

CHAPTER 1

Basic Wavefront Aberration Theory
for Optical Metrology

JAMES C. WYANT

Optical Sciences Center, University of Arizona

and

WYKO Corporation, Tucson, Arizona

KATHERINE CREATH

*Optical Sciences Center**University of Arizona, Tucson, Arizona*

| | |
|--|----|
| I. Sign Conventions | 2 |
| II. Aberration-Free Image | 4 |
| III. Spherical Wavefront, Defocus, and Lateral Shift | 9 |
| IV. Angular, Transverse, and Longitudinal Aberration | 12 |
| V. Seidel Aberrations | 15 |
| A. Spherical Aberration ($\Delta W = W_{040}\rho^4 = W_{040}(x^2 + y^2)^2$) | 18 |
| B. Coma ($\Delta W = W_{131}x_0\rho^3 \cos \theta = W_{131}x_0x(x^2 + y^2)$) | 22 |
| C. Astigmatism ($\Delta W = W_{222}x_0^2\rho^2 \cos^2 \theta = W_{222}x_0^2x^2$) | 24 |
| D. Field Curvature ($\Delta W = W_{220}x_0^2\rho^2 = W_{220}x_0^2(x^2 + y^2)$) | 26 |
| E. Distortion ($\Delta W = W_{311}x_0^3\rho \cos \theta = W_{311}x_0^3x$) | 28 |
| VI. Zernike Polynomials | 28 |
| VII. Relationship between Zernike Polynomials and Third-Order Aberrations | 35 |
| VIII. Peak-to-Valley and RMS Wavefront Aberration | 36 |
| IX. Strehl Ratio | 38 |
| X. Chromatic Aberrations | 40 |
| XI. Aberrations Introduced by Plane Parallel Plates | 40 |
| XII. Aberrations of Simple Thin Lenses | 46 |
| XIII. Conics | 48 |
| A. Basic Properties | 48 |
| B. Spherical Aberration | 50 |
| C. Coma | 51 |
| D. Astigmatism | 52 |
| XIV. General Aspheres | 52 |
| References | 53 |

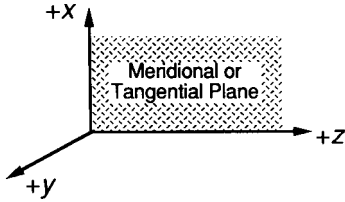


FIG. 1. Coordinate system.

The principal purpose of optical metrology is to determine the aberrations present in an optical component or an optical system. To study optical metrology, the forms of aberrations that might be present need to be understood. The purpose of this chapter is to give a brief introduction to aberrations in an optical system. The goal is to provide enough information to help the reader interpret test results and set up optical tests so that the accessory optical components in the test system will introduce a minimum amount of aberration. To obtain a more detailed description of aberrations, the reader should see the references given at the end of the chapter.

I. SIGN CONVENTIONS

The coordinate system is defined as **right-handed coordinate** and shown in Fig. 1. The z axis is the direction of propagation, the x axis is the meridional or tangential direction, and the y axis is the sagittal direction. This definition for the meridional plane is not universally accepted. It is common to use the y axis as the meridional plane; however, using the x axis as the meridional plane makes the mathematics in this chapter more consistent.

The optical path difference (OPD) is defined as the difference between the aberrated and the ideal unaberrated wavefronts. The **OPD is positive if the aberrated wavefront leads the ideal unaberrated wavefront** as shown in Fig. 2. Also, if the aberrated wavefront curves in more than the unaberrated wavefront, the OPD is positive. Therefore, **a negative focal shift will introduce a positive aberration**, so that **a positive aberration will focus in front of the**

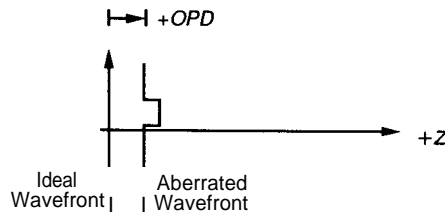


FIG. 2. Definition of OPD.

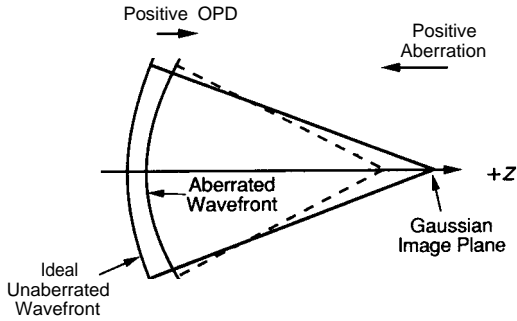
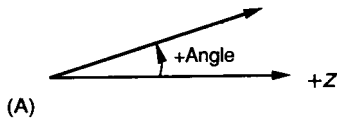


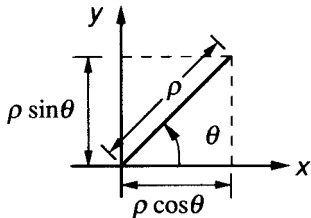
FIG. 3. OPD for positive aberration.

Gaussian image plane, as shown in Fig. 3. In the case of interferometric optical testing, this means that a bump on a test mirror (aberrated beam) will be represented by a bump in the OPD.

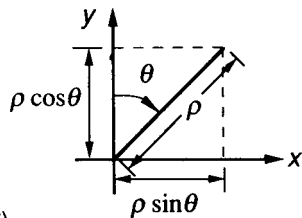
Angles are defined to be positive if they correspond to a positive slope. In the x - z or y - z plane, a positive angle is defined relative to the z axis in a counterclockwise direction, as shown in Fig. 4a. In the pupil plane, angles are



(A)



(B)



(C)

FIG. 4. Angle conventions. (a) Angles in xz (meridional plane). (b) Relationship between rectangular and polar coordinates as used in this chapter.

defined as shown in Fig. 4b. Polar coordinates are used for most of the aberration formulas, because the Seidel aberrations and Zernike coefficients are circularly symmetric. The relationship between rectangular and polar coordinates as used in this chapter is shown in Fig. 4b. **Angles increase in the counterclockwise direction, -and the zero angle is along the x axis when looking at the pupil plane from the image plane.** This is the definition used for interferogram analysis by many of the computer-aided lens design and wavefront analysis programs. Another common definition that will not be used in this chapter is shown in Fig. 4c. With this definition, the zero angle is along the y axis rather than the x axis, and positive angles are in a clockwise direction in the pupil plane. This definition is used in aberration theory (for an example, see Welford's (1986) book), and is the convention most likely used in the classroom. We have chosen the definition used in computer programs for interferogram analysis (Fig. 4b) so that the expressions for the different aberrations can be compared to those results.

Tilt in a wavefront affects the image by causing a shift of its center location in the Gaussian image plane. **A tilt causing a positive OPD change in the +x direction will cause the image to shift in the -x direction, as shown in Fig. 5.**

II. ABERRATION-FREE IMAGE

Before looking at aberrations, and the effect of aberrations on images, we will look at the image (point spread **function**) that would be obtained of a **"point" object if no aberration were present.** The image, which we will call an **aberration-free image**, is more commonly called a **diffraction-limited image.** We will avoid the use of the term diffraction-limited image when referring to an image formed in the absence of aberration because strictly speaking any image is diffraction limited, regardless of how much aberration is present in the optical system used to form the image.

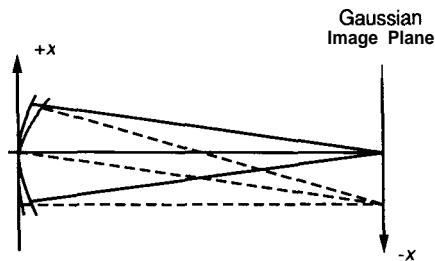


FIG. 5. Wavefront tilt causing an image shift.

For a uniformly illuminated circular aperture the **power per unit solid angle** in the image plane can be expressed in terms of the Bessel function of the first order $J_1[.]$ as

$$I(\alpha) = \frac{E_A(\pi d^2/4)^2}{\lambda^2} \left[\frac{2J_1(\pi d\alpha/\lambda)}{\pi d\alpha/\lambda} \right]^2, \tag{1}$$

where

d = exit pupil diameter,

α = angle between observation point and center of diffraction pattern as measured from the center of the circular aperture,

λ = wavelength of radiation,

E_A = irradiance (power/area) incident upon the aperture, and

$E_A(\pi d^2/4)$ = total power incident upon aperture.

Instead of expressing the distribution of the aberration-free diffraction pattern in units of power per unit solid angle, it is often more convenient to express the distribution in units of **power per unit area**. For a lens whose exit pupil diameter is d and whose image plane is a distance L from the exit pupil, the power per unit area (irradiance) of the diffraction pattern is given by

$$E(r) = \frac{E_A(\pi d^2/4)^2}{\lambda^2 L^2} \left\{ \frac{2J_1[\pi dr/(\lambda L)]}{\pi dr/(\lambda L)} \right\}^2, \tag{2}$$

where r is the radial distance from the center of the diffraction pattern to the observation point.

The effective F-number of the lens $f \#$ can be defined as L/d . With this definition, Eq. (2) becomes

$$E(r) = \frac{E_A \pi^2 d^2}{16 \lambda^2 (f \#)^2} \left\{ \frac{2J_1[\pi r/(\lambda f \#)]}{\pi r/(\lambda f \#)} \right\}^2. \tag{3}$$

Figure 6 is a plot of the normalized irradiance for the diffraction pattern of a

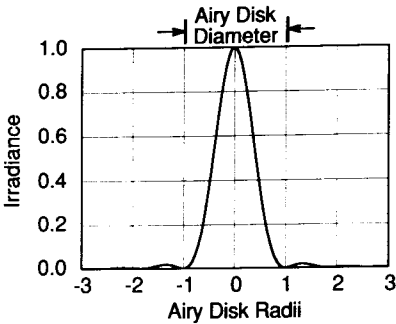


FIG. 6. Normalized irradiance of diffraction pattern of uniformly illuminated circular aperture.

uniformly illuminated circular aperture. This diffraction pattern is called an Airy disk. The distance between the two zero intensity regions on each side of the central maximum is called the Airy disk diameter. For a lens illuminated with collimated light, this Airy disk diameter is given by $2.44 \lambda f \#$. For visible light, λ is on the order of $0.5 \mu\text{m}$ and the Airy disk diameter $2.44 \lambda f \#$ is approximately equal to the effective F-number expressed in micrometers. For example, in visible light, the Airy disk diameter for an $f/5$ lens is approximately $5 \mu\text{m}$.

Table I gives the distribution of energy in the diffraction pattern at the focus of a perfect lens. Nearly 84% of the total energy is contained within the central ring. The fraction of the total energy contained in a circle of radius r about the diffraction pattern center is given by

$$\text{Fraction of encircled energy} = 1 - J_0^2\left(\frac{\pi r}{\lambda f \#}\right) - J_1^2\left(\frac{\pi r}{\lambda f \#}\right). \quad (4)$$

$J_0(\cdot)$ is the Bessel function of zero order while $J_1(\cdot)$ is the Bessel function of first order. Equation (4) is plotted in Fig. 7.

Generally a mirror system will have a central obscuration. If ε is the ratio of the diameter of the central obscuration to the mirror diameter d , and if the entire circular mirror of diameter d is uniformly illuminated, the power per unit solid angle is given by

$$I(\mathbf{R}) = \frac{E_A(\pi d^2/4)^2}{\lambda^2} \left\{ \frac{2J_1(1.22\pi R)}{1.22\pi R} - \varepsilon^2 \left[\frac{2J_1(1.22\pi\varepsilon R)}{1.22\pi\varepsilon R} \right] \right\}^2, \quad (5)$$

where \mathbf{R} is given in units of the Airy disk radius for an unobstructed aperture of equal diameter. The other quantities are the same as defined above.

TABLE I
DISTRIBUTION OF ENERGY IN THE DIFFRACTION PATTERN AT THE
Focus OF A PERFECT LENS

| Ring | $r/(\lambda f \#)$ | Peak Illumination | Energy in Ring (%) |
|-----------------|--------------------|----------------------|-----------------------|
| Central max | 0 | 1 | 83.9 |
| 1st dark ring | 1.22 | 0 | |
| 1st bright ring | 1.64 | 0.017 | 7.1 |
| 2nd dark | 2.24 | 0 | |
| 2nd bright | 2.66 | 0.0041 | 2.8 |
| 3rd dark | 3.24 | 0 | |
| 3rd bright | 3.70 | 0.0016 | 1.5 |
| 4th dark | 4.24 | 0 | |
| 4th bright | 4.72 | 0.00078 | 1.0 |
| 5th dark | 5.24 | 0 | |

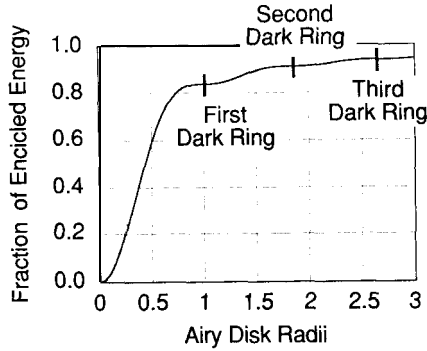


FIG. 7. Encircled energy for aberration-free system.

Equation (5) shows that the peak intensity of the diffraction pattern goes as $(1 - \epsilon^2)^2$ and the diameter of the central maximum of the diffraction pattern shrinks as the obscuration ratio ϵ increases. If β is the distance to the first zero in units of the Airy disk radius, β is obtained by solving the equation

$$\epsilon J_1(1.22\pi\epsilon\beta) = J_1(1.22\pi\beta). \quad (6)$$

Figure 8 gives a plot of the distance to the first zero in the diffraction pattern in units of the Airy disk radius as a function of obscuration ratio.

As the obscuration ratio increases, and the size of the central core of the diffraction pattern decreases, less light is contained within the central core. Figure 9 gives a plot of the encircled energy distribution as a function of obscuration ratio. The curves are normalized so that for each case the same total energy is transmitted through the aperture.

In modern optical metrology, it is common for the optical system under test to be illuminated with a laser beam having a Gaussian intensity distribution. If a lens of diameter d and obscuration ratio ϵ is illuminated with

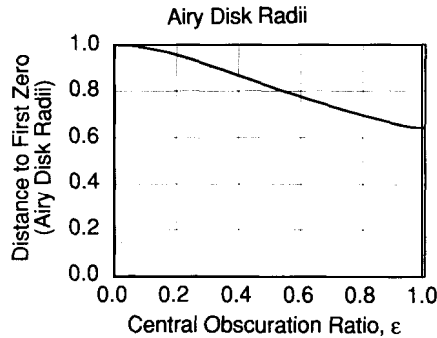


FIG. 8. Position of first zero of diffraction pattern as a function of obscuration ratio.

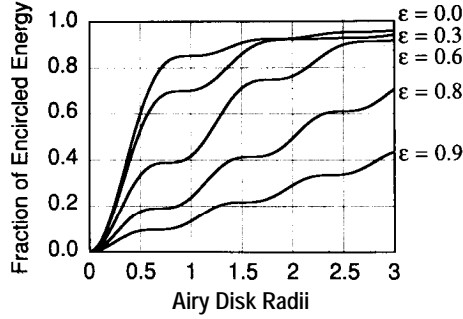


FIG. 9. Encircled energy for aberration-free system and different obscuration ratios ϵ .

a collimated Gaussian beam of total power P , the irradiance of the diffraction pattern is given by

$$E(r) = \frac{P\pi d^2}{2\omega^2 \lambda^2 (f\#)^2} \left| \int_{\epsilon}^1 e^{-\rho^2(d/2\omega)^2} J_0 \left[\left(\frac{\pi r}{\lambda f\#} \right) \rho \right] \rho d\rho \right|^2, \quad (7)$$

where $J_0[.]$ is a zero order Bessel function and ω is the distance from the axis to the $1/e^2$ intensity point.

For comparison purposes, the irradiance of the diffraction pattern of a circular aperture illuminated with a uniform beam of total power P contained within a circular aperture of diameter d can also be written in terms of $J_0[.]$ as

$$E(r) = \frac{P\pi}{\lambda^2 (f\#)^2} \left| \int_{\epsilon}^1 J_0 \left[\left(\frac{\pi r}{\lambda f\#} \right) \rho \right] \rho d\rho \right|^2. \quad (8)$$

It follows from Eqs. (7) and (8) that, for an unobscured aperture, the ratio of the peak irradiance of the diffraction pattern of the Gaussian beam to the

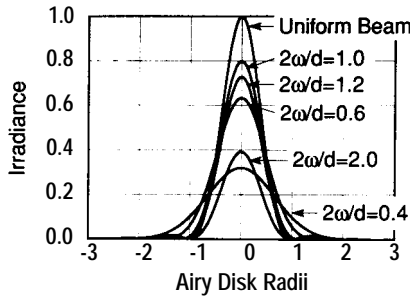
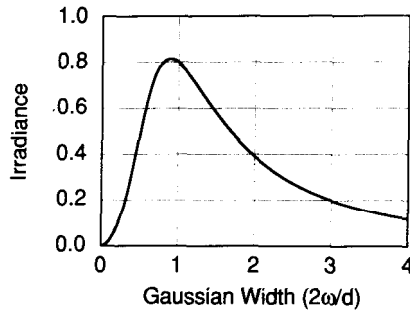


FIG. 10. Diffraction pattern of a circular aperture for different Gaussian widths.



11. Ratio of Gaussian and uniform peak irradiances versus Gaussian width.

peak irradiance of the diffraction pattern of the uniform beam is given by

$$\frac{E_G}{E_u} = \frac{2[1 - e^{-(d/2w)^2}]^2}{(d/2w)^2}. \quad (9)$$

Figure 10 shows plots of Eq. (7) for different Gaussian widths $2w/d$ as well as a plot of Eq. (8) for $\epsilon = 0$. It is interesting to note that if the $1/e^2$ intensity point falls at the edge of the aperture ($d/2w = 1$), the intensity of the diffraction pattern drops to $1/e^2$ of its maximum value at a radius approximately equal to $3/4$ of the Airy disk radius. Figure 11 is a plot of Eq. (9) for different values of $2w/d$.

III. SPHERICAL WAVEFRONT, DEFOCUS, AND LATERAL SHIFT

The action of a perfect lens is to produce in its exit pupil a spherical wave whose center of curvature coincides with the origin of the (x_0, y_0, z_0) coordinate system, as shown in Fig. 12. If R is the radius of curvature of the spherical wavefront and the center of the exit pupil is at the origin of the (x, y, z) coordinate system, then

$$x = x_0,$$

$$y = y_0,$$

and

$$z = z_0 - R. \quad (10)$$

The equation of the spherical wavefront in the (x, y, z) system is thus

$$x^2 + y^2 + (z - R)^2 = R^2. \quad (11)$$

If x and y are small compared to R (i.e., if the size of the exit pupil is small

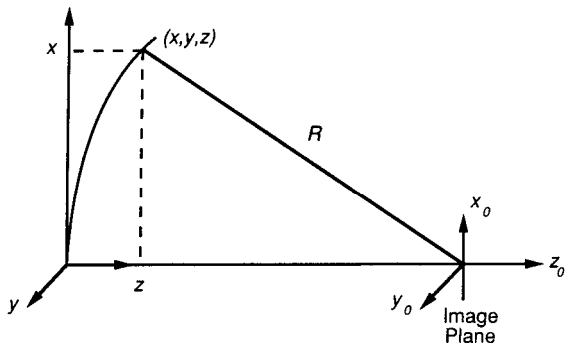


FIG. 12. Relation between image and exit pupil coordinates.

compared to the radius of curvature of the emergent wave), and z is sufficiently small so that the term z^2 can be neglected, the equation of the spherical wavefront in the exit pupil can be approximated by a parabola

$$z = \frac{x^2 + y^2}{2R}. \quad (12)$$

Therefore, a wavefront distribution (optical path difference) of

$$W(x, y) = \frac{x^2 + y^2}{2R} \quad (13)$$

in the exit pupil of a lens represents a spherical wavefront converging to a point a distance R from the exit pupil.

Often in optical testing, we are given a wavefront distribution in the exit pupil and must find the effect of such a distribution in a plane other than the paraxial image plane. For example, Fig. 13 shows a situation where the

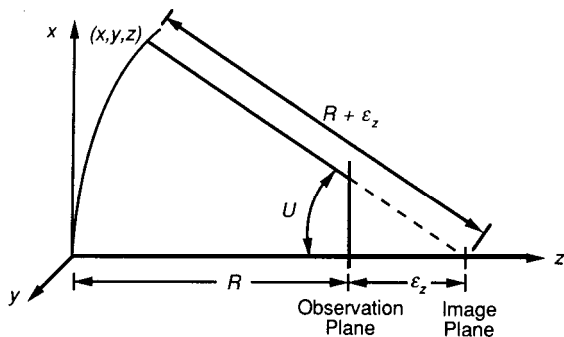


FIG. 13. Focus shift of a spherical wavefront.

observation plane is a distance R from the exit pupil, but the wavefront in the exit pupil has a radius of curvature $R + \varepsilon_z$. That is,

$$W(x, y) = \frac{x^2 + y^2}{2(R + \varepsilon_z)} = \frac{x^2 + y^2}{2R} \left[\frac{1}{1 + \varepsilon_z/R} \right]. \quad (14)$$

If ε_z is much smaller than R , Eq. (14) can be written as

$$W(x, y) = \frac{x^2 + y^2}{2R} - \varepsilon_z \frac{x^2 + y^2}{2R^2}. \quad (15)$$

Therefore, if a defocus term with OPD of $A(x^2 + y^2)$ is added to a spherical wave of radius of curvature R , a focal shift results, given by

$$\varepsilon_z = -2R^2 A. \quad (16)$$

Note that for the sign convention chosen, increasing the OPD moves the focus toward the exit pupil in the negative z direction.

From another viewpoint, if the image plane is shifted along the optical axis toward the lens an amount ε_z (ε_z is negative), a change in the wavefront relative to the original spherical wavefront given by

$$\Delta W(x, y) = -\varepsilon_z \frac{x^2 + y^2}{2R^2} \quad (17)$$

is required to change the focus position by this amount. If U is the maximum half-cone angle of the converging beam having an F-number $f \#$, the defocus changes the OPD at the edge of the pupil an amount

$$\Delta W_{\text{defocus}} = -\frac{1}{2} \varepsilon_z \sin^2 U = -\frac{1}{2} \varepsilon_z (\text{NA})^2 = -\frac{\varepsilon_z}{8(f \#)^2}, \quad (18)$$

where NA is the numerical aperture.

By use of the Rayleigh criterion, it is possible to establish a rough allowance for a tolerable depth of focus. Setting the OPD, due to defocus, at the edge of the pupil equal to $\lambda/4$ yields

$$\Delta W_{\text{defocus}} = \pm \frac{\lambda}{4} = \pm \frac{\varepsilon_z}{8(f \#)^2}, \quad (19)$$

or

$$\varepsilon_z = \pm 2\lambda(f \#)^2.$$

For visible light $\lambda \cong 0.5 \mu\text{m}$. Hence, ε_z (in micrometers) $\cong \pm (f \#)^2$.

Suppose that instead of being shifted along the optical axis, the center of the sphere is moved along the x_0 axis an amount ε_x as shown in Fig. 14. The equation of the spherical wavefront in the exit pupil can then be written as

$$y^2 + (x - \varepsilon_x)^2 + (z - R)^2 = R^2. \quad (20)$$

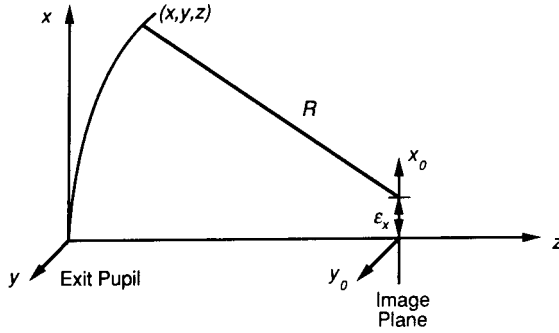


FIG. 14. Lateral shift of image position.

Assuming that ϵ_x is sufficiently small so that the term ϵ_x^2 can be neglected, and if z^2 is neglected as before, the wavefront can be written as

$$W(x, y) = \frac{x^2 + y^2}{2R} - \frac{x\epsilon_x}{R}. \quad (21)$$

Thus, an OPD term B times x in the wavefront distribution in the exit pupil represents a transverse shift in the position of the focus by an amount

$$\epsilon_x = -RB. \quad (22)$$

Likewise, a term C times y in the wavefront distribution in the exit pupil represents a shift of the focus along the y, axis by an amount

$$\epsilon_y = -RC. \quad (23)$$

Combining the results of the preceding analyses, it is seen that a wavefront distribution in the exit pupil of an optical system

$$W(x, y) = \frac{x^2 + y^2}{2R} - \epsilon_z \frac{x^2 + y^2}{2R^2} - \epsilon_x \frac{x}{R} - \epsilon_y \frac{y}{R} \quad (24)$$

represents a spherical wave whose center of curvature is located at the point $(\epsilon_x, \epsilon_y, \epsilon_z)$ in an image plane coordinate system (x_0, y_0, z_0) whose origin is located a distance R from the vertex of the exit pupil.

IV. ANGULAR, TRANSVERSE, AND LONGITUDINAL ABERRATION

In general, the wavefront in the exit pupil is not a perfect sphere, but is an aberrated sphere, so [different parts of the wavefront come to focus in different places](#). It is often desirable to know where these focus points are located; i.e.,

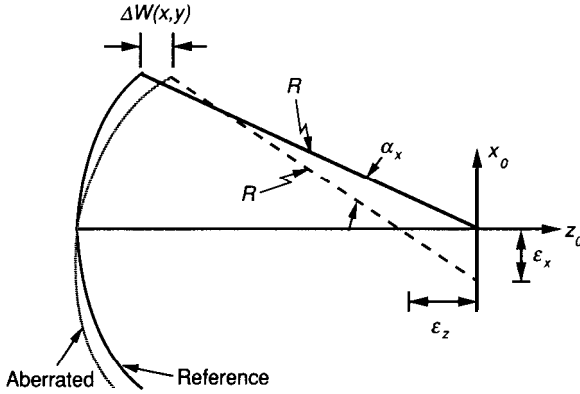


FIG. 15. Transverse and longitudinal aberrations. $\Delta W(x, y)$ = distance between reference wavefront and aberrated wavefront, ϵ_x = transverse or lateral aberration, ϵ_z = longitudinal aberration.

find $(\epsilon_x, \epsilon_y, \epsilon_z)$ as a function of (x, y) given some $W(x, y)$ that is not necessarily a spherical wavefront. This is depicted in Fig. 15. The figure shows a spherical reference wavefront having a focus at the origin of the coordinate system (x_0, y_0, z_0) and an aberrated wavefront that may not have a unique focus, but that has rays in the area enclosed by the small circle that intersects the paraxial image plane at the point $x_0 = -\epsilon_x$. We assume that the angle these rays make with the optical axis is sufficiently small so that we can approximate its sine by the angle itself, and the cosine of the angle by unity. The OPD $\Delta W(x, y)$ is the distance between the reference wavefront and the aberrated wavefront. It is assumed that $\Delta W(x, y)$ is sufficiently small so that the angle α_x between the aberrated wavefront and the reference wavefront is also small, in which case α_x , called the **angular aberration** of the ray, is given by

$$\alpha_x = \frac{-\partial \Delta W(x, y)}{n \partial x}. \quad (25)$$

In air, the refractive index n can be set equal to 1. The distance ϵ_x called the transverse aberration of the ray, is given by

$$\epsilon_x = R\alpha_x = -R \frac{\partial \Delta W(x, y)}{\partial x}, \quad (26)$$

where the refractive index has been set equal to 1. Similar formulas hold for the y components of the angular and transverse aberrations.

Sometimes it is useful to know where an aberrated ray intersects the

optical axis. Referring to Fig. 15, which is an exaggerated drawing, it is seen that

$$\frac{\varepsilon_z}{\varepsilon_x} \approx \frac{R}{x - \varepsilon_x}. \quad (27)$$

Since $\varepsilon_x \ll x$, the longitudinal aberration can be written as

$$\varepsilon_z \approx \frac{R}{x} \varepsilon_x = -\frac{R^2}{x} \frac{\partial \Delta W(x, y)}{\partial x}. \quad (28)$$

The quantity ε_z is called the longitudinal aberration of the ray.

Sometimes in writing the wavefront aberration, the exit pupil coordinates are expressed in normalized coordinates such that $(x^2 + y^2)^{1/2} = 1$ at the edge of the exit pupil. In this case, the transverse and longitudinal aberrations become

$$\varepsilon_x = -\frac{R}{h} \frac{\partial \Delta W}{\partial x} \quad (29)$$

and

$$\varepsilon_z = -\frac{R^2}{xh^2} \frac{\partial \Delta W}{\partial x},$$

where h is the geometrical pupil radius.

In practice, the transverse aberrations for individual rays through an optical system are often determined by tracing the individual rays through the system and finding the intersection of the rays with the image plane. The result is a spot diagram that gives a useful impression of the geometrical image quality.

In many testing systems, either the transverse or the longitudinal aberration is measured, and from this measurement the wavefront aberration is to be determined. In the case where the aberration is symmetric about the optical axis, only rays in the x - z meridional plane need be determined, and $\Delta W = \Delta W(x)$. Hence, in normalized coordinates,

$$\Delta W = \Delta W(x') = -\frac{h}{R} \int_0^{x'} \varepsilon_x dx,$$

or

$$\Delta W = -\frac{h^2}{R^2} \int_0^{x'} x \varepsilon_z dx. \quad (30)$$

A convenient unit for wavefront tilt is [waves/radius](#). Consider the wavefront shown in Fig. 16. A tangent drawn to the curve at a radius of 0.75

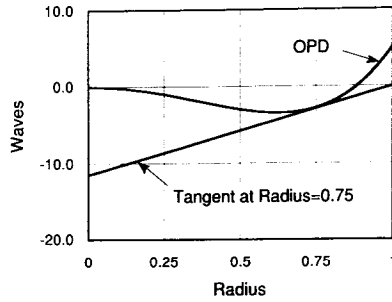


FIG. 16. Wavefront with tangent drawn parallel to 0.75 radius point.

goes from an OPD of - 11.5 waves to 0 waves. Thus, if the wavefront had this same slope across the entire radius, the OPD would be 11.5 waves and therefore the wavefront slope at this point is 11.5 waves/radius. Figure 17 shows a plot of both a wavefront in terms of OPD and its slope in terms of waves/radius.

V. SEIDEL ABERRATIONS

The particular form of the wavefront emerging from a real lens can be exceedingly complex, since it is generally the result of a number of random errors in the design, fabrication, and assembly of the lens. Nevertheless, lenses that are well made and carefully assembled do possess certain aberrations that are inherent in their design. These aberrations give rise to well defined test patterns, and it is conventional to refer to such test patterns as evidence of the existence of a certain aberration, regardless of whether the particular pattern was caused by an inherent defect of the lens or a circumstantial

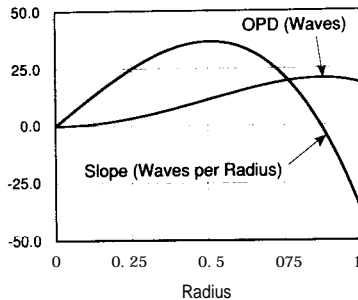


FIG. 17. OPD and slope of a general wavefront

combination of manufacturing errors. The test patterns caused by the basic aberrations of lenses are thus of considerable importance. In this section, the general characteristics of the primary monochromatic aberrations of rotationally symmetrical optical systems are reviewed. To describe the aberrations of such a system, we specify the shape of the wavefront emerging from the exit pupil. For each object point, there will be a quasi-spherical wavefront converging toward the paraxial image point. A particular image point is specified by giving the paraxial image coordinates (x_0, y_0) , as shown in Fig. 12. The wavefront can be expanded as a power series in the four variables (x, y) (the exit pupil coordinates) and (x_0, y_0) (the coordinates of the image point). Due to rotational symmetry, the wavefront distribution W must not change if there is a rigid rotation of the x_0, y_0 and x, y axes about the z axis. It is possible to select the coordinate system such that the image point lies in the plane containing the x and z axes so y_0 is identically equal to zero. Hence, W can be expanded as

$$\begin{aligned}
 W(x, y, x_0) &\equiv W(x^2 + y^2, xx_0, x_0^2) \\
 &= a_1(x^2 + y^2) + a_2xx_0 + a_3x_0^2 + b_1(x^2 + y^2)^2 \\
 &\quad + b_2xx_0(x^2 + y^2) + b_3x^2x_0^2 + b_4x_0^2(x^2 + y^2) \\
 &\quad + b_5xx_0^3 + b_6x_0^4 + \dots \\
 &\quad + \text{fifth and higher order terms.}
 \end{aligned} \tag{31}$$

(See Welford (1974), p. 87, or Welford (1986), p. 107 for more details of the justification of writing W in the form given in Eq. (31).)

As described in Section III, the **first term** of Eq. (31) represents what is commonly called **defocus**, a longitudinal shift of the center of the reference sphere, while the **second term** represents a transverse shift of the center of the reference sphere, commonly called **tilt**. The **third term** gives rise to a **phase shift** that is constant across the exit pupil. It does not affect the shape of the wavefront, and consequently has no effect on the image. When considering monochromatic light, these three terms normally have coefficients equal to zero. However, these terms have nonzero coefficients for broadband illumination, and give rise to the chromatic aberrations of Section X.

The six terms with coefficients b_1 to b_6 are of fourth degree in variables x, y , and x_0 when expressed as wavefront aberrations, and are of third degree when expressed as transverse ray aberrations. Because of this, these terms are either known as fourth-order or third-order aberrations. Since they are most commonly known as third-order aberrations, we will be referring to them as third-order aberrations. Succeeding groups of higher order terms are fifth- and seventh-order aberrations. The first five third-order aberrations are often

called Seidel aberrations after L. Seidel, who in 1856 gave explicit formulae for calculating them.

In discussing the Seidel aberrations, the exit pupil coordinates will be expressed in polar coordinates (see Fig. 4b), so that

$$x = \rho \cos \theta \quad \text{and} \quad y = \rho \sin \theta. \quad (32)$$

The radial coordinate ρ is usually normalized so that it is equal to 1 at the edge of the exit pupil. The field coordinate x_0 is also usually normalized to be equal to 1 at the maximum field position. Normalized coordinates will be used throughout the remainder of this chapter. Using polar coordinates, the wavefront expansion of Eq. (31) can be written either in terms of wavefront aberration coefficients W_{jmn} :

$$\begin{aligned} W(x_0, \rho, \theta) &= \sum_{j,m,n} W_{klm} x_0^k \rho^l \cos^m \theta, \quad \text{with } k = 2j + m, \quad l = 2n + m, \\ &= W_{200} x_0^2 + W_{111} x_0 \rho \cos \theta + W_{020} \rho^2 \\ &\quad + W_{040} \rho^4 + W_{131} x_0 \rho^3 \cos \theta + W_{222} x_0^2 \rho^2 \cos^2 \theta \\ &\quad + W_{220} x_0^2 \rho^2 + W_{311} x_0^3 \rho \cos \theta; \end{aligned} \quad (33)$$

or Seidel aberration coefficients S_i :

$$\begin{aligned} W(x_0, \rho, \theta) &= \frac{1}{8} S_I \rho^4 + \frac{1}{2} S_{II} x_0 \rho^3 \cos \theta + \frac{1}{2} S_{III} x_0^2 \rho^2 \cos^2 \theta \\ &\quad + \frac{1}{4} (S_{III} + S_{IV}) x_0^2 \rho^2 + \frac{1}{2} S_V x_0^3 \rho \cos \theta. \end{aligned} \quad (34)$$

The names of the wavefront aberration coefficients and their relationship to the Seidel aberration terms are given in Table II. Note that piston, tilt, and

TABLE II
WAVEFRONT ABERRATION COEFFICIENTS AND RELATIONSHIP TO
SEIDEL ABERRATIONS

| Wavefront Aberration Coefficient | Seidel Aberration Coefficient | Functional Form | Name |
|--|-------------------------------------|------------------------------|-----------------|
| W_{200} | | x_0^2 | piston |
| W_{111} | | $x_0 \rho \cos \theta$ | tilt |
| W_{020} | | ρ^2 | focus |
| W_{040} | $= \frac{1}{8} S_I$ | ρ^4 | spherical |
| W_{131} | $= \frac{1}{2} S_{II}$ | $x_0 \rho^3 \cos \theta$ | coma |
| W_{222} | $= \frac{1}{2} S_{III}$ | $x_0^2 \rho^2 \cos^2 \theta$ | astigmatism |
| W_{220} | $= \frac{1}{4} (S_{III} + S_{IV})$ | $x_0^2 \rho^2$ | field curvature |
| W_{311} | $= \frac{1}{2} S_V$ | $x_0^3 \rho \cos \theta$ | distortion |

focus are first-order properties of the wavefront and are not Seidel aberrations.

Figure 18 shows the departure of the wavefront ΔW from an ideal spherical wavefront form for the five Seidel aberrations.

A. SPHERICAL ABERRATION $(\Delta W = W_{040}\rho^4 = W_{040}(x^2 + y^2)^2)$

Since the field variable x_0 , does not appear in this term, the effect is constant over the field of the system. Figure 19 shows the contours of constant primary spherical aberration.

The rays from an axial object point that make an appreciable angle with the axis will intersect the axis in front of or behind the Gaussian focus. The point at which the rays from the edge of the aperture (marginal rays) intersect

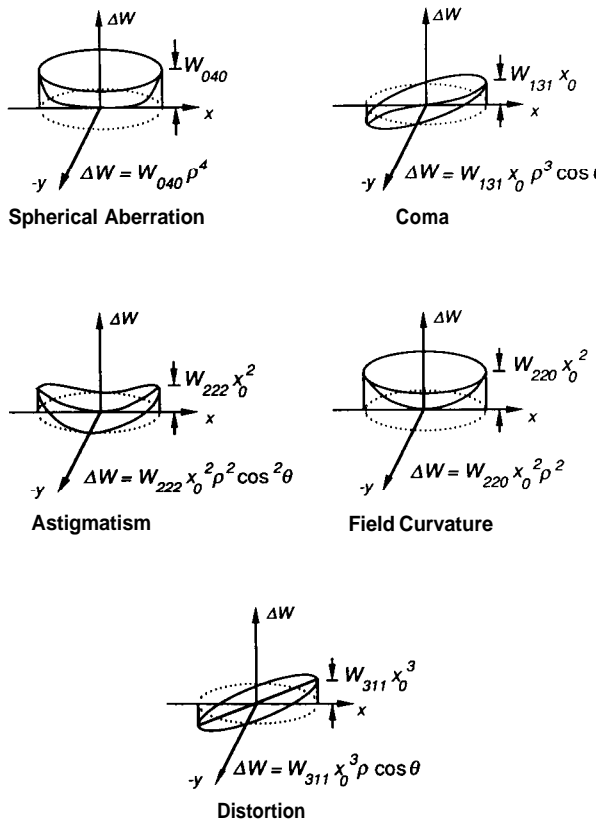


FIG. 18. Seidel aberrations.

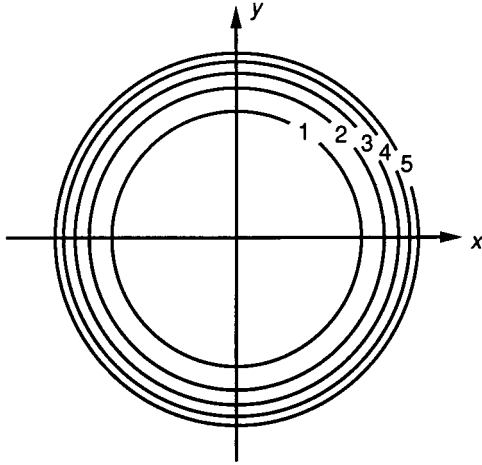


FIG. 19. Levels or contours of constant primary spherical aberration.

the axis is called the marginal focus, as shown in Fig. 20. The point at which the rays from the region near the center of the aperture (paraxial rays) intersect the axis is called the paraxial or Gaussian focus.

If we now go to the transverse ray aberration representation, we find from Eq. (29) that in normalized coordinates the aberration components are

$$\varepsilon_x = -\frac{4RW_{040}}{h} x(x^2 + y^2)$$

(35)

and

$$\varepsilon_y = -\frac{4RW_{040}}{h} y(x^2 + y^2),$$

where h is again the geometrical pupil radius. Since the aberration is

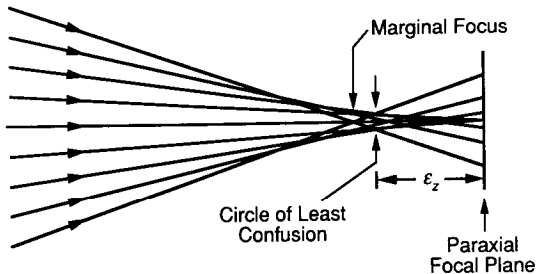


FIG. 20. Circle of least confusion for spherical aberration.

symmetrical about the principal ray, the transverse aberration due to third-order spherical aberration can be written as

$$\varepsilon_x = \varepsilon_y = -\frac{4RW_{040}}{h}\rho^3. \quad (36)$$

A focal shift changes the size of the image patch. It follows from Eq. (17) that if the observation plane is shifted a distance ε_z , the wavefront aberration with respect to the new reference sphere is

$$\Delta W = W_{040}\rho^4 + \frac{\varepsilon_z h^2}{2R^2}\rho^2, \quad (37)$$

where ε_z is positive if the shift is in the + z direction (i.e., away from the lens). The effect of the focal shift on the size of the image patch is equivalent to adding a linear term to Eq. (36) for the transverse aberration:

$$\varepsilon_x = \varepsilon_y = -\frac{4RW_{040}\rho^3}{h} - \frac{\varepsilon_z h \rho}{R}. \quad (38)$$

As the image plane is shifted, there is a position for which the circular image is a minimum; this minimal image is called the circle of least confusion. The focal shift ε_z for which the transverse aberration has equal positive and negative excursion corresponds to the focal shift for the circle of least confusion. As shown in Fig. 20, this image position corresponds to the position where the ray from the rim of the pupil (marginal ray) meets the caustic. It can be shown that the [image position for the circle of least confusion occurs three-quarters of the way from the paraxial focus to the marginal focus](#), and hence its radius, from Eq.(38), RW_{040}/h , one-quarter the radius of the image at the paraxial focus. The resulting wavefront aberration is

$$\Delta W = W_{040}(\rho^4 - 1.5\rho^2). \quad (39)$$

Figure 21 shows the contours of constant wavefront aberration for spherical aberration with defocus.

The diffraction pattern for a rotationally symmetric aberration such as spherical aberration can be calculated quite easily using the [Hankel transform](#). That is, the irradiance of the diffraction pattern of a circular aperture having a rotationally symmetric pupilfunction $\wp(\rho)$, which is illuminated with a uniform beam of total power P contained within the circular aperture, can be written as

$$E(r) = \frac{P\pi}{\lambda^2(f\#)^2} \left| \int_0^1 \wp(\rho) J_0 \left[\left(\frac{\pi r}{\lambda f\#} \right) \rho \right] \rho d\rho \right|^2. \quad (40)$$

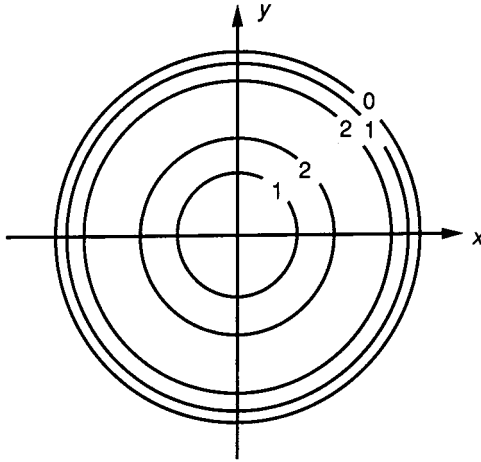


FIG. 21. Contours of constant wavefront aberration for spherical aberration with defocus.

For spherical aberration,

$$\varphi(\rho) = \exp\left(i \frac{2\pi}{\lambda} W_{040}\rho^4\right). \tag{41}$$

If defocus is included, then

$$\varphi(\rho) = \exp\left[i \frac{2\pi}{\lambda} (W_{040}\rho^4 + W_{020}\rho^2)\right]. \tag{42}$$

Equation (40) is well fitted for computer calculations of the point spread function. Figures 22 through 24 illustrate the normalized point spread function for different amounts of third-order spherical aberration and defocus.

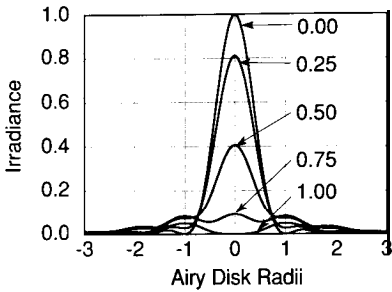


FIG. 22. Diffraction pattern of circular pupil for different values of defocus, W_{020}/λ . (No spherical aberration present.)

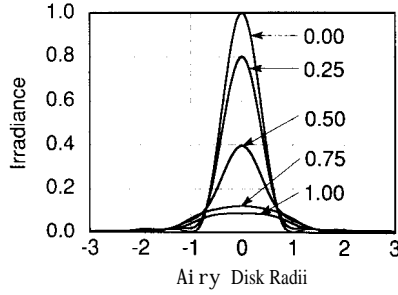


FIG. 23. Diffraction pattern of circular pupil for different values of third-order spherical aberration, W_{040}/λ . (No defocus present.)

$$\text{B. COMA} \quad (\Delta W = W_{131}x_0\rho^3 \cos \theta = W_{131}x_0x(x^2 + y^2))$$

Coma is absent on the axis and increases with field angle or distance. Figure 25 shows contours of wavefront aberration for primary coma. The transverse aberration components are

$$\begin{aligned} \varepsilon_y &= -\frac{2R}{h} W_{131}x_0xy = -\frac{R}{h} W_{131}x_0\rho^2 \sin 2\theta, \\ \varepsilon_x &= -\frac{R}{h} W_{131}x_0(3x^2 + y^2) = -\frac{R}{h} W_{131}x_0\rho^2(2 + \cos 2\theta). \end{aligned} \quad (43)$$

If x_0 is fixed and ρ is kept constant, the image point describes a circle twice over as θ runs through the range from 0 to 2π . The radius of the circle is $RW_{131}x_0\rho^2/h$ and the center is a distance $2RW_{131}x_0\rho^2/h$ from the Gaussian image, as shown in Fig. 26. The circle therefore touches two straight lines that pass through the Gaussian image and are inclined to the x axis at an angle of 30° .

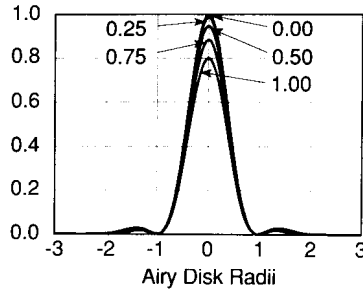


FIG. 24. Diffraction pattern of circular pupil for equal and opposite values of third-order spherical aberration W_{040}/λ and defocus W_{020}/λ .

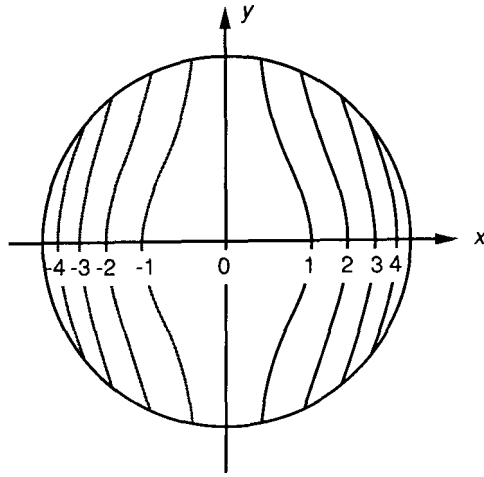


FIG. 25. Contours of wavefront aberration for primary coma.

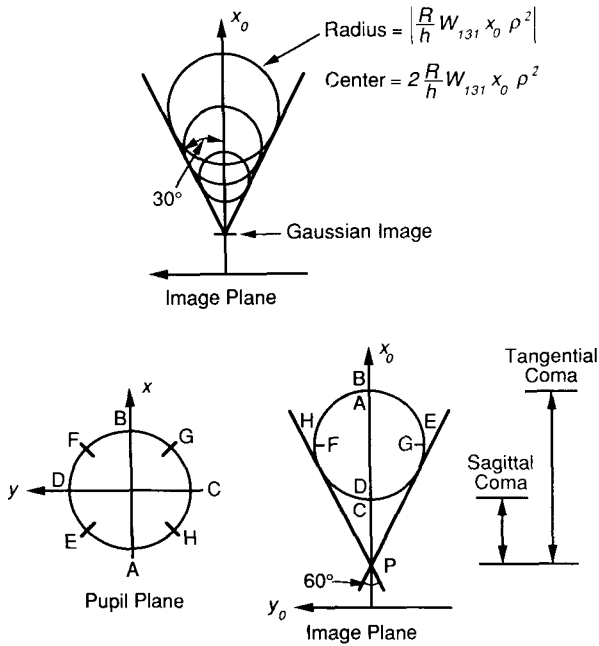


FIG. 26. Transverse ray aberration for coma.

A shift of focus does not improve a comatic image. A lateral shift (tilt) decreases the general departure of the wavefront from the reference sphere as illustrated by Fig. 27, which shows the contour of a coma wavefront with a linear term added. However, this does not really improve the image; it is simply selecting a point other than the Gaussian image point to represent the best center of light concentration in the point image.

In testing an optical system, coma may appear on axis, where coma should be zero. This on-axis aberration is not dependent upon field position, and is additive. It can result from tilted and/or decentered optical components in the system due to misalignment.

C. ASTIGMATISM ($\Delta W = W_{222}x_0^2\rho^2 \cos^2\theta = W_{222}x_0^2x^2$)

In an optical system, a plane containing the optical axis is called the meridional or tangential plane. While the tangential plane can be any plane through the optical axis, the tangential plane for any off-axis object point contains the object point. Rays that lie in this tangential plane are called meridional or tangential rays. The meridional ray through the center of the entrance pupil is called the principal or chief ray (see Fig. 28). The sagittal plane contains the chief ray and is perpendicular to the tangential plane. Rays that do not lie in either the tangential or sagittal planes are called skew rays. For astigmatism, there is no wavefront aberration in the sagittal plane, but in the meridional or tangential plane there is an increment of curvature since the

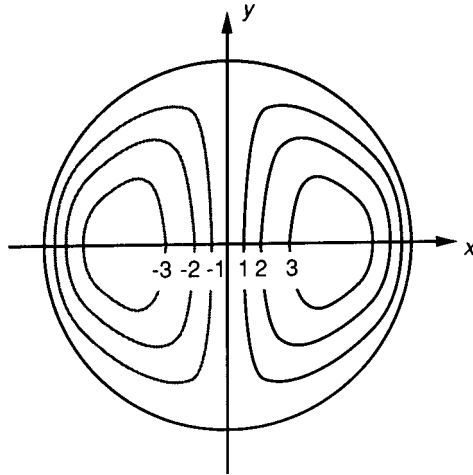


FIG. 27. Contours of comatic wavefront with linear term.

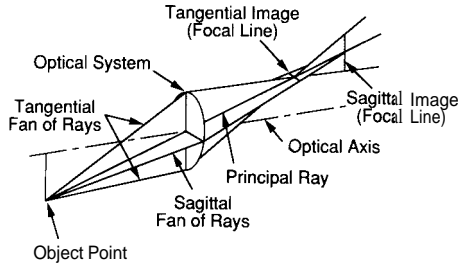


FIG. 28. Sagittal and tangential foci.

aberration depends quadratically on x , as shown in Fig. 29. For conventional astigmatism, the increment in curvature in the x section depends upon the square of the field angle. A common error in the manufacturing of optical surfaces is for a surface to be slightly cylindrical, instead of perfectly spherical, in which case the wavefront leaving the surface will have a different radius of curvature in two orthogonal directions, and the wavefront is said to be astigmatic, even though the curvature difference may not depend upon the square of the field angle.

For astigmatism, the components of transverse aberration are

$$\begin{aligned}\varepsilon_y &= 0, \\ \varepsilon_x &= -\frac{2R}{h} W_{222} x_0^2 x = -\frac{2R}{h} W_{222} x_0^2 \rho \cos \theta.\end{aligned}\quad (44)$$

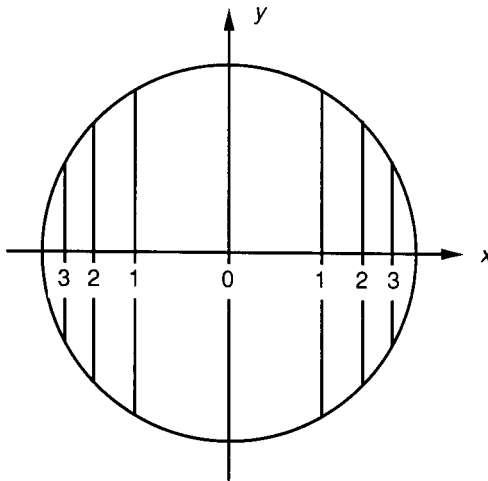


FIG. 29. Contours of astigmatic wavefront.

Thus, a line image of length $4RW_{222}x_0^2/h$ is formed in the meridional plane. For many applications, a focal shift improves the image. Given a wavefront aberration

$$\Delta W = W_{222}x_0^2x^2 + \frac{\varepsilon_z h^2}{2R^2}(x^2 + y^2), \quad (45)$$

the transverse aberration becomes

$$\begin{aligned} \varepsilon_y &= -\left(\frac{\varepsilon_z h}{R}\right)y, \\ \varepsilon_x &= -\left(\frac{2R}{h}W_{222}x_0^2 + \frac{\varepsilon_z h}{R}\right)x. \end{aligned} \quad (46)$$

If the focal shift is chosen so that

$$\frac{\varepsilon_z}{R} = -\frac{2RW_{222}x_0^2}{h^2}, \quad (47)$$

a line image of length $4RW_{222}x_0^2/h$ is formed in the sagittal plane. Thus, the three-dimensional distribution of rays in the image region is such that they all pass through two orthogonal lines, as shown in Fig. 28. The vertical line in Fig. 28 can be regarded as the focus for rays from the y section of the pupil since ε_y is zero at this focus, and the horizontal line is the focus for the x section. The vertical line is called the sagittal astigmatic focal line and the horizontal line is called the meridional or tangential focal line. **It should be remembered that the sagittal focal line is in the meridional plane and vice versa.**

The best image is generally taken to be halfway between the astigmatic foci. At this focal position, all the rays pass through a circular patch of diameter $2RW_{222}x_0^2/h$, half the length of a focal line, and this is called the circle of least confusion. Figure 30 shows the contours of the wavefront defocused to give the circle of least confusion. For an object such as a spoked wheel, the image of the rim is sharp on the tangential focal plane and the image of the spokes is sharp on the sagittal focal plane as shown in Fig. 31.

D. FIELD CURVATURE ($\Delta W = W_{220}x_0^2\rho^2 = W_{220}x_0^2(x^2 + y^2)$)

The fourth Seidel aberration has the same pupil dependence as a focal shift, but it depends upon the square of the field height. If no other aberrations are present, the images are true point images on a curved image surface, called the Petzval surface, of curvature $4R^2(W_{220}/h^2)$, as shown in Fig. 32. Thus, this aberration is called field curvature. Astigmatism and field curvature are often grouped together since they both depend upon the square of the field height.

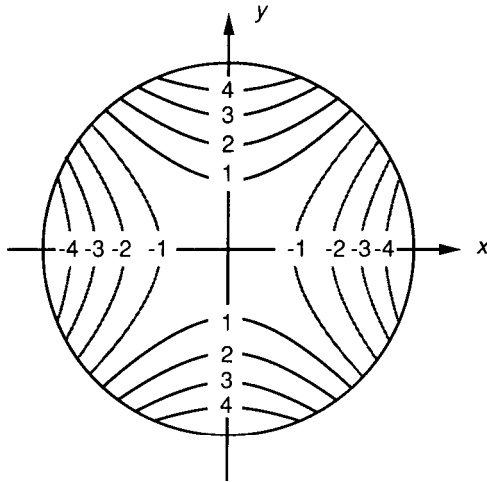
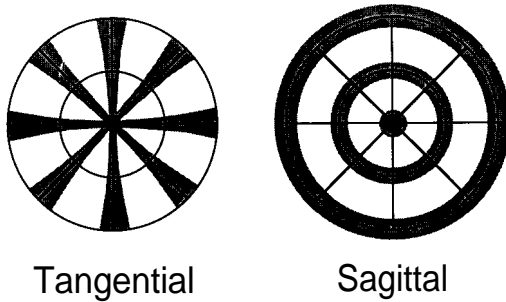


FIG. 30. Contours of astigmatic wavefront defocused to give circle of least confusion



Tangential

Sagittal

FIG. 31. Astigmatic images of a spoked wheel.

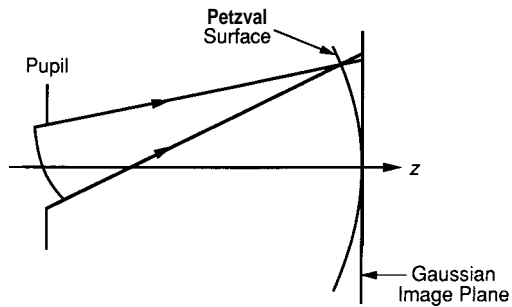


FIG. 32. Field curvature.

If no astigmatism is present, the sagittal and tangential image surfaces coincide and lie on the Petzval surface. When primary astigmatism is present, both the tangential and the sagittal image surfaces are on the same side of the Petzval surface with the tangential image surface three times as far from the Petzval surface as the sagittal image surface, as shown in Fig. 33. The curvature of the sagittal image surface is $4R^2(W_{220}/h^2)$, the curvature of the tangential image surface is $4R^2(W_{220} + W_{222})/h^2$, and the curvature of the Petzval surface is $4R^2(W_{220} - W_{222}/2)/h^2$.

$$\text{E. DISTORTION} \quad (\Delta W = W_{311}x_0^3\rho \cos \theta = W_{311}x_0^3x)$$

The last Seidel aberration, distortion, has the effect of shifting the image position a distance

$$\varepsilon_x = -\frac{R}{h} W_{311}x_0^3. \quad (48)$$

This is not simply a change in magnification, since the shift is proportional to the cube of the Gaussian image height. The effect is that the shape of the image is distorted, and consequently this aberration is known as distortion. The image of any straight line that meets the origin is a straight line, but the image of any other straight line is curved. We can see the general effect by taking a square placed symmetrically with respect to the optical axis and shifting each point radially an amount proportional to the cube of its distance from the center, as shown in Fig. 34.

VI. ZERNIKE POLYNOMIALS

Often, to aid in the interpretation of optical test results it is convenient to express wavefront data in polynomial form. Zernike polynomials are often

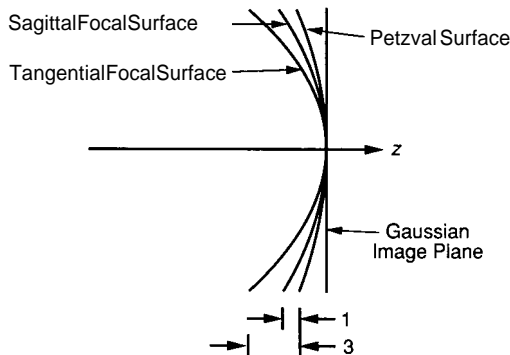


FIG. 33. Focal surfaces in presence of field curvature and astigmatism.

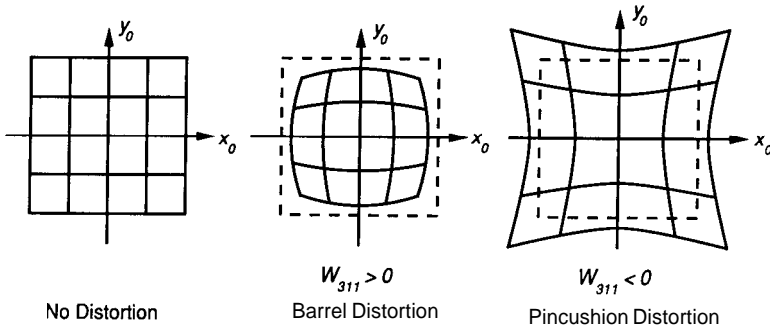


FIG. 34. Distortion,

used for this purpose since they are made up of terms that are of the **same form as the types of aberrations often observed in optical tests** (Zernike, 1934). This is not to say that Zernike polynomials are the best polynomials for fitting test data. Sometimes Zernike polynomials give a terrible representation of the wavefront data. For example, Zernikes have little value when air turbulence is present. Likewise, fabrication errors present in the single-point diamond turning process cannot be represented using a reasonable number of terms in the Zernike polynomial. In the testing of conical optical elements, additional terms must be added to Zernike polynomials to accurately represent alignment errors. Thus, **the reader should be warned that the blind use of Zernike polynomials to represent test results can lead to disastrous results.**

Zernike polynomials have several interesting properties. First, they are one of an infinite number of complete sets of polynomials in two real variables, ρ and θ' that are orthogonal in a continuous fashion over the interior of a unit circle. It is important to note that **the Zernikes are orthogonal only in a continuous fashion over the interior of a unit circle, and in general they will not be orthogonal over a discrete set of data points within a unit circle.**

Zernike polynomials have three properties that distinguish them from other sets of orthogonal polynomials. First, they have simple rotational symmetry properties that lead to a polynomial product of the form

$$R(\rho)G(\theta'), \quad (49)$$

where $G(\theta')$ is a continuous function that repeats itself every 2π radians and satisfies the requirement that rotating the coordinate system by an angle α does not change the form of the polynomial. That is,

$$G(\theta' + \alpha) = G(\theta')G(\alpha). \quad (50)$$

The set of trigonometric functions

$$G(\theta') = e^{\pm im\theta'}, \quad (51)$$

where m is any positive integer or zero, meets these requirements.

The second property of Zernike polynomials is that the radial function must be a polynomial in ρ of degree n and contain no power of ρ less than m . The third property is that $R(\rho)$ must be even if m is even, and odd if m is odd.

The radial polynomials can be derived as a special case of Jacobi polynomials, and tabulated as $R_n^m(\rho)$. Their orthogonality and normalization properties are given by

$$\int_0^1 R_n^m(\rho) R_n^m(\rho) \rho d\rho = \frac{1}{2(n+1)} \delta_{nn'} \quad (52)$$

and

$$R_n^m(1) = 1. \quad (53)$$

It is convenient to factor the radial polynomial into

$$R_{2n-m}^m(\rho) = Q_n^m(\rho) \rho^m, \quad (54)$$

where $Q_n^m(\rho)$ is a polynomial of order $2(n-m)$. $Q_n^m(\rho)$ can be written generally as

$$Q_n^m(\rho) = \sum_{s=0}^{n-m} (-1)^s \frac{(2n-m-s)!}{s!(n-s)!(n-m-s)!} \rho^{2(n-m-s)}. \quad (55)$$

In practice, the radial polynomials are combined with sines and cosines rather than with a complex exponential. The final Zernike polynomial series for the wavefront OPD ΔW can be written as

$$W = \overline{\Delta W} + \sum_{n=1}^{\infty} \left[A_n Q_n^0(\rho) + \sum_{m=1}^n Q_n^m(\rho) \rho^m (B_{nm} \cos m\theta' + C_{nm} \sin m\theta') \right], \quad (56)$$

where $\overline{\Delta W}$ is the mean wavefront OPD, and A_n , B_{nm} , and C_{nm} are individual polynomial coefficients. For a symmetrical optical system, the wave aberrations are symmetrical about the tangential plane and only even functions of θ' are allowed. In general, however, the wavefront is not symmetric, and both sets of trigonometric terms are included.

Table III gives a list of 36 Zernike polynomials, plus the constant term. (Note that the ordering of the polynomials in the list is not universally accepted, and different organizations may use a different ordering.) Figures 35 through 39 show contour maps of the 36 terms. Term #0 is a constant or piston term, while terms #1 and #2 are tilt terms. Term #3 represents focus. Thus, terms #1 through #3 represent the Gaussian or paraxial properties of

TABLE III
ZERNIKE RADIAL POLYNOMIALS

| n | m | No. | Polynomial |
|-----|-----|---|---|
| 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | $\rho \cos \theta'$ |
| | | 2 | $\rho \sin \theta'$ |
| | 0 | 3 | $2\rho^2 - 1$ |
| 2 | 2 | 4 | $\rho^2 \cos 2\theta'$ |
| | | 5 | $\rho^2 \sin 2\theta'$ |
| | 1 | 6 | $(3\rho^2 - 2)\rho \cos \theta'$ |
| | | 7 | $(3\rho^2 - 2)\rho \sin \theta'$ |
| | 0 | 8 | $6\rho^4 - 6\rho^2 + 1$ |
| 3 | 3 | 9 | $\rho^3 \cos 3\theta'$ |
| | | 10 | $\rho^3 \sin 3\theta'$ |
| | 2 | 11 | $(4\rho^2 - 3)\rho^2 \cos 2\theta'$ |
| | | 12 | $(4\rho^2 - 3)\rho^2 \sin 2\theta'$ |
| | 1 | 13 | $(10\rho^4 - 12\rho^2 + 3)\rho \cos \theta'$ |
| | | 14 | $(10\rho^4 - 12\rho^2 + 3)\rho \sin \theta'$ |
| | 0 | 15 | $20\rho^6 - 30\rho^4 + 12\rho^2 - 1$ |
| 4 | 4 | 16 | $\rho^4 \cos 4\theta'$ |
| | | 17 | $\rho^4 \sin 4\theta'$ |
| | 3 | 18 | $(5\rho^2 - 4)\rho^3 \cos 3\theta'$ |
| | | 19 | $(5\rho^2 - 4)\rho^3 \sin 3\theta'$ |
| | 2 | 20 | $(15\rho^4 - 20\rho^2 + 6)\rho^2 \cos 2\theta'$ |
| | | 21 | $(15\rho^4 - 20\rho^2 + 6)\rho^2 \sin 2\theta'$ |
| | 1 | 22 | $(35\rho^6 - 60\rho^4 + 30\rho^2 - 4)\rho \cos \theta'$ |
| | | 23 | $(35\rho^6 - 60\rho^4 + 30\rho^2 - 4)\rho \sin \theta'$ |
| | 0 | 24 | $70\rho^8 - 140\rho^6 + 90\rho^4 - 20\rho^2 + 1$ |
| | 5 | 5 | 25 |
| 26 | | | $\rho^5 \sin 5\theta'$ |
| 4 | | 27 | $(6\rho^2 - 5)\rho^4 \cos 4\theta'$ |
| | | 28 | $(6\rho^2 - 5)\rho^4 \sin 4\theta'$ |
| 3 | | 29 | $(21\rho^4 - 30\rho^2 + 10)\rho^3 \cos 3\theta'$ |
| | | 30 | $(21\rho^4 - 30\rho^2 + 10)\rho^3 \sin 3\theta'$ |
| 2 | | 31 | $(56\rho^6 - 105\rho^4 + 60\rho^2 - 10)\rho^2 \cos 2\theta'$ |
| | | 32 | $(56\rho^6 - 105\rho^4 + 60\rho^2 - 10)\rho^2 \sin 2\theta'$ |
| 1 | | 33 | $(126\rho^8 - 280\rho^6 + 210\rho^4 - 60\rho^2 + 5)\rho \cos \theta'$ |
| | | 34 | $(126\rho^8 - 280\rho^6 + 210\rho^4 - 60\rho^2 + 5)\rho \sin \theta'$ |
| 0 | 35 | $252\rho^{10} - 630\rho^8 + 560\rho^6 - 210\rho^4 + 30\rho^2 - 1$ | |
| 6 | 0 | 36 | $924\rho^{12} - 2772\rho^{10} + 3150\rho^8 - 1680\rho^6 + 420\rho^4 - 42\rho^2 + 1$ |

n=1

FIRST-ORDER PROPERTIES

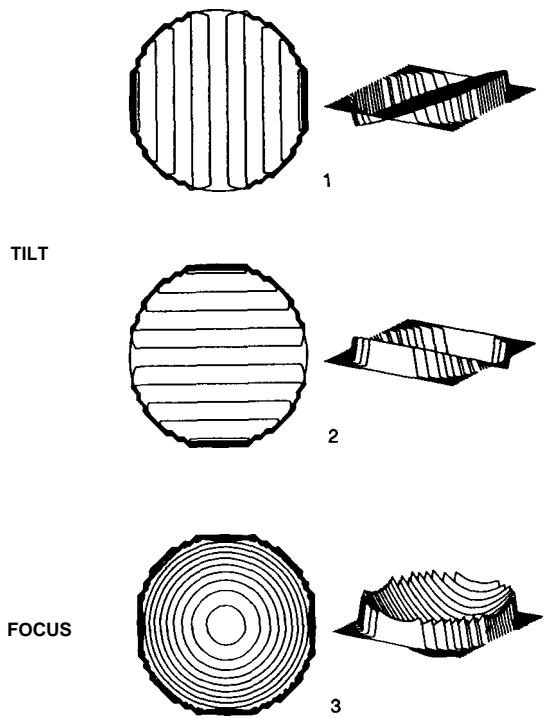


FIG. 35. Two- and three-dimensional plots of Zernike polynomials # 1 to # 3.

n=2

THIRD-ORDER ABERRATIONS

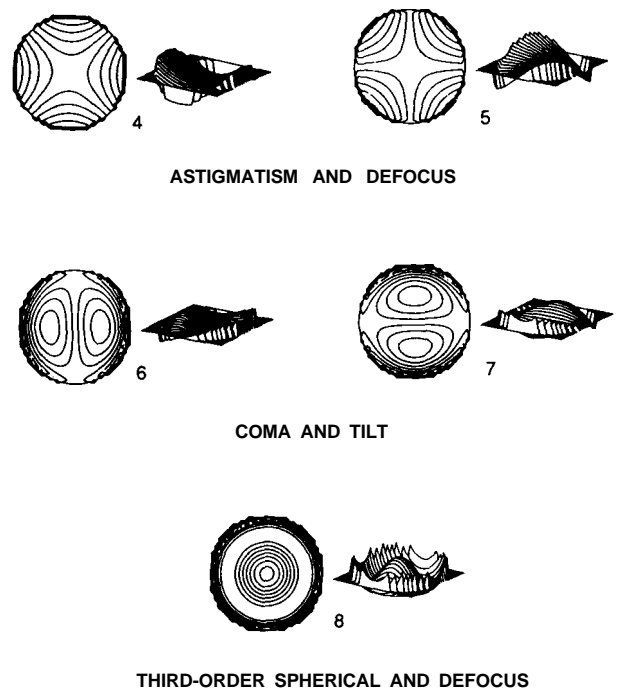


FIG. 36. Two- and three-dimensional plots of Zernike polynomials #4 to # 8.

THIRD-ORDER SPHERICAL AND DEFOCUS

n=3

FIFTH-ORDER ABERRATIONS

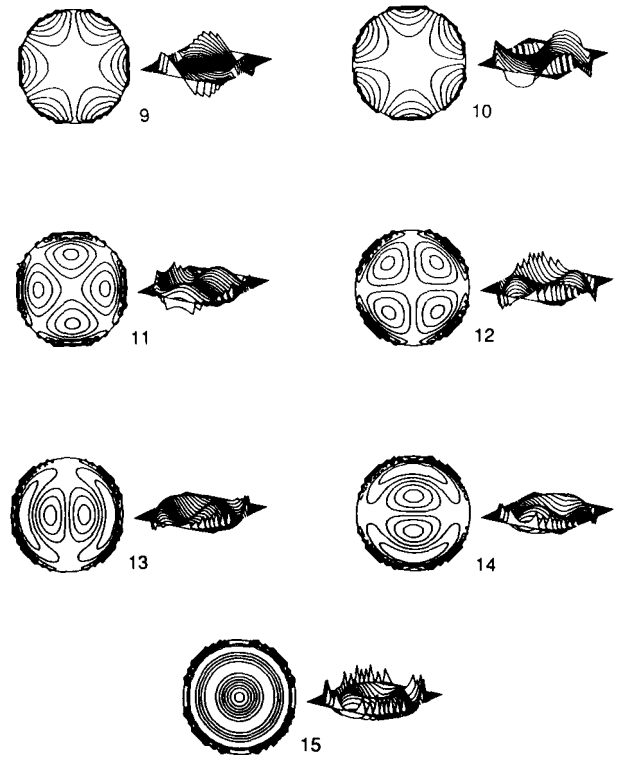


FIG. 37. Two- and three-dimensional plots of Zernike polynomials # 9 to # 15.

n=4

SEVENTH-ORDER ABERRATIONS

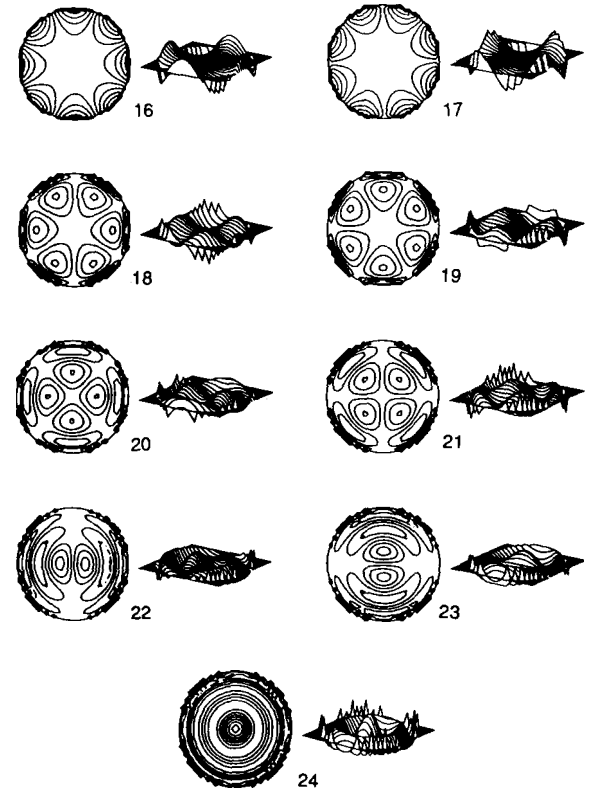


FIG. 38. Two- and three-dimensional plots of Zernike polynomials # 16 to # 24.

n=5,6 NINTH- & ELEVENTH-ORDER ABERRATIONS

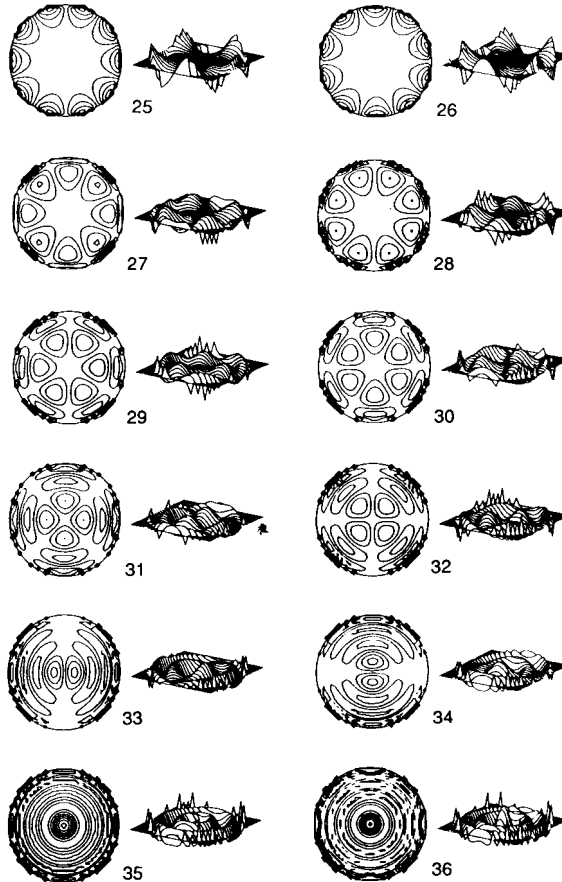


FIG. 39. Two- and three-dimensional plots of Zernike polynomials # 25 to # 36.

the wavefront. Terms # 4 and # 5 are astigmatism plus defocus. Terms # 6 and #7 represent coma and tilt, while term #8 represents third-order spherical and focus. Likewise, terms # 9 through # 15 represent fifth-order aberration, # 16 through # 24 represent seventh-order aberrations, and # 25 through # 35 represent ninth-order aberrations. Each term contains the appropriate amount of each lower order term to make it orthogonal to each lower order term. Also, each term of the Zernikes minimizes the rms wavefront error to the order of that term. Adding other aberrations of lower order can only increase the rms error. Furthermore, the average value of each term over the unit circle is zero.

VII. RELATIONSHIP BETWEEN ZERNIKE POLYNOMIALS AND THIRD-ORDER ABERRATIONS

First-order wavefront properties and third-order wavefront aberration coefficients can be obtained from the Zernike polynomials coefficients. Using the first nine Zernike terms Z_0 to Z_8 , shown in Table III, the wavefront can be written as

$$\begin{aligned}
 W(\rho, \theta) = & Z_0 + Z_1\rho \cos \theta' + Z_2\rho \sin \theta' + Z_3(2\rho^2 - 1) \\
 & + Z_4\rho^2 \cos 2\theta' + Z_5\rho^2 \sin 2\theta' \\
 & + Z_6(3\rho^2 - 2)\rho \cos \theta' + Z_7(3\rho^2 - 2)\rho \sin \theta' \\
 & + Z_8(6\rho^4 - 6\rho^2 + 1).
 \end{aligned} \tag{57}$$

The aberrations and properties corresponding to these Zernike terms are shown in Table IV. Writing the wavefront expansion in terms of field-independent wavefront aberration coefficients, we obtain

$$\begin{aligned}
 W(\rho, \theta) = & W_{11}\rho \cos \theta + W_{20}\rho^2 \\
 & + W_{40}\rho^4 + W_{31}\rho^3 \cos \theta + W_{22}\rho^2 \cos^2 \theta.
 \end{aligned} \tag{58}$$

Because there is no field dependence in these terms, they are not true Seidel aberrations. Wavefront measurement using an interferometer only provides data at a single field point. This causes field curvature to look like focus, and distortion to look like tilt. Therefore, a number of field points must be measured to determine the Seidel aberrations.

Rewriting the Zernike expansion of Eq. (57), first- and third-order field-independent wavefront aberration terms are obtained. This is done by

TABLE IV
ABERRATIONS CORRESPONDING TO THE FIRST NINE ZERNIKE
TERMS

| | |
|-------|----------------------------------|
| Z_0 | piston |
| Z_1 | x-tilt |
| Z_2 | y-tilt |
| Z_3 | focus |
| Z_4 | astigmatism @ 0° & focus |
| Z_5 | astigmatism @ 45° & focus |
| Z_6 | coma & x-tilt |
| Z_7 | coma & y-tilt |
| Z_8 | spherical & focus |

grouping like terms, and equating them with the wavefront aberration coefficients:


$$\begin{aligned}
 W(\rho, \theta') &= Z_0 - Z_3 + Z_8 && \text{piston} \\
 &+ (Z_1 - 2Z_6)\rho \cos \theta' + (Z_2 - 2Z_7)\rho \sin \theta' && \text{tilt} \\
 &+ (2Z_3 - 6Z_8 + Z_4 \cos 2\theta' + Z_5 \sin 2\theta')\rho^2 && \text{focus + astigmatism} \\
 &+ 3(Z_6 \cos \theta' + Z_7 \sin \theta')\rho^3 && \text{coma} \\
 &+ 6Z_8\rho^4. && \text{spherical}
 \end{aligned} \tag{59}$$

Equation (59) can be rearranged using the identity

$$a \cos \alpha + b \sin \alpha = \sqrt{a^2 + b^2} \cos\left[\alpha - \tan^{-1}\left[\left(\frac{b}{a}\right)\right]\right], \tag{60}$$

yielding terms corresponding to field-independent wavefront aberration coefficients:

$$\begin{aligned}
 W(\rho, \theta') &= Z_0 - Z_3 + Z_8 && \text{piston} \\
 &+ \rho\sqrt{(Z_1 - 2Z_6)^2 + (Z_2 - 2Z_7)^2} \\
 &\times \cos\left[\theta' - \tan^{-1}\left(\frac{Z_2 - 2Z_7}{Z_1 - 2Z_6}\right)\right] && \text{tilt} \\
 &+ \rho^2(2Z_3 - 6Z_8 \pm \sqrt{Z_4^2 + Z_5^2}) && \text{focus} \\
 &\pm 2\rho^2\sqrt{Z_4^2 + Z_5^2} \cos^2\left[\theta' - \frac{1}{2}\tan^{-1}\left(\frac{Z_5}{Z_4}\right)\right] && \text{astigmatism} \\
 &+ 3\rho^3\sqrt{Z_6^2 + Z_7^2} \cos\left[\theta' - \tan^{-1}\left(\frac{Z_7}{Z_6}\right)\right] && \text{coma} \\
 &+ 6\rho^4Z_8. && \text{spherical}
 \end{aligned} \tag{61}$$

The magnitude, sign, and angle of these field-independent aberration terms are listed in Table V. Note that focus has the sign chosen to minimize the magnitude of the coefficient, and astigmatism uses the sign opposite that chosen for focus. 

VIII. PEAK-TO-VALLEY AND RMS WAVEFRONT ABERRATION

If the wavefront aberration can be described in terms of third-order aberrations, it is convenient to specify the wavefront aberration by stating the number of waves of each of the third-order aberrations present. This method

TABLE V

THIRD-ORDER ABERRATIONS IN TERMS OF ZERNIKE COEFFICIENTS

| Term | Description | Magnitude | Angle |
|----------|-------------|---|---|
| W_{11} | tilt | $\sqrt{(Z_1 - 2Z_6)^2 + (Z_2 - 2Z_7)^2}$ | $\tan^{-1}\left(\frac{Z_2 - 2Z_7}{Z_1 - 2Z_6}\right)$ |
| W_{20} | focus | $2Z_3 - 6Z_8 \pm \sqrt{Z_4^2 + Z_5^2}$ sign chosen to minimize absolute value of magnitude | |
| W_{22} | astigmatism | $\pm 2\sqrt{Z_4^2 + Z_5^2}$ sign opposite that chosen in focus term | $\frac{1}{2}\tan^{-1}\left(\frac{Z_5}{Z_4}\right)$ |
| W_{31} | coma | $3\sqrt{Z_6^2 + Z_7^2}$ | $\tan^{-1}\left(\frac{Z_7}{Z_6}\right)$ |
| W_{40} | spherical | $6Z_8$ | |

Note: For angle calculations, if denominator < 0, then angle ← angle + 180°.

for specifying a wavefront is of particular convenience if only a single third-order aberration is present. For more complicated wavefront aberrations it is convenient to state the **peak-to-valley (P-V)** sometimes called peak-to-peak (P-P) wavefront aberration. This is simply the maximum departure of the actual wavefront from the desired wavefront in both positive and negative directions. For example, if the maximum departure in the positive direction is +0.2 waves and the maximum departure in the negative direction is -0.1 waves, then the P-V wavefront error is 0.3 waves.

While using P-V to specify wavefront error is convenient and simple, it can be misleading. Stating P-V is simply stating the maximum wavefront error, and it is telling nothing about the area over which this error is occurring. An optical system having a large P-V error may actually perform better than a system having a small P-V error. **It is generally more meaningful to specify wavefront quality using the rms wavefront error.**

Equation (62) defines the rms wavefront error σ for a circular pupil, as well as the variance σ^2 . $\Delta W(\rho, \theta)$ is measured relative to the best fit spherical wave, and it generally has the units of waves. $\overline{\Delta W}$ is the mean wavefront OPD.

$$\begin{aligned}\sigma^2 &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 [\Delta W(\rho, \theta) - \overline{\Delta W}]^2 \rho \, d\rho \, d\theta = \overline{\Delta W^2} - (\overline{\Delta W})^2 \\ &= \left[\frac{1}{\pi} \int_0^{2\pi} \int_0^1 \Delta W^2(\rho, \theta) \rho \, d\rho \, d\theta \right] - \frac{1}{\pi^2} \left[\int_0^{2\pi} \int_0^1 \Delta W(\rho, \theta) \rho \, d\rho \, d\theta \right]^2. \quad (62)\end{aligned}$$

If the wavefront aberration can be expressed in terms of Zernike polynomials,

the wavefront variance can be calculated in a simple form by using the orthogonality relations of the Zernike polynomials. The final result for the entire unit circle is

$$\sigma^2 = \sum_{n=1}^{\infty} \left[\frac{A_n^2}{2n+1} + \frac{1}{2} \sum_{m=1}^n \frac{B_{nm}^2 + C_{nm}^2}{2n+1-m} \right]. \quad (63)$$

Table VI gives the relationship between σ and mean wavefront aberration for the third-order aberrations of a circular pupil. While Eq. (62) could be used to calculate the values of σ given in Table VI, it is easier to use linear combinations of the Zernike polynomials to express the third-order aberrations, and then use Eq. (63).

IX. STREHL RATIO

While in the absence of aberrations, the intensity is a maximum at the Gaussian image point. If aberrations are present this will in general no longer be the case. The point of maximum intensity is called diffraction focus, and for small aberrations is obtained by finding the appropriate amount of tilt and defocus to be added to the wavefront so that the wavefront variance is a minimum.

The ratio of the intensity at the Gaussian image point (the origin of the reference sphere is the point of maximum intensity in the observation plane) in the presence of aberration, divided by the intensity that would be obtained if no aberration were present, is called the Strehl ratio, the Strehl definition, or the Strehl intensity. The Strehl ratio is given by

$$\text{Strehl ratio} = \frac{1}{\pi^2} \left| \int_0^{2\pi} \int_0^1 e^{i2\pi\Delta W(\rho,\theta)} \rho \, d\rho \, d\theta \right|^2, \quad (64)$$

TABLE VI
RELATIONSHIPS BETWEEN WAVEFRONT ABERRATION MEAN AND RMS FOR
FIELD-INDEPENDENT THIRD-ORDER ABERRATIONS

| Aberration | ΔW | $\overline{\Delta W}$ | σ |
|-----------------------|--|-----------------------|---------------|
| Defocus | $W_{20}\rho^2$ | $\frac{1}{2}W_{20}$ | $0.289W_{20}$ |
| Spherical | $W_{40}\rho^4$ | $\frac{1}{3}W_{40}$ | $0.298W_{40}$ |
| Spherical & Defocus | $W_{40}[\rho^4 - \rho^2]$ | $-\frac{1}{6}W_{40}$ | $0.075W_{40}$ |
| Astigmatism | $W_{22}\rho^2 \cos^2 \theta$ | $\frac{1}{4}W_{22}$ | $0.250W_{22}$ |
| Astigmatism & Defocus | $W_{22}[\rho^2 \cos^2 \theta - \frac{1}{2}\rho^2]$ $= \frac{1}{2}W_{22}\rho^2 \cos 2\theta$ | 0 | $0.204W_{22}$ |
| Coma | $W_{31}\rho^3 \cos \theta$ | 0 | $0.354W_{31}$ |
| Coma & Tilt | $W_{31}[\rho^3 - \frac{2}{3}\rho] \cos \theta$ | 0 | $0.118W_{31}$ |

where ΔW in units of waves is the wavefront aberration relative to the reference sphere for diffraction focus. Equation (64) may be expressed in the form

$$\text{Strehl ratio} = \frac{1}{\pi^2} \left| \int_0^{2\pi} \int_0^1 [1 + i2\pi \Delta W + \frac{1}{2}(i2\pi \Delta W)^2 + \dots] \rho \, d\rho \, d\theta \right|^2. \quad (65)$$

If the aberrations are so small that the third-order and higher-order powers of $2\pi \Delta W$ can be neglected, Eq. (65) may be written as

$$\begin{aligned} \text{Strehl ratio} &\approx \left| 1 + i2\pi \overline{\Delta W} - \frac{1}{2}(2\pi)^2 \overline{\Delta W^2} \right|^2 \\ &\approx 1 - (2\pi)^2 [\overline{\Delta W^2} - (\overline{\Delta W})^2] \\ &\approx 1 - (2\pi\sigma)^2, \end{aligned} \quad (66)$$

where σ is in units of waves.

Thus, when the aberrations are small, the Strehl ratio is independent of the nature of the aberration and is smaller than the ideal value of unity by an amount proportional to the variance of the wavefront deformation.

Equation (66) is valid for Strehl ratios as low as about 0.5. The Strehl ratio is always somewhat larger than would be predicted by Eq. (66). A better approximation for most types of aberration is given by

$$\text{Strehl ratio} \approx e^{-(2\pi\sigma)^2} \approx 1 - (2\pi\sigma)^2 + \frac{(2\pi\sigma)^4}{2!} + \dots, \quad (67)$$

which is good for Strehl ratios as small as 0.1.

Once the normalized intensity at diffraction focus has been determined, the quality of the optical system may be ascertained using the Maréchal criterion. The Maréchal criterion states that a system is regarded as well corrected if the normalized intensity at diffraction focus is greater than or equal to 0.8, which corresponds to an rms wavefront error $\leq \lambda/14$.

As mentioned in Section VI, a useful feature of Zernike polynomials is that each term of the Zernikes minimizes the rms wavefront error to the order of that term. That is, each term is structured such that adding other aberrations of lower orders can only increase the rms error. Removing the first-order Zernike terms of tilt and defocus represents a shift in the focal point that maximizes the intensity at that point. Likewise, higher order terms have built into them the appropriate amount of tilt and defocus to minimize the rms wavefront error to that order. For example, looking at Zernike term 8 in Table III shows that for each wave of third-order spherical aberration present, one wave of defocus should be subtracted to minimize the rms wavefront error and find diffraction focus.

X. CHROMATIC ABERRATIONS

Because the index of refraction varies as a function of the wavelength of light, the properties of optical elements also vary with wavelength. There are two general types of chromatic aberration: (1) chromatic variations in paraxial image-forming properties of a system, and (2) wavelength dependence of monochromatic aberrations. It is generally the first type of chromatic aberration that is of most interest.

The Gaussian optics properties of any image-forming system depend solely on the location of principal planes and focal planes. Chromatic aberrations result when the location of any of the planes is wavelength-dependent. It is possible to discuss the results in terms of a variation in image distance along the axis and a variation in transverse magnification.

Longitudinal chromatic aberration is the variation of focus (or image position) with wavelength. In general, the refractive index of optical materials is larger for short wavelengths than for long wavelengths. Hence, the short wavelengths are more strongly refracted at each surface of a lens, as shown in Fig. 40. The distance along the axis between the two focal points is called the axial or longitudinal chromatic aberration. When the short wavelength rays are brought to focus nearer the lens than the long wavelength rays, the lens is said to be undercorrected. Thus, for an undercorrected visual instrument, the image has a yellowish dot (formed by the orange, yellow, and green rays) and a purplish halo (due to the red and blue rays). If the viewing screen is moved toward the lens, the central dot becomes blue; if it is moved away, the central dot becomes red. When a lens system forms images of different sizes for different wavelengths or spreads the image of an off-axis point into a rainbow, the difference between the image heights for different colors is called lateral color or chromatic difference of magnification, as shown in Fig. 41 for a simple lens.

XI. ABERRATIONS INTRODUCED BY PLANE PARALLEL PLATES

Often, in an optical test setup a spherical wavefront is transmitted through a plane parallel plate. While a ray transmitted through a plane parallel plate has exactly the same slope angle that it has before passing

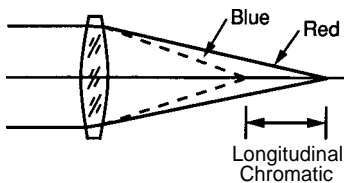


FIG. 40. Longitudinal chromatic aberration (variation of focus with wavelength).

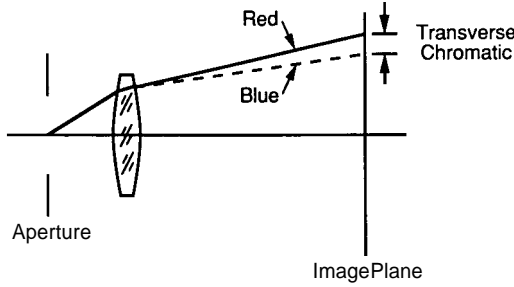


FIG. 41. Lateral or transverse chromatic aberration (variation of magnification with wavelength).

through the plate, the plate causes both longitudinal and lateral displacement of the focal point of a spherical beam, as well as aberrations. (For further information, see Smith (1990), p. 96.) The longitudinal displacement produced by passage through a plate of thickness T and refractive index n is easily found by use of Snell's law for small angle of incidence to be

$$\text{longitudinal displacement} = \frac{(n - 1)}{n} T. \tag{68}$$

This longitudinal displacement shown in Fig. 42 is independent of the position of the plate in the spherical beam. The effective thickness of the plate compared to air (the equivalent air thickness) is less than the actual thickness T by the amount of this shift. The equivalent air thickness is thus equal to

$$T - \frac{(n - 1)}{n} T = \frac{T}{n}. \tag{69}$$

If a collimated beam of light is incident upon a plane parallel plate at an angle I , as shown in Fig. 43, the beam is laterally displaced an amount D given by

$$\begin{aligned} D &= T \cos I(\tan I - \tan I') \\ &= T \sin I \left(1 - \sqrt{\frac{1 - \sin^2 I'}{n^2 - \sin^2 I'}} \right). \end{aligned} \tag{70}$$

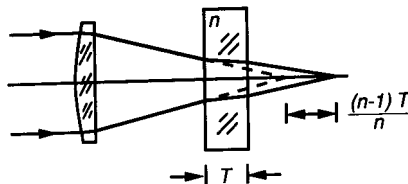


FIG. 42. Longitudinal displacement of image produced by a plane parallel plate.

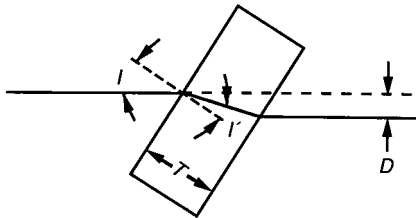


FIG. 43. Lateral displacement of beam produced by plane parallel plate.

For small angles, D can be approximated as

$$D \approx \frac{TI(n-1)}{n}. \quad (71)$$

It is often convenient to remember that for $n \cong 1.5$ and $I \cong 45^\circ$, D is approximately $T/3$.

When used in collimated light, the plane parallel plate does not introduce any aberration; however, when used with either converging or diverging light, aberrations are introduced. The source of the aberration is that rays that have a larger angle of incidence with respect to the normal to the plate are displaced more than rays that have a smaller angle of incidence.

Let U be the slope angle of the ray to the axis and \bar{U} be the tilt of the plate, as shown in Fig. 44; then, the longitudinal spherical aberration for a plate of refractive index n in air is given by

$$\begin{aligned} L' - l' &= \frac{T}{n} \left[1 - \frac{n \cos U}{\sqrt{n^2 - \sin^2 U}} \right] \quad (\text{exact}) \\ &\approx \frac{TU^2(n^2 - 1)}{2n^3} \quad (\text{third-order}), \end{aligned} \quad (72)$$

where L' is the distance from the plate to the focus of the marginal rays, and l' is the distance from the plate to the focus for the paraxial rays.

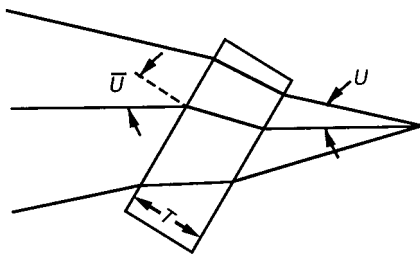


FIG. 44. Spherical wavefront passing through tilted plane parallel plate.

The third-order spherical wavefront aberration with $\rho = 1$ is given by

$$\begin{aligned}\Delta W_{\text{sph}} &= -\frac{TU^4(n^2 - 1)}{8n^3} \\ &= -\frac{T}{(f\#)^4} \left[\frac{(n^2 - 1)}{128n^3} \right],\end{aligned}\quad (73)$$

where the effective F-number $f\#$ is equal to $1/2U$. Note that the $f\#$ is negative in a converging beam. Figure 45 gives two graphs of ΔW for third-order spherical, in micrometers, as a function of $f\#$, and the plate thickness T in millimeters, for $n = 1.5$.

It is interesting to note from Eq. (73) that if T and $f\#$ are kept fixed, as n increases from 1 to $\sqrt{3}$ the amount of third-order spherical aberration increases, while as n increases beyond $\sqrt{3}$ the amount of third-order spherical aberration decreases, and in the limit as $n \rightarrow \infty$ the amount of spherical aberration goes to zero. While at first it may seem strange that as the refractive index becomes very large the aberration becomes small, it becomes less puzzling when we consider that as the index becomes very large, the rays

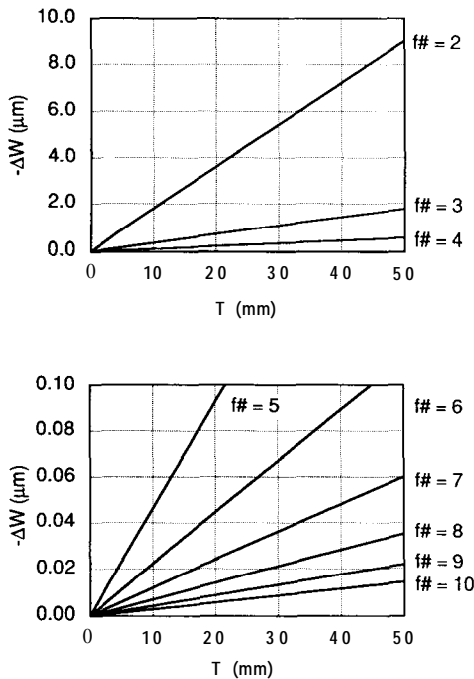


FIG. 45. Spherical aberration produced by plane parallel plate ($n = 1.5$).

inside the plate travel nearly along the normals to the surface, and we have very nearly an exact mapping of the wavefront (or rays) between the two surfaces. That is, all rays are displaced approximately the same amount.

The sagittal coma, which is equal to 1/3 the tangential coma (see Fig. 26), is given by

$$\text{Sagittal coma} \approx \frac{TU^2\bar{U}(n^2 - 1)}{2n^3} \quad (\text{third-order}). \quad (74)$$

The corresponding third-order wavefront aberration for coma with $\rho = 1$ and $x_0 = 1$ is

$$\begin{aligned} \Delta W_{\text{coma}} &= -\frac{TU^3\bar{U}(n^2 - 1)}{2n^3} \cos \theta \\ &= -\frac{T\bar{U}}{(f\#)^3} \left[\frac{(n^2 - 1)}{16n^3} \right] \cos \theta. \end{aligned} \quad (75)$$

Coma will be positive in a converging beam. Figure 46 gives four plots of the

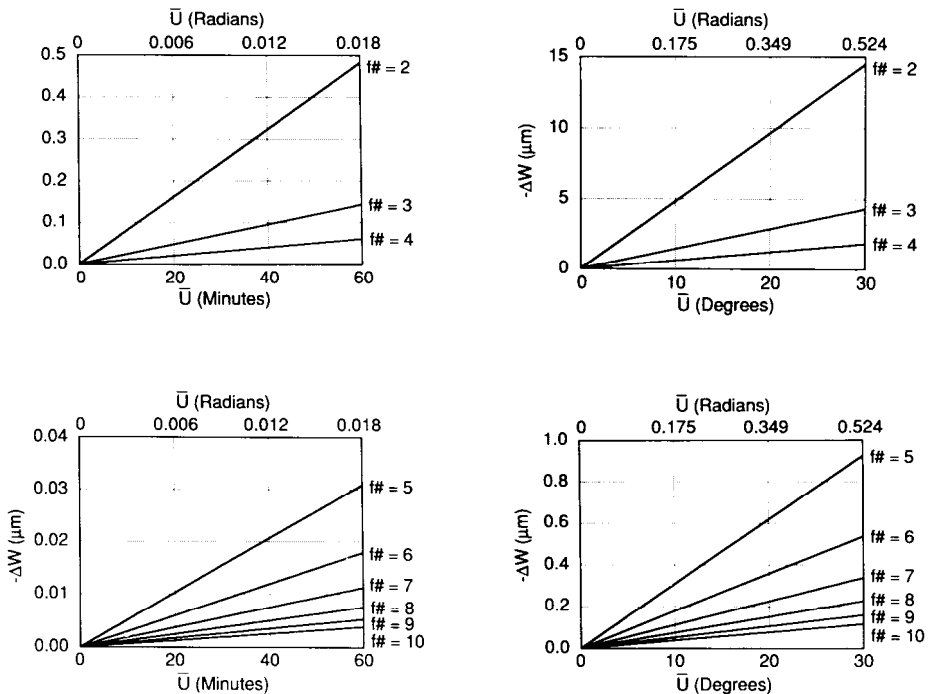


FIG. 46. Coma produced by 10 mm thick tilted plane parallel plate ($n = 1.5$).

zero to peak value of coma for a plate of refractive index 1.5, 10 mm thick, and several values of $f\#$ and \bar{U} .

The longitudinal astigmatism is given by

$$l'_s - l'_t = \frac{T}{\sqrt{n^2 - \sin^2 \bar{U}}} \left[\frac{n^2 \cos^2 \bar{U}}{n^2 - \sin^2 \bar{U}} - 1 \right] \quad (\text{exact})$$

$$\approx -\frac{T\bar{U}^2(n^2 - 1)}{n^3} \quad (\text{third-order}). \quad (76)$$

The third-order wavefront aberration for astigmatism with $\rho = 1$ and $y_0 = 1$ is

$$\Delta W_{\text{astig}} = -\frac{TU^2\bar{U}^2(n^2 - 1)}{2n^3} \cos^2 \theta$$

$$= -\frac{T\bar{U}^2}{(f\#)^2} \left[\frac{(n^2 - 1)}{8n^3} \right] \cos^2 \theta. \quad (77)$$

Figure 47 gives a plot of third-order astigmatism for a plate of index 1.5, 10 mm thick, for various values of \bar{U} and $f\#$.

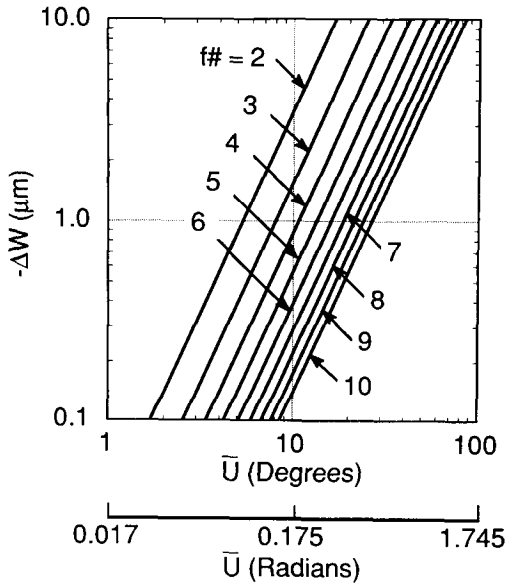


FIG. 47. Astigmatism produced by tilted plane parallel plate ($n = 1.5$)

The lateral or transverse chromatic aberration is

$$h'_F - h'_C = \frac{T\bar{U}(n_d - 1)}{n_d^2 V_d}, \quad (78)$$

where the Abbe V_d number is given by

$$V_d = \frac{n_d - 1}{n_F - n_C}. \quad (79)$$

n_d , n_F , and n_C are the indices of refraction for the helium d line (587.6 nm), the hydrogen F line (486.1 nm), and the hydrogen C line (656.3 nm), respectively. The longitudinal chromatic aberration is given by

$$l'_F - l'_C = -\frac{T(n_d - 1)}{n_d^2 V_d}. \quad (80)$$

XII. ABERRATIONS OF SIMPLE THIN LENSES

Often, it is necessary to ray trace an optical test setup to determine aberrations introduced by lenses or mirrors. In many instances, the aberration comes from a simple thin lens used at infinite conjugates with the stop at the lens, in which case the third-order aberration can be calculated using the equations given in this section. (For further information, see Welford (1986), Chapter 12.)

The third-order spherical aberration produced by a thin lens is

$$\Delta W_{\text{sph}} = \frac{h^4 \Phi^3}{32n_0} \left\{ \left(\frac{n}{n-1} \right)^2 + \frac{(n+2)}{n(n-1)^2} \left[B + \frac{2(n^2-1)}{n+2} C \right]^2 - \frac{n}{n+2} C^2 \right\}, \quad (81)$$

where

h = semidiameter of pupil,

n_o = refractive index of material surrounding lens,

n = refractive index of lens,

Φ = power = $(n - n_o)(C_1 - C_2)$,

C_1 and C_2 are curvatures of two surfaces (sign is + if center of curvature is on opposite side of incident light),

B = shape factor = $(C_1 + C_2)/(C_1 - C_2)$,

C = conjugate variable = $(U_1 + U_2)/(U_1 - U_2)$, and

U_1 and U_2 are defined in Fig. 48.

The shape factor B is equal to zero for an equi-convex lens, equals 1 for a

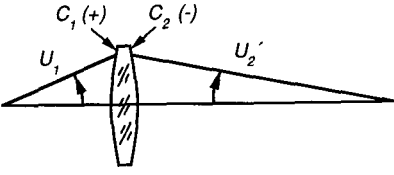


FIG. 48. Slope angles for ray incident on a lens and after refracted by lens.

plano convex lens where the light is incident upon the convex surface, and equals -1 for a plano convex lens where the light is incident upon the plano surface. For equal conjugates, the conjugate variable C equals zero. If the object is at the first principal focus $C = 1$, and if the object is at infinity $c = -1$.

If the lens is bent for minimum spherical aberration,

$$B = -\frac{2(n^2 - 1)}{n + 2} C. \quad (82)$$

If $n_0 = 1$, the minimum spherical is given by

$$\begin{aligned} \Delta W_{\text{sph-min}} &= \frac{h^4 \Phi^3}{32} \left[\left(\frac{n}{n-1} \right)^2 - \frac{n}{n+2} C^2 \right] \\ &= \frac{h}{256(f\#)^3} \left[\left(\frac{n}{n-1} \right)^2 - \frac{n}{n+2} C^2 \right]. \end{aligned} \quad (83)$$

If $C = \pm 1$ and $n = 1.5$, Eq. (83) reduces to

$$\Delta W_{\text{sph-min}} = \frac{0.033h}{(f\#)^3}. \quad (84)$$

The coma for a thin lens in air at infinite conjugates with the stop at the lens is

$$= \frac{h^3 \Phi^2 \bar{U}}{4} \left[\frac{n+1}{n(n-1)} B + \frac{2n+1}{n} C \right], \quad (85)$$

where \bar{U} is the field angle.

For minimum spherical aberration, the coma is

$$\Delta W_{\text{coma}} = \frac{h\bar{U}}{16(f\#)^2} \left(\frac{1}{n+2} \right). \quad (86)$$

For infinite conjugates, the astigmatism for a thin lens in air is given by

$$\Delta W_{\text{astig}} = \frac{1}{4} \left(\frac{h\bar{U}^2}{f\#} \right). \quad (87)$$

XIII. CONICS

A. BASIC PROPERTIES

Conics are aspherical surfaces of revolution with special properties that can be advantageous. (For further information, see Smith (1990) p. 445.) A conic is defined by the equation

$$s^2 - 2rz + (K + 1)z^2 = 0, \quad (88)$$

where the z axis is the axis of revolution, and r is the radius of curvature at the vertex of the surface. In this equation, $s^2 = x^2 + y^2$, and is not normalized. s is used in place of ρ because ρ is usually normalized. When $K = -1$,

$$z = \frac{s^2}{2r}. \quad (89)$$

Otherwise, solving for z yields

$$\begin{aligned} z &= \frac{r - \sqrt{r^2 - (K + 1)s^2}}{K + 1} \\ &= \frac{s^2/r}{1 + \sqrt{1 - (K + 1)(s/r)^2}}, \end{aligned} \quad (90)$$

where the conic constant $K = -\varepsilon^2$, and ε is the eccentricity. Equation (90) can be written as

$$\begin{aligned} z &= \frac{s^2}{2r} + \frac{(K + 1)s^4}{2^2 2! r^3} + \frac{1 \cdot 3(K + 1)^2 s^6}{2^3 3! r^5} + \frac{1 \cdot 3 \cdot 5(K + 1)^3 s^8}{2^4 4! r^7} \\ &\quad + \frac{1 \cdot 3 \cdot 5 \cdot 7(K + 1)^4 s^{10}}{2^5 5! r^9} + \dots \end{aligned} \quad (91)$$

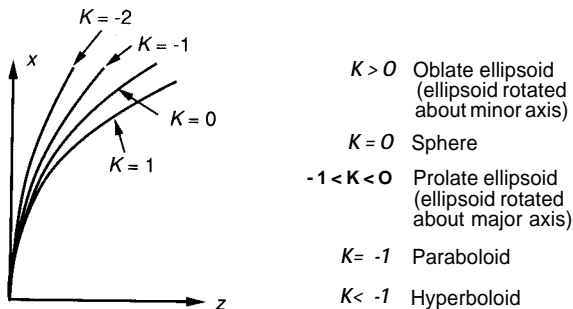


FIG. 49. Types of conics.

The type of conic is determined by K , as shown in Fig. 49.

The positions of the foci for the conic surfaces are functions of r and K as given in Eqs. (92)-(96) and illustrated in Fig. 50:

$$d_1 = \frac{r}{K + 1}, \tag{92}$$

$$d_2 = \frac{r}{K + 1} (2\sqrt{K}), \tag{93}$$

$$d_3, d_4 = \frac{r}{K + 1} (1 \pm \sqrt{-K}), \tag{94}$$

$$d_5 = \frac{r}{2}, \tag{95}$$

$$d_6, d_7 = \frac{r}{K + 1} (\sqrt{-K} \pm 1). \tag{96}$$

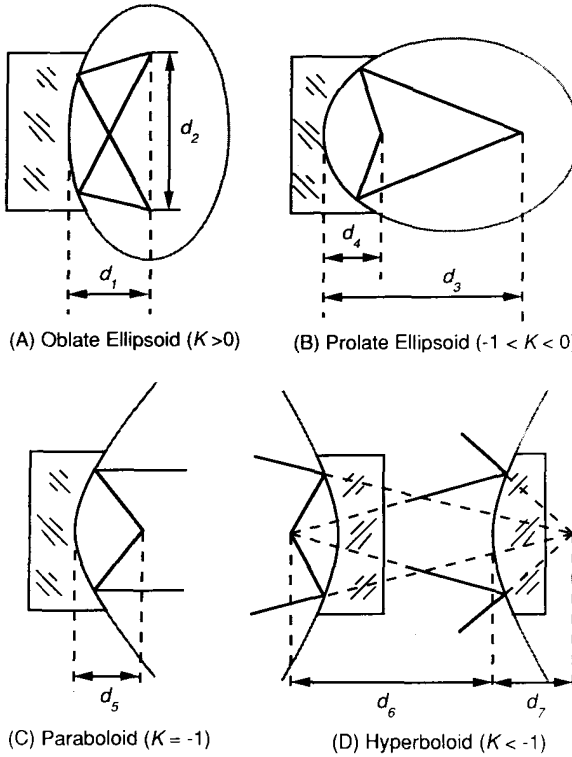


FIG.50. Foci for conic surfaces.

The difference between a conic and a sphere (i.e., departure from a sphere) is given by

$$\begin{aligned} \Delta z &= z(K) - z(0) \\ &= \frac{[(K+1) - 1]s^4}{2^2 2! r^3} + \frac{1 \cdot 3[(K+1)^2 - 1]s^6}{2^3 3! r^5} + \frac{1 \cdot 3 \cdot 5[(K+1)^3 - 1]s^8}{2^4 4! r^7} + \dots \end{aligned} \quad (97)$$

Some properties of normals to the surface of a conic can be found from Fig. 51. The angle α between the normal to the conic and the z axis is given by

$$\alpha = \tan^{-1} \left[\frac{-x}{r - (K+1)z} \right], \quad (98)$$

$$\sin \alpha = \frac{-x}{\sqrt{x^2 + [r - (K+1)z]^2}}. \quad (99)$$

The longitudinal aberration of the normal is

$$\begin{aligned} \delta z &= -Kz \quad (\text{exact}) \\ &\approx -\frac{Ks^2}{2r} - \frac{K(K+1)s^4}{8r^3} \quad (\text{third-order}), \end{aligned} \quad (100)$$

and the length of the normal \mathbf{R} is given by

$$R^2 = r^2 - Ks^2 \quad (\text{exact}). \quad (101)$$

B. SPHERICAL ABERRATION

No spherical aberration is produced if a conic is used at proper conjugates, i.e., its foci. For example, no spherical aberration is produced if a

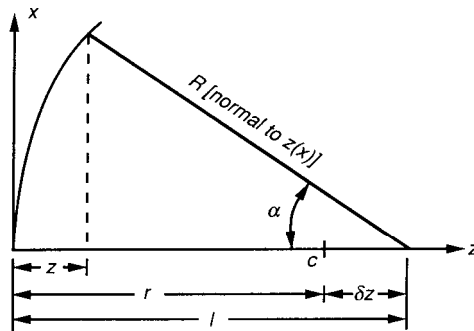


FIG. 51. Properties of normal to conic surface.

paraboloid is used with one conjugate at infinity, while spherical aberration does result if a sphere is used at infinite conjugates. The amount of spherical aberration is determined by finding the difference between the sphere and the paraboloid, and multiplying this result by 2. The spherical aberration obtained using a sphere at infinite conjugates is found from Eq. (97) to be

$$\Delta W_{\text{sph}}(\mu\text{m}) = s(\text{cm}) \left[\frac{10^4}{256(f\#)^3} + \frac{10^4}{8192(f\#)^5} + \frac{10^4}{2097152(f\#)^7} + \dots \right], \quad (102)$$

where ΔW has units of micrometers if s has units of centimeters. $f\#$ is defined as $r/2D$, where D is the mirror diameter.

This technique for finding the spherical aberration wavefront error introduced by using the mirror at the wrong conjugates can be extended to other Conics. For example, for a prolate ellipsoid, Eq. (94) gives the distance to the foci in terms of r and K :

$$\text{distance to foci} = d_3, d_4 = \frac{r}{K+1} (1 \pm \sqrt{-K}).$$

Likewise, the distance to the foci for a hyperboloid given by Eq. (96) is the negative of the right-hand side of Eq. (94). Let's say we want to find K such that the distance to the foci is Nr , where N may be positive or negative. From Eq. (94) or (96) then

$$Nr = \frac{r}{K+1} (1 \pm \sqrt{-K}). \quad (103)$$

Solving Eq. (103) for the K of the prolate ellipsoid ($-1 < K < 0$) yields

$$K = \frac{2N-1}{N^2} - 1. \quad (104)$$

Thus, if K_c is the conic constant for the conic we have, and the conic is being used at the conjugate Nr , the spherical aberration produced is given by

$$\begin{aligned} \Delta W_{\text{sph}} &= 2[z(K_c) - z(K)] \\ &= 2 \left\{ \frac{[(K_c + 1) - (K + 1)]s^4}{2^2 2! r^3} + \frac{1 \cdot 3[(K_c + 1)^2 - (K + 1)^2]s^6}{2^3 3! r^5} + \dots \right\} \end{aligned} \quad (105)$$

where K is obtained from Eq. (104).

C. COMA

The expression for third-order coma with the stop at the mirror is independent of the conic constant. If the mirror is operated at infinite

conjugates, the third-order coma is given by

$$\Delta W_{\text{coma}} = \frac{s\bar{U}}{16(f\#)^2} \cos \theta, \quad (106)$$

where as before \bar{U} is the field angle. If the mirror is used at 1:1 conjugates, the coma is zero.

D. ASTIGMATISM

If the stop is at the mirror, the third-order astigmatism is also independent of the conic constant, and will be given by

$$\Delta W_{\text{astig}} = \frac{s\bar{U}^2}{4(f\#)} \cos^2 \theta \quad (107)$$

for all conjugates.

The sagittal surface is always flat, and the tangential surface has a curvature given by $-4/r$.

XIV. GENERAL ASPHERES

Many aspheric surfaces are not conic surfaces. For an even figure of revolution about the optical axis, we can write

$$z = \frac{s^2/r}{1 + \sqrt{1 - (K + 1)(s/r)^2}} + A_4 s^4 + A_6 s^6 + \dots,$$

where the first term is the conic surface. [If the first term is nonzero, the \$A\$ term is redundant; the \$A\$ term is required for aspheric surfaces having zero power. To the level of third-order aberrations, all aspherics with power are indistinguishable from Conics. The higher order wavefront departures \(\$> 4\$ \) from a conic affect only higher order aberrations.](#)

There is an aspheric surface, commonly called a ‘‘potato chip’’ surface, which has bilateral symmetry in both x and y but not necessarily rotational symmetry. The equation of such a surface is

$$z = \frac{x^2/rX + y^2/rY}{1 + [1 - (1 + KX)(x/rX)^2 - (1 + KY)(y/rY)^2]} + AR[(1 - AP)x^2 + (1 + AP)y^2]^2 + BR[(1 - BP)x^2 + (1 + BP)y^2]^3 + CR[(1 - CP)x^2 + (1 + CP)y^2]^4 + DR[(1 - DP)x^2 + (1 + DP)y^2]^5. \quad (109)$$

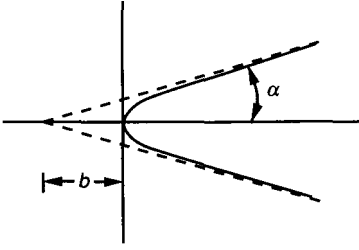


FIG. 52. Axicon surface is shown with dotted line. Hyperbolic approximation given by solid line.

rX and rY are the radii of curvature in the x and y planes, respectively, while KX and KY are the conic constants in the x and y planes.

Note that this equation reduces to Eq. (108), the equation for a standard aspheric surface, if $rX = rY$, $KX = KY$ and $AP = BP = CP = DP = 0$. Also note that if $AP = BP = CP = DP = \pm 1$, the higher order aspherizing is purely in y or x , respectively. The AP , BP , CP , and DP represent nonrotational components, while the AR , BR , CR , and DR are the rotationally symmetric portion of the higher order expressions.

Another surface of interest is the axicon. The axicon surface illustrated in Fig. 52 (dotted line) has the shape of a cone that can be represented by means of a hyperboloid with an extremely large curvature (solid line). Following the notation of Fig. 52,

$$K = -(1 + \tan^2 \alpha) < -1, \quad (110)$$

and

$$r = (K + 1)b. \quad (111)$$

For a nonlinear axicon, the profile would be that of a general asphere or a conic, often a paraboloid.

REFERENCES

- Born, M., and Wolf, E. (1964). "Principles of Optics." Pergamon Press, New York.
- Malacara, D., ed. (1978). "Optical Shop Testing." Wiley, New York.
- Smith, W. J. (1990). "Modern Optical Engineering: The Design of Optical Systems," Second Edition. McGraw-Hill, New York.
- Welford, W. T. (1974). "Aberrations of the Symmetrical Optical System." Academic Press, New York.
- Welford, W. T. (1986). "Aberrations of Optical Systems." Adam Hilger, Bristol, England.
- Wetherell, W. B. (1980). In "Applied Optics and Optical Engineering," Vol. VIII (R. R. Shannon and J. C. Wyant, eds.), pp. 171-315. Academic Press, New York.
- Zernike, F. (1934). *Physica* 1, 689.