UCRL-CONF-213323

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Calibration of symmetric and non-symmetric errors for interferometry of ultra-precise imaging systems

Donald W Phillion, Gary E Sommargren, Michael A Johnson, Todd A Decker, Yoshio Gomei, Osamu Kakuchi, Seiji Takeuchi

June 29, 2005

SPIE Annual Meeting
San Diego, CA, United States
July 31, 2005 through August 4, 2005

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Calibration of symmetric and non-symmetric errors for interferometry of ultra-precise imaging systems

Donald W. Phillion[a], Gary E. Sommargren[a], Michael A. Johnson[a], Todd A. Decker[a], John S. Taylor[a], Yoshio Gomei[b], Osamu Kakuchi[b], Seiji Takeuchi[b]

[a]Lawrence Livermore National Laboratory, P. O. Box 808, Livermore, CA, USA 94550-0808
[b]Canon Inc., 23-10 Kiyohara-Kogyo-Danchi, Utsunomiya-Shi, Tochigi 321-3298, Japan

ABSTRACT

The azimuthal Zernike coefficients for shells of Zernike functions with shell numbers n<N may be determined by making measurements at N equally spaced rotational positions. However, these measurements do not determine the coefficients of any of the purely radial Zernike functions. Label the circle that the azimuthal Zernikes are measured in as circle A. Suppose that the azimuthal Zernike coefficients for n<N are also measured in a smaller circle B which is inside circle A but offset so that it is tangent to circle A and so that it has the center of circle A just inside its circular boundary. The diameter of circle B is thus only slightly larger than half the diameter of circle A. From these two sets of measurements, all the Zernike coefficients may be determined for n<N. However, there are usually unknown small rigid body motions of the optic between measurements. Then all the Zernike coefficients for n<N except for piston, tilts, and focus may be determined. We describe the exact mathematical algorithm that does this and describe an interferometer which measures the complete wavefront from pinholes in pinhole aligners. These pinhole aligners are self-contained units which include a fiber optic, focusing optics, and a "pinhole mirror". These pinhole aligners can then be used in another interferometer so that its errors would then be known. Physically, the measurements in circles A and B are accomplished by rotating each pinhole aligner about an aligned axis, then about an oblique axis. Absolute measurement accuracies better than 0.2 nm were achieved.

Keywords: interferometer, wavefront measurement, phase-shifting interferometry

1. INTRODUCTION

Visible light interferometry has been used to both to characterize individual EUVL mirrors and to characterize entire EUVL projection optical systems[1, 2]. For EUVL projection optical systems that will be used at the 32-nm lithography node, allowable wavefront errors will be at the one angstrom level. Here we describe a point-source-diffraction, phase-shifting visible light interferometer (PSDI) that is absolutely accurate to better than two angstroms after correcting for wavefront errors. It is designed to characterize a 0.3 NA EUVL projection optical system that is being developed by Canon Inc.

The PSDI achieves its extreme accuracy by eliminating all conventional optics[1]. A nearly spherical measurement wave is emitted from a pinhole at the object plane and is relayed by the optical system under test to a focus on a reflecting mirror with a pinhole nearby in the reflecting surface. This pinhole emits the nearly spherical reference wave which interferes with the reflected measurement wave to produce the fringe pattern which is measured by a CCD camera. Phase-shifting interferometry determines the complete complex wavefront from a set of intensity frames with different phase shifts. Phase-shifting interferometry is capable of measuring extremely small fractions of a wavelength. This complex wavefront is numerically back-propagated to the aperture plane.

In order to make the light from a pinhole be a nearly perfect spherical wave, the pinhole must be made small and extremely round. It must be illuminated by circular polarized light that is focused onto the back side of the pinhole with a spot diameter that is much larger than the pinhole diameter. There are tight tolerances on the allowable aberrations in the relay optics. However, even this nearly perfect spherical wave still isn't perfect enough and must be accurately measured to calibrate its deviation from sphericity.

The portable package of fiber optic connector, relay lenses, and pinhole mirror is called either a "wavefront reference source" (WRS) or a "pinhole aligner" (PA)[3]. Its wavefront is measured on what is called the Canon Test Stand Interferometer, which uses a complex multi-element lens to replace the EUVL projection optical system. The wavefronts of two wavefront reference sources must be measured since one is needed to emit the nearly spherical measurement wave and the other is needed both to reflect the focused measurement wave and to emit the nearly spherical reference wave.

The azimuthally varying Zernike terms for a WRS wavefront may be determined by making measurements at different rotation angles about the optical axis. However, the axisymmetric purely radial terms will be totally undetermined by this method. It turns out that by making additional measurements while rotating about an oblique axis through the pinhole, that the axisymmetric terms may be exactly measured as well in principle.

2. DESCRIPTION OF THE ALGORITHM

Suppose one is measuring a test optic in an interferometer with unknown errors. Let circle A be the test optic's circular aperture. By making measurements at a set of N_{rot} equally spaced rotational positions for the test optic, the azimuthally varying Zernike coefficients for all values of m that are not multiples of N may be determined. This is done by rotating all the measurements to a common orientation and taking their average. All the Zernike functions have either a $\cos(m\phi)$ or $\sin(m\phi)$ azimuthal dependence multiplied by a purely radial function. The average of either of these functions rotated to N_{rot} equally spaced rotational positions is zero unless m is a multiple of N_{rot}, in which case the rotational average is the original function. This averaging procedure averages out the components of the interferometer error for which m is not a multiple of N_{rot}. In practice there are unknown small rigid body motions other than the rotations between the measurements, and thus piston, tilts, and focus cannot be determined.

Note that the Zernike functions having a $\cos(m\phi)$ azimuthal dependence with m=0 are not measured since zero is the zero multiple of N_{rot}. Thus the coefficients of the purely radial Zernikes cannot be determined by this method. However, both the radial Zernike coefficients and the azimuthally varying Zernikes coefficients for the test optic's wavefront can be exactly determined by making measurements in two circles with the smaller circle offset with respect to the larger circle but still contained within the larger circle. How this is done is the subject of the rest of this section.

Let circle A be the larger circle and let circle B be the smaller circle that is contained within circle A. The figure shows the geometry. Suppose that we make N_{rot} equally spaced measurements about the center of circle A and that we also make N_{rot} equally spaced measurements about the center of circle B. By the averaging procedure just described we can obtain the coefficients of the azimuthally varying Zernikes in circle A for which m is not a multiple of N_{rot}. These coefficients are for a set of Zernike functions which have the center of circle A as their center and the radius of circle A as their normalizing pupil radius. Similarly, we can obtain the coefficients of the azimuthally varying Zernikes in circle B for which m is not a multiple of N_{rot}. These coefficients are for a set of Zernike functions which have the center of circle B as their center and the radius of circle B as their normalizing pupil radius. From these two sets of measurements we wish to determine the coefficients of the purely radial Zernikes in circle A.
There are various normalizations and orderings for the circular Zernike functions in the literature[4, 5, 6]. We use what are called the FRINGE Zernike polynomials that were originally developed at the University of Arizona[6]. This set of polynomials, possibly normalized, has increasingly become the standard. They are also the most natural set (appendix A). Within each n shell, the ordering is m=n, m=-n, …, m=1, m=-1, m=0. Positive values of m denote Zernike functions whose azimuthal dependence is $\cos(m\phi)$ and negative values of m denote Zernike functions whose azimuthal dependence is $\sin(|m|\phi)$.

We work with Zernike shells $n < N_{rot}$. In the recursion relations (appendix A), the Zernike functions are partitioned into shells. The number of Zernikes in shells 0, 1, …, n is $(n+1)^2$. Within each shell, m varies from –n to +n. The convention is that minus values of m denote Zernikes having a $\sin(m\phi)$ azimuthal dependence while positive values of m denote Zernikes having a $\cos(m\phi)$ azimuthal dependence. The n=0 shell consists only of piston and the n=1 shell consists of x and y tilts and focus. The reason for the concern with shells is that a Zernike function in one coordinate system can always be expressed in terms of Zernike functions in another coordinate system in the same or lower shells. This property gives closure to the problem. If it happens that the wavefront can be exactly represented by Zernikes in

shells with n lower than N_{rot}, then the mathematical algorithm to be described here will exactly recover these Zernike coefficients in the absence of noise.

Appendices A and B describe the mathematics necessary to transform from one Zernike representation to another. The closure property is also proved in these appendices. Although the transformations in these appendices are not directly used in the algorithm to be described here, they were used to test this algorithm.

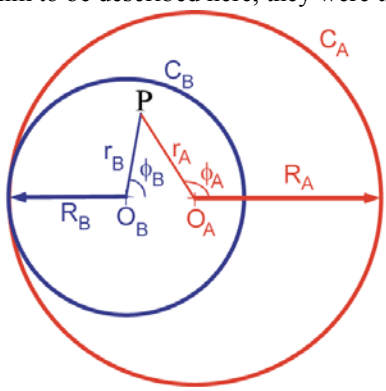

Figure 1 Geometry for the determination of the azimuthally symmetric and non-symmetric errors by making measurements in two circles, one in circle A and the other in circle B, which is off center with respect to circle A but is contained in circle A.

We are given the non-radial Zernike coefficients for circle A about its origin and with its pupil radius and for circle B about its origin and for its pupil radius. If the unknown radial Zernike terms are included, the two Zernike expansions must agree in their overlap region, which is all of circle B. We thus have:

$$\sum_{n=0}^{N}\sum_{\substack{m=-n\\ m\neq 0}}^{n} c_{nm}^{(A)} Z_{n,m}\left(\frac{r_A}{R_A},\phi_A\right)+\sum_{n=0}^{N} c_{n0}^{(A)} Z_{n,0}\left(\frac{r_A}{R_A},\phi_A\right)= \sum_{n=0}^{N}\sum_{\substack{m=-n\\ m\neq 0}}^{n} c_{nm}^{(B)} Z_{n,m}\left(\frac{r_B}{R_B},\phi_B\right)+\sum_{n=0}^{N} c_{n0}^{(B)} Z_{n,0}\left(\frac{r_B}{R_B},\phi_B\right) \tag{1}$$

We may rewrite this as

$$\sum_{n=0}^{N} c_{n0}^{(A)} Z_{n,0}\left(\frac{r_A}{R_A},\phi_A\right)-\sum_{n=0}^{N} c_{n0}^{(B)} Z_{n,0}\left(\frac{r_B}{R_B},\phi_B\right)= \sum_{n=0}^{N}\sum_{\substack{m=-n\\ m\neq 0}}^{n} c_{nm}^{(B)} Z_{n,m}\left(\frac{r_B}{R_B},\phi_B\right)-\sum_{n=0}^{N}\sum_{\substack{m=-n\\ m\neq 0}}^{n} c_{nm}^{(A)} Z_{n,m}\left(\frac{r_A}{R_A},\phi_A\right) \tag{2}$$

where the unknown terms are now on the left and the known terms are now on the right. The upper limit for the shell number n in the sums is $N=N_{rot}-1$. We might think to use least squares to solve for the unknown coefficients but there is one complication that must be addressed first. There may be piston, tilts and focus between the A and B measurements. Thus we really wish to do least squares with the following expansion:

$$\sum_{n=2}^{N} c_{n0}^{(A)} Z_{n,0}\left(\frac{r_A}{R_A},\phi_A\right)-\sum_{n=2}^{N} c_{n0}^{(B)} Z_{n,0}\left(\frac{r_B}{R_B},\phi_B\right)+f_1+f_2\,x_A+f_2\,y_A+f_3\left(2r_A^{\,2}-1\right)= \sum_{n=2}^{N}\sum_{\substack{m=-n\\ m\neq 0}}^{n} c_{nm}^{(B)} Z_{n,m}\left(\frac{r_B}{R_B},\phi_B\right)-\sum_{n=2}^{N}\sum_{\substack{m=-n\\ m\neq 0}}^{n} c_{nm}^{(A)} Z_{n,m}\left(\frac{r_A}{R_A},\phi_A\right) \tag{3}$$

Note that only the Zernike terms for n>=2 have been kept. The least squares functions are the piston, the x tilt, the y tilt, the focus and the coefficients of the radial Zernikes for n>=2 for both circle A and circle B. The least squares functions are on the left hand side of the equation.

3. TESTING OF THE ALGORITHM

The mathematical algorithm has been tested by assuming a Zernike expansion in circle A, then finding the Zernike expansion in circle B using the translation and scaling transformations described in the appendices, then by setting the coefficients in both expansions for the radial Zernikes and the x and y tilts to zero, and then by doing least squares fitting in circle B. The procedure is exact when there is no noise. The least squares fitting gives the radial Zernike coefficients for both circles A and B.

Let $f_i(x)$ be a set of functions for which the equation $\sum c_i f_i(x_k) = g(x_k)$ is to be least squares fit. Here x_k is a set of (x,y) points on a grid filling circle B, the $f_i(x)$ are the functions on the left hand side of Eq (7), and g(x) is the right hand side of Eq (3). The denser the grid the more accurate the result. We find that a 100x100 square grid masked to circle B is more than fine enough. This can be cast as the matrix equation

$$\sum_j A_{ij} c_j = B_i \qquad \text{where} \qquad A_{ij} = \sum_k f_i(x_k) f_j(x_k) \qquad \text{and} \qquad B_i = \sum_k f_i(x_k) g(x_k)$$

Since g(x) is linear in the non-radial Zernike coefficients for $n \geq 2$ for circles A and B, so are the coefficients in the solution vector c_i. This solution vector gives the radial Zernike coefficients for $n \geq 2$ for circles A and B and the coefficients of the piston, tilts, and focus terms on the left-hand side of Eq (7). These last coefficients are of course of no physical interest.

We next present a numerical example to demonstrate the numerical well-behavedness of this method. Define the circle A coordinate system so that circle A is centered at the origin and has unity pupil radius. Let circle B have origin $(X_{0B},Y_{0B})=(0.33,0)$ and pupil radius $R_{pupilB}=0.5$ in this circle A coordinate system. Choose an arbitrary set of circle A coefficients c_A and solve for the circle B Zernike coefficients c_B using the appendices to obtain the below table.

N	n	m	c_A	c_B	Zernike
1	0	0	+1.00	+0.780591445	1
2	1	1	+0.20	-0.566567000	X
3	1	-1	+0.30	+0.150000000	Y
4	1	0	+0.50	-0.051850000	$2 R^2-1$
5	2	2	-1.00	-0.004975000	X^2-Y^2
6	2	-2	+0.00	+0.000000000	2 X Y
7	2	1	+0.00	0.2475000000	$(3 R^2-2) X$
8	2	-1	+0.00	+0.000000000	$(3 R^2-2) Y$
9	2	0	+0.75	+0.0468750000	$6 R^4 - 6 R^2 + 1$

Table 1 Numerical example for the algorithm. The table gives the Zernike coefficients in the circle A and B coordinate systems. The circle B Zernike coefficients are derived from the circle A coefficients, which have been arbitrarily chosen.

If we zero the coefficients for the tilts and all the purely radial Zernikes and use least squares as described in section II, we exactly recover the purely radial Zernike coefficients for both circles A and B except for piston and focus. We also get the error sensitivity matrix. We have studied the sensitivity of the solved-for radial Zernike coefficients for measurement errors in any quantity including the individual coefficients for the azimuthally varying Zernikes for either circle A or circle B and including errors in the origin or pupil radius of circle B. We find the algorithm to be numerically robust. This algorithm has been extensively tested with different Zernike coefficients, different circle B origins and pupil radii, and different maximum Zernike shell numbers.

4. WAVEFRONT REFERENCE SOURCE CALIBRATIONS

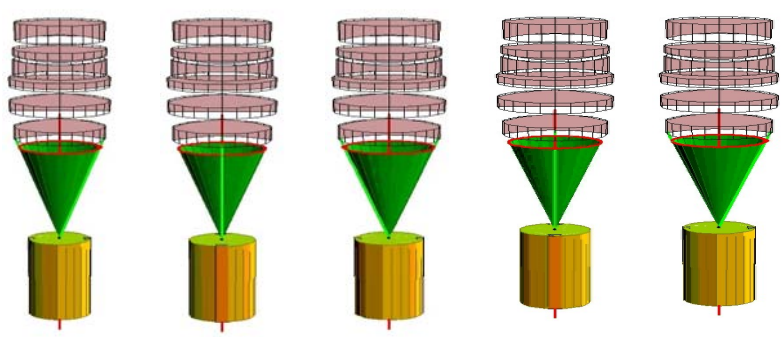

Figure 2 Rotation of the WRS about the aligned axis which is the optical axis of the Canon test lens. This axis passes through the pinhole. Four equally spaced rotations were performed. The last rotation on the far right returns the WRS to its original orientation.

Three WRS wavefronts were measured using N_{rot}=4. For each, four rotations were done about the aligned axis (figure 2) and four rotations were done about the oblique axis (figure 3). The wavefront reference source under test is the bottom WRS in the Canon Test Lens Interferometer, which is shown in figure 4. The outer vacuum vessel is shown cut away.

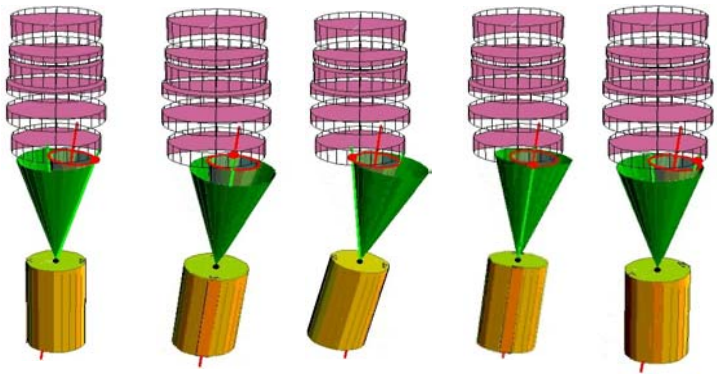

Figure 3 Rotation of the WRS about the oblique axis. This axis passes through the pinhole but is at an angle to the optical axis of the test lens. Four equally spaced rotations were performed. The last rotation on the far right returns the WRS to its original orientation.

The bottom WRS provides the measurement wave which is the phase-shifted wave in this phase-shifting, point-source-diffracting interferometer (PSDI). The complex multi-element test lens (center) focuses the light from the bottom pinhole to a spot near the top WRS's pinhole, which emits the reference wave. The measurement and reference waves interfere at the CCD camera, forming a fringe pattern with nearly linear fringes.

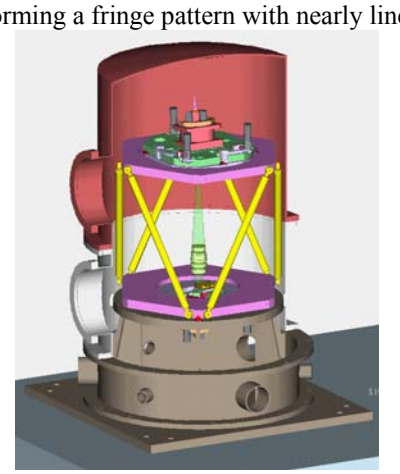

Figure 4 Canon Test Lens Interferometer

The top WRS is inclined at a small angle so that the reflected light may be directed onto a CCD chip which is out of the way of the incident measurement wave. The CCD chip is remote from most of the electronics for the CCD camera and is mounted at the end of a small carrier board which has minimal electronics. The carrier board is only a little wider than the CCD chip. In this way, minimal clearance is needed and the top WRS needs to be inclined only at a small angle.

The twelve-frame, $\pi/4$ phase-shifting interferometry phase-shifting algorithm[7] enables both the phase and amplitude to be determined. The particular algorithm that is used reduces the errors caused by the PZT voltage ramp having the wrong slope or being curved and reduces the errors caused by a linear variation in time of the background light or the laser power. The algorithm is also insensitive to CCD power law nonlinearities up to and including the sixth power. Numerical back-propagation is used to propagate the wavefront from the CCD camera to the aperture plane at the bottom of the test lens.

For each rotational position of the bottom WRS, the average of a set of measurements over a "tilt grid" is done. A "tilt grid" is a set of focus points on the top WRS mirror for the measurement wave. The focus point is moved with respect to the pinhole by translating the top WRS. The set of points in a "tilt grid" is always chosen to have inversion symmetry with respect to the top WRS's pinhole. This property makes the errors due to distortion average out. A tilt grid also provides averaging over the roughness of the reflecting surface and provides averaging for the fringe print-through.

Each WRS provides rotation about its own axis by using inner and outer cylindrical shells which may be rotated with respect to each other using high precision bearings. This gives the rotation about the "aligned" axis. In addition, the bottom WRS is mounted on an inclined rotation stage whose rotation axis passes through the pinhole. There is a wedge block between the bottom WRS and the rotation stage. This wedge has the same wedge angle as the rotation stage's inclination angle. At the 0° rotation angle, the body z axis of the WRS coincides with the optical axis of the test lens. This gives the rotation about the "oblique" axis.

The use of circular polarization is crucial to the algorithm. The measurement wave diffracted from the bottom WRS pinhole is RH circular while the reference wave diffracted from the top WRS pinhole is LH circular. After the measurement wave is reflected from the top WRS's reflecting surface, it is LH circular polarized. If a linear polarization were used that was fixed with respect to the interferometer, then it would no longer necessarily be true that the wavefront errors from the bottom WRS would rotate with the bottom WRS. If a linear polarization were used that was fixed with respected to the bottom WRS so that the direction of linear polarization rotated with the bottom WRS, then it would no longer necessarily be true that the interferometer errors would be fixed.

For all the wavefront reference sources, the principal wavefront error was astigmatism. The WRS wavefronts had about 0.2 to 0.4 nm of aberration which could be measured with about 0.1 to 0.15 nm reproducibility (one sigma with respect to the mean). The results of several measurement sets were averaged to obtain each WRS's wavefront.

5. SUMMARY

By rotating about an oblique axis as well as an aligned axis, the complete wavefront of the wavefront reference sources could be measured. By calibrating the wavefront for each WRS, an absolute accuracy of 0.2 nm can be achieved for the interferometer.

REFERENCES

1. G. E. Sommargren, et. al., *100–picometer interferometry for EUVL*, in Proc. SPIE, (**4688**) 316 (2002)
2. D. A. Tichenor, et. al., *System Integration and Performance of the EUVL Engineering Test Stand*, Emerging Lithographic Technologies V, Proc. SPIE (**4243**) 19 (2001)
3. M. J. Johnson, et. al., *Construction and testing of wavefront reference sources for interferometry of ultra-precise imaging systems*, in Proc. SPIE, August meeting, 2005.
4. Max Born and Emil Wolf, *Principles of Optics, 7th (expanded) edition*, Cambridge University Press (1999), Appendix VII.
5. Daniel Malacara, *Optics Shop Testing*, John Wiley & Sons, Inc. (New York, 1992), sections 13.2.2 and 13.2.3 .
6. Optical Research Associates, *Code V Reference Manual Version 9.5 Volume II*, pp 10-21 to 10-22.
7. D. W. Phillion, *General methods for generating phase-shifting interferometry algorithms*, Appl. Opt. **36** (31) 8098 (1997).

Appendix A Translational transformation of the Zernike functions

Since the Zernike functions are a complete set, a Zernike function in one coordinate system can be expressed in terms of the Zernike functions in any other coordinate system. Here we wish to consider only a change of origin or a translation. Scaling transformations due to changes in the normalizing radius are discussed in appendix B, while rotational transformations are trivial. The Zernike functions have the form $Z_{nm}(\rho,\phi) = f_{nm}(\rho^2)\ \rho^m \cos m\phi$ with $n \geq m$ for $m \geq 0$ and $f_{n|m|}\ \rho^{|m|} \sin |m|\phi$ with $n \geq |m|$ for $m < 0$. It is more convenient to work with $N = n-|m|$ rather than n since the functions $f_N^{(m)}(x)$ are of degree N in $x = \rho^2$ and are defined for $N \geq 0$. The Rodrigues formula for these functions also has a simpler form. For each m, the functions $f_N^{(m)}(x)$ form a complete set over x. The Zernike functions are defined to be orthogonal on the unit circle with the normalization condition $f_N^{(m)}(1) = 1$. The factor $f_N^{(m)}(\rho^2)$ of the Zernike

function $Z_{nm}(\rho,\phi)$ is actually a Jacobi polynomial whose argument has been scaled and shifted. A large literature on the properties of Jacobi polynomials exists. Note that this function does not contain all the radial dependence of the corresponding Zernike function except when m=0. Any polynomial in X and Y can be represented as a sum of Zernike functions since they can be shown to form a complete set of functions in the XY plane.

The reason that the factor $\rho^{|m|}$ exists and that the remainder of the radial dependence depends only upon ρ^2 can be seen as follows: only terms in X^kY^l with $k+l \geq m \geq 0$ can give rise to terms in cos $m\phi$ and sin $m\phi$. Thus if the azimuthal dependence is either cos $m\phi$ or sin $m\phi$, there must be a ρ^m factor. The polynomial in ρ multiplying ρ^mcos $m\phi$ or ρ^msin $m\phi$ must then depend only upon ρ^2 for we require invariance under the identity transformation $\rho \to -\rho$ and $\phi \to 180°+\phi$, which maps a point into itself.
The generators for the translations in X and Y are $\partial/\partial X$ and $\partial/\partial Y$, respectively, where X and Y are the normalized x and y coordinates. We wish to find expansions for $\partial/\partial X\ Z_{nm}(\rho,\phi)$ and $\partial/\partial Y\ Z_{nm}(\rho,\phi)$ in terms of the $Z_{nm}(\rho,\phi)$. Define $x = \rho^2$. Orthogonality on the unit circle requires that:

$$\int_0^1 f_N^{(m)}(x) f_{N'}^{(m)}(x) x^m dx = 0 \ for\ N \neq N' \tag{A.1}$$

These are orthogonal polynomials on the unit interval [0,1] with weighting factor x^m. This property completely determines these polynomials except for their normalization. It also determines the differential equation they obey and their generating function. Using the theory of orthogonal polynomials, they can be shown to obey the differential equation:

$$\left\{ x(1-x)\frac{d^2}{dx^2} + \left[(m+1)-(m+2)x\right]\frac{d}{dx} + \lambda_{Nm} \right\} f_N^{(m)}(x) = 0 \tag{A.2}$$

where the eigenvalue $\lambda_{Nm} = N^2+(m+1)\ N$.

The theory of orthogonal polynomials starts with a differential equation of the form:

$$\left(p(x)\frac{d^2}{dx^2} + q(x)\frac{d}{dx} + \lambda_n \right) f_n(x) = 0 \tag{A.3}$$

Here $p(x) = p_2x^2 + p_1x + p_0$ is a polynomial of no higher than the second degree and $q(x) = q_1x + q_0$ is a polynomial of no higher than the first degree. It may be proved, except for normalization, that:

$$f_n(x) = \frac{1}{w(x)}\frac{d^n}{dx^n}\left[w(x)p^n(x)\right] \tag{A.4}$$

where

$$w(x) = \frac{1}{p(x)}\exp\left[\int \frac{q(x)}{p(x)}dx\right] \tag{A.5}$$

It may also be proved that the eigenvalue λ_n is determined by $p_2n(n-1)+q_1n+\lambda_n=0$. From Sturm-Liouville theory, orthogonality on an interval [a,b] with weighting factor w(x) requires that

$$e^{\int \frac{q(x)}{p(x)}dx} W(f_n, f_m)\Big|_a^b = 0 \tag{A.6}$$

where $W(f_n,f_m)$ is the Wronskian of $f_n(x)$, and $f_m(x)$. Since $w(x)=x^m$ for $f_N^{(m)}(x)$:

$$\exp\left[\int \frac{q(x)}{p(x)}dx\right] = x^m p(x) \tag{A.7}$$

If p(x) has zeros at 0 and 1, then the polynomials will be orthogonal on [0,1] with weighting factor x^m for $m \geq 0$. Thus p(x) has the form x(1-x) up to a factor. The polynomial q(x) may be determined so as to make x^m the weighting factor

The functions $f_N^{(m)}(x)$ are given by the Rodrigues formula:

$$f_N^{(m)}(x) = \frac{1}{N!x^m}\frac{d^N}{dx^N}\left[x^{N+m}(x-1)^N\right] \tag{A.8}$$

They obey the orthogonality relation:

$$\int_0^1 f_N^{(m)}(x)\, f_{N'}^{(m)}(x)\, x^m\, dx = \frac{1}{2N+|m|+1}\delta_{N,N'} \tag{A.9}$$

Now we are ready to find the expansion for the gradient of a Zernike function in terms of the Zernike functions. Since X = ρ cos ϕ and Y = ρ sin ϕ, we have:

$$\frac{\partial}{\partial X} = \cos\phi\frac{\partial}{\partial\rho} - \frac{\sin\phi}{\rho}\frac{\partial}{\partial\phi}$$
$$\frac{\partial}{\partial Y} = \sin\phi\frac{\partial}{\partial\rho} + \frac{\cos\phi}{\rho}\frac{\partial}{\partial\phi} \tag{A.10}$$

Thus:

$$\frac{\partial}{\partial X}f(\rho^2)\cos m\phi = f'(\rho^2)\rho^{m+1}\cos(m+1)\phi + \left[\rho^2 f'(\rho^2) + mf(\rho^2)\right]\cos(m-1)\phi$$
$$\frac{\partial}{\partial X}f(\rho^2)\sin m\phi = f'(\rho^2)\rho^{m+1}\sin(m+1)\phi + \left[\rho^2 f'(\rho^2) + mf(\rho^2)\right]\sin(m-1)\phi$$
$$\frac{\partial}{\partial Y}f(\rho^2)\cos m\phi = f'(\rho^2)\rho^{m+1}\sin(m+1)\phi - \left[\rho^2 f'(\rho^2) + mf(\rho^2)\right]\sin(m-1)\phi$$
$$\frac{\partial}{\partial Y}f(\rho^2)\sin m\phi = -f'(\rho^2)\rho^{m+1}\cos(m+1)\phi + \left[\rho^2 f'(\rho^2) + mf(\rho^2)\right]\cos(m-1)\phi$$

Here $f(\rho^2)$ represents $f_N^{(m)}(\rho^2)$. We wish to expand $\frac{d}{dx}f_N^{(m)}(x)$ in terms of $f_{N'}^{(m+1)}(x)$ with varying N' and to expand $\frac{d}{dx}f_N^{(m)}(x) + mf_N^{(m)}(x)$ in terms of $f_{N'}^{(m-1)}(x)$ with varying N'. Note that there is a problem here when m=0 since $f_N^{(m)}(x)$ is not defined for negative m. For the m=0 case, only the cos mϕ functions exist and we only need $\frac{d}{dx}f_N^{(0)}(x)$ in terms of $f_N^{(1)}(x)$.

We will prove that for $m \geq 0$ and $N \geq 0$

$$\frac{d}{dx} f_N^{(m)}(x) = \begin{array}{l} [2(N-1)+m+2] f_{N-1}^{(m+1)}(x) \\ [2(N-2)+m+2] f_{N-2}^{(m+1)}(x) \\ \vdots \\ [2(N-N)+m+2] f_0^{(m+1)}(x) \end{array} \tag{A.12}$$

We will also prove that for m > 0 and N ≥ 0:

$$x\frac{d}{dx} f_N^{(m)}(x) + m f_N^{(m)}(x) = \begin{array}{l} [2(N-0)+m] f_N^{(m-1)}(x) \\ [2(N-1)+m] f_{N-1}^{(m-1)}(x) \\ \vdots \\ [2(N-N)+m] f_0^{(m-1)}(x) \end{array} \tag{A.13}$$

The expressions $\frac{d}{dx} f_N^{(m)}(x)$ and $x\frac{d}{dx} f_N^{(m)}(x) + m f_N^{(m)}(x)$ are polynomials of degree N-1 and N, respectively. The expansion (A.12) is for m ≥ 0 while the expansion (A.13) is for m > 0. The coefficients in these two expansions are given by:

$$\frac{\int_0^1 \left(\frac{d}{dx} f_N^{(m)}(x) \right) f_{N'}^{(m+1)}(x) x^{m+1} dx}{\int_0^1 f_{N'}^{(m+1)}(x) f_{N'}^{(m+1)}(x) x^{m+1} dx}$$

and

$$\frac{\int_0^1 \left(x\frac{d}{dx} f_N^{(m)}(x) + f_N^{(m)}(x) \right) f_{N'}^{(m-1)}(x) x^{m-1} dx}{\int_0^1 f_{N'}^{(m-1)}(x) f_{N'}^{(m-1)}(x) x^{m-1} dx}$$

The first integral may be easily evaluated by using the Rodrigues formula formula given earlier and integrating repeatedly by parts. The same approach may be used on the second integral after making the substitution

$$x\frac{d}{dx} f_N^{(m)}(x) + f_N^{(m)}(x) = \frac{1}{x^{m-1}} \frac{d}{dx} \left[x^m f_N^{(m)}(x) \right] \tag{A.14}$$

The explicit equations for the expansion of the gradient of a Zernike function when m ≠ 0 are:

$$\begin{aligned} &\frac{\partial}{\partial X} Z_{n,\pm m}(\rho,\phi) = \\ &+ \left\{ [2n-m] Z_{n,\pm(m+1)} + [2(n-1)-m] Z_{n-1,\pm(m+1)} + \cdots + [2(m+1)-m] Z_{m+1,\pm(m+1)} \right\} \\ &+ \left\{ [2(n-1)-m+2] Z_{n-1,\pm(m-1)} + [2(n-2)-m+2] Z_{n-2,\pm(m-1)} \right. \\ &\left. + \cdots + [2(m-1)-m+2] Z_{m-1,\pm(m-1)} \right\} \left(1 - \delta_{\pm m,-1} \right) \end{aligned}$$

$$\frac{\partial}{\partial Y} Z_{n,\pm m}(\rho,\phi) =$$
$$\pm\left\{2n-m]Z_{n,\mp(m+1)} + [2(n-1)-m]Z_{n-1,\mp(m+1)} + \cdots + [2(m+1)-m]Z_{m+1,\mp(m+1)}\right\}$$
$$\mp\left\{2(n-1)-m+2]Z_{n-1,\mp(m-1)} + [2(n-2)-m+2]Z_{n-2,\mp(m-1)}\right.$$
$$\left.+\cdots+[2(m-1)-m+2]Z_{m-1,\mp(m-1)}\right\}\left(1-\delta_{\pm m,+1}\right) \tag{A.15}$$

The factors of $(1-\delta_{\pm m,-1})$ and $(1-\delta_{\pm m,+1})$ are present since there is no Zernike function in sin mϕ with m=0. For m=0:

$$\frac{\partial}{\partial X} Z_{n,0}(\rho,\phi) = 4nZ_{n,+1}(\rho,\phi) + 4(n-1)Z_{n-1,+1}(\rho,\phi) + \cdots + 4Z_{1,+1}(\rho,\phi)$$
$$\frac{\partial}{\partial Y} Z_{n,0}(\rho,\phi) = 4nZ_{n,-1}(\rho,\phi) + 4(n-1)Z_{n-1,-1}(\rho,\phi) + \cdots + 4Z_{1,-1}(\rho,\phi) \tag{A.16}$$

Let X_c and Y_c be the location of the origin of the coordinate system for the Zernike expansion measured in units of the normalizing radius. Note that $\partial/\partial X_c = -\partial/\partial X$ and $\partial/\partial Y_c = -\partial/\partial Y$. The operators $\partial/\partial X_c$ and $\partial/\partial Y_c$ are the generators for infinitesimal translations of the origin of the coordinate system in the X and Y directions, respectively. The operator representing the finite translation $(\Delta X_c, \Delta Y)_c$ is $\exp(\Delta X_c\, \partial/\partial X_c + \Delta Y_c\, \partial/\partial Y_c)$. Let the matrices A and B be given by:

$$\frac{\partial}{\partial X_c} Z_{nm}(\rho,\phi) = A_{n,m;n',m'} Z_{n'm'}(\rho,\phi)$$
$$\frac{\partial}{\partial Y_c} Z_{nm}(\rho,\phi) = B_{n,m;n',m'} Z_{n'm'}(\rho,\phi) \tag{A.17}$$

The matrices A and B are infinite dimensional. However, note that $A_{n,m;n',m'}$ and $B_{n,m;n',m'}$ are zero if n'>n. If the expansion in Zernike functions is truncated at some shell n, then translation will not couple to Zernike functions in higher shells. The matrices A and B will then be finite-dimensional and nilpotent. Nilpotency for a matrix A means that there is some non-negative integer r such that $A^r=0$. The operators $\partial/\partial X_c$ and $\partial/\partial Y_c$ couple a Zernike function $Z_{nm}(\rho,\phi)$ either to lower shells or to higher magnitude |m|. The operators $\partial/\partial X_c$ and $\partial/\partial Y_c$ have the property that $(\partial/\partial X_c)^r\,(\partial/\partial Y_c)^s\, Z_{nm}(\rho,\phi)=0$ whenever r+s > 2n-|m|. This means that $(\Delta X_c\, \partial/\partial X_c + \Delta Y_c\, \partial/\partial Y_c)^t\, Z_{nm}(\rho,\phi)=0$ whenever t > 2n-|m|. Define the matrix $Q=\Delta X_c\, A + \Delta Y_c\, B$. Q is nilpotent since $Q^t=0$ whenever t > 2n, where n is the highest shell kept in the expansion. In evaluating $e^Q=1+Q+Q^2/2!+...+Q^k/k!+...$, only the first 2n+1 terms have to be kept.

Suppose one has a set of functions $f_N(P;X_c,Y_c)$ where P is the point and X_c,Y_c is the origin for the expansion. Also suppose that matrices A and B are defined so that $\partial/\partial X_c\, f_N(P;X_c,Y_c) = A_{N,N'}\, f_{N'}(P;X_c,Y_c)$ and that $\partial/\partial Y_c\, f_N(P;X_c,Y_c) = B_{N,N'}\, f_{N'}(P;X_c,Y_c)$. Note that $\partial P/\partial X_c=0$ and $\partial P/\partial Y_c=0$. If some function F(P) is expanded in terms of these functions with coefficients a_N such that $F(P) = \Sigma\, a_N\, f_N(P;X_c,Y_c)$, then $\partial/\partial X_c\, a_N = -a_{N'}\, A_{N',N}$ and $\partial/\partial Y_c\, a_N = -a_{N'}\, B_{N',N}$. since $\partial/\partial X_c\, F(P)=0$ and $\partial/\partial Y_c\, F(P)=0$.

It is common to number the Zernike functions. They are ordered by shells n where $n \geq 0$ and $-n \leq m < n$. Within each shell, the Zernikes are ordered as follows: m = +n, m = -n, m = +(n-1), m = -(n-1), ..., m = +1, m = -1, m=0. The numbering is usually 1-based so that the purely radial Zernikes $Z_{00}(\rho,\phi)$, $Z_{10}(\rho,\phi)$, $Z_{20}(\rho,\phi)$, ... , $Z_{n0}(\rho,\phi)$, ... have N=1, 4, 9,..., $(n+1)^2$, ... Note that this N differs from the N used as a subscript on the functions $f_N^{(m)}(\rho^2)$. Eqn (A.17) can alternatively be written as:

$$\frac{\partial}{\partial X_c} Z_N(\rho,\phi) = A_{N,N'} Z_{N'}(\rho,\phi)$$
$$\frac{\partial}{\partial Y_c} Z_N(\rho,\phi) = B_{N,N'} Z_{N'}(\rho,\phi) \tag{A.18}$$

Appendix B Scaling transformation of the Zernike functions

The Zernike functions $Z_{nm}(\rho,\phi)$ depend upon the normalized radius $\rho=r/R$ and the azimuthal angle ϕ. We wish to express the Zernike functions with the normalizing radius R_{old} in terms of the Zernike functions with a different normalizing radius R_{new}. The operator $R\,\partial/\partial R$ is the generator for an infinitesimal scale transformation. Note that $R\,\partial/\partial R\, Z_{nm}(\rho,\phi) = -\rho\,\partial/\partial\rho\, Z_{nm}(\rho,\phi)$. The Zernike functions have the form $Z_{nm}(\rho,\phi)=f_{nm}(\rho^2)\ \rho^m \cos m\phi$ with $n \geq m$ for $m \geq 0$ and $f_{n|m|}\ \rho^{|m|} \sin |m|\phi$ with $n \geq m$ for $m < 0$. The operator $-\rho\,\partial/\partial\rho$ only couples Zernike functions of the same m. Restrict attention to $m \geq 0$ since the coefficients of the expansion depend only upon $|m|$. Now for $m \geq 0$

$$-\rho\frac{\partial}{\partial\rho} Z_{nm}(\rho,\phi) = \left(-\rho\frac{\partial}{\partial\rho} f_{nm}(\rho^2)\right)\rho^m \cos m\phi - mZ_{nm}(\rho,\phi) \tag{B.1}$$

so that we need only to expand $\rho d/d\rho\ f_{nm}(\rho^2)$ in terms of $f_{nm}(\rho^2)$ of the same m but differing n. Let $x = \rho^2$ so that $-\rho\, d/d\rho = -1/2\ x\, d/dx$. Also, it is more convenient to work with N = n-m rather than n since the functions $f_N^{(m)}(x)$ obey the Rodriguez formula

$$f_N^{(m)}(x) = \frac{1}{N!\,x^m}\frac{d^N}{dx^N}\left[x^{N+m}(x-1)^N\right] \tag{B.2}$$

The functions $f_N^{(m)}(x)$ are normalized so that

$$\int_0^1 f_N^{(m)}(x) f_{N'}^{(m)}(x)\, x^m\, dx = \frac{1}{2N+m+1} \tag{B.3}$$

We wish to find the coefficients $a_{N'}$ defined by

$$x\frac{d}{dx} f_N^{(m)}(x) = \sum_{N'=0}^{N} a_{N'}(N,m) f_{N'}^{(m)}(x) \tag{B.4}$$

The set of orthogonal polynomials $f_N^{(m)}(x)$ is a complete set for each m. N varies from zero to infinity and is the degree of the polynomial. Since $f_N^{(m)}(1) = 1$ for all $m\geq 0$ and $N\geq 0$, we can integrate by parts to obtain:

$$a_{N'}(N,m) = (2N'+m+1)\left(1 - \int_0^1 \frac{1}{N!}\frac{d^N}{dx^N}\left[x^{N+m}(x-1)^N\right]\left(\frac{1}{x^m}\frac{d}{dx}\left[x^{m+1} f_{N'}^{(m)}(x)\right]\right)\right)dx$$

Since $f_{N'}^{(m)}(x)$ is a polynomial of degree N', it follows that the factor in parentheses in the integrand is also a polynomial of degree N which we choose to call $p_{N'}(x)$. Repeated integration by parts gives:

$$a_{N'}(N,m) = (2N'+m+1)\left(1 - (-1)^N \int_0^1 \frac{1}{N!} x^{N+m}(x-1)^N \frac{d^N}{dx^N} p_{N'}(x)\right)dx \tag{B.6}$$

For N'<N, the Nth derivative of $p_{N'}(x)$ vanishes. Thus $a_{N'}(N,m)=2N'+m+1$ for N'<N and, of course, $a_{N'}(N,m)=0$ for N'>N. Thus we need to evaluate $a_{N'}(N,m)$ only for the remaining case N'=N. Now the leading coefficient in x^N for $x\frac{d}{dx}f_N^{(m)}(x)$ is N times the leading coefficient of x^N for $f_N^{(m)}(x)$. Thus

$$x\frac{d}{dx}f_N^{(m)}(x)=N\,f_N^{(m)}(x)+\sum_{N'=0}^{N-1}(2N'+m+1)f_{N'}^{(m)}(x) \tag{B.7}$$

Since N=n-m and since x d/dx=1/2 ρ d/dρ,we have

$$\rho\frac{d}{d\rho}f_{nm}(\rho^2)=2(n-m)\,f_{nm}(\rho^2)+2\sum_{N'=0}^{N-1}(2n'-m+1)f_{n'm}(\rho^2) \tag{B.8}$$

or

$$-R\frac{\partial}{\partial R}Z_{nm}(\rho,\phi)=(2n-m)Z_{nm}(\rho,\phi)+2\sum_{N'=0}^{N-1}(2n'-m+1)Z_{n'm}(\rho,\phi) \tag{B.9}$$

We thus have

$$-R\frac{\partial}{\partial R}Z_{nm}(\rho,\phi)=\sum_{n'}T_{nn'}(m)Z_{n'm}(\rho,\phi) \tag{B.10}$$

The matrix $T_{nn'}(m)$ is a lower triangular matrix which depends upon m. Let some function F(ρ,ϕ) be expanded in the Zernike functions, so that $F(\rho,\phi)=\Sigma\, a_{nm}\, Z_{nm}(\rho,\phi)$. Both the coefficients a_{nm} and the Zernike functions $Z_{nm}(\rho,\phi)$ are implicitly functions of R, but F(ρ,ϕ) does not depend upon R. This implies

$$R\frac{\partial}{\partial R}a_{nm}=-\sum_{n'}a_{n'm}T_{n'n}(m) \tag{B.11}$$

Since the diagonal elements of the lower triangular matrix $T_{n'n}(m)$ are all distinct, it may be diagonalized by some similarity transformation S with the property

$$\sum_{k,l}\left(S^{-1}\right)_{ik}T_{kl}S_{lj}=T_{ii}\,\delta_{ij} \tag{B.12}$$

Let $a^{(m)}$ be the vector

$$a^{(m)}=\begin{bmatrix}a_{m+0,m}\\ a_{m+1,m}\\ a_{m+2,m}\\ \vdots\end{bmatrix} \tag{B.13}$$

Then

$$R\frac{\partial}{\partial R}a^{(m)}=-a^{(m)}S\,T^{(diag)}\,S^{-1} \tag{B.14}$$

This has the solution

$$a^{(m)}(R)=a^{(m)}(R_o)e^{\log(R/R_o)\,S^{-1}T^{(diag)}S} \tag{B.15}$$

or

$$a^{(m)}(R)=a^{(m)}(R_o)\,S^{-1}\left(\frac{R}{R_o}\right)^{T^{(diag)}}S \tag{B.16}$$