

# Pixel-based OPC optimization based on conjugate gradients

Xu Ma<sup>1,\*</sup> and Gonzalo R. Arce<sup>2,3</sup>

<sup>1</sup>Key Laboratory of Photoelectronic Imaging Technology and System of Ministry of Education of China, School of Optoelectronics, Beijing Institute of Technology, Beijing, China

<sup>2</sup>Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware, U.S.A.

<sup>3</sup>arce@ece.udel.edu

\*maxu@bit.edu.cn

**Abstract:** Optical proximity correction (OPC) methods are resolution enhancement techniques (RET) used extensively in the semiconductor industry to improve the resolution and pattern fidelity of optical lithography. In pixel-based OPC (PBOPC), the mask is divided into small pixels, each of which is modified during the optimization process. Two critical issues in PBOPC are the required computational complexity of the optimization process, and the manufacturability of the optimized mask. Most current OPC optimization methods apply the steepest descent (SD) algorithm to improve image fidelity augmented by regularization penalties to reduce the complexity of the mask. Although simple to implement, the SD algorithm converges slowly. The existing regularization penalties, however, fall short in meeting the mask rule check (MRC) requirements often used in semiconductor manufacturing. This paper focuses on developing OPC optimization algorithms based on the conjugate gradient (CG) method which exhibits much faster convergence than the SD algorithm. The imaging formation process is represented by the Fourier series expansion model which approximates the partially coherent system as a sum of coherent systems. In order to obtain more desirable manufacturability properties of the mask pattern, a MRC penalty is proposed to enlarge the linear size of the sub-resolution assist features (SRAFs), as well as the distances between the SRAFs and the main body of the mask. Finally, a projection method is developed to further reduce the complexity of the optimized mask pattern.

© 2011 Optical Society of America

**OCIS codes:** (110.5220) Photolithography; (110.4980) Partial coherence in imaging; (100.3190) Inverse problems.

---

## References and links

1. A. K. Wong, *Resolution Enhancement Techniques* (SPIE Press, 2001), Vol. 1.
2. X. Ma and G. R. Arce, *Computational Lithography*, Wiley Series in Pure and Applied Optics, 1st ed. (John Wiley and Sons, 2010).
3. S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication*, 2nd ed. (Publishing House of Electronics Industry, 2003).
4. F. Schellenberg, "Resolution enhancement technology: the past, the present, and extensions for the future," Proc. SPIE **5377**, 1–20 (2004).
5. F. Schellenberg, *Resolution Enhancement Techniques in Optical Lithography* (SPIE Press, 2004).
6. L. Liebmann, S. Mansfield, A. Wong, M. Lavin, W. Leipold, and T. Dunham, "TCAD development for lithography resolution enhancement," IBM J. Res. Dev. **45**, 651–665 (2001).

7. U. Okoroanyanwu, *Chemistry and Lithography*, 1st ed. (in review, SPIE Press, 2010).
8. S. Sherif, B. Saleh, and R. Leone, "Binary image synthesis using mixed integer programming," *IEEE Trans. Image Proc.* **4**(9), 1252–1257 (1995).
9. Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," *IEEE Trans. Semicond. Manuf.* **5**(2), 138–152 (1992).
10. A. Erdmann, R. Farkas, T. Fühner, B. Tollkühn, and G. Kokai, "Towards automatic mask and source optimization for optical lithography," *Proc. SPIE* **5377**, 646–657 (2004).
11. Y. Granik, "Solving inverse problems of optical microlithography," *Proc. SPIE* **5754**, 506–526 (2004).
12. Y. Granik, "Fast pixel-based mask optimization for inverse lithography," *J. Microlithogr., Microfabr., Microsyst.* **5**, 043002 (2006).
13. A. Poonawala and P. Milanfar, "Fast and low-complexity mask design in optical microlithography—an inverse imaging problem," *IEEE Trans. Image Process.* **16**(3), 774–788 (2007).
14. X. Ma and G. R. Arce, "Binary mask optimization for inverse lithography with partially coherent illumination," *J. Opt. Soc. Am. A* **25**(12), 2960–2970 (2008).
15. X. Ma and G. R. Arce, "Binary mask optimization for forward lithography based on boundary layer model in coherent systems," *J. Opt. Soc. Am. A* **26**(7), 1687–1695 (2009).
16. X. Ma and G. R. Arce, "Pixel-based simultaneous source and mask optimization for resolution enhancement in optical lithography," *Opt. Express* **17**(7), 5783–5793 (2009).
17. J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, 2006).
18. B. E. A. Saleh and M. Rabbani, "Simulation of partially coherent imagery in the space and frequency domains and by modal expansion," *Appl. Opt.* **21**(15), 2770–2777 (1982).
19. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," *Opt. Express* **15**, 15066–15079 (2007).
20. X. Ma and G. R. Arce, "PSM design for inverse lithography with partially coherent illumination," *Opt. Express* **16**(24), 20126–20141 (2008).
21. K. Kato, K. Nishizawa, and T. Inoue, "Advanced mask rule check (MRC) tool," *Proc. SPIE* **6283**, 62830O (2006).
22. K. Kato, Y. Taniguchi, K. Nishizawa, and M. Endo, "Mask rule check using priority information of mask patterns," *Proc. SPIE* **6730**, 67304F (2007).
23. B. Salik, J. Rosen, and A. Yariv, "Average coherent approximation for partially coherent optical systems," *J. Opt. Soc. Am. A* **13**(10), 2086–2090 (1996).
24. M. Born and E. Wolfe, *Principles of Optics* (Cambridge University Press, 1999).
25. R. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics* (John Wiley and Sons, 1995).
26. M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Natl. Bur. Stand.* **49**, 409–436 (1952).
27. R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *Comput. J.* **7**, 149–154 (1964).
28. H. P. Crowder and P. Wolfe, "Linear convergence of the conjugate gradient method," *IBM J. Res. Dev.* **16**, 431–433 (1972).
29. A. Cohen, "Rate of convergence of several conjugate gradient algorithms," *SIAM J. Numer. Anal.* **9**, 248–259 (1972).
30. K. Ritter, "On the rate of superlinear convergence of a class of variable metric methods," *Numerische Mathematik* **35**, 293–313 (1980).
31. C. Vogel, *Computational Methods for Inverse Problems* (SIAM Press, 2002).
32. M. P. McLoughlin and G. R. Arce, "Deterministic properties of the recursive separable median filter," *IEEE Trans. Acoust., Speech, Signal Process.* **35**, 98–106 (1987).
33. S. Kalluri and G. R. Arce, "Fast algorithms for weighted myriad computation by fixed-point search," *IEEE Trans. Signal Process* **48**(1), 159–171 (2000).
34. K. Barner and G. R. Arce, "Permutation filters - a class of nonlinear filters based on set permutations," *IEEE Trans. Signal Process* **42**, 782–798 (1994).

---

## 1. Introduction

Due to the resolution limits of optical lithography systems, the electronics industry has relied on resolution enhancement techniques (RET) to compensate and minimize mask distortions as they are projected onto semiconductor wafers [1, 2]. Resolution in optical lithography obeys the Rayleigh resolution limit  $R = k \frac{\lambda}{NA}$ , where  $\lambda$  is the wavelength,  $NA$  is the numerical aperture, and  $k$  is the process constant which can be minimized through RET methods [3–6]. In optical proximity correction (OPC), mask amplitude patterns are modified by the addition of the sub-resolution assistant features (SRAFs) that can pre-compensate for imaging distortions [1, 2]. In PBOPC, the mask is divided into small pixels, each of which is modified during the optimiza-

tion process. The optimized PBOPC is then fabricated by two approaches: the raster scanning e-beam method and the vector scanning e-beam method [7]. In raster scanning, each pixel on the mask pattern is exposed by the electron-beam moving back and forth row by row. Due to the low throughput of the raster scanning method, the vector scanning approach is now preferred for mask fabrication. In the vector scanning method, the turned off electron-beam first moves on a vector path and directly reaches the region to be exposed. Then, the electron-beam is turned on to expose the nether feature on the PBOPC. After that, the electron-beam is turned off and moves again until the entire pattern is completely exposed [7]. Tiny features are either unmanufacturable or disadvantageous for the mask writing accuracy. Since each pixel on the PBOPC may be flipped independently, the two critical issues in PBOPC are the computational complexity in the design and the manufacturability of the optimized mask.

Several OPC optimization algorithms have been proposed in the literature. These range from heuristic and empirically-based design rules to optimization-based inverse algorithms. Sherif, et al. derived an iterative approach to generate binary masks in incoherent diffraction-limited imaging systems [8]. Liu and Zakhor developed a binary and phase shifting mask (PSM) design strategy based on the branch and bound algorithm and simulated annealing [9]. Erdmann proposed the optimization of the mask and illumination parameters with a genetic algorithm [10]. Granik described and compared solutions of inverse mask problems [11, 12]. All of the methods mentioned above, however, are not based on gradient type optimization and thus the searching process for a suitable solution is either computationally expensive or not efficient.

Poonawala and Milanfar recently introduced a PBOPC optimization framework for coherent illumination sources relying on the SD algorithm [13]. Ma and Arce generalized this algorithm to partially coherent illumination (PCI) systems [14]. Subsequently, Ma and Arce proposed a three-dimensional OPC optimization algorithm, and a simultaneous source and mask optimization algorithm [15, 16]. In each iteration of the gradient-based OPC optimization, the aerial image needs to be recalculated due to the updated pixel values on the mask [14]. The aerial image calculation is, in particular, the most time-consuming step in the OPC optimization process. Thus, the overall runtime of the gradient-based OPC optimization is approximately proportional to the iteration number. It has been proven that the SD solution linearly converges to the optimal solution with respect to the iteration number, while the Newton's method has a quadratic convergence speed, and the nonlinear CG method converges in linear or superlinear speed [17]. In practice, the CG method has much faster convergence than SD. Since the SD algorithm has very slow convergence characteristics, as compared to other efficient optimization algorithms, the OPC SD optimization framework is not efficient, and other approaches such as the CG method and Newton's method are therefore needed [17]. Furthermore, unlike the Newton-class of optimization algorithms, the CG method only needs the first derivative of the cost function [17]. This property makes the CG method as simple to implement as the SD algorithm. In our future work, we plan to study and develop an in-depth analysis of the convergence properties of CG algorithms, and compose a separate manuscript devoted to this topic exclusively. To this end, this paper develops OPC optimization based on the CG method to seek optimal OPC mask patterns. Compared to the SD algorithm, the CG method often leads to greater image fidelity using much less iterations, thus effectively reducing the overall runtime. A partially coherent illumination system model is assumed the partially coherent imaging system is represented by the Fourier series expansion model [14, 18].

A second contribution of the paper is to introduce a mask rule check (MRC) penalty to obtain more desirable manufacturability characteristics for the optimized mask. Currently, a set of regularization and post-processing techniques have been proposed to reduce the mask complexity [13, 19, 20]. However, they are inadequate to generate MRC-favorable optimized masks. MRC is used currently in the mask design process to constrain the mask geometries

[21, 22]. Usually, the minimum linear size of the SRAFs, and the distance from the SRAFs to the main body of the mask are forced to be above some prescribed thresholds. However, most of the existing MRC methods are applied to the post-optimized mask pattern, and serve as post-processing steps, which lead to sub-optimal solutions. In contrast, the proposed approach in this paper embeds the MRC penalty in the cost function resulting in MRC-favorable optimal masks.

A third contribution of this paper is the development of a projection method to further removes the small SRAFs. The projection method projects the pixel values on the mask between the subspace of the optimal solution and the subspace supported by the low frequency components in two-dimensional discrete cosine transform (2D-DCT) domain. Simulation shows that the resulting optimized OPC has much lower complexity.

The remainder of the paper is organized as follows. The partially coherent imaging model is discussed in Section 2. The OPC optimization process based on the CG method under partially coherent system is developed in Section 3. The MRC penalty is introduced in Section 4. The projection method is proposed in Section 5. Conclusions are provided in Section 6.

## 2. Partially Coherent Imaging Model

### 2.1. Partially Coherent Illumination System

Partially coherent illumination has been shown to improve the theoretical resolution limit in lithography. Thus, practical lithography systems often operate under partially coherent illumination due to non-zero width sources and off-axis illumination from spatially extended sources. A schematic of an optical lithography system with partially coherent illumination is illustrated

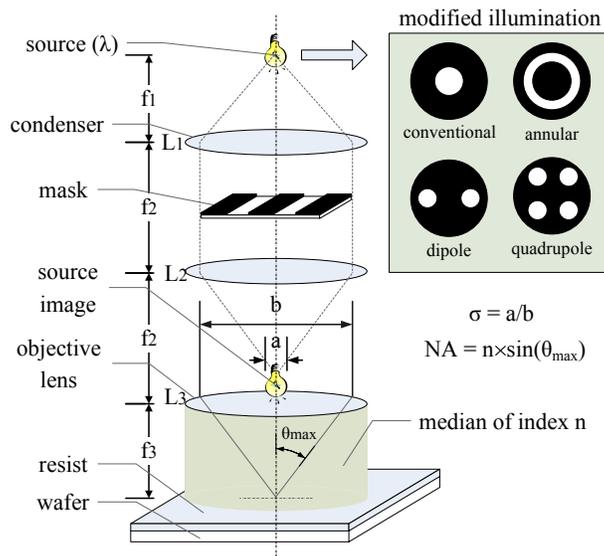


Fig. 1. Optical lithography system with partially coherent illuminations.

in Fig. 1. The light source with a wavelength of  $\lambda$  is placed at the focal plane of the first condenser, illuminating the mask. Common illuminations sources include dipole, quadrupole and annular shapes, all introducing partial coherence. The image of the photomask is formed by the projection optics onto the wafer [1]. The partial coherence factor  $\sigma = \frac{a}{b}$  is defined as the ratio between the size of the source image and the pupil. Radiation of partially coherent light

has been shown to be described as an expansion of coherent modes added incoherently in the image plane [18].

According to the Hopkins imaging model, the light intensity distribution exposed on the wafer in the PCI system is bilinear and described by [2]

$$I(\mathbf{r}) = \iint M(\mathbf{r}_1)M(\mathbf{r}_2)\gamma(\mathbf{r}_1 - \mathbf{r}_2)h^*(\mathbf{r} - \mathbf{r}_1)h(\mathbf{r} - \mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2, \quad (1)$$

where  $\mathbf{r} = (x, y)$  is the coordinate on the image plane located on the wafer.  $\mathbf{r}_1 = (x_1, y_1)$  and  $\mathbf{r}_2 = (x_2, y_2)$  are the coordinates on the object plane located on the mask.  $M(\mathbf{r})$  is the mask pattern,  $\gamma(\mathbf{r}_1 - \mathbf{r}_2)$  is the complex degree of coherence, and  $h(\mathbf{r})$  represents the amplitude impulse response of the optical system. The complex degree of coherence  $\gamma(\mathbf{r}_1 - \mathbf{r}_2)$  is generally a complex number, whose magnitude represents the extent of optical interaction between two spatial locations  $\mathbf{r}_1 = (x_1, y_1)$  and  $\mathbf{r}_2 = (x_2, y_2)$  of the light source [1]. The complex degree of coherence in the spatial domain is the inverse 2-D Fourier transform of the illumination shape. The Hopkins imaging model assumes that the light diffraction from the masks is invariant with respect to the direction of the incident light. The diffraction spectrum for oblique incidence is obtained by a simple shift of the diffraction spectrum for vertical incident light. This assumption is less accurate for modern high NA projection imaging systems using strong off-axis illuminations. In such cases, PBOPC optimization with more rigorous imaging models needs to be explored as this case is left for future research work. In general, Eqn. (1) is tedious to compute, both analytically and numerically [23]. In the following, the Fourier series expansion model is applied to formulate the imaging process of a partially coherent system in the discrete domain, and accelerate the calculation of Eqn. (1) [18].

## 2.2. Fourier Series Expansion Model

Assume the mask is constrained in the square area  $A$  defined by  $x, y \in [-\frac{D}{2}, \frac{D}{2}]$ . Thus, for the computations involved in Eqn. (1), the only values of  $\gamma(\mathbf{r})$  needed are those inside the square area  $A_\gamma$  defined by  $x, y \in [-D, D]$ . Applying the 2-D Fourier series expansion,  $\gamma(\mathbf{r})$  can be rewritten as

$$\gamma(\mathbf{r}) = \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} \exp(j\omega_0 \mathbf{m} \cdot \mathbf{r}), \quad (2)$$

and

$$\Gamma_{\mathbf{m}} = \frac{1}{D^2} \int_{A_\gamma} \gamma(\mathbf{r}) \exp(j\omega_0 \mathbf{m} \cdot \mathbf{r}) d\mathbf{r}, \quad (3)$$

where  $\omega_0 = \pi/D$ ,  $\mathbf{m} = (m_x, m_y)$ ,  $m_x$  and  $m_y$  are integers, and  $\cdot$  is the inner-product operation. Substituting Eqn. (2) into Eqn. (1), the light intensity on the wafer is given by

$$I(\mathbf{r}) = \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |M(\mathbf{r}) \otimes h^{\mathbf{m}}(\mathbf{r})|^2, \quad (4)$$

where

$$h^{\mathbf{m}}(\mathbf{r}) = h(\mathbf{r}) \exp(j\omega_0 \mathbf{m} \cdot \mathbf{r}). \quad (5)$$

It is observed from Eqn. (4) that PCI leads to the superposition of coherent systems. Since the Fourier series expansion model is based on direct discretization of the Hopkins imaging model, they have the same accuracy. For the annular illumination, the complex degree of coherence is

$$\gamma(\mathbf{r}) = \frac{J_1(2\pi r/2D_{cu})}{2\pi r/2D_{cu}} - \frac{D_{cu}^2}{D_{cl}^2} \frac{J_1(2\pi r/2D_{cl})}{2\pi r/2D_{cl}}, \quad (6)$$

where  $r = \sqrt{x^2 + y^2}$ . The corresponding Fourier series coefficients are

$$\Gamma_{\mathbf{m}} = \begin{cases} \frac{4D_{cu}^2 D_{cl}^2}{\pi D^2 (D_{cl}^2 - D_{cu}^2)} & \text{for } D/2D_{cl} \leq |\mathbf{m}| \leq D/2D_{cu} \\ 0 & \text{elsewhere} \end{cases}, \quad (7)$$

where  $D_{cl}$  and  $D_{cu}$  are the coherent lengths of the inner and outer circles respectively.  $\sigma_{inner} = \frac{\lambda}{2D_{cl}NA}$  and  $\sigma_{outer} = \frac{\lambda}{2D_{cu}NA}$  are the corresponding inner and outer partial coherence factors. The convolution kernel  $h(\mathbf{r})$  is defined as the Fourier transform of the circular lens aperture with cutoff frequency  $NA/\lambda$  [24, 25]; therefore,

$$h(\mathbf{r}) = \frac{J_1(2\pi r NA/\lambda)}{2\pi r NA/\lambda}. \quad (8)$$

### 3. Inverse Lithography Optimization

Let  $M(x, y)$  be the input binary mask to an optical lithography system  $T\{\cdot\}$ , with a partially coherent illumination. The PCI optical system is approximated by a Hopkins imaging model. The effect of the photoresist is modeled by a hard threshold operation. The output pattern is denoted as  $Z(x, y) = T\{M(x, y)\}$ . Given a  $N \times N$  desired output pattern  $\tilde{Z}(x, y)$ , the goal of OPC design is to find the optimized  $M(x, y)$  called  $\hat{M}(x, y)$  such that the distance

$$F = d(Z(x, y), \tilde{Z}(x, y)) = d(T\{M(x, y)\}, \tilde{Z}(x, y)) \quad (9)$$

is minimized, where  $d(\cdot, \cdot)$  is the square of the  $L_2$  norm. The OPC optimization problem can thus be formulated as the search of  $\hat{M}(x, y)$  over the  $N \times N$  real space  $\mathfrak{R}^{N \times N}$  such that

$$\hat{M}(x, y) = \arg \min_{M(x, y) \in \mathfrak{R}^{N \times N}} d(T\{M(x, y)\}, \tilde{Z}(x, y)). \quad (10)$$

The forward imaging process is illustrated in Fig. 2, where the Fourier series expansion model is used in the image formation stage, and a hard threshold function is used to approximate the photoresist effect. The output pattern of the optical system is binary. Further, since the

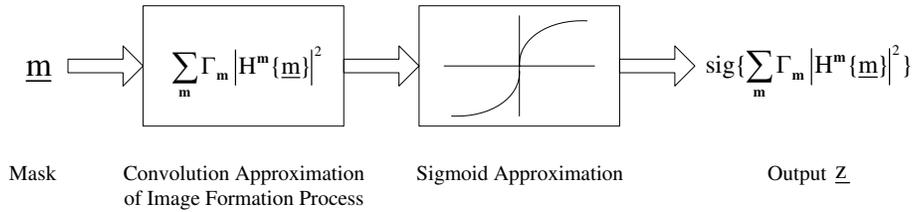


Fig. 2. Approximated forward process model.

derivative of the sigmoid function exists, it is used to approximate the hard threshold function. The sigmoid function is

$$\text{sig}(x) = \frac{1}{1 + \exp[-a(x - t_r)]}, \quad (11)$$

where  $t_r$  is the process threshold, and  $a$  dictates the steepness of the sigmoid function.

Following the definitions above, the following notations are used:

1) The  $M_{N \times N}$  matrix represents the mask pattern with a  $N^2 \times 1$  equivalent raster scanned vector representation, denoted as  $\underline{m}$ .

2) A convolution matrix  $H^{\mathbf{m}}$  is a  $N^2 \times N^2$  matrix with an equivalent two-dimensional filter  $h^{\mathbf{m}}$ .

3) The desired  $N \times N$  binary output pattern is denoted as  $\tilde{Z}$ . It is the desired light distribution sought on the wafer. Its vector representation is denoted as  $\tilde{z}$ .

4) The output of the sigmoid function is the  $N \times N$  image denoted as:

$$Z = \text{sig}\left\{\sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |H^{\mathbf{m}}\{\underline{m}\}|^2\right\}. \quad (12)$$

The equivalent vector is denoted as  $\underline{z}$ .

5) The hard threshold version of  $Z$  is the binary output pattern denoted as  $Z_b$ . Its equivalent vector is denoted as  $\underline{z}_b$ , with all entries constrained to 0 or 1.

6) The optimized  $N \times N$  gray mask denoted as  $\hat{M}$  minimizes the distance between  $Z$  and  $\tilde{Z}$ , ie,

$$\hat{M} = \arg \min_M d(\text{sig}\{\sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |H^{\mathbf{m}}\{\underline{m}\}|^2\}, \tilde{Z}). \quad (13)$$

Its equivalent vector is denoted as  $\hat{m} \in [0, 1]$ .

7) The binary optimized mask  $\hat{M}_b$  is the quantization of  $\hat{M}$ . Its equivalent vector is denoted as  $\hat{m}_b$ , with all entries constrained to 0 or 1.

Given the gray level pattern  $\underline{z} = \text{sig}\{\sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |H^{\mathbf{m}}\{\underline{m}\}|^2\}$ , the  $i$ th entry in this vector can be represented as

$$z_i = \frac{1}{1 + \exp[-a \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |\sum_{j=1}^{N^2} h_{ij}^{\mathbf{m}} m_j|^2 + at_r]}, \quad i = 1, \dots, N^2, \quad (14)$$

where  $h_{ij}^{\mathbf{m}}$  is the  $i, j$ th entry of the filter. In the optimization process,  $\hat{m}$  is searched to minimize the square of the  $L_2$  norm of the difference between  $\underline{z}$  and  $\tilde{z}$ . Therefore,

$$\hat{m} = \arg \min_{\hat{m}} \{F(\underline{m})\}, \quad (15)$$

where the cost function  $F(\cdot)$  is defined as:

$$F(\underline{m}) = \|\tilde{z} - \underline{z}\|_2^2 = \sum_{i=1}^{N^2} (\tilde{z}_i - z_i)^2, \quad (16)$$

where  $\|\cdot\|_2$  is the  $L_2$  norm and  $z_i$  is represented in Eqn. (14). In order to reduce the above bound-constrained optimization problem to an unconstrained optimization problem, we adopt the parametric transformation [13]. Let

$$\underline{m}_j = \frac{1 + \cos(\underline{\theta}_j)}{2}, \quad j = 1, \dots, N^2, \quad (17)$$

where  $\underline{\theta}_j \in (-\infty, \infty)$  and  $\underline{m}_j \in [0, 1]$ . Defining the vector  $\underline{\theta} = [\underline{\theta}_1, \dots, \underline{\theta}_{N^2}]^T$ , the optimization problem is formulated as

$$\begin{aligned} \hat{\underline{\theta}} &= \arg \min_{\underline{\theta}} \{F(\underline{\theta})\} \\ &= \arg \min_{\underline{\theta}} \left\{ \sum_{i=1}^{N^2} \left( \tilde{z}_i - \frac{1}{1 + \exp[-a \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |\sum_{j=1}^{N^2} h_{ij}^{\mathbf{m}} \frac{1 + \cos \theta_j}{2}|^2 + at_r]} \right)^2 \right\}. \quad (18) \end{aligned}$$

The CG method is used to optimize the above problem. CG method was originally designed to solve linear systems with positive definite coefficient matrices [26]. Fletcher and Reeves generalized the linear CG method to nonlinear problems [27]. The CG method generates a set of gradient estimate vectors  $\{P_0, P_1, \dots, P_i\}$  with some properties, along which the cost function is successively minimized. Each vector  $P_i$  is updated from  $P_{i-1}$ , and is the linear combination of the past gradients. Thus, each new minimizing direction  $P_i$  inherits the useful information from its ancestor. In the past, a variety of CG methods have been proposed, such as Fletcher-Reeves method, Polak-Ribière method, and so on [17]. In this paper, we adopt the Fletcher-Reeves method. The comparison among these CG algorithms and the method to select the proper algorithm are out of the scope of this paper.

Let  $\Theta \in \Re^{N \times N}$  is the equivalent matrix of  $\underline{\theta}$ . The gradients  $\nabla F(\Theta)$  can be calculated as follows [14]:

$$\begin{aligned} \nabla F(\Theta) = & a \times \sin\Theta \odot \left\{ \sum_{\mathbf{m}} \Gamma_{\mathbf{m}}(h^{\mathbf{m}})^{\circ} \otimes [(\tilde{Z} - Z) \odot Z \odot (\mathbf{1}_{N \times N} - Z) \odot ((h^{\mathbf{m}})^* \otimes M)] \right\} \\ & + a \times \sin\Theta \odot \left\{ \sum_{\mathbf{m}} \Gamma_{\mathbf{m}}(h^{\mathbf{m}})^{\circ*} \otimes [(\tilde{Z} - Z) \odot Z \odot (\mathbf{1}_{N \times N} - Z) \odot (h^{\mathbf{m}} \otimes M)] \right\}, \end{aligned} \quad (19)$$

where  $\nabla F(\Theta) \in \Re^{N \times N}$ ,  $\odot$  is the element-by-element multiplication operator,  $(h^{\mathbf{m}})^{\circ}$  rotates  $h^{\mathbf{m}}$  by  $180^{\circ}$  degrees in both row and column directions,  $*$  is the conjugate operation, and  $\mathbf{1}_{N \times N} \in \Re^{N \times N}$  has all entries equal to 1. Assuming  $\Theta^k$  is the  $k^{\text{th}}$  iteration result, then the CG method is summarized as following:

**Iteration 0:** Given the desired binary mask pattern  $\tilde{Z}$ , calculate

$$\Theta^0(i, j) = \begin{cases} \frac{4\pi}{5} & \text{for } \tilde{Z}(i, j) = 0 \\ \frac{\pi}{5} & \text{for } \tilde{Z}(i, j) = 1 \end{cases}, \quad i, j = 1, 2, \dots, N, \quad (20)$$

According to Eqn. (19), assigning  $\Theta^0(i, j) = \pi$  or  $\Theta^0(i, j) = 0$  would degrade the gradient to 0 and therefore the iteration would be terminated. Thus, we select the angle values  $\frac{4\pi}{5}$  and  $\frac{\pi}{5}$ , which are close to the ideal values. In addition, we initialize

$$P^0 = -\nabla F(\Theta^0), \quad (21)$$

where  $\Theta^0, P^0 \in \Re^{N \times N}$ .

**Iteration k:**

**Step 1:** Line search the step length  $s^k \in \Re$ , such that

$$s^k = \underset{s}{\arg \min} F(\Theta^k + sP^k). \quad (22)$$

**Step 2:** Update  $\Theta^{k+1} \in \Re^{N \times N}$

$$\Theta^{k+1} = \Theta^k + s^k P^k. \quad (23)$$

**Step 3:** Calculate  $\beta^k \in \Re$

$$\beta^k = \frac{\|\nabla F(\Theta^{k+1})\|_2^2}{\|\nabla F(\Theta^k)\|_2^2}. \quad (24)$$

**Step 4:** Update  $P^{k+1} \in \Re^{N \times N}$

$$P^{k+1} = -\nabla F(\Theta^{k+1}) + \beta^k P^k. \quad (25)$$

**Stop:**

Until the termination condition is reached.

The iterative optimization above, in general, leads to gray mask with pixel values between 0 and 1. Therefore a post-processing step is needed to obtain the binary optimized mask:

$$\hat{m}_{bi} = U(\hat{m}_i - t_m), i = 1, \dots, N^2, \quad (26)$$

where  $t_m$  is a global threshold. We define the pattern error  $E$  as the square of the  $L_2$  norm of the distance between the desired output image  $\tilde{Z}$  and the actual binary output pattern  $Z_b$ :

$$E = \sum_{i=1}^{N^2} \|\tilde{z}_i - z_{bi}\|_2^2 = \sum_{i=1}^{N^2} \left\| \tilde{z}_i - U\left(\sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |H^{\mathbf{m}} m_{bi}|^2 - t_r\right) \right\|_2^2. \quad (27)$$

In general, lower pattern errors result in lower critical dimension (CD) errors. There are some other metrics to evaluate the performance of the optical lithography, such as the process window and so on. However this topic is out of the scope of this paper, and will be studied in the future work. In the proposed algorithm, when the pattern error is reduced to a tolerable level, or the iteration number reaches the prescribed upper limit, the CG iteration is terminated. In the following simulations, the tolerable error level is set to 0. That means we would like to seek a printed image on the wafer that is the same as the desired pattern. Therefore, the CG algorithm is terminated when the printed image is the same as the desired pattern, or the iteration number reaches the prescribed upper limit. The CG method above is tailored for the quadratic problem [17]. However, the inverse lithography (ILT) optimization formulated in Eqn. (18) is a non-quadratic problem, and the cost function possesses numerous local minima. Thus, the CG method is easily trapped in these local minima. For the SD method used in prior work, the step length is a constant. Thus, the SD method also may get stuck in a local minimum, but a different one. Our extensive simulations show that setting the step length as a constant makes the CG method have much faster convergence. It also leads to lower pattern errors than the results obtained by the SD algorithm. Intuitively, the constant step length helps the CG method skip the local minimum. Thus, in the following simulations, we replace the **Step 1** above with a constant step length. In practice, the proposed CG method often leads to linear or superlinear convergence speed [17, 28–30].

In the following, we address the advantages of the CG method, by comparing it to the performance of SD method. In this simulation, annular illumination is applied with the inner and outer partial coherence factors  $\sigma_{inner} = 0.3$  and  $\sigma_{outer} = 0.4$ . For both methods, the convolution kernel is shown in Eqn. (8) with  $NA = 1.25$  and  $\lambda = 193nm$ . We assume  $h(\mathbf{r})$  vanishes outside the area  $A_h$  defined by  $x, y \in [-56.25nm, 56.25nm]$ . In the sigmoid function, we assign the parameters  $a = 25$  and  $t_r = 0.19$ . The global threshold is  $t_m = 0.5$ , the pixel size is  $5.625nm \times 5.625nm$ . Generally, smaller step lengths lead to stable but slower convergence, while larger step lengths lead to faster convergence or divergence. Based on this guideline, we set the step length equal to 1, which is moderate for the underlying optimization problem.

ILT optimization is an ill-posed problem, where numerous inputs may result in the same output. Therefore, regularization approaches are usually applied to bias the solution space to sample solutions with some favorable properties [31]. Regularization is formulated as follows:

$$\hat{m} = \arg \min_{\underline{m}} \{F(\underline{m}) + \gamma R(\underline{m})\}, \quad (28)$$

where  $F(\underline{m})$  is the data-fidelity term and  $R(\underline{m})$  is the regularization term which is used to reduce the solution space and constrain the optimized results.  $\gamma$  is the user-defined parameter to reveal the weight of the regularization. In the above simulation, the quadratic penalty [13] and

the wavelet penalty [19] are applied to reduce the pattern error introduced by post-processing described in Eqn. (26), and to reduce the mask complexity, respectively. The formulae of the quadratic penalty and the wavelet penalty are summarized in Appendix A. Thus, the cost function of Eqn. (16) is adjusted as

$$J(\underline{m}) = F(\underline{m}) + \gamma_Q R_Q(\underline{m}) + \gamma_W R_W(\underline{m}). \quad (29)$$

However, making  $\gamma_Q$  and  $\gamma_W$  too large can result in divergence of the optimization process. In the following simulations, the regularization weights of the quadratic penalty and wavelet penalty are  $\gamma_Q = 0.01$  and  $\gamma_W = 0.025$ , respectively.

Figure 3 shows the performance comparison between the masks optimized with the SD and CG methods, respectively. Figure 3(a) is the desired pattern  $\tilde{Z}$  with dimension of  $1035nm \times 1035nm$ , and Fig. 3(d) is its output pattern on the wafer with pattern error of 5052. Figure 3(b) is the optimized OPC pattern from the SD method, and Fig. 3(e) is its output pattern with pattern error of 1368. Figure 3(c) is the optimized OPC pattern from the CG method, and Fig. 3(f) is its output pattern with pattern error of 1288. Compared to the SD method, the CG method reduces the pattern error by 6%

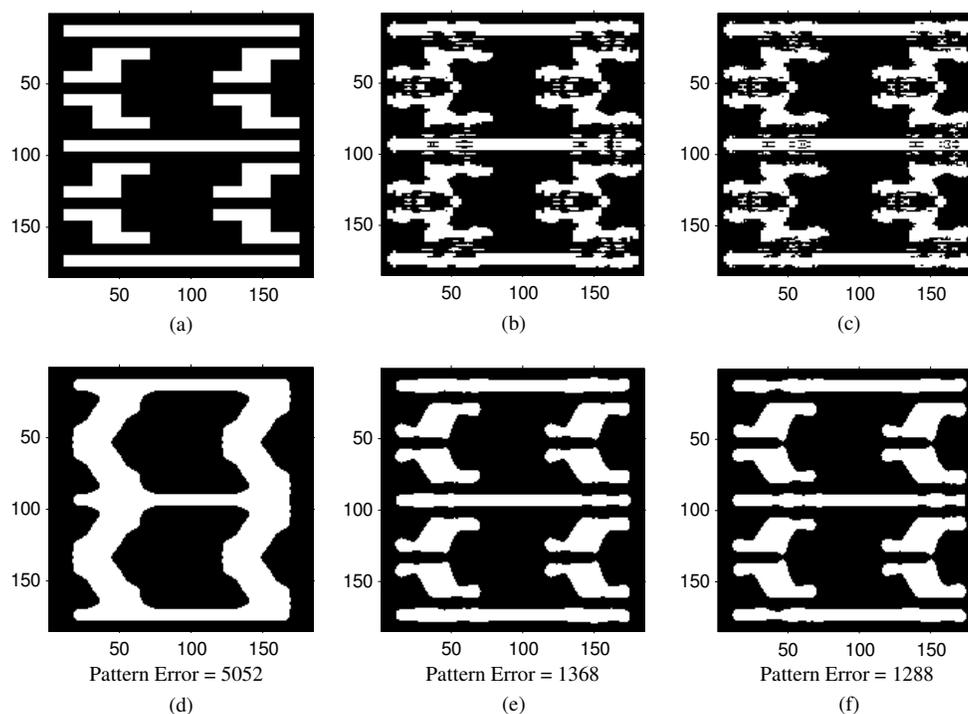


Fig. 3. Performance comparison between the SD method and the CG method; (a) desired pattern  $\tilde{Z}$ ; (b) optimized OPC pattern from the SD method; (c) optimized OPC pattern from the CG method; (d) the output pattern when (a) is used as input; (e) the output pattern when (b) is used as input; (f) the output pattern when (c) is used as input.

Figure 4 illustrates the convergence comparison, where  $x$ -axis is the iteration number and  $y$ -axis is the pattern error. In Fig. 4, the convergence of SD and CG methods are shown in black solid line and blue dashed line, respectively. The SD method converges to pattern error of 1368 after 150 iterations using 996 seconds, while the CG method converges to a pattern error

Table 1. Comparison of performance and convergence

|                | Pattern error | Iteration number | Runtime (s) |
|----------------|---------------|------------------|-------------|
| SD             | 1368          | 150              | 996         |
| CG             | 1288          | 33               | 220         |
| CG + MRC       | 1492          | 33               | 222         |
| Project method | 1880          | 54               | 352         |

of 1288 after 33 iterations using 220 seconds. The performance and convergence comparison between the two approaches are summarized in the 2<sup>nd</sup> to 3<sup>rd</sup> rows of Table 1.

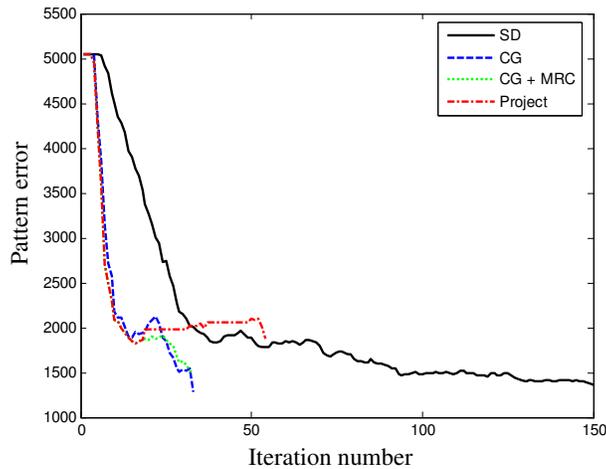


Fig. 4. Convergence comparison of the SD method (black solid line), the CG method (blue dashed line), the CG method with MRC penalty (green dotted line), and the projection method (red dot-dashed line).

#### 4. MRC Penalty

Although the wavelet penalty has been applied in the above simulation, numerous small opening and blocking areas are still contained in Fig. 3(b) and 3(c). In addition, the SRAFs are too close to the main body of the mask pattern. These characteristics make the mask fabrication more difficult. Figure 5 illustrates the influence of the wavelet penalty weight on the optimized OPCs and their output patterns. Figures 5(a), 5(b) and 5(c) show the optimized OPCs with  $\gamma_w = 0.025$ , 0.035 and 0.045, respectively. Figures 5(d), 5(e) and 5(f) show the output patterns corresponding to their above OPC patterns. It is shown that increasing the wavelet penalty weight  $\gamma_w$  cannot enlarge the dimensions of the SRAFs or the distances between the SRAFs and the main body of the mask. In addition, larger wavelet penalty weight will increase the pattern error. Particularly, Fig. 5(f) shows that when  $\gamma_w = 0.045$ , some bridges appear in the output pattern. Therefore, the wavelet penalty falls short to generate MRC-favorable optimized masks.

In order to make the solution have more desirable manufacturability properties, a MRC penalty is proposed in this section. MRC is used currently in the mask design process to constrain the mask geometries. Usually, the minimum linear size of the SRAFs, and the distance

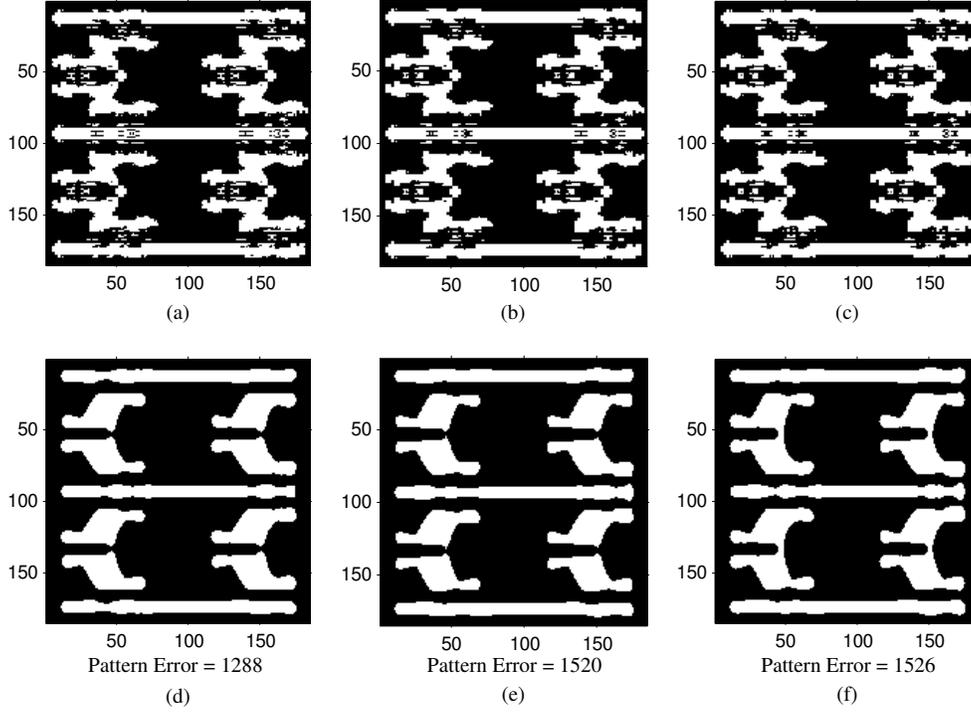


Fig. 5. Influence of the wavelet penalty weight on the optimized OPCs and their output patterns; (a) optimized OPC pattern with  $\gamma_W = 0.025$ ; (b) optimized OPC pattern with  $\gamma_W = 0.035$ ; (c) optimized OPC pattern with  $\gamma_W = 0.045$ ; (d) the output pattern when (a) is used as input; (e) the output pattern when (b) is used as input; (f) the output pattern when (c) is used as input.

from the SRAFs to the main body of the mask are forced to be above some prescribed thresholds. However, most of the existing MRC methods are applied to the post-optimized mask pattern, and serve as post-processing steps, which lead to sub-optimal solutions. In contrast, the proposed MRC penalty in this paper is embedded in the optimization cost function. The MRC penalty term is formulated as

$$R_M(M) = -\mathbf{1}_{N \times 1}^T [(M - \mathbf{t}_{mN \times N}) \odot (g \otimes M)] \mathbf{1}_{N \times 1}, \quad (30)$$

where  $\mathbf{t}_{mN \times N} \in \mathfrak{R}^{N \times N}$  has all entries equal to  $t_m$ ,  $g$  is referred to as the MRC filter,  $g \otimes M$  is a weighted sum of each pixel value and its neighbors, and  $M - \mathbf{t}_{mN \times N}$  judges whether the pixel value is above  $t_m$  and carved on the binary mask or not.  $t_m$  is a constant typically defined as 0.5. The meaning of the MRC penalty is explained as follows. Consider the value of a pixel located at  $(x_0, y_0)$  that is larger than  $t_m$ , i.e.,  $p(x_0, y_0) > t_m$ . Then,  $(M - \mathbf{t}_{mN \times N})(x_0, y_0) > 0$ . If we want to obtain a low penalty value, we need to keep  $(g \otimes M)(x_0, y_0)$  as large as possible, which means that we want all pixel values around  $p(x_0, y_0)$  to be larger than  $t_m$ . The similar analysis applies for the case of  $p(x_0, y_0) < t_m$ . In other words, it is preferable to have the value of  $p(x, y) - t_m$  around the location of  $(x_0, y_0)$  have the same sign of  $p(x_0, y_0) - t_m$ , thus both of the mask opening and blocking areas are concentrated and enlarged. In our simulations,  $t_m = 0.5$ , and  $g = \mathbf{1}_{3 \times 3}$ , which means that we equally sum each pixel value with its eight neighboring pixel values. Through this way, the linear sizes of the SRAFs, and the distances from the SRAFs to

the main body of the mask will be simultaneously increased. Eqn. (30) can be reformulated as

$$R_M(\underline{m}) = \sum_{k=1}^{N^2} (t_m - \underline{m}_k) \left( \sum_{j=1}^{N^2} g_{k,j} \underline{m}_j \right). \quad (31)$$

As derived in Appendix B, the gradient of  $R_M(\Theta)$  is

$$\nabla R_M(\Theta) = (4.5 - 2g \otimes M) \odot (-0.5 \sin \Theta), \quad (32)$$

thus, the cost function in Eqn. (29) is adjusted as

$$J(\underline{m}) = F(\underline{m}) + \gamma_Q R_Q(\underline{m}) + \gamma_W R_W(\underline{m}) + \gamma_M R_M(\underline{m}), \quad (33)$$

Figure 6(a) shows the resulting OPC pattern by optimizing Eqn. (33), where the desired pattern is the same as Fig. 3(a) and  $\gamma_M = 0.005$ . Figure 6(b) shows the output pattern corresponding to Fig. 6(a). The pattern error, iteration number and the runtime are summarized in the 4<sup>th</sup> row of Table 1. The convergence of this simulation is shown by the green dotted line in Fig. 4. It is observed that the MRC penalty effectively enlarges the linear sizes of SRAFs and the distances from the SRAFs to the main body of the mask. As a tradeoff, the pattern error is increased by 21%.

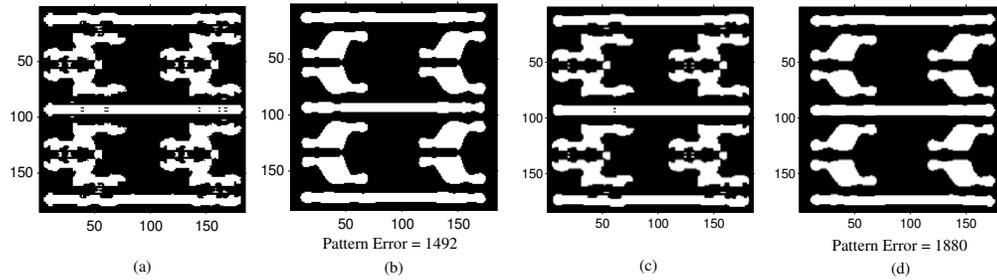


Fig. 6. Performance comparison between the the CG method with MRC penalty and the projection method; (a) optimized OPC pattern from the CG method with MRC penalty; (b) the output pattern when (a) is used as input; (c) optimized OPC pattern from the projection method; (d) the output pattern when (c) is used as input.

## 5. Projection Method

The MRC penalty proposed in Section 4 is an effective way to seek the MRC-favorable optimal solution of the ILT problem. However, it is observed from Fig. 6(a) that some small SRAFs still remain in the OPC pattern. If we extend the dimension of the MRC filter  $g$  in Eqn. (30), the pattern error will be dramatically increased, while the number of the small SRAFs cannot be further reduced. This phenomenon is illustrated in Fig. 7. Figures 7(a), 7(b) and 7(c) show the optimized OPCs with  $g = \mathbf{1}_{3 \times 3}$ ,  $\mathbf{1}_{5 \times 5}$  and  $\mathbf{1}_{7 \times 7}$ , respectively. Figures 7(d), 7(e) and 7(f) show the output patterns corresponding to their above OPC patterns. It is shown that increasing the MRC filter dimension leads to more details on the optimized OPC. In addition, larger MRC filter dimension will increase the pattern error. One approach to ameliorate the pattern distortions as the window size of the MRC filter is increased is to replace the linear filter by a median-type edge-preserving filter [32–34]. The cost function representations of such nonlinear filters, however, are not differentiable and thus are not easily integrated into the optimization.

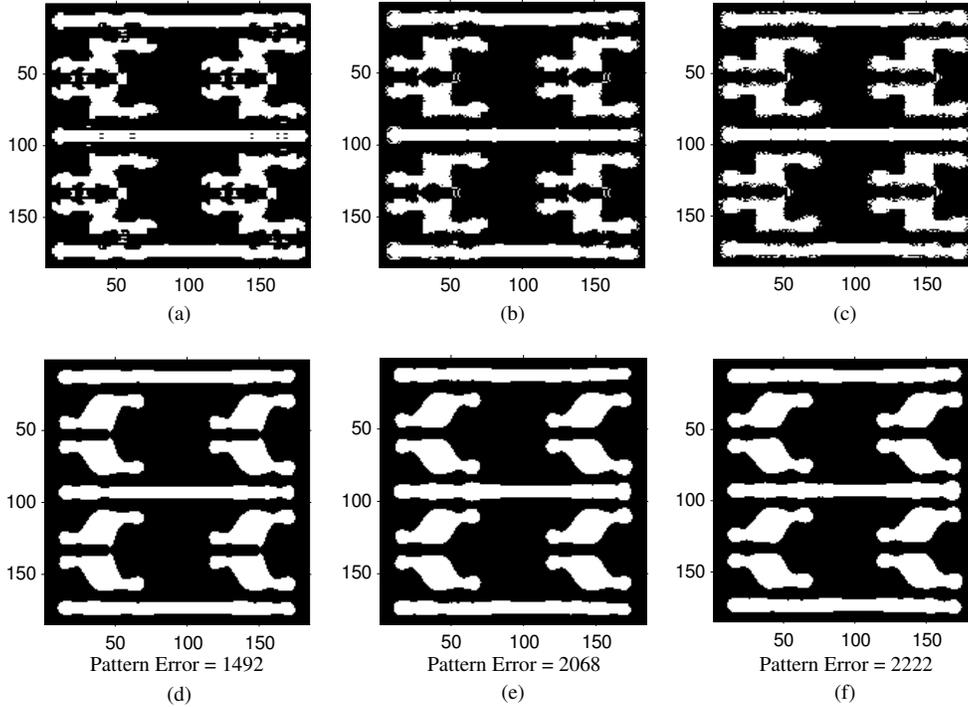


Fig. 7. Influence of the MRC filter dimension on the optimized OPCs and their output patterns; (a) optimized OPC pattern with  $g = \mathbf{1}_{3 \times 3}$ ; (b) optimized OPC pattern with  $g = \mathbf{1}_{5 \times 5}$ ; (c) optimized OPC pattern with  $g = \mathbf{1}_{7 \times 7}$ ; (d) the output pattern when (a) is used as input; (e) the output pattern when (b) is used as input; (f) the output pattern when (c) is used as input.

In order to further improve the manufacturability of the optimized OPC pattern, a projection method is proposed hereafter. Optical lithography system has a low-pass filter property that cuts off the high frequency components of the mask pattern, and constructs the print image on the wafer by the low frequency components. It has been shown that only low frequency components of the mask in the 2D-DCT domain significantly influence the image formation of the optical lithography system [20]. Based on this principle, the proposed projection method is depicted as Fig. 8.

In Fig. 8, the input is the initial pattern  $\Theta_0$  as described in Eqn. (20). In stage I, the CG method with the MRC penalty is applied to obtain a raw optimized gray-level mask pattern  $\hat{M}$ . If  $\hat{M}$  satisfies the prescribed termination conditions, we threshold  $\hat{M}$  by  $t_m$  to obtain the final optimized binary mask  $\hat{M}_b$ . Otherwise, we move to stage II. In stage II, we cut off the high frequency components of  $\hat{M}$  in the 2D-DCT domain, resulting in  $\hat{M}'$ . Subsequently, in stage III we threshold  $\hat{M}'$  to obtain  $\hat{M}'_b$ . In addition, we update  $\Theta$  according to  $\hat{M}'_b$  using method described in Eqn. (20). The updated  $\Theta$  is used as the input of the stage I to continue the loop. Let  $S_F$  represent the subspace of the optimal solutions of Eqn. (10).  $S_{DCT}$  represents the subspace supported by the low frequency components in the 2D-DCT domain. The proposed method iteratively projects the pixel values on the mask between  $S_F$  and  $S_{DCT}$ , and drives the final solution into the intersection between the two subspaces.

Figure 6(c) shows the resulting OPC pattern from the projection method, and Fig. 6(d) shows the corresponding output pattern. The pattern error, iteration number and the runtime are sum-

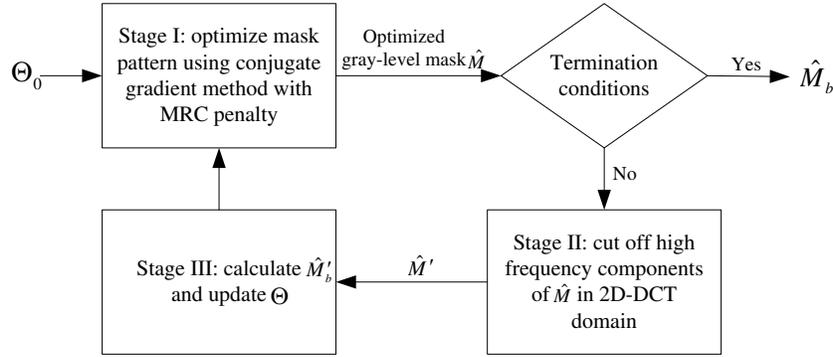


Fig. 8. Scheme of the projection method.

marized in the 5<sup>th</sup> row of Table 1. In stage I, the CG method is iterated for 18 iterations in each projection loop. In stage II, we maintain 5995 low frequency components. The projection loop is terminated when the pattern error is reduced to a tolerable level, or the overall iteration number reaches the prescribed upper limit. The convergence of this simulation is shown by the red dot-dashed line in Fig. 4. It is observed that the projection method effectively improves the manufacturability of the mask. On the other hand, the pattern error is unavoidably increased by 20.6% compared to the CG method with MRC penalty.

## 6. Conclusion

This paper proposes a set of approaches for inverse lithography based on the CG method. The Fourier series expansion model is used to formulate the optimization problem. Compared to the SD algorithm, the CG method results in lower pattern error and faster convergence speed. Subsequently, the MRC penalty and the projection method are proposed to further reduce the mask complexity. As a tradeoff, the pattern errors are increased caused by the absence of numerous SRAFs. Simulations illustrate that our approaches are effective and practical.

### A. Appendix A

The quadratic penalty term is  $R_Q(\underline{m}) = 4\mathbf{m}^T(\mathbf{1} - \underline{m})$ . For each pixel value, the corresponding penalty is the quadratic function  $r(\underline{m}_i) = 1 - (2\underline{m}_i - 1)^2$ ,  $i = 1, \dots, N^2$ . The gradient of  $R_q(\underline{m})$  is  $\nabla R_Q(\Theta) = -\frac{1}{2}(-8M + 4) \cdot \sin(\Theta)$ .

The wavelet penalty term is

$$R_W(\underline{m}) = h_{11}^2 + h_{12}^2 \cdots + h_{\left(\frac{N}{2}\right)\left(\frac{N}{2}\right)}^2 + v_{11}^2 + v_{12}^2 \cdots + v_{\left(\frac{N}{2}\right)\left(\frac{N}{2}\right)}^2 + d_{11}^2 + d_{12}^2 \cdots + d_{\left(\frac{N}{2}\right)\left(\frac{N}{2}\right)}^2, \quad (34)$$

where

$$h_{ij} = m_{(2(i-1)+1)(2(j-1)+1)} - m_{(2(i-1)+1)(2(j-1)+2)} + m_{(2(i-1)+2)(2(j-1)+1)} - m_{(2(i-1)+2)(2(j-1)+2)}, \quad (35)$$

$$v_{ij} = m_{(2(i-1)+1)(2(j-1)+1)} + m_{(2(i-1)+1)(2(j-1)+2)} - m_{(2(i-1)+2)(2(j-1)+1)} - m_{(2(i-1)+2)(2(j-1)+2)}, \quad (36)$$

$$d_{ij} = m_{(2(i-1)+1)(2(j-1)+1)} - m_{(2(i-1)+1)(2(j-1)+2)} - m_{(2(i-1)+2)(2(j-1)+1)} + m_{(2(i-1)+2)(2(j-1)+2)}, \quad (37)$$

for  $i, j = 1, \dots, \frac{N}{2}$ . The gradient of the wavelet penalty is given as

$$\frac{\partial R_W}{\partial \theta_{(2(i-1)+p)(2(j-1)+q)}} = -\frac{1}{2} \sin \theta_{(2(i-1)+p)(2(j-1)+q)} \times (3\underline{m}_{(2(i-1)+p)(2(j-1)+q)} - \underline{m}_{(2(i-1)+p_1)(2(j-1)+q)} - \underline{m}_{(2(i-1)+p)(2(j-1)+q_1)} - \underline{m}_{(2(i-1)+p_1)(2(j-1)+q_1)}), \quad (38)$$

where  $i, j = 1, \dots, \frac{N}{2}$ ,  $p, q = 1$  or  $2$ ,  $p_1 = (p+1) \bmod 2$  and  $q_1 = (q+1) \bmod 2$ .

## B. Appendix B

The derivation of Eqn. (32) is as following:

$$\begin{aligned} \frac{\partial R_M(\underline{m})}{\partial \underline{m}_k} &= \partial \sum_{p=1}^{N^2} [(t_m - \underline{m}_p) (\sum_{j=1}^{N^2} g_{pj} \underline{m}_j)] / \partial \underline{m}_k \\ &= \partial [(t_m - \underline{m}_k) (\sum_{j=1}^{N^2} g_{kj} \underline{m}_j)] / \partial \underline{m}_k + \partial \sum_{p=1, p \neq k}^{N^2} [(t_m - \underline{m}_p) (\sum_{j=1}^{N^2} g_{pj} \underline{m}_j)] / \partial \underline{m}_k \\ &= \partial [(t_m - \underline{m}_k) (\sum_{j=1, j \neq k}^{N^2} g_{kj} \underline{m}_j)] / \partial \underline{m}_k + \partial [(t_m - \underline{m}_k) g_{kk} \underline{m}_k] / \partial \underline{m}_k \\ &+ \partial \sum_{p=1, p \neq k}^{N^2} [(t_m - \underline{m}_p) g_{pk} \underline{m}_k] / \partial \underline{m}_k. \end{aligned}$$

Since  $g = \mathbf{1}_{3 \times 3}$ ,  $g_{kk} = 1$  and  $g_{pk} = g_{kp}$ . Therefore,

$$\begin{aligned} \frac{\partial R_M(\underline{m})}{\partial \underline{m}_k} &= -(\sum_{j=1, j \neq k}^{N^2} g_{kj} \underline{m}_j) + \partial (t_m \underline{m}_k - \underline{m}_k^2) / \partial \underline{m}_k \\ &+ \sum_{p=1, p \neq k}^{N^2} g_{pk} (t_m - \underline{m}_p) \\ &= -(\sum_{j=1, j \neq k}^{N^2} g_{kj} \underline{m}_j) + t_m - 2\underline{m}_k - (\sum_{p=1, p \neq k}^{N^2} g_{kp} \underline{m}_p) + t_m \sum_{p=1, p \neq k}^{N^2} h_{pk} \\ &= -2(\sum_{j=1}^{N^2} g_{kj} \underline{m}_j) + 2g_{kk} \underline{m}_k + t_m - 2\underline{m}_k + t_m \sum_{p=1}^{N^2} h_{pk} - t_m h_{kk} \\ &= -2(\sum_{j=1}^{N^2} g_{kj} \underline{m}_j) + t_m \sum_{p=1}^{N^2} h_{pk}. \end{aligned}$$

Since  $t_m = 0.5$  and  $\sum_{p=1}^{N^2} h_{pk} = 9$ ,  $\nabla R_M(M) = (4.5 - 2g \otimes M)$ . Therefore,

$$\begin{aligned} \nabla R_M(\Theta) &= \nabla R_M(M) \odot \nabla M(\Theta) \\ &= (4.5 - 2g \otimes M) \odot (-0.5 \sin \Theta). \end{aligned} \quad (39)$$