Optimization of pupil sampling scheme for aerial image based aberration measurement of projection optics in lithographic tools

Lijuan WANG [a], Shiyuan LIU [*, b], Wei LIU [a], Tingting ZHOU [a]

[a] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China.

[b] State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China.

ABSTRACT

In this paper, the aberration measurement technique using aerial image sensor (AIS) is further discussed, and an approach to optimize the pupil sampling scheme for this technique is proposed. The accuracy of this technique heavily relies on the pupil sampling scheme as it has a significant impact on the random error propagation of the Zernike coefficients. We formulate the optimization problem using a continuous function and using the gradient information to search the solution space. We also employ the regularization framework using penalty functions to restrain the complexity of the sampling scheme. The simulation work has demonstrated that the pupil sampling scheme obtained by the proposed optimization approach are more suitable than those by the trial and error method.

Keywords: lithography, wavefront aberration metrology, aerial image sensor, pupil sampling, optimization

1. INTRODUCTION

Wavefront aberration of projection lens is one of the most important indicators to evaluate the imaging quality of optical lithographic tools [1-2]. As numerical aperture (NA) yields to the manufacturing limit adapting to a higher resolution, it is necessary to ensure the wavefront aberration is no more than 10 mλ. Thus there is a need to develop in-situ techniques and systems to accurately measure aberrations up to the 37th or even higher-order Zernike coefficient.

Due to the advantage of lower cost and easier implement in tools without a portable phase measurement interferometer PMI [3-5], aerial image based techniques, such as transmission image sensor (TIS) at multiple illumination settings (TAMIS) [6, 7] and Z37 aerial image sensor (AIS) [8-11] have been widely used for in-situ measurement of lens aberrations. In these techniques, binary grating marks are utilized to perform pupil sampling. For example, Z37 AIS technique is able to measure aberrations up to the 37th Zernike coefficient by introducing a set of 36 binary grating marks with different pitches and orientations. As these gratings are corresponding to 72 pupil sampling points, the wavefront aberration at each sampling point over the pupil plane can be obtained from the spectrum of the aerial image intensity, and finally the Zernike coefficients can be calculated from the over-determined equations by the least-square method.

The optimization of the sampling scheme in the pupil plane is crucial in aerial image based aberration measurement techniques, since the sampling scheme greatly affects the measurement error in Zernike coefficients. The existing approach to optimize the pupil sampling scheme is mostly on the basis of trial and error method [11], which evaluates the pupil sampling schemes by the maximum error propagation ratio that has a relationship of parabolic dependence with the condition number. However, the trial and error method relies heavily on the list with limited pupil sampling schemes, and thus the pupil sampling scheme finally chosen for the aberration measurement may not be the best one.

In this paper, the aberration measurement technique using aerial image sensor is further discussed, and a popular approach to optimize the pupil sampling scheme for this technique is proposed. The proposed approach simplifies the optimization of the pupil sampling scheme by using a continuous function and searches the solution space by using the gradient information based on Powell's optimization method. Since this optimization method is suitable for non-constraint problems, a regularization framework is introduced by using penalty functions to restrain the complexity of the

* Contact author: shyliu@mail.hust.edu.cn; phone: +86 27 87792409; fax: +86 27 87792413.

2009 International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology Systems, edited by Shenghua Ye, Guangjun Zhang, Jun Ni, Proc. of SPIE Vol. 7511, 751106 · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.837771

pupil sampling scheme. Simulation work was conducted to demonstrate the validity and advantage of the proposed approach by comparing to the trial and error method.

2. ABERRATION MASUREMENT METHOD

The in-situ technique based on aerial image sensor can measure aberrations up to the 37th Zernike coefficient. A set of 36 binary grating marks with different pitches and orientations are used in this technique. Figure 1 shows schematically the optical imaging system of this technique in lithographic tools. When the light from the illumination system passes through the grating mark, the light beam is diffracted by the binary grating. As the duty of the grating is 50%, the diffracted rays have no even harmonics except for the 0th-order one, and they form an aerial image on the focal plane via the projection lens. The image sensor integrated into the wafer stage is used to collect the light intensity and convert it into electrical signals.

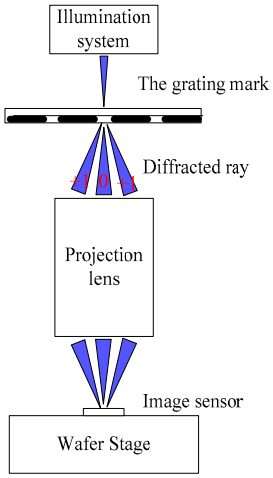


Fig.1 Imaging of a binary grating in a lithographic tool.

According to the Hopkins imaging theory, the aerial image intensity can be expressed in the scalar form [14]:

$$i(x_i, y_i, \sigma, d) = \iiiint TCC(f', g', f'', g'', \sigma, d) O(f', g') O^*(f'', g'') \times \exp[2\pi j x_i (f' - f'') + y_i (g' - g'')] df' df'' dg' dg'' , \quad (1)$$

where d is the axial shift of the aerial image sensor at the image plane, σ is the partial coherent factor and defined as the effective source filling factor in the projection optics pupil, $TCC(f', g', f'', g'', \sigma, d)$ is the transmission cross coefficient:

$$TCC(f', f'', g', g'', \sigma, d) = \iint_{\infty} J(f_c, g_c, \sigma) H(f' + f_c, g' + g_c, d) \times H^*(f'' + f_c, g'' + g_c, d) df_c dg_c . \quad (2)$$

Here, $O(f', g')$ is the diffraction spectrum of the binary grating, $J(f_c, g_c, \sigma)$ represents the effective source in optical lithography as:

$$J(f_c, g_c, \sigma) = \frac{1}{\pi \sigma^2} \mathrm{circ}\left(\frac{\sqrt{f_c + g_c}}{\sigma}\right). \quad (3)$$

The pupil function $H(f, g, d)$ stands for the transmission function in the pupil plane and can be written as:

$$H(f, g, d) = \exp[-jkW(f, g, d)] \mathrm{circ}\left(\sqrt{f^2 + g^2}\right), \quad (4)$$

where $k = 2\pi/\lambda$ is the wave number, λ is the wavelength of the monochromatic light source, and $W(f, g, d)$ is the total aberrated wavefront including the lens aberration.

The $+1^{st}$-order spectrum of the aerial image intensity can be determined by the interaction of the 0^{th}-order and $+1/-1^{st}$-order diffraction lights of the object:

$$I(\rho_m, \theta_m, \sigma, d) = \frac{1}{2\pi}\left[TCC(0,0,x_m, y_m, \sigma, d) + TCC(-x_m, -y_m, 0,0,\sigma, d)\right], \tag{5}$$

where (x_m, y_m) represents the normalized Cartesian coordinates in the pupil plane, and can be transformed from the corresponding normalized polar coordinates (ρ_m, θ_m) by:

$$x_m = \rho_m \cos\theta, \;\; y_m = \rho_m \sin\theta \,. \tag{6}$$

The wavefront aberration at each sampling point $W(\rho_m, \theta_m)$ can be divided into even aberration and odd aberration:

$$W(\rho_m, \theta_m) = W_{even}(\rho_m, \theta_m) + W_{odd}(\rho_m, \theta_m), \tag{7}$$

By making some assumptions, Eq. (5) can be further simplified. It is noted that the phase shift $\varphi(\rho_m, \theta_m)$ of the $+1^{st}$-order spectrum of the aerial image intensity at the ideal focal plane corresponding to $d = 0$ is proportional to the odd aberration, while the defocus $D(\rho_m, \theta_m)$ at which the amplitude of the $+1^{st}$-order spectrum reaches the extreme value is proportional to even aberration:

$$W_{odd}(\rho_m, \theta_m) = \varphi(\rho_m, \theta_m)/k \,, \tag{8}$$

$$W_{even}(\rho_m, \theta_m) = -D(\rho_m, \theta_m)\left(\sqrt{1 - NA^2\rho^2} - 1\right), \tag{9}$$

In order to calculate higher-order Zernike coefficients, enough sampling points in the pupil plane or enough binary grating marks with different pitches and orientations are essential. Assuming there are 36 grating marks, the wavefront aberrations at 72 corresponding sampling points can be obtained, and then can be used to calculate the Zernike coefficients up to the 37^{th} term by the least-square method:

$$\begin{bmatrix} W(\rho_1,\theta_1) \\ W(\rho_2,\theta_2) \\ \vdots \\ W(\rho_{72},\theta_{72}) \end{bmatrix} = \begin{bmatrix} R_2(\rho_1,\theta_1) & R_3(\rho_1,\theta_1) & \cdots & R_{37}(\rho_1,\theta_1) \\ R_2(\rho_2,\theta_2) & R_3(\rho_2,\theta_2) & \cdots & R_{37}(\rho_2,\theta_2) \\ \vdots & \vdots & \cdots & \vdots \\ R_2(\rho_{72},\theta_{72}) & R_3(\rho_{72},\theta_{72}) & \cdots & R_{37}(\rho_{72},\theta_{72}) \end{bmatrix} \begin{bmatrix} Z_2 \\ Z_3 \\ \vdots \\ Z_{37} \end{bmatrix}, \tag{10}$$

In a more compact notation Eq. (10) can be rewritten as:

$$\mathbf{W} = \mathbf{R}\mathbf{Z} \,. \tag{11}$$

where $\mathbf{W}$ is the vector containing the wavefront aberrations at the 72 sampling points; $\mathbf{R}$ is the matrix of Zernike polynomials; $\mathbf{Z}$ is the unknown Zernike coefficient vector to be measured; $R_n(\rho_m, \theta_m)$ and Z_n are the n^{th} Zernike polynomial over the pupil plane and its corresponding n^{th} Zernike coefficient, respectively.

To conclude, the aerial image based aberration measurement of projection optics in lithographic tools can be carried out in the process shown in Fig.2.

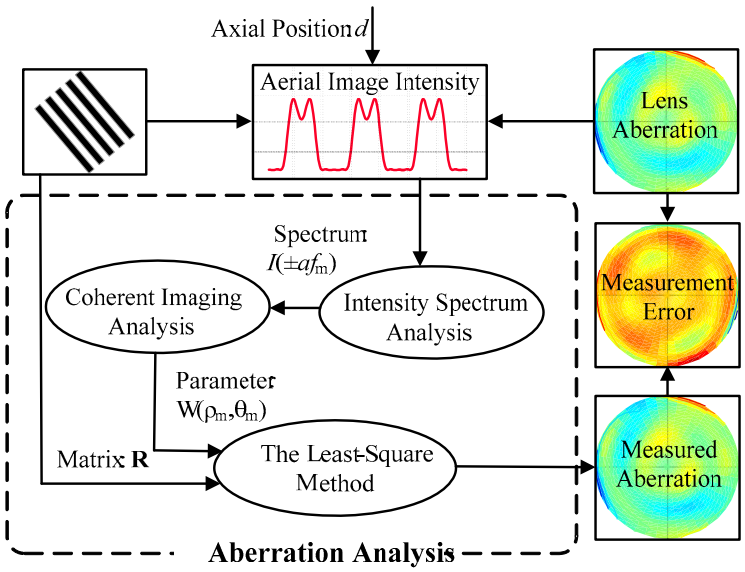
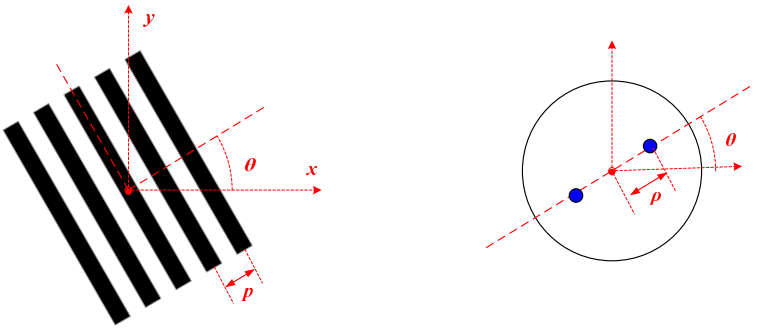


Fig.2 Process of aberration measurement based on aerial image sensor.

3. OPTIMAZATION OF PUPUIL SAMPLING SCHEME

The pupil sampling scheme is determined by the set of binary grating marks. Figure 3 shows schematically one of these 50% duty marks and the two corresponding sampling points at the pupil plane. Considering this relationship between the binary grating and its corresponding sampling points, the optimization of pupil sampling scheme can be divided into two steps, namely optimization of orientations and optimization of normalized radii.



(a) A binary grating mark (b) Two corresponding sampling points

Fig.3 (a) A binary grating mark with orientation θ and pitch p, and (b) the two corresponding sampling points in the pupil plane with orientation θ and normalized radius ρ.

3.1 Optimization of orientations

As mentioned in Eq. (10), the wavefront aberration at a pupil sampling point (ρ_m, θ_m) can be expressed as the sum of orthogonal Zernike polynomials:

$$W(\rho_m,\theta_m)=\sum_n R_n(\rho_m,\theta_m)Z_n\,. \tag{12}$$

where the Zernike polynomials $R_n(\rho_m, \theta_m)$ is a sinusoidal or cosine function of 0-θ term, 1-θ term, 2-θ term, 3-θ term and so on. For example, the polynomial of Z_7 term is expressed as:

$$R_7(\rho,\theta)=(3\rho^3-2\rho)\cos\theta\,, \tag{13}$$

This means that the polynomial R_7 has a cosine function of 1-θ term, while its derivative to θ has a sinusoidal function:

$$R_7'(\rho,\theta) = -(3\rho^3 - 2\rho)\sin\theta\,. \tag{14}$$

In order to measure the Zernike coefficient Z_n with the less error, the angle θ should be chosen such that the derivative of R_n to θ is zero. At this angle, R_n reaches its extreme value for a given radius ρ and the wavefront aberration varies the most sharply along the change of ρ. In the case of Z_7, it demands Eq. (14) to be zero, which means that θ is 0˚. The sampling point with this angle is said to be sensitive to Z_7. Table 1 depicts all the angles sensitive to Zernike coefficients up to the 37th term.

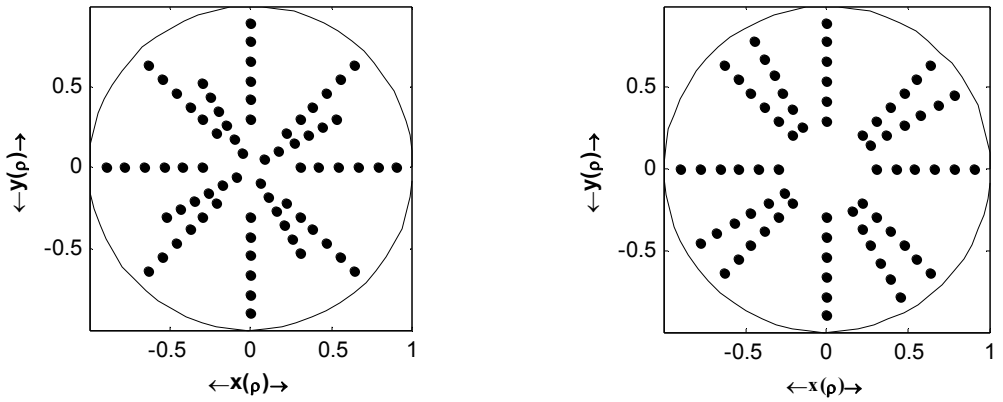
Table 1 Angles sensitive to Zernike coefficients up to the 37th term.

θ terms	Classification of aberration	Function of θ	Sensitive to Zernike coefficients	Angles
0-θ	Spherical	Non	Z_1, Z_4, Z_9, Z_{16}, Z_{25}, Z_{36}	All
1-θ	Coma	cosθ	Z_2, Z_7, Z_{14} Z_{23}, Z_{34}	0˚
		sinθ	Z_3, Z_8, Z_{15}, Z_{24}, Z_{35}	90˚
2-θ	Astigmatism	cos2θ	Z_5, Z_{12}, Z_{21}, Z_{32}	0˚, 90˚
		sin2θ	Z_6, Z_{13}, Z_{22}, Z_{33}	45˚, 135˚
3-θ	3-foil	cos3θ	Z_{10}, Z_{19}, Z_{30}	0˚, 60˚, 120˚
		sin3θ	Z_{11}, Z_{20}, Z_{31}	30˚, 90˚, 150˚
4-θ	4-foil	cos4θ	Z_{17}, Z_{28}	0˚, 45˚, 90˚, 135˚
		sin4θ	Z_{18}, Z_{29}	22.5˚, 67.5˚, 112.5˚, 157.5˚
5-θ	5-foil	cos5θ	Z_{26}	0˚, 36˚, 72˚, 108˚, 144˚
		sin5θ	Z_{27}	18˚, 54˚, 90˚, 126˚, 162˚

Considering the simplicity for grating fabrication, 6 orientations with angles of 0˚, 30˚, 45˚, 90˚, 120˚, and 135˚ are preferred for the measurement of aberrations up to the 37th Zernike coefficient.

3.2 Optimization of normalized radii

The trial and error method [11] was adopted in Z37 AIS technique to optimize the pupil sampling scheme, in which a list of several sampling schemes is generated and evaluated. Figure 4 depicts two typical examples of the sampling schemes.



(a) Sampling scheme 1 (b) Sampling scheme 2

Fig.4 Two typical examples of sampling schemes in the pupil plane.

The way to evaluate the pupil sampling schemes is to calculate the maximum error propagation ratio of the matrix **R**, which has a relationship of parabolic dependence with the condition number [11]. Here, the condition number is defined as a square root of ratio between the maximum and the minimum singular values of **R** and can be used to describe the error

or uncertainty in solving Eq. (11). It is noted that the smaller the condition number is, the more accurate the outcome is, owing to the smaller maximum error propagation ratio. Therefore, the sampling scheme with a minimal value of condition number is chosen to be the optimal one. However, it is clear that this trial and error method relies heavily on the list of limited pupil sampling schemes and the sampling scheme finally chosen for the aberration measurement may not be the best one.

In order to achieve an optimal result of pupil sampling scheme, here we adopt a popular approach from the field of optimization theory. By formulating the optimization problem, we introduce a continuous cost function:

$$f(\mathbf{x}) = \mathrm{cond}(\mathbf{R}(\mathbf{x})) \tag{15}$$

where $\mathbf{x}$ is a vector of the parameters to be optimized, $\mathbf{R}(\mathbf{x})$ represents the matrix of Zernike polynomials derived from $\mathbf{x}$, and **cond** means the function to obtain the condition number of $\mathbf{R}$. Thus, the optimization problem turns into a minimization problem:

$$\begin{aligned} &\min[f(\mathbf{x})], \quad \mathbf{x} = \boldsymbol{\rho} \\ &s.t.: \max(\boldsymbol{\rho}) \le 1 \quad \min(\boldsymbol{\rho}) \ge 0 \end{aligned} \tag{16}$$

where $\boldsymbol{\rho}$ is a vector consisting of the 36 normalized radii of pupil sampling points corresponding to the 36 pitches of binary gratings. In Eq. (16), there are too many parameters in the vector $\mathbf{x}$ to be optimized. On the assumption that the distances $\Delta\rho$ between two adjacent sampling points of the same orientation are identical, the model can be simplified as:

$$\begin{aligned} &\min[f(\mathbf{x})], \quad \mathbf{x} = [\rho_1, \rho_2, \ldots, \rho_6, \Delta\rho_1, \Delta\rho_2, \ldots, \Delta\rho_6] \\ &s.t.: \max(\boldsymbol{\rho}(\mathbf{x})) \le 1 \quad \min(\boldsymbol{\rho}(\mathbf{x})) \ge 0 \end{aligned} \tag{17}$$

where ρ_1, $\rho_{2,\ldots,}$ ρ_6 represent the smallest normalized radii for the 6 orientations, and $\Delta\rho_1$, $\Delta\rho_2$,…, $\Delta\rho_6$ represent the distances between two adjacent sampling points of each orientations. Thus the number of parameters in vector $\mathbf{x}$ to be optimized is reduced from 36 to 12, while other normalized radii in vector $\boldsymbol{\rho}$ can be obtained from vector $\mathbf{x}$. By employing the regularization framework using the penalty function $P(\mathbf{x})$, the bound constrained optimization problem is reduced to an unconstrained one:

$$\begin{aligned} &\min[F(\mathbf{x})] \\ &F(\mathbf{x}) = f(\mathbf{x}) + \alpha_1 P_1(\mathbf{x}) + \alpha_2 P_2(\mathbf{x}) \end{aligned} \tag{18}$$

where α_1, α_2 are the user-defined parameters to reveal the weights of the regularization; $P_1(\mathbf{x})$ and $P_2(\mathbf{x})$ are penalty functions:

$$\begin{aligned} P_1(\mathbf{x}) &= \{\max[\boldsymbol{\rho}(\mathbf{x}), 1]\}^2 \\ P_2(\mathbf{x}) &= \{\max[-\boldsymbol{\rho}(\mathbf{x}), 0]\}^2 \end{aligned} \tag{19}$$

Conjugate gradient methods are widely used for optimization. The basic idea is to search the n-dimensional space in non-interfering directions for n parameters iteratively by using the gradient information. Powell has developed the conjugate gradient method into an efficient method without calculating derivatives [15]. Noting that the derivative of the cost function in Eq. (18) cannot be achieved, we adopt Powell's method to search the solution space. This method is able to guarantee the efficient convergence of each iterative procedure when searching along 12 independent directions that relate to the 12 parameters to be optimized. That is to say, there are 12 changes in each iteration cycle. The procedure of the iteration is as follows. The search starts from an initial value $\mathbf{x}_0^{(k)}, k = 1$ which represents the iteration number and the initial direction $\mathbf{s}_i (i = 1,2,\ldots,12)$ which represent the coordinate directions.

Step 1: For i = 1, 2, ..., 12, calculate the optimal step α_i satisfying the condition $F(\mathbf{x}_{i-1}^{(k)} + \alpha_i \mathbf{s}_i^{(k)}) = \min_{\alpha} F(\mathbf{x}_{i-1}^{(k)} + \alpha \mathbf{s}_i^{(k)})$ and define $\mathbf{x}_i = \mathbf{x}_{i-1} + \alpha_i \mathbf{s}_i$. The optimal step is obtained by golden section method which is based on interval elimination principle.

Step 2: Find the integer $r, 1 \le r \le 12$ meeting the requirement that $\left|F(\mathbf{x}_{r-1}^{(k)}) - F(\mathbf{x}_r^{(k)})\right|$ is a maximum, and define $\Delta F_r^{(k)} = \max_{r=1,2,..12} \left|F(\mathbf{x}_{r-1}^{(k)}) - F(\mathbf{x}_r^{(k)})\right|$.

Step 3: Calculate $F_3 = F(\mathbf{x}_{12+1}^{(k)})$, $\mathbf{x}_{12+1}^{(k)} = 2\mathbf{x}_{12}^{(k)} - \mathbf{x}_0^{(k)}$, and define $F_1 = F(\mathbf{x}_0^{(k)})$, $F_2 = F(\mathbf{x}_{12}^{(k)})$

Step 4: If either $F_3 < F_1$ or $(F_1 + F_3 - 2F_2)(F_1 - F_2 - \Delta F_r^{(k)})^2 < 0.5\Delta F_r^{(k)}(F_1 - F_3)^2$, do not change the current directions for the next iteration and let $\mathbf{x}_0^{k+1} = \mathbf{x}_{12}^{(k)}$. Otherwise, let $\mathbf{s}_{12+1} = \mathbf{x}_r^{(k)} - \mathbf{x}_0^{(k)}$ and replace $\mathbf{s}_r$ with $\mathbf{s}_{12+1}$, which is the steeply direction corresponding to $\Delta F_r^{(k)}$.

Step 5: If the result meets the optimization goal, the iteration terminates. Otherwise, let the iteration number $k=k+1$, and go to Step 1.

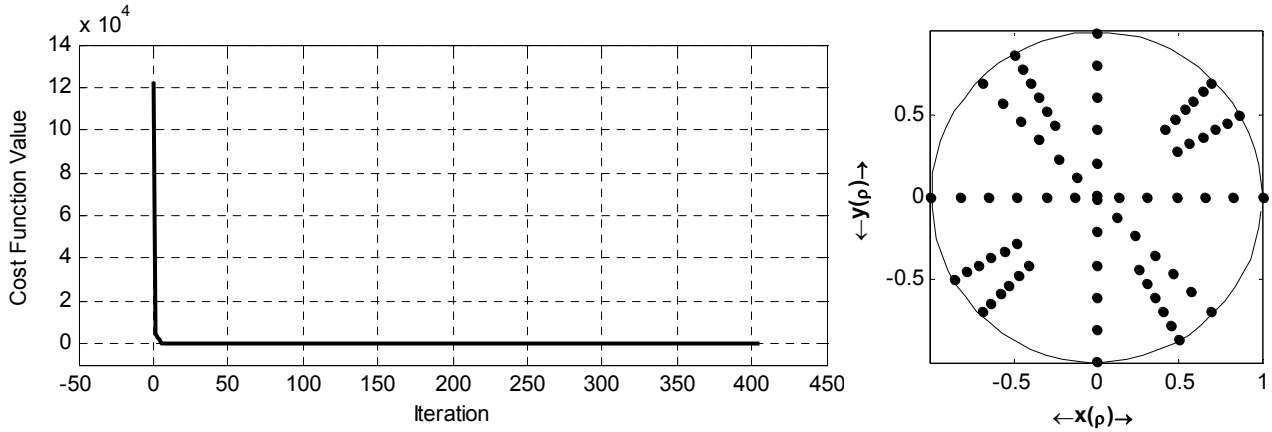
The output of the optimization iteration gives out the best pupil sampling scheme. Once the sampling scheme is determined, the periods of the binary gratings p can be obtained from the corresponding normalized radii ρ:

$$p = \frac{\lambda}{\rho \cdot M \cdot NA}. \tag{20}$$

where M is the magnification factor of the projection lens.

4. SIMULATION AND DISCUSSION

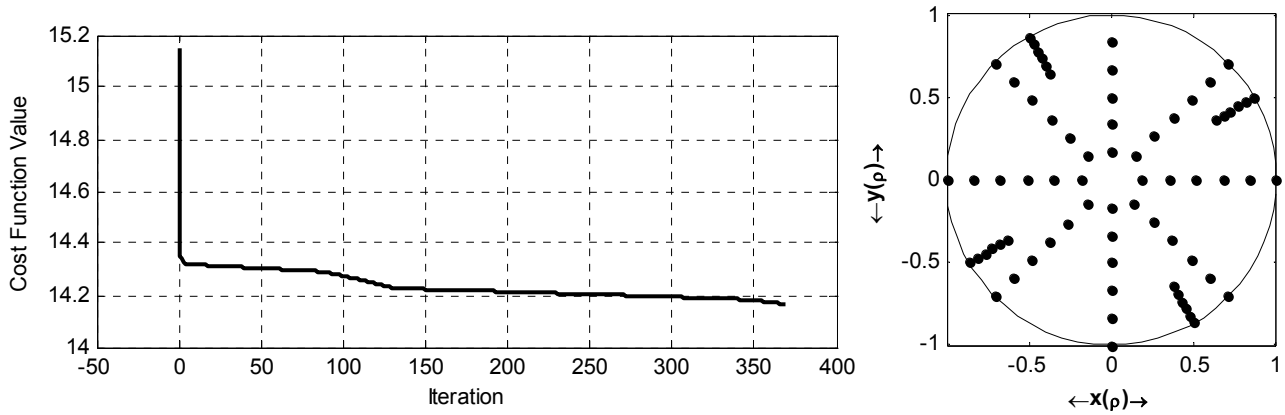
Simulations were conducted to testify the proposed optimization method. As described above, 6 angles of 0˚, 30˚, 45˚, 90˚, 120˚, and 135˚ were determined as the orientations of the sampling points at the pupil plane. An initial pupil sampling scheme was generated from randomly selected radii. During the optimization process, the penalty coefficients α_1 and α_2 were both set to be 10. The iteration of the optimization process and the optimized result are shown in Fig.5. It is clear that the initial pupil sampling scheme is quite poor as the value of cost function is as high as 1.236×10^5. The cost function value decreases sharply to the order of tens only after the first several iterations, and then it converges gradually to a minimal value of 14.21 after 405 iterations. This optimized sampling scheme is quite satisfactory as its corresponding condition number is as low as 4.21.



(a) The convergence of the cost function (b) The optimized sampling scheme

Fig.5 The optimization process with a very poor initial pupil sampling scheme.

In order to further verify the proposed optimization method, another pupil sampling scheme with a much lower value of cost function was generated as the initial input. The iteration of the optimization process and the optimized result are shown in Fig.6. As expected, the cost function decreases sharply from the initial value of 15.18 only after the first several iterations, and then it converges gradually to a minimal value of 14.16 after 369 iterations. This means that the proposed optimization method can achieve a further optimized result even though the initial input is good enough, and of course the trial and error method does not have this merit.



(a) The convergence of the cost function

(b) The optimized sampling scheme

Fig.6 The optimization process with a good enough initial pupil sampling scheme.

From Fig.5 and Fig.6, it is interesting to note that the optimized results are different due to the different initial sampling schemes, but the cost function values of the optimized results both approximate to the minimal value of 14. This means that whether the cost function value of the initial scheme is high or low, the optimization process can guarantee a desired output.

In order to demonstrate the influence of pupil sampling scheme on the error of aberration measurement, we further performed simulations using the lithographic simulator PROLITH. The wavelength in the simulations is 193nm, NA is 0.75, and the partial coherent factor σ is set to be 0.001 to simulate the nearly fully coherent illumination condition. The optimized sampling scheme (Scheme A) with cost function value of 14.16 shown in Fig.6(b) and a randomly selected sampling scheme (Scheme B) with cost function value of 31.68 shown in Fig.4(a) are used for comparison. The aberrated wavefront used as input is the sum of Zernike terms from Z_2 to Z_{37}. The measured Zernike coefficients for these two schemes and their errors to the input values are depicted in Fig.7.

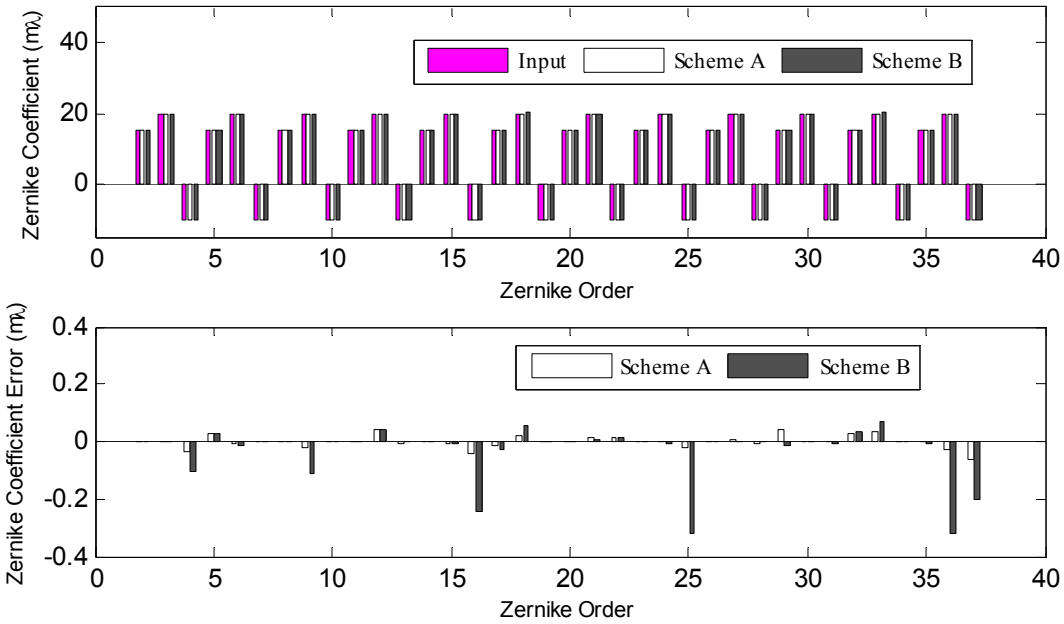


Fig.7 The measured Zernike coefficients and measurement errors for the two sampling schemes. Scheme A is shown in Fig.6(b) and Scheme B is shown in Fig.4(a).

From the results shown in Fig.8, it is obvious that the maximal measurement error of all Zernike coefficients is below 0.1mλ for Scheme A, but is nearly as high as 0.4mλ for Scheme B. The corresponding root mean squares (RMS) of the measurement error for these two sampling schemes are 0.021mλ and 0.098mλ respectively. These results demonstrate that the cost function value (or the condition number) of the sampling scheme has a significant influence on the accuracy of aberration measurement. This observation is further verified in Fig.8, in which 20 sets of Zernike coefficients from Z_2 to Z_{37} are generated randomly as inputs of PROLITH, and the measurement results are all combined together to calculate the RMS errors of each Zernike coefficient for the two schemes. Obviously, the errors detected by Scheme A with the lower cost function value are balanced, while some terms of errors induced by Scheme B are so high that will definitely decrease the accuracy of aberration measurement.

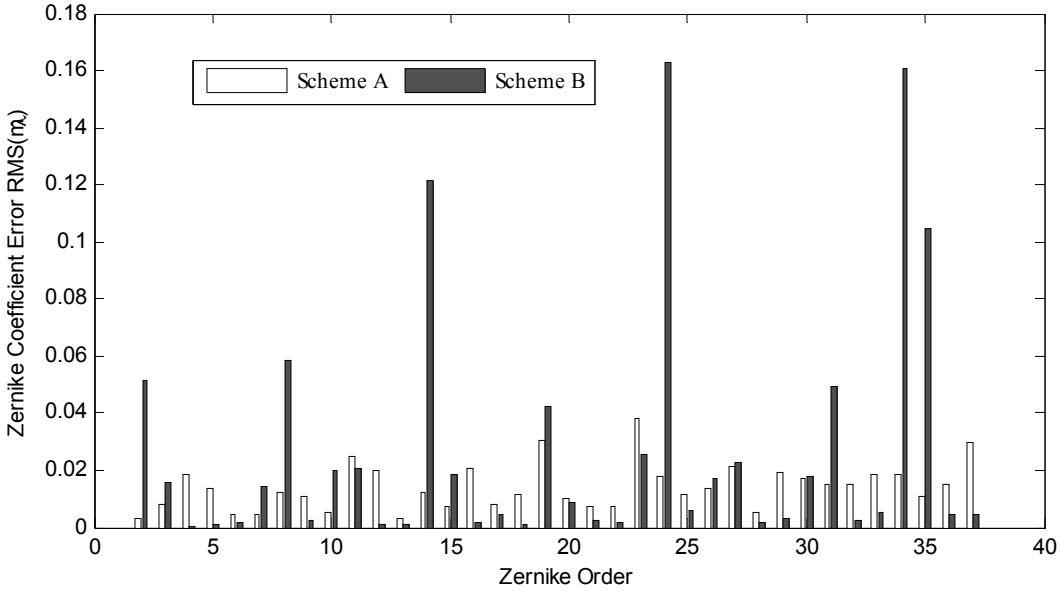


Fig.8 The RMS of measurement errors with 20 inputs for the two sampling schemes. Scheme A is shown in Fig.6(b) and Scheme B is shown in Fig.4(a).

5. CONCLUSIONS

In this paper, we have analyzed the aberration measurement of optics lens in lithographic tools based on aerial image sensor, and have focused on the optimization of pupil sampling schemes for this technique. Considering the simplicity for grating fabrication, 6 orientations with angles of 0˚, 30˚, 45˚, 90˚, 120˚, and 135˚ are preferred for the measurement of aberrations up to the 37th Zernike coefficient. A new approach based on Powell's optimization method has been proposed for the optimization of normalized radii, which introduces a continuous cost function and searches the solution space with gradient information. The simulation work has demonstrated that the pupil sampling scheme obtained by the proposed optimization approach are more suitable than those by the trial and error method. The results simulated by PROLITH have further verified that the measurement errors of Zernike coefficients by the optimized scheme with a lower cost function value are much less than those by the schemes with higher cost function values. It is expected that the proposed optimization approach will have potential applications not only in the Z37 AIS technique, but also in the other aerial image based techniques such as TAMIS.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China (Grant No. 50775090), National Basic Research Program of China (Grant No. 2009CB724204), National Hi-Tech Research and Development Program of China (Grant No. 2006AA04Z325), and Program for New Century Excellent Talents in University of China (Grant No. NCET-06-0639). The authors would like to thank National Engineering Research Center for Lithographic Equipment of

China for supporting this work and KLA-Tencor Corporation for providing an academic usage product of PROLITH™ software.

REFERENCES

[1] Smith, B. W. and Schlief, R., "Understanding lens aberration and influences to lithographic imaging," Proc. SPIE 4000, 294-306 (2000).
[2] Wang, F., Wang, X. and Ma, M., "Measurement technique for in situ characterizing aberrations of projection optics in lithographic tools," Appl. Opt. 45, 6086-6093 (2006).
[3] Van de Kerkhof, M., de Boeij, W., Kok, H., et al, "Full optical column characterization of DUV lithographic projection tools," Proc. SPIE 5377, 1960-1970 (2004).
[4] Fujii, T., Suzuki, K., Mizuno, Y. and Kita, N., "Integrated projecting optics tester for inspection of immersion ArF scanner," Proc. SPIE 6152, 615237 (2006).
[5] Ohsaki, Y., Mori, T., Koga, S., Ando, M., et al, "A new on-machine measurement system to measure wavefront aberrations of projection optics with hyper-NA," Proc. SPIE 6154, 615424 (2006).
[6] Schenau, K. I., Bakker, H., Zellenrath, M., et al, "System qualification and optimization for imaging performance on the 0.80-NA 248nm step and scan systems," Proc. SPIE 4691, 637-651 (2002).
[7] Laan, H., Dierichs, M., Greevenbroek, H., et al, "Aerial image measurement methods for fast aberration setup and illumination pupil verification," Proc. SPIE 4346, 394–407 (2001).
[8] Tyminski, J. K., Hagiwara, T., Kondo, N. and Irihama, H. "Aerial image sensor: in-situ scanner aberration monitor, " Proc. SPIE 6152, 61523D (2006).
[9] Higashibata, S., Hagiwara, T., Mizutani, H., Kondo, et al, "Aerial image sensor for self- calibration of wafer steppers," Proc.SPIE 4346, 1635-1643 (2001).
[10] Tsuneyuki, H., Masato, H., Naoto, K, Kosuke, S., et al, "Self calibration of wafer scanners using aerial image sensor," Proc. SPIE 4691, 871-881 (2002).
[11] Hagiwara, T., Kondo, N., Hiroshi, I., Suzuki, K. and Magome, N., "Development of aerial image based aberration measurement technique," Proc. SPIE 5754, 1659-1669 (2005).
[12] Zernike, F., "Beugungstheorie des Schneidenverfahrens und seiner verbesserten form, der Phasenkontrastmethode," Physica 1, 689-704 (1934).
[13] Ma, M., Wang, X., and Wang, F., "Aberration measurement of projection optics in lithographic tools based on two-beam interference theory," Appl. Opt. 45, 8200-8208 (2006).
[14] Hopkins, H., "Canonical pupil coordinates in geometrical and diffraction image theory," Jpn. J. Appl. Phys. 4, 31-35 (1965).
[15] Powell, M. J. D., "An efficient method for finding the minimum of a function of several variables without calculating derivatives," Comput. J. 7, 155-162 (1964).