Efficient source mask optimization with Zernike polynomial functions for source representation

Xiaofei Wu,[1] **Shiyuan Liu,**[2] **Jia Li,**[1] **and Edmund Y. Lam**[1,*]

[1]*Imaging Systems Laboratory, Department of Electrical and Electronic Engineering,*
The University of Hong Kong, Pokfulam, Hong Kong
[2]*State Key Laboratory of Digital Manufacturing Equipment and Technology,*
Huazhong University of Science and Technology, Wuhan, China

[*]*elam@eee.hku.hk*

Abstract: In 22 nm optical lithography and beyond, source mask optimization (SMO) becomes vital for the continuation of advanced ArF technology node development. The pixel-based method permits a large solution space, but involves a time-consuming optimization procedure because of the large number of pixel variables. In this paper, we introduce the Zernike polynomials as basis functions to represent the source patterns, and propose an improved SMO algorithm with this representation. The source patterns are decomposed into the weighted superposition of some well-chosen Zernike polynomial functions, and the number of variables decreases significantly. We compare the computation efficiency and optimization performance between the proposed method and the conventional pixel-based algorithm. Simulation results demonstrate that the former can obtain substantial speedup of source optimization while improving the pattern fidelity at the same time.

References and links

1. A. K. Wong, *Resolution Enhancement Technologies in Optical Lithography* (SPIE, 2001).
2. A. E. Rosenbluth, D. O. Melville, K. Tian, S. Bagheri, J. Tirapu-Azpiroz, K. Lai, A. Waechter, T. Inoue, L. Ladanyi, F. Barahona, K. Scheinberg, M. Sakamoto, H. Muta, E. Gallagher, T. Faure, M. Hibbs, A. Tritchkov, and Y. Granik, "Intensive optimization of masks and sources for 22nm lithography," Proc. SPIE **7274**, 727409 (2009).
3. M. Mulder, A. Engelen, O. Noordman, G. Streutker, B. van Drieenhuizen, C. van Nuenen, W. Endendijk, J. Verbeeck, W. Bouman, A. Bouma, R. Kazinczi, R. Socha, D. Jürgens, J. Zimmermann, B. Trauter, J. Bekaert, B. Laenens, D. Corliss, and G. McIntyre, "Performance of FlexRay: a fully programmable illumination system for generation of freeform sources on high NA immersion systems," Proc. SPIE **7640**, 76401P (2010).
4. K. Tian, M. Fakyry, A. Dave, A. Tritchkov, J. Tirapu-Azpiroz, A. E. Rosenbluth, D. Melville, M. Sakamoto, T. Inoue, S. Mansfield, A. Wei, Y. Kim, B. Durgan, K. Adam, G. Berger, G. Bhatara, J. Meiring, H. Haffner, and B.-S. Kim, "Applicability of global source mask optimization to 22/20nm node and beyond," Proc. SPIE **7973**, 79730C (2011).
5. N. Jia and E. Y. Lam, "Pixelated source mask optimization for process robustness in optical lithography," Opt. Express **19**, 19384–19398 (2011).
6. T. Fühner and A. Erdmann, "Improved mask and source representations for automatic optimization of lithographic process conditions using a genetic algorithm," Proc. SPIE **5754**, 415–426 (2005).
7. J. Li, S. Liu, and E. Y. Lam, "Efficient source and mask optimization with augmented Lagrangian methods in optical lithography," Opt. Express **21**, 8076–8090 (2013).

8. N. Jia and E. Y. Lam, "Machine learning for inverse lithography: using stochastic gradient descent for robust photomask synthesis," J. Opt. **12**, 045601 (2010).

9. K. Lai, A. E. Rosenbluth, S. Bagheri, J. Hoffnagle, K. Tian, D. Melville, J. Tirapu-Azpiroz, M. Fakhry, Y. Kim, S. Halle, G. McIntyre, A. Wagner, G. Burr, M. Burkhardt, D. Corliss, E. Gallagher, T. Faure, M. Hibbs, D. Flagello, J. Zimmermann, B. Kneer, F. Rohmund, F. Hartung, C. Hennerkes, M. Maul, R. Kazinczi, A. Engelen, R. Carpaij, R. Groenendijk, J. Hageman, and C. Russ, "Experimental result and simulation analysis for the use of pixelated illumination from source mask optimization for 22-nm logic lithography process," Proc. SPIE **7274**, 72740A (2009).

10. J. Bekaert, P. van Adrichem, R. Socha, O. Mouraille, J. Zimmermann, P. Grupner, K. Schreel, S. Hsu, B. Laenens, S. Verhaegen, H.-Y. Liu, M. Dusa, J. T. Neumann, L. V. Look, D. Trivkovic, F. Lazzarino, and G. Vandenberghe, "Experimental verification of source-mask optimization and freeform illumination for 22-nm node static random access memory cells," J. Micro/Nanolith. MEMS MOEMS **10**, 013008 (2011).

11. E. Y. Lam and A. K. Wong, "Computation lithography: virtual reality and virtual virtuality," Opt. Express **17**, 12259–12268 (2009).

12. E. Y. Lam and A. K. Wong, "Nebulous hotspot and algorithm variability in computation lithography," J. Micro/Nanolith. MEMS MOEMS **9**, 033002 (2010).

13. D. G. Flagello and D. G. Smith, "Calculation and uses of the lithographic aerial image," Adv. Opt. Technol. **1**, 237–248 (2012).

14. Y. Granik, "Source optimization for image fidelity and throughput," J. Microlith. Microfab. Microsys. **3**, 509–522 (2004).

15. A. Poonawala, W. Stanton, and C. Sawh, "Source mask optimization for advanced lithography nodes," Proc. SPIE **7640**, 76401M (2010).

16. A. E. Rosenbluth, S. Bukofsky, C. Fonseca, M. Hibbs, K. Lai, A. F. Molless, R. N. Singh, and A. K. K. Wong, "Optimum mask and source patterns to print a given shape," J. Microlith. Microfab. Microsys. **1**, 13 (2002).

17. S. Liu, W. Liu, X. Zhou, and P. Gong, "Kernel-based parametric analytical model of source intensity distributions in lithographic tools," Appl. Opt. **51**, 1479–1486 (2012).

18. J. Li, Y. Shen, and E. Y. Lam, "Hotspot-aware fast source and mask optimization," Opt. Express **20**, 21792–21804 (2012).

19. Y. Peng, J. Zhang, Y. Wang, and Z. Yu, "Gradient-based source and mask optimization in optical lithography," IEEE Trans. Image Process. **20**, 2856–2864 (2011).

20. X. Ma, C. Han, Y. Li, L. Dong, and G. R. Arce, "Pixelated source and mask optimization for immersion lithography," J. Opt. Soc. Am. A **30**, 112–123 (2013).

21. T. Mülders, V. Domnenko, B. Küchler, T. Klimpel, H.-J. Stock, A. A. Poonawala, K. N. Taravade, and W. A. Stanton, "Simultaneous source-mask optimization: a numerical combining method," Proc. SPIE **7823**, 78233X (2010).

22. J.-C. Yu, P. Yu, and H.-Y. Chao, "Library-based illumination synthesis for critical CMOS patterning," IEEE Trans. Image Process. **22**, 2811–2821 (2013).

23. J. Y. Wang and D. E. Silva, "Wave-front interpretation with Zernike polynomials," Appl. Opt. **19**, 1510–1518 (1980).

24. M. Born and E. Wolf, *Principle of Optics*, 7th ed. (Cambridge University, 1999).

25. S. Liu, X. Zhou, W. Lv, S. Xu, and H. Wei, "Convolution-variation separation method for efficient modeling of optical lithography," Opt. Lett. **38**, 2168–2170 (2013).

26. J. Aluizio Prata and W. V. T. Rusch, "Algorithm for computation of Zernike polynomials expansion coefficients," Appl. Opt. **28**, 749–754 (1989).

27. S. Liu, X. Wu, W. Liu, and C. Zhang, "Fast aerial image simulations using one basis mask pattern for optical proximity correction," J. Vac. Sci. Technol. B **29**, 06FH03 (2011).

28. A. K. Wong, *Optical Imaging in Projection Microlithography* (SPIE, 2005).

29. J. L. Sturtevant, L. Hong, S. Jayaram, S. P. Renwick, M. McCallum, and P. D. Bisschop, "Impact of illumination source symmetrization in OPC," Proc. SPIE **7028**, 70283M (2008).

30. J.-C. Yu, P. Yu, and H.-Y. Chao, "Fast source optimization involving quadratic line-contour objectives for the resist image," Opt. Express **20**, 8161–8174 (2012).

31. M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx (2013).

1. Introduction

As the critical dimension (CD) continues to shrink in the semiconductor industry, the continuation of ArF optical lithography depends heavily on resolution enhancement techniques (RETs) [1]. Source mask optimization (SMO) as one of the RETs becomes critical in 22 nm technological node and beyond since it provides a viable and powerful approach to scale down

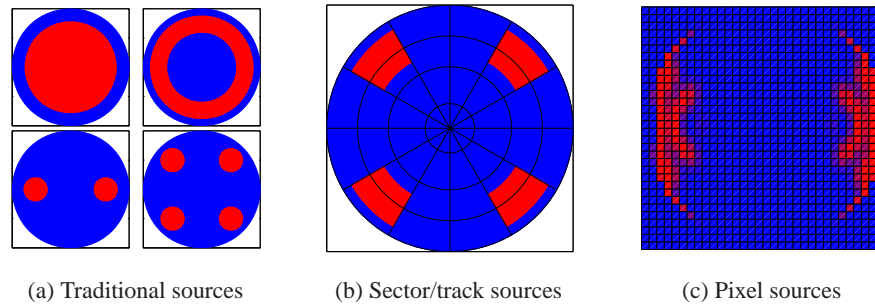(a) Traditional sources   (b) Sector/track sources   (c) Pixel sources

Fig. 1. Source representation methods in optical lithography.

the resolution [2]. This is because highly customized sources are available by using diffractive optical element (DOE) or programmable illumination, which can shape the light to free-form with little throughput loss [3, 4]. At the same time, the SMO process is carried out by various algorithms including the gradient-based method, the genetic algorithm, and more recently the augmented Lagrangian method for speed enhancement [5–7]. The algorithm is extended to take robustness to process variations into account [8]. Simulations and experiments of SMO are also performed to demonstrate its applicability in integrated circuits fabrication [9, 10].

SMO is usually carried out through analysis of the aerial image generated on the wafer plane and inverse optimization for the mask and source designs [11–13]. In this process, source representation methods play a critical role, and they affect the optimization performance and the efficiency significantly [14]. As shown in Fig. 1, the source pattern in optical lithography has evolved from the traditional circular, annular, dipole, and quadrupole sources, to more complicated shapes such as sectors/track, and more recently to pixel-based sources. The traditional source patterns need only one or several parameters for its description [15]; more customized sources represented by arcs, sectors/tracks and so on, however, use dozens of variables, and bring larger flexibility [6, 16]. For these sources, the number of variables to represent the source patterns is small, and the source optimization problem is also of a limited scale. However, these representation methods lead to a nonlinear relationship between the aerial image and the variables, and thus the optimization requires a nonlinear optimization problem. More importantly, the source patterns described by these methods are binary, largely limiting the freedom of the source patterns and the optimization performance [14]. A kernel based parametric model is also proposed that can represent the physical distribution of real-world illumination sources [17]. However, the nonlinear relationship still remains, which make it difficult to be incorporated in source optimization.

In contrast, most of the recent SMO algorithms compute a source pattern represented by grayscale pixels [18–20]. In these algorithms, the source patterns are discretized into matrices according to a specified pixel size, and each entry of the matrices is a variable [21]. The grayscale pixel can represent a continuum of real numbers from 0 to 1, and the freedom of the solution space is then greatly enlarged. These methods can take advantage of the linear relationship between the aerial image and the source patterns, and the source optimization can be formulated as a quadratic problem. The drawback of the pixel-based method is that the number of pixel variables can be very large, leading to a computationally intensive optimization problem.

Recently, a library-based method is proposed by Yu *et al.* for efficient source optimization with a large mask pattern [22]. This method employs the illumination cross coefficient (ICC) for source optimization for small mask patterns, setting up a library with the optimized source

patterns and optimizing the source pattern within this library for large mask patterns. Yet, the steps in constructing the library still amount to substantial computation. Thus, it is still highly desirable to find a source representation method that can better leverage the representation freedom and optimization efficiency.

In this paper, we develop an SMO algorithm by representing the source patterns using the superposition of weighted Zernike polynomial functions. In optical lithography, these functions are widely used for the phase representation of the wavefront aberration [23, 24], and can be easily incorporated in efficient modeling of the imaging process [25]. The coefficients can be computed through matrix inversion or 2-D integration, making it straightforward to transform between the pixel images and the Zernike coefficients [26]. In addition, we can take advantage of the source pattern characteristics, such as symmetry, to choose the Zernike polynomials. Thus the number of variables in source optimization can be significantly reduced compared with pixel-based algorithms. The relationship between the aerial image and the Zernike coefficients is also linear, and the former can be calculated efficiently.

In the following sections, we first model the forward aerial image formation with partially coherent imaging systems using the Zernike polynomials. We then develop an algorithm for a fast transmission cross coefficient (TCC) calculation with a similar linear relationship, which helps to reduce the computation of mask optimization by introducing the sum of coherent systems (SOCS) [27]. Next, we cast the source optimization as a quadratic problem, which can be efficiently solved through convex optimization tools. We further analyze the source optimization results with different terms of the Zernike polynomial functions. Finally, simulations of a sequential SMO are performed to show the computation efficiency and optimization performance improvement of the proposed algorithm compared with the pixel-based SMO algorithm.

2. Fast aerial image calculation

The optical lithography imaging process is usually modeled as a partially coherent system, which consists of an extended source, a condenser, a mask pattern, a projection lens, and an aerial image on the wafer plane. Generally, the aerial image I on the wafer plane can be expressed, using Abbe's formulation, as [28]

$$I(x,y) = \iint\limits_{-\infty}^{+\infty} J(f,g) \left| \mathfrak{F}\left[O(f',g')H(f'+f,g'+g) \right] \right|^2 \, \mathrm{d}f \, \mathrm{d}g, \tag{1}$$

or with Hopkins' formulation as

$$I(x,y) = \iiiint\limits_{-\infty}^{+\infty} T(f_1,g_1;f_2,g_2)O(f_1,g_1)O^\dagger(f_2,g_2)\mathrm{e}^{-i2\pi\left[(f_1-f_2)x+(g_1-g_2)y\right]} \, \mathrm{d}f_1 \, \mathrm{d}g_1 \, \mathrm{d}f_2 \, \mathrm{d}g_2, \tag{2}$$

where (x,y) are the spatial coordinates, (f,g) are the spatial frequency coordinates, J is the illumination source, O is the mask spectrum, H is the projection pupil, $\mathfrak{F}$ denotes the Fourier transform, $\dagger$ denotes the complex conjugate, and T is the transmission cross coefficient (TCC) defined by

$$T(f_1,g_1;f_2,g_2) = \iint\limits_{-\infty}^{+\infty} J(f,g)H(f+f_1,g+g_1)H^\dagger(f+f_2,g+g_2) \, \mathrm{d}f \, \mathrm{d}g. \tag{3}$$

In this system, $J(f,g)$ is conventionally a circular function with radius σ, also known as the partial coherence factor. To improve the resolution, lithographers have developed off-axis illumination, including annular, dipole and quadrupole sources, and more recently the technology
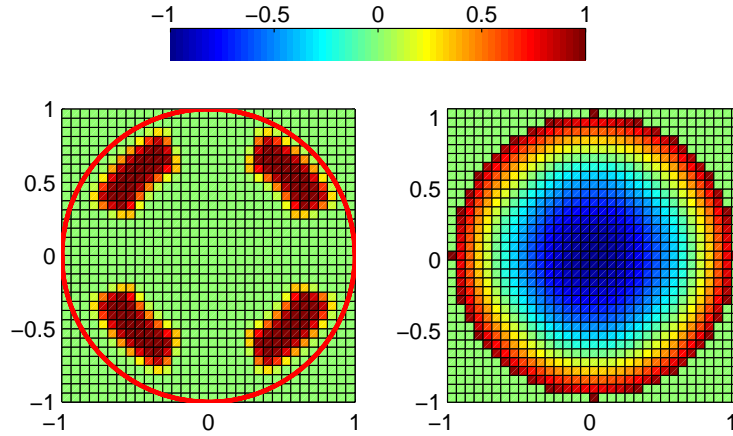
Fig. 2. The sampling method of the source patterns and the Zernike polynomial functions.

has enabled the use of more customized sources with a variety of shapes. These customized sources can be represented by several methods, including sector/track-based, and pixel-based. The former uses a small number of variables to represent the source patterns, but with limited flexibility. The latter permits much more flexible designs, leading to the highest possibility to improve the resolution. However, this representation method requires a large number of pixel variables.

To cope with this, we make use of the Zernike polynomial functions, which are a sequence of orthogonal basis functions [24, 28]. We denote the source patterns as the expansion of P terms of Zernike polynomials, i.e.

$$J(f,g) = \sum_{l=1}^{P} \psi_l Z_l(f,g), \tag{4}$$

where ψ_l is the corresponding Zernike coefficient, and Z_l is the lth Zernike polynomial. For convenience, we can change this to a matrix representation using lexicographic ordering. Equation (4) can then be rewritten as

$$\mathbf{J} = \mathbf{Z}\Psi, \tag{5}$$

where Ψ is the vector of Zernike coefficients $\Psi = \begin{bmatrix} \psi_1 & \psi_2 & \ldots \end{bmatrix}^{\mathsf{T}}$, and $\mathbf{Z}$ is a $N_s^2 \times P$ matrix generated by stacking the vectors of the Zernike polynomials together.

In the pixel-based SMO method, the source patterns are discretized to a square grid, and the intensity of each grid location is represented by a pixel value. In order to better represent the source patterns with the Zernike polynomial functions, we also discretize these basis functions into matrices of the same size as the source patterns, as shown in Fig. 2. For our purposes here, we assume that they are of size $N_s \times N_s$. Moreover, we note that for a partially coherent imaging system, the effective source intensity is limited to a unit disk, and only the pixels within the circle shown in Fig. 2 are of interest. Thus, we only need to consider the values within the unit disk for the Zernike polynomial functions.

The number of terms P can be quite large to represent a free-form source, which would require significant computation as a result. However, realistic source patterns in optical lithography often have some characteristics such as symmetry to reduce the pattern placement shift [9, 29]. This property can reduce the number of polynomials because we can restrict ourselves to those symmetrical to both the horizontal and vertical axes. The first 21 Zernike polynomial functions satisfying this requirement are shown in Fig. 3.
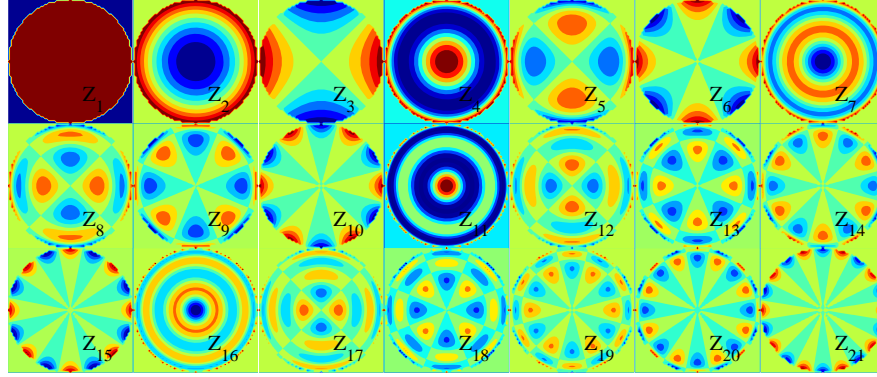
Fig. 3. The first 21 Zernike polynomial functions chosen to represent the source patterns.

Using the Zernike polynomial representation, we substitute Eq. (4) into Abbe's aerial image formulation in Eq. (1). Since the aerial image is linearly related to the source pattern, we separate the Zernike coefficients by changing the position of integration and summation such that

$$I(x,y) = \sum_{l=1}^{P} \psi_l \hat{I}_l(x,y), \qquad (6)$$

where

$$\hat{I}_l(x,y) = \iint_{-\infty}^{+\infty} Z_l(f,g) \left| \mathfrak{F} \left[O(f',g') H(f'+f,g'+g) \right] \right|^2 \, \mathrm{d}f \, \mathrm{d}g. \qquad (7)$$

This equation can be considered as the basis aerial image corresponding to the Zernike polynomial function Z_l. In matrix form, Eq. (6) can be written as

$$\mathbf{I} = \hat{\mathbf{I}}\Psi. \qquad (8)$$

where $\mathbf{I}$ is the vector form of I, and $\hat{\mathbf{I}}$ is a matrix generated in the same way as $\mathbf{Z}$.

In a similar way, we also substitute Eq. (4) in the TCC expression, getting

$$T(f_1,g_1;f_2,g_2) = \sum_{l=1}^{P} \psi_l \hat{T}_l(f_1,g_1;f_2,g_2), \qquad (9)$$

where the basis TCC matrix $\hat{T}_l$ is

$$\hat{T}_l(f_1,g_1;f_2,g_2) = \iint_{-\infty}^{+\infty} Z_l(f,g) H(f+f_1,g+g_1) H^{\dagger}(f+f_2,g+g_2) \, \mathrm{d}f \, \mathrm{d}g. \qquad (10)$$

The equivalent matrix form of Eq. (9) is

$$\mathbf{T} = \hat{\mathbf{T}}\Psi. \qquad (11)$$

Again, the generation of this matrix form of TCC follows the same method as Eqs. (5) and (8).

We would like to point out that the basis TCCs can be pre-computed because they only involve the corresponding basis Zernike polynomial functions and the pupil functions. Then, only the coefficients ψ_p vary when the source pattern changes, and the new TCC can be calculated

(a) Source coefficients

(b) Zernike polynomials (c) Basis TCCs (d) Basis images
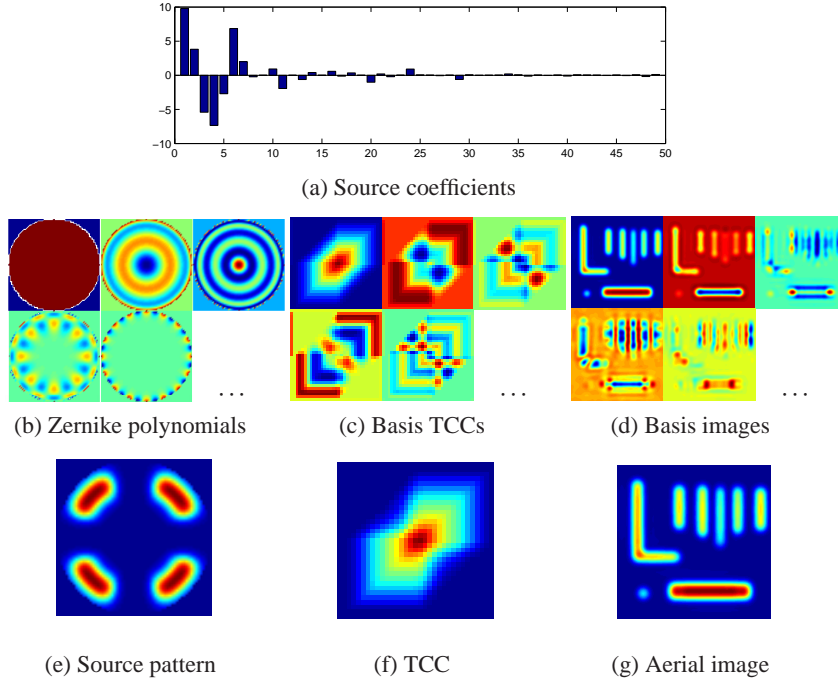

(e) Source pattern (f) TCC (g) Aerial image

Fig. 4. Theory of aerial image simulation with Zernike polynomial-based source representation.

through a linear combination of these bases, which can be very efficient. We depict this in Fig. 4, where Fig. 4(a) shows the Zernike coefficients, Figs. 4(b), 4(c), and 4(d) are the basis Zernike polynomial functions, basis TCCs, and basis images, respectively, and Figs. 4(e), 4(f), and 4(g) are correspondingly the source pattern, the TCC, and the aerial image in the imaging systems.

3. Inverse problem formulation

3.1. Mask optimization

In this section, we formulate the inverse optimization scheme for SMO based on the forward imaging model derived above. First, we consider the mask optimization. In previous SMO algorithms, Abbe's formulation is usually used for aerial image calculation, since the calculation of the TCC involves multiple integrations. Here, we can use the sum of coherent systems (SOCS) theory for fast aerial image calculation since the TCC can be calculated efficiently through Eq. (11) without computationally intensive integrations [27].

In this theory, the TCC matrix is decomposed into kernels through singular value decomposition (SVD), and the singular values descend rapidly. Only the few largest singular values and their corresponding eigenfunctions, which are considered as kernels in the imaging system, are maintained for the aerial image calculation. Thus the amount of computation can be significantly reduced. Let K be the number of singular values used for the computation. After the TCC is computed efficiently through Eq. (9), the decomposition can be expressed as

$$T(f_1, g_1; f_2, g_2) = \sum_{n=1}^{K} \lambda_n \phi_n(f_1, g_1) \phi_n^{\dagger}(f_2, g_2), \tag{12}$$

where λ_n is the nth eigenvalue, and ϕ_n is the corresponding eigenvector. Then the aerial image can be calculated through

$$I(x,y) = \sum_{n=1}^{K} \lambda_n \left| \mathfrak{F}\left[O(f,g)\phi_n(f,g) \right] \right|^2 \tag{13}$$

$$= \sum_{n=1}^{K} \lambda_n |M(x,y) * \Phi_n(x,y)|^2, \tag{14}$$

where M is the mask pattern in the spatial domain, Φ_n is the Fourier transform of the eigenfunction $\phi_n(f,g)$, and $*$ denotes 2-D convolution.

For the cost function, we define the difference between the resist image I_r and the target pattern I_t as a measure of the image fidelity. The resist image is obtained from the aerial image with a sigmoid function modeled as the resist effect

$$I_r(x,y) = \text{sig}\left\{I(x,y)\right\} = \frac{1}{1+\mathrm{e}^{-\alpha[I(x,y)-t_r]}}, \tag{15}$$

where t_r is the threshold in the photoresist effect, and α indicates the steepness of the sigmoid function. Then the image fidelity term $\mathscr{R}_m$ is given by

$$\mathscr{R}_m\left\{M(x,y)\right\} = \sum_{x,y} \left\| I_r(x,y) - I_t(x,y) \right\|_2^2. \tag{16}$$

In addition, to enhance image contrast, we define a penalty term $\mathscr{R}_a$ on the aerial image as

$$\mathscr{R}_a\left\{M(x,y)\right\} = \sum_{x,y} \left\| I(x,y) - 2t_r I_t(x,y) \right\|_2^2. \tag{17}$$

This term can force the aerial image to be 0 while the target is 0, and the aerial image to be $2t_r$ while the target is 1. More explanations about this term can be found in Ref. [5]. The overall cost function $\mathscr{L}_m$ for mask optimization can be represented as

$$\mathscr{L}_m\left\{M(x,y)\right\} = \mathscr{R}_m\left\{M(x,y)\right\} + \tau \mathscr{R}_a\left\{M(x,y)\right\}, \tag{18}$$

where τ is a weight assigned to the image contrast term. Therefore the mask optimization can be formulated as

$$\hat{M}(x,y) = \arg\min_{M} \mathscr{L}_m\left\{M(x,y)\right\}. \tag{19}$$

A conjugate gradient method can be employed to optimize the mask pattern iteratively [5].

3.2. Source optimization

We formulate the source optimization as a quadratic problem as introduced by Yu *et al.* [30]. Note, however, that in our method, the source patterns are fully characterized by the Zernike coefficients, and therefore we only need to optimize them instead of the pixel variables. The cost function consists of two terms, namely the contour awareness term $\mathscr{R}_c$ and the side-lobes compressing term $\mathscr{R}_0$. The former forces the intensity on the contour to be equal to the threshold, as defined by

$$\mathscr{R}_c\{\Psi\} = \|\hat{\mathbf{I}}_c \Psi - \mathbf{t}_r\|_2^2, \tag{20}$$

where $\hat{\mathbf{I}}_c$ is a $N_c \times P$ matrix denoting the aerial images extracted from $\hat{\mathbf{I}}$ by choosing those located on the mask edge position, N_c is the number of points on these position, and $\mathbf{t}_r$ is a

N_c-element vector whose values are all t_r. The latter suppresses the side-lobes by forcing the aerial image around the main features to be small, as given by

$$\mathscr{R}_0\{\Psi\} = \|\hat{\mathbf{I}}_0\Psi - \boldsymbol{\epsilon}\|_2^2, \tag{21}$$

where $\hat{\mathbf{I}}_0$ is a $N_0 \times P$ matrix denoting the aerial images located on a closed curve surrounding the main features of the mask, N_0 is the number of points on the curve, and $\boldsymbol{\epsilon}$ is a N_0 length vector with all its values equal to ε, which is a small positive value. The distance between the curve and the main features is half a pitch for periodic patterns and $0.61\lambda/\mathrm{NA}$ for isolated and semi-isolated patterns, where λ is the wavelength of the source, and NA is the numerical aperture.

Adding these two items together, we get the overall cost function

$$\mathscr{L}_s\{\Psi\} = \mathscr{R}_c\{\Psi\} + \mu\mathscr{R}_0\{\Psi\}, \tag{22}$$

where μ signifies the relative importance of the two terms. This cost function can be written as a quadratic form

$$\begin{aligned}\mathscr{L}_s\{\Psi\} &= (\hat{\mathbf{I}}_c\Psi - \mathbf{t}_r)^{\mathsf{T}}(\hat{\mathbf{I}}_c\Psi - \mathbf{t}_r) + (\hat{\mathbf{I}}_0\Psi - \boldsymbol{\epsilon})^{\mathsf{T}}(\hat{\mathbf{I}}_0\Psi - \boldsymbol{\epsilon}) \\ &= \Psi^{\mathsf{T}}\mathbf{Q}\Psi - \mathbf{b}^{\mathsf{T}}\Psi + c,\end{aligned} \tag{23}$$

where $\mathbf{Q} = \hat{\mathbf{I}}_c^{\mathsf{T}}\hat{\mathbf{I}}_c + \mu\hat{\mathbf{I}}_0^{\mathsf{T}}\hat{\mathbf{I}}_0$, $\mathbf{b} = 2\hat{\mathbf{I}}_c^{\mathsf{T}}\mathbf{t}_r + 2\mu\hat{\mathbf{I}}_0^{\mathsf{T}}\boldsymbol{\epsilon}$, $c = \mathbf{t}_r^{\mathsf{T}}\mathbf{t}_r + \mu\boldsymbol{\epsilon}^{\mathsf{T}}\boldsymbol{\epsilon}$. The sizes of $\mathbf{Q}$ and $\mathbf{b}$ are $P \times P$ and $P \times 1$, respectively, and c is a scalar.

The source intensities in lithography tools are non-negative, real-value functions. On the other hand, the Zernike polynomial functions contain negative values, and their summation is not necessarily positive. Another issue is that to avoid sharp spikes that can damage the lenses, the source intensities are limited to some value $S_{\max}$. As the source patterns with the Zernike representation can be expressed as $\mathbf{Z}\Psi$, these two requirements can be satisfied by setting a linear constraint on the source patterns as $0 \leq \mathbf{Z}\Psi \leq S_{\max}$.

In addition, the dose variation in optical lithography can be characterized by either the total intensity of the illumination source, or the threshold value t_r in the resist model. Here, we fix the threshold value, and limit the total intensity of the illumination source to a certain value $D_{\max}$. The total intensity of the source can be calculated as the summation of all the pixel values of the source patterns, that is, $\mathbf{E}\mathbf{Z}\Psi$, where $\mathbf{E}$ is an N_s^2 row vector whose values are all 1. Thus, this requirement can be expressed as a linear constraint $\mathbf{E}\mathbf{Z}\Psi \leq D_{\max}$ in the optimization process.

Overall, the source optimization can be formulated as

$$\begin{aligned}\underset{\Psi}{\text{minimize}} \quad &\mathscr{L}_s\{\Psi\}, \\ \text{subject to} \quad &0 \leq \mathbf{Z}\Psi \leqslant S_{\max} \\ &\mathbf{E}\mathbf{Z}\Psi \leq D_{\max}.\end{aligned} \tag{24}$$

This is a quadratic problem with linear constraints, and can be conveniently solved by convex optimization tools such as CVX [31].

4. Simulations

4.1. Selection of the Zernike polynomials

In the above derivations, we take advantage of the prior information of the source patterns to choose the symmetric Zernike polynomial functions. There is a tradeoff between the amount of computation and the optimization performance, i.e., fewer terms can lead to faster computation,
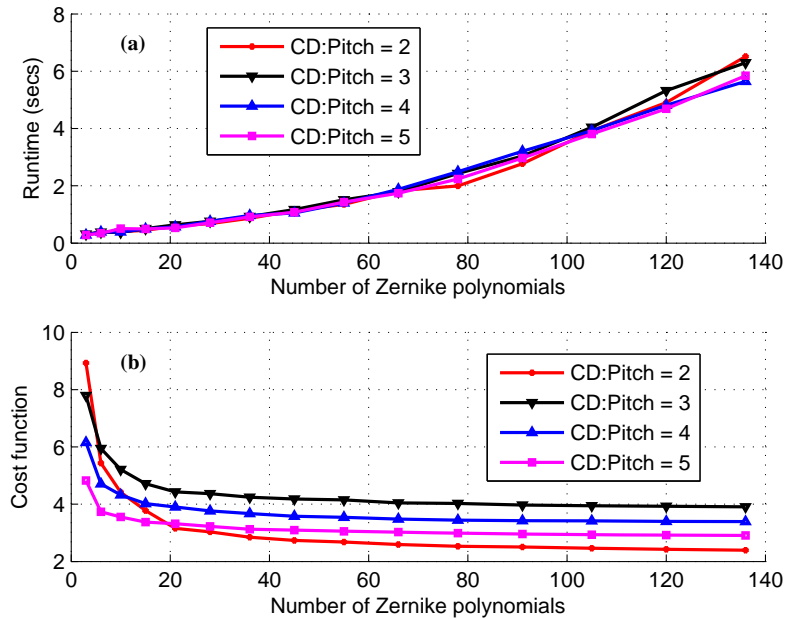
Fig. 5. The tradeoff between optimization performance and computation speed.

at the expense of lower source pattern flexibility. Here, we first perform simulations to evaluate this quantitatively.

We compute the source optimization using different numbers of Zernike polynomial functions on several line array mask patterns with different densities. The line width of the line array patterns, also known as the CD, is 51.5 nm. For different densities, the ratio of CD to pitch ranges from 2 to 5, thus we have four mask patterns in total. It is well known that the Zernike polynomials have radial components and axial components [24]. For the chosen symmetric Zernike polynomials, the number $P = L(L+1)/2$ if the first L orders of radial components are selected. Here, we set L ranges from 2 to 17, and the corresponding P equals to $3, 6, 10, \cdots, 136$. Each Zernike polynomial, and therefore the source pattern, is represented by a 65×65 pixel image. The number of source pixels of interest located within the unit disk is 3785. The source wavelength is 193 nm, and the NA is 1.35. The t_r and α in the sigmoid function to calculate the resist image are 0.3 and 85, respectively. We also set the weight $\mu = 0.1$, $\varepsilon = 0.001$ in the source optimization, and the maximum pixel values $S_{max} = 1$ and $D_{max} = 500$.

Figure 5 plots the source optimization runtime and the optimized cost function value versus the number of Zernike polynomials. Each line in Fig. 5(a) plots the relationship between the runtime and the number of Zernike polynomials, while Fig. 5(b) displays the optimized values of the cost function corresponding to mask patterns at various densities. The runtime increases almost linearly in the shown simulation region, while the optimized value decreases with more Zernike terms. This demonstrates the tradeoff mentioned above between the amount of computation and the optimization performance. It is also noted that when the number of Zernike polynomials increases to a certain number, the optimized cost function reduces slowly. It indicates that the Zernike polynomials of low order contribute to the main features of the optimized source patterns. The Zernike polynomials of high order have little contribution to the reduction of the cost function, and can be neglected in source optimization.

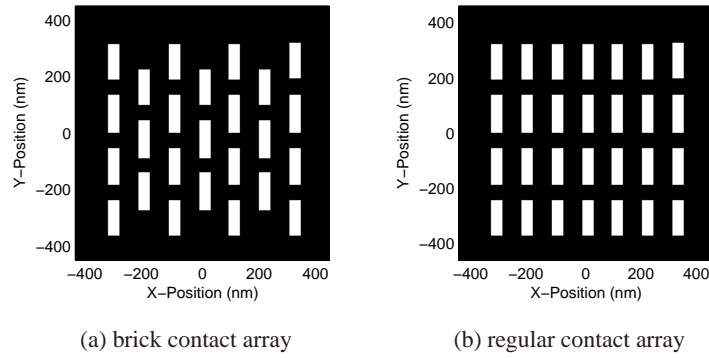(a) brick contact array   (b) regular contact array

Fig. 6. Two test mask patterns for source mask optimization.

To better balance the amount of computation and the optimization performance, we choose $P = 78$ in our following simulations of source mask optimization. This is because the optimization process can approximate the largest solution space when $P = 78$, and the increase of terms can only add computational burden.

4.2. Optimization results

Source optimization alone is insufficient to obtained the required pattern fidelity in computational lithography. Thus, we apply the above technique to a sequential SMO process, and compare the performance and efficiency between the pixel-based algorithm and the Zernike polynomial based algorithm under the same conditions. The sequential SMO process is carried out by first performing source optimization with the target pattern as the initial step, and then performing mask optimization with the optimized source obtained earlier [5]. This process can be repeated several times until convergence. Note that the source representation method here can be applied to other SMO algorithms, such as the simultaneous and hybrid algorithms.

We evaluate the optimization performance by measuring both the pattern fidelity and the robustness to process variations. The pattern fidelity is evaluated by computing the pattern error (PE) defined as the difference between the output pattern and the target pattern, and the edge placement errors (EPE) at critical places. In order to evaluate the robustness to process variations, we estimate the process windows and normalized image log slopes (NILS) at critical places. We also assess the optimization efficiency by calculating the runtime of the optimization process.

Figure 6 shows the two target mask patterns, namely, a brick contact array and a regular contact array, which we use to test the SMO algorithm. Both patterns are of size 201×201 pixels, and each pixel represents 4.47 nm. The size of each contact is $40.2 \, \text{nm} \times 125.1 \, \text{nm}$, and the distances between neighboring contacts are 107.2 nm and 180.9 nm in the horizontal and vertical directions, respectively. The critical places to calculate the EPE are located at the cutlines, which are the central places of each contact region. As stated above, we choose 78 terms of Zernike polynomial functions as the basis functions to represent the source patterns. Each is represented as a size 65×65 pattern. The source pattern is therefore also of this size, hence $N_s = 65$, and the spatial frequencies ranges from -1 to 1 after normalization by NA/λ. Similar to the earlier simulation, we set $S_{\max} = 1$, $D_{\max} = 500$, and $\mu = \tau = 0.1$. Furthermore, in mask optimization, the number of kernels maintained for the aerial image calculation is $K = 10$.

The optimization results of the brick contact array are shown in Fig. 7. The source optimization results using the pixel-based (PB) method and the Zernike polynomial-based (ZPB) method

(a) Optimized PB source

(b) Optimized mask
with source (a)

(c) Output pattern

(d) Optimized ZPB source

(e) Optimized mask
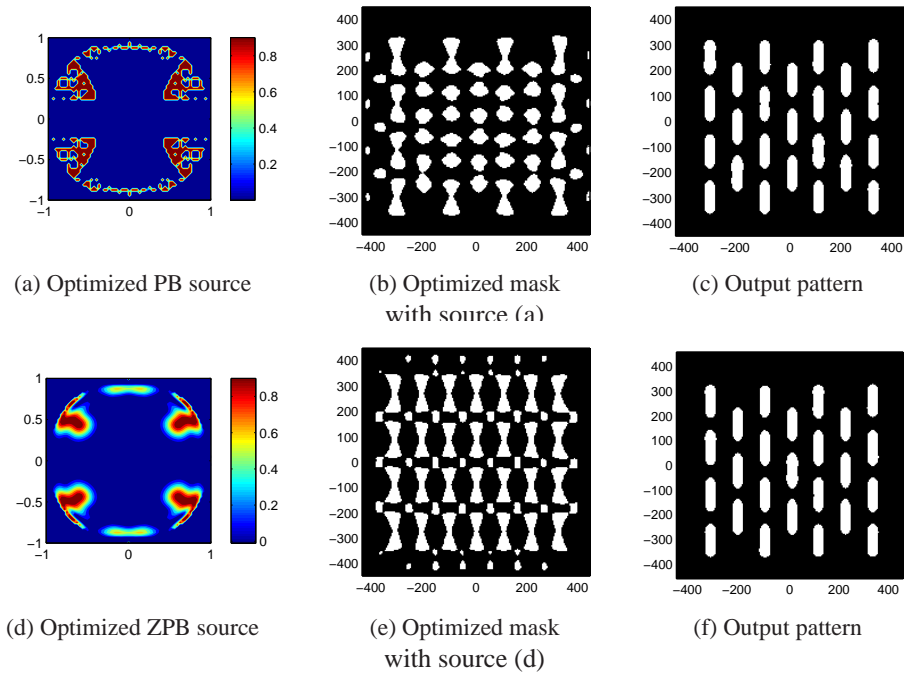with source (d)

(f) Output pattern

Fig. 7. Simulation results of SMO for the brick contact array.

are shown in Figs. 7(a) and 7(d), respectively. As we can see, the locations of the main features of the optimized source patterns are similar, located at the 45 degree quadrupole position. The PB source pattern contains some isolated points and holes, while the ZPB source is smooth and less complicated, which is advantageous because this means the source can be manufactured more easily. The corresponding optimized masks are shown in Figs. 7(b) and 7(e), which are quite similar with one another. Finally, the output patterns on the wafer are given in Figs. 7(c) and 7(f), which are obtained at the best focus plane. Numerically, the PEs of the two cases are 502 and 492, and the average EPE are 2.66 nm and 2.49 nm, respectively. The results for the regular contact array pattern are shown in Fig. 8, where the locations of the main features in both the optimized source pattern and mask patterns from PB and ZPB are similar. The PEs of the two output patterns are 484 and 496, and the average EPE are 1.98 nm and 2.02 nm, respectively. In both cases, the numerical values indicate that the ZPB method can deliver comparable results with the more conventional PB method.

We also depict the average exposure-defocus (E-D) window of the optimization results in Fig. 9 to evaluate the robustness to process variations. The E-D windows are measured at the critical places, which are the same as those in EPE measurement. The curves show the maximum and minimum doses with 10% line width changes and −40 nm to 40 nm focus variations for both the PB and ZPB algorithms. We also measure the depth of focus (DoF) as the largest focus range where the vertical length of an ellipse is tangent to the curve pair for a particular E-D window. It is shown that the optimization results obtained by the Zernike polynomial-based method have similar E-D window sizes compared with the pixel-based algorithm. Quantitatively, the DoF obtained from the PB algorithm is 54 nm for the brick contact array and 68 nm for the regular contact array, while the DoF of the ZPB algorithm is 56 nm and 66 nm, respectively. The NILS are also measured and summarized in Table 1, which also demonstrates similar optimization performance. It indicates the ZPB algorithm can obtain competitive optimization
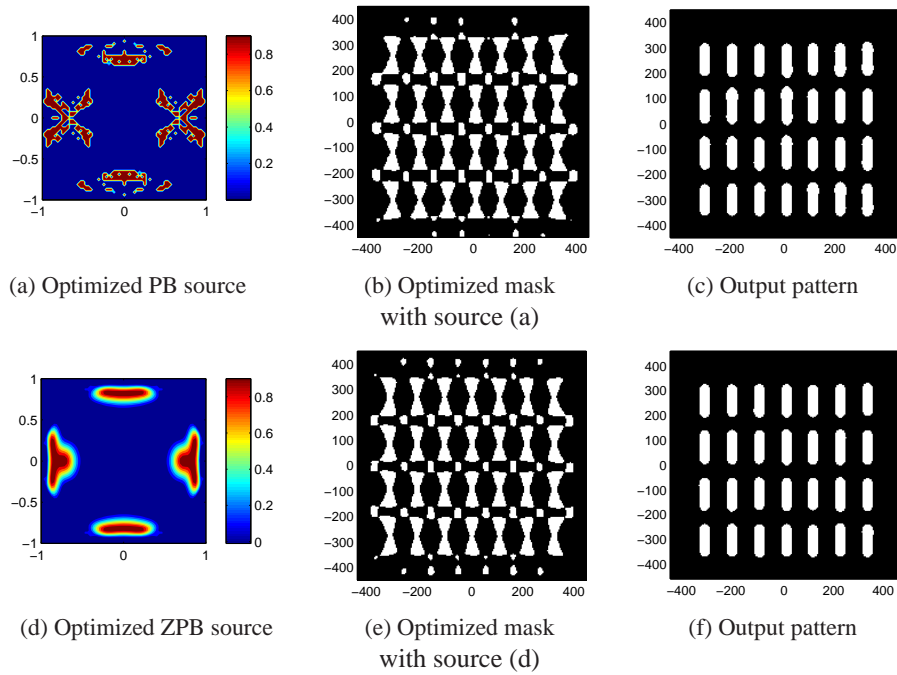
(a) Optimized PB source

(b) Optimized mask
with source (a)

(c) Output pattern

(d) Optimized ZPB source

(e) Optimized mask
with source (d)

(f) Output pattern

Fig. 8. Simulation results of SMO for the regular contact array.

(a) PB, DoF = 54 nm; ZPB, DoF = 56 nm
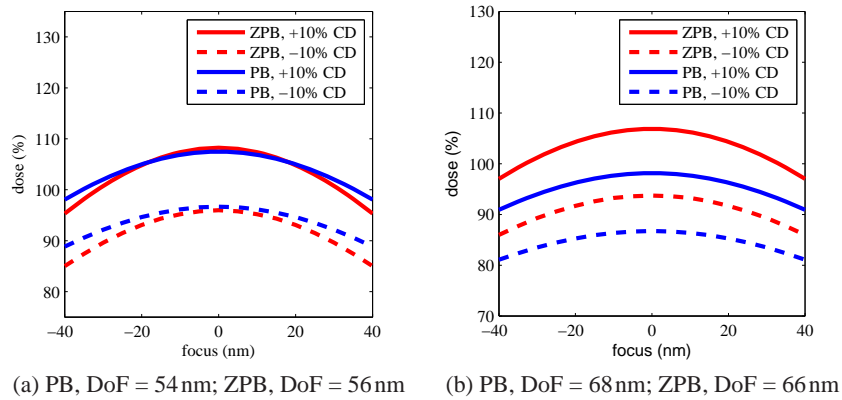

(b) PB, DoF = 68 nm; ZPB, DoF = 66 nm

Fig. 9. Comparison of process window for (a) brick contact array and (b) regular contact array.

results compared with the PB algorithm in terms of robustness to process variations.

After evaluating the optimized image performance of the two algorithms, we now assess the optimization efficiency. Table 1 summarizes the runtime for both the source optimization (SO) and the mask optimization (MO) steps with the two test patterns. With both algorithms, the SO converges to a global solution in about 30 iterations. The total runtime of the ZPB algorithm is about $40 - -50$ times shorter than the PB algorithm. This is attributed to the fact that the number of source variables in the PB algorithm is 3785, including all the pixels inside the unit circle represented by a 65×65 image, which is about 48 times the number of bases used in the

Table 1. Comparison of the optimization performance and efficiency.

Test patterns	Methods	PE (nm)	EPE (nm)	NILS	Runtime (sec)	
					SO	MO
Brick contact array	PB	502	2.66	0.93	110.0	54.9
	ZPB	492	2.49	0.95	2.45	38.6
Regular contact array	PB	484	1.98	1.32	116.4	52.2
	ZPB	496	2.02	1.28	2.34	39.8

ZPB algorithm. In addition, the MO also records a slight speedup with the latter. This is because the TCC can be calculated from the linear equations in Eq. (11) efficiently, while ordinarily it would need multiple integrations.

5. Conclusions

In this paper, we propose an efficient SMO algorithm using the Zernike polynomial functions to represent the source patterns. We demonstrate that the source patterns can be represented with a small number of Zernike polynomials, and the source optimization problem can be formulated as a quadratic problem. We show that this can deliver similar performance to that provided by the pixel-based algorithm in enhancing both the pattern fidelity and robustness; at the same time, the optimization efficiency can be significantly improved due to the smaller number of source variables in source optimization and the use of the linear relationship to calculate the TCC in mask optimization.