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

1

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
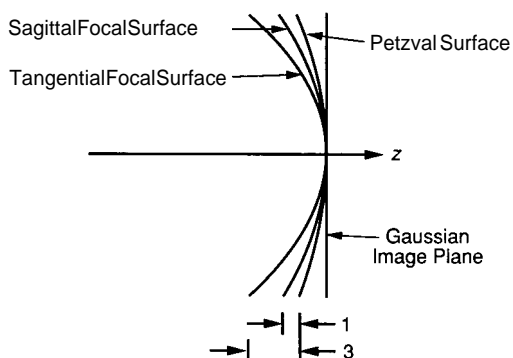

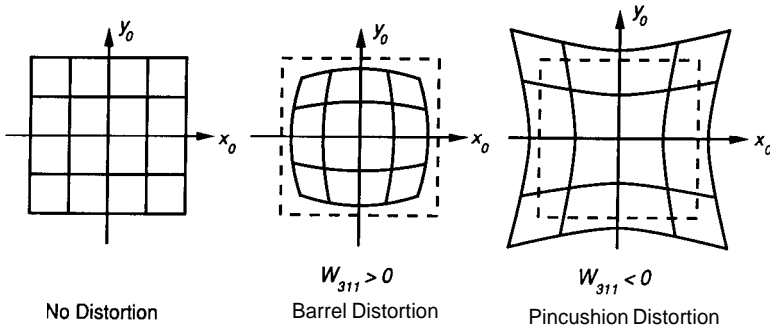FIG. 33. Focal surfaces in presence of field curvature and astigmatism.

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

$$R(\rho)G(\theta'), \tag{49}$$

where G (θ') 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,

$$G(\theta' + \alpha) = G(\theta')G(\alpha). \tag{50}$$

The set of trigonometric functions

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

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(p) 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

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

and

$$R_n^m(1) = 1. \tag{53}$$

It is convenient to factor the radial polynomial into

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

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

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

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

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

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	$\rho^5 \cos 5\theta'$
		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**

n=2   **THIRD-ORDER ABERRATIONS**

TILT

ASTIGMATISM AND DEFOCUS

FOCUS

COMA AND TILT

THIRD-ORDER SPHERICAL AND DEFOCUS

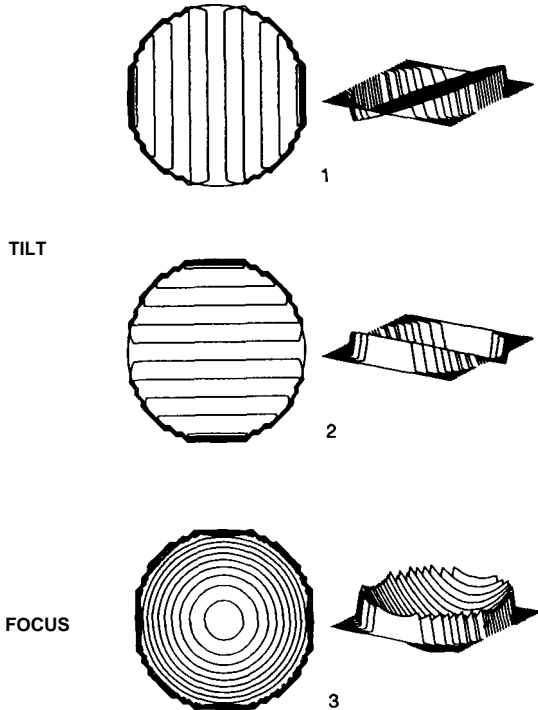**FIG. 35.** Two- and three-dimensional plots of Zernike poly-
nomials # 1 to # 3.

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

FIFTH-ORDER ABERRATIONS

n = 3

SEVENTH-ORDER ABERRATIONS

n = 4

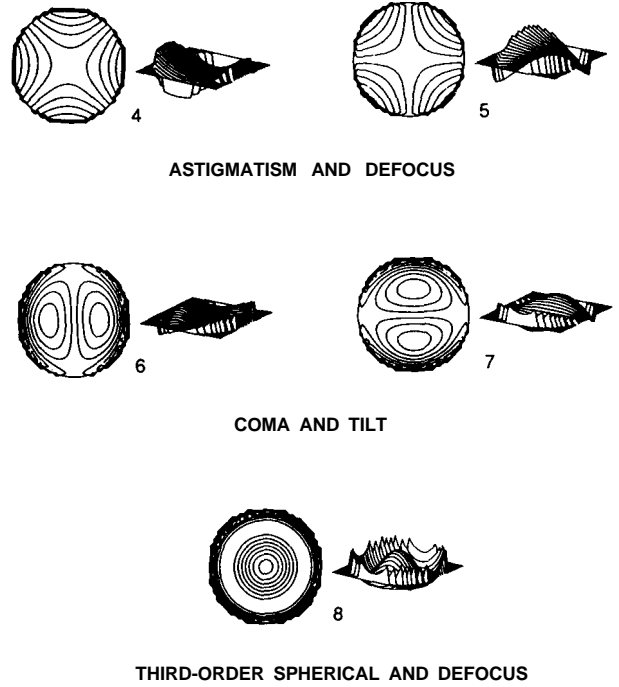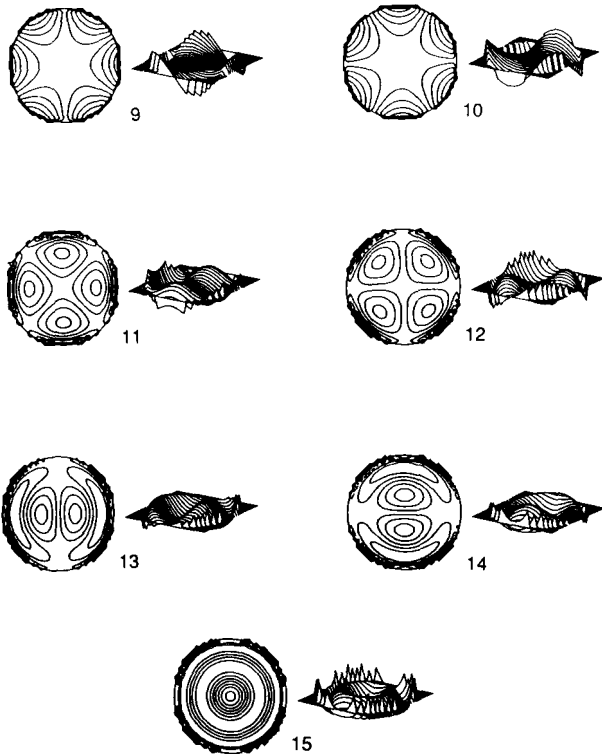FIG. 37. Two- and three-dimensional plots of Zernike polynomials # 9 to # 15.
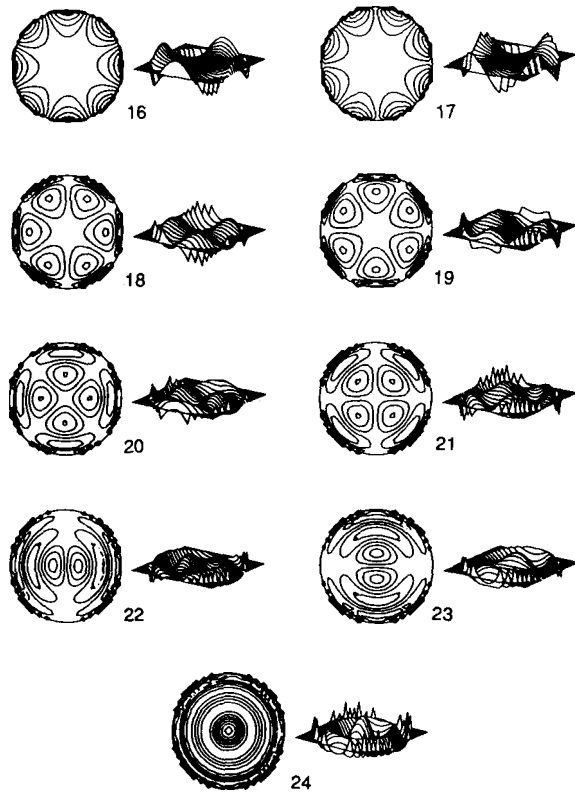
FIG. 38. Two- and three-dimensional plots of Zernike polynomials # 16 to # 24.

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:

$$W(\rho, \theta') = Z_0 - Z_3 + Z_8 \qquad \qquad \text{piston}$$

$$+ (Z_1 - 2Z_6)\rho \cos \theta' + (Z_2 - 2Z_7)\rho \sin \theta' \qquad \text{tilt}$$

$$+ (2Z_3 - 6Z_8 + Z_4 \cos 2\theta' + Z_5 \sin 2\theta')\rho^2 \qquad \text{focus + astigmatism}$$

$$+ 3(Z_6 \cos \theta' + Z_7 \sin \theta')\rho^3 \qquad \text{coma}$$

$$+ 6Z_8\rho^4. \qquad \text{spherical}$$

$$(59)$$

Equation (59) can be rearranged using the identity

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

yielding terms corresponding to field-independent wavefront aberration coefficients:

$$W(\rho, \theta') = Z_0 - Z_3 + Z_8 \qquad \qquad \text{piston}$$

$$+ \rho\sqrt{(Z_1 - 2Z_6)^2 + (Z_2 - 2Z_7)^2}$$

$$\times \cos\left[\theta' - \tan^{-1}\left(\frac{Z_2 - 2Z_7}{Z_1 - 2Z_6}\right)\right] \qquad \text{tilt}$$

$$+ \rho^2 (2Z_3 - 6Z_8 \pm \sqrt{Z_4^2 + Z_5^2}) \qquad \text{focus}$$

$$\pm 2\rho^2\sqrt{Z_4^2 + Z_5^2} \cos^2\left[\theta' - \frac{1}{2}\tan^{-1}\left(\frac{Z_5}{Z_4}\right)\right] \qquad \text{astigmatism}$$

$$+ 3\rho^3\sqrt{Z_6^2 + Z_7^2} \cos\left[\theta' - \tan^{-1}\left(\frac{Z_7}{Z_6}\right)\right] \qquad \text{coma}$$

$$+ 6\rho^4 Z_8. \qquad \text{spherical}$$

$$(61)$$

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	$\sqrt{(Z_1 - 2Z_6)^2 + (Z_2 - 2Z_7)^2}$	$\tan^{-1}\left(\dfrac{Z_2 - 2Z_7}{Z_1 - 2Z_6}\right)$
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	$\dfrac{1}{2}\tan^{-1}\left(\dfrac{Z_5}{Z_4}\right)$
W_{31}	coma	$3\sqrt{Z_6^2 + Z_7^2}$	$\tan^{-1}\left(\dfrac{Z_7}{Z_6}\right)$
W_{40}	spherical	$6Z_8$	

Note: For angle calculations, if denominator < 0, then angle $\leftarrow$ angle $+ 180^\circ$.

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.

$$\sigma^2 = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \left[\Delta W(\rho, \theta) - \overline{\Delta W}\right]^2 \rho \, d\rho \, d\theta = \overline{\Delta W^2} - \left(\overline{\Delta W}\right)^2$$

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

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

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

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

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

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

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

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

$$\text{Strehl ratio} \approx \left| 1 + i2\pi\,\overline{\Delta W} - \tfrac{1}{2}(2\pi)^2\,\overline{\Delta W^2} \right|^2$$

$$\approx 1 - (2\pi)^2[\overline{\Delta W^2} - (\overline{\Delta W})^2]$$

$$\approx 1 - (2\pi\sigma)^2, \quad (66)$$

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

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

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.