Robust and efficient inverse mask synthesis with basis function representation

Xiaofei Wu,[1] Shiyuan Liu,[2] Wen Lv,[1,2] and Edmund Y. Lam[1,*]

[1]*Imaging Systems Laboratory, Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China*
[2]*State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China*
**Corresponding author: elam@eee.hku.hk*

Received July 7, 2014; accepted August 18, 2014;
posted September 2, 2014 (Doc. ID 216337); published September 26, 2014

Mask optimization is essential in the resolution scaling of optical lithography due to its strong ability to overcome the optical proximity effect. However, it often demands extensive computation in solving the nonlinear optimization problem with a large number of variables. In this paper, we use a set of basis functions to represent the mask patterns, and incorporate this representation into the mask optimization at both the nominal plane and various defocus conditions. The representation coefficients are updated according to the gradient to the coefficients, which can be easily obtained from the gradient to the pixel variables. To ease the computation of the gradient, we use an adaptive method that divides the optimization into two steps, in which a small number of kernels is used as the first step, and more kernels are used for fine optimization. Simulations performed on two test patterns demonstrate that this method can improve the optimization efficiency by several times, and the optimized patterns have better manufacturability compared with regular pixel-based representation.

OCIS codes: (110.5220) Photolithography; (110.1758) Computational imaging; (110.4235) Nanolithography.
http://dx.doi.org/10.1364/JOSAA.31.0000B1

1. INTRODUCTION

Optical lithography has been the main approach for large-scale integrated circuit manufacturing for several decades due to its low cost and high volume. The shrinking resolution of optical lithography relies heavily on resolution enhancement techniques (RETs), such as off-axis illumination, optical proximity correction, and source mask optimization [1–3]. Mask optimization is among the most important because it is able to overcome the optical proximity effects caused by diffraction.

Generally, current mask optimization methods can be categorized as polygon-based optical proximity correction (OPC) and the pixel-based inverse lithography technique (ILT) [4,5]. With the same objective to obtain a mask pattern that can generate a print pattern with high fidelity to the target pattern, they differ in the representation methods of the mask patterns. The former represents the mask pattern as polygons, and the optimization is carried out by moving the edges of the polygons. This technique is very successful in 65 nm technological node, but encounters problems of limited freedom in higher resolutions [6]. This leads to the introduction of inverse lithography, which represents the mask patterns as pixel-based images [7–9]. This representation method increases the flexibility and can further scale down the lithography resolution. However, the trade-off is the large amount of computation caused by the nonlinear optimization and the large number of pixel variables [10,11].

In this work, we focus mainly on pixel-based inverse lithography. The research on this technique dates back to the 1990s, when Liu and Zakhor conduct pioneering work on optimal binary and phase-shifting mask design [12,13]. This technique gained more interest a decade later. The inverse mask design problem was analyzed by Granik for incoherent, coherent, and partially coherent systems [14]. This problem was also approached through the level-set-based algorithm [8,15,16] and the genetic algorithm [17]. Meanwhile, Poonawala and Milanfar formulated this to an inverse imaging problem with regularization, and solved with the steepest descent algorithm [18]. Based on this framework, the gradient-based methods were extended to the conjugate gradient (CG) method with the objective of optimizing the phase-shifting mask patterns under partially coherent imaging systems [19,20]. Mask optimization was also conducted with the more rigorous vectorial model and with mask topography effects taken into consideration [21,22].

Extending the mask optimization to robust ILT, considering the process variations in the imaging system, can be even more challenging. The shrinking dimension of the mask patterns intensifies the necessity of considering multiple process conditions due to its higher sensitivity to process variations [23]. However, this consideration requires increasing modeling of the aerial image processes, and significantly reduces the computational efficiency. The mask optimization algorithm under process variations was developed by Yu *et al.*, and a quadratic model was employed to accelerate the aerial image calculation [24]. However, this method deals with the polygon-based OPC problem, and is not suitable for pixel-based inverse mask optimization, which requires computation of both aerial image and gradient for multiple process conditions. The pixel-based mask optimization under different

defocus conditions was formulated by Jia and co-workers and solved with statistical methods [25–27]. The algorithm is also extended to source mask optimization [28], and accelerated through several approaches, such as the augmented Lagrangian algorithm and the multigrid algorithm [29,30].

Nevertheless, these algorithms still need to handle all the pixel variables, which can be a large number in current mask optimization. These variables can cause stagnation in the inverse optimization, and slow down the convergence [31]. In this work, we reduce the number of variables in mask optimization by decomposing the mask patterns into a linear combination of a set of basis functions. This idea of using decomposition for acceleration has been used in many applications of optical lithography, such as polygon-based OPC [32,33], thick mask modeling [34], and source optimization [35]. The discrete cosine transform (DCT)-based method is also employed in the mask optimization to increase manufacturability [36,37].

In this paper, we develop a mask representation method with a set of basis functions, and incorporate this representation into robust ILT. Specifically, we use the 2D discrete cosine basis as the basis functions. The higher order coefficients are neglected to reduce the number of variables and avoid stagnation in the iterative optimization method. Instead of updating the pixel values in the mask patterns, we update the representation coefficients according to the gradient of the objective function to the coefficients. The gradient can be calculated with a simple DCT from the gradient of the objective function to the mask patterns. To ease computation, we use an adaptive method to calculate the gradient of the objective function to the mask patterns. This method divides the optimization into two steps according to the objective function, and changes the parameters in the optimization.

To evaluate the performance of the mask optimization method in this paper, we perform simulations on two typical mask patterns on the meta layer. Both the runtime and the aerial image quality are evaluated and compared with regular pixel-based optimization method under the nominal and various defocus conditions. The simulation results demonstrate that the proposed method can be several times faster in both cases. It is also observed that the optimized patterns with the basis representation are less complicated compared with the pixel-based representation.

2. ROBUST INVERSE LITHOGRAPHY FORMULATION

The objective of inverse lithography is to derive a mask that can print a pattern close to the target one through a forward model. In this technique, the forward model includes mainly two parts, namely, aerial imaging formation and the photoresist process. In inverse lithography, the sum of coherent system (SOCS) method is introduced to approximate the partially coherent imaging system with a number of coherent systems, namely, kernels, to accelerate the forward modeling [35]. Letting H_n be the nth kernel preserved for imaging modeling, and K is the number of kernels, the aerial image I_a above the wafer can be calculated through

$$I_a(\mathbf{x}) = \sum_{n=1}^{K} \lambda_n \left| H_n(\mathbf{x}) * M(\mathbf{x}) \right|^2, \tag{1}$$

where $\mathbf{x} = (x, y)$ is the spatial coordinate of the imaging system, $M(\mathbf{x}) \in \mathcal{R}^{N\times N}$ is the mask pattern, N is the number of pixels in each direction for 2D mask pattern representation, and * is the 2D convolution.

After this imaging formation, the aerial image undergoes a photoresist process to obtain the printed image as the wafer pattern. This process is usually modeled as a sigmoid function, and the printed pattern can be calculated as

$$I(\mathbf{x}) = \mathrm{sig}\{I_a(\mathbf{x})\} = \frac{1}{1 + \mathrm{e}^{-\alpha[I_a(\mathbf{x}) - t_r]}}, \tag{2}$$

where t_r is the threshold value for the photoresist process, and α is the steepness of the sigmoid function.

To derive an optimized mask pattern, we define an objective function to form an optimization problem. Usually, the objective function for inverse lithography contains the pattern fidelity term and the regularization term. The Euclidean distance between the printed pattern and the target pattern I_t is defined as the pattern fidelity term:

$$R_m\{M(\mathbf{x})\} = \sum_{\mathbf{x}} \|I(\mathbf{x}) - I_t(\mathbf{x})\|_2^2. \tag{3}$$

We also minimize the total variation of the mask patterns as a regularization term to improve the manufacturability as

$$R_{\mathrm{TV}}\{M(\mathbf{x})\} = \|Q_x M(\mathbf{x})\|_1 + \|Q_y M(\mathbf{x})\|_1, \tag{4}$$

where Q_x and Q_y are the differential matrix in the x and y directions, respectively.

To extend the mask optimization to robust ILT, the process variations need to be considered. We consider the defocus as the main process variation in this paper. We assume the defocus as normally distributed around the nominal plane, that is, the probability density function of the defocus is

$$p(h_m) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{h_m^2}{2\sigma^2}\right),$$

where h_m is the defocus, and σ is the standard deviation. For each defocus h_m, the pattern fidelity term can be expressed as

$$\tilde{R}_m\{M(\mathbf{x}), h_m\} = \sum_{\mathbf{x}} \|I(\mathbf{x}, h_m) - I_t(\mathbf{x})\|_2^2, \tag{5}$$

where $I(\mathbf{x}, h_m)$ denotes the printed pattern at the h_m defocus plane. Then, the expectation of the pattern errors over all the defocus conditions is considered as the pattern fidelity term, expressed as

$$E_m\{M(\mathbf{x}), h_m\} = \sum_m p(h_m) \tilde{R}_m\{M(\mathbf{x}), h_m\}. \tag{6}$$

The mask optimization at nominal plane can be considered a special case of the formulation above, in which $h_m = 0$, and $p(h_m) = 1$.

Then, the overall objective function for the mask optimization is the combination of the pattern fidelity term and the total variation term as

$$\mathcal{L}_m\{M(\mathbf{x}), h_m\} = E_m\{M(\mathbf{x}), h_m\} + \mu R_{\mathrm{TV}}\{M(\mathbf{x})\}, \tag{7}$$

where μ is the Lagrangian multiplier to define the aggressiveness of the regularization term.

The transmission for a binary mask is either 0 or 1. In the optimization, we relax them to gray-scale values in the region of [0,1] to facilitate the computation of the gradient. Thus, the optimized mask pattern $M_{\text{opt}}(\mathbf{x})$ is obtained by solving a combinational optimization problem as

$$M_{\text{opt}}(\mathbf{x}) = \arg \min_{0 \le M(\mathbf{x}) \le 1} \mathcal{L}_m\{M(\mathbf{x}), h_m\}. \tag{8}$$

The final optimized mask pattern is obtained by binarizing the mask with a threshold of 0.5.

3. INVERSE LITHOGRAPHY WITH BASIS REPRESENTATION

A. Mask Representation

Usually, the problem formulated above is solved through gradient-based algorithms. Here, we introduce the basis functions to represent the mask patterns during the optimization. The basis functions are a set of bases that can linearly represent a pattern. Let $\mathcal{D} = [D_1 \quad D_2 \quad \ldots]^{\mathsf{T}}$ to represent a particular set of bases, where D_l is the lth basis function. Then, the mask patterns can be expressed as a linear combination of these bases:

$$M(\mathbf{x}) = \sum_{l=1}^{P} \psi_l D_l(\mathbf{x}) = \mathcal{D}\Psi, \tag{9}$$

where $\Psi = [\psi_1 \quad \psi_2 \quad \ldots]^{\mathsf{T}}$ are the representation coefficients. The spatial coordinate $\mathbf{x}$ is omitted in the matrix form. It is also omitted in the following sections if there is no confusion.

In this paper, we use the 2D discrete cosine basis as the basis functions to represent the mask patterns. This is done for two reasons. First, the discrete cosine basis functions are suitable to represent the real numbers, and can be easily handled with DCT due to its popularity in image processing. Statistical analysis also demonstrates that the DCT coefficients are superior in energy compaction in representing 2D images [38,39]. Second, in optical lithography, the high frequency of the mask patterns is cut off due to diffraction. With off-axis illumination, the highest frequency that can be preserved in the mask patterns is $2\ \text{NA}/\lambda$, where λ is the wavelength, and NA is the numerical aperture of the imaging system [40]. Since the DCT is similar to the Fourier transform, the coefficients of the higher DCT frequencies can also be dropped during optimization. This will lead to fewer number of variables, and can be used to accelerate optimization.

For simplicity, we use the notation $\mathfrak{D}$ to represent the DCT transform, which transforms a pattern to its discrete cosine basis coefficients, and $\mathfrak{D}^{-1}$ for the inverse process. Thus, the basis representation and reconstruction of mask patterns can be expressed as

$$\Psi = \mathfrak{D}^{-1}\{M\} = \mathcal{D}^{-1}M, \tag{10}$$

where $\mathcal{D}$ is the set of discrete cosine basis functions. It is also known that for DCT bases, $\mathcal{D}^{-1} = \mathcal{D}^{\mathsf{T}}$, thus the above equation can be written as

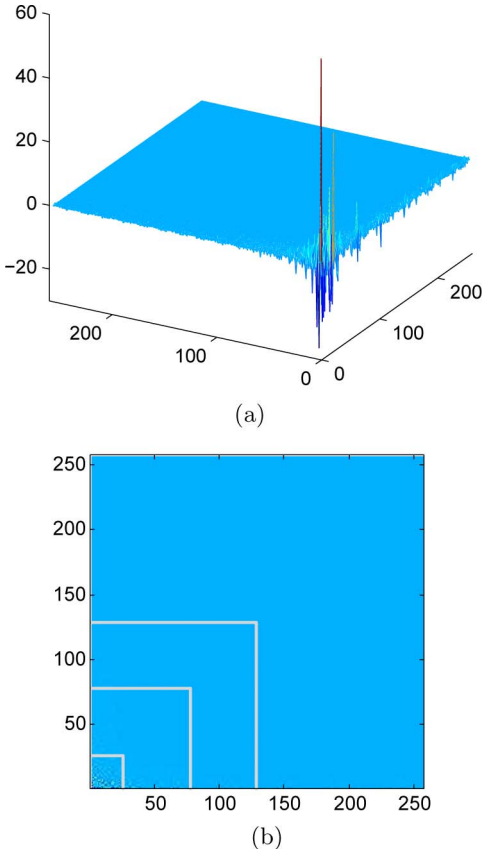

Fig. 1. DCT coefficients and the sampling scheme for a common mask pattern: (a) distribution of the DCT coefficients and (b) selection of the DCT coefficients.

$$\Psi = \mathcal{D}^{\mathsf{T}}M. \tag{11}$$

We depict the DCT coefficients for a common mask pattern and the sampling scheme in Fig. 1. It is obviously shown in Fig. 1(a) that the higher order coefficients are very small, and are almost flat at zero plane. Thus, we can select the lower order coefficients to represent the mask patterns. A further illustration of this point is shown in Fig. 2, where we depict the mask patterns represented by its partial DCT coefficients and the corresponding aerial images. The mask patterns are represented as 251×251 pixel images, and the DCT coefficients are contained in 251×251 matrices. By manually

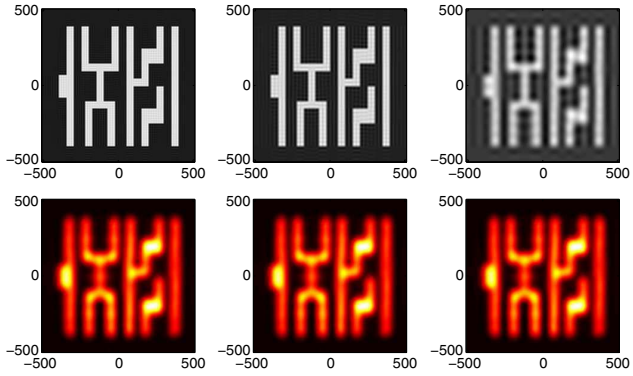

Fig. 2. Mask patterns and the corresponding aerial images represented with partial DCT coefficients. The unit of the spatial coordinates in the mask patterns and aerial images is nanometers.

setting the higher order coefficients to zeros, we reconstruct the mask patterns from the lower coefficients, and calculate the aerial images corresponding to these mask patterns. The mask patterns in the upper row are represented with the lower 125×125, 76×76, and 26×26 orders of the coefficients as shown in Fig. 1(b). We make comparison of the reconstructed mask patterns and aerial images with the original ones. The mean square errors of the aerial images are 9.97×10^{-7}, 7.12×10^{-6}, and 7.12×10^{-4}. It is shown that even if most of the coefficients are ignored, the aerial images calculated using the reconstructed mask patterns with partial coefficients can still be very small. Thus, it can be reasonable to drop the higher order coefficients in mask representation.

B. Optimization with Basis Representation

In many cases, the problem formulated in Eq. (8) can be solved with the gradient-based methods. Here, we show that the basis function representation can be incorporated into the CG method for mask optimization at the nominal plane. The method can be extended to robust ILT.

In the CG method, the mask pattern is updated as

$$M^{k+1} = M^k + \varepsilon_m \boldsymbol{v}_m^k, \tag{12}$$

where ε_m is the step size, and $\boldsymbol{v}_m^k$ is the searching direction, which can be calculated from the gradient of the objective function to the mask patterns $\nabla \mathcal{L}_m\{M\}$ [28,30].

With the basis function representation, we need to update the DCT coefficients instead of the mask patterns. Here, we show how the gradient of the DCT coefficients can be calculated from the gradient of the mask patterns. Since $M = \mathcal{D}\Psi$, the gradient of the objective function to the coefficients is

$$\nabla \mathcal{L}_m\{\Psi\} = \mathcal{D}^{\mathsf{T}} \nabla \mathcal{L}_m\{M\} = \mathfrak{D}^{-1}\{\nabla \mathcal{L}_m\{M\}\}, \tag{13}$$

and the DCT coefficients can be updated similarly:

$$\Psi^{k+1} = \Psi^k + \varepsilon_c \boldsymbol{v}_c^k, \tag{14}$$

where ε_c is the step size for updating the coefficients, and $\boldsymbol{v}_c^k = \mathfrak{D}^{-1}\{\boldsymbol{v}_m^k\}$. To drop the high-order coefficients, we force them to zeros:

$$\psi_l = \begin{cases} \psi_l, & \text{if } l \le C \\ 0, & \text{otherwise} \end{cases}, \tag{15}$$

where C denotes the highest frequency preserved for optimization.

Thus, we need to calculate the gradient of the objective function to the mask patterns to update the coefficients. The overall gradient is composed of two items, namely, the gradient of the pattern fidelity term, and the gradient of the regularization term, and can be expressed as

$$\nabla \mathcal{L}_m\{M\} = \nabla R_m\{M\} + \mu \nabla R_{\mathrm{TV}}\{M\}, \tag{16}$$

where ∇R_m is the gradient of the pattern fidelity term:

$$\nabla R_m\{M\} = \mathrm{Re}\left\{\sum_{\mathrm{n}=1}^{\mathrm{K}} \lambda_{\mathrm{n}} [2\alpha(\mathrm{I} - \mathrm{I_t}) \odot \mathrm{I} \odot (1 - \mathrm{I}) \odot (M * H_n)^{\dagger}] * H_n\right\}, \tag{17}$$

where $\dagger$ denotes the complex conjugate, and ∇R_{TV} is the gradient of the total regularization term:

$$\nabla R_{\mathrm{TV}}\{M\} = Q_x^{\mathsf{T}} \mathrm{sgn}(Q_x M) + Q_y^{\mathsf{T}} \mathrm{sgn}(Q_y M). \tag{18}$$

The detailed derivation of the these two terms can be found in [28,30].

Furthermore, to further ease the computational cost, we devise an adaptive method to calculate the gradient. The adaptive method can monitor the optimization process, and change the parameters in it. Here, we change the number of kernels used in calculating the gradient by monitoring the convergence behavior. As shown in Eq. (17), the computation amount of the gradient information for the pattern fidelity terms is proportional to the number of kernels. To reduce the computation, we can use fewer kernels for gradient computation, similar to the SOCS theory used for aerial image calculation. We divide the mask optimization into two steps. In the first step, we use a small number of kernels in the gradient computation, and the gradient is calculated by

$$\nabla R_m\{M\} = \mathrm{Re}\{\lambda_1 [2\alpha(\mathrm{I} - \mathrm{I_t}) \odot \mathrm{I} \odot (1 - \mathrm{I}) \odot (M * H_1)^{\dagger}] * H_1\}. \tag{19}$$

Then, for a fine optimization, we use more kernels to calculate the gradient in Eq. (17) in the second step.

In these two steps, the convergence criteria to terminate the iterations can be set according to the pattern errors. We terminate the iteration if the difference between several consecutive pattern errors is small. We also terminate the iteration if it reaches a pattern error that is smaller than a predefined one.

4. SIMULATIONS

In this section, we perform simulations to evaluate the efficiency and performance of the mask optimization method introduced above. The efficiency is evaluated in terms of runtime and convergence behavior, and the performance is evaluated by measuring the pattern fidelity and the process windows commonly used in other mask optimization methods. We make comparison of the proposed method with the regular CG method, which does not use the acceleration approach, to show the efficiency improvement.

Figure 3 shows the two test patterns used in the simulations. Both of the mask patterns are designed at the 45 nm technological node on the meta layer. The first pattern is a coarse pattern, which is represented by a 251×251 pixel image, and the size of each pixel is 4 nm. The second pattern is a dense pattern, which is represented by a 601×601 image, and the pixel size is 2.5 nm. The mask optimizations are performed in a partially coherent imaging system with the wavelength 193 nm, and NA = 1.35. The illumination source is fixed as a quasar one with $\sigma_{\mathrm{in}} = 0.68$, and $\sigma_{\mathrm{out}} = 0.92$. The Lagrange multiplier to define the aggressiveness of the total variation term is set as $\mu = 0.01$. The threshold for the sigmoid function in the resist model is set as $t_r = 0.3$, and the steepness is set as $\alpha = 85$. In the optimizations, the step size is set as 3 for the coarse pattern, and 4 for the dense pattern. The initial patterns in all the simulations are set as the corresponding target patterns to facilitate comparison.

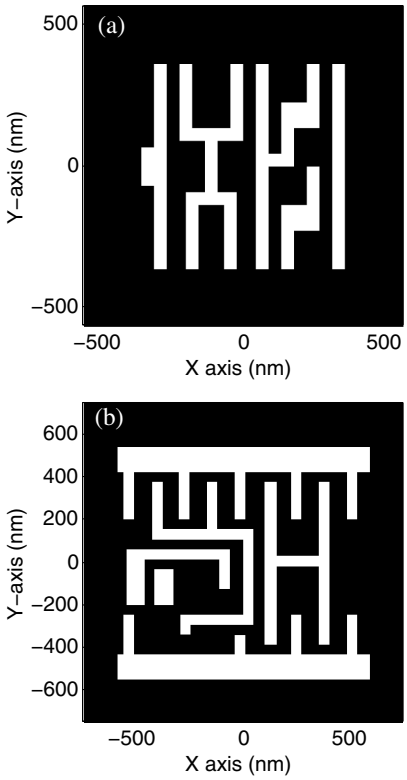

Fig. 3. Two test mask patterns for mask optimization, where (a) is the coarse test pattern and (b) is the dense one.

To show the performance of the proposed method, we carry out the mask optimization in three approaches: (a) the regular CG method, in which mask optimization is carried out through the regular CG method with a fixed number of kernels; (b) the adaptive CG method, in which the optimization is conducted with the regular conjugate method, but uses one kernel to obtain an initial pattern, and then uses more kernels to continue the optimization; and (c) the adaptive DCT CG method, in which the optimization is carried out similar to that in (b), while the mask pattern is represented as the combination of discrete cosine basis.

A. Optimization at the Nominal Plane

We first perform simulations of mask optimization at the nominal plane. The number of kernels used in the optimization for the first approach is seven. For the latter two approaches, we use one kernel to calculate the gradient at the first stage, and then use seven kernels for fine optimization. In mask optimization with DCT basis representation, the mask patterns are represented with the lower 76×76 DCT coefficients for the first mask pattern, and with the 180×180 ones for the latter. The optimization results of the optimized patterns and the printed patterns for the first test pattern are shown in Fig. 4. We measure the edge placement errors (EPEs) at the centers and a distance of 90 nm away from it at the edges. To analyze the mask optimization results, we depict the values of the pattern fidelity in each iteration in Fig. 5.

As shown in Fig. 5, the three approaches have similar performance in terms of pattern fidelity. All three methods can achieve an optimized pattern that can print the main features of the target pattern. In terms of computation efficiency, it shows that the adaptive DCT CG approach converges faster than the other two methods. Both optimization approaches using the regular pixel-based representation undergo a stagnation at the initial stage of the mask optimization. However, the objective function of mask optimization that uses the adaptive DCT CG approach decreases rapidly without any stagnation. Even in the latter iterations, the pattern error still reduces faster than those in the other two approaches. For the optimization without using basis representation, the adaptive CG method converges a little faster than the regular CG method. More importantly, most of the iterations in the adaptive approaches are completed with only one kernel. It means the adaptive CG method can produce the optimized mask pattern much faster than the regular CG method.

Quantitatively, the average EPE for the printed pattern for the three methods are 2.50, 2.43, and 2.60 nm, respectively, which are very similar. The total numbers of iterations for

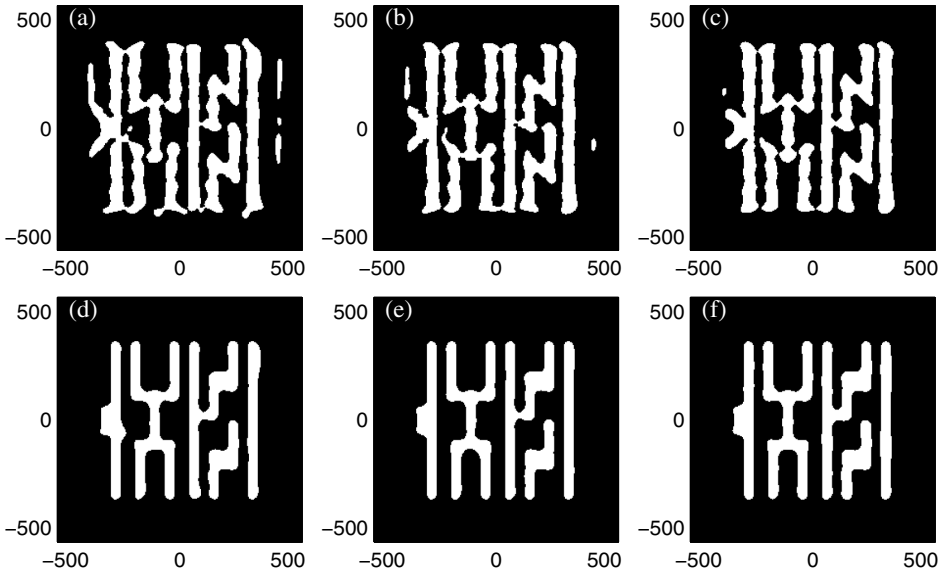

Fig. 4. Simulation results on the coarse pattern, where (a), (b), and (c) are the optimized mask patterns, and (d), (e), and (f) are the printed patterns for the regular CG, adaptive CG, and adaptive DCT CG approaches, respectively.

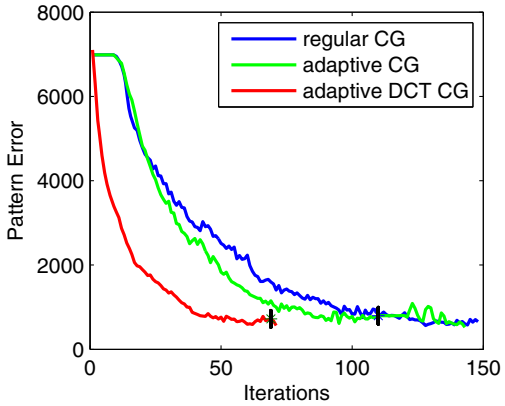

Fig. 5. Pattern errors in the iterations for the three approaches.

the three methods are 148, 143, and 71. Among them, 110 and 69 iterations are conducted with one kernel. The total runtimes are 29.6, 18.7, and 8.6 s. This means that the adaptive DCT CG approach converges with significant fewer iterations compared with the regular CG method. It can increase the speed by several times, while the basis representation can be 2 times faster.

Similarly, the optimization results for the second test pattern are shown in Fig. 6. Compared with the first test pattern, this pattern has more features. Thus, the optimized mask patterns shown in the figure are also more complicated than those for the first test pattern. This illustrates that the optimized pattern generated with the basis representation has more smooth edges, and fewer isolated points and holes. This is consistent with the fact that the high frequency of the mask patterns is truncated in the optimization; thus, the mask pattern has better manufacturability.

In terms of optimization performance, the average EPEs for the three approaches are 1.74, 1.70, and 1.73 nm, respectively. The optimization results are obtained through 242, 203, and 140 iterations, as shown in Fig. 7. Again, in the latter two adaptive approaches, 202 and 139 iterations are completed using one kernel. The runtimes for these three methods are 224.7, 116.5, and 104.8 s. This again demonstrates that the adaptive DCT CG approach can reduce the number of iterations significantly. At the same time, it can increase the speed of mask optimization by about 2 times, while the mask optimization with basis representation can further reduce the runtime by 11.3%.

B. Optimization at Various Defocus Planes

We next perform simulations of mask optimization considering the defocus variations in the same imaging system. Again, we conduct the simulations with the three approaches introduced above. Since the influence of positive defocus on the image quality is similar to that of negative defocus, we consider only positive defocus in the optimization. Thus, we set the defocus in [0,100] nm with a step size of 20 nm, which means the mask optimizations are performed at six defocus planes. The standard deviation is set as $\sigma = 0.1$ for normal distribution. The target patterns are the same as those in the previous simulations.

We depict the optimized patterns for both test patterns in Fig. 8, where the upper panels are for the coarse test pattern, and the lower panels are for the dense mask pattern. Compared with the optimized mask patterns shown in Figs. 4 and 6, we can see in robust mask optimization, much stronger assistant features are generated. For the coarse pattern, long-line assistant features are placed around the initial pattern. For the dense pattern, many more features are generated compared with the optimization patterns at the nominal plane shown in Fig. 6. Again, the optimized mask patterns have similar features in the three approaches for each test pattern.

We also evaluate the performance in terms of image quality and efficiency for the three approaches. To evaluate the image quality, we measure the exposure–defocus (E-D) window at critical places, such as the center of each line, and a distance of 300 nm from the center for both patterns. The range of the defocus is set from −60 to 60 nm for all the measurements.

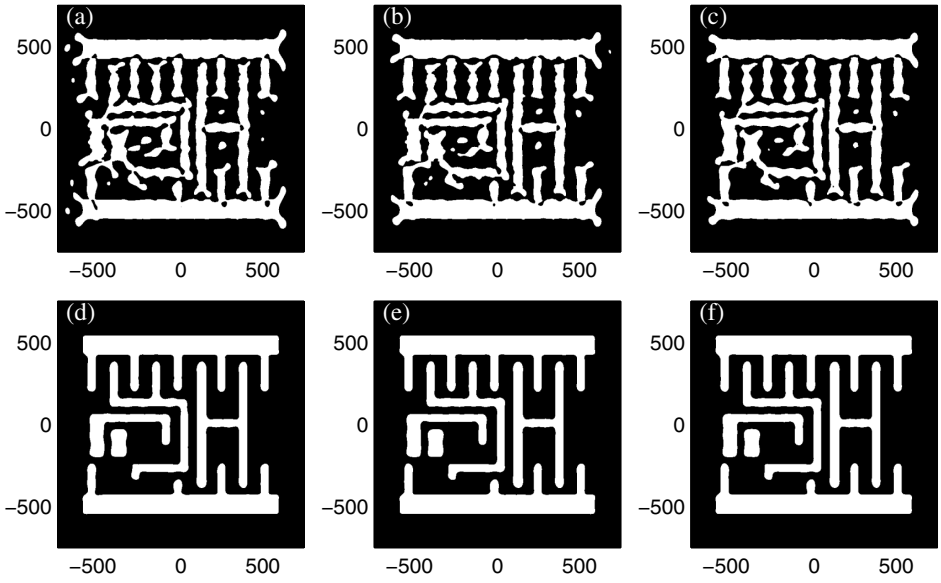

Fig. 6. Simulation results for the dense pattern. The optimized mask patterns are shown in (a), (b), and (c), and the printed patterns at the nominal plane are shown in (d), (e), and (f) for the regular CG, adaptive CG, and adaptive DCT CG approaches, respectively.

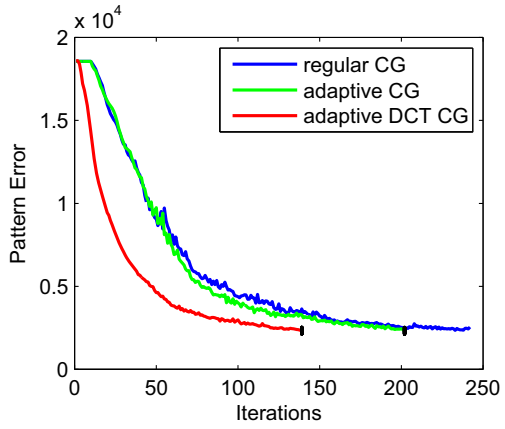

Fig. 7. Pattern errors in the iterations for the dense pattern.

Each E-D window is represented with two curves, with the upper one corresponding to the doses when the critical dimension is 10% smaller than 45 nm, and the lower one is 10% larger than it. To evaluate efficiency, we assess the convergence behavior in the iterations and the runtime to similarly carry out optimizations.

Figure 9 depicts the E-D windows measured for the three mask patterns obtained with the three approaches; Fig. 9(a) is for the coarse pattern, and Fig. 9(b) is for the dense pattern. It is observed that the E-D windows are quite close to each other for both test patterns. The E-D window for the adaptive DCT CG approach is a little larger than the other two approaches in the coarse pattern, with the lowest curve. This indicates the three approaches have similar ability to optimize the mask when taking robustness to defocus into consideration. By making comparison of the two E-D windows for the two test patterns, it is obvious that the one for the dense mask pattern is larger than that for the coarse pattern. This may be caused by the smaller pixel size in the dense pattern. More pixel variables lead to larger freedom, and can have stronger ability to overcome the optical proximity effect.

In terms of efficiency, the total number of iterations for the three approaches are 177, 188, and 145 for the coarse pattern, in which 151 and 141 iterations are conducted with one kernel

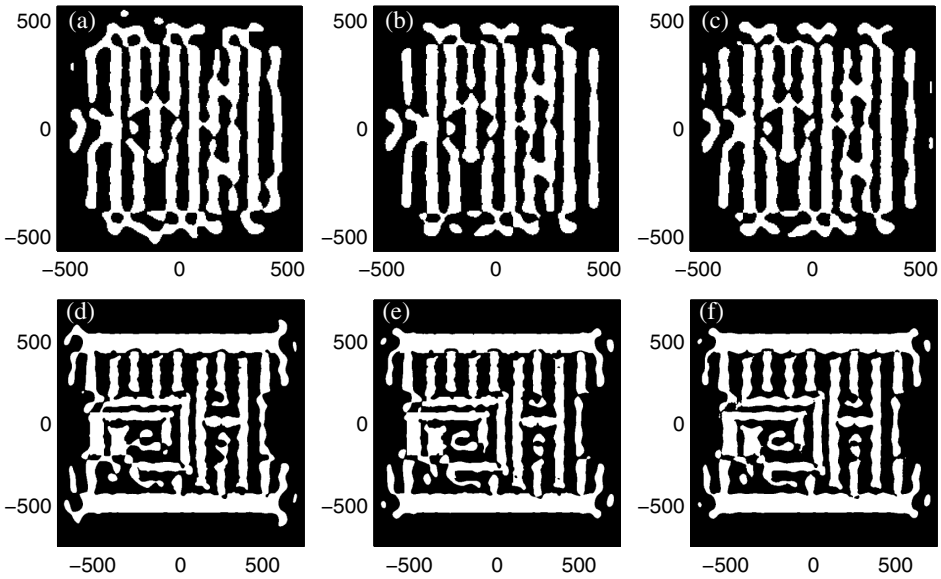

Fig. 8. Optimized mask patterns at various defocus conditions, where (a), (b), and (c) are for the coarse pattern, and (d), (e), and (f) are for the dense pattern.

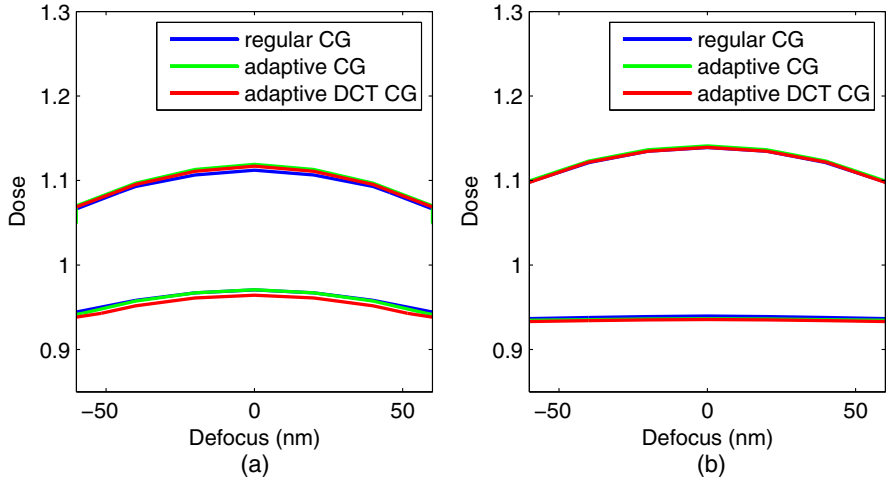

Fig. 9. Process windows for the optimized mask patterns.

for the latter two approaches. Again, the mask optimization with the DCT representation again converges faster than those represented with pixel variables. The optimization with basis representation can reduce the iteration number by 22.8% compared with the adaptive CG approach. The runtimes for the three approaches are 147.4, 109.5, and 78.5 s, respectively. The adaptive CG method can improve the speed by 34.6%, and the adaptive DCT CG approach can make a further 39.5% improvement.

For the dense pattern, the numbers of iterations for the three approaches are 237, 230, and 143, and the runtimes are 1032, 553.7, and 316.4 s, respectively. The reduction of the iteration number for the adaptive DCT CG approach is 37.8% over the adaptive CG. The speed improvement for the adaptive CG approach is about 1 time over the regular CG method, and the adaptive DCT CG can make a further 75% improvement. This again demonstrates that the mask representation with DCT basis together with the adaptive method can be useful in improving the efficiency of mask optimization at various defocus conditions. These simulations indicate that mask optimization with basis representation can have a competitive ability to optimize the mask patterns, while improving efficiency.

5. CONCLUSIONS

In this paper, we propose a mask optimization algorithm with basis function representation, and use an adaptive optimization method to accelerate the algorithms. We use the 2D discrete cosine basis functions to represent the mask patterns, and incorporate this representation into the CG algorithm for optimization. We also employ an adaptive method which uses a small number of kernels to get initial patterns, and then use more kernels for fine optimization to accelerate the the mask optimization. Simulations performed on two test patterns at both the nominal plane and various defocus planes for robust ILT demonstrate that the proposed method can improve the efficiency by approximately 2 times, while achieving optimized mask patterns with competitive performance compared with the regular pixel-based method.

ACKNOWLEDGMENTS

This work was supported in part by the UGC Areas of Excellence project Theory, Modeling, and Simulation of Emerging Electronics; and by the State Key Laboratory of Digital Manufacturing Equipment and Technology under Project DMETKF2013003.

REFERENCES

1. A. K. Wong, *Resolution Enhancement Technologies in Optical Lithography* (SPIE, 2001).
2. F. M. Schellenberg, "Resolution enhancement technology: the past, the present, and extensions for the future," Proc. SPIE **5377**, 1–20 (2004).
3. D. O. S. Melville, A. E. Rosenbluth, A. Waechter, M. Millstone, J. Tirapu-Azpiroz, K. Tian, K. Lai, T. Inoue, M. Sakamoto, K. Adam, and A. Tritchkov, "Computational lithography: exhausting the resolution limits of 193-nm projection lithography systems," J. Vac. Sci. Technol. B **29**, 06FH04 (2011).
4. Y. Granik, "Solving inverse problems of optical microlithography," Proc. SPIE **5754**, 506–526 (2005).
5. N. B. Cobb and A. Zakhor, "Fast, low-complexity mask design," Proc. SPIE **2440**, 313–327 (1995).
6. N. B. Cobb and Y. Granik, "New concepts in OPC," Proc. SPIE **5377**, 680–690 (2004).
7. D. S. Abrams and L. Pang, "Fast inverse lithography technology," Proc. SPIE **6154**, 61541J (2006).
8. L. Pang, Y. Liu, and D. Abrams, "Inverse lithography technology (ILT): a natural solution for model-based SRAF at 45 nm and 32 nm," Proc. SPIE **6607**, 660739 (2007).
9. M. L. Kempsell, E. Hendrickx, A. Tritchkov, K. Sakajiri, K. Yasui, S. Yoshitake, Y. Granik, G. Vandenberghe, and B. W. Smith, "Inverse lithography for 45-nm-node contact holes at 1.35 numerical aperture," J. Micro/Nanolith. MEMS MOEMS **8**, 043001 (2009).
10. E. Y. Lam and A. K. Wong, "Computation lithography: virtual reality and virtual virtuality," Opt. Express **17**, 12259–12268 (2009).
11. S. Liu, X. Zhou, W. Lv, S. Xu, and H. Wei, "Convolution-variation separation method for efficient modeling of optical lithography," Opt. Lett. **38**, 2168–2170 (2013).
12. Y. Liu and A. Zakhor, "Optimal binary image design for optical lithography," Proc. SPIE **1264**, 401–412 (1990).
13. Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," IEEE Trans. Semicond. Manuf. **5**, 138–152 (1992).
14. Y. Granik, "Fast pixel-based mask optimization for inverse lithography," J. Micro/Nanolith. MEMS MOEMS **5**, 043002 (2006).
15. Y. Shen, N. Wong, and E. Y. Lam, "Level-set-based inverse lithography for photomask synthesis," Opt. Express **17**, 23690–23701 (2009).
16. W. Lv, S. Liu, Q. Xia, X. Wu, Y. Shen, and E. Y. Lam, "Level-set-based inverse lithography for mask synthesis using the conjugate gradient and an optimal time step," J. Vac. Sci. Technol. B **31**, 041605 (2013).
17. T. Fühner, A. Erdmann, and S. Seifert, "Direct optimization approach for lithographic process conditions," J. Micro/Nanolith. MEMS MOEMS **6**, 031006 (2007).
18. A. Poonawala and P. Milanfar, "Mask design for optical microlithography—An inverse imaging problem," IEEE Trans. Image Process. **16**, 774–788 (2007).
19. J.-C. Yu and P. Yu, "Impacts of cost functions on inverse lithography patterning," Opt. Express **18**, 23331–23342 (2010).
20. Y. Peng, J. Zhang, Y. Wang, and Z. Yu, "Gradient-based source and mask optimization in optical lithography," IEEE Trans. Image Process. **20**, 2856–2864 (2011).
21. X. Ma, Y. Li, and L. Dong, "Mask optimization approaches in optical lithography based on a vector imaging model," J. Opt. Soc. Am. A **29**, 1300–1312 (2012).
22. J. Li and E. Y. Lam, "Robust source and mask optimization compensating for mask topography effects in computational lithography," Opt. Express **22**, 9471–9485 (2014).
23. N. B. Cobb and Y. Granik, "OPC methods to improve image slope and process window," Proc. SPIE **5042**, 116–125 (2003).
24. P. Yu, S. X. Shi, and D. Z. Pan, "True process variation aware optical proximity correction with variational lithography modeling and model calibration," J. Micro/Nanolith. MEMS MOEMS **6**, 031004 (2007).
25. N. Jia, A. K. Wong, and E. Y. Lam, "Robust mask design with defocus variation using inverse synthesis," Proc. SPIE **7140**, 71401W (2008).
26. N. Jia and E. Y. Lam, "Machine learning for inverse lithography: using stochastic gradient descent for robust photomask synthesis," J. Opt. **12**, 045601 (2010).
27. Y. Shen, N. Jia, N. Wong, and E. Y. Lam, "Robust level-set-based inverse lithography," Opt. Express **19**, 5511–5521 (2011).
28. N. Jia and E. Y. Lam, "Pixelated source mask optimization for process robustness in optical lithography," Opt. Express **19**, 19384–19398 (2011).
29. J. Li, S. Liu, and E. Y. Lam, "Efficient source and mask optimization with augmented Lagrangian methods in optical lithography," Opt. Express **21**, 8076–8090 (2013).
30. W. Lv, E. Y. Lam, H. Wei, and S. Liu, "Cascadic multigrid algorithm for robust inverse mask synthesis in optical lithography," J. Micro/Nanolith. MEMS MOEMS **13**, 023003 (2014).
31. X. Ma and G. R. Arce, "Pixel-based OPC optimization based on conjugate gradients," Opt. Express **19**, 2165–2180 (2011).

32. Y. C. Pati, A. A. Ghazanfarian, and R. F. Pease, "Exploiting structure in fast aerial image computation for integrated circuit patterns," IEEE Trans. Semicond. Manuf. **10**, 62–74 (1997).
33. S. Liu, X. Wu, W. Liu, and C. Zhang, "Fast aerial image simulations using one basis mask pattern for optical proximity correction," J. Vac. Sci. Technol. B **29**, 06FH03 (2011).
34. K. Adam and A. R. Neureuther, "Domain decomposition methods for the rapid electromagnetic simulation of photomask scattering," J. Microlith. Microfab. Microsyst. **1**, 253–269 (2002).
35. X. Wu, S. Liu, J. Li, and E. Y. Lam, "Efficient source mask optimization with Zernike polynomial functions for source representation," Opt. Express **22**, 3924–3937 (2014).
36. S. Shen, P. Yu, and D. Z. Pan, "Enhanced DCT2-based inverse mask synthesis with initial SRAF insertion," Proc. SPIE **7122**, 712241 (2008).
37. P. Gao, A. Gu, and A. Zakhor, "Optical proximity correction with principal component regression," Proc. SPIE **6924**, 69243N (2008).
38. E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," IEEE Trans. Image Process. **9**, 1661–1666 (2000).
39. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, 2003).
40. S. R. J. Brueck and X. Chen, "Spatial frequency analysis of optical lithography resolution enhancement techniques," J. Vac. Sci. Technol. B **17**, 908–920 (1999).