Automated aberration extraction using phase wheel targets

Lena Zavyalova, Anatoly Bourov, Bruce W. Smith
Rochester Institute of Technology, 82 Lomb Memorial Dr., Rochester, NY 14623

ABSTRACT

An approach to in-situ wavefront aberration measurement is explored. The test is applicable to sensing aberrations from the image plane of a microlithography projection system or a mask inspection tool. A set of example results is presented which indicate that the method performs well on lenses with a Strehl ratio above 0.97. The method uses patterns produced by an open phase figure[1] to determine the deviation of the target image from its ideal shape due to aberrations. A numerical solution in the form of Zernike polynomial coefficients is reached by modeling the object interaction with aberrated pupil function using the nonlinear optimization routine over the possible deformations to give an accurate account of the image detail in 2-D. The numerical accuracy for the example below indicated superb performance of the chosen target shapes with only a single illumination setup.

Keywords: aberrations, phase wheel, lithography

1. INTRODUCTION

Aberrations are directly related to loss of contrast in the image. Modern lithography lenses require optical path tolerances on the order of several nanometers over extremely large apertures. If the primary aberrations are well-corrected, it is important to know how much wavefront distortion remains. Higher order aberrations are therefore of consequence due to their increased contribution to the total root mean square (rms) wavefront error in large pupils.

Measuring the optical aberrations in microlithography projection systems, such as stepper or aerial image microscope, when working from the image space is not trivial. Zernike polynomials are commonly used for the wavefront description. For many inverse aberration extraction methods however, it is generally thought that as long as the Zernike set is used there will be some aliasing[2] in the terms (i.e. higher order terms are indistinguishable from low order counterparts of the same azimuthal order and symmetry). Overall, the approaches currently being used to measure aberrations in-situ are image-based and estimate the wavefront phase either from the gradients[4,5,6] or from the intensity point spread functions.[7] These techniques have been modified and adapted in numerous ways by a number of contributors,[3–10] and include analysis of resist images or aerial images in their specific context. These processes often require a linear system representation when analyzing the response to aberrations. While the image formation in lithography systems is always non-linear, some linearity assumptions are valid to some degree. In certain tests, correction factors are applied to enforce linearity in the model, including lumping the aberration terms of the same type together.

We investigate a nonlinear solution to this problem of obtaining high order coefficients using numerical methods. The number of parameters to be estimated is large. We want to reliably measure up to and including the 9th order Zernike terms (35 individual aberrations) using a compact (small footprint) target and fast setup. The goal is to detect low levels of aberration (of Strehl above 0.9) at high resolution (less than 0.002 waves). Various methods, where image data is used for aberration retrieval, may require very intensive test cycles. For practical applications it is necessary to minimize the amount of data collection for input. In some situations, steppers are detuned and additional aberrations are introduced as a quantitative way to characterize the method performance in the optical system. In order to estimate the sensitivity of our model, a simulation study is performed with a synthetic lens, whose aberrations are precisely known.

The organization of this paper is as follows. We start with an introduction to the use of Zernike polynomials in studying optical aberrations (section 2). Our analysis and discussion will be centered on the corresponding wavefront aberrations. In section 3 the method of nonlinear least squares will be introduced as a way of obtaining the aberration coefficients.

An application of the proposed method and a simulation study will be presented in sections 4 and 5. It will be shown that Zernike coefficients can be estimated directly from the image data by way of object–image conjugation. Finally, the overall performance of our reconstruction method will be reviewed for accuracy and errors will be evaluated.

2. ZERNIKE POLYNOMIALS

We use the Zernike aberration polynomials to expand the phase as they are the standard set for analyzing the optical elements used in lithography applications. The wavefront shape is modeled as a weighted sum of polynomial functions:

$$W(\rho,\theta) = \sum_i \alpha_i Z_i(\rho,\theta)$$

where $Z_i(\rho,\theta)$ is the i^{th} Zernike polynomial function and α_i its coefficient. Here ρ and θ are the coordinates in the exit pupil. The Zernike aberrations have been described by many authors[11–13] although there are considerable variations in notation conventions. The index i, degree n, angular frequency m, and the polynomial in polar representation are listed in Table 1 up to 9^{th} order, while more extensive lists are available from other references. The ordering of polynomials in Table 1 follows a singular indexing scheme of Optical Sciences Center, University of Arizona.[11] The piston is the 0^{th} term on the list. Figure 1 shows plots of the Zernike polynomial functions $Z_i(\rho,\theta)$ ordered such that the radial degree n changes vertically and frequency m changes horizontally. It can be seen that as the index increases, the shapes vary slower on the inner part and faster on the periphery of the pupil. This means that if the expansion is truncated, it would not carry the maximum amount of information about the wavefront.

Table 1: Zernike terms Z_i for i=0 to 35 (up to the ninth order) in polar coordinates

i	Form of the polynomial	(n,m)	Name
0	1	(0,0)	Piston (not used)
1	$2\rho\cos(\theta)$	(1,1)	Tilt x (about y-axis)
2	$2\rho\sin(\theta)$	(1,1)	Tilt y (about x-axis)
3	$\sqrt{3}\,(2\rho^2-1)$	(1,0)	Power
4	$\sqrt{6}\,\rho^2\cos(2\theta)$	(2,2)	Astigmatism x (0°)
5	$\sqrt{6}\,\rho^2\sin(2\theta)$	(2,2)	Astigmatism y (45°)
6	$\sqrt{8}\,(3\rho^3-2\rho)\cos(\theta)$	(2,1)	Coma x
7	$\sqrt{8}\,(3\rho^3-2\rho)\sin(\theta)$	(2,1)	Coma y
8	$\sqrt{5}\,(6\rho^4-6\rho^2+1)$	(2,0)	Primary Spherical
9	$\sqrt{8}\,\rho^3\cos(3\theta)$	(3,3)	Trefoil x
10	$\sqrt{8}\,\rho^3\sin(3\theta)$	(3,3)	Trefoil y
11	$\sqrt{10}\,(4\rho^4-3\rho^2)\cos(2\theta)$	(3,2)	Secondary Astigmatism x
12	$\sqrt{10}\,(4\rho^4-3\rho^2)\sin(2\theta)$	(3,2)	Secondary Astigmatism y
13	$\sqrt{12}\,(10\rho^5-12\rho^3+3\rho)\cos(\theta)$	(3,1)	Secondary Coma x
14	$\sqrt{12}\,(10\rho^5-12\rho^3+3\rho)\sin(\theta)$	(3,1)	Secondary Coma y
15	$\sqrt{7}\,(20\rho^6-30\rho^4+12\rho^2-1)$	(3,0)	Secondary Spherical
16	$\sqrt{10}\,\rho^4\cos(4\theta)$	(4,4)	Tetrafoil x
17	$\sqrt{10}\,\rho^4\sin(4\theta)$	(4,4)	Tetrafoil y
18	$\sqrt{12}\,(5\rho^5-4\rho^3)\cos(3\theta)$	(4,3)	Secondary Trefoil x
19	$\sqrt{12}\,(5\rho^5-4\rho^3)\sin(3\theta)$	(4,3)	Secondary Trefoil y
20	$\sqrt{14}\,(15\rho^6-20\rho^4+6\rho^2)\cos(2\theta)$	(4,2)	Tertiary Astigmatism x
21	$\sqrt{14}\,(15\rho^6-20\rho^4+6\rho^2)\sin(2\theta)$	(4,2)	Tertiary Astigmatism y
22	$4\,(35\rho^7-60\rho^5+30\rho^3-4\rho)\cos(\theta)$	(4,1)	Tertiary Coma x
23	$4\,(35\rho^7-60\rho^5+30\rho^3-4\rho)\sin(\theta)$	(4,1)	Tertiary Coma y
24	$3\,(70\rho^8-140\rho^6-90\rho^4-20\rho^2+1)$	(4,0)	Tertiary Spherical
25	$\sqrt{12}\,\rho^5\cos(5\theta)$	(5,5)	Pentafoil x
26	$\sqrt{12}\,\rho^5\sin(5\theta)$	(5,5)	Pentafoil y
27	$\sqrt{14}\,(6\rho^6-5\rho^4)\cos(4\theta)$	(5,4)	Secondary Tetrafoil x
28	$\sqrt{14}\,(6\rho^6-5\rho^4)\sin(4\theta)$	(5,4)	Secondary Tetrafoil y
29	$4\,(21\rho^7-30\rho^5+10\rho^3)\cos(3\theta)$	(5,3)	Tertiary Trefoil x
30	$4\,(21\rho^7-30\rho^5+10\rho^3)\sin(3\theta)$	(5,3)	Tertiary Trefoil y
31	$\sqrt{18}\,(56\rho^8-105\rho^6+60\rho^4-10\rho^2)\cos(2\theta)$	(5,2)	Quaternary Astigmatism x
32	$\sqrt{18}\,(56\rho^8-105\rho^6+60\rho^4-10\rho^2)\sin(2\theta)$	(5,2)	Quaternary Astigmatism y

33	$\sqrt{20}\,(126\rho^9-280\rho^7+210\rho^5-60\rho^3+5\rho)\cos(\theta)$	(5,1)	Quaternary Coma x
34	$\sqrt{20}\,(126\rho^9-280\rho^7+210\rho^5-60\rho^3+5\rho)\sin(\theta)$	(5,1)	Quaternary Coma y
35	$\sqrt{11}\,(252\rho^{10}-630\rho^8+560\rho^6-210\rho^4+30\rho^2-1)$	(5,0)	Quaternary Spherical

Zernike coefficients α_i are directly related to the rms wavefront error, which in turn is related to image quality via the Strehl ratio.[13] Zernike polynomials are orthogonal over the unit circle (exit pupil with radius one), the significance of which is reflected in the polynomial coefficients. Since the wavefront variance is equal to the sum of the Zernike coefficients squared, each coefficient can be independently minimized to reduce the total wavefront rms value and theoretically should not affect coefficients for the other terms in the expansion.

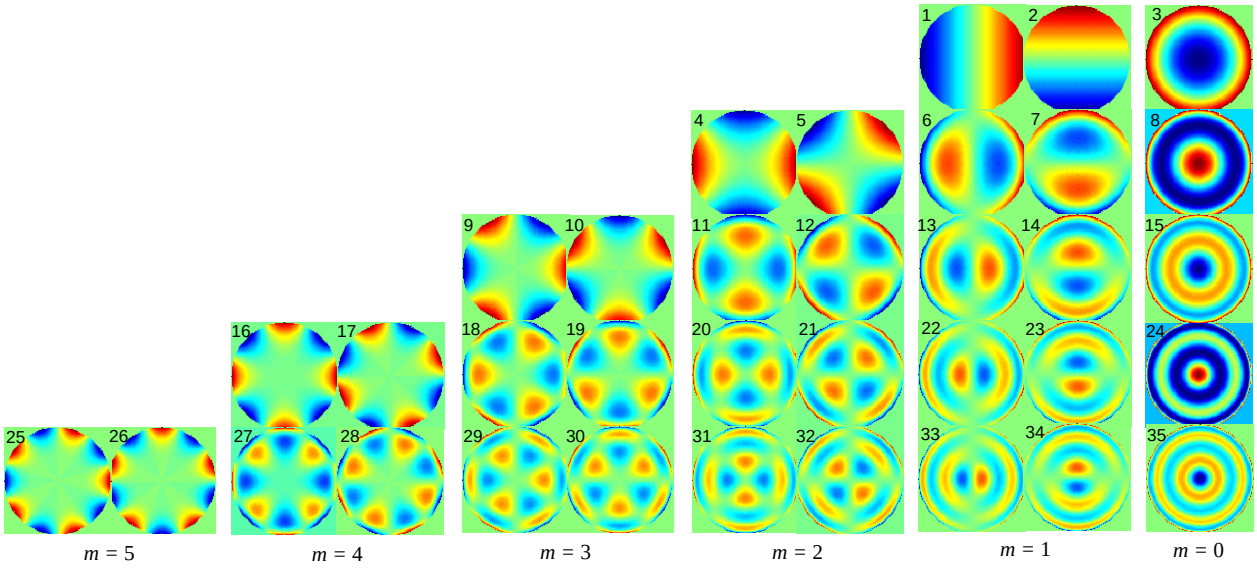

Fig. 1. Zernike polynomials up to the 9th order. Piston is excluded.

3. METHOD

The aberration effects on the image can be estimated by studying the image intensity changes coupled with defocus. We use a phase wheel target, described in a previous report[1], which allows a regular sampling across the pupil. The target contains transparent mask features phase-shifted by 180 degrees relative to a background. It is a complex-valued object function that has an infinite support in the frequency domain and therefore can sample the pupil continuously. The method does not impose a restriction on the target choice so long as the pupil is being adequately sampled. Ideally, the method must work with a target that has both a high sensitivity to aberrations and a low cross-correlation between different fit parameters.

The procedure involves precisely modeling the propagation of light from target to the detector in an image-forming system, and subsequently comparing the simulated images to the measured image. In microlithography, the image measurements may be derived either from the image intensity or from the image in photoresist. Once the initial set of images is compared, an iterative algorithm is employed to arrive at an estimate of aberration coefficients that give the best match between the model and the measured image data.

The aberration extraction is approached as an optimization[14] problem. The algorithm estimates the aberration coefficients by finding the minimum norm solution to a nonlinear least squares problem to quantify the difference between experimental and model-generated data. The wavefront is parameterized into the orthogonal basis functions (Zernike polynomials), coefficients for which are accepted into our parametric model as the arguments. On the input

we have $\{ f_1(\mathbf{z}), \ldots, f_M(\mathbf{z}) \}$, which is a set of input image functions that depend on a set of n variables obtained at M different focal planes. Let $\hat{\mathbf{f}}$ be an estimate of **f**, derived from the model which utilizes the first n Zernike functions. To find a set of 'optimal' parameters $\hat{\mathbf{z}}$ that reconstructs the image of the target, a fitting procedure is set up as follows:

$$\text{given } f(\mathbf{z}), \text{ find } \min_{\mathbf{z}\in\Re^n} \|e\|^2 = \sum_j \left| f(\mathbf{z}) - \hat{f}(\mathbf{z}) \right|^2 \quad \text{for the range of defocus values, where}$$

$$\mathbf{z} = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n\}, \quad \mathbf{z} \in \Re^n, \quad f \in \Re^n$$

In terms of the given basis, each $\mathbf{z} \in R^n$ can be represented by the n-tuple of aberration coefficients $\{\alpha_1, \ldots, \alpha_n\}$. An error metric is computed from the residual differences between modeled and recorded images, and the wavefront estimate is refined iteratively to drive the metric to a desired minimum. Parameter estimation is an n-dimensional problem requiring a nonlinear optimization algorithm to find a numeric solution. We make use of large scale gradient-based methods of optimization such as Gauss-Newton algorithm. A simplex search method that minimizes the difference without using the derivatives, while more computationally expensive, also performs well for this problem.

The fitting procedure is implemented in Matlab[15] (to perform image processing and calculations) and is interfaced with the Prolith[16] lithography simulator.

The method has potential to be further refined. Besides the Zernike coefficients, the important degrees of freedom of the imaging model include the illumination coherence factor, numerical aperture, focus offset, target geometry, etc. All but the Zernikes are usually known for the system or, in practice, can be determined from independent measurements. Additional flexibility can be gained in the model by including the variation in any of these – a partial coherence, for instance. This would then take into account an uncertainty in the partial coherence measurement.

The next example will explore a best-fit numerical solution for 35 parameters simultaneously.

4. EXAMPLE

To test the accuracy of the method, we perform a simulated experiment evaluating a 193nm 0.9NA aberrated lens with Strehl ratio of 0.975, chosen to closely match an actual system. The 2-D aerial image intensity that has been degraded by the known amount of aberration is input to the model. Synthetic aerial images are taken from a single target, for single illumination condition (σ=0.3), at three planes of focus.

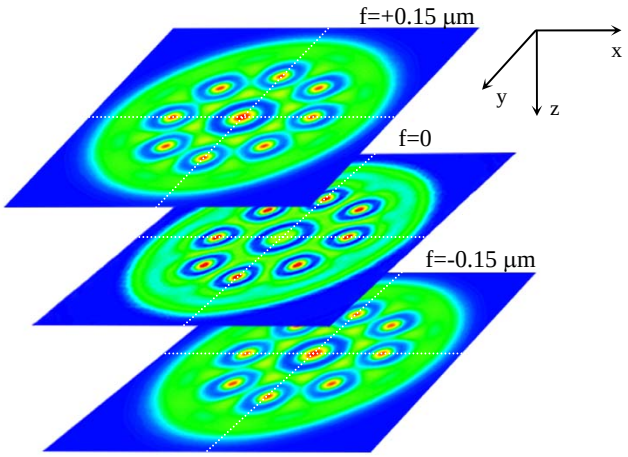

Fig. 2. Phase wheel target (aerial image) perturbation through focus

As is shown in Figure 2, one of the input images is the in-focus image while the others are at 0.15 microns to either side of the best focus. Defocusing an image perturbs it with an additional known phase error. The relative quadratic phase difference that results is encoded within the each image. As the fitting procedure is performed to determine the coefficients, the response at three focal planes is minimized jointly.

In this example we derive the first 35 Zernike terms up through the 9th order, excluding the piston term. The coefficients for tilts x and y (Z_1 and Z_2) are kept zero. The results obtained from the fit are presented in next section.

5. RESULTS AND DISCUSSION

The simulated results demonstrate that the method yields a superior quality of the wavefront estimate for zero measurement noise. This is illustrated in Figure 3 as a height map of the wavefront aberration over the pupil. The residual rms wavefront error is very low at 10^{-4} λ (or 0.02 nm). The agreement between the best-fit and the input aberration functions is further illustrated in Figures 4 and 5.

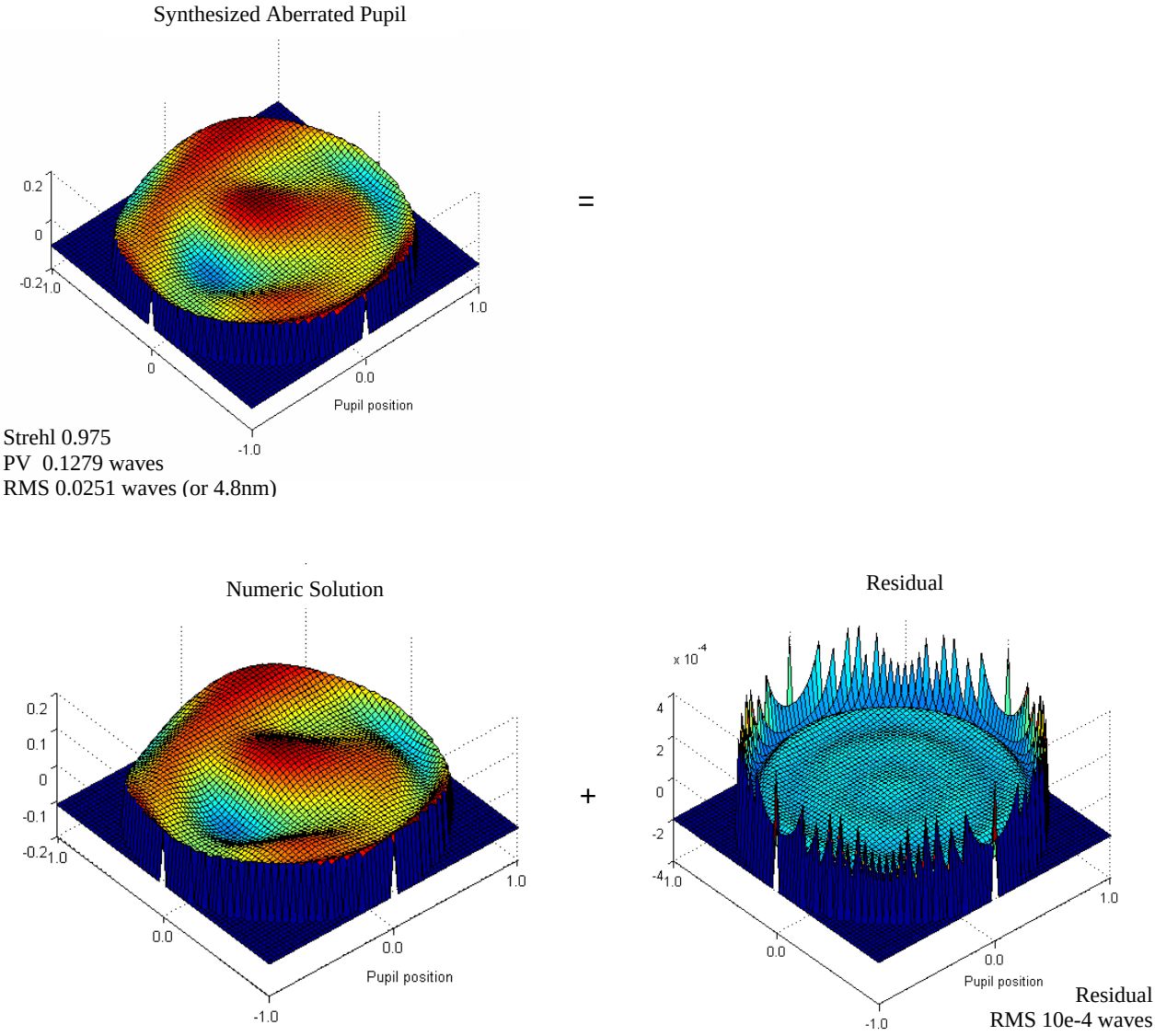

Fig. 3. Simulated performance: pupil wavefront maps of the actual wavefront (top), the generated solution (bottom left), and the residual surface (bottom right).

The individual Zernike coefficients obtained from the fit are given in Figure 4. The fit was performed with 35 coefficients. The coefficients are expressed in multiples of the wavelength used (λ=193 nm). The estimates of the Zernike coefficients are seen to coincide closely with the real values, confirming that the model performs a unique reconstruction of the parameters.

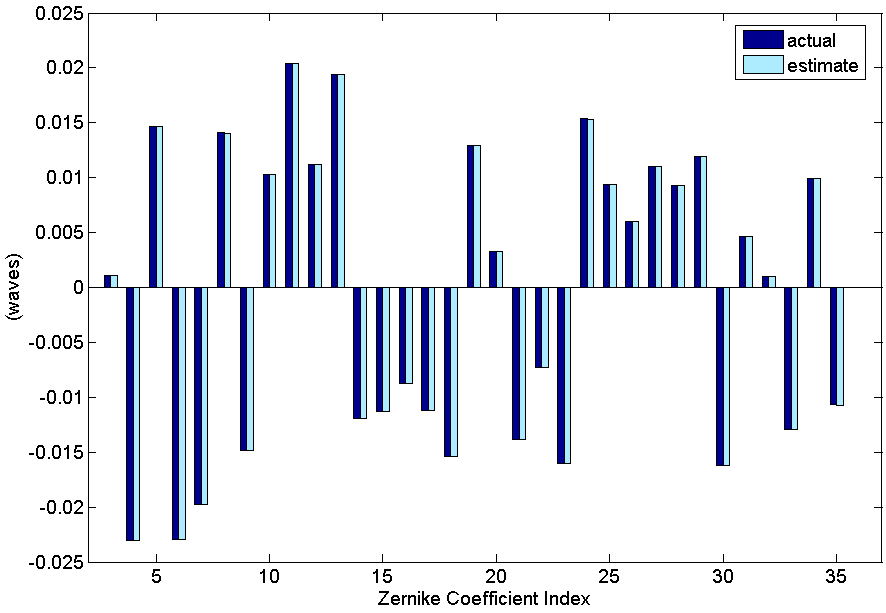

Fig. 4. Calculated vs. actual aberration coefficient values

Next, Figure 5 shows how the total rms error is distributed among the individual terms. The parameter differences look rather uniform, and on average are slightly higher for the spherical terms. The largest error observed between the true and the retrieved Zernike coefficients is less than 0.0001 waves. Hence, the theoretically achievable model accuracy is on the order of 0.01%.

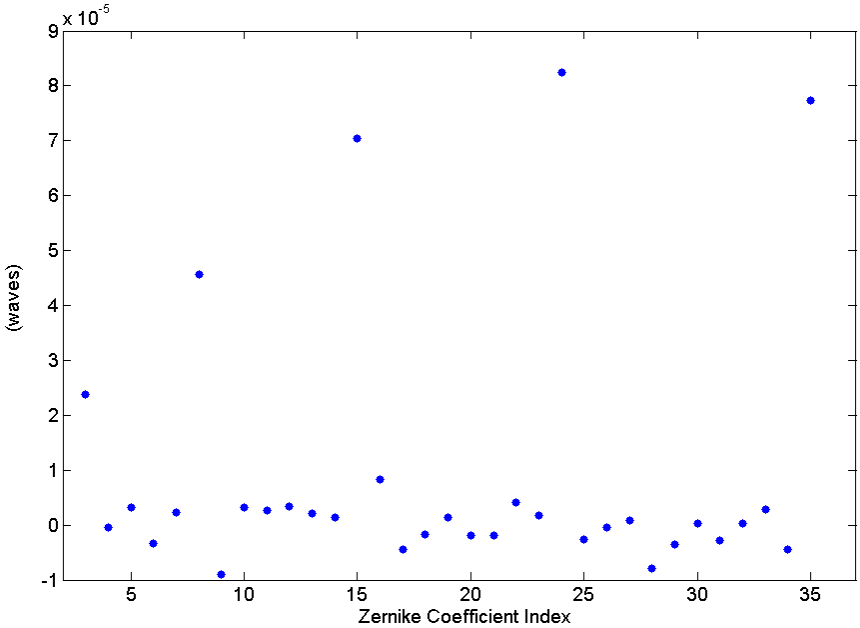

Fig. 5. Difference real vs. fitted wavefront coefficient values

Figure 6 shows the representative standard errors on the individual parameters from the fit. In general, the standard error tends to increase for the higher order terms but all are considered to be very small. As the number of terms in the expansion becomes large we will expect a limitation in the precision of the estimate, however the impact is only secondary. While reconstructing a larger number of terms increases noise, the system that is allowed to model higher order terms provides more degrees of freedom for a better fit to a real wavefront. So in practice, the estimates will still be tied to a total number of coefficients used in a fit.

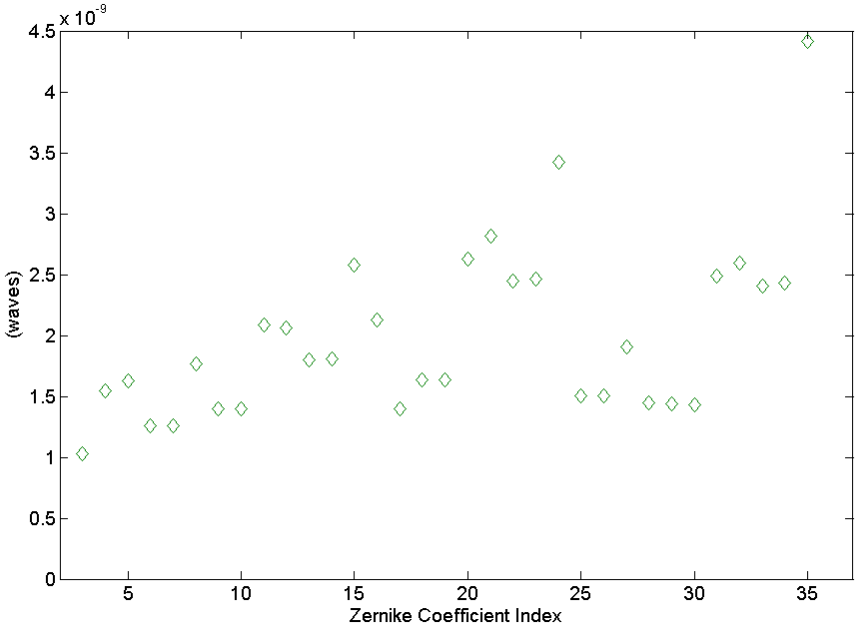

Fig. 6. Standard error on fitted aberration coefficients

The method's convergence is good and its stability gives some reassurance that the pupil has been adequately sampled by the target's spectra. This analysis is extendable to resist imaging as well. The use of multiple measurements (additional images of the same object) in addition to a different defocus is expected to further increase the robustness of model if needed and help ensure that the algorithm converges reliably.

If a robust method of least squares is used to obtain the coefficients, the quality of the fit will depend only on the number of terms used. To model the wavefront aberration as closely as possible, the Zernike fit to a wavefront should contain a sufficient number of terms, so long as the error of the fit associated with each coefficient does not become as significant as the coefficient itself. Hence, in practice, depending on the noise level there will be an optimal number of terms providing a minimum fit error. However, the parameter-fits still have to be used with caution as the performance will depend on how well the assumed expansion matches the reality. Due to some redundancy[2] in the Zernike decomposition, some aliasing would presumably be found in that the value of each coefficient could change when the expansion is truncated. In fact, the main effect is seen when the sum is truncated early, i.e. when higher order aberrations are present in the pupil but are not modeled in the fit. We look at this issue in the next example.

A simulated analysis is done next where the polynomial order of the original wavefront is 35, but only a limited number of coefficients are used in the extraction (the first 24). Hereby, there is a finite content in the high order terms $Z_{25} - Z_{35}$ that is not extracted. The simulation reveals that the presence of high order aberrations causes systematic errors in the estimation of low order coefficients. Figure 7 gives some insight into the possible magnitude of such errors. The chart displays the coefficient values for several subgroups. Within each subgroup there are terms of the same azimuthal frequency m of increasing radial degree n (just as they are along the columns in Figure 1). The similarity between the Zernike polynomials along the columns implies the stability of the coefficient values. In this specific example, the

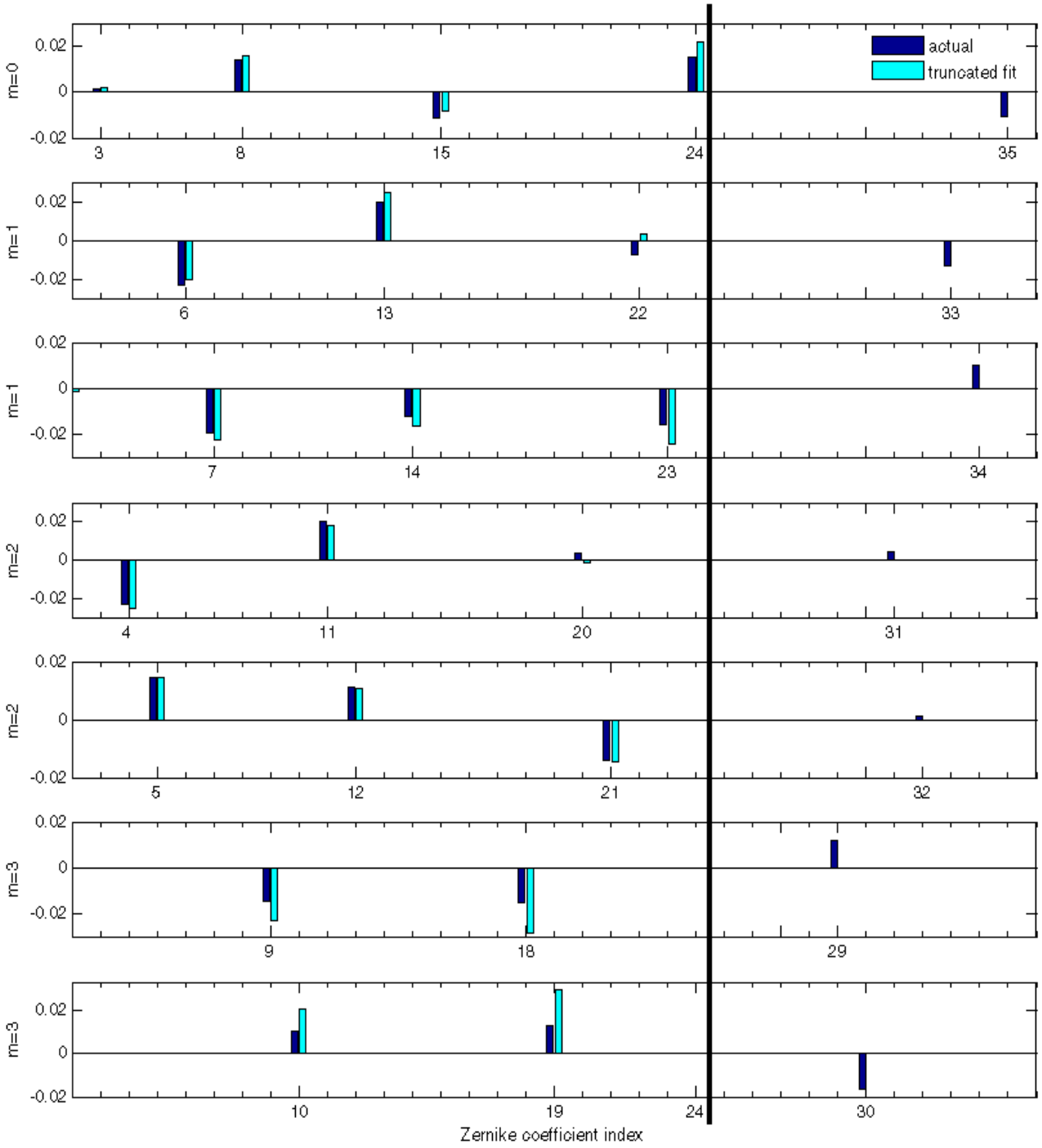

Fig. 7. Here we check the validity of truncating the fit early, which could lead to interaction between orders. The wavefront solution is given when expansion is truncated at index #24. Higher order coefficients (25–35) while present were excluded from the fit. Charted by subgroup (about the individual columns in Fig. 1), m=0 represents spherical aberration terms (3, 8, 5, and 24), m=1 comatic x terms (6, 13, 22, and 33), m=2 astigmatic, and so on. The effect of high orders aberration is noticeable at low orders. The solution is generally unsatisfactory and higher order coefficients must be included in the fit.

impact of higher order terms on the estimation of the first 24 Zernikes is as follows. It can be seen form the corresponding graph for spherical aberrations that Z_{35} is shifted into the corresponding lower terms 3, 8, 15, and 24. With only 24 terms in the expansion, adding the fifth spherical component (9^{th} order polynomial) to the wavefront changes the lower order spherical coefficient estimates by 15 to 40% on average. The coefficient for 7^{th} order y-coma (Z_{23}) is also significantly overestimated, by about 50% in magnitude according to the simulation. The next lower order term Z_{14}, representing 5^{th} order y-coma aberration, is overestimated by about 30%. Here the 3^{rd} order y-coma (Z_7) estimate is contaminated by 15%. The effect on the higher order estimates is stronger, diminishing toward the 3^{rd} (lower) order. This phenomenon affects each group more or less equally as well as raises the total rms wavefront error by 1.6 nm. Depending on the amount of high order aberration present, the error in extracted coefficients is expected to shoot to a 100% (as in the m=3 subgroup). This behavior is the form of aliasing, which in this context is a contamination of low order aberration estimates by the presence of higher order errors that occurs in the truncated fits. When the expansion is truncated early the fit accuracy will suffer. We know that we can minimize this error by extracting a complete set of coefficients (35 coefficients). Therefore, in cases where the high order aberrations exist (as they do in fact) and the accurate knowledge of low order terms is needed, the former must also be included in the fit.

On the other hand, when the aberration content of the pupil is represented completely by some finite number of polynomials, the higher order terms do not affect the lower terms. As the number of the fitted coefficients increases, the value of each coefficient will essentially be independent of the other terms in the expansion. This means that higher order components should be included whenever possible, and may be measurable when included in the fit (unless the effect of variations is below the method's ability to sense these). Hereby, the accuracy of the model will depend on the number of coefficients in the fit, and there may be a trade-off between accuracy and noise. Selecting an appropriate number of fit parameters is critical to final accuracy. There is always a possibility that real aberrations of very high order could affect the fit. There must be a large enough number of coefficients in the model to distinguish the magnitude of terms in lenses with significant amounts of higher order aberrations.

6. CONCLUSION

A technique for automated aberration extraction has been developed for lithography applications. The fitting algorithm includes the nonlinear optimization of the fit parameters. Simulation results using synthetic images show that Zernike aberration coefficients can be recovered from intensity information, given appropriate target choice. Nonlinear optimization proved to be useful for fitting a variety of targets. Phase wheel targets provide high sensitivity and yield good results. The example was given for the method, characterizing a typical lens. The results indicate that it is feasible to use the phase wheel aberration monitor for characterizing the high order Zernike space (35 terms). The model has been validated to provide a high sensitivity (wave errors at $\lambda_{193}/100$) as well as high accuracy. This performance compares favorably with other methods. In the noiseless case, the fitting error as predicted is less than 0.01%, which is below the typical noise level of any current system. The effects of image measurement noise will be quantified in follow-on studies. Further work will also address the tests with various amounts of random errors. The method is extendable to incorporate resist imaging into the model, and is capable of using SEM images of actual resist for fitting the aberrations.

7. REFERENCES

[1] Zavyalova, L., Smith, B., Suganaga, T., Matsuura, S., Itani, T., Cashmore, J. "In-situ aberration monitoring using phase wheel targets," *Proc. SPIE,* **5377**, 2004.

[2] Herrmann, J. "Cross coupling and aliasing in modal wave-front estimation," *J. Opt. Soc. Am.* **71**, 989, 1981.

[3] Shack, R. and Platt, B. "Production and use of a lenticular Hartmann screen," *J. Opt. Soc. Am.* **61**, 656, 1971.

[4] Farrar, N., Smith, A., Busath, A., Taitano, D. "In-situ measurement of lens aberrations," *Proc. SPIE,* **4000**, 2000.

[5] van der Laan, H., Dierichs, M., van Greevenbroeck, H., McCoo, E., Stoffels, F., Pongers, R., Willekers, R. "Aerial image measurement methods for fast aberration set-up and illumination pupil verification," *Proc. SPIE,* **4346**, 2001.

[6] Kirk, J., Progler, C. "Application of blazed gratings for determination of equivalent primary azimuthal aberrations," *Proc. SPIE,* **3679**, 1999.

[7] Dirksen, P., Braat, J., Janssen, A., Juffermans, C., Leeuwestein, A. "Experimental determination of lens aberrations from the intensity point spread function in the focal region," *Proc. SPIE,* **5040**, 2003.

[8] Kirk, J. "Review of photoresist based lens evaluation methods," *Proc. SPIE,* **4000**, 2000.

[9] Dirksen, P., Juffermans, C., Pellens, R., Maenhoudt, M., De Bisschop, P. "Novel aberration monitor for optical lithogarphy," *Proc. SPIE,* **3679**, 1999.

[10] Gordon, R., Flagello, D., McCallum, M. "Deducing aerial image behavior from AIMS data," *Proc. SPIE,* **4000**, 2000.

[11] Wyant, J. C. and Creath, K. "Basic Wavefront Aberration Theory for Optical Metrology." Ch.1 in *Applied Optics and Optical Engineering, Vol. XI* (R. R. Shannon and J. C. Wyant, Eds.). New York: Academic Press, 1992.

[12] Born, M. and Wolf, E. *Principles of Optics*, 7th ed. Cambridge, UK: Cambridge University Press, 1999.

[13] Malacara, D. and Malacara, Z. *Handbook of Lens Design*. New York: Marcel Dekker, 1994.

[14] Dennis, J. and Schnabel, R. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[15] MATLAB 7, The MathWorks Inc., Natick, MA.

[16] Lithography simulator PROLITH version 8.1, KLA-Tencor Corp.