from a series of spectral orders $\nu + p$ and $\mu + p$ agree to about 10 Angstroms.

16.20.1.8 One should not form the impression that the method of channeled spectra is restricted to analysis of fringes produced by multiple beam interferometry. We have seen, for example, that order numbers ν are associated with Fizeau fringes as in Equation (61). By projecting Fizeau fringes formed in white light upon the slit of a wavelength monochromator as shown in Figure 16.29 or by adapting the modification illustrated in Figure 16.29, a series of bands will be seen at the eyepiece. Comparison of Equations (61) and (131) shows that one deals with the simpler case $f = 0$ in applying the method of channeled spectra to Fizeau fringes.

16.21 HUYGENS' PRINCIPLE

16.21.1 Introduction. Although Huygens' principle is less general than, for example, Kirchhoff's law, its applications are far simpler to follow and yield predictions that are in resonable close accord with experiment with respect to the phenomena that we shall consider.

16.21.1.1 Huygens' principle supposes that as a wave travels through a homogeneous, isotropic space, each point in the space is excited as the wave passes through it and serves as origin for a spherical wave that expands with the velocity of light in the medium. Requirements such as conservation of energy require that the amplitude of the spherical wave decrease as 1/r where the distance r is measured from the point of expansion. Furthermore, the principle supposes that the propagation of the wave itself through space is a consequence of the interference effects that take place between the infinite set of expanding spherical wavelets. Close examination of this interference process shows, for example, that the reconstructed wave thus obtained from an assumed plane wave travelling to the right is, in turn, a plane wave that travels to the right. The wave that tends to travel to the left is destroyed, in effect, by destructive interference. The development of a wavefront as the envelope of the spherical wavelets that expand from the original wavefront at $z = z_o$ is illustrated in Figure 16.30

16.21.1.2 The construction of Figure 16.31 enables one to deduce Snell's law of refraction from Huygen's principle. If t_o is the time required for light to travel from C to B in the 0^{th} medium,

$$CB = v_o\ t_o = \frac{ct_o}{n_o} \ .$$

The spherical wave starting from A travels the distance AD in time t_o such that

$$AD = v_1\ t_o = \frac{ct_o}{n_1} \ .$$

But

$$\sin i = \frac{CB}{AB} \ ; \quad \sin i' = \frac{AD}{AB} \ .$$

Hence,

$$\frac{\sin i}{\sin i'} = \frac{CB}{AD} = \frac{n_1}{n_o}$$

This demonstration shows that the most basic law of geometrical optics can be explained by diffraction.

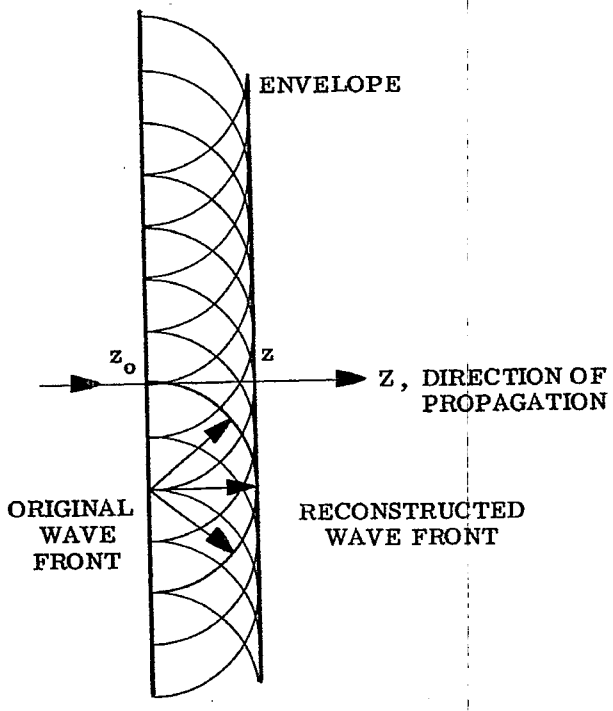

FIGURE 16. 30- Propagation of a plane wave in accordance with Huygens' Principle.

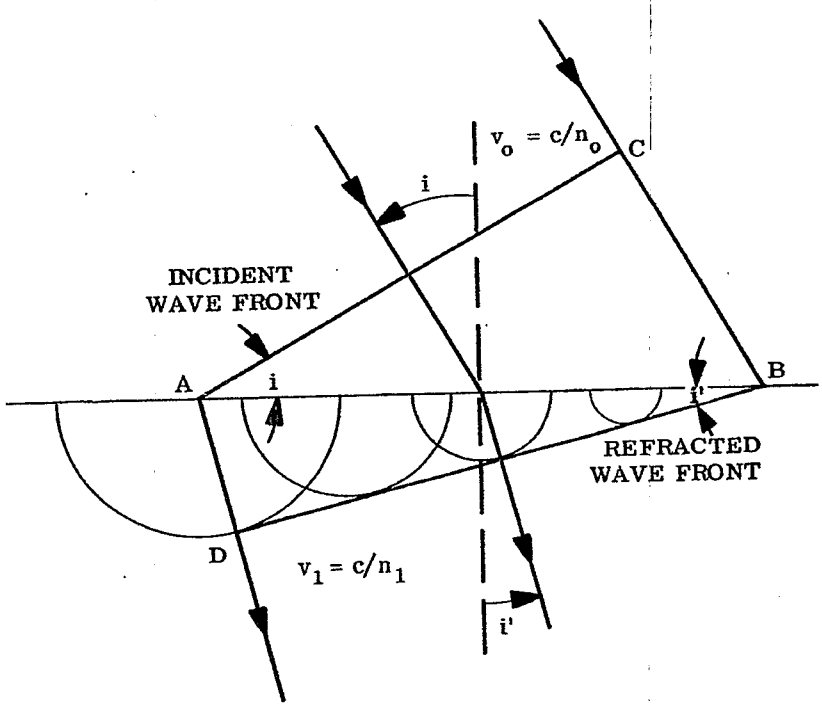

FIGURE 16. 31- Construction for obtaining Snell's Law of Refraction from Huygens' Principle.

16.21.1.3 We shall need an analytical statement of Huygens' principle. The amplitude and phase variation of the electric vector of a spherical wave that expands from any point O in a space whose refractive index is n is given by

$$E = \frac{e}{r}^{i\,(knr\,-\omega t)} \qquad (141)$$

where $k \equiv 2\pi/\lambda$; $\omega \equiv 2\pi/T$; and r is distance measured from point O. The physical meaning of Equation (141) is, of course, in doubt at the point r = 0 - but not elsewhere.

16.22 FRAUNHOFER DIFFRACTION

16.22.1 Discussion of theory.

16.22.1.1 Fortunately, the theory and interpretation of diffraction phenomena become much simpler when these phenomena are considered at relatively large distances from the diffracting aperture or obstacle. When a lens is placed between the aperture and the plane at infinity, the diffraction phenomena at infinity are brought into the focal plane of the lens. This consideration leads one to suspect that diffraction phenomena that occur at the focal plane of lenses are likely to be Fraunhofer diffraction phenomena. Since diffraction effects associated with focal planes belong to the classification known as Fraunhofer diffraction phenomena, these diffraction phenomena are of primary fundamental interest to the designer of optical (or radar) instruments.

16.22.1.2 Simplified arguments based upon Huygens' construction can be used to locate maxima and minima in the energy densities associated with Fraunhofer diffraction effects, but such arguments do not predict the distribution of energy density. The following diffraction integrals become so simple and direct that we shall omit the elementary and less instructive theory. The diffraction integral governing Fraunhofer diffraction is easily integrated or applied to a large number of practical cases.

16.22.1.3 We suppose that the aperture or obstacle from which diffraction occurs is located at the $\zeta\eta$ plane of Figure 16.32 and that the observation plane xy is located at distance D from the $\zeta\eta$ plane. Huygens wavelets leave each element of area $d\zeta\, d\eta$ of the $\zeta\eta$ plane and arrive at point P of the plane of observation after traversing the distance r where

$$r = \left[(x - \zeta)^2 + (y - \eta)^2 + D^2 \right]^{1/2} \qquad (142)$$

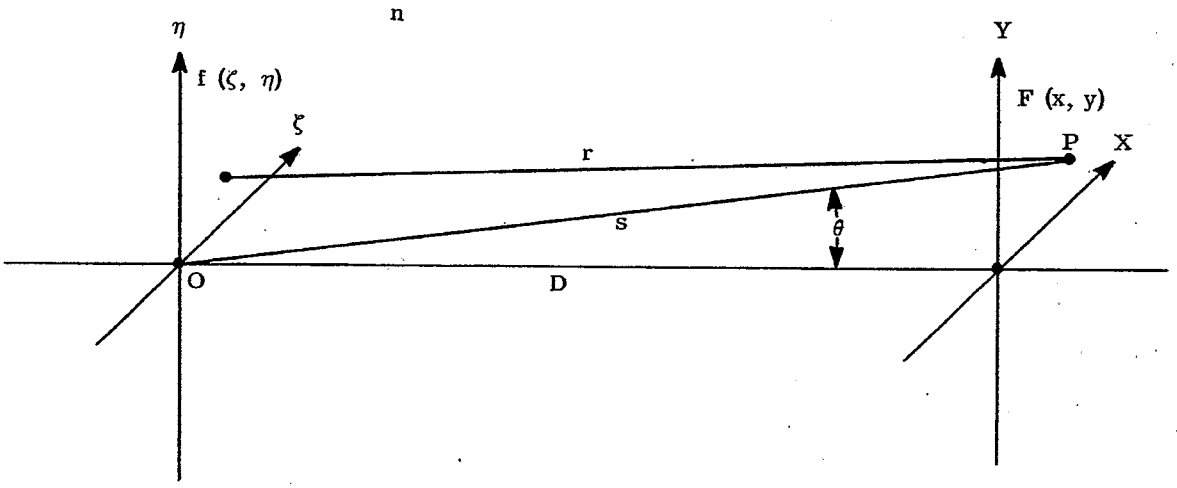

FIGURE 16.32 -Convention with respect to the integral statement of Huygens' Principle.

These Huygens wavelets expand from point (ζ, η) as described by Equations (141) and (142). Our problem is to sum the Huygens wavelets that leave all points (ζ, η) of the plane of the aperture and arrive at point P.

16.22.1.4 To formulate the problem a bit more generally without adding unduly to the complexity of presenting the problem, we can suppose that $f(\zeta, \eta)\, d\zeta\, d\eta$ is a complex number that specifies the amplitude and phase of the coherent Huygens wavelets that leave the area $d\zeta\, d\eta$. (We shall deal mainly with the simple cases in which $f(\zeta, \eta) = 1$.) According to Equation (141), the Huygens wavelets that leave the area $d\zeta\, d\eta$ with the amplitude and phase expressed by $f(\zeta, \eta)\, d\zeta\, d\eta$ arrive at point P with the amplitude and phase given by

$$f(\zeta, \eta)\, d\zeta\, d\eta\, \frac{e^{i(knr - \omega t)}}{r} .$$

Let $F(x, y)$ be the complex number that denotes the sum of all of the interfering Huygens wavelets that arrive at the point of observation P of Figure 16. 32. From the theory of integral calculus this sum is given at once by the integral

$$F(x,y) = e^{-i\omega t} \int\int f(\zeta, \eta)\, \frac{e^{iknr}}{r}\, d\zeta\, d\eta \qquad (143)$$

in which the integration extends over the illuminated area of the $\zeta\eta$ plane and in which r is given by Equation (142).

16.22.1.5 Before passing on to the Fraunhofer form of the integral given in Equation (143), we remark that the term Fresnel diffraction (as distinguished from Fraunhoffer diffraction) is applied to the cases in which the distance D from the plane of the aperture to the plane of observation is relatively small. Equation (143) is the most general statement of Huygens' principle. It includes Fresnel and Fraunhofer diffraction as special cases.

16.22.1.6 The Fraunhofer specialization diffraction integral is obtained in the following way from the integral of Equation (143) and the supposition that D is large. By expanding the squares in Equation (142) and defining

$$S \equiv \left[D^2 + x^2 + y^2\right]^{1/2} , \qquad (144)$$

one finds that

$$r = S\left[\, 1 + \frac{\zeta^2 + \eta^2}{S^2} - \frac{2(x\zeta + y\eta)}{S^2} \,\right]^{1/2} \qquad (145)$$

in which S has the geometrical meaning illustrated in Figure 16. 32. We suppose that D becomes great but that the aperture opening at the $\zeta\eta$ plane remains finite. Equivalently, but somewhat more generally, we may say that $f(\zeta, \eta) = 0$ when $(\zeta^2 + \eta^2)^{1/2}$ exceeds some finite value and that D can approach infinity. Under these circumstances, the quantity $(\zeta^2 + \eta^2)/S^2$ in Equation (145) is surely negligible. Since x and y can become infinite at $D = \infty$, the quantity $2(x\zeta + y\eta)/S^2$ is not entirely negligible. Because ζ and η will be small in comparison to S, $1 >> 2(x\zeta + y\eta)/S^2$. Hence, with excellent approximation,

$$\left[1 - 2\left(\frac{x\zeta + y\eta}{S^2}\right)\right]^{1/2} = 1 - \frac{x\zeta + y\eta}{S^2} \qquad (146)$$

Therefore,

$$r = S - \frac{x\zeta + y\eta}{S} \qquad (147)$$

16.22.1.7 Upon introducing r from Equation (147) into Equation (143), it will suffice to set $r = S$ in the denominator since $(x\zeta + y\eta)/S$ will be very small. However, the quantity $(x\zeta + y\eta)/S$ is multiplied by the large factor $k = 2\pi/\lambda$ in the exponent. We now introduce $r = S$ in the denominator of Equation (143) and r from Equation (147) into the exponent and thus obtain

$$F(x, y) = e^{-i\omega t}\, \frac{e^{iknS}}{S} \int\int f(\zeta, \eta)\, e^{-ikn\frac{x\zeta + y\eta}{S}}\, d\zeta\, d\eta \qquad (148)$$

in which

$$S = (D^2 + x^2 + y^2)^{1/2} \qquad (149)$$
$$k = 2\pi/\lambda$$
$$\omega = 2\pi/T .$$

$f(\zeta, \eta)$ specifies the amplitude and phase of the disturbance as it leaves the plane of the aperture. The integration extends over the plane of the aperture. In case the aperture consists, for example, of an opaque screen

with a hole in it, the integration with respect to $d\zeta\, d\eta$ extends over the area of the hole. $F(x, y)$ is a complex number that specifies the amplitude and phase of the so-called Fraunhofer region.

16.22.1.8 The energy density, $W(x, y)$, is proportional to $|F(x, y)|^2$. Since $|e^{-i\omega t}|^2 = 1$ and $|e^{iknS}|^2 = 1$, it follows from Equation (148) that

$$W(x, y) = \frac{1}{S^2} |F_o(x, y)|^2 \tag{150}$$

where

$$F_o(x, y) = \int\int_{\text{over plane of aperture}} f(\zeta, \eta)\, e^{-ikn\frac{x\zeta + y\eta}{S}}\, d\zeta\, d\eta . \tag{151}$$

It suffices therefore to compute the slightly simpler integral, $F_o(x, y)$, of Equation (151) when one wishes to determine the time-averaged distribution $W(x, y)$ of energy density produced at point (x, y) by the radiation in a coherent wave that illuminates the $\zeta\eta$ plane of the aperture.

16.23 FRAUNHOFER DIFFRACTION FROM A RECTANGULAR APERTURE

16.23.1 Discussion of principles.

16.23.1.1 We suppose for simplicity that the rectangular aperture is illuminated as in Figure 16. 32 by a plane wave at normal incidence. It suffices to set

$$f(\zeta, \eta) = \text{constant} = 1 . \tag{152}$$

Then, from Equation (151),

$$F_o(x, y) = \int_{-a}^{a}\int_{-b}^{b} e^{\frac{-iknx\zeta}{S}}\, e^{\frac{-ikny\eta}{S}}\, d\zeta\, d\eta$$

$$= \int_{-a}^{a} e^{\frac{-iknx\zeta}{S}}\, d\zeta \int_{\cdot}^{b} e^{\frac{-ikny\eta}{S}}\, d\eta$$

$$= \frac{e^{\frac{iknxa}{S}} - e^{\frac{-iknyb}{S}}}{(iknx)/S} \cdot \frac{e^{\frac{iknxa}{S}} - e^{\frac{-iknyb}{S}}}{(ikny)/S} .$$

Since $k = 2\pi/\lambda$ and $\sin z = (e^{iz} - e^{-iz})/2i$,

$$F_o(x, y) = 4ab\left[\frac{\sin(2\pi anx/S\lambda)}{2\pi anx/S\lambda}\right]\left[\frac{\sin(2\pi bny/S\lambda)}{2\pi bny/S\lambda}\right] . \tag{153}$$

From Equations (152) and (150), the corresponding time-averaged distribution of energy density in the observation plane is given by

$$W(x, y) = \frac{16a^2b^2}{S^2}\left[\frac{\sin(2\pi anx/S\lambda)}{2\pi anx/S\lambda}\right]^2\left[\frac{\sin(2\pi bny/S\lambda)}{2\pi bny/S\lambda}\right]^2 . \tag{154}$$

Along, for example, the line $y = 0$,

$$W(x, 0) \equiv W(x) = \frac{A^2}{S^2}\left[\frac{\sin(2\pi anx/S\lambda)}{2\pi anx/S\lambda}\right]^2 \tag{155}$$

because $(\sin u)/u = 1$ when $u = 0$. $A \equiv 4ab$ is the area of the rectangular aperture. We can take $S \doteq D$ for most purposes. $W(x)$ assumes its greatest value $W = A^2/S^2$ at $x = 0$. $W(x)$ decreases as $1/x^2$. The energy density is zero whenever $(2\pi anx)/S\lambda = \nu\pi$ where ν is an integer. Hence, the zeros of $W(x)$ occur at the points x for which

$$\frac{x_\nu}{S} = \sin\theta_\nu = \frac{\nu\lambda}{2an} \tag{156}$$

where $\nu = \pm 1, \pm 2, \pm 3$, etc.; $2a$ is the width of the rectangular aperture along the x-direction; n is the refractive index of the space; and θ_ν is the angle θ (Figure 16. 33) that corresponds to x_ν along the line $y = 0$.

16.23.1.2 Similiar conclusions hold along the line $x = 0$. One has only to substitute y for x and b for a in Equations (155) and (156). The width of the aperture along the y-direction is 2b.

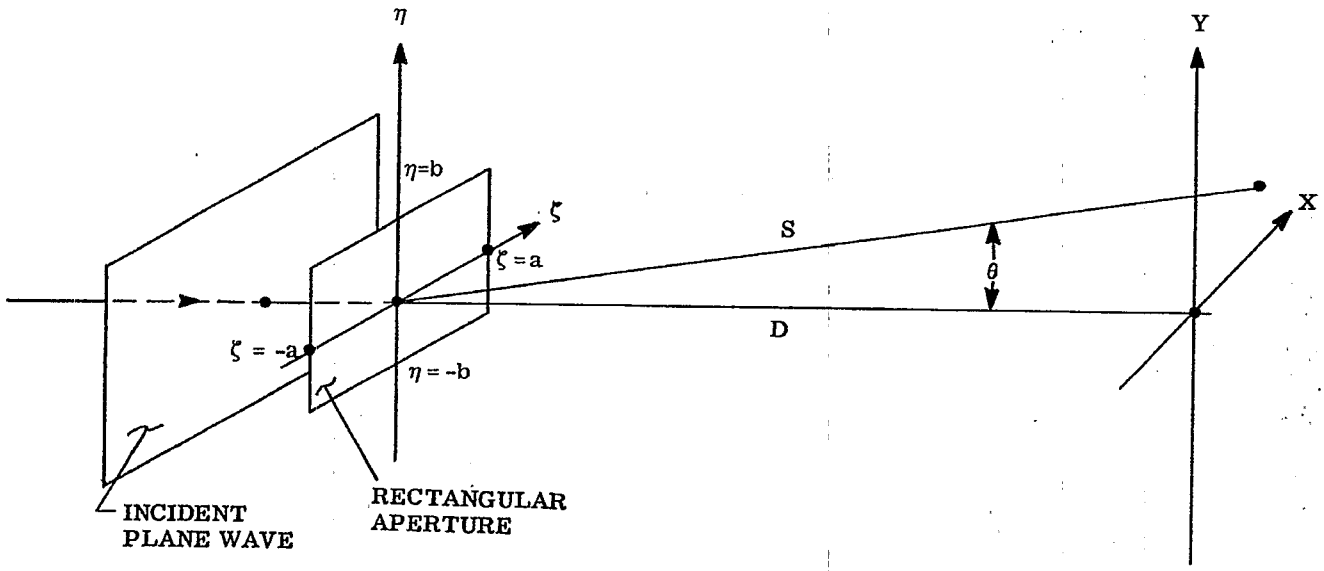

FIGURE 16.33-Notation with respect to diffraction from a rectangular aperture illuminated at normal incidence.

16.24 FRAUNHOFER DIFFRACTION FROM CIRCULAR APERTURES

16.24.1 Discussion of principles.

16.24.1.1 We suppose that a plane wave is incident normally upon the circular aperture so that $f(\zeta, \eta) = 1$. It is convenient to replace ζ, η and x, y by polar coordinates because the aperture is circular. Let

$$\zeta = u \cos \phi \quad ; \quad \eta = u \sin \phi \quad ; \tag{157}$$

$$x = r \cos \alpha \quad ; \quad y = r \sin \alpha \quad ; \tag{158}$$

in which the geometrical meanings of u, ϕ, r, and α are illustrated in Figure 16.34. Upon introducing Equations (157) and (158) into Equation (151) and setting $f(\zeta, \eta) = 1$, one obtains

$$F_o(x, y) \equiv F_o(r) = \int_0^a \int_0^{2\pi} e^{\frac{-iknru}{S}\cos(\phi - \alpha)} \, u \, du \, d\phi \tag{159}$$

in which

$$S = (D^2 + x^2 + y^2)^{1/2} = (D^2 + r^2)^{1/2} . \tag{160}$$

16.24.1.2 One can prove that $F_o(r)$ must be independent of α because the integrand of Equation (159) is periodic in the angle ϕ. However, it is clear from Figure 16.34 that $F_o(r)$ should be independent of the angle α because the system has complete axial symmetry. Hence, we can set

$$\alpha = 0 \tag{161}$$

in Equation (159).

16.24.1.3 Now one finds from almost all text books treating the elementary theory of Bessel functions that

$$\int_0^{2\pi} e^{\pm iz\cos\phi} \, d\phi = 2\pi J_o(z) \tag{162}$$

where $J_o(z)$ is a Bessel function of zero order and first kind. From Equations (159) and (162)

$$F_o(r) = 2\pi \int_0^a J_0\left(\frac{knru}{S}\right) u \, du . \tag{163}$$

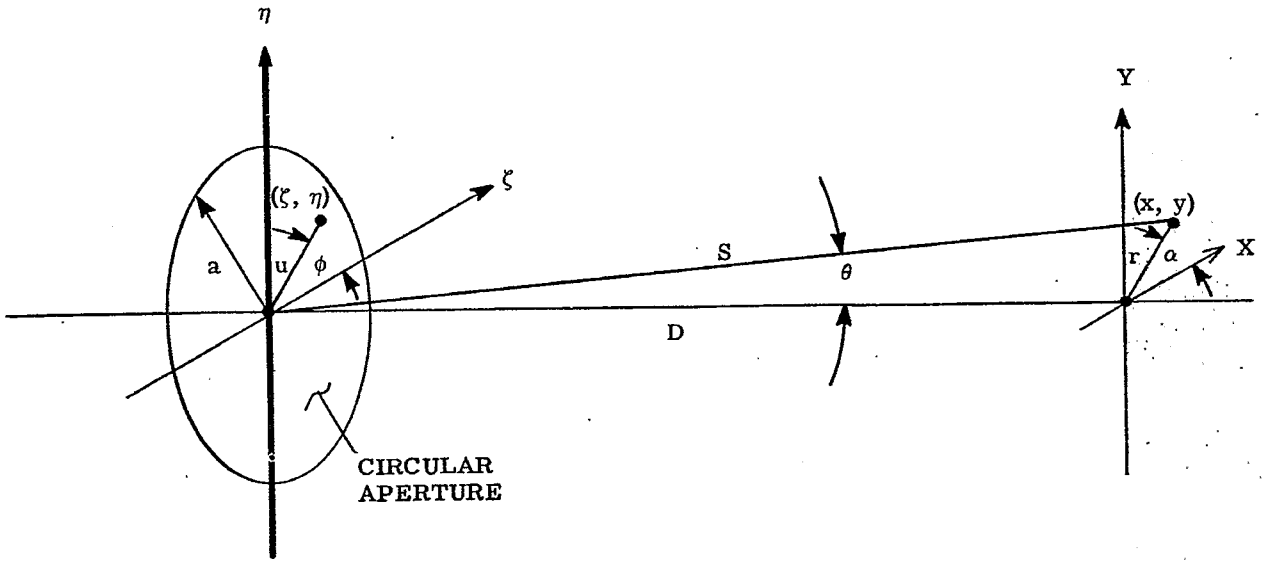

FIGURE 16. 34- Notation with respect to Fraunhofer Diffraction from circular apertures.

Introduce the change of variable

$$v \equiv \frac{knru}{S} \quad , \text{ or } \quad u \equiv \frac{S}{knr} v \ . \tag{164}$$

Then,

$$F_o(r) \equiv 2\pi \left(\frac{S}{knr} \right)^2 \int_0^{\frac{knra}{S}} v\, J_o(v)\, dv \ . \tag{165}$$

It is another elementary proposition in Bessel Functions that

$$\int_0^z v\, J_o(v)\, dv = z\, J_1(z) \tag{166}$$

where $J_1(z)$ is a Bessel function of first order and first kind. Since $J_1(z) = 0$ at $z = 0$, one finds directly from Equations (165) and (166) that

$$F_o(r) = 2\pi \left(\frac{S}{knr} \right)^2 \frac{knra}{S} \ J_1 \left(\frac{knra}{S} \right) \ .$$

Whence

$$F_o(r) = 2\pi a^2 \ \frac{J_1(2\pi na\ r/\lambda S)}{(2\pi nar)/\lambda S} \ . \tag{167}$$

16.24.1.4 We note from Figure 16.34 that

$$\sin\theta = r/S \ . \tag{168}$$

Alternatively, one may therefore write

$$F_o(\theta) \equiv 2\pi a^2 \cdot \frac{J_1(2\pi na \sin\theta/\lambda)}{2\pi na \sin\theta/\lambda} \tag{169}$$

The energy density in the Fraunhofer diffraction image or pattern from a circular aperture of radius $\underline{a}$ is now given by Equation (150) in which one introduces F_o from Equation (167) or from Equation (169).

16.24.1.5 The function $[J_1(z)]/z$ is 1/2 at $z = 0$ and assumes its first zero at $z = 3.8317$. Therefore, the energy density in the Fraunhofer diffraction pattern has its first zero minimum at $(2\pi na \sin \theta_1) / \lambda = 3.8317$ or at

$$\sin \theta_1 = \frac{r_1}{S} = \frac{0.61\lambda}{na} = \frac{1.22\lambda}{2an} \qquad (170)$$

in which 2a is the diameter of the aperture. It is instructive to compare Equations (170) and (156) at the first zero where $\nu = 1$. We see that the central maximum in the diffraction pattern is 22 per cent larger in linear dimension for the circular aperture than for the rectangular aperture whose width is equal to the diameter of the circular aperture. The Bessel function $J_1(z)$ oscillates with increasing z in such a way that successive maxima and minima of $J_1(z)$ decrease numerically. Hence, the energy density

$$W(r) = \frac{4\pi^2 a^4}{S^2} \left[\frac{J_1(2\pi nar / S\lambda)}{2\pi nar / S\lambda} \right]^2 \qquad (171)$$

in the diffraction pattern produced by a circular aperture decreases considerably faster with increasing distance r from the diffraction head than does the energy density $W(x)$ produced by a rectangular aperture. (Compare Equations (155) and (171).) One must expect that circular apertures are preferable to rectangular apertures for lenses because the diffraction images produced by circular apertures are, on the whole, more concentrated.

16.25 DIFFRACTION FROM SPHERICAL WAVEFRONTS

16.25.1 General. Whereas the methods of paragraphs 16.23 and 16.24 can be utilized as a basis for discussing the diffraction images produced by lenses, the adaptation of these methods is a bit too artificial and leads, awkwardly, to the predictions that resolving power is related to the tangent of certain axial angles rather than to the sine of these angles.

16.25.1.1 It is the purpose of a well corrected lens to convert a spherical wave that diverges from an object point into a spherical wave that converges upon the conjugate image point as in Figure 16.34. We suppose for

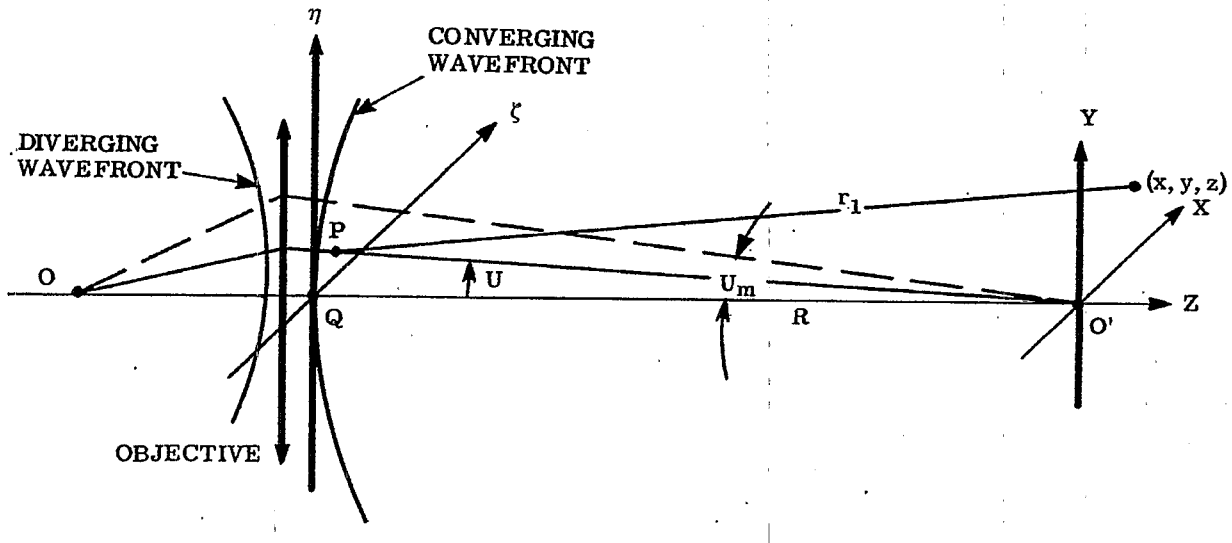

FIGURE 16.35- Convention with respect to the formation of a diffraction image, O', of O by a lens system.

simplicity of presentation that the object point O is located upon the axis. Let V be the optical path from O to O' . We draw a reference sphere of radius R about the point O' such that this sphere touches the tangent plane $\zeta\eta$ at point Q on the axis. The optical path from O to Q is now V - nR where n is the refractive index of the image space. Similarly, the optical path from point O to any point P on the $\zeta\eta$ plane from point O to any point P on the $\zeta\eta$ plane is

$$V - n\,O'\,P = V - n\,(R^2 + \zeta^2 + \eta^2)^{1/2} \tag{172}$$

in the absence of spherical aberration. The Huygens wavelets now leave the $\zeta\eta$ plane with an amplitude-phase distribution given by

$$f(\zeta, \eta) = \frac{e^{ikV}\, e^{-ikn\,(R^2 + \zeta^2 + \eta^2)^{1/2}}}{(R^2 + \zeta^2 + \eta^2)^{1/2}} \tag{173}$$

16.25.1.2 We choose the origin of the coordinates X, Y, Z at the point O' with O' conjugate to O . Thus, the plane z = 0 is the sharply focused image plane. The problem is to find the amplitude-phase distribution F (x, y, z) produced by all the Huygens wavelets that leave the $\zeta\eta$ plane. From Equations (143) and (173),

$$F(x,y,z) = e^{-i\omega t}\, e^{ikV} \iint \frac{e^{-ikn\,(R^2 + \zeta^2 + \eta^2)^{1/2}}}{(R^2 + \zeta^2 + \eta^2)^{1/2}} \cdot \frac{e^{iknr_1}}{r_1}\, d\zeta\, d\eta \tag{174}$$

where the distance r_1 of Figure 16. 35 is

$$r_1 = [(x - \zeta)^2 + (y - \eta)^2 + (R + z)^2]^{1/2} \tag{175}$$

However, one finds after slight rearrangement that

$$r_1 = (R^2 + \zeta^2 + \eta^2)^{1/2} \left[\frac{1 + x^2 + y^2 + z^2 - 2(x\zeta + y\eta - Rz)}{R^2 + \zeta^2 + \eta^2}\right]^{1/2} \tag{176}$$

16.25.1.3 In order to obtain the approximation to r_1 that leads to the conventional diffraction integral for lenses, we have to suppose that the field of view is so small that one can afford to neglect the term $(x^2 + y^2 + z^2) / (R^2 + \zeta^2 \eta^2)$ in Equation (176). This means that the following theory holds best for small fields of view. We have to suppose also that the dimensions of the aperture at the $\zeta\eta$ plane and the out-of-focus distance z are small enough for us to be willing to accept the approximation

$$\left[1 - 2\,\frac{(x\zeta + y\eta - Rz)}{R^2 + \zeta^2 + \eta^2}\right]^{1/2} = 1 - \frac{x\zeta + y\eta - Rz}{R^2 + \zeta^2 + \eta^2} \tag{177}$$

Under these approximations,

$$r_1 = (R^2 + \zeta^2 + \eta^2)^{1/2} - \frac{x\zeta + y\eta - Rz}{(R^2 + \zeta^2 + \eta^2)} \tag{178}$$

16.25.1.4 Upon substituting r_1 from Equation (178) into the integral (174) it suffices to set $r_1 = (R^2 + \zeta^2 + \eta^2)^{1/2}$ in the denominator. Our approximation for F (x, y, z) becomes

$$F(x,y,z) = e^{-i\omega t}\, e^{iknV} \iint \frac{e^{-ikn\,\frac{x\zeta + y\eta - Rz}{\sqrt{R^2 + \zeta^2 + \eta^2}}}}{R^2 + \zeta^2 + \eta^2}\, d\zeta\, d\eta \tag{179}$$

in which the integration extends over the aperture of the objective, Figure 16. 35.

16.25.1.5 Since $|F(x, y, z)|^2$ is independent of ωt and V, it is convenient to drop the external exponentials in Equation (179) and to write again

$$F_o(x,y,z) \equiv \iint \frac{e^{-ikn\,\frac{x\zeta + y\eta - Rz}{\sqrt{R^2 + \zeta^2 + \eta^2}}}}{R^2 + \xi^2 + \eta^2}\, d\zeta\, d\eta \tag{180}$$

The time-averaged energy density in the diffraction image of an object point located upon the axis is

$$W(x,y,z) = |F(x,y,z)|^2 = |F_o(x,y,z)|^2 \tag{181}$$

The plane z = 0 is the sharply focused image plane.

16.26 PRIMARY DIFFRACTION INTEGRALS WITH OBJECTIVES HAVING CIRCULAR APERTURES

16.26.1 Introduction. We shall call the integral $F_o(x, y, z)$ of Equation (180) the primary diffraction integral and shall refer to the corresponding distribution of energy density $W(x, y, z)$ as the primary diffraction image. These two quantities are of fundamental importance to the diffraction theory of optical instruments. In this section, the primary diffraction integral will be specialized to the great class of objectives that have circular apertures. Thus far, the objective has been assumed free of spherical aberration.

16.26.1.1 Corresponding elements of Figures 16. 35 and 16. 36 are labeled alike. One notes from Figure 16. 36 that

$$\zeta = R \tan U \cos \phi \quad ; \quad \eta = R \tan U \sin \phi \quad ; \tag{182}$$

and that

$$\cos U = R / \sqrt{R^2 + \zeta^2 + \eta^2} \quad . \tag{183}$$

Hence,

$$\frac{\zeta}{\sqrt{\zeta^2 + \eta^2 + R^2}} = \sin U \cos \phi \quad ; \quad \frac{\eta}{\sqrt{\zeta^2 + \eta^2 + R^2}} = \sin U \sin \phi \, . \tag{184}$$

16.26.1.2 It is convenient to change the variables of integration from ζ and η to U and ϕ. Since

$$dA = \begin{vmatrix} \frac{\partial \zeta}{\partial U} & \frac{\partial \eta}{\partial U} \\ \frac{\partial \zeta}{\partial \phi} & \frac{\partial \eta}{\partial \phi} \end{vmatrix} dU \, d\phi = d\zeta \, d\eta \, ,$$

wherein ζ and η are given by Equation (182)

$$d\zeta \, d\eta = R^2 \frac{\sin U}{\cos^3 U} \, dU \, d\phi \quad . \tag{185}$$

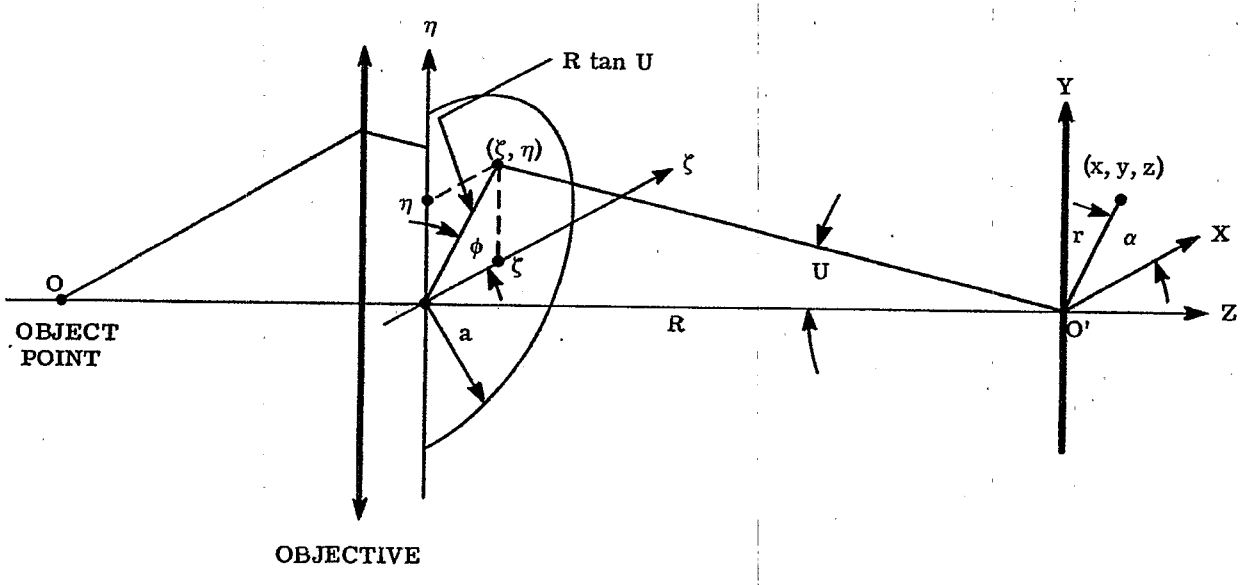

FIGURE 16. 36--Notation with respect to objectives that have axial symmetry and circular apertures of radius a.

Upon substituting from Equations (183), (184), and (185) into Equation (180), one obtains the result

$$F_o(x, y, z) = \int_0^{U_m} \int_0^{2\pi} e^{-ikn\left[\sin U\ (x \cos\phi + y \sin\phi)\right] - z \cos U} \frac{\sin U}{\cos U}\, dU\, d\phi \quad (186)$$

in which U_m is the largest value of U in the cone of axial rays that pass from the object point O to the conjugate point O' . Equation (186) is the Luneburg-Debye statement of the primary diffraction integral.

16.26.1.3 A change of variable from U to the zonal numerical apertures ρ where

$$\rho \equiv \sin U \ ; \quad \rho_m = \sin U_m \ ; \quad (187)$$

renders both the form and the physical interpretation of the primary diffraction integral somewhat simpler. One obtains from Equations (186) and (187) the result

$$F_o(x, y, z) = \int_0^{\rho_m} \int_0^{2\pi} e^{ikz\sqrt{1-\rho^2}}\, e^{-ikn\rho\,(x \cos\phi + y \sin\phi)} \frac{\rho\, d\rho\, d\phi}{1-\rho^2} \quad (188)$$

Equation (188) is known to hold well for the image space of microscope objectives, telescopes, etc., in which ρ_m is so small that one can set $1 - \rho^2 = 1$ in the denominator. For example, with microscope objectives $\rho_m \approx 3/150 = 0.02$ so that $\rho^2 \leq 0.0004$, a quantity that can be ignored in the denominator of (188) but not in the exponential of the numerator except when $z = 0$, i. e., except when one focuses upon the plane which is conjugate to the object point. In computing $F_o(x, y, 0) = F_o(x, y)$ for the conjugate plane $z = 0$, one obtains the Fraunhofer type of diffraction integral

$$F_o(x, y) = \int_0^{\rho_m} \int_0^{2\pi} e^{-ikn\rho\,(x \cos\phi + y \sin\phi)}\, \rho\, d\rho\, d\phi \quad (189)$$

upon neglecting ρ^2 in the denominator.

16.26.1.4 Typical of diffraction integrals of the Fraunhofer type, the integral (189) is easily integrated. Introduce polar coordinates r, α such that

$$x = r \cos\alpha \ ; \quad y = r \sin\alpha \ . \quad (190)$$

Then from Equation (189)

$$F_o(x, y) = F_o(r) = \int_0^{\rho_m} \int_0^{2\pi} e^{-ikn\rho r \cos(\phi - \alpha)}\, \rho\, d\rho\, d\phi \ . \quad (191)$$

As in the integral (159), $F_o(r)$ is independent of α . Furthermore, from Equations (162) and (191)

$$F_o(r) = 2\pi \int_0^{\rho_m} J_o(kn\rho r)\, \rho\, d\rho \ . \quad (192)$$

Comparison of Equations (192) and (163) shows that the integral (192) is obtained from the integral (163) by setting $S = 1$ and $a = \rho_m$. Hence, we conclude at once from Equation (157) that

$$F_o(r) = 2\pi\, \rho_m^2\, \frac{J_1(2\pi n \rho_m r / \lambda)}{2\pi n \rho_m r / \lambda} \quad (193)$$

wherein r is the distance from the diffraction head, and

$$n\rho_m = n \sin U_m \quad (194)$$

is the zonal numerical aperture of the objective with respect to its image space of refractive index n . We see that $F_o(r)$ is a real number when it is evaluated at the sharply focused image plane $z = 0$ for objectives that have negligible spherical aberration. The time-averaged-energy density $W(r) = |F_o(r)|^2$. Thus,

$$W(r) = 4\pi^2 \rho_m^4 \left[\frac{J_1(2\pi n \rho_m r / \lambda)}{2\pi n \rho_m r / \lambda}\right]^2 \ . \quad (195)$$

a result that should be compared with that of Equation (171).

16.26.1.5 The primary diffraction integrals (191), (192), and (193) are called the Airy type, and the corresponding, idealized objectives are distinguished as the Airy type of objective. As in the discussion leading to Equation (170), the first zero of $W(r)$ occurs at $(2\pi n \rho_m r) / \lambda = 3.8317$ or at

$$r = r_1 = \frac{3.8317}{2\pi n \rho_m} \lambda = \frac{0.6098}{n \rho_m} \lambda \ . \quad (196)$$

r_1 is the distance from the diffraction head (where $W(r) = W(0) = \pi^2 \rho_m^4$, its maximum value) to the first

zero of $W(r)$ in the image space. The distance r_1 is frequently utilized as unit distance and is called the Airy unit with respect to the image space. The quantity $n\rho_m = n \sin U_m$ is the numerical aperture of the objective with respect to its image space.

16.27 RESOLUTION WITH CIRCULAR APERTURES

16.27.1 General. It is not possible to specify a universal limit of resolution that applies to all kinds of details in an object field. Resolving power varies with the type of details that are to be resolved, with the manner in which the object is illuminated, with the wavelength utilized for illumination, with the numerical aperture of the objective, and with the degree of correction of the objective. Resolution can depend upon the type of optical system. For example, it can be shown theoretically that an ordinary microscope cannot resolve two nonabsorbing particles, irrespective of their separation, when the optical path difference Δ between the particle and its surround becomes so small that $\sin \Delta$ can be replaced by Δ. The chief reason for this peculiarity is that with such particles, contrast in the image becomes so poor that one cannot actually observe the particles. When the ordinary microscope is replaced by a phase microscope, contrast in the image is increased enormously. Consequently, the phase microscope can exhibit resolving power when the ordinary microscope does not. Finally, resolving power depends upon the criterion that one is willing to accept in concluding from the observation of the image that the details in question are distinct, i.e., are resolved.

16.27.1.1 We shall restrict our considerations of resolving power to the resolution of two self-luminous particles whose dimensions are neglibible. Let one particle be located at point O on the axis as in Figure 16.37 According to Rayleigh's criterion of resolution, two object points O and P will be resolved provided their separation equals or exceeds the separation r_o for which the maximum energy density in the diffraction image of one particle falls upon the first minimum in the diffraction image of the second particle as illustrated in Table 16.1. We have seen in the previous section that the distance r_1 from the central maximum to the first minimum is given by Equation (196) for objectives of the Airy type. Hence, the linear limit of resolution is r_1 (or one Airy unit) in the image space where

$$r_1 = \frac{0.6098}{n\rho_m} \lambda \ . \tag{197}$$

Therefore,

$$r_o = \frac{r_1}{|M|} = \frac{0.6098}{|M| n\rho_m} \lambda \tag{198}$$

is the linear limit of resolution in the object space where M denotes the magnification ratio. If the objective obeys the Abbe sine condition,

$$|M| n\rho_m = |M| n \sin U_m = n_o \sin U_{o,m} \cong \text{N.A.} \tag{199}$$

where N.A. denotes the numerical aperture of the objective with respect to its object space. Therefore,

$$r_o = \frac{0.6098}{\text{N.A.}} \lambda \ . \tag{200}$$

The linear limit of resolution, r_o, for two self-luminous object points is one Airy unit with respect to the object space of the objectives that approximate the Airy type.

16.27.1.2 The corresponding angular limit of resolution, θ_1, is given by

$$\theta_1 = \frac{r_1}{V} = \frac{0.6098 \lambda}{n\rho_m V} \tag{201}$$

in which V is the image distance, Figure 16.37. When U_m is small as in the image space of telescopes and microscope objectives

$$\tan U_m = \frac{D}{2V} \rightarrow \sin U_m = \rho_m \ . \tag{202}$$

Hence

$$\theta_1 = \frac{(2)(0.6098 \lambda)}{nD} = \frac{1.22\lambda}{nD} \tag{203}$$

where D is the diameter of the objective and n is the refractive index of the image space.

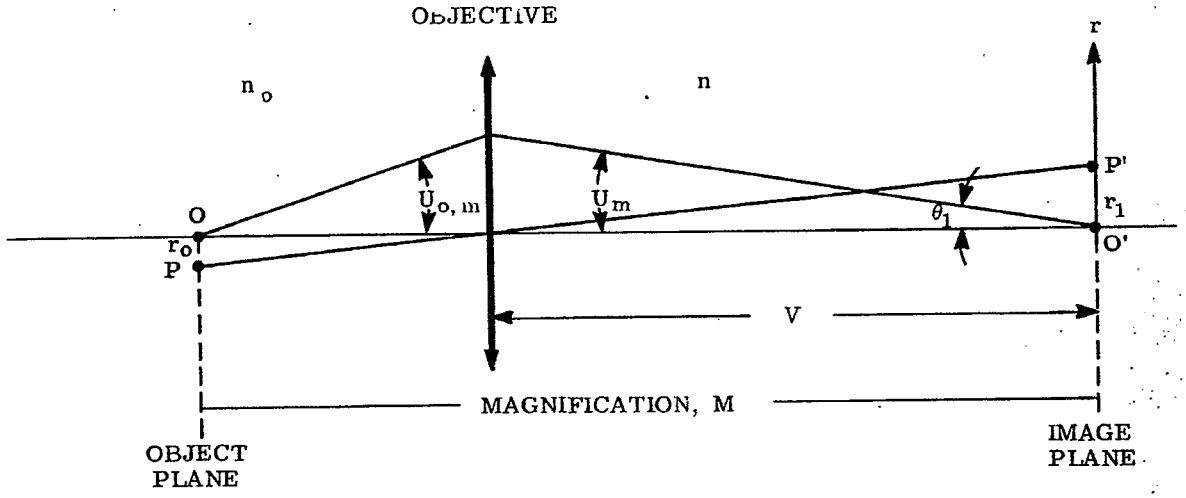

FIGURE 16. 37- Notation with respect to the resolution of two self-luminous object points by objectives having circular apertures.

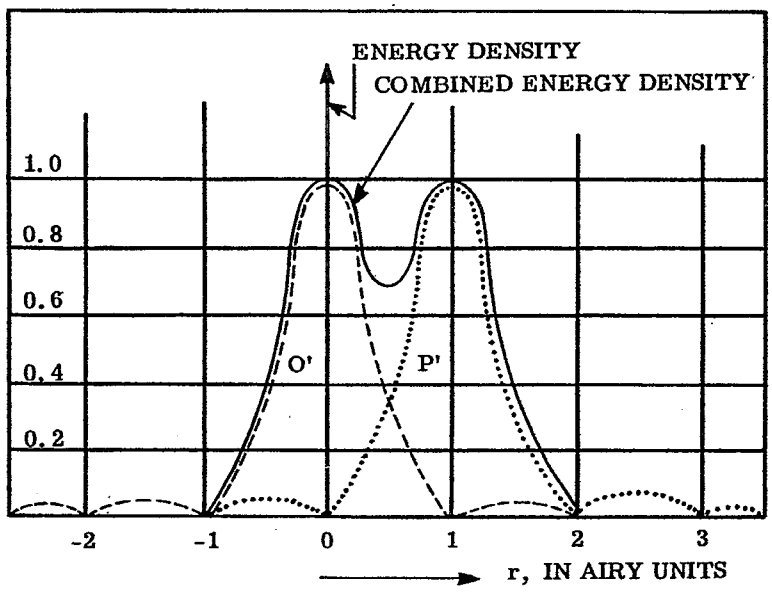

TABLE 16. 1- Physical situation at the limit of resolution based on Rayleigh's criterion. O' and P' are the curves of the energy densities in the image of two, like, self-luminous particles, O and P, respectively. The solid curve is the sum of the energy densities due to the two particles. This solid curve displays an easily seen dip at 0. 5 Airy Unit, the mid-point between the geometrical images of the two particles.

16.27.1.3 The limit of resolution obtained from Rayleigh's criterion is a conservative limit with highly corrected objectives. The Sparrow * or physical limit of resolution is 0.78 Airy units for Airy type objectives. In principle, this limit can be approached but not realized. Many observations have indicated that resolutions near 0.81 Airy units have been achieved with highly corrected objectives.

16.28 OUT-OF-FOCUS ABERRATION

16,28.1 General.

16.28.1.1 The out-of-focus aberrations for axial object points are included in Equation (188) in which the system is out-of-focus by the amount z. The integration with respect to $d\phi$ can be carried out just as in the argument leading to the integral (192) even when $z \neq 0$. One obtains instead of (192) the integral

$$F_o(r) = 2\pi \int_o^{\rho_m} e^{iknz\sqrt{1-\rho^2}} J_o(knr\rho)\ \rho\, d\rho \qquad (204)$$

when ρ^2 is ignored in the denominator of the primary diffraction integral (188).

16.28.1.2 In the presence of spherical aberration and out-of-focus aberration, one finds in general that $F_o(r)$ is of the form

$$F_o(r) = 2\pi \int_o^{\rho_m} P(\rho)\ J_o(knr\rho)\ \rho\, d\rho \qquad (205)$$

for axial object points where $P(\rho)$ is called the pupil function. We see that the pupil function $P_z(\rho)$ corresponding to out-of-focus aberration is

$$P_z(\rho) = e^{iknz\sqrt{1-\rho^2}} \qquad (206)$$

Whenever $\rho^2 << 1$, it is usual to accept the approximation

$$\sqrt{1-\rho^2} = 1 - \frac{\rho^2}{2} \qquad (207)$$

and to write

$$F_o(r) = 2\pi\, e^{iknz} \int_o^{\rho_m} e^{i\pi nz\rho^2/\lambda} J_o(knr\rho)\ \rho\, d\rho\ , \qquad (208)$$

a result that follows from Equations (205), (206), and (207).

16.28.1.3 The following is one of the simplest methods for estimating the maximum tolerable amount z that an objective of given numerical aperture $n\rho_m$ can be out-of-focus. Equation (208) is easily integrated for any axial image point r = 0 because $J_o(0) = 1$. Thus

$$F_o(0) = 2\pi\, e^{iknz} \int_o^{\rho_m^2} e^{\frac{i\pi nz\rho^2}{\lambda}}\ \rho\, d\rho$$

$$= 2\pi\, e^{iknz} \left[e^{\frac{i\pi nz\rho_m^2}{\lambda}} - 1 \right] \Big/ \frac{i\pi nz}{\lambda}\ . \qquad (209)$$

The corresponding energy density $W(0) = |F_o(0)|^2$ is now

$$W(0) = 4\pi^2 \left(e^{\frac{i\pi nz\rho_m^2}{\lambda}} - 1 \right)\left(e^{\frac{-i\pi nz\rho_m^2}{\lambda}} - 1 \right) \Big/ \frac{\pi^2 n^2 z^2}{\lambda^2}$$

$$W(0) = 8\pi^2 \left[1 - \cos\left(\pi nz\rho_m^2/\lambda \right) \right] \Big/ \frac{\pi^2 n^2 z^2}{\lambda^2}\ , \text{ or}$$

$$W(0) = 4\pi^2 \rho_m^4 \left[\frac{\sin\left(\pi nz\rho_m^2/2\lambda \right)}{\pi nz\rho_m^2/2\lambda} \right]^2\ , \qquad (210)$$

where $W(0)$ is the energy density at the diffraction head when the objective is out-of-focus by the amount z. $n\rho_m$ is the numerical aperture of the objective with respect to its image space.

16.28.1.4 When z = 0,

$$W(0) = W_o = 4\pi\rho_m^4\ , \qquad (211)$$

* See H. Osterburg, Microscope Imagary and Interpretation, J. Opt. Soc. Amer,, 40, 299 (1950).

a result that agrees, as it should, with W(0) from Equation (195). Let

$$K = \frac{W(0)}{W_o} = \left[\frac{\sin(\pi z n \rho_m^2 / 2\lambda)}{\pi z n \rho_m^2 / 2\lambda}\right]^2 \quad (212)$$

where K is the ratio of the energy density at the diffraction head when the objective is out-of-focus by the amount z to the energy density at the diffraction head when the objective is in focus. The ratio K is, we note, an even function of z when no spherical aberration is present. The assigned value of K becomes a criterion for the maximum tolerable out-of-focus distance z.

16.28.1.5 Suppose that

$$\pi |z| n \rho_m^2 / \lambda \leq \pi / 2 \quad . \quad (213)$$

This means (see Equation (208)) that the phase aberration due to being out of focus shall not exceed one-fourth wavelength. By introducing $(\pi z n \rho_m^2) / 2\lambda = \pi / 4$ into Equation (212), one finds that K = 0.8106. Hence, the criterion

$$K \geq 0.8106 \quad (214)$$

is equivalent to the criterion* of Equation (213). We learn from Equations (213) and (214) that if

$$|z| \leq \frac{1}{2} \frac{n\lambda}{(n\rho_m)^2} \quad , \quad (215)$$

the central energy density in the out-of-focus image of a self-luminous object point located upon the axis of the objective will not fall below 81.06 per cent of the maximum central energy density which occurs at the state of sharpest focus z = 0. $n\rho_m$ is the numerical aperture of the objective with respect to its image space. Consider, for example, the case in which the refractive index n of the image space is unity and in which $\rho_m \equiv \sin U_m = 0.1$. From Equation (215), $|z| \leq 0.5\lambda / 0.01 = 50$ wavelengths.

16.28.1.6 This diffraction theory for out-of-focus images will become less reliable as ρ_m becomes large; but within the range of applicability of the theory, the depth of focus should vary inversely as the square of the numerical aperture of the objective and directly as the wavelength. This conclusion is quite different from that based upon the more elementary notions of geometrical optics.

16.28.1.7 The reader who wishes to examine the applications of the more general primary diffraction integral (205) to cases in which the pupil function $P(\rho)$ includes spherical aberration and in which $r \neq 0$ may consult an excellent, detailed publication** by Guy Lansraux.

*This criterion is known as Rayleigh's criterion for phase aberrations.

**Guy Lansraux, "Calcul des Figures de Diffraction des Pupilles de Revolution," Revue D' Optique, 26, 24-45 (1947).