Binary mask optimization for inverse lithography with partially coherent illumination

Xu Ma* **and Gonzalo Arce**

Department of Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716, USA
**Corresponding author: maxu@udel.edu*

Received June 30, 2008; revised September 16, 2008; accepted September 21, 2008;
posted September 29, 2008 (Doc. ID 97804); published November 11, 2008

Recently, a set of generalized gradient-based optical proximity correction optimization methods have been developed to solve for the inverse lithography problem under coherent illumination. Most practical lithography systems, however, operate under partially coherent illumination. This paper focuses on developing gradient-based binary mask optimization methods that account for the inherent nonlinearities of partially coherent systems. Two nonlinear models are used in the optimization. The first relies on a Fourier representation that approximates the partially coherent system as a sum of coherent systems. The second model is based on an average coherent approximation that is computationally faster. To influence the solution patterns toward more desirable manufacturability properties, wavelet regularization is added to the optimization framework.
© 2008 Optical Society of America
OCIS codes: 220.3740, 100.3190, 110.4980, 100.7410.

1. INTRODUCTION

Because of the resolution limits of optical lithographic 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]. Resolution in optical lithography obeys the Rayleigh resolution limit $R=k\lambda/NA$, where λ is the wavelength, NA is the numerical aperture, and k is the process constant that can be minimized through RET methods [2–5]. In optical proximity correction (OPC), mask amplitude patterns are modified by the addition of subresolution features that can precompensate for imaging distortions [6].

Several approaches to inverse lithography have been proposed in the literature. These range from heuristic and empirically based design rules to computationally expensive optimization-based inverse algorithms. Sherif *et al.* derived an iterative approach to generate binary masks in incoherent diffraction-limited imaging systems [7]. Liu and Zakhor developed a binary and phase shifting mask (PSM) design strategy based on the branch and bound algorithm and simulated annealing [8]. Pati and Kailath developed suboptimal projections onto convex sets for PSM designs [9]. In addition, Erdmann *et al.* proposed automatic optimization of the mask and illumination parameters with a genetic algorithm [10]. Pang *et al.* gave an overview of inverse lithography technology (ILT) and provided some simulations to demonstrate the benefit of ILT [11]. Granik described and compared solutions of inverse mask problems [12], and solved the general nonlinear formulation by the local variations and gradient descent methods [13]. However, the searching process of the methods mentioned above for a suitable solution is either computationally expensive or not efficient.

Poonawala and Milanfar recently introduced a novel optimization framework for inverse lithography relying on a pixel-based, continuous function formulation well suited for gradient-based search [6,14]. Ma and Arce generalized this algorithm so as to admit multiphase components having arbitrary PSM patterns [15,16]. These algorithms are computationally effective; however, they focused on coherent illumination systems. Most practical illumination sources have a nonzero line width and their radiation is more generally described as partially coherent [17]. While the inverse lithography methods derived in [6,14–16] are effective in coherent illumination, these algorithm will not produce adequate results when applied to a partially coherent illumination system. Note that the application of the gradient-based optimization approach to solve for the partially coherent illumination problem was suggested in [14]; the derivation of the iterative algorithm was not developed or studied at length. Partially coherent illumination (PCI) is desired, since it can improve the theoretical resolution limit. PCI is thus introduced in practice through modified illumination sources having large coherent factors or through annular-shaped illumination.

In partially coherent imaging, the mask is illuminated by light traveling 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 [1,18]. According to the Hopkins diffraction model, the light intensity distribution exposed on the wafer in PCI is bilinear and described by [19]

$$I(\mathbf{r}) = \int\int M(\mathbf{r}_1)M(\mathbf{r}_2)\gamma(\mathbf{r}_1-\mathbf{r}_2)h^*(\mathbf{r}-\mathbf{r}_1)h(\mathbf{r}-\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2, \tag{1}$$

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 coher-

1084-7529/08/122960-11/$15.00 © 2008 Optical Society of America

ence, 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, this equation is tedious to compute, both analytically and numerically [17]. 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

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

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

$$I(\mathbf{r}) = |M(\mathbf{r})|^2 \otimes |h(\mathbf{r})|^2. \quad (3)$$

The imaging synthesis and analysis of partially coherent systems are thus more complex than those of the coherent or incoherent imaging systems.

The contribution of this paper is the development of computationally effective inverse optimization algorithms for the design of binary masks with OPC for lithography under partially coherent illumination. The sum-of-coherent-system (SOCS) model introduces a decomposition approach to represent a partially coherent system as a summation of several coherent systems, and thus the gradient-based optimization results in a summation of terms for the gradient search that must be taken into account. As the degree of coherence approaches the value of one, the iterative optimization solution reduces to the methods developed in [6] for the coherent illumination problem. This is accomplished as follows: First, an iterative optimization framework is formulated in which the bilinear Hopkins diffraction model of partially coherent imaging systems is represented by a SOCS model based on a Fourier series expansion [9,19]. The iterative algorithm used is a generalization of the work in [6], but in this case, the SOCS nonlinear model is used to obtain the iterative optimization algorithm. To control the complexity of the mask pattern, wavelet regularization is added to the optimization framework.

A second contribution of the paper is the introduction of a fast algorithm for binary mask optimization in PCI. Although the SOCS model is an accurate representation of PCI, it is computationally expensive, as numerous terms of the Fourier series expansion are needed to obtain an adequate representation. The fast optimization algorithm uses the so-called average coherence approximation model [17] to represent the underlying PCI system. The inverse lithography optimization algorithm using this latter model is significantly faster than the first approach. The drawback of the fast algorithm is that the associated error of the output pattern can be larger than that of the slower approach, since the accuracy of the average coherence approximation model depends on the spatial coordinates, mask pattern, optical system kernel, and the complex degree of coherence [17]. Extensive simulations, however, show that the errors suffered by both approaches do not differ significantly.

The remainder of the paper is organized as follows. Two partially coherent imaging models are discussed in Section 2. Mask optimization processes for PCI lithography using these models are developed in Section 3. Several inverse regularization methods are described in Section 4 with the objective of obtaining optimized masks with better manufacturability properties. Simulations are illustrated in Section 5, where the effects of the degree of coherence on the mask synthesis are shown. Conclusions are provided in Section 6. The derivations of the gradient-based optimization solution are presented in the appendixes.

2. PARTIALLY COHERENT IMAGING MODELS

PCI has been shown to improve the theoretical resolution limit in lithography. Thus, practical lithography systems often operate under PCI emitted by nonzero-width sources and off-axis illumination from spatially extended sources. Common PCI modes include dipole, quadrupole, and annular illumination. To formulate the optimization problem of ILT with PCI, two kinds of partially coherent imaging models are discussed in this section. The first is a Fourier series expansion model that approximates the partially coherent system as a sum of coherent systems [19]. The second is based on an average coherent approximation [17]. Both methods are described next, as they are central to the development of the inverse lithography optimization algorithm.

A. Fourier Series Expansion Model

A schematic of an optical lithography system with PCI is illustrated in Fig. 1. The light source with a wavelength of

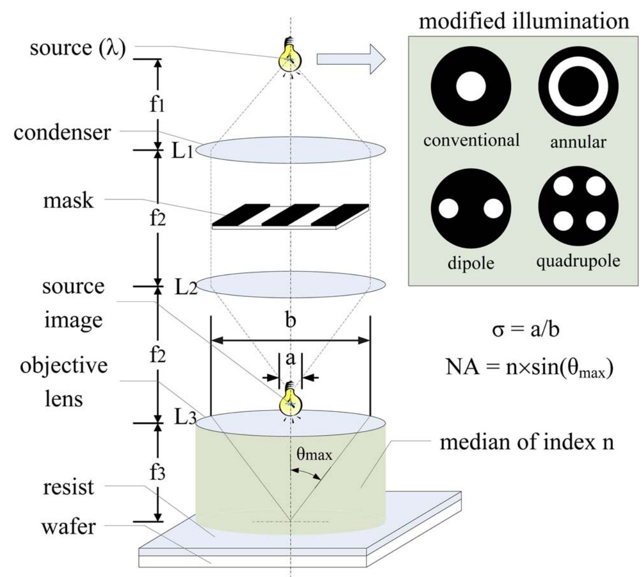


Fig. 1. (Color online) Optical lithography system with partially coherent illumination.

λ is placed at the focal plane of the first condenser, illuminating the mask. Common illumination 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 $\sigma = a/b$ is defined as the ratio between the size of the source image and that of 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 [9,19]. In typical imaging applications, the size of the mask is much larger than the width of the complex degree of coherence $\gamma(\mathbf{r})$. In this first approach, the Fourier series expansion model is applied to reduce the computation cost of partially coherent imaging.

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

$$\gamma(\mathbf{r}) = \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} \exp(j\omega_0 \mathbf{m} \cdot \mathbf{r}), \quad (4)$$

$$\Gamma_{\mathbf{m}} = \frac{1}{D^2} \int_{A_\gamma} \gamma(\mathbf{r}) \exp(j\omega_0 \mathbf{m} \cdot \mathbf{r}) d\mathbf{r}, \quad (5)$$

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 Eq. (4) into Eq. (1), the light intensity on the wafer is given by

$$I(\mathbf{r}) = \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |M(\mathbf{r}) \otimes h^{\mathbf{m}}(\mathbf{r})|^2, \quad (6)$$

where

$$h^{\mathbf{m}}(\mathbf{r}) = h(\mathbf{r}) \exp(j\omega_0 \mathbf{m} \cdot \mathbf{r}). \quad (7)$$

It is observed from Eqs. (6) and (2) that the PCI is equal to the superposition of coherent systems. Since the Fourier series expansion model is based on direct discretization of the Hopkins diffraction model, they have the same accuracy. For the annular illumination, the complex degree of coherence is

$$\gamma(\mathbf{r}) = \frac{J_1(2\pi r/2D_{cu})}{2\pi r/2D_{cu}} - \frac{D_{cu}^2}{D_{cl}^2} \frac{J_1(2\pi r/2D_{cl})}{2\pi r/2D_{cl}}, \quad (8)$$

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

$$\Gamma_{\mathbf{m}} = \begin{cases} \dfrac{4D_{cu}^2 D_{cl}^2}{\pi D^2 (D_{cl}^2 - D_{cu}^2)} & \text{for } D/2D_{cl} \leq |\mathbf{m}| \leq D/2D_{cu} \\ 0 & \text{elsewhere} \end{cases}, \quad (9)$$

where D_{cl} and D_{cu} are the coherent lengths of the inner and outer circles, respectively. $\sigma_{inner} = \lambda/2D_{cl}NA$ and $\sigma_{outer} = \lambda/2D_{cu}NA$ 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/λ [20,21]; therefore,

$$h(\mathbf{r}) = \frac{J_1(2\pi r NA/\lambda)}{2\pi r NA/\lambda}. \quad (10)$$

Annular illuminations are classified by the sizes of their inner and outer partial coherence factors. The larger the partial coherence factor, the higher the resolvable spatial frequency. Thus, large partial coherence factors lead to improvements in resolution and contrast. Small partial coherence factors, on the other hand, have the advantage of forming sparse patterns, which can be exploited effectively by phase-shifting masks. Medium partial coherence factors are preferred for mask patterns containing both sparse and dense patterns [1].

Figure 2 illustrates annular illumination sources having large, medium, and small partial coherence factors. For the large partial coherence factor illumination in Fig. 2, $\sigma_{inner} = 0.8$ and $\sigma_{outer} = 0.975$. For the medium illumination, $\sigma_{inner} = 0.5$ and $\sigma_{outer} = 0.6$. For the small illumination, $\sigma_{inner} = 0.3$ and $\sigma_{outer} = 0.4$. The dashed circles represent the dimension of the pupil. The number of terms used in the Fourier series expansion in Eq. (6) plays a critical role in the computational complexity of the model and consequently in the computational complexity of the inverse optimization. The number of terms in the expansion in Eq. (6) will be referred to as T. According to Eq. (9), $D/2D_{cl} \leq |\mathbf{m}| \leq D/2D_{cu}$. In addition, $D_{cl} = \lambda/2\sigma_{inner}$, $D_{cu} = \lambda/2\sigma_{outer}$, and $D = N \times p$, where $p \times p$ is the pixel size. Thus,

$$T \sim \pi[(D/2D_{cu})^2 - (D/2D_{cl})^2] \sim CN^2, \quad (11)$$

where the constant $C = \pi p^2 NA^2 (\sigma_{\text{outer}}^2 - \sigma_{\text{inner}}^2)/\lambda^2$. The parameter T is larger for sources with larger partial coherence factors. As an example, the values of T for the sources in Fig. 2 are 12, 12, and 52 as σ increases from smaller to larger partial coherence factors.

Figure 3 illustrates a mask of dimensions 1035 nm $\times$ 1035 nm, and the corresponding aerial images formed by the annular illuminations having large, medium, and small partial coherence factors. The mask consists of 45 nm features. The pitch $p = 90$ nm is indicated by dashed lines. The aerial images are synthesized by the Fourier series expansion model. In these simulations, $NA = 1.25$, $\lambda = 193$ nm, and $h(\mathbf{r})$ is assumed to vanish outside the area A_h defined by $x,y \in [-56.25 \text{ nm}, 56.25 \text{ nm}]$. The constant $k = 0.29$ and the pixel size is 5.63 nm $\times$ 5.63 nm. Note that the aerial images increasingly be-

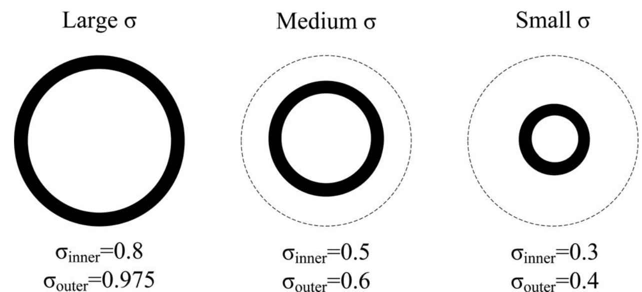


Fig. 2. Annular illumination with large, medium, and small partial coherence factors. The dashed circles represent the dimension of the pupil.

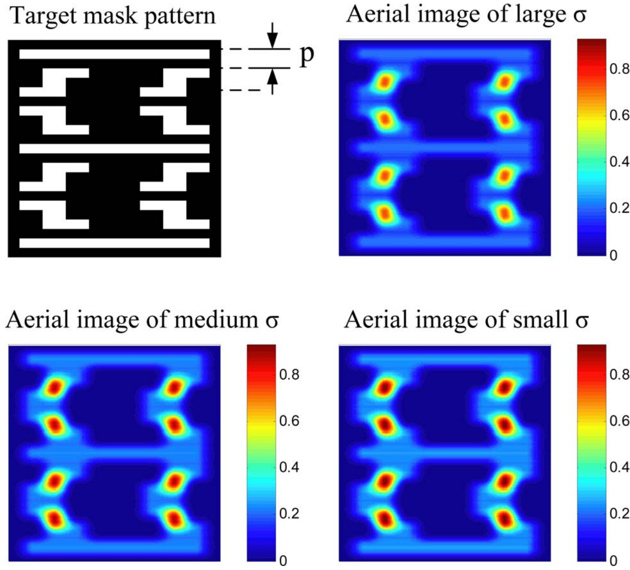


Fig. 3. (Color online) (Top left) Target mask pattern containing 45 nm features with pitch $p=90$ nm is indicated by dashed lines. Aerial images formed by annular illuminations with large (top right: $\sigma_{inner}=0.8$, $\sigma_{outer}=0.975$), medium (bottom left: $\sigma_{inner}=0.5$, $\sigma_{outer}=0.6$), and small (bottom right: $\sigma_{inner}=0.3$, $\sigma_{outer}=0.4$) partial coherence factors. Here NA=1.25 and $k=0.29$.

come more blurred as the partial coherence factor is decreased.

B. Average Coherent Approximation Model

Although the accuracy of the Fourier series expansion is the same as that of the discrete version of the Hopkins diffraction model, it is computationally expensive. The computational cost is a polynomial of the number of the Fourier series term used to represent the SOCS model, and in general, numerous terms are needed to attain an adequate representation. Therefore, the development of an optimization algorithm that uses a more computationally efficient model is desired. To this end, the second inverse lithography optimization method developed in this paper uses the model introduced by Salik *et al.* and referred to as "average approximation" for PCI [17].

The principal idea of the average approximation of PCI is to approximately decompose the contribution of the PCI into a coherent illumination component and an incoherent illumination component. Therefore,

$$I(\mathbf{r}) = \int\int M(\mathbf{r}_1)M(\mathbf{r}_2)\gamma(\mathbf{r}_1-\mathbf{r}_2)h^*(\mathbf{r}-\mathbf{r}_2)h(\mathbf{r}-\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \approx \left|\int M(\mathbf{r}')h_C(\mathbf{r}',\mathbf{r})\mathrm{d}\mathbf{r}'\right|^2 + \int |M(\mathbf{r}')|^2|h_I(\mathbf{r}',\mathbf{r})|^2\mathrm{d}\mathbf{r}', \tag{12}$$

where $\mathbf{r}'=(x',y')$, $h_C(\mathbf{r}',\mathbf{r})$, and $h_I(\mathbf{r}',\mathbf{r})$ are the equivalent amplitude impulse responses of the coherent and incoherent components, respectively. Furthermore,

$$h_C(\mathbf{r}',\mathbf{r}) = f(\mathbf{r}',\mathbf{r})^{1/2}h(\mathbf{r}',\mathbf{r}), \tag{13}$$

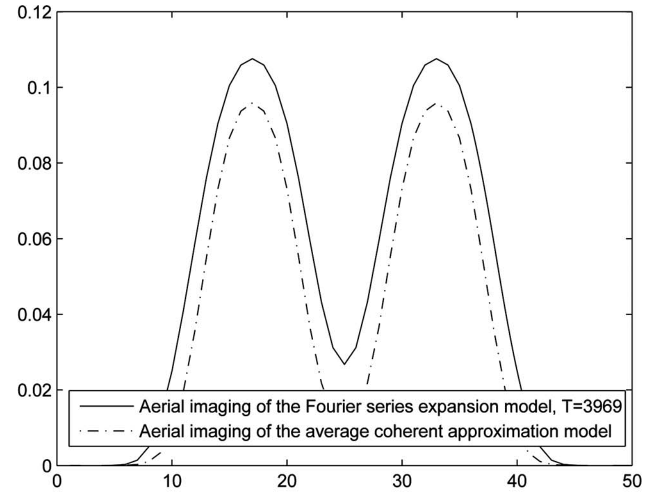
$$h_I(\mathbf{r}',\mathbf{r}) = [1-f(\mathbf{r}',\mathbf{r})]^{1/2}h(\mathbf{r}',\mathbf{r}), \tag{14}$$

where

$$f(\mathbf{r}',\mathbf{r}) = \frac{\int|h(\mathbf{r}',\hat{\mathbf{r}})|^2\mu(\mathbf{r},\tilde{\mathbf{r}})\mathrm{d}\tilde{\mathbf{r}}}{\int|h(\mathbf{r}',\hat{\mathbf{r}})|^2\mathrm{d}\mathbf{r}}, \tag{15}$$

$$\mu(\mathbf{r},\hat{\mathbf{r}}) = \frac{\gamma(\mathbf{r},\hat{\mathbf{r}})}{[\gamma(\mathbf{r},\mathbf{r})\gamma(\hat{\mathbf{r}},\hat{\mathbf{r}})]^{1/2}}. \tag{16}$$

In the equations above $\hat{\mathbf{r}}$ and $\tilde{\mathbf{r}}$ are dummy variables. f is the fraction of coherent incident power with $0 \leqslant f \leqslant 1$. For the annular illuminating sources, the function f is obtained by substituting Eqs. (8) and (10) into Eqs. (15) and (16), leading to



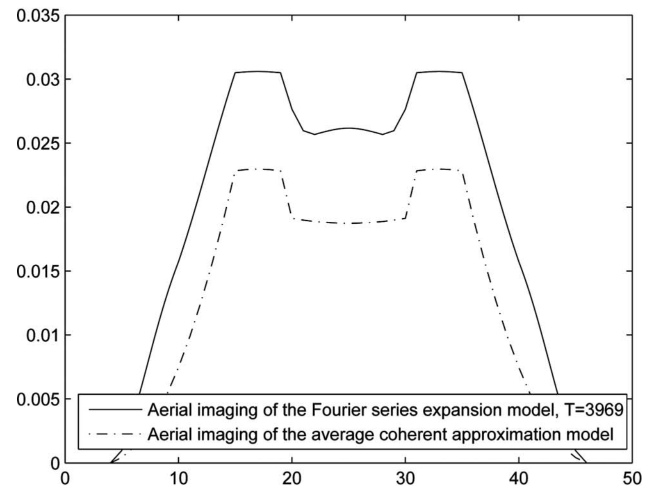


Fig. 4. ACAA gives more accurate aerial imaging for sharper amplitude impulse response. (a) NA=1.35, corresponding to a sharper amplitude impulse response, SNR=18.7. (b) NA=0.15, corresponding to a smoother amplitude impulse response, SNR =10.2.

$$f(\mathbf{r}',\mathbf{r}) = \frac{1}{\int|[J_1(\pi\mathbf{r}/d)]/(\pi\mathbf{r}/d)|^2 d\mathbf{r}} \int \left| \frac{[J_1(\pi\hat{\mathbf{r}}/d)]}{(\pi\hat{\mathbf{r}}/d)} \right|^2 \times \left(\frac{[J_1(\pi(\hat{\mathbf{r}}-\dot{\mathbf{r}})/a_u)]}{[\pi(\hat{\mathbf{r}}-\dot{\mathbf{r}})/a_u]} - \frac{a_u^2}{a_l^2} \frac{[J_1(\pi(\hat{\mathbf{r}}-\dot{\mathbf{r}})/a_l)]}{[\pi(\hat{\mathbf{r}}-\dot{\mathbf{r}})/a_l]} \right) d\hat{\mathbf{r}}, \quad (17)$$

where $\dot{\mathbf{r}}=\mathbf{r}-\mathbf{r}'$, $d=\lambda/2NA$, $a_u=2D_{cu}$, and $a_l=2D_{cl}$. Applying the Fourier transform, Eq. (17) becomes

$$f(\mathbf{r}',\mathbf{r}) = \frac{1}{\int|[J_1(\pi\mathbf{r}/d)]/(\pi\mathbf{r}/d)|^2 d\mathbf{r}} \cdot \mathrm{IFFT} \times [\mathrm{FFT}(|h(\mathbf{r})|^2) \cdot \mathrm{FFT}(\gamma(\mathbf{r}))], \quad (18)$$

where FFT(·) and IFFT(·) are the FFT and inverse FFT operations, respectively.

It is noted that for some specific $h(\mathbf{r})$ and $\gamma(\mathbf{r})$, the condition $0 \leqslant f \leqslant 1$ may not be satisfied. According to Eq. (13), negative values of f will introduce complex pixel values in $h_C(\mathbf{r}',\mathbf{r})$. Similarly, Eq. (14) indicates that values of f larger than 1 will introduce complex pixel values in $h_I(\mathbf{r}',\mathbf{r})$. Nevertheless, in the following binary mask optimization simulations, the average coherent approximation model leads to similar aerial image contours as those of the Fourier series expansion model. Substituting Eq. (18) into Eqs. (13) and (14), the equivalent amplitude impulse responses of the coherent and incoherent components can be found. Therefore, the PCI is approximately divided into a summation of a coherent system and an incoherent system.

As mentioned above, the accuracy of the average coherence approximation model depends on the spatial coordinates, mask pattern, optical system kernel, and the complex degree of coherence [17]. Specifically, the error of the average coherence approximation model becomes smaller as the amplitude impulse response $h(\mathbf{r})$ becomes sharper, or its energy is more concentrated. Figure 4 illustrates the cross sections of the aerial imaging of two vertical bars based on the two models introduced above. The mask dimension is 600 nm × 600 nm and $\lambda=193$ nm. The Fourier series coefficients of the circular illumination are

$$\Gamma_{\mathbf{m}} = \begin{cases} 4D_c^2/\pi D^2 & \text{for } |\mathbf{m}| \leqslant D/2D_c \\ 0 & \text{elsewhere} \end{cases}, \quad (19)$$

where $D=600$ nm and $D_c=8.6$ nm. $T=3969$ is the number of Fourier series terms to represent the SOCS model.

In Fig. 4, the solid curves and dashed–dotted curves represent the aerial imaging of the Fourier series expansion model and average coherent approximation model respectively. Since the accuracy of the Fourier series expansion model is the same as that of the discrete version of the Hopkins diffraction model, the former is chosen as the criterion by which to measure the accuracy of the average coherent approximation model. The SNR is defined as the ratio between the energy of the accurate imaging and the error energy. In Fig. 4(a), the numerical aperture $NA=1.35$ and SNR=18.7, while in Fig. 4(b), $NA=0.15$ and SNR=10.2. It can be observed from Eq. (10) that larger NA corresponds to a sharper amplitude impulse response. The simulations show that the average coherent approximation model gives more accurate aerial imaging for sharper amplitude impulse response.

3. INVERSE LITHOGRAPHY OPTIMIZATION

A. Optimization Using the Fourier Series Expansion Model

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

$$D = d[Z(x,y),Z^*(x,y)] = d[T\{M(x,y)\},Z^*(x,y)] \quad (20)$$

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

$$\hat{M}(x,y) = \arg \min_{M(x,y)\in\Re^{N\times N}} d\{T[M(x,y)],Z^*(x,y)\}. \quad (21)$$

The forward imaging process is illustrated in Fig. 5, where the Fourier series expansion model is used in the image formation stage. The binary mask is the input of the system. Light propagating through the mask pattern is affected by diffraction and mutual interference—a phenomenon described by the Hopkins diffraction model [6,20,22]. Light that is transmitted through the mask reaches a light-sensitive photoresist that 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

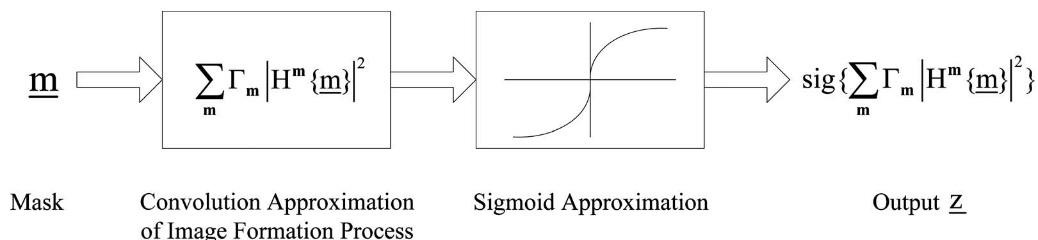


Fig. 5. Approximated forward process model.

a high-exposure area; otherwise, it is a low-exposure area. Thus, a hard threshold operation can adequately represent the exposure effect described above, and the output pattern of the optical system is binary.

Figure 5 depicts the approximated forward process model [14], where $|\cdot|$ 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. 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 $U(x-t_r)$ that is approximated by the sigmoid function

$$\mathrm{sig}(x) = \frac{1}{1+\exp[-a(x-t_r)]}, \tag{22}$$

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

Following the definitions above, the following notation is 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 an $N^2\times N^2$ matrix with an equivalent two-dimensional filter $h^{\mathbf{m}}$.
3. The desired $N\times N$ binary output pattern is denoted as 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

$$Z = \mathrm{sig}\left\{\sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |H^{\mathbf{m}}(\underline{m})|^2\right\}. \tag{23}$$

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

5. The hard threshold version of $\bar{Z}$ 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 Z^*, that is,

$$\hat{M} = \arg\min_{M} d\left\{\mathrm{sig}\left[\sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |H^{\mathbf{m}}(\underline{\mathbf{m}})|^2\right], Z^*\right\}. \tag{24}$$

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

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

Given the gray level pattern $\underline{z} = \mathrm{sig}[\Sigma_{\mathbf{m}}\Gamma_{\mathbf{m}}|H^{\mathbf{m}}(\underline{m})|^2]$, the ith entry in this vector can be represented as

$$\underline{z}_i = \frac{1}{1+\exp(-a\Sigma_{\mathbf{m}}\Gamma_{\mathbf{m}}|\Sigma_{j=1}^{N^2} h_{ij}^{\mathbf{m}} \underline{m}_j|^2 + at_r)} \qquad i=1,\ldots N^2, \tag{25}$$

where h_{ij} is the i,jth entry of the filter. 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,

$$\hat{\underline{m}} = \arg\min_{\hat{\underline{m}}}[F(\underline{m})], \tag{26}$$

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

$$F(\underline{m}) = \|\underline{z}^* - \underline{z}\|_2^2 = \sum_{i=1}^{N^2} (\underline{z}_i^* - \underline{z}_i)^2, \tag{27}$$

where $\underline{z}_i$ in Eq. (27) is represented in Eq. (25). To reduce the above bound-constrained optimization problem to an unconstrained optimization problem, we adopt the parametric transformation [6]. Let

$$\underline{m}_j = \frac{1+\cos(\underline{\theta}_j)}{2} \qquad j=1,\ldots,N^2, \tag{28}$$

where $\underline{\theta}_j \in (-\infty,\infty)$, and $\underline{m}_j \in [0,1]$. Defining the vector $\underline{\theta} = (\theta_1,\ldots,\theta_{N^2})^T$, the optimization problem is formulated as

$$(\hat{\underline{\theta}}) = \arg\min_{\underline{\theta}}[F(\underline{\theta})] = \arg\min_{\underline{\theta}}\left\{\sum_{i=1}^{N^2}\left[z_i^* - \frac{1}{1+\exp\left(-a\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}\left|\sum_{j=1}^{N^2} h_{ij}^{\mathbf{m}}\frac{1+\cos\underline{\theta}_j}{2}\right|^2 + at_r\right)}\right]^2\right\}. \tag{29}$$

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

$$\begin{aligned}\nabla F(\underline{\theta}) = \underline{d}_{\underline{\theta}} &= a\times\sin\underline{\theta}\odot\left\{\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}(H^{\mathbf{m}})^{*T}[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{\mathbf{1}}-\underline{z})\odot(H^{\mathbf{m}})^*(\underline{m})]\right\} \\ &+ a\times\sin\underline{\theta}\odot\left\{\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}(H^{\mathbf{m}})^{T}[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{\mathbf{1}}-\underline{z})\odot(H^{\mathbf{m}})(\underline{m})]\right\},\end{aligned} \tag{30}$$

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

$$\underline{\theta}^{k+1} = \underline{\theta}^k - s_{\underline{\theta}}\underline{d}_{\underline{\theta}}^k, \tag{31}$$

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

The iterative optimization above, in general, leads to gray masks with pixel values between 0 and 1. Therefore a postprocessing step is needed to obtain the binary optimized mask $\hat{\underline{m}}_{bi}=U(\hat{\underline{m}}_i-t_m)$, $i=1,\ldots,N^2$, 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 Z_b:

$$E=\sum_{i=1}^{N^2}|z_i^*-z_{bi}|=\sum_{i=1}^{N^2}\left|z_i^*-U\left(\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}|H^{\mathbf{m}}m_{bi}|^2-t_r\right)\right|. \quad (32)$$

When the pattern error is reduced to a tolerable level, the steepest-descent iteration is stopped.

Note that if the complex degree of coherence approaches the value of one, then $\gamma\to 1$, the system becomes completely coherent, and $H^{\mathbf{m}}\to H$ in Eq. (30). The gradient in Eq. (30) then reduces to

$$\nabla F(\underline{\theta})=\underline{d}_\theta=a\times\sin\underline{\theta}\odot\{H^T[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{1}-\underline{z})\odot H(\underline{m})]\}, \quad (33)$$

which is the result obtained in Eq. (6) for mask optimization in the completely coherent case.

B. Optimization Using the Average Coherent Approximation Model

The optimization process based on the Fourier series expansion model suffers from expensive computational cost. For a mask pattern with dimension $N\times N$, the number of multiplication operations in each iteration is $(2T^2+4T)N^4+(10T+2)N^2+2$, where again T is the number of Fourier series terms used to represent the PCI. According to Eq. (11), $T\sim CN^2$. The computational cost is thus of the order of $O(2C^2N^8)\sim O(C'N^8)$, where $C'=2C^2$. In general, N is large. Therefore, development of a fast algorithm is desired. Based on the average coherent approximation model, a fast algorithm, referred to as the average coherence approximation algorithm (ACAA) is proposed in this section to reduce the computational complexity.

Equations(25), (27), and (30) are thus modified to account for the use of the average coherence approximation, leading to

$$\underline{z}_i=\frac{1}{1+\exp[-a(|\Sigma_{j=1}^{N^2}h_{Cij}\underline{m}_j|^2+\Sigma_{j=1}^{N^2}|h_{Iij}|^2|\underline{m}_j|^2)+at_r']}$$

$$i=1,\ldots,N^2, \quad (34)$$

$$F(\underline{\theta})=\|\underline{z}^*-\underline{z}\|_2^2=\sum_{i=1}^{N^2}(z_i^*-z_i)^2=\sum_{i=1}^{N^2}\left\{z_i^*-\frac{1}{1+\exp[-a(|\Sigma_{j=1}^{N^2}h_{Cij}\underline{m}_j|^2+\Sigma_{j=1}^{N^2}|h_{Iij}|^2|\underline{m}_j|^2)+at_r']}\right\}^2, \quad (35)$$

$$\nabla F(\underline{\theta})=\underline{d}_\theta=2a\times\sin\underline{\theta}\odot\{H_C^T[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{1}-\underline{z})\odot(H_C(\underline{m})]\}+2a\times\sin\underline{\theta}\odot\{H_I^{2T}[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{1}-\underline{z})\odot(\underline{m})]\}, \quad (36)$$

where t_r' is the process threshold for the average coherent approximation model. In general $t_r'\neq t_r$ and must be estimated *a priori*, such that it leads to a binary output pattern similar to that of the Fourier series expansion model. The aerial imaging synthesis can be implemented based on the two models. Thus, given t_r for the Fourier series expansion model, t_r' can be found using a line search process. The convolution matrices H_C and H_I are each of size $N^2\times N^2$ with equivalent two-dimensional filters h_c and h_i, respectively.

For a mask pattern with dimension $N\times N$, the number of multiplication operations in each iteration is equal to $8N^4+12N^2+4$. The computational cost of this second approach is of the order of $O(8N^4)$. Compared with the computational complexity of the algorithm based on Fourier series expansion model, and ignoring the lower order of large numbers, the reduction of the computational complexity is of the order of $8/C'N^4$. When N is much larger than 1, the fast algorithm is significantly more efficient. The drawback of ACAA is that the error of the corresponding optimized output pattern can be higher, because of the inaccuracy of the average coherent approximation model. Nevertheless, the simulations in Section 5 show that ACAA is effective for the inverse lithography problem.

4. REGULARIZATION

In the prior simulation settings, the fact that the estimated output pattern should be binary is not considered. An additional postprocessing (binarization) of the gray optimized mask pattern is suboptimal with no guarantee that the pattern error is under the goal [6]. Furthermore, the optimized mask patterns contain numerous details that may bring difficulty to mask fabrication. One approach to overcome the two disadvantages is through regularization during the optimization process [6,23]. Regularization is formulated as follows:

$$\hat{\underline{m}}=\arg\min_{\hat{\underline{m}}}[F(\underline{m})+\gamma R(\underline{m})], \quad (37)$$

where $F(\underline{m})$ is the data-fidelity term and $R(\underline{m})$ is the regularization term that 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, we will discuss quadratic penalty and wavelet penalty.

A. Quadratic Penalty

To obtain near-binary gray optimized mask patterns through the optimization process, we adopt the quadratic penalty [6]. The formulation of the quadratic penalty is summarized as follows. The quadratic penalty term is

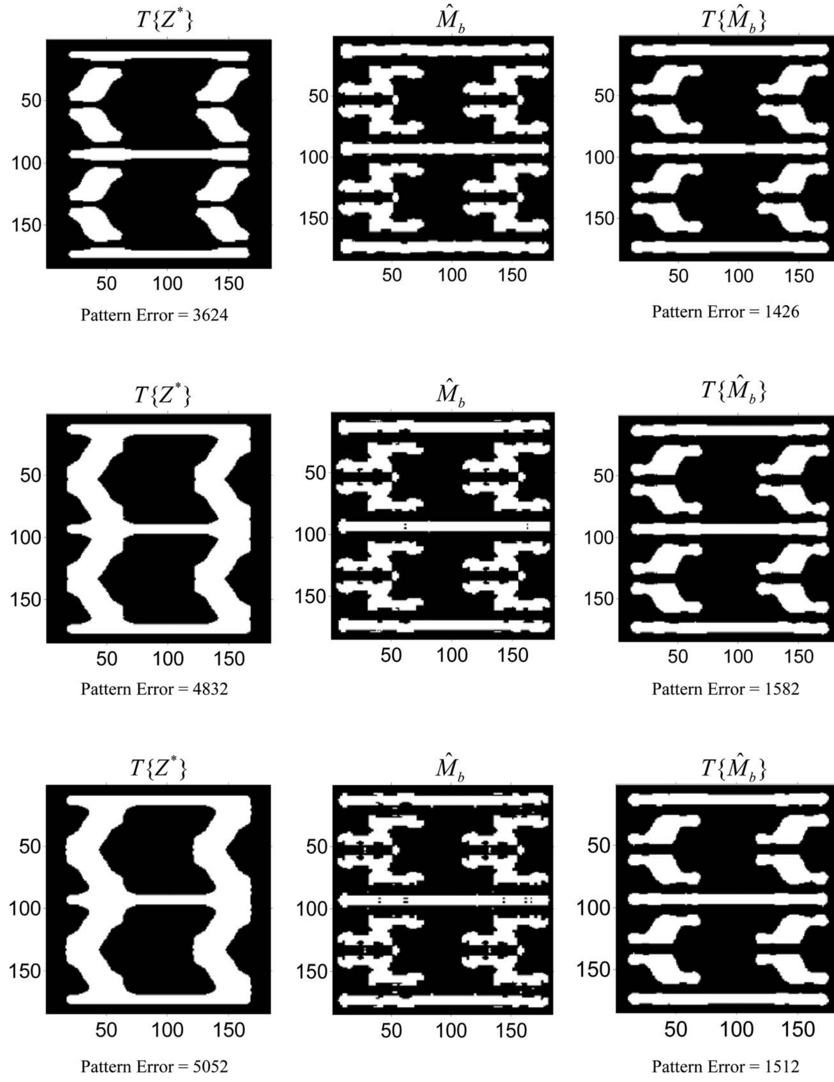


Fig. 6. Binary mask optimization using the Fourier series expansion model and wavelet penalty. Left to right: the output pattern when target pattern is used as input, the binary optimized mask, and the output pattern of binary optimized mask. Top row illustrates the simulations using the annular illumination with large partial coherence factor ($\sigma_{inner}=0.8$, $\sigma_{outer}=0.975$); middle row, with medium partial coherence factor ($\sigma_{inner}=0.5$, $\sigma_{outer}=0.6$); bottom row, with small partial coherence factor ($\sigma_{inner}=0.3$, $\sigma_{outer}=0.4$).

$$R_Q(\underline{m}) = 4\underline{m}^T(\underline{\mathbf{1}} - \underline{m}). \tag{38}$$

For each pixel value, the corresponding penalty is the quadratic function

$$r(\underline{m}_i) = 1 - (2\underline{m}_i - 1)^2 \qquad i = 1, \ldots, N^2. \tag{39}$$

According to Eq. (21), the gradient of $R_q(\underline{m})$ is $\nabla R_Q(\underline{m}) = (-8\underline{m}+4)$; thus, the cost functions in Eqs. (27) and (35) are adjusted as

$$J(\underline{m}) = F(\underline{m}) + \gamma_Q R_Q(\underline{m}). \tag{40}$$

B. Wavelet Penalty

A novel wavelet penalty based on Haar wavelets was introduced by Ma and Arce [15,16] to remove those details in the mask patterns. The formulas of the wavelet penalty are summarized as follows. The wavelet penalty term is

$$R_W = h_{11}^2 + h_{12}^2 \cdots + h_{(N/2)(N/2)}^2 + v_{11}^2 + v_{12}^2 \cdots + v_{(N/2)(N/2)}^2 + d_{11}^2 + d_{12}^2 \cdots + d_{(N/2)(N/2)}^2, \tag{41}$$

where

$$h_{ij} = m_{[2(i-1)+1][2(j-1)+1]} - n_{[2(i+1)+1][2(j-1)+2]} + m_{[2(i-1)+2][2(j-1)+1]} - m_{[2(i-1)+2][2(j-1)+2]}, \tag{42}$$

$$v_{ij} = m_{[2(i-1)+1][2(j-1)+1]} + m_{[2(i-1)+1][2(j-1)+2]} - m_{[2(i-1)+2][2(j-1)+1]} - m_{[2(i-1)+2][2(j-1)+2]}, \tag{43}$$

$$d_{ij} = m_{[2(i-1)+1][2(j-1)+1]} - m_{[2(i-1)+1][2(j-1)+2]} - m_{[2(i-1)+2][2(j-1)+1]} + m_{[2(i-1)+2][2(j-1)+2]}, \tag{44}$$

for $i,j=1,\ldots,N/2$. The gradient of the wavelet penalty is given as

$$\frac{\partial R_W}{\partial \underline{\theta}_{[(2-1)+p][2(j-1)+q]}} = -\frac{1}{2} \sin \underline{\theta}_{[2(i-1)+p][2(j-1)+q]} \times \{3\underline{m}_{[2(i-1)+p][2(j-1)+q]} - \underline{m}_{[2(i-1)+p_1][2(j-1)+q]} - \underline{m}_{[2(i-1)+p][2(j-1)+q_1]} - \underline{m}_{[2(i-1)+p_1][2(j-1)+q_1]}\}, \tag{45}$$

where $i,j=1,\ldots,N/2,p,q=1$ *or* 2, $p_1=(p+1)$ mod 2 and $q_1=(q+1)$ mod 2. Therefore, the cost function is adjusted as

$$J(\underline{m}) = F(\underline{m}) + \gamma_Q R_Q(\underline{m}) + \gamma_W R_W(\underline{m}). \tag{46}$$

Another regularization that can be used for detail removal is the total variation penalty used by Poonalwala and Milanfar [6]. The wavelet penalty was proved to outperform the total variation penalty in the setting of PSM design with arbitrary discrete phase levels [15,16]. In the following simulations, the wavelet penalty is exploited to constrain the details in the optimized mask patterns.

5. SIMULATIONS

A. Simulations Using the Fourier Series Expansion Model

To demonstrate the validity of the optimization algorithms, consider the same desired pattern as that shown in Fig. 3, with dimension of 1035 nm × 1035 nm. To prove the universality of the algorithm for different sizes of illuminations, the simulations are repeated based on annular illuminations with large, medium, and small partial coherence factors. The values of the inner and outer partial coherence factors of the illuminations are the same as those in Fig. 2. In Fig. 6, the top row illustrates the simulation results applying the large annular illumination with $\sigma_{inner}=0.8$ and $\sigma_{inner}=0.975$: (left) the output pattern when the desired pattern is inputted $[T(Z^*)]$, (center) the binary optimized mask $(\hat{M}_b)$ using the Fourier series expansion model, and (right) the output pattern of the binary optimized mask $[T(\hat{M}_b)]$. The middle row shows the simulation results applying the medium annular illumination with $\sigma_{inner}=0.5$ and $\sigma_{inner}=0.6$. The bottom row

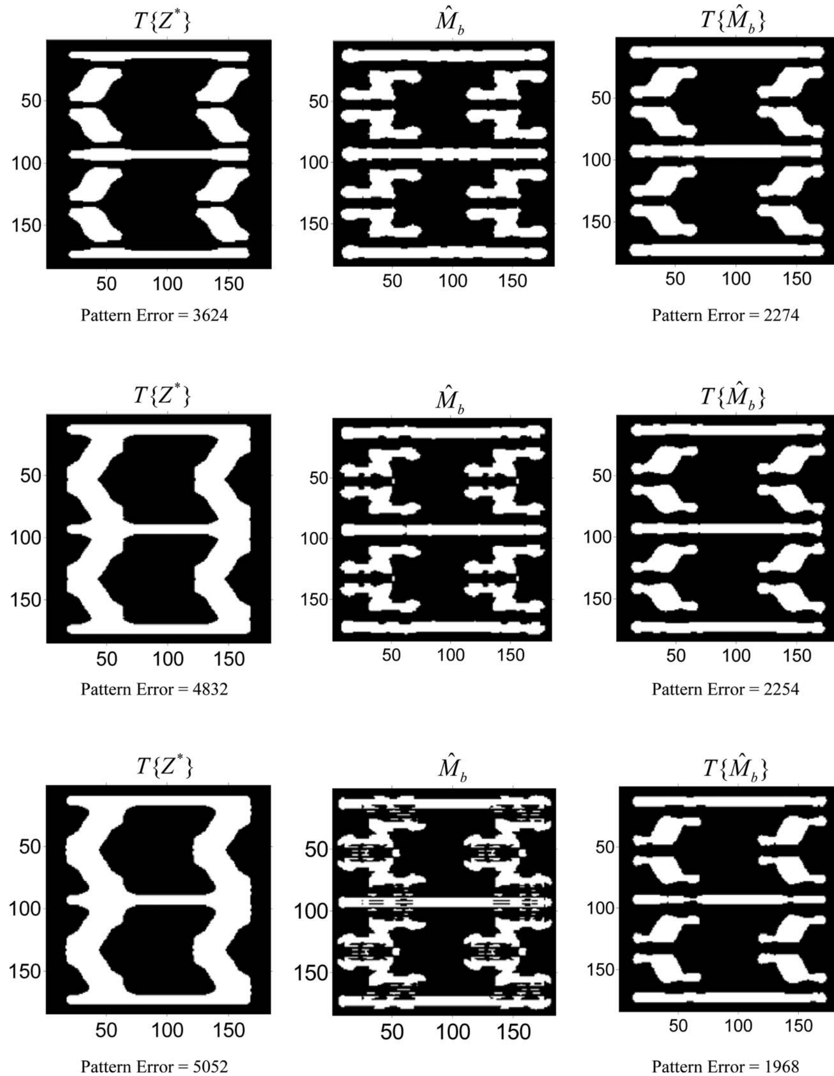


Fig. 7. Binary mask optimization using the ACAA and wavelet penalty. Left to right: the output pattern when target pattern is used as input, the binary optimized mask, and the output pattern of binary optimized mask. Top row illustrates the simulations using the annular illumination with large partial coherence factor ($\sigma_{inner}=0.8$, $\sigma_{outer}=0.975$); middle row, with medium partial coherence factor ($\sigma_{inner}=0.5$, $\sigma_{outer}=0.6$); bottom row, with small partial coherence factor ($\sigma_{inner}=0.3$, $\sigma_{outer}=0.4$).

shows the simulation results applying the small annular illumination with $\sigma_{inner}=0.3$ and $\sigma_{inner}=0.4$. The Fourier series expansion coefficients of the annular illumination are as shown in Eq. (9). The convolution kernel is as shown in Eq. (10) with $NA=1.25$ and $\lambda=193$ nm. In the simulations, we assume $h(\mathbf{r})$ vanishes outside the area A_h defined by $x,y\in(-56.25\text{ nm},56.25\text{ nm})$. In the sigmoid function, we assign parameters $a=25$ and $t_r=0.19$. The global threshold is $t_m=0.5$, the constant $k=0.29$, and the pixel size is $5.625\text{ nm}\times5.625\text{ nm}$. Step length and regularization weights are $s_{\underline{\theta}}=2$, $\gamma_Q=0.025$, and $\gamma_W=0.025$. The initial mask pattern is the same as the desired binary output pattern Z^*. For $\underline{\theta}$, we assign the phase of $\pi/5$ corresponding to the areas having a magnitude of one and the phase of $4\pi/5$ the areas with magnitude of zero. The number of Fourier series terms to represent the SOCS model is 52 for the large annular illumination, 12 for the medium one, and 12 for the small one. As shown in Fig. 6, this approach has been proved efficient in our extensive simulation analysis.

B. Simulations Using the Average Coherent Approximation Model

Although the accuracy of the ACAA depends on various parameters, it is shown in the following simulations that it is in general effective for the inverse lithography problem. In Fig. 7, the simulations shown in Fig. 6 are repeated using ACAA, with $s_{\underline{\theta}}=0.5$, $\gamma_Q=0.025$, $\gamma_W=0.025$. $t_r'=0.09$ for the large annular illumination, $t_r'=0.095$ for the medium one, and $t_r'=0.17$ for the small one. The simulation time is effectively reduced by an order of magnitude, indicating that the the ACAA is indeed computationally efficient. However, it is found that the optimized output pattern errors increase somewhat because of the reduced accuracy of the ACAA. Nevertheless, this approach can also effectively add subresolution blocks into the optimized mask patterns and compensate for distortion in the optical system.

To illustrate the computational efficiency of both algorithms, the computational running time is illustrated here. The construction of the inverse mask shown in Fig. 6 and 7 with medium partial coherence factor took 18 minutes for the first approach, and 55 seconds for the second approach. The computation was done on an Intel Pentium4 CPU, 3.40 GHz, 1.00 GB of RAM. This difference would be scaled with the dimension of the mask being constructed. It should be noted that other factors such as the treatment of boundary regions and hierarchy management can affect the overall run time and should be taken into account for large-scale mask designs. Such considerations fall outside the scope of our paper and are topics for future work.

6. CONCLUSION

This paper studies and models the PCI for inverse lithography. First, two types of partially coherent imaging models are described for the optimization problem: the Fourier series expansion model and the ACAA. The Fourier series expansion model gives an accurate representation of PCI, while the ACAA supplies approaches of faster imaging synthesis and analysis. Based on these models, ILT optimization processes are formulated. It is noted that AACA is capable of reducing the computation complexity. As a tradeoff, the optimized output pattern errors are increased because of the inaccuracy of the ACAA. To control the amount of details, wavelet penalty terms are applied in the optimization framework. Simulations illustrate that our approaches are effective and practical.

APPENDIX A: DERIVATION OF GRADIENT $\nabla F(\underline{\theta})$ FOR THE FOURIER SERIES EXPANSION MODEL

The derivation of Eq. (30) from Eq. (29) is as follows:

$$\frac{\partial F(\underline{\theta})}{\partial \underline{\theta}_k}=2\sum_{i=1}^{N^2}\left[z_i^*-\frac{1}{1+\exp\left(-a\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}\left|\sum_{j=1}^{N^2}h_{ij}^{\mathbf{m}}\frac{1+\cos\underline{\theta}_j}{2}\right|^2+at_r\right)}\right]\frac{1}{\left[1+\exp\left(-a\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}\left|\sum_{j=1}^{N^2}h_{ij}^{\mathbf{m}}\frac{1+\cos\underline{\theta}_j}{2}\right|^2+at_r\right)\right]^2}$$

$$\times\exp\left(-a\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}\left|\sum_{j=1}^{N^2}h_{ij}^{\mathbf{m}}\frac{1+\cos\underline{\theta}_j}{2}\right|^2+at_r\right)\times(-a)\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}\left[\left(\sum_{j=1}^{N^2}h_{ij}^{\mathbf{m}}\frac{1+\cos\underline{\theta}_j}{2}\right)h_{ik}^{\mathbf{m}*}\left(-\frac{\sin\underline{\theta}_m}{2}\right)\right.$$

$$\left.+\left(\sum_{j=1}^{N^2}h_{ij}^{m*}\frac{1+\cos\underline{\theta}_j}{2}\right)h_{ik}^{\mathbf{m}}\left(-\frac{\sin\underline{\theta}_m}{2}\right)\right].$$

Using Eqs. (25) and (28), the gradient above can be written as

$$\nabla F(\underline{\theta})=\underline{d}_{\underline{\theta}}=a\times\sin\underline{\theta}\odot\left\{\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}(H^{\mathbf{m}})^{*T}[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{1}-\underline{z})\odot(H^{\mathbf{m}})^*(\underline{m})]\right\}$$

$$+a\times\sin\underline{\theta}\odot\left\{\sum_{\mathbf{m}}\Gamma_{\mathbf{m}}(H^{\mathbf{m}})^{T}[(\underline{z}^*-\underline{z})\odot\underline{z}\odot(\underline{1}-\underline{z})\odot(H^{\mathbf{m}})(\underline{m})]\right\}.$$

APPENDIX B: DERIVATION OF GRADIENT $\nabla F(\underline{\theta})$ FOR THE AVERAGE COHERENT APPROXIMATION MODEL

The derivation of Eq. (36) from Eq. (35) is as follows:

$$\frac{\partial F(\underline{\theta})}{\partial \underline{\theta}_k} = 2\sum_{i=1}^{N^2}\left\{ z_i^* - \frac{1}{1+\exp[-a(|\Sigma_{j=1}^{N^2} h_{Cij}\underline{m}_j|^2 + \Sigma_{j=1}^{N^2}|h_{Iij}|^2|\underline{m}_j|^2) + at_r']}\right\} \frac{1}{\{1+\exp[-a(|\Sigma_{j=1}^{N^2} h_{Cij}\underline{m}_j|^2 + \Sigma_{j=1}^{N^2}|h_{Iij}|^2|\underline{m}_j|^2) + at_r']\}^2}$$
$$\times \exp\left[-a\left(\left|\sum_{j=1}^{N^2} h_{Cij}\underline{m}_j\right|^2 + \sum_{j=1}^{N^2}|h_{Iij}|^2|\underline{m}_j|^2\right) + at_r'\right] \times (-a)\left[\left(\sum_{j=1}^{N^2} 2h_{Cij}\frac{1+\cos\underline{\theta}_j}{2}\right) h_{Cik}\left(-\frac{\sin\underline{\theta}_m}{2}\right)\right.$$
$$\left. + 2\left(\frac{1+\cos\underline{\theta}_k}{2}\right) h_{Iik}^2\left(-\frac{\sin\underline{\theta}_m}{2}\right)\right].$$

Using Eqs. (28) and (34), the gradient above can be written as

$$\nabla F(\underline{\theta}) = \underline{d}_{\underline{\theta}} = 2a \times \sin\underline{\theta} \odot \{H_C^T[(\underline{z}^* - \underline{z})\odot\underline{z}\odot(\underline{\mathbf{1}} - \underline{z})\odot H_C(\underline{m})]\} + 2a \times \sin\underline{\theta}\odot\{H_I^{2T}[(\underline{z}^* - \underline{z})\odot\underline{z}\odot(\underline{\mathbf{1}} - \underline{z})\odot(\underline{m})]\}.$$

ACKNOWLEDGMENTS

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

REFERENCES

1. A. K. Wong, *Resolution Enhancement Techniques*, Vol. 1 (SPIE Press, 2001).
2. S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication*, 2nd ed. (Publishing House of Electronics Industry, 2003).
3. F. Schellenberg, "Resolution enhancement technology: The past, the present, and extensions for the future," 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, "Mask design for optical microlithography—An inverse imaging problem," IEEE Trans. Image Process. **16**, 774–788 (2007).
7. S. Sherif, B. Saleh, and R. Leone, "Binary image synthesis using mixed integer programming," IEEE Trans. Image Process. **4**, 1252–1257 (1995).
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. Erdmann, R. Farkas, T. Fuhner, B. Tollkuhn, and G. Kokai, "Towards automatic mask and source optimization for optical lithography," Proc. SPIE **5377**, 646–657 (2004).
11. L. Pang, Y. Liu, and D. Abrams, "Inverse lithography technology (ILT): What is the impact to the photomask industry?" Proc. SPIE **6283**, 62830X-1 (2006).
12. Y. Granik, "Illuminator optimization methods in microlithography," Proc. SPIE **5524**, 217–229 (2004).
13. Y. Granik, "Fast pixel-based mask optimization for inverse lithography," J. Microlithogr., Microfabr., Microsyst. **5**, 043002 (2006).
14. A. Poonawala and P. Milanfar, "Opc and psm design using inverse lithography: A non-linear optimization approach," Proc. SPIE **6154**, 1159–1172 (2006).
15. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," Proc. SPIE **6520**, 65200U (2007).
16. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," Opt. Express **15**, 15066–15079 (2007).
17. B. Salik, J. Rosen, and A. Yariv, "Average coherent approximation for partially coherent optical systems," J. Opt. Soc. Am. A **13**, 2086–2090 (1996).
18. A. K. Wong, *Optical Imaging in Projection Microlithography* (SPIE Press, 2005).
19. 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**, 2770–2777 (1982).
20. M. Born and E. Wolf, *Principles of Optics* (Cambridge U. Press, 1999).
21. R. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics* (Wiley, 1995).
22. N. Cobb and A. Zakhor, "Fast sparse aerial image calculation for opc," Proc. SPIE **2440**, 313–327 (1995).
23. C. Vogel, *Computational Methods for Inverse Problems* (SIAM Press, 2002).