

# DIRECT: An Efficient Optimization Scheme for Mask Generation Using Inverse Lithography

Wei Xiong<sup>\*</sup>, Min-Chun Tsai<sup>\*\*</sup>, Jinyu Zhang<sup>\*</sup> and Zhiping Yu<sup>\*</sup>

<sup>\*</sup>Institute of Microelectronics, Tsinghua University, Beijing, 100084, China  
xiongwei00@mails.tsinghua.edu.cn

<sup>\*\*</sup>Advanced Technology Group, Synopsys Inc., Mountain View, CA, 94043, USA  
min-chun.tsai@synopsys.com

## ABSTRACT

Resolution Enhancement Technologies (RETs) are widely used to cope with the severe optical effects that are manifests in sub-wavelength lithography. Inverse Lithography Technique (ILT) has recently been proposed as an effective RET for sub-wavelength technology. ILT increases the degree-of-freedom in mask data manipulation, and allows automatic correction to 2D pattern distortion. In this work, a realistic aerial image model with an efficient optimization scheme is developed to pattern metal layers for the 65nm technology node. Simulation results show that the optimized masks provide good fidelity in patterning. We called our method DIScrete RETiCle Technique, DIRECT.

**Keywords:** ilt, ret, mask, simulated annealing, psm

## 1 INTRODUCTION

As semiconductor manufacturing reaches 90nm and 65nm technology nodes, and is moving towards 45nm, 32nm, and below, one of the greatest challenges is in lithography. The mostly used exposure steppers use 193nm wavelength, however the critical dimension (CD) on wafer keeps shrunk, approaching a quarter of the illumination wavelength. Under such circumstances, due to the optical diffraction and interference, serious distortions occurred when translating the patterns on the mask to the wafer, causing the failure of printed circuits. The pattern distortion is generally called optical proximity effect (OPE) [1], [2]. In order to compensate OPE, people have proposed a series of techniques to improve the litho-system's resolution. The resolution is determined by the Rayleigh's criterion,

$$R = k\lambda / NA \quad (1)$$

where  $\lambda$  is the illumination wavelength,  $NA$  is the numerical aperture of the imaging lenses, and  $k$  is the process constant affected by the process conditions. Using a new exposure source may decrease the wavelength, and putting the imaging lenses into water or other liquid with a high refraction index can increase the numerical aperture. These two approaches can both improve the resolution effectively,

but they require changing the existing lithography infrastructure and have high costs. An alternative approach is using light's wave nature to decrease the process constant  $k$ . This method is called RETs (Resolution Enhancement Techniques). Based on exploring the optical wave's amplitude, phase and direction, RETs are accordingly classified as optical proximity correction (OPC), phase-shifting masks (PSM) and off-axis illumination (OAI) [3]. These methods have extended the lifespan of current optical projection systems. However, finding an optimum mask becomes increasingly complex, as less intuitive relationships between optimum masks and resulted images apply. More robust optimization techniques need to be developed to improve the image formation.

Back in the early 1980s, B. E. A. Saleh and S. Sayegh [4] considered the design of masks as an inverse problem, and proposed a rigorous mathematical approach to solve the inverse problem and find the optimal mask for a given process. In their work, mask is discretized to pixels, and then values 0 and 1 are tried. If the image fidelity improves, the pixel value is accepted, otherwise it is rejected, and the next pixel is then tried, and so on. Later, Liu [5] and Sherif [6] applied simulated annealing (SA) algorithm and branch-bound (BB) algorithm to the mask optimization, respectively. But the early SA algorithm is inadequate in the convergency efficiency, which leaves room for improving this algorithm, while BB algorithm would encounter difficulties when the pixels have more than two values. Recently, Granik [7] has considered the inverse problem as a non-linear programming problem, and demonstrated it is a possible solution method.

In this article, we discuss our optimization method DIRECT, which is composed of two stages: modeling of the imaging system and the synthesis of masks. A realistic lithography system can be approximated by a partially coherent model, which is based on famous Hopkins' Equation [8]. We treat the partially coherent system as sum of coherent systems with different weight. SA algorithm is applied to optimize masks, and an efficient optimization scheme is developed to accelerate the convergency process and to reduce the calculation time.

The optical imaging model is formulated in Section 2. The optimization scheme for SA algorithm is explained in

Section 3. Some simulation results are shown in Section 4. Finally, we provide conclusive remarks in Section 5.

## 2 OPTICAL IMAGING MODEL

The imaging mechanism of a stepper can be modeled by Hopkins equation [8]:

$$I(f, g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T(f'+f, g'+g, f', g') \cdot M(f'+f, g'+g) M^*(f', g') df' dg' \quad (2)$$

where  $I(f, g)$  is the forward Fourier transform of the output image intensity  $i(x, y)$ ,  $M(f, g)$  is the forward Fourier transform of mask transmission function  $m(x, y)$ , and  $T(f, g, f', g')$  is the Transmission Cross-Coefficient (TCC) of the optical system, which characterize all the features about the imaging system and illumination.

In our model, 2D mask pattern is quantized into small square pixels, as shown in Fig. 2(a), with  $a_{ik}$  representing the transmission variable at the  $(i, k)$ -th pixel, then the mask transmission function can be expressed as:

$$m(x, y) = \sum_{i=1}^{N1} \sum_{k=1}^{N2} a_{ik} \psi_{ik}(x, y) \quad (3)$$

where  $N1, N2$  are the numbers of pixels in each dimension and  $\psi_{ik}$  is a unit square pulse function located at the  $(i, k)$ -th pixel. For different kinds of masks,  $a_{ik}$  may have different values. It takes values  $\{0, 1\}$  for binary mask,  $\{-0.245, 1\}$  for EPSM and  $\{1, 0, -1\}$  for APSM, the minus sign means that the light's phase has changed 180 degrees. EPSM and APSM are two types of special phase-shifting masks for 65 CMOS technology node.

It is difficult and time-consuming to calculate the output image intensity using Eq. (2) directly. We need to decompose the TCC function as follows [8], [9]:

$$T(f'+f, g'+g, f', g') \approx \sum_{l=1}^M \sigma_l \Phi_l(f'+f, g'+g) \Phi_l^*(f', g') \quad (4)$$

This transformation is based on the singular value decomposition (SVD) of matrices and  $\{\sigma_l\}$  is the singular value set of TCC matrix,  $\{\Phi_l\}$  is the kernel set corresponding to the singular value set, and  $M$  is the order of decomposition. These singular values and kernels are only determined by the characteristic of the imaging system. Substituting Eq. (4) in Eq. (2) and taking inverse Fourier transform, we obtain the following expression for the output image intensity:

$$i(x, y) = \mathcal{F}^{-1}[I(f, g)] = \sum_{l=1}^M \sigma_l |(\phi_l \otimes m)(x, y)|^2 \quad (5)$$

where  $\{\phi_l\}$  is the inverse Fourier transform of  $\{\Phi_l\}$ , and notation  $\otimes$  represents the convolution operator.  $\{\phi_l\}$  and  $\{\sigma_l\}$  are obtained before the synthesis of masks, and we can use Eq. (5) to calculate the output image intensity distribution. Eq. (5) also has an obvious physical meaning that the original partially coherent system can be considered as sum of several weighted coherent systems, as shown in Fig. 1.

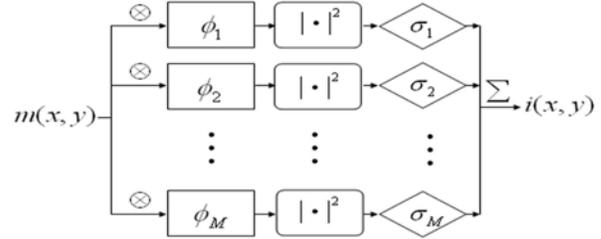


Fig. 1: Approximation of the partially coherent system.

## 3 SA ALGORITHM

Simulated Annealing is a global optimization algorithm to solve complex combinatorial problems [10], [11]. A brief description of the basic algorithm is provided below, and the emphasis is on the way we apply it to the mask design problem.

Assume different patterns of the mask to be the state of a statistical system, denoted by a random vector  $X$ . If we introduce a control temperature  $T$  and the energy function or error function corresponding to a state  $X$  as  $H(X)$ , then according to Boltzmann distribution the probability of being in state  $X$  at thermal equilibrium is:

$$P(\mathbf{X}) = \frac{1}{Z(T)} \exp\{-[H(\mathbf{X})/T]\} \quad (6)$$

where  $Z$  is a normalization constant, or the partition function in statistical physics. At high temperature, this distribution is almost uniform, and the system is equally likely to be in any state. Then we gradually decrease  $T$ , allowing the system to reach thermal equilibrium at each  $T$ . As the temperature decreases, the Boltzmann distribution concentrates more on the states with low energy. Finally, when  $T$  approaches zero, the system will reach a ground state, namely the optimum solution of mask design problem.

In computer simulations, the system reaches the thermal equilibrium at each  $T$  by randomly choosing a pixel and by flipping it to the other state, continuously. For binary mask, assuming the  $(i, k)$ -th pixel to be selected, if we denote the state before its flipping as  $X_m$ , and the state after the flipping as  $X_n$ , and the corresponding energy functions by  $H(X_m)$  and  $H(X_n)$ , then the probability for state transition of  $X_m \rightarrow X_n$  is:

$$P(X_m \rightarrow X_n) = \begin{cases} 1 & H(X_n) - H(X_m) < 0 \\ \exp\left(\frac{-[H(X_n) - H(X_m)]}{T}\right) & H(X_n) - H(X_m) \geq 0 \end{cases} \quad (7)$$

This flipping is called an intent transition. If this flipping is accepted, we call it a success transition and  $a_{ik}$  changes value, otherwise we call it a failure transition and  $a_{ik}$  remains unchanged. For PSM with 3 values which each pixel can take, we still randomly choose a pixel, assuming  $(i, k)$ -th pixel to be selected. If these three different states corresponding to the pixel values are  $X^1$ ,  $X^0$  and  $X^{-1}$ , respectively, and corresponding energy functions are  $H(X^1)$ ,  $H(X^0)$  and  $H(X^{-1})$ , then we calculate  $\Delta H_{1 \rightarrow 0} = H(X^0) - H(X^1)$ ,  $\Delta H_{1 \rightarrow -1} = H(X^{-1}) - H(X^1)$  and  $\Delta H_{1 \rightarrow 1} = 0$ . So the probability for state transition of  $X^1 \rightarrow X^0$  is:

$$P(X^1 \rightarrow X^0) = \frac{\exp(-\Delta H_{1 \rightarrow 0}/T)}{\exp(-\Delta H_{1 \rightarrow 0}/T) + \exp(-\Delta H_{1 \rightarrow -1}/T) + 1} \quad (8)$$

Also we can obtain the probability for state transition of  $X^1 \rightarrow X^{-1}$  and  $X^1 \rightarrow X^1$ . We notice that the probability of making a transition from state  $X_m$  to  $X_n$  does not depend on the history of how state  $X_m$  was reached, hence the generated sequence  $\{X_m\}$  for each  $T$  is a Markov chain [12].

When applying SA to the practical issues, a cooling schedule should be developed to maximize the performance of this algorithm. We briefly describe our cooling schedule as follows:

*Initial Value of Control Temperature:* The initial value should be high enough to make all transitions be accepted almost equally, but not be too high to save computing time. So we find a suitable initial value as follows:

1. Define variable  $T$  and initial temperature  $t_0$  with any given value;  $\chi = 0.9$ ,  $R_0 = 0$ ; iteration order  $k = 1$ .
2. Try  $L$  intent transitions ( $L$  is a predefined number), and compute the acceptance ratio  $R_k$  which is the number of success transitions divided by the number of intent transitions.
3. If  $|R_k - \chi| < \varepsilon$  ( $\varepsilon$  is a given small value), terminate the iteration. Otherwise, if  $R_{k-1}$  and  $R_k < \chi$ , then  $k = k + 1$ ,  $t_k = t_{k-1} + T$ , return to step 2; if  $R_{k-1}$  and  $R_k \geq \chi$ , then  $k = k + 1$ ,  $t_k = t_{k-1} - T$ , return to step 2; if  $R_{k-1} \geq \chi$  and  $R_k \leq \chi$ , then  $k = k + 1$ ,  $t_k = t_{k-1} + T/2$ ,  $T = T/2$ , return to step 2; if  $R_{k-1} \leq \chi$  and  $R_k \geq \chi$ , then  $k = k + 1$ ,  $t_k = t_{k-1} - T/2$ , return to step 2.

Finally we obtain the suitable initial temperature value.

*Decreasing the Control Temperature:* In order to achieve the final ground state, the control temperature is supposed to be reduced slowly. But the slower the temperature reduces, the lower the convergence efficiency is. So we use

$$T_m = \alpha T_{m-1}, \quad \alpha = 0.8; \quad m = 1, 2, \dots \quad (9)$$

as our decreasing rule.

*Length of Markov Chains:* The Markov chain is terminated when the system reaches the thermal equilibrium. Our termination rule is: define  $U_{\text{int}}$  and  $L_{\text{int}}$  ( $U_{\text{int}} > L_{\text{int}}$ ) as the ceiling and limit of the number of the intent transitions. At each temperature, the number of intent transitions tried must be larger than  $L_{\text{int}}$ . If the acceptance ratio is over certain predefined value, the Markov chain is terminated. Meanwhile, we record the state with a minimum energy function in this Markov chain, and use it as the start of next Markov chain.

*Termination of the Simulated Annealing:* A lot of termination criteria can be used. For example, the energy function declines to certain value; the temperature reaches certain value; the number of intent transition tried exceeds the limit. These criterions can also be used in a combinatorial form.

## 4 RESULTS AND ANALYSIS

In the following simulations, three different types of masks for 65nm CMOS technology node are used to validate the optimization method described above. The common parameters for the image model are as follows: for the annular illumination, the wavelength is 193nm, and the outer coherence is 0.8, the inner coherence is 0.56 [13]; the numerical aperture is 0.8; the threshold of the resist is 0.247; the pixel size is  $25 \times 25 \text{nm}^2$ ; the feature size is 100nm.

Fig. 2 shows the distortion occurred in the lithography due to the strong OPE. The regular binary mask to be optimized is plotted in Fig. 2(a). The white pixels represent the transparent area with a transmission value of 1, while the gray pixels represent the opaque area with a transmission value of 0. In Fig. 2(b), the black bold line represents the desired image, and the gray line shows the output image corresponding to the mask shown in Fig. 2(a).

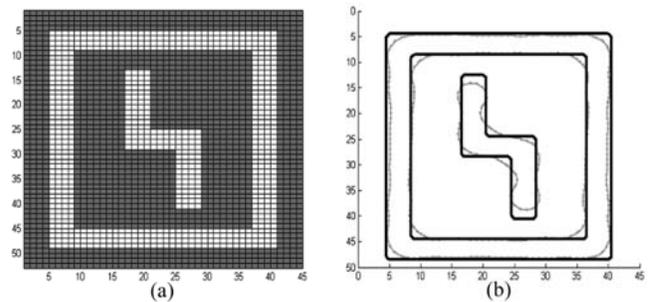


Fig. 2: (a) The regular mask to be optimized; (b) Output image due to the input mask shown in (a).

From Fig. 2(b), an obvious distortion occurred between the desired image and the output image. The effects of line-end shortening and corner rounding are serious. Then we use our optimization method to improve the image performance. The simulation results are illustrated in Fig. 3 and Fig.

4. The convergence process is shown in Fig. 3, which explains the relationship between the image performance and the number of intent transitions. The vertical axis represents the deviation rate, which means the area of where the output image mismatches the desired image divided by the area of the desired image. With the number of intent transitions increasing, the deviation rate decreases, thus the image performance is improved. Noticed that the deviation rate is basically unchanged when the number of intent transitions is larger than 4500. The optimum mask is thus obtained and is shown in Fig. 4(a), and its corresponding output image is displayed in Fig. 4(b). Comparing Fig. 4(b) with Fig. 2(b), we can find that the image fidelity improves greatly.

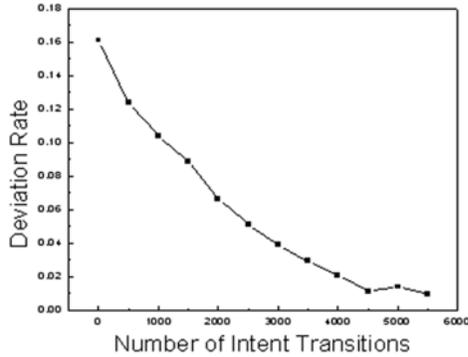


Fig. 3: The convergence process of optimizing binary mask.

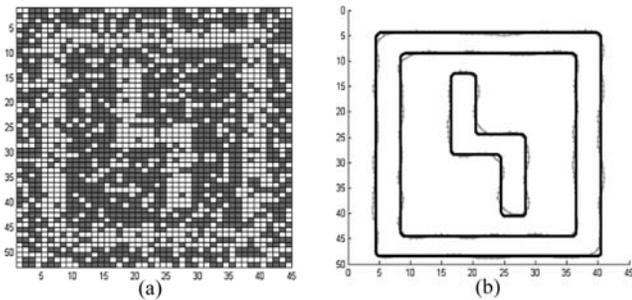


Fig. 4: (a) The optimized binary mask;  
(b) Output image due to the input mask shown in (a).

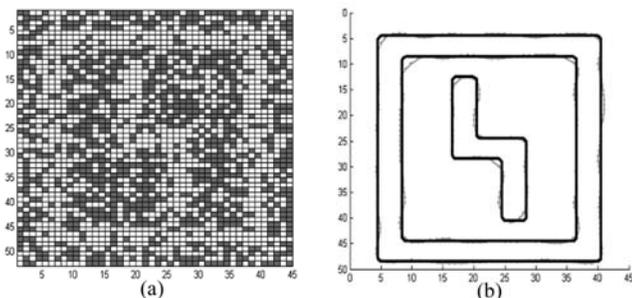


Fig. 5: (a) The optimized EPSM mask;  
(b) Output image due to the input mask shown in (a).

Not only the binary mask but also the advanced masks like 6% EPSM and APSM can be optimized through our method. The simulation results are shown in Fig. 5 and Fig. 6. In Fig. 5(a), the white pixels represent the transparent area with a transmission value of 1, while the gray ones re-

present the opaque area with a transmission value of -0.245. In Fig. 6(a), the white, gray and black pixels represent the area with a transmission value of 1, 0 and -1, respectively. Fig. 5(b) and Fig. 6(b) show the excellent image performance. And we obtain the results after 7000 and 9000 intent transitions, respectively.

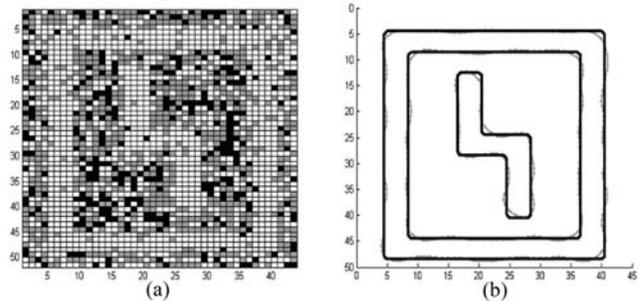


Fig. 6: (a) The optimized APSM mask;  
(b) Output image due to the input mask shown in (a).

## 5 SUMMARY

A refined optimization scheme based on simulated annealing algorithm is applied to synthesize masks in sub-wavelength lithography simulation. Three different types of masks for 65nm CMOS technology node are used to validate the method. The results shows the optimized masks provide good fidelity in patterning, and this algorithm is proved to be of high accuracy, and fast speed.

## REFERENCES

- [1] A. B. Kahng and Y. C. Pati, Proc. ACM Intl. Symp. On Physical Design, pp. 112-119, 1999.
- [2] C. Dolainsky and W. Maurer, Proc. SPIE, vol. 3051, pp. 774, 1997.
- [3] L.Liebmann, S. Mansfield, et al, IBM Journal of Research and Development 45, pp. 651-665, 2001.
- [4] B.E.A. Saleh and S.I. Sayegh, Optical Eng. Vol. 20, pp. 781-784, 1981
- [5] Yong Liu and Avideh Zakhori, IEEE Trans. Semi. Manufacturing, vol. 5, no. 2, pp. 138-152, 1992
- [6] Sherif Sherif and Bahaa Saleh, IEEE Trans. Image Processing, vol. 4, no. 9, pp. 1252-1257, 1995
- [7] Yuri Granik, SPIE, vol. 5754, pp. 506-526, 2005
- [8] Y.C. Pati and T. Kailath, J.Opt.Soc.Am.A, vol.11, pp. 2438-2452, 1994
- [9] N. Cobb, "Fast Optical and Process Proximity Correction Algorithms For Integrated Circuit Manufacturing", Ph.D. Dissertation, U.C. Berkeley, 1998.
- [10] B.Hajek, in Proc. 24th Conf. on Decision and Control, Ft. Lauderdale, pp. 775-760, 1985.
- [11] S.Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Science, vol. 220, pp. 671-680, 1983.
- [12] A. J. Thomason, "Random process," Dept. of E.E. and C. S., U. C. Berkeley, text manuscript, 1990.
- [13] A. K. Wong, "Resolution Enhancement Techniques in Optical Lithography," In Tutorial Texts in Optical Engineering, TT47. SPIE Press, 2001.