

PSM design for inverse lithography with partially coherent illumination

Xu Ma^a and Gonzalo R. Arce^b

Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, 19716, U.S.A.

maxu@udel.edu and arce@ece.udel.edu

Abstract: Phase-shifting masks (PSM) are resolution enhancement techniques (RET) used extensively in the semiconductor industry to improve the resolution and pattern fidelity of optical lithography. Recently, a set of gradient-based PSM optimization methods have been developed to solve for the inverse lithography problem under coherent illumination. Most practical lithography systems, however, use partially coherent illumination due to non-zero width and off-axis light sources, which introduce partial coherence factors that must be accounted for in the optimization of PSMs. This paper thus focuses on developing a framework for gradient-based PSM optimization methods which account for the inherent nonlinearities of partially coherent illumination. In particular, the singular value decomposition (SVD) is used to expand the partially coherent imaging equation by eigenfunctions into a sum of coherent systems (SOCS). The first order coherent approximation corresponding to the largest eigenvalue is used in the PSM optimization. In order to influence the solution patterns to have more desirable manufacturability properties and higher fidelity, a post-processing of the mask pattern based on the 2D discrete cosine transformation (DCT) is introduced. Furthermore, a photoresist tone reversing technique is exploited in the design of PSMs to project extremely sparse patterns.

© 2008 Optical Society of America

OCIS codes: (110.5220) Photolithography; (110.4980) Partial coherence in imaging; (050.5080) Phase shift; (100.3190) Inverse problems.

References and links

1. A. K. Wong, *Resolution enhancement techniques*, **1** (SPIE Press, 2001).
2. S. A. Campbell, *The science and engineering of microelectronic fabrication*, 2nd ed. (Publishing House of Electronics Industry, Beijing, China, 2003).
3. F. Schellenberg, "Resolution enhancement technology: The past, the present, and extensions for the future, Optical Microlithography," in *Proc. SPIE*, **5377**, 1–20 (2004).
4. F. Schellenberg, *Resolution enhancement techniques in optical lithography* (SPIE Press, 2004).
5. 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).
6. A. Poonawala and P. Milanfar, "Fast and low-complexity mask design in optical microlithography - An inverse imaging problem," *IEEE Trans. Image Process.* **16**, 774–788 (2007).
7. M. D. Levenson, N. S. Viswanathan, and R. A. Simpson, "Improving resolution in photolithography with a phase-shifting mask," *IEEE Trans. Electron Devices* **ED-29**, 1828–1836 (1982).
8. Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," *IEEE Trans. Semicond. Manuf.* **5**, 138–152 (1992).
9. Y. C. Pati and T. Kailath, "Phase-shifting masks for microlithography: Automated design and mask requirements," *J. Opt. Soc. Am. A* **11**, 2438–2452 (1994).

10. A. Poonawala and P. Milanfar, "OPC and PSM design using inverse lithography: A non-linear optimization approach," in *Proc. SPIE*, **6154**, 1159–1172 (San Jose, CA, 2006).
11. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," in *Proc. SPIE* (San Jose, CA, 2007).
12. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," *Opt. Express* **15**, 15,066–15,079 (2007).
13. B. Salik, J. Rosen, and A. Yariv, "Average coherent approximation for partially coherent optical systems," *J. Opt. Soc. Am. A* **13** (1996).
14. E. Wolf, "New spectral representation of random sources and of the partially coherent fields that they generate," *Opt. Commun.* **38** (1981).
15. P. S. Davids and S. B. Bollepalli, "Generalized inverse problem for partially coherent projection lithography," in *Proc. SPIE* (San Jose, CA, 2008).
16. X. Ma and G. R. Arce, "Binary mask optimization for inverse lithography with partially coherent illumination," in *Proc. SPIE* (Taiwan, 2008).
17. X. Ma and G. R. Arce, "Binary mask optimization for inverse lithography with partially coherent illumination," *J. Opt. Soc. Am. A* **25** (2008).
18. N. Cobb, "Fast optical and process proximity correction algorithms for integrated circuit manufacturing," Ph.D. thesis, University of California at Berkeley (1998).
19. J. Zhang, W. Xiong, M. Tsai, Y. Wang, and Z. Yu, "Efficient mask design for inverse lithography technology based on 2D discrete cosine transformation (DCT)," in *Simulation of Semiconductor Processes and Devices*, **12** (2007).
20. 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** (1982).
21. M. Born and E. Wolfe, *Principles of optics* (Cambridge University Press, United Kingdom, 1999).
22. R. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics* (John Wiley and Sons, New York, 1995).
23. N. Cobb and A. Zakhor, "Fast sparse aerial image calculation for OPC," in *BACUS Symposium on Photomask Technology, Proc. SPIE*, **2440**, 313–327 (1995).
24. C. Vogel, *Computational methods for inverse problems* (SIAM Press, 2002).

1. Introduction

Due to the resolution limits of optical lithography, the electronics industry has relied on resolution enhancement techniques (RET) to compensate and minimize mask distortions as they are projected onto semiconductor wafers [1]. Resolution in optical lithography obeys the Rayleigh resolution limit $R = k \frac{\lambda}{NA}$, where λ is the wavelength, NA is the numerical aperture, and k is the process constant which can be minimized through RET methods [2, 3, 4, 5]. In optical proximity correction (OPC) methods, mask amplitude patterns are modified by the addition of sub-resolution features that can pre-compensate for imaging distortions [6]. Phase-shifting mask (PSM) methods, commonly attributed to Levenson [7], induce phase shifts in the transmitted field which have a favorable constructive or destructive interference effect. Thus, a suitable modulation of both the intensity and the phase of the incident light can be used to effectively compensate for some of the resolution-limiting phenomena in optical diffraction.

Several approaches to PSM for inverse lithography have been proposed in the literature. Liu and Zakhor developed a binary and phase shifting mask design strategy based on the branch and bound algorithm and simulated annealing [8]. Pati-Kailath developed sub-optimal projections onto convex sets for PSM designs [9]. Both of the methods, however, are not based on gradient type optimization and thus the searching process for a suitable solution is either computationally expensive or not efficient. Recently, Poonawala and Milanfar introduced a novel PSM optimization framework for inverse lithography relying on a pixel-based, continuous function formulation, well suited for gradient-based search [10]. Ma and Arce generalized this algorithm so as to admit multi-phase components having arbitrary PSM patterns [11, 12]. Although these algorithms are computationally effective, they focused on coherent illumination systems.

Most practical illumination sources, however, have a nonzero line width and their radiation is more generally described as partially coherent [13]. Partially coherent illumination (PCI) is

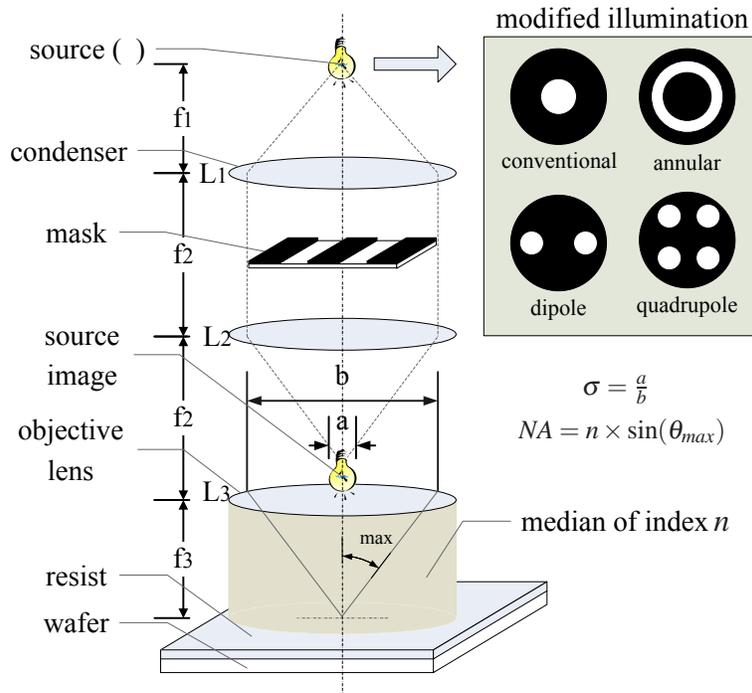


Fig. 1. Optical lithography system with partially coherent illuminations

desired, as it can improve the theoretical resolution limit [1]. In partially coherent imaging, the mask is illuminated by light travelling in various directions. The source points giving rise to these incident rays are incoherent with one another, such that there is no interference that could lead to nonuniform light intensity impinging on the mask [1]. A schematic of an optical lithography system with partially coherent illumination is illustrated in Fig. 1. The light source with a wavelength of λ is placed at the focal plane of the first condenser (L_1), illuminating the mask. 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 [14].

The optical lithography system reduces to simple forms in the two limits. When the illumination source is at a single point, the system reduces to the completely coherent case. On the other hand, when the illumination source is of infinite extent, the system reduces to the completely incoherent case. Phase-shifting masks provide no advantages in the completely incoherent case, while they make their most significant contributions to the output intensity in the completely coherent case. Common partially coherent illumination modes lie between these two limits, and include dipole, quadrupole and annular shapes, which provide small to large partial coherence factors. Illumination with large partial coherence factors is closer to the completely incoherent illumination case, while small partial coherence factors approach completely coherent illumination. There are some tradeoffs in the extent that partial coherence is used. Large partial coherence factors, such as $\sigma = 0.9$, lead to improvements on resolution and contrast. On the other hand, small partial coherence factors, such as $\sigma = 0.3$, have the advantage to form sparse patterns, which can be exploited effectively by phase-shifting masks. Medium partial coherence factors such as $\sigma = 0.5$ and $\sigma = 0.6$ are preferred for mask pattern containing both sparse and dense patterns. The smallest usable partial coherence factor is approximately

$\sigma = 0.3$ [1].

While inverse lithography optimization methods have been studied extensively in the past two decades for the case of coherent illumination, equivalent methods for inverse lithography under partially coherent illumination have not been addressed until recently [15]. In [16, 17], Ma and Arce used the sum of coherent systems (SOCS) model and the average coherent approximation model for partially coherent illumination to develop gradient-based binary mask design algorithms for inverse lithography under partially coherent illuminations. The goal of this paper is to extend the concepts in [16, 17] to focus on the development of gradient-based inverse optimization algorithms for the design of PSM under partially coherent illumination. As it will be described shortly, the proposed methods are most effective with small to medium partial coherence factors. Since the imaging synthesis and analysis of partially coherent systems are much more complex than the coherent case, the singular value decomposition (SVD) is applied to expand the partially coherent imaging equation by eigenfunctions into a sum of coherent systems [9, 18]. An iterative optimization framework for the PSM design is formulated when the partially coherent imaging system is approximated by the first order coherent approximation corresponding to the largest eigenvalue. The first order coherent approximation removes the influence among different coherent components during the inverse optimization process and reduces the computational complexity of the algorithms.

In order to attain simpler manufacturability properties, the reduction of mask pattern complexity is desired. To this end, we extend the method in [19] to develop a 2D DCT post-processing method, that reduces the detail complexity of the optimized PSM, yet preserves or improves the fidelity of the output pattern. In essence, the approach is to reduce the solution space by cutting off the high frequency components of the desired pattern in the discrete cosine spectrum. Low frequency components of the mask pattern are proved to have more influence to the fidelity of the output pattern than the high frequency components. Thus, a fraction of the high frequency components of the optimized PSM is therefore deleted. It is interesting to show that the fidelity of the output pattern may be improved by cutting of high frequency components. The relationships among the number of maintained DCT low frequency components, mask complexity and the output pattern fidelity are discussed.

The final contribution of this paper is the use of photoresist tone reversing in PSM design so as to project extremely sparse patterns [1]. Reversion of the photoresist tone thus allows for optimized PSM which can produce targets with higher spatial frequency than the resolution limit without application of the photoresist tone reversing. The remainder of the paper is organized as follows: Partially coherent imaging models are discussed in Section 2. The PSM optimization processes for partially coherent illumination with small to medium partial coherence factor is discussed in Section 3. The post-processing of the mask pattern based on 2D DCT is described in Section 4. The photoresist tone reversing technique for projecting extremely sparse patterns is described in Section 5. Conclusions are provided in Section 6.

2. Partially coherent imaging models

Practical lithography systems most often operate under partially coherent illuminations due to non-zero width sources and off-axis illuminations from spatially extended sources. Common partially coherent illumination modes include dipole, quadrupole and annular shapes, as well as small, medium, and large partial coherence factors. In order to formulate the optimization problem of ILT with partially coherent illuminations, the Hopkins diffraction model and the SOCS model based on SVD are discussed in this section.

2.1. Hopkins diffraction model

According to the Hopkins diffraction model, the light intensity distribution exposed on the wafer with PCI is bilinear and described by [20]

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

where $\mathbf{r} = (x, y)$, $\mathbf{r}_1 = (x_1, y_1)$ and $\mathbf{r}_2 = (x_2, y_2)$. $M(\mathbf{r})$ 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. In general, the intensity distribution equation in Eq. (1) is tedious to compute, both analytically and numerically [13]. The system reduces to simple forms in the two limits of complete coherence or complete incoherence. For the completely coherent case, the illumination source is at a single point, thus, $\gamma(\mathbf{r}) = 1$. In this case, the intensity distribution in Eq. (1) is separable on \mathbf{r}_1 and \mathbf{r}_2 , and thus $I(\mathbf{r}) = |M(\mathbf{r}) \otimes h(\mathbf{r})|^2$, where \otimes is the convolution operation. For the completely incoherent case, the illumination source is of infinite extent and thus, $\gamma(\mathbf{r}) = \delta(\mathbf{r})$. In this case, the intensity distribution reduces to $I(\mathbf{r}) = |M(\mathbf{r})|^2 \otimes |h(\mathbf{r})|^2$. In the frequency domain, Eq. (1) is translated as

$$I(x, y) = \int \int \int \int_{-\infty}^{+\infty} TCC(f_1, g_1; f_2, g_2) \tilde{M}(f_1, g_1) \tilde{M}^*(f_2, g_2) \times \exp\{-i2\pi[(f_1 - f_2)x + (g_1 - g_2)y]\} df_1 dg_1 df_2 dg_2, \quad (2)$$

where $\tilde{M}(f_1, g_1)$ and $\tilde{M}(f_2, g_2)$ are the Fourier transforms of $M(x_1, x_2)$ and $M(x_2, y_2)$, respectively. $TCC(f_1, g_1; f_2, g_2)$ is the transmission cross-coefficient, which indicates the interaction between $\tilde{M}(f_1, g_1)$ and $\tilde{M}(f_2, g_2)$. Specifically,

$$TCC(f_1, g_1; f_2, g_2) = \int \int_{-\infty}^{+\infty} \tilde{\gamma}(f, g) \tilde{h}(f + f_1, g + g_1) \tilde{h}^*(f + f_2, g + g_2) df dg, \quad (3)$$

where $\tilde{\gamma}(f, g)$, referred to as the effective source, is the Fourier transform of $\gamma(x, y)$. $\tilde{h}(f, g)$ is the Fourier transform of $h(x, y)$.

2.2. SVD model as the sum of coherent systems

The singular value decomposition (SVD) model, described in this section, decomposes the Hopkins diffraction model of Section 2.1 into a sum of coherent systems [18]. The result is a bank of linear systems whose outputs are squared, scaled and summed. The SVD model is summarized as follows.

Given the discretization of the mask pattern $M(x, y)$, referred to as $M(m, n)$, $m, n = 1, 2, \dots, N$, the intensity distribution on the wafer shown in Eq. (2) can be reformulated as a function of matrices

$$I(m, n) = \tilde{\mathbf{s}}^H A \tilde{\mathbf{s}}, \quad m, n = 1, 2, \dots, N, \quad (4)$$

where H is the conjugate transposition operator. $\tilde{\mathbf{s}}$ is an $N^2 \times 1$ vector, and the i th entry of $\tilde{\mathbf{s}}$ is

$$\tilde{\mathbf{s}}_i = \tilde{M}(p, q) \exp[i2\pi(pm + qn)] \quad i = 1, 2, \dots, N^2, \quad (5)$$

where $\tilde{M}(p, q) = FFT\{M(m, n)\}$, and $FFT\{\cdot\}$ is the FFT operator. $p = i \bmod N$, $q = \lceil \frac{i}{N} \rceil$, and $\lceil \cdot \rceil$ is the smallest integer larger than the argument. A is an $N^2 \times N^2$ matrix including the

information of the transmission cross-coefficient TCC . Specifically, the i th row and j th column entry of A is $A_{ij} = TCC(p, q; r, u)$, where $p = i \bmod N$, $q = \lceil \frac{i}{N} \rceil$, $r = j \bmod N$, and $u = \lceil \frac{j}{N} \rceil$. In order to reformulate Eq. (4) into the sum of coherent systems, the variable pairs of (p, q) and (r, u) in the argument of TCC should be separated by the singular value decomposition. The result of the singular value decomposition of A is $A = \sum_{k=1}^{N^2} \alpha_k V_k V_k^*$, where α_k is the k th eigenvalue, and $\alpha_1 > \alpha_2 > \dots > \alpha_{N^2}$. The $N^2 \times 1$ vector V_k is the eigenfunction corresponding to α_k . Thus Eq. (4) becomes

$$I(m, n) = \sum_{k=1}^{N^2} \alpha_k |\tilde{\mathbf{s}}^T V_k|^2. \quad (6)$$

Let $S^{-1}(\cdot)$ be the inverse column stacking operation which converts the $N^2 \times 1$ column vector V_k into a $N \times N$ square matrix $S^{-1}(V_k)$. In particular,

$$\tilde{h}_k(p, q) = S^{-1}(V_k) = \begin{pmatrix} V_{k,1} & V_{k,N+1} & \cdots & V_{k,N(N-1)+1} \\ V_{k,2} & V_{k,N+2} & \cdots & V_{k,N(N-1)+2} \\ \vdots & \vdots & \ddots & \vdots \\ V_{k,N} & V_{k,2N} & \cdots & V_{k,N^2} \end{pmatrix}, \quad (7)$$

where $V_{k,i}$ is the i th entry of V_k . Taking the inverse FFT of $\tilde{h}_k(p, q)$ leads to the k th equivalent kernel of the SOCS model,

$$h_k(m, n) = \text{IFFT}\{\tilde{h}_k(p, q)\}, \quad m, n = 1, 2, \dots, N. \quad (8)$$

Substituting Eq. (7) and Eq. (8) into Eq. (6),

$$I(m, n) = \sum_{k=1}^{N^2} \alpha_k |h_k(m, n) \otimes M(m, n)|^2. \quad (9)$$

Note that the partially coherent system is decomposed into the superposition of N^2 coherent systems. The scheme of the SOCS decomposition by SVD is depicted in Fig. 2. The i th order

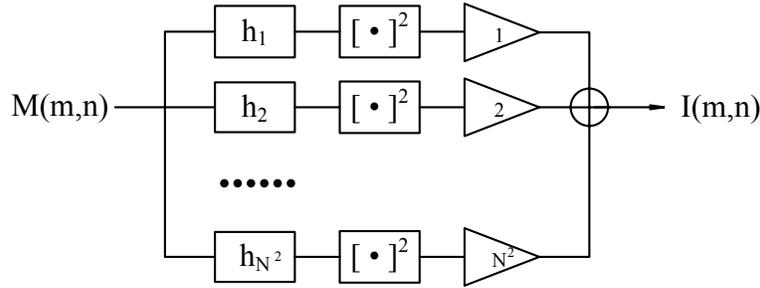


Fig. 2. A partially coherent system represented by a SVD decomposition of a sum of coherent systems.

coherent approximation to the partially coherent system is defined as

$$I(m, n) \approx \sum_{k=1}^i \alpha_k |h_k(m, n) \otimes M(m, n)|^2, \quad i = 1, 2, \dots, N^2. \quad (10)$$

An example of the first 50 eigenvalues of the SOCS decomposition with a small and medium partial coherence factors is illustrated in Fig. 3. In this simulation, the partially coherent illuminations are circular illuminations with partial coherence factors $\sigma = 0.3$ and $\sigma = 0.6$. The

dimension of the discretization is $N = 51$. The pixel size is $11\text{nm} \times 11\text{nm}$. Thus, the effective source is

$$\begin{aligned}\tilde{\gamma}(f, g) &= \frac{\lambda^2}{\pi(\sigma NA)^2} \text{circ}\left(\frac{\lambda\sqrt{f^2 + g^2}}{\sigma NA}\right) \\ &= \begin{cases} \frac{\lambda^2}{\pi(\sigma NA)^2} & \text{for } \sqrt{f^2 + g^2} \leq \frac{\sigma NA}{\lambda} \\ 0 & \text{elsewhere} \end{cases},\end{aligned}\quad (11)$$

where $NA = 1.35$ and $\lambda = 193\text{nm}$. The amplitude impulse response is defined as the Fourier transform of the circular lens aperture with cutoff frequency NA/λ [21, 22]; therefore,

$$h(\mathbf{r}) = h(x, y) = \frac{J_1(2\pi r NA/\lambda)}{2\pi r NA/\lambda}.\quad (12)$$

The Fourier transform of $h(x, y)$ is

$$\tilde{h}(f, g) = \frac{\lambda^2}{\pi(NA)^2} \text{circ}\left(\frac{\lambda\sqrt{f^2 + g^2}}{NA}\right) = \begin{cases} \frac{\lambda^2}{\pi(NA)^2} & \text{for } \sqrt{f^2 + g^2} \leq \frac{NA}{\lambda} \\ 0 & \text{elsewhere} \end{cases}.\quad (13)$$

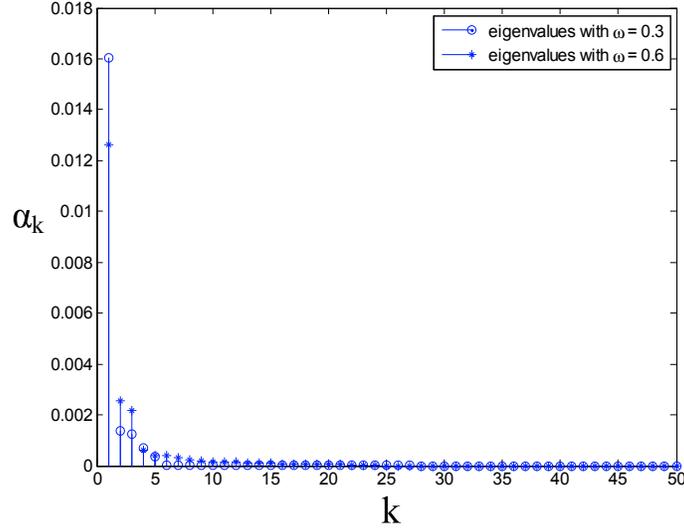


Fig. 3. Eigenvalues α_k of sum of coherent systems decomposition by SVD.

The amplitude of the first and second equivalent kernels corresponding to the first and second largest eigenvalues with $\sigma = 0.3$ are illustrated in Figs. 4(a) and 4(b), respectively. It is noted that for the illuminations having small partial coherence factors, the eigenvalues decay very rapidly. It was proved that for partial coherence factors $\sigma \leq 0.5$, a partially coherent imaging system may be approximated to within 10% error by the first order coherent approximation [9].

3. PSM Optimization Using Inverse Lithography

3.1. Optimization using the SOCS model

Let $M(m, n)$ be the input phase-shifting mask to an optical lithography system $T\{\cdot\}$, with a partially coherent illumination. The PCI optical system is approximated by a Hopkins diffraction

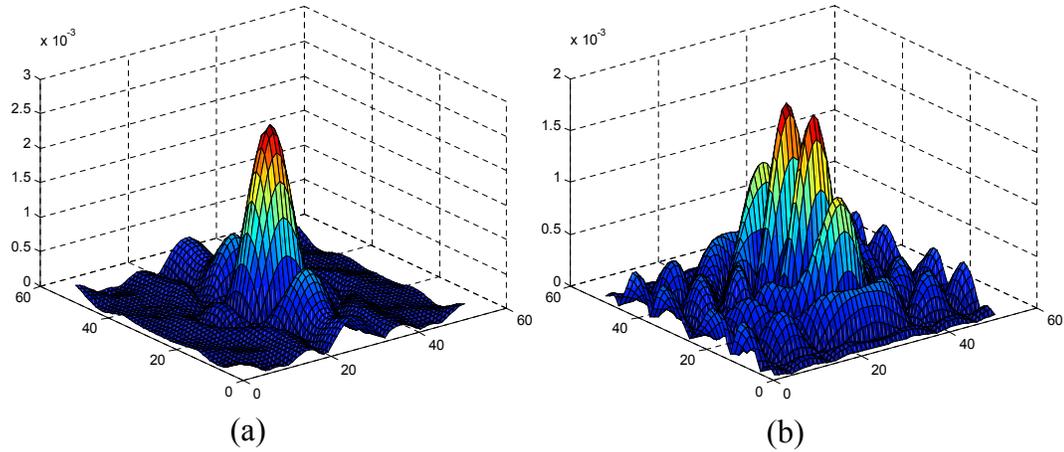


Fig. 4. (a) The amplitude of the first equivalent kernel corresponding to the largest eigenvalue $|\phi_1(x, y)|$, (b) second equivalent kernel corresponding to the second largest eigenvalue $|\phi_2(x, y)|$.

model described in Eq. (1). The effect of the photoresist is modelled by a hard threshold operation. The output pattern is denoted as $Z(m, n) = T\{M(m, n)\}$. Given a $N \times N$ desired output pattern $Z^*(m, n)$, the goal of PSM design is to find the optimized $M(m, n)$ called $\hat{M}(m, n)$ such that the distance

$$D = d(Z(m, n), Z^*(m, n)) = d(T\{M(m, n)\}, Z^*(m, n)) \quad (14)$$

is minimized, where $d(\cdot, \cdot)$ is the mean square error criterion. The PSM inverse lithography optimization problem can thus be formulated as the search of $\hat{M}(m, n)$ over the $N \times N$ real space $\mathfrak{R}^{N \times N}$ such that

$$\hat{M}(m, n) = \arg \min_{M(m, n) \in \mathfrak{R}^{N \times N}} d(T\{M(m, n)\}, Z^*(m, n)). \quad (15)$$

Figure 5 depicts the approximated forward process model [10]. The phase-shifting mask is the input of the system. Light propagating through the mask pattern is affected by diffraction and mutual interference—a phenomena described by the Hopkins Diffraction Model [6, 21, 23]. In this optimization approach, the PCI optical system is approximated by the first order coherent approximation of the SOCS model. $|\cdot|$ is the element-by-element absolute operation, and the output of the convolution and the absolute operation model is the intensity distribution of the aerial image. The aerial image is projected on a light-sensitive photoresist, which is subsequently developed through the use of solvents. The thickness of the remaining photoresist after development is proportional to the exposure dose exceeding a given threshold intensity. In a positive photoresist process, almost all the photoresist material remains in the low-exposure area on the wafer and is removed in the high-exposure area. Between these two extremes is the transition region. For mathematical simplicity, it is assumed that when the light field exceeds a threshold, the exposed area becomes a high-exposure area, otherwise, a low-exposure area. Thus, a hard threshold operation can adequately represent the exposure effect described above. Further, since the derivative of the sigmoid function exists, it is used to approximate the hard threshold function. The hard threshold function is a shifted unit step function $U(x - t_r)$, which is approximated by the sigmoid function

$$\text{sig}(x) = \frac{1}{1 + \exp[-a(x - t_r)]}, \quad (16)$$

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

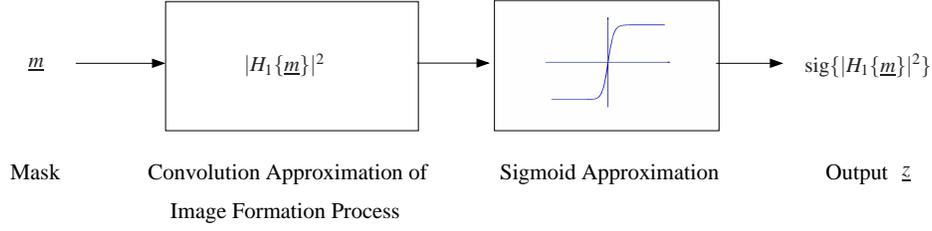


Fig. 5. Approximated forward process model.

Following the definitions above, the following notation is used:

- 1) The $M_{N \times N}$ a real-valued 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_1 is a $N^2 \times N^2$ matrix. Its equivalent two-dimensional filter is the first equivalent kernel $h_1(m, n)$ of the SOCS model.
- 3) The desired $N \times N$ binary output pattern is denoted as Z^* . It is the desired light distribution sought on the wafer. Its vector representation is denoted as \underline{z}^* .
- 4) The output of the sigmoid function is the $N \times N$ image denoted as: $Z = \text{sig}\{|H_1\{\underline{m}\}|^2\}$. 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 $N \times N$ real-valued mask denoted as \hat{M} minimizes the distance between Z and Z^* , ie,

$$\hat{M} = \arg \min_M d(\text{sig}\{|H_1\{\underline{m}\}|^2\}, Z^*). \quad (17)$$

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

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

Given the gray level pattern $\underline{z} = \text{sig}\{|H_1\{\underline{m}\}|^2\}$, the i th entry in this vector can be represented as

$$\underline{z}_i = \frac{1}{1 + \exp[-a \sum_{j=1}^{N^2} h_{1,ij} \underline{m}_j]^2 + at_r]}, \quad i = 1, \dots, N^2, \quad (18)$$

where $h_{1,ij}$ is the i, j th entry of the first equivalent kernel $h_1(m, n)$. In the optimization process, $\hat{\underline{m}}$ is searched to minimize the L_2 norm of the difference between \underline{z} and \underline{z}^* . Therefore,

$$\hat{\underline{m}} = \arg \min_{\hat{\underline{m}}} \{F(\hat{\underline{m}})\}, \quad (19)$$

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

$$F(\hat{\underline{m}}) = \|\hat{\underline{z}}^* - \hat{\underline{z}}\|_2^2 = \sum_{i=1}^{N^2} (\hat{z}_i^* - \hat{z}_i)^2, \quad (20)$$

where \hat{z}_i in Eq. (20) is represented in Eq. (18). In order to reduce the above bound-constrained optimization problem to an unconstrained optimization problem, we adopt the parametric transformation [10]. Let $\hat{\underline{m}}_j = \cos(\underline{\theta}_j)$, $j = 1, \dots, N^2$, where $\underline{\theta}_j \in (-\infty, \infty)$ and $\hat{\underline{m}}_j \in [-1, 1]$. Defin-

ing the vector $\underline{\theta} = [\theta_1, \dots, \theta_{N^2}]^T$, the optimization problem is formulated as

$$\begin{aligned} (\hat{\underline{\theta}}) &= \arg \min_{\underline{\theta}} \{F(\underline{\theta})\} \\ &= \arg \min_{\underline{\theta}} \left\{ \sum_{i=1}^{N^2} \left(z_i^* - \frac{1}{1 + \exp[-a | \sum_{j=1}^{N^2} h_{1,ij} \cos \theta_j |^2 + at_r]} \right)^2 \right\}. \end{aligned} \quad (21)$$

The steepest-descent method is used to optimize the above problem. The gradients $\nabla F(\underline{\theta})_{\underline{\theta}}$ can be calculated as follows:

$$\begin{aligned} \nabla F(\underline{\theta}) = \underline{d}_{\underline{\theta}} &= 2a \times \sin \underline{\theta} \odot \{ (H_1^{*H} [(z^* - \underline{z}) \odot \underline{z} \odot (\mathbf{1} - \underline{z}) \odot (H_1^* \underline{m})]) \} \\ &\quad + 2a \times \sin \underline{\theta} \odot \{ (H_1^H [(z^* - \underline{z}) \odot \underline{z} \odot (\mathbf{1} - \underline{z}) \odot (H_1 \underline{m})]) \}, \end{aligned} \quad (22)$$

where $\nabla F(\underline{\theta}) \in \Re^{N^2 \times 1}$, \odot is the element-by-element multiplication operator. $\mathbf{1} = [1, \dots, 1]^T \in \Re^{N^2 \times 1}$. Assuming $\underline{\theta}^k$ is the k^{th} iteration result, then at the $k+1^{\text{th}}$ iteration:

$$\underline{\theta}^{k+1} = \underline{\theta}^k - s_{\underline{\theta}} \underline{d}_{\underline{\theta}}^k, \quad (23)$$

where $s_{\underline{\theta}}$ is the step-size.

The iterative optimization above, in general, leads to a real-valued mask with pixel values between -1 and 1. Therefore a post-processing step is needed to obtain the trinary optimized mask, $\hat{m}_{tri,i} = \text{sgn}(\hat{m}_i) U(|\hat{m}_i| - t_m)$, $i = 1, \dots, N^2$, where t_m is a global threshold. We define the pattern error E as the distance between the desired output image Z^* and the actual binary output pattern \hat{Z}_b evaluated by the equation Eq. (2) and a hard threshold operator,

$$E = \sum_{i=1}^{N^2} |z_i^* - z_{bi}|^2 = \sum_{i=1}^{N^2} |z_i^* - T\{M_{tri}\}|^2. \quad (24)$$

3.2. Discretization regularization

In the prior simulation settings, the fact that the estimated output pattern should be trinary is not considered. An additional post-processing (trinarization) of the gray optimized mask pattern is suboptimal with no guarantee that the pattern error is under the goal [10]. One approach to overcome this disadvantage is through regularization during the optimization process [10, 24]. Regularization is formulated as follows:

$$\hat{m} = \arg \min_{\underline{m}} \{F(\underline{m}) + \gamma R(\underline{m})\}, \quad (25)$$

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. γ is the user-defined parameter to reveal the weight of the regularization. In the following, the discretization penalty is discussed.

In order to attain near-trinary optimized mask pattern through the optimization process, we adopt the discretization penalty [10]. The formulation of the discretization penalty is summarized as following. For each pixel value, the discretization penalty term is

$$r_D(m_i) = -4.5m_i^4 + m_i^2 + 3.5, \quad i = 1, \dots, N^2. \quad (26)$$

Thus, the cost function in Eq. (20) is adjusted as $J(\underline{m}) = F(\underline{m}) + \gamma_D R_D(\underline{m})$. In our simulations for $\sigma = 0.3$, discretization regularization attains near-trinary optimized mask and reduce 30% output pattern error.

3.3. Simulations

In order to demonstrate the validity of the optimization algorithms, consider the desired pattern shown in Fig. 6, with dimension of $561nm \times 561nm$. The matrices representing all of the patterns have dimension of $N \times N$, where $N = 51$. The pixel size is $11nm \times 11nm$. The partially coherent illumination is a circular illumination with small partial coherence factor $\sigma = 0.3$. In

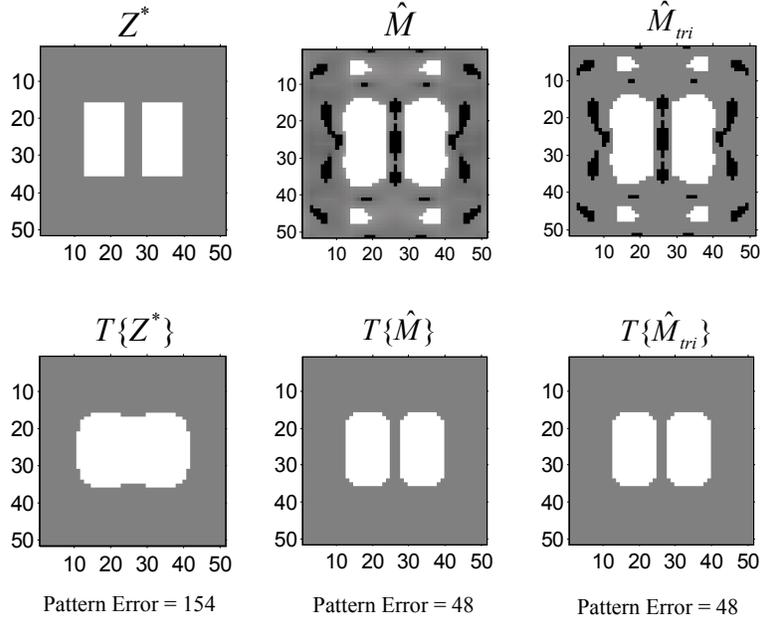


Fig. 6. Top row (input masks), left to right: desired pattern, optimized real-valued mask and optimized trinary mask. The bottom row illustrates the corresponding binary output patterns. White, grey and black represent 1, 0 and -1, respectively. $\sigma = 0.3$.

Figure 6, the top row illustrates the input masks of (left) the desired pattern Z^* , (center) the optimized real-valued mask \hat{M} , and (right) the optimized trinary mask \hat{M}_{tri} . The optimized trinary mask, referred to as the alternating phase-shifting mask, includes clear areas and shifting areas, which introduce 180° phase difference with each other. In particular, the transmission coefficients of the clear area and shifting area are assigned to 1 and -1, respectively. The binary output patterns are shown in the bottom row. White, grey and black represent 1, 0 and -1, respectively. The effective source and the amplitude impulse response are shown in Eq. (11) and Eq. (12) with $NA = 1.35$ and $\lambda = 193nm$. In the sigmoid function, we assign parameters $a = 200$ and $t_r = 0.003$. The binary output patterns in the bottom row are evaluated by the equation Eq. (2) followed by a hard threshold operator with threshold $\bar{t}_r = tr \times \sum_{k=1}^{N^2} \alpha_k$. The global threshold is $t_m = 0.33$. The step length and the regularization weights are $s_{\underline{\theta}} = 0.2$ and $\gamma_D = 0.1$. The initial mask pattern is the same as the desired binary output pattern Z^* . For $\underline{\theta}$, we assign the phase of $\frac{\pi}{2}$ corresponding to the areas having a magnitude of 1 and the phase of $\frac{\pi}{2}$ for the areas with magnitude of 0. As shown in Fig. 6, this approach is effective for small partial coherence factors. These results are consistent with those obtained in other numerous simulations we have ran with small partial coherence factors.

These simulations, are then repeated in Fig. 7 with the same parameters except for the value of the partial coherence factor which in this case was raised to $\sigma = 0.6$. The output pattern error corresponding to the optimized trinary mask in this case is increased compared with that

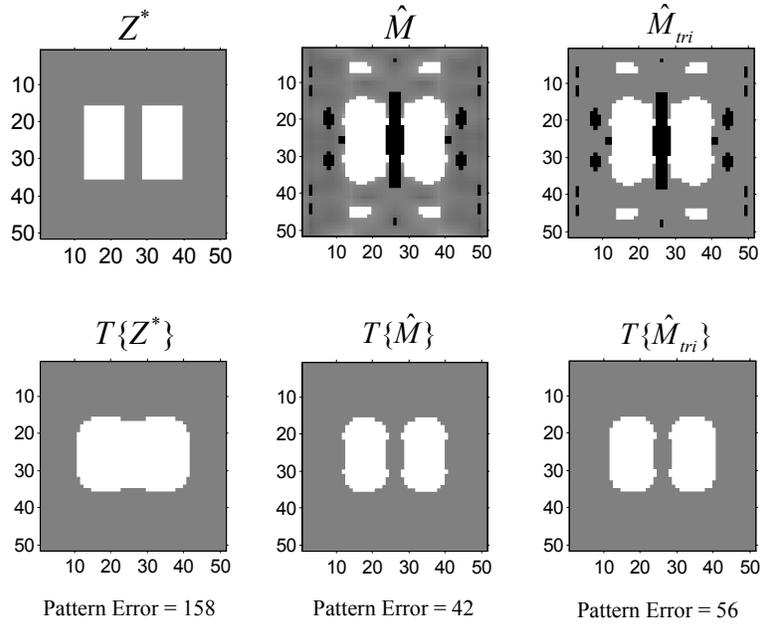


Fig. 7. Top row (input masks), left to right: desired pattern, optimized real-valued mask and optimized trinary mask. The bottom row illustrates the corresponding binary output patterns. White, grey and black represent 1, 0 and -1, respectively. $\sigma = 0.6$.

of Fig. 6. This degradation results from the less accurate first order coherent approximation to the partially coherent system when medium or large partial coherence factors are used. In fact, the SVD approach taken here gradually degrades as the partial coherent factor increases from small to large values. Nevertheless, the optimized PSM attains a 65% reduction of the output pattern error even with the medium partial coherent factor values.

4. Post-processing of Mask Pattern Based on 2D DCT

In order to attain simpler manufacturability properties, the reduction of the complexity of mask patterns is desired. Recently, J. Zhang, et al., proposed an efficient mask design for inverse lithography based on 2D DCT [19]. The solution space is greatly reduced by cutting off the high frequency components of the desired pattern in discrete cosine spectrum. Low frequency components of the mask pattern were proved to have more influence to the fidelity of the output pattern than the high frequency components. From this point of view, a post-processing of the mask pattern based on 2D DCT is introduced in this Section. It is observed that most of the energy of the mask pattern concentrates on the low frequency components. In order to reduce the complexity of the mask pattern, a subset of high frequency components of the optimized real-valued mask are cut off. The post-processed real-valued mask \hat{M}' is the inverse 2D DCT of the maintained low frequency components. The post-processed trinary mask \hat{M}'_{tri} is the discretization of \hat{M}' .

Figure 8 illustrated the relationship between the number of maintained DCT low frequency components and the output pattern errors with partial coherence factors $\sigma = 0.3$ (solid line) and $\sigma = 0.6$ (dot line). Since the inverse lithography is an ill-posed problem, numerous input patterns can lead to the same binary output pattern. Thus the post-processing based on 2D DCT can even simultaneously reduce the complexity of the masks and the output pattern errors. It is

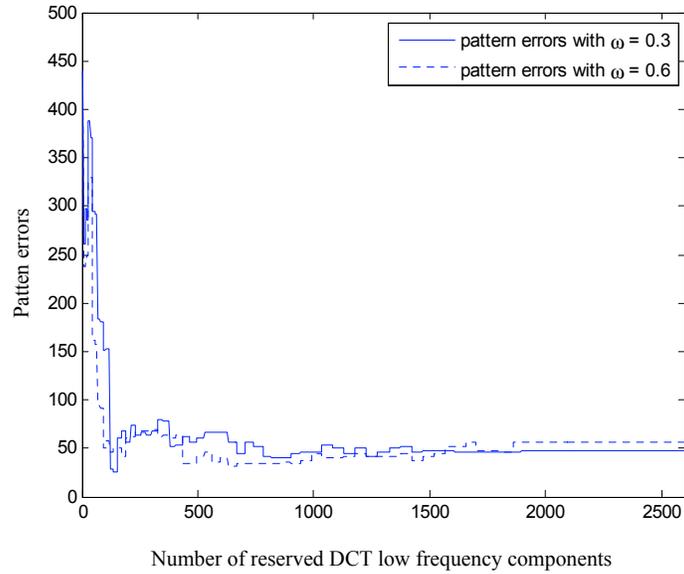


Fig. 8. The relationship between the number of maintained DCT low frequency components and the output pattern errors.

shown in Fig. 8 that the fidelity of the output patterns is improved by maintaining just 136 low frequency components with $\sigma = 0.3$ and 666 low frequency components with $\sigma = 0.6$. Figure 9 illustrates the simulations of the PSM optimization using the DCT post-processing. All of the parameters are the same as the simulations shown in Fig. 6. In Fig. 9, the left figure shows the output pattern of the desired pattern. The middle figure shows the post-processed trinary PSM mask using DCT post-processing maintaining 136 low frequency components. The right figure shows the binary output pattern of the post-processed trinary mask. White, grey and black represent 1, 0 and -1, respectively.

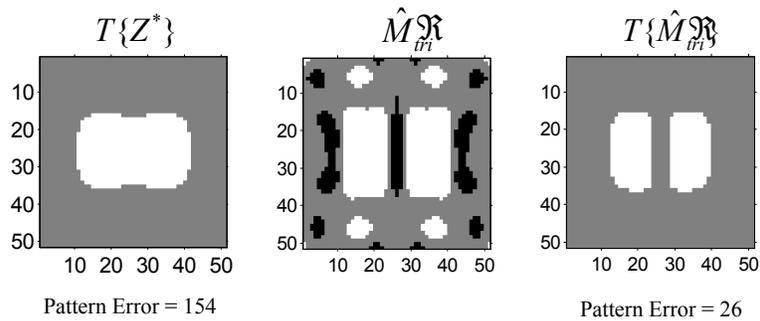


Fig. 9. Left to right: output pattern when the desired pattern is inputted, post-processed trinary mask with the DCT post-processing maintaining 136 low frequency components, and the binary output pattern of the post-processed trinary mask. White, grey and black represent 1, 0 and -1, respectively.

5. PSM Optimization with Photoresist Tone Reversing

As a final extension of the PSM design method introduced in this paper, we focus on inverse lithography optimization where photoresist tone reversing is used. Photoresist can be divided by its polarity. In a positive photoresist process, more photoresist material remains in the low-exposure area on the wafer and less in the high-exposure area. Negative photoresist responds in the opposite manner. Photoresist tone reversing method exploits both kinds of the photoresist materials on the wafer. Reversion of the photoresist tone is proposed to improve lithography performance of sub-resolution features [1]. In this section, the photoresist tone reversing is used to image a desired resolution and contrast for the sparse pattern, whose resolution limit is much higher than the traditional case without application of the photoresist tone reversing. Consider an optical lithography system applying monotonous positive photoresist with the parameters $k = 0.29$, $\lambda = 193nm$ and $NA = 1.35$. The resolution limit is

$$R = k \frac{\lambda}{NA} = 41.5nm. \quad (27)$$

PSM optimization without use of photoresist tone reversing fails to print an aerial image containing features with dimensions smaller than the limit in Eq. (27). An example is illustrated in Fig. 10, where the desired image of dimension $561nm \times 561nm$ contains two pairs of vertical bars, each with pitch width of $22nm < R$. In this simulation, all of the parameters are the same as the simulation shown in Fig. 6, except for $t_r = 0.0003$. In Fig. 10, from left to right, the first figure shows the desired pattern. The second figure shows the binary output pattern of the desired pattern. The third figure shows the optimized trinary PSM. The last figure shows the binary output pattern of the optimized PSM. White, grey and black represent 1, 0 and -1, respectively. Note that the binary output pattern of the optimized PSM is totally different from the desired pattern, indicating that the PSM optimization approach cannot attain the desired output pattern on the wafer. The reason is that the dimension of the features in the desired pattern is smaller than the resolution limit without application of the photoresist tone reversing.

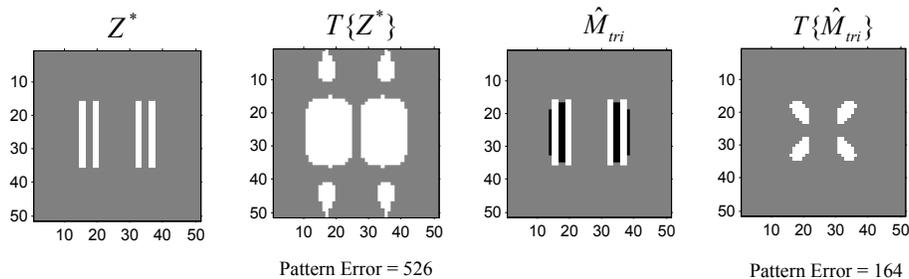


Fig. 10. Left to right: desired pattern, output pattern when the desired pattern is inputted, optimized trinary mask, and the output pattern of the optimized trinary mask. White, grey and black represent 1, 0 and -1, respectively. $\sigma = 0.3$.

In order to overcome the limit, photoresist tone reversing is exploited to find an adequate distribution of positive and negative photoresist on the wafer. One possibility of the distribution is shown in the left figure in Fig. 11. White and black represent positive and negative photoresist, respectively. Assign negative photoresist in the gaps in each pair of the bars and positive photoresist in other areas. If the optimized mask is able to expose a rectangular aerial image in the area laying over each pair of bars, the negative photoresist will prevent the exposure in the gaps. Thus, the binary output pattern is the same as the desired pattern. Therefore, the PSM optimization approach is used to search the binary output pattern of two rectangles on the

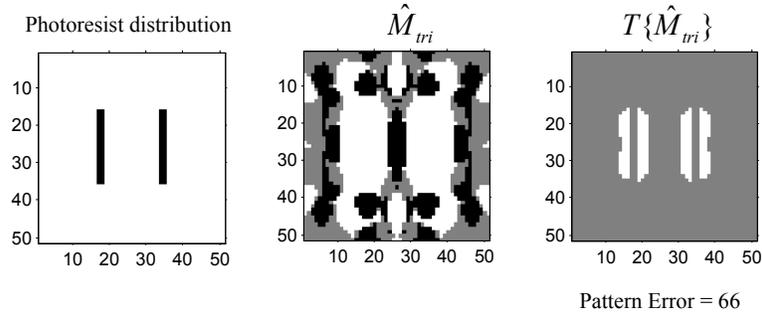


Fig. 11. Left to right: Photoresist distribution, optimized trinary mask using the photoresist tone reversing method, and the output pattern of the optimized trinary mask. White and black represent positive and negative photoresist, respectively in the first figure. White, grey and black represent 1, 0 and -1, respectively in the second and the third figures. $\sigma = 0.3$.

wafer. In Fig. 11, $t_r = 0.01$, and the optimized trinary mask is shown in the middle figure. The binary output pattern of the optimized trinary mask is shown in the right figure. White, grey and black represent 1, 0 and -1, respectively. It is obvious that photoresist tone reversing method is effective to expose a sparse feature, whose resolution limit is much higher than the traditional case without application of photoresist tone reversing.

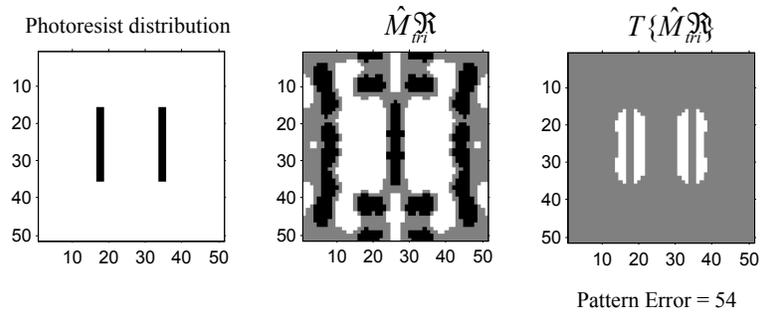


Fig. 12. Left to right: Photoresist distribution, post-processed trinary mask using the DCT post-processing maintaining 276 low frequency components, and the binary output pattern of the post-processed trinary mask. White and black represent positive and negative photoresist, respectively in the first figure. White, grey and black represent 1, 0 and -1, respectively in the second and the third figures. $\sigma = 0.3$.

The DCT post-processing of the mask developed in Section 4 can be simultaneously exploited with the photoresist tone reversing method. The simulation is illustrated in Fig. 12. The left figure shows the photoresist distribution. The middle figure shows the post-processed trinary mask using the DCT post-processing maintaining 276 low frequency components. The right figure shows the binary output pattern of the post-processed trinary mask. Note that the DCT post-processing successfully reduces the error of the binary output pattern from 66 to 54.

The heuristic photoresist distribution design approach described above is suitable for simple target patterns. The joint optimization of the photoresist distribution and PSMs is desirable and is of interest for future research work.

6. Conclusions

This paper focuses on PSM inverse lithography under partially coherent illuminations. Firstly, two kinds of partially coherent imaging models are described: the Hopkins diffraction model and the sum of coherent systems decomposition model by SVD. Based on the second model, a first order coherent approximation is used to represent the Hopkins diffraction model, and inverse lithography technology PSM optimization processes is formulated. In order to simultaneously reduce the complexity of the mask and the output pattern error, the post-processing based on 2D DCT is introduced. Finally, the photoresist tone reversing technique is exploited to design PSMs capable of projecting extremely sparse patterns, whose resolution limit is much higher than the traditional case without application of photoresist tone reversing.

Acknowledgments

We wish to thank Christof Krautschik, Yan Borodovsky and the TCAD group at the Intel corporation for their comments and support.