Alternating phase-shifting mask design for low aberration sensitivity

Giuseppe Y. Mak
The University of Hong Kong
Department of Electrical and Electronic
    Engineering
Pokfulam Road, Hong Kong
E-mail: yhgmak@eee.hku.hk

Alfred K. Wong
Fortis Systems, Inc.
1850 Beacon St., #604
Brookline, Massachusetts 02445

Edmund Y. Lam
The University of Hong Kong
Department of Electrical and Electronic
    Engineering
Pokfulam Road, Hong Kong

Abstract. The aberration present in the lenses of exposure systems can cause placement errors to the images produced by alternating phase-shifting masks (PSMs). In reality, when the aberration signature varies from one lens to another, the magnitude of placement error also varies. It remains a question of how the alternating PSM should be designed, so that the image placement error, on average, can be minimized. To achieve this goal, we are interested in optimizing the phase width of an alternating PSM with a fixed critical dimension (CD). The constraint of the optimization is the mean of root mean square (rms) aberrations for a set of interest of exposure systems. To begin the analysis, the image placement error is expressed as a function of illumination, mask spectrum, and wave aberration. A Monte Carlo technique is then applied to produce random samples of wave aberration and image placement error. This analysis shows that a global minimum of mean image placement error is likely to occur at phase widths between $0.2[\lambda/\text{numerical aperture}$ (NA)] and $0.4(\lambda/\text{NA})$. This is further confirmed by analytically considering the expected value of the square of the image placement error. The methodology of finding the optimal phase width is applicable to the design of all alternating PSMs. © *2005 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.1861732]

Subject terms: aberration; alternating phase-shifting mask; image placement error; phase width; Monte Carlo analysis; Zernike coefficient.

Paper 04022 received May 4, 2004; revised manuscript received Jul. 8, 2004; accepted for publication Jul. 23, 2004; published online Feb. 4, 2005.

1 Introduction

In the production of integrated circuits, the miniaturization of devices and the rapid increase in integration density have led to the development of resolution enhancement techniques (RETs) in lithography.[1] Examples include modified illumination[2,3] (such as annular, dipole and quadrupole illumination), an alternating phase-shifting mask[4] (PSM), and an attenuated PSM (Ref. 5). An alternating PSM is one of the RETs that offers superior image quality for printing small and dark features. Making use of the destructive interference of light rays that are 180 deg out of phase, alternating PSMs are capable of both light-field and dark-field applications. Linewidth as low as 0.1 has been achieved[6] using alternating PSMs.

Current issues such as image intensity imbalance, aberration sensitivity, and mask defects present a challenge to the design of alternating PSMs. We focus on aberration sensitivity in this paper. Aberration is the departure from sphericity of image-forming light rays. There has been an extensive study on the relationship between aberration and PSMs. For example, spherical aberration is found[7] to degrade the depth of focus (DoF) of semirandomly aligned patterns printed by alternating PSMs. Coma, on the other hand, causes[8] critical dimension (CD) asymmetry in multiphase PSMs. Although it has been demonstrated that alter-

nating PSMs incur less intrafield linewidth variations than conventional chromium-on-glass (CoG) masks,[9] they generally result[10] in higher image placement error (i.e., the lateral shift of printed features). It is thus necessasry to design an alternating PSM that is less susceptible to image placement error.

From our previous study, the image placement error can be expressed as a function of effective light source, photomask structure, and wave aberration.[11] The formula enables us to predict the image placement error incurred by a particular exposure system. Given that the exposure system is unchanged, we can use the formula to optimize the alternating PSM easily. However, in the real world, there are many types of exposure tools. Each of them has a different aberration signature and produces varying degree of image placement error (Fig. 1). A mask optimized for one exposure tool may not be optimized for another. It is also cost-ineffective and time-consuming to redesign the mask when a new exposure system is installed. Therefore, it is our interest to optimize the alternating PSM so that the image placement error can be minimized in the average sense. To limit the scope of the problem, we consider an alternating PSM with symmetric 0- and 180-deg phase regions (transparent regions) and a fixed CD (the width of the opaque line between the two transparent regions) (Fig. 2). We aim at optimizing the width of the phase regions (hereafter as phase width). The constraint of the optimization is the mean of root mean square (rms) aberrations for an inter-
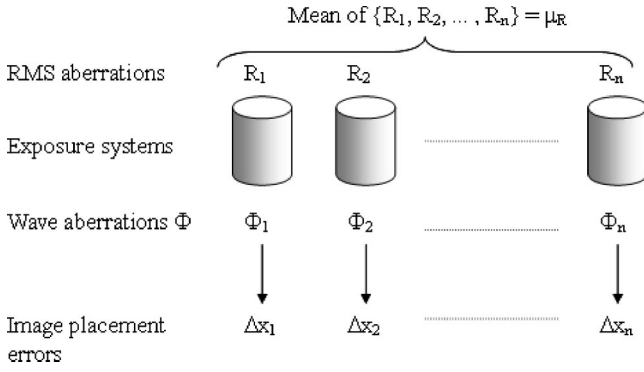
Fig. 1 Illustration of the optimization problem.

Fig. 3 Image formation by light rays from Gaussian reference sphere S and aberrated wavefront W.

ested set of exposure systems (Fig. 1). Coherent imaging is assumed here to simplify the analysis.[12]

2 Theory

2.1 Wave Aberration and rms Aberration

To begin, we take a closer look at how aberration is expressed mathematically. Figure 3 shows the formation of a point image. Without aberration, all transmitted light rays converge to the point image P_I. Taking P_I as the center, a spherical wavefront is formed at the exit pupil E. This wavefront is called Gaussian reference sphere S.

When aberration is present, the light rays no longer converge to a single point on the image plane. The aberrated wavefront W at the exit pupil deviates from the surface S. The optical path difference (OPD) between W and S is known[13] as the wave aberration Φ. Since Φ can be regarded as a surface over the exit pupil E, it is possible to use polynomials to fit this surface. The most common candidates are Zernike polynomials. Let Z_i denote the i'th Zernike polynomial (rms normalized), and C_i is the i'th Zernike coefficient. Sufficient for current lithographic applications, $\{C_5,\ldots,C_{37}\}$ is chosen to express the wave aberration Φ in this paper. The wave aberration at an arbitrary point (f,g) on plane E is then given by[14]

$$\Phi(f,g)=\sum_{i=5}^{37} C_i Z_i(f,g).\tag{1}$$

[Note that (f,g) are normalized by r/k, where r is the distance from the center of the object to (f,g) and k is the propagation number of the light rays.] In Fourier optics,
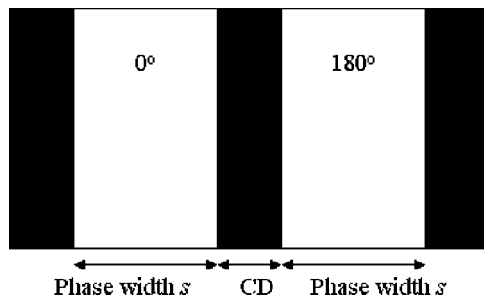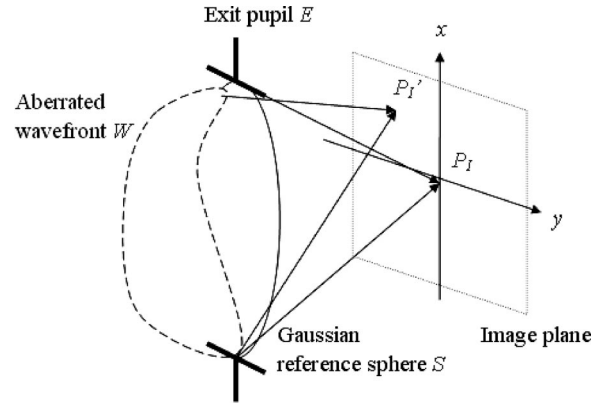

Fig. 2 Model of alternating PSM.

(f,g) are called spatial frequencies and the points on the image plane are called[15] spatial coordinates (x,y). Following the convention of projection lithography, the spatial frequencies are normalized by (NA/λ), while the spatial coordinates are normalized by (λ/NA), where NA is the numerical aperture, and λ is the wavelength. In the subsequent discussion, this is denoted by a caret ($\hat{}$) over the corresponding variables.

The rms aberration R is defined in terms of the Zernike coefficients as

$$R=\left(\sum_{i=5}^{37} C_i^2\right)^{1/2}.\tag{2}$$

This parameter characterizes the overall wavefront quality of an exposure system. From Eq. (2), we can see that one set of Zernike coefficients corresponds to one value of R. Different exposure systems have different sets of Zernike coefficients. When the number of exposure systems becomes very large, each Zernike coefficient can be treated as a continuous random variable with unknown probability density functions (pdfs). Since there is not much statistical research on how the Zernike coefficients are distributed in exposure systems, each coefficient is modeled as an independent, normally distributed random variable with zero mean and nonzero variance σ_c^2, i.e., $C_i \sim \mathcal{N}(0,\sigma_c^2)$ for $5 \leq i \leq 37$. It implies that the aberration present in various exposure systems is most likely very small. With this assumption, R^2 becomes a χ^2 random variable with n degrees of freedom.[16] The given mean value of rms aberrations in our constraint is the population mean of R (denoted as μ_R). Since μ_R is given by[17]

$$\mu_R=\sqrt{2}\sigma_c\frac{\Gamma[1/2(n+1)]}{\Gamma(1/2n)},\tag{3}$$

where n is the number of normal random variables ($n = 33$ in our case), and $\Gamma(p)$ is the gamma function, and σ_c is

$$\sigma_c=\frac{\mu_R\Gamma(16.5)}{\sqrt{2}\Gamma(17)}.\tag{4}$$

2.2 Image Placement Error Formula

To facilitate the optimization, we must express the image placement error as a function of effective light source, 1-D mask spectrum, and wave aberration. [Note that the alternating PSM is a mask with features that vary in one dimension only. A photomask with mask features varying in the $\hat{x}$ direction can be represented only by the mask transmission function $\hat{O}_x(\hat{x})$. This function defines the amplitude and phase of the transmitted light wave at every point of $\hat{x}$. The mask spectrum $\widetilde{O}_x(\hat{f})$ is the Fourier transform of the mask transmission function, where the tilde (˜) denotes a frequency-domain function.] Let the image placement error be $\Delta\hat{x}$. Also let the effective light source be $J(\hat{f}_s,\hat{g}_s)$, let the mask spectrum be $\widetilde{O}_x(\hat{f})$, and let the wave aberration be $\Phi(\hat{f},\hat{g})$. The image intensity is given by

$$I(\hat{x}) = \frac{\int\int_S J(\hat{f}_s,\hat{g}_s)I_s(\hat{f}_s;\hat{x})\mathrm{d}\hat{f}_s\,\mathrm{d}\hat{g}_s}{\int\int_S J(\hat{f}_s,\hat{g}_s)\mathrm{d}\hat{f}_s\,\mathrm{d}\hat{g}_s}, \tag{5}$$

where

$$I_s(\hat{f}_s,\hat{x}) = \int_{-1}^{1}\int_{-1}^{1}\widetilde{O}_x(\hat{f}_1-\hat{f}_s)\widetilde{O}_x^*(\hat{f}_2-\hat{f}_s)$$
$$\times\exp[-i2\pi(\hat{f}_{12}\hat{x}+\phi_{12})]\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2,$$

$\hat{f}_{12}=\hat{f}_1-\hat{f}_2$; and $\phi_{12}=\Phi(\hat{f}_1,0)-\Phi(\hat{f}_2,0)$. By setting $\mathrm{d}I/\mathrm{d}\hat{x}=0$, we have

$$\int\int_S J(\hat{f}_s,\hat{g}_s)(\mathrm{d}I_s/\mathrm{d}\hat{x})\mathrm{d}\hat{f}_s\,\mathrm{d}\hat{g}_s = 0. \tag{6}$$

We can write $I_s(\hat{f}_s,\hat{x})$ as

$$I_s(\hat{f}_s,\hat{x}) = \int_{-1}^{1}|\widetilde{O}_x(\hat{f}-\hat{f}_s)|^2\mathrm{d}\hat{f}$$
$$+ 2\,\mathrm{Re}\Bigg\{\int_{-1}^{1}\int_{\hat{f}_2}^{1}\widetilde{O}_x(\hat{f}_1-\hat{f}_s)\widetilde{O}_x^*(\hat{f}_2-\hat{f}_s)$$
$$\times\exp[-i2\pi(\hat{f}_{12}\hat{x}+\phi_{12})]\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2\Bigg\}, \tag{7}$$

where $\mathrm{Re}(\cdot)$ denotes the real part of $(\cdot)$. Then,

$$\frac{\mathrm{d}I_s}{\mathrm{d}\hat{x}} = 4\pi\int_{-1}^{1}\int_{\hat{f}_2}^{1}\mathrm{Im}\{\widetilde{O}_x(\hat{f}_1-\hat{f}_s)\widetilde{O}_x^*(\hat{f}_2-\hat{f}_s)$$
$$\times\exp[-i2\pi(\hat{f}_{12}\hat{x}+\phi_{12})]\}\hat{f}_{12}\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2, \tag{8}$$

where $\mathrm{Im}(\cdot)$ denotes the imaginary part of $(\cdot)$. Denoting $\mathrm{Im}(\cdot)$ in Eq. (8) as A, A can be simplified as

$$A = D_{12}(\hat{f}_s)\cos[2\pi(\hat{f}_{12}\hat{x}+\phi_{12})]$$
$$- S_{12}(\hat{f}_s)\sin[2\pi(\hat{f}_{12}\hat{x}+\phi_{12})], \tag{9}$$

where

$$D_{12}(\hat{f}_s) = \mathrm{Im}[\widetilde{O}_x(\hat{f}_1-\hat{f}_s)]\mathrm{Re}[\widetilde{O}_x(\hat{f}_2-\hat{f}_s)]$$
$$- \mathrm{Re}[\widetilde{O}_x(\hat{f}_1-\hat{f}_s)]\mathrm{Im}[\widetilde{O}_x(\hat{f}_2-\hat{f}_s)],$$

$$S_{12}(\hat{f}_s) = \mathrm{Re}[\widetilde{O}_x(\hat{f}_1-\hat{f}_s)]\mathrm{Re}[\widetilde{O}_x(\hat{f}_2-\hat{f}_s)]$$
$$+ \mathrm{Im}[\widetilde{O}_x(\hat{f}_1-\hat{f}_s)]\mathrm{Im}[\widetilde{O}_x(\hat{f}_2-\hat{f}_s)].$$

By expanding Eq. (9) into Taylor series and retaining the first-order terms, we have

$$A = D_{12}(\hat{f}_s)[\cos(2\pi\phi_{12}) - 2\pi\hat{f}_{12}\sin(2\pi\phi_{12})\hat{x}] - S_{12}(\hat{f}_s)$$
$$\times[\sin(2\pi\phi_{12}) + 2\pi\hat{f}_{12}\cos(2\pi\phi_{12})\hat{x}]. \tag{10}$$

Substituting Eq. (10) into Eq. (8) and putting the resulting equation into Eq. (6), we get a linear equation in $\hat{x}$. Solving for $\hat{x}$:

$$\hat{x} = \frac{\int\int_S J(\hat{f}_s,\hat{g}_s)M(\hat{f}_s)\mathrm{d}\hat{f}_s\,\mathrm{d}\hat{g}_s}{\int\int_S J(\hat{f}_s,\hat{g}_s)N(\hat{f}_s)\mathrm{d}\hat{f}_s\,\mathrm{d}\hat{g}_s}, \tag{11}$$

where

$$M(\hat{f}_s) = \int_{-1}^{1}\int_{\hat{f}_2}^{1}[D_{12}(\hat{f}_s)\cos(2\pi\phi_{12})$$
$$- S_{12}(\hat{f}_s)\sin(2\pi\phi_{12})]\hat{f}_{12}\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2,$$

$$N(\hat{f}_s) = 2\pi\int_{-1}^{1}\int_{\hat{f}_2}^{1}[D_{12}(\hat{f}_s)\sin(2\pi\phi_{12})$$
$$+ S_{12}(\hat{f}_s)\cos(2\pi\phi_{12})]\hat{f}_{12}^2\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2.$$

Without loss of generality, the image intensity extremum is assumed to exist at $\hat{x}=0$. Equation (11) is then equivalent to the image placement error $\Delta\hat{x}$.

We can simplify the equation by noting the following points. In coherent imaging, $J(\hat{f}_s,\hat{g}_s)=\delta(\hat{f}_s,\hat{g}_s)$, where $\delta(\hat{f}_s,\hat{g}_s)$ is the 2-D Dirac delta function. Equation (11) is then reduced to

$$\Delta\hat{x} = \frac{M(0)}{N(0)}. \tag{12}$$

Furthermore, taking a thin mask approximation, where there is no transmission or phase error in the phase regions, the mask spectrum of an alternating PSM with phase width $\hat{s}$ and critical dimension $\widehat{CD}$ is a purely imaginary function:

$$\widetilde{O}_x(\hat{f}) = \frac{i2}{\pi\hat{f}}\sin(\pi\hat{f}\hat{s})\sin[\pi\hat{f}(\widehat{CD}+\hat{s})], \tag{13}$$

where $i=\sqrt{-1}$. With this spectrum, the factor $D_{12}(0)$ in the numerator and the denominator of Eq. (12) is reduced to zero. The image placement error formula is now given by

Table 1 Parameters in Monte Carlo analysis.

Wavelength (λ)	248 nm (KrF laser)
NA	0.68
CD	0.3(λ/NA)
Mean rms aberration	0.025λ
Number of trials (i.e., number of exposure systems)	10,000

$$\Delta\hat{x} = \frac{\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)\sin(2\pi\phi_{12})\hat{f}_{12}\,\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2}{2\pi\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)\cos(2\pi\phi_{12})\hat{f}_{12}^2\,\mathrm{d}\hat{f}_1\,\mathrm{d}\hat{f}_2}. \tag{14}$$

Note that the value of $\Delta\hat{x}$ can be either positive or negative, which represents a shift toward the $+\hat{x}$ or $-\hat{x}$ direction, respectively. Since we are interested in only the absolute amount of image placement error in our optimization, we ignore the sign of $\Delta\hat{x}$ and consider $|\Delta\hat{x}|$ instead. Alternatively, we can also consider the quantity $(\Delta\hat{x})^2$. The reason why $(\Delta\hat{x})^2$ is considered is further explained in Sec. 4.

Due to the randomness of Zernike coefficients, the placement error becomes a random quantity (hereafter as $\Delta\hat{X}$, with capital letter meaning random variable). A phase width optimized for a particular combination of Zernike coefficients may not be optimal for another combination. This makes it necessary to perform the optimization in an average sense. Monte Carlo analysis is our first attempt to determine the relationship between mean image placement error and phase width. This is described in the next section.

3 Monte Carlo Analysis

The aim of Monte Carlo analysis on Zernike coefficients is to obtain some preliminary evidence on the existence of global minimum of average placement error at certain phase width. The results underscore the possibility of theoretical analysis in the subsequent sections.

The parameters used in the simulation are listed in Table 1. In the analysis, the mean rms aberration is taken to be 0.025λ. This is suggested in Ref. 18 as a guideline for lithographers to achieve the best-quality wavefront control. Using Eq. (4), the pdf of each coefficient is found to be $\mathcal{N}(0, 1.9228\times10^{-5}\lambda^2)$.

In each trial of Monte Carlo analysis, a set of Zernike coefficients is randomly sampled according to the normal distribution just mentioned. By substituting the coefficients (i.e., the wave aberration Φ) into Eq. (14), we get a sample of $|\Delta\hat{X}|$ and a sample of $(\Delta\hat{X})^2$ as functions of $\hat{s}$. After taking 10,000 trials in the simulation, the sample means of $|\Delta X|$ and $(\Delta X)^2$ [without normalization by $(\lambda$/NA)] are plotted against the phase width in Figs. 4 and 5. Observing the two plots, we see that there is remarkable similarity in the shape of the plots. They both peak at $\hat{s}=1(\lambda$/NA). Besides, both plots have their global minima occurring between $\hat{s}=0.2(\lambda$/NA) and $0.4(\lambda$/NA). The optimal phase width is likely to lie in this range. In Sec. 4, we return to our image placement error formulas to obtain the optimal phase width analytically.
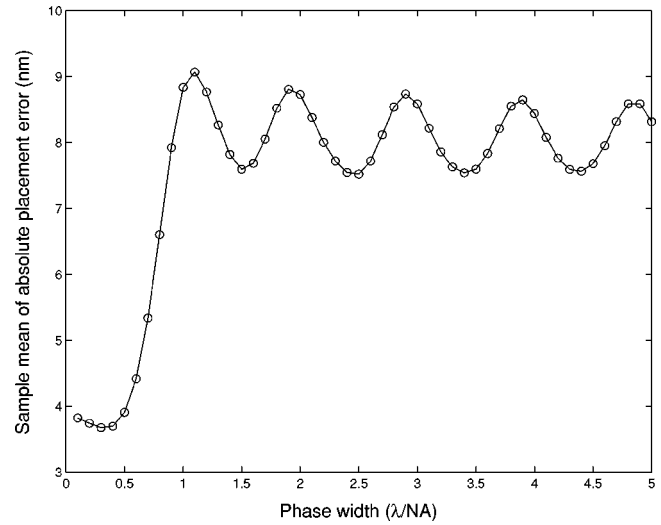
Fig. 4 Sample mean of $|\Delta X|$ as a function of $\hat{s}$ for $0.1\leq\hat{s}\leq5$.

4 Expected Placement Error and Optimal Phase Width

The encouraging results from Monte Carlo analysis prompt for a theoretical way in obtaining the optimal phase width. To this end, it is natural to consider the expected value of $|\Delta\hat{X}|$ [denoted by $E(|\Delta\hat{X}|)$, where $E(\cdot)$ stands for expected value operation]. However, without knowing the pdf of $\Delta\hat{X}$, it is a daunting task to determine its expected value. This is because from the mathematical point of view, expectation is an integral, but the absolute value operation is nonlinear. It is not possible to interchange the order of expectation and absolute value operation, i.e., $E(|\Delta\hat{X}|)$ is not identical to $|E(\Delta\hat{X})|$.

To overcome this problem, we consider the expected value of $(\Delta\hat{X})^2$ instead. It is possible to determine $E[(\Delta\hat{X})^2]$ without knowing the pdf of $\Delta\hat{X}$. The idea is as follows. Referring to Eq. (14), we note that both the numerator and the denominator depend on ϕ_{12}, which is the
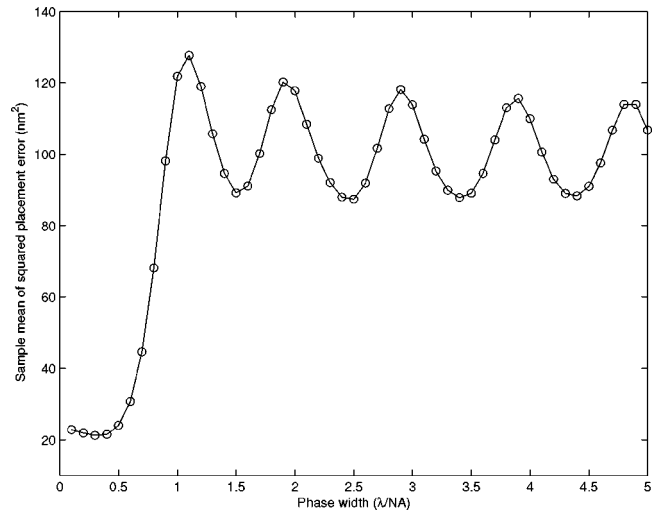
Fig. 5 Sample mean of $(\Delta X)^2$ as a function of $\hat{s}$ for $0.1\leq\hat{s}\leq5$.

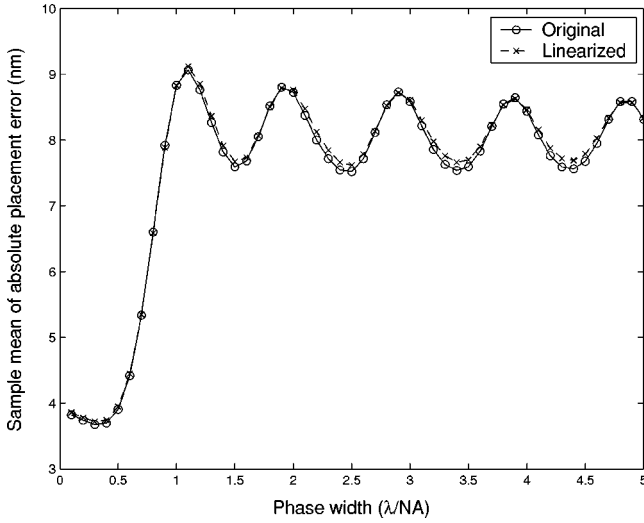**Fig. 6** Comparison between the sample means of $|\Delta X|$ and its linearized version.

Fig. 7 Comparison between the sample means of $(\Delta X)^2$ and its linearized version.

difference of the wave aberration function values evaluated at $(\hat{f}_1, 0)$ and $(\hat{f}_2, 0)$. The statistical parameters of ϕ_{12} (e.g., mean, variance, etc.) can be determined from the means and variances of the Zernike coefficients. If the expectation can be performed on ϕ_{12}, then our problem is mostly solved. The difficulty here is that Eq. (14) is a nonlinear operation on ϕ_{12}. With ϕ_{12} in the denominator, the order of expectation and the double integral in the numerator cannot be interchanged. To linearize Eq. (14), i.e., to remove the dependence on ϕ_{12} in the denominator, the following empirical approximation has been made to $\cos(2\pi\phi_{12})$ in the denominator of Eq. (14):

$$\cos(2\pi\phi_{12}) \approx \alpha(\hat{f}_1, \hat{f}_2)$$
$$= \begin{cases} 0.89 & 0.95 \leq |\hat{f}_1| \leq 1 \quad \text{or} \quad 0.95 \leq |\hat{f}_2| \leq 1 \\ 0.98 & \text{otherwise}. \end{cases}$$
(15)

The numbers in Eq. (15) are estimated from the sample mean and sample variance of $\cos(2\pi\phi_{12})$ (number of samples = 10,000). Their validity is verified by means of Monte Carlo analysis. All the parameters are the same as those in Table 1. We compare the sample mean of $|\Delta\hat{X}|$ obtained from Eq. (14) and the linearized version of $|\Delta\hat{X}|$, as well as the sample mean of $(\Delta\hat{X})^2$ and its linearized version. The results are shown in Figs. 6 and 7. In general, the plots from the linearized equations match the plots from the original equations. The match is better for low phase widths ($0 \leq \hat{s} \leq 1$).

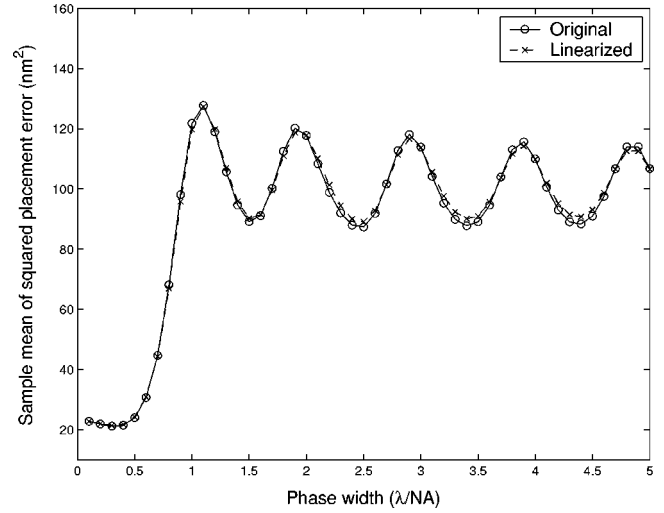With the linearization, $(\Delta\hat{X})^2$ can be expanded in the following manner:

$$(\Delta\hat{X})^2 = \left(\frac{\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)\sin(2\pi\phi_{12})\hat{f}_{12}\,d\hat{f}_1\,d\hat{f}_2}{2\pi\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)\alpha(\hat{f}_1,\hat{f}_2)\hat{f}_{12}^2\,d\hat{f}_1\,d\hat{f}_2} \right)^2$$

$$= \frac{1}{K}\int_{-1}^{1}\int_{\tilde{f}_4}^{1}\int_{-1}^{1}\int_{\tilde{f}_2}^{1} S_{12}(0)S_{34}(0)\sin(2\pi\phi_{12})$$
$$\times \sin(2\pi\phi_{34})\hat{f}_{12}\hat{f}_{34}\,d\hat{f}_1\,d\hat{f}_2\,d\hat{f}_3\,d\hat{f}_4,$$
(16)

where K is the denominator of the first step of Eq. (16). Note that K does not depend on the wave aberration Φ.

The integrand of Eq. (16) is continuous for all independent variables present in it. The order of expectation and integration can thus be interchanged. This is shown as follows:

$$E[(\Delta\hat{X})^2] = E\left[\frac{1}{K}\int_{-1}^{1}\int_{\hat{f}_4}^{1}\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)S_{34}(0)\sin(2\pi\phi_{12}) \right.$$
$$\left. \times \sin(2\pi\phi_{34})\hat{f}_{12}\hat{f}_{34}\,d\hat{f}_1\,d\hat{f}_2\,d\hat{f}_3\,d\hat{f}_4 \right]$$

$$= \frac{1}{K}\int_{-1}^{1}\int_{\hat{f}_4}^{1}\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)S_{34}(0)E[\sin(2\pi\phi_{12})$$
$$\times \sin(2\pi\phi_{34})]\hat{f}_{12}\hat{f}_{34}\,d\hat{f}_1\,d\hat{f}_2\,d\hat{f}_3\,d\hat{f}_4$$

$$= \frac{1}{2K}\int_{-1}^{1}\int_{\hat{f}_4}^{1}\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)S_{34}(0)$$
$$\times E\{\cos[2\pi(\phi_{12}-\phi_{34})] - \cos[2\pi(\phi_{12}+\phi_{34})]\} \cdot \hat{f}_{12}\hat{f}_{34}\,d\hat{f}_1\,d\hat{f}_2\,d\hat{f}_3\,d\hat{f}_4$$

$$= \frac{1}{2K}\int_{-1}^{1}\int_{\hat{f}_4}^{1}\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)S_{34}(0)$$
$$\times (E\{\cos[2\pi(\phi_{12}-\phi_{34})]\} - E\{\cos[2\pi(\phi_{12}+\phi_{34})]\})\hat{f}_{12}\hat{f}_{34}\,d\hat{f}_1\,d\hat{f}_2\,d\hat{f}_3\,d\hat{f}_4.$$
(17)

The expected values of the two cosine functions in Eq. (17) must be evaluated. First, we consider the pdf of $(\phi_{12} - \phi_{34})$ and $(\phi_{12} + \phi_{34})$. Since $C_i \sim \mathcal{N}(0, \sigma_c^2)$ for all i, $\Phi(\hat{f}, \hat{g}) \sim \mathcal{N}(0, \sigma_c^2 \Sigma_i [Z_i(\hat{f}, \hat{g})]^2)$. Hence, $(\phi_{12} - \phi_{34})$ and $(\phi_{12} + \phi_{34})$ are also normally distributed, with the mean and variance as

$$E(\phi_{12} - \phi_{34}) = E\left[\sum_i C_i(Z_{i1} - Z_{i2} - Z_{i3} + Z_{i4})\right]$$

$$= \sum_i (Z_{i1} - Z_{i2} - Z_{i3} + Z_{i4})E(C_i) = 0,$$

$$E(\phi_{12} + \phi_{34}) = E\left[\sum_i C_i(Z_{i1} - Z_{i2} + Z_{i3} - Z_{i4})\right]$$

$$= \sum_i (Z_{i1} - Z_{i2} + Z_{i3} - Z_{i4})E(C_i) = 0,$$

$$\mathrm{var}(\phi_{12} - \phi_{34}) = \mathrm{var}\left[\sum_i C_i(Z_{i1} - Z_{i2} - Z_{i3} + Z_{i4})\right]$$

$$= \sum_i (Z_{i1} - Z_{i2} - Z_{i3} + Z_{i4})^2 \, \mathrm{var}(C_i)$$

$$= \sigma_c^2 P_{1234},$$

$$\mathrm{var}(\phi_{12} + \phi_{34}) = \mathrm{var}\left[\sum_i C_i(Z_{i1} - Z_{i2} + Z_{i3} - Z_{i4})\right]$$

$$= \sum_i (Z_{i1} - Z_{i2} + Z_{i3} - Z_{i4})^2 \, \mathrm{var}(C_i)$$

$$= \sigma_c^2 Q_{1234},$$

where $Z_{ij} = Z_i(\hat{f}_j, 0)$, for $j = 1, 2, 3, 4$, and

$$P_{1234} = \sum_i (Z_{i1} - Z_{i2} - Z_{i3} + Z_{i4})^2,$$

$$Q_{1234} = \sum_i (Z_{i1} - Z_{i2} + Z_{i3} - Z_{i4})^2,$$

i.e., functions of the Zernike polynomials only.

Back to our original problem in Eq. (17), we must consider the expected value of the cosine of a normal random variable. Let $Y \sim \mathcal{N}(\mu, \sigma^2)$. The moment generating function of Y is[19]

$$E(e^{tY}) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right), \tag{18}$$

for any complex t. Then,

$$E(\cos aY) = E\left[\frac{\exp(iaY) + \exp(-iaY)}{2}\right]$$

$$= \frac{1}{2}\{E[\exp(iaY)] + E[\exp(-iaY)]\}$$

$$= \frac{1}{2}\left[\exp\left(ia\mu - \frac{a^2\sigma^2}{2}\right) - \exp\left(-ia\mu - \frac{a^2\sigma^2}{2}\right)\right]$$

$$= \exp\left(-\frac{a^2\sigma^2}{2}\right)\cos a\mu, \tag{19}$$

where a is a constant. By substituting $a = 2\pi$, $Y = \phi_{12} - \phi_{34}$ (or $\phi_{12} + \phi_{34}$), we have

$$E\{\cos[2\pi(\phi_{12} - \phi_{34})]\} = \exp[-2\pi^2 \, \mathrm{var}(\phi_{12} - \phi_{34})]$$
$$\times \cos[2\pi E(\phi_{12} - \phi_{34})]$$
$$= \exp(-2\pi^2\sigma_c^2 P_{1234}),$$

$$E\{\cos[2\pi(\phi_{12} + \phi_{34})]\} = \exp[-2\pi^2 \, \mathrm{var}(\phi_{12} + \phi_{34})]$$
$$\times \cos[2\pi E(\phi_{12} + \phi_{34})]$$
$$= \exp(-2\pi^2\sigma_c^2 Q_{1234}). \tag{20}$$

Finally, $E[(\Delta\hat{X})^2]$ is given by

$$E[(\Delta\hat{X})^2] = \frac{A}{8\pi^2 B}, \tag{21}$$

where

$$A = \int_{-1}^{1}\int_{\hat{f}_4}^{1}\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)S_{34}(0)[\exp(-2\pi^2\sigma_c^2 P_{1234})$$
$$- \exp(-2\pi^2\sigma_c^2 Q_{1234})]\hat{f}_{12}\hat{f}_{34}\,d\hat{f}_1\,d\hat{f}_2\,d\hat{f}_3\,d\hat{f}_4,$$

$$B = \left[\int_{-1}^{1}\int_{\hat{f}_2}^{1} S_{12}(0)\alpha(\hat{f}_1, \hat{f}_2)\hat{f}_{12}^2\,d\hat{f}_1\,d\hat{f}_2\right]^2.$$

The essence of Eq. (21) is that $E[(\Delta\hat{X})^2]$ is a function of two parameters only—the phase width $\hat{s}$ and the variance of the Zernike coefficients σ_c^2:

$$E[(\Delta\hat{X})^2] = u(\hat{s}, \sigma_c^2). \tag{22}$$

This is a very pleasant result because σ_c^2 is related to the mean rms aberration by Eq. (4). In other words, given a certain mean rms aberration, $E[(\Delta\hat{X})^2]$ depends on the phase width only.

In Fig. 8, the values of $E[(\Delta\hat{X})^2]$ computed from the Monte Carlo analysis of Eq. (14) and those computed from Eq. (21) are plotted. A similar plot with the linearized version is shown in Fig. 9. This is done to show the validity of Eq. (21). In both figures, the results generally follow the same trend, with better agreement at low phase widths. Since we are interested in only low phase widths, Eq. (21) remains valid.
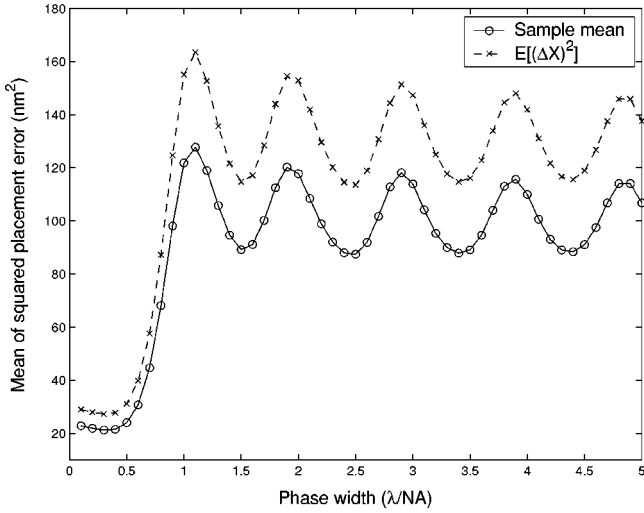
Fig. 8 Comparison between the sample mean of $(\Delta X)^2$ obtained from the Monte Carlo analysis of Eq. (14) and $E[(\Delta X)^2]$ from Eq. (21).

Fig. 10 Optimal phase width as a function of rms aberration; CD $=0.3(\lambda/\text{NA})$.

Equipped with Eq. (21), we can use numerical methods to determine the position of the global minimum. We do not proceed with the differentiation of Eq. (21) with respect to $\hat{s}$, because Eq. (21) is still a complicated function of $\hat{s}$. We chose the Golden Section Search method for its quick convergence. The range of search is limited to $0 \leqslant \hat{s} \leqslant 1$. After iterations, the optimal phase width is found to be $0.3217(\lambda/\text{NA})$ (i.e., approximately 117 nm) at 0.025λ mean rms aberration. The optimal phase width as a function of mean rms aberration is also plotted in Fig. 10. This curve shows that the optimal phase width decreases with increasing aberration level.

5 Discussion

There are several points to note in the foregoing analysis. First, the theory can be extended to any masks whose spec-
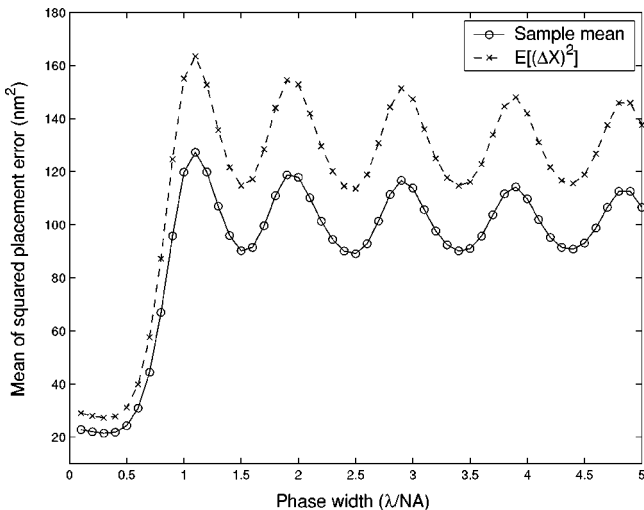

Fig. 9 Comparison between the sample mean of $(\Delta X)^2$ obtained from the Monte Carlo analysis of the linearized Eq. (14) and $E[(\Delta X)^2]$ from Eq. (21).
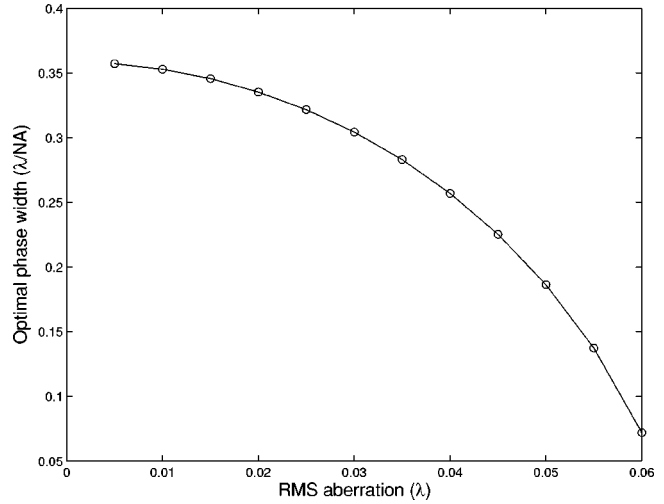
tra are, in general, complex functions with nonzero real parts. Examples include symmetric alternating PSMs with phase error in the phase shifters and asymmetric alternating PSMs. The method is briefly outlined as follows. By considering Eq. (14) again, we can see that if $\text{Re}[\widetilde{\hat{O}_x}(\hat{f}-\hat{f}_s)] \neq 0$, $D_{12}(\hat{f}_s) \neq 0$. By retaining $D_{12}(\hat{f}_s)$ in the derivation, we can arrive at a new and more complicated expression for $E[(\Delta\hat{X})^2]$. This expressions helps us to analyze the aberration sensitvity for any mask patterns under coherent imaging.

Second, without taking $J(\hat{f}_s,\hat{g}_s) = \delta(\hat{f}_s,\hat{g}_s)$, the theory can also be extended from coherent imaging to a general imaging condition, such as partially coherent imaging and other modified illuminations. Together with the first point, a theory applied to any mask patterns and any light sources can be obtained.

Finally, the optimality of phase width is also determined by the necessity to maintain adequate image quality for CD control. Figure 11 plots the simulated exposure latitude of a $0.3(\lambda/\text{NA})$ line as a function of phase width. The exposure latitude decreases from 30 to 18% as the phase width decreases from $0.8(\lambda/\text{NA})$ to $0.1(\lambda/\text{NA})$. A trade-off exists between placement sensivity and process window. A manufacturable process requires at least a 15% exposure latitude.[20] If this requirement is tightened, the optimal phase width should be adjusted accordingly.

6 Summary

Optimization was performed on the phase widths of alternating PSMs. The aim is to minimize the mean image placement error toward aberration under coherent imaging. The constraint was a given mean rms aberration for a set of exposure systems. We first expressed the image placement error as a function of effective light source, mask spectrum, and wave aberration. By randomly generating wave aberrations that conform to our constraint, we performed Monte Carlo analysis to the absolute image placement error $|\Delta\hat{X}|$ and the square of placement error $(\Delta\hat{X})^2$. From the results of the Monte Carlo analysis, on average, a global minimum
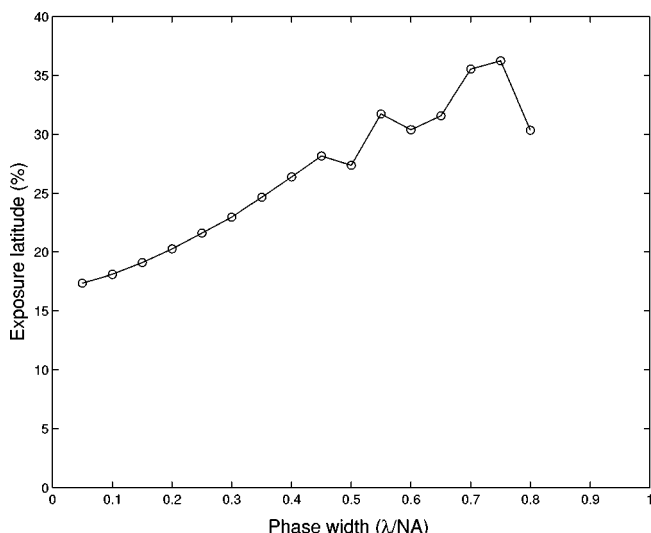
Fig. 11 Exposure latitude of a 0.3(λ/NA) line printed with alternating PSM increases with phase width. The image was simulated with partial coherence factor=0.5, and CD=0.3(λ/NA).

of placement error is likely to occur at a phase width between 0.2(λ/NA) and 0.4(λ/NA). By the theoretical consideration of the expected value of $(\Delta \hat{X})^2$, the optimal phase width of alternating PSM is obtained as a function of mean rms aberration. The results are generally applicable to the design of all alternating PSMs.

References

1. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, WA (2001).
2. A. Yen, W. N. Partlo, S. R. Palmer, M. A. Hanratty, and M. C. Tipton, "Quarter-micron lithography using a deep-UV stepper with modified illumination," *Proc. SPIE* **1927**, 158–166 (1993).
3. M. D. Himel, M. K. Poutous, J. D. Stack, and J. L. Leonard, "Microfabrication of controlled angle diffusers used for resolution enhancement in microlithography," in *Micromachining Technology for Micro-Optics and Nano-Optics*, E. G. Johnson, Ed., *Proc. SPIE* **4984**, 230–233 (Jan. 2003).
4. M. Levenson, N. Viswanathan, and R. Simpson, "Improving resolution in photolithography with a phase-shifting mask," *IEEE Trans. Electron Devices* **29**, 1812–1846 (Dec. 1982).
5. B. J. Lin, "The attenuated phase-shifting mask," *Solid State Technol.* **35**, 43–47 (Jan. 1992).
6. S. Nakao, J. Itoh, A. Nakae, I. Kanai, T. Saitoh, H. Matsubara, K. Tsujita, I. Arimoto, and W. Wakamiya, "Extension of KrF lithography to sub-50-nm pattern formation," in *Optical Microlithography XIII*, C. J. Progler, Ed., *Proc. SPIE* **4000**, 358–365 (July 2000).
7. A. Nakae, S. Nakao, and Y. Matsui, "Proposal for pattern layout rule in application of alternating phase-shift mask," in *Photomask and X-Ray Mask Technology IV*, N. Aizaki, Ed., *Proc. SPIE* **3096**, 362–374 (July 1997).
8. R. T. Schmidt, C. A. Spence, L. Capodieci, Z. Krivokapic, B. Geh, and D. G. Flagello, "Impact of coma on CD control for multiphase PSM designs," in *Optical Microlithography XI*, L. V. den Hove, Ed., *Proc. SPIE* **3334**, 15–24 (June 1998).
9. H. Y. Liu, "The application of alternating phase-shifting masks to 140 nm gate patterning: II. Mask design and manufacturing tolerances," in *Optical Microlithography XI*, L. V. den Hove, Ed., *Proc. SPIE* **3334**, 2–14 (June 1998).
10. R. J. Socha, W. Conley, X. Shi, M. V. Dusa, J. S. Petersen, J. F. Chen, K. E. Wampler, T. L. Laidig, and R. F. Caldwell, "Resolution enhancement with high-transmission attenuating phase-shift masks," in *Photomask and X-Ray Mask Technology VI*, H. Morimoto, Ed., *Proc. SPIE* **3748**, 290–314 (Aug. 1999).
11. G. Y. H. Mak, A. K. Wong, and E. Y. Lam, "Placement sensitivity to aberration in optical imaging," in *Proc. IEEE Conf. on Electron Devices and Solid-State Circuits*, pp. 475–478 (Dec. 2003).
12. G. Y. Mak, A. K. Wong, and E. Y. Lam, "Alternating phase-shifting mask design for low aberration sensitivity," in *Optical Microlithography XVII*, B. W. Smith, Ed., *Proc. SPIE* **5377**, 581–591 (Feb. 2004).
13. J. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996).
14. V. N. Mahajan, "Zernike polynomials and aberration balancing," in *Current Developments in Lens Design and Optical Engineering IV*, P. Z. Mouroulis, W. J. Smith, and R. B. Johnson, Eds., *Proc. SPIE* **5173**, 1–17 (Nov. 2003).
15. E. Hecht, *Optics*, Chap. 11. Addison Wesley, San Francisco, CA (2002).
16. J. Goodman, *Statistical Optics*, Wiley, New York (2000).
17. J. G. Proakis, *Digital Communications*, 3rd ed., pp. 46–47, McGraw-Hill, New York (1995).
18. C. J. Progler and D. C. Wheeler, "Optical lens specifications from the user's perspective," in *Optical Microlithography XI*, L. V. den Hove, Ed., *Proc. SPIE* **3334**, 256–268 (June 1998).
19. R. J. Larsen and M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, 3rd ed., p. 281. Prentice Hall, Upper Saddle River, NJ (2001).
20. A. K. Wong, "Theoretical discussion on reduced aberration sensitivity of enhanced alternating phase-shifting masks," in *Optical Microlithography XV*, A. Yen, Ed., *Proc. SPIE* **4691**, 359–368 (July 2002).

Giuseppe Y. Mak received his B Eng degree in electrical and electronic engineering from the University of Hong Kong, HK-SAR in 2002. He is currently a Master of Philosophy student in the Department of Electrical and Electronic Engineering at the University of Hong Kong. His research interests include microlithography and optical imaging system design.

Alfred K. Wong received his BS, MS, and PhD degrees in electrical engineering from the University of California, Berkeley, in 1990, 1992, and 1994 respectively. His research focused on electromagnetic simulation using the time-domain finite-difference (TDFD) method on massively parallel computers. He was affiliated with the Advanced Silicon Processing Division at IMEC in Belgium from 1994 to 1995, the IBM Semiconductor Research and Development Center from 1995 to 2000 in Hopewell Junction, New York, USA, and the Department of Electrical and Electronic Engineering at the University of Hong Kong from 2000 to 2003. His current research interests include resolution enhancement techniques and synergism between circuit design and fabrication.

Edmund Y. Lam received his BS degree in 1995, MS degree in 1996, and PhD degree in 2000, all in electrical engineering from Stanford University, Stanford, California. At Stanford, he conducted research for the Programmable Digital Camera project. He also consulted for industry in the areas of digital camera systems design and algorithms development. Before returning to academia, he was affiliated with the Reticle and Photomask Inspection Division (RAPID) of KLA-Tencor Corporation in San Jose, California, as a senior engineer. He was involved in the design of defect detection tools for the core die-to-die and die-to-database inspections. He is now an assistant professor of electrical and electronic engineering at the University of Hong Kong. His research interests include optics and imaging, and their use in the semiconductor manufacturing process.