3 CONSIDERATIONS OF PHYSICAL OPTICS

3.1 INTRODUCTION

3.1.1 Diffraction nature of optical images.

3.1.1.1 The goal in designing a lens system on the basis of geometrical optics is to find a combination of lenses for which all rays in a specified cone of rays that diverges from an object point P are converged upon the corresponding image point P' such that the optical paths of all rays from P to P' are equal. Other requirements are added. For example, it may be required that points P and P' shall belong to a single object plane and a single image plane, respectively. Even when the design satisfies all these requirements to a high degree, the image P' of a self-luminous object point P is not a point but consists of a central bright spot surrounded by systematically distributed dark and bright fringes whose contour and width depend upon the contour and dimensions of the aperture of the lens. If, for example, the lens aperture is circular and if the self-luminous object point is located upon or near the optic axis, the image consists of a circular, central bright spot surrounded alternately by dark and bright rings. The central bright spot is called the Airy disk. Its diameter decreases as the diameter of the lens aperture is increased. The actual image of the object point is modified to such a degree by diffraction from the finite lens aperture that this image is appropriately called a diffraction image.

3.1.1.2 The diffractive nature of the image may not be so apparent with, for example, high-speed objectives in which compromises among the geometrical corrections and tolerable aberrations must be made. However, the image will generally exhibit effects due to diffraction, i. e., effects that cannot be explained from Snell's law of refraction or reflection alone. In any case, the image of a point will not be a point; an exact point by point similarity between object and image cannot be achieved. Resolution of details in the image of the object is restricted first by the degree of correction of the optical system and finally by the laws of diffraction, i. e., by the laws governing the bending of light rays from the paths consistent with Snell's law of refraction and reflection.

3.1.1.3 Whereas the action of most optical systems can be explained by the principles of geometrical optics, the action of other systems such as phase microscopy can be understood only as a proposition in diffraction. However, in any system, the ultimate resolving power and contrast in the fine-grained details of an image are determined by diffraction.

3.1.2 Diffraction and interference.

3.1.2.1 Broadly, diffraction is the phenomenon whereby waves are modified in direction, amplitude, and in phase by interaction with an object or obstacle. In its most general sense, diffraction includes the phenomena of refraction and reflection but these two phenomena are ordinarily considered apart from diffraction. However, when the dimensions of the object become comparable to the wavelength, the concepts of refraction and reflection become useless. With such small objects, even scattering becomes a direct aspect of diffraction.

3.1.2.2 Interference is the process by which two or more overlapping waves interact so as to re-enforce one another in some regions and to oppose one another in other regions. This process is essentially one of addition of the instantaneous amplitudes of the overlapping waves. It matters a great deal whether or not the overlapping waves are coherent. In case the added waves are incoherent, the time-averaged energy density is simply the sum of the time-average of the energy density associated with each wave, i. e., the resulting energy follows the law of superposition of energy. Conversely, it may be concluded that if the time-average of the energy densities follows the law of superposition of energy, the interfering waves are essentially incoherent. Interference includes the process by which a given wave is split or decomposed into two or more waves (often called component waves). These component waves are automatically coherent since they belong to the same wave-train. The action of interferometers can usually (but not always) be explained adequately by considering the sum of two or more waves.

3.1.2.3 Diffraction and interference are related processes, but diffraction is the more inclusive. In fact, diffraction effects can include interference effects as special cases. For example, in explaining the "interference fringes" produced with monochromatic light leaving two small pinholes that are illuminated coherently from a third pinhole, it is natural to regard the formation of the interference fringes as an interference effect, i. e., as a process of adding the two well defined spherical waves that emerge from the pair of pinholes. However, as the area of the pinholes is increased, the location of the origin of the spherical waves that leave different portions of the pinholes begins to matter. The process of summing the effects of the infinite many wavelets that leave the pinholes is now carried out most conveniently by means of integrals that characterize diffraction processes.

3.2 THE PHYSICAL NATURE OF LIGHT

3.2.1 The wave theory

3.2.1.1 Much evidence supports the view that light is propagated as electromagnetic waves whose wavelengths λ fall in the visible range from 0.38 to 0.76 microns. The transverse nature of electromagnetic waves is illustrated in Figure 3.1 in which E and H denote the electric and magnetic vectors, respectively. The electric

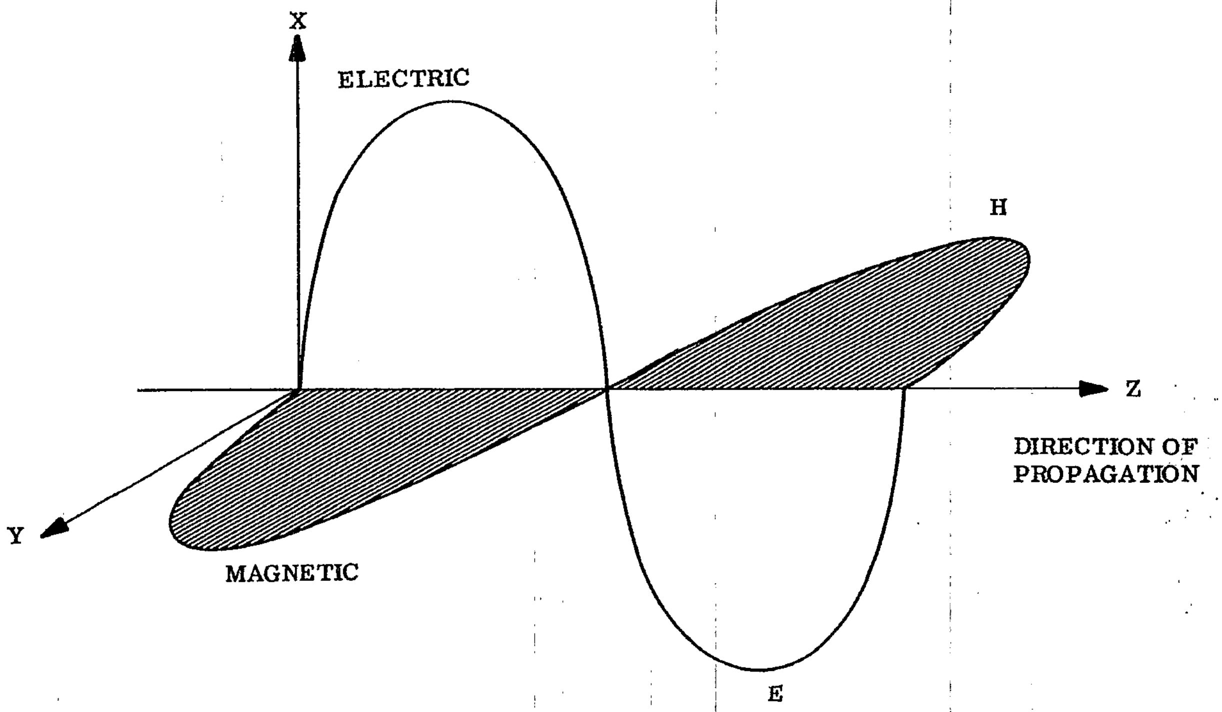

Figure 3. 1—The electromagnetic nature of a plane polarized light wave. The electric vector E and the magnetic vector H oscillate at right angles to the direction of propagation and at right angles to one another.

and magnetic vectors are ordinarily perpendicular to each other and to the direction of propagation. The electric vector describes an electric force field that will cause an electric charge to vibrate along the E-direction. Thus, the electric vector produces displacements of ions or electrons along the positive or negative E-direction, respectively. The vectors E and H are inseparable and are mutually dependent. For this reason it usually suffices to specify only the electric vector. The luminous flux can be computed whenever the radiant flux of the electromagnetic waves is known (as it is when the E-vector is specified).

3.2.1.2 The velocity of all electromagnetic waves in vacuum is a constant = c = 299792.5 kilometers per second. The velocity of monochromatic waves in non-vacuum media invariably depends upon the wavelength and is accordingly called the phase velocity to distinguish it from the group velocity of a group of monochromatic waves. The refractive index n of a medium is defined such that

$$n = \frac{\text{velocity in vacuum}}{\text{phase velocity in the medium}} \qquad (1)$$

Let T denote the period of vibration of a monochromatic wave. Let $\nu = 1/T$ denote the frequency ν of vibration. Then if v denotes the phase velocity

$$v = \nu \lambda = \frac{c}{n} \qquad (2)$$

As an electromagnetic wave moves from one medium into another, its frequency remains fixed. Hence its wavelength must change such that the wavelength λ in a medium of refractive index n varies according to the law

$$\lambda = \frac{c}{n\nu} = \frac{cT}{n} = \frac{\lambda_0}{n} \qquad (3)$$

where λ_o = cT = wavelength in vacuum.

3.2.2 Plane-polarized light waves.

3.2.2.1 A plane-polarized light wave is one whose electric vector vibrates in a fixed plane (which we shall call the plane of polarization) in homogeneous media that do not rotate the plane of polarization. The wave illustrated in Figure 3. 1 is plane-polarized. If the direction of propagation is the Z-axis, the magnitude E (z, t) of the electric vector can be specified as the trigonometric function

$$E\ (z,\ t) = a \cos\ (knz + \phi - \omega t) \tag{4}$$

where

z = distance measured along Z
t = time
k = $2\pi/\lambda$
ω = $2\pi/T$
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.

The phase angle ϕ is needed for specifying the phase of one wave relative to another. If, for example,

$$E_1 = a_1 \cos\ (knz + \phi_1 - \omega t) \tag{5}$$

$$E_2 = a_2 \cos\ (knz + \phi_2 - \omega t) \tag{6}$$

the corresponding waves differ in phase by the amount $\phi_1 - \phi_2$ at like values of t and z.

3.2.2.2 The state of vibration or polarization is the same for all points that belong to a wavefront. On each wavefront

$$knz + \phi - \omega t = \text{constant} = w \tag{7}$$

where w is different for each wavefront. The wavefront moves so as to satisfy Equation (7). By differentiating the members of Equation (7) with respect to the time t, one finds that

$$\frac{dz}{dt} = v = \frac{\omega}{kn} = \frac{1}{n}\,\frac{\lambda}{T} = \frac{c}{n^2};\ \frac{\lambda}{T} = \frac{c}{n} \therefore \frac{1}{n} \cdot \frac{c}{n} = \frac{c}{n^2} \tag{8}$$

3.2.2.3 The wavefronts of the plane-polarized wave described by Equation (4) are perpendicular to the Z-axis, the direction of propagation. If the plane-polarized plane wave is propagated along an arbitrary direction OP, Figure 3.2, the magnitude E of the electric vector assumes the form

$$E = a \cos \left[kn\ (px + qy + rz) + \phi - \omega t \right] \tag{9}$$

where p, q and r are the direction cosines of OP with respect to X, Y, and Z, respectively. Thus,

$$p^2 + q^2 + r^2 = 1. \tag{10}$$

Equation (9) reduces to Equation (4) when the direction of propagation OP is the Z-direction only, for then $p = q = o$ and $r = 1$. It is important to observe that the wave motion of Equations (4) and (9) is of the form

$$E = a \cos\ (\Phi - \omega t) \tag{11}$$

where

$$\Phi = kn\ (px + qy + rz) + \phi \tag{12}$$

with p, q, and r defined as the direction cosines of the direction of propagation of the plane-polarized, plane wave. The electric vector vibrates in the wavefront.

3.2.3 Energy in a single wave. The instantaneous energy, W_i (whether energy flux or energy density) in the wave is proportional to E^2, where E denotes the instantaneous magnitude of the electric vector. We take the factor of proportionality as unity and write from Equation (11)

$$W_i = E^2 = a^2 \cos^2 (\Phi - \omega t). \tag{13}$$

The oscillations of light waves are so rapid that the eye or other known detectors are unable to follow the in-

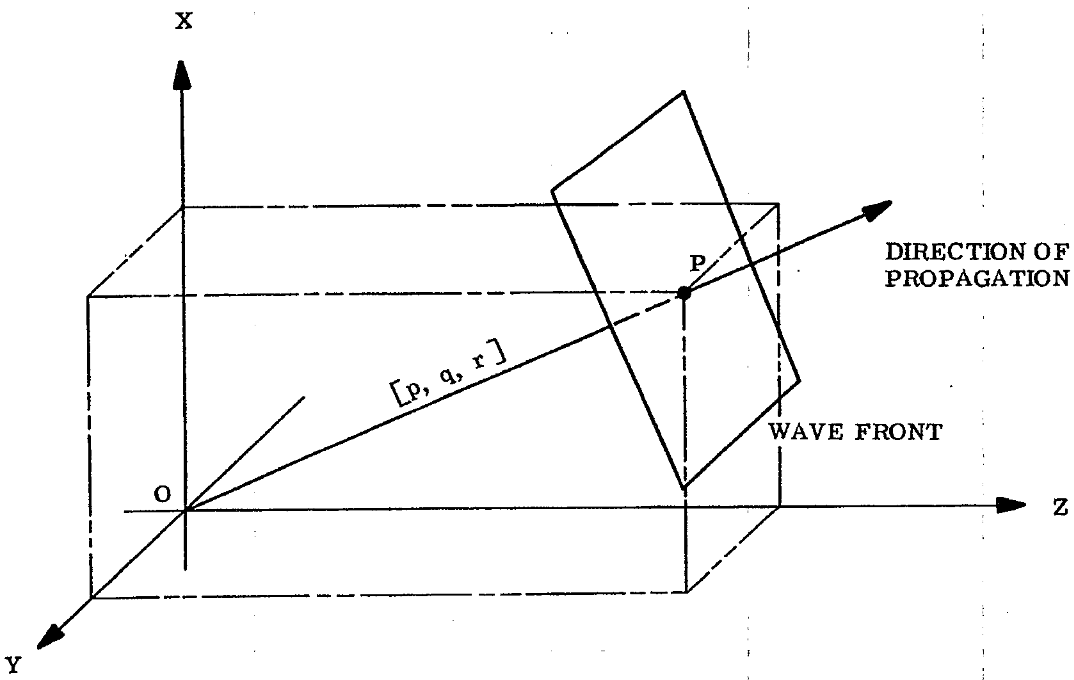

Figure 3. 2—Notation with respect to the propagation of a plane wave.

stantaneous values. Rather, the time average W of W_i is detected and measured. It suffices to average over one period T of oscillation. Thus,

$$W = \frac{1}{T} \int_o^T a^2 \cos^2 (\Phi - \omega t) \, dt$$

$$= \frac{a^2}{2T} \int_o^T \left[1 + \cos 2 (\Phi - \omega t) \right] dt. \tag{14}$$

Since $\omega = 2\pi/T$, it follows almost directly that

$$\int_o^T \cos 2 (\Phi - \omega t) \, dt = 0. \tag{15}$$

Hence,

$$W = a^2/2, \tag{16}$$

i.e. the time-averaged energy density or energy flux in a single wave is proportional to the square of its amplitude. W is independent of, for example, the phase angle ϕ of the single plane wave.

3.3 INTERFERENCE BETWEEN WAVES

3.3.1 Collinear, coherent waves.

3.3.1.1 Two waves will be called collinear when they are propagated in the same direction. We consider the interference of two plane-polarized,* plane waves that are propagated in the same direction with a constant phase

* The electric vectors of these two plane polarized waves are assumed parallel, i.e., are assumed to vibrate in the same fixed plane.

difference δ . The magnitudes E_1 and E_2 of the electric vectors of two unlike plane waves assume from Equation (11) the form

$$E_1 = a_1 \cos (\Phi_1 - \omega t) ; \qquad E_2 = a_2 (\Phi_2 - \omega t) . \tag{17}$$

From Equation (12)

$$\Phi_1 - \Phi_2 = \delta , \tag{18}$$

the phase difference between the two waves.

3.3.1.2 Let E denote the magnitude of the electric vector formed by the sum of E_1 and E_2, i.e., formed by the interference of the two waves. Then,

$$E = a_1 \cos (\Phi_1 - \omega t) + a_2 \cos (\Phi_2 - \omega t) . \tag{19}$$

Let W be the time-averaged energy density formed by the two interfering waves. As in paragraph 3.2.3,

$$\begin{aligned} W &= \frac{1}{T} \int_0^T E^2 \, dt . \\ &= \frac{a_1^2}{T} \int_0^T \cos^2 (\Phi_1 - \omega t) \, dt + \frac{a_2^2}{T} \int_0^T \cos^2 (\Phi_2 - \omega t) \, dt \\ &\quad + \frac{2 a_1 a_2}{T} \int_0^T \cos (\Phi_1 - \omega t) \cos (\Phi_2 - \omega t) \, dt \\ &= \frac{a_1^2}{2} + \frac{a_2^2}{2} + 2 a_1 a_2 I \end{aligned} \tag{20}$$

where

$$I = \frac{1}{T} \int_0^T \cos (\Phi_1 - \omega t) \cos (\Phi_2 - \omega t) \, dt . \tag{21}$$

But

$$\cos (\Phi_1 - \omega t) \cos (\Phi_2 - \omega t) = \frac{1}{2} \left[\cos (\Phi_1 + \Phi_2 - 2\omega t) + \cos (\phi_1 - \phi_2) \right] . \tag{22}$$

As in Equation (15),

$$\int_0^T \cos (\Phi_1 + \Phi_2 - 2\omega t) \, dt = 0 .$$

Hence,

$$I = \frac{\cos (\Phi_1 - \Phi_2)}{2T} \int_0^T dt = \frac{\cos (\Phi_1 - \Phi_2)}{2} = \frac{\cos \delta}{2} \tag{23}$$

Finally, from Equations (23) and (20) we find that the time-averaged density, W , produced by the interference of two, plane-polarized, collinear, plane waves having amplitudes a_1 and a_2 and phase difference ($\phi_1 - \phi_2$) is

$$W = \frac{1}{2} \left[a_1^2 + 2 a_1 a_2 \cos \delta + a_2^2 \right] . \tag{24}$$

3.3.1.3 For constructive interference, the phase difference $\Phi_1 - \Phi_2 = \delta$ between the two waves is 0, 2π , 4π , etc., so that

$$W = \frac{1}{2} (a_1 + a_2)^2 . \tag{25}$$

For destructive interference, $\delta = m\pi$ where m is an odd integer. Correspondingly,

$$W = \frac{1}{2} (a_1 - a_2)^2 . \tag{26}$$

It should be noted from Equation (26) that $W = 0$ when the two waves have equal amplitudes and are out of phase. Thus, two plane waves that are propagated in the same direction can cancel one another everywhere, or they can re-enforce one another everywhere provided that their phase difference δ is a suitably chosen constant. The time-averaged energy density of the resultant wave is not merely the sum of the time-averaged energy densities of the two separate waves except in the special cases $\cos \delta = 0$. See Equations (25) and (16). The waves are coherent when δ is constant.

3.3.2 Collinear, incoherent waves.

3.3.2.1 One should expect that when light or any other radiation from two independent sources overlap, the resulting energy density is simply the sum of the overlapping energy densities, i. e., the law of superposition of energy should apply. The interfering waves ought to be incoherent. The following somewhat oversimplified argument brings to bear the essential physics underlying the interference of incoherent waves.

3.3.2.2 The time-averaged energy density, produced by two interfering waves that have amplitudes a_1 and a_2 and the phase difference δ , is given by Equation (24). We shall avoid considering the sum of a large number of waves having randomly distributed phase differences δ (as will occur with independent sources) by supposing that in a short interval of time the phase differences δ between the two interfering waves are distributed with equal probability in the interval $0 \leqq \delta \leqq 2\pi$. Then from Equation (24)

$$W = \frac{1}{2} \left[a_1^2 + 2\, a_1 a_2 \overline{\cos \delta} + a_2^2 \right] \tag{27}$$

where $\overline{\cos \delta}$ is the average value of $\cos \delta$ when all values of δ are equally probable in the interval $0 \leqq \delta \leqq 2\pi$. One can show that

$$\overline{\cos \delta} = 0 \ . \tag{28}$$

In this manner we conclude that

$$W = \frac{1}{2} \left(a_1^2 + a_2^2 \right) \tag{29}$$

so that the interference between incoherent waves is of that degenerate variety to which the law of superposition of energy applies.

3.3.3. Non-collinear, coherent waves.

3.3.3.1 The theory of paragraph 3.3 is almost but not quite adequate for explaining and interpreting the interference fringes that appear in Twyman Green and other double-beam interferometers; for in these interferometers the mirrors are usually tilted so that the two interfering waves are not propagated in the same direction. It is well known that a series of straight and parallel interference fringes are seen when the interfering waves are not collinear and when the reflecting surfaces are optical flats.

3.3.3.2 We may suppose without essential loss of generality that one wave is propagated along the direction OP that makes any angle θ with Z but is oriented so that the direction cosine $q = 0$. The two interfering waves are described by Equation (17); but $\Phi_1 - \Phi_2$ will not be given by Equation (18). Instead,

$$\begin{aligned} \Phi_1 &= knz + \phi_1 \\ \Phi_2 &= kn \left(x \sin \theta + z \cos \theta \right) + \phi_2 \end{aligned} \tag{30}$$

so that

$$\Phi_1 - \Phi_2 = \phi_1 - \phi_2 - knx \sin \theta + knz \left(1 - \cos \theta \right) \ . \tag{31}$$

From Equations (20) and (23) the time-averaged energy density formed by the two interfering, coherent waves is

$$W = \frac{1}{2} \left[a_1^2 + 2\, a_1 a_2 \cos \left(\Phi_1 - \Phi_2 \right) + a_2^2 \right] . \tag{32}$$

Substituting $\Phi_1 - \Phi_2$ from Equation (31) and setting $\phi_1 - \phi_2 = \delta$, the fixed phase difference between the two waves, one obtains

$$2W = a_1^2 + a_2^2 + 2\, a_1 a_2 \cos \left[\delta - knx \sin \theta + knz \left(1 - \cos \theta \right) \right] \tag{33}$$

in which θ is the angle indicated in Figure 3.3, $k = 2\pi/\lambda$ and n is the refractive index of the medium. δ is the phase difference between the two interfering waves having amplitude a_1 and a_2 at the point $x = 0$, $z = 0$.

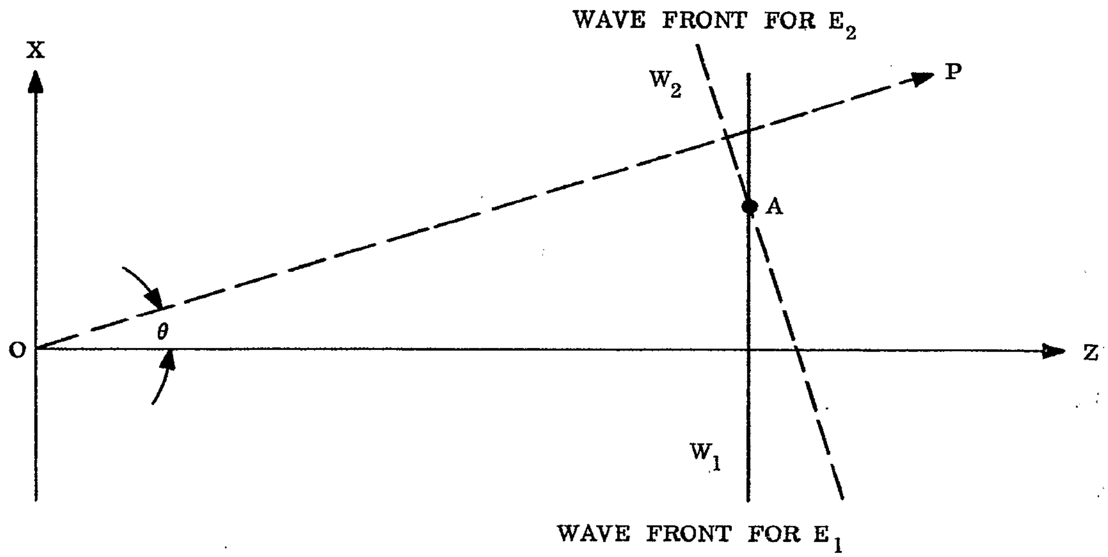

Figure 3.3— Interference between two plane wavefronts W_1 and W_2 that are propagated along different directions.

3.3.3.3 In double beam interferometry, the angle θ is usually so small that one can set $\sin\theta = \theta$ and $1 - \cos\theta = \theta^2/2$. If, then, one makes observations in planes z near z = 0, Figure 3.3, the term containing z in Equation (30) can be neglected. The approximation thus obtained is the usual interference formula

$$2W = a_1^2 + a_2^2 + 2a_1a_2\cos(\delta - 2\pi nx\theta/\lambda) . \tag{34}$$

The fringes are repeated whenever x is increased by an amount Δx such that

$$kn\Delta x \sin\theta = 2\pi .$$

The fringe width h is therefore given by

$$h = \Delta x = \frac{\lambda}{n \sin\theta} . \tag{35}$$

The greater fringe widths belong to the longer wavelengths.

3.3.3.4 In case the fringes are photographed with a camera that images a plane into a plane, the interference fringes will be straight. Suppose, however, that the camera has curvature of field. In this case a plane z = constant will not be focused upon the photographic plate. Consequently, one has to expect from Equation (33) that the photographed fringes will be curved and that the distortion of the fringes should increase as θ and z are increased.