

Hotspot-aware fast source and mask optimization

Jia Li, Yijiang Shen, and Edmund Y. Lam*

*Imaging Systems Laboratory, Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong, China*

[*elam@eee.hku.hk](mailto:elam@eee.hku.hk)

Abstract: Source mask optimization (SMO) is a useful technique for printing the integrated circuit (IC) on a wafer with increasingly smaller feature size. However, complex SMO algorithms generally lead to undesirably long runtime resulting from an optimization of largely identical regions over the whole mask pattern. In this work, a weighted SMO scheme incorporating both an awareness of the hotspots and robustness against process variations is proposed. We show how optimal solutions are reached with fewer iterations by applying various degrees of correction in the corresponding regions. The proposed method includes identifying the hotspots and combining a weight matrix to the cost function for adjustment and control. Simulation results are compared with the mask optimization (under a fixed source) and conventional SMO to illustrate the performance improvement in terms of pattern fidelity, convergence rate and process window size.

© 2012 Optical Society of America

OCIS codes: (110.3960) Microlithography; (110.5220) Photolithography; (110.1758) Computational imaging.

References and links

1. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, (SPIE, Washington, 2001).
2. S. H. Chan, A. K. Wong, and E. Y. Lam, "Initialization for robust inverse synthesis of phase-shifting masks in optical projection lithography," *Opt. Express* **16**, 14746–14760 (2008).
3. S. Sherif, B. Saleh, and R. De Leone, "Binary images synthesis using mixed linear integer programming," *IEEE Trans. Image Process.* **4**, 1252–1257 (1995).
4. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," *Opt. Express* **15**, 15066–15079 (2007).
5. Y. Shen, N. Wong, and E. Y. Lam, "Level-set-based inverse lithography for photomask synthesis," *Opt. Express* **17**, 23690–23701 (2009).
6. Y. Shen, N. Jia, N. Wong, and E. Y. Lam, "Robust level-set-based inverse lithography," *Opt. Express* **19**, 5511–5521 (2011).
7. T. H. Dam, X. Zhou, D. Chen, A. Adamov, D. Peng, and B. Gleason, "Validation and application of a mask model for inverse lithography," in *Design for Manufacturability through Design-Process Integration II*, V. K. Sing, M. L. Rieger eds., *Proc. SPIE* **6925**, 69251J (2008).
8. Y. Deng, Y. Zou, K. Yoshimoto, Y. Ma, C. E. Tabery, J. Kye, L. Capodieci, and H. J. Levinson, "Considerations in source-mask optimization for logic applications," in *Optical Microlithography XXIII*, M. V. Dusa and W. Conley, eds., *Proc. SPIE* **7640**, 76401J (2010).
9. D. Melville, A. Rosenbluth, K. Tian, K. Lai, S. Bagheri, J. Tirapu-Azpiroz, J. Meiring, S. Halle, G. McIntyre, T. Faure, D. Corliss, A. Krasnoperova, L. Zhuang, P. Strenski, A. Waechter, L. Ladanyi, F. Barahona, D. Scarpazza, J. Lee, T. Inoue, M. Sakamoto, H. Muta, A. Wagner, G. Burr, Y. Kim, E. Gallagher, M. Hibbs, A. Trichtkov, Y. Granik, M. Fakhry, K. Adam, G. Berger, M. Lam, A. Dave, and N. Cobb, "Demonstrating the benefits of source-mask optimization and enabling technologies through experiment and simulations," in *Optical Microlithography XXIII*, M. V. Dusa and W. Conley, eds., *Proc. SPIE* **7640**, 764006 (2010).

10. Y. Granik, "Source optimization for image fidelity and throughput," *J. Microlith. Microfab. Microsys.* **3**, 509–522 (2004).
11. K. Iwase, P. D. Bisschop, B. Laenens, Z. Li, K. Gronlund, P. V. Adrichem, and S. Hsu, "A new source optimization approach for 2X node logic," in *Photomask Technology 2011*, W. Maurer and F. E. Abboud, eds., Proc. SPIE **8166**, 81662A (2011).
12. H. Hu, Y. Zou, and Y. Deng, "Optimization on illumination source with design of experiments," in *Optical Microlithography XXIII*, M. V. Dusa and W. Conley, eds., Proc. SPIE **7640**, 764027 (2010).
13. T. Mülders, V. Domnenko, B. Küchler, T. Klimpel, H.-J. Stock, A. A. Poonawala, K. N. Taravade, and W. A. Stanton, "Simultaneous source-mask optimization: a numerical combining method," in *Photomask Technology 2010*, M. W. Montgomery and W. Maurer, eds., Proc. SPIE **7823**, 78233X (2010).
14. M. Fakhry, Y. Granik, K. Adam, and K. Lai, "Total source mask optimization: high-capacity, resist modeling, and production-ready mask solution," in *Photomask Technology 2011*, W. Maurer and F. E. Abboud, eds., Proc. SPIE **8166**, 81663M (2011).
15. T. Dam, V. Tolani, P. Hu, K.-H. Baik, L. Pang, B. Gleason, S. D. Slonaker, and J. K. Tyminski, "Source-mask optimization (SMO): from theory to practice," in *Optical Microlithography XXIII*, M. V. Dusa and W. Conley, eds., Proc. SPIE **7640**, 764006 (2010).
16. Y. Deng, T. H. Coskun, J. Kye, and H. J. Levinson, "Lithography target optimization with source-mask optimization," in *Optical Microlithography XXV*, W. Conley, ed., Proc. SPIE **8326**, 83262P (2012).
17. X. Ma and G. R. Arce, "Pixel-based simultaneous source and mask optimization for resolution enhancement in optical lithography," *Opt. Express* **17**, 5783–5793 (2009).
18. J.-C. Yu and P. Yu, "Gradient-based fast source mask optimization (SMO)," in *Optical Microlithography XXIV*, M. V. Dusa, ed., Proc. SPIE **7973**, 797320 (2011).
19. J.-C. Yu, P. Yu, and H. Y. Chao, "Fast source optimization involving quadratic line-contour objectives for the resist image," *Opt. Express* **20**, 8161–8174 (2012).
20. E. Y. Lam and A. K. Wong, "Computation lithography: virtual reality and virtual virtuality," *Opt. Express* **17**, 12259–12268 (2009).
21. S. K. Choy, N. Jia, C. S. Tong, M. L. Tang, and E. Y. Lam, "A robust computational algorithm for inverse photomask synthesis in optical projection lithography," *SIAM J. Imaging Sciences* **5**, 625–651 (2012).
22. A. K. Wong, *Optical Imaging in Projection Microlithography*, (SPIE, Washington, 2005).
23. N. Jia and E. Y. Lam, "Pixelated source mask optimization for process robustness in optical lithography," *Opt. Express* **19**, 19384–19398 (2011).
24. 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).
25. A. Poonawala and P. Milanfar, "Mask design for optical microlithography — an inverse imaging problem," *IEEE Trans. Image Process.* **16**, 774–788 (2007).
26. J. Kim and M. Fan, "Hotspot detection on post-OPC layout using full chip simulation based verification tool : a case study with aerial image simulation," in *23rd Annual BACUS Symposium on Photomask Technology*, K. R. Kimmel and W. Staud, eds., Proc. SPIE **5256**, 919–925 (2003).
27. M. L. Kempell, 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. Microlith. Microfab. Microsys.* **8**, 043001 (2009).
28. J.-C. Yu and P. Yu, "Choosing objective functions for inverse lithography patterning," in *Optical Microlithography XXIV*, M. V. Dusa, ed., Proc. SPIE **7973**, 79731N (2011).
29. K. Lai, A. E. Rosenbluth, S. Bagheri, J. Hoffnagle, K. Tian, D. Melville, J. Tirapu-Azpiroz, M. Fakhry, Y. Kim, S. Halle, G. McIntyre, A. Wagner, G. Burr, M. Burkhardt, D. Corliss, E. Gallagher, T. Faure, M. Hibbs, D. Flagello, J. Zimmermann, B. Kneer, F. Rohmund, F. Hartung, C. Hennerkes, M. Manu, R. Kazinczi, A. Engelen, R. Carpaij, R. Groenendijk, J. Hageman, and C. Russ, "Experimental result and simulation analysis for the use of pixelated illumination from source mask optimization for 22nm logic lithography process," in *Optical Microlithography XXIV*, H. J. Levinson, M. V. Dusa eds., Proc. SPIE **7274**, 72740A (2009).
30. S. Hsu, Z. Li, L. Chen, K. Gronlund, H.-Y. Liu, and R. Socha, "Source-mask co-optimization: optimize design for imaging and impact of source complexity on lithography performance," in *Lithography Asia 2009*, A. C. Chen, W.-S. Han, B. J. Lin, and A. Yen eds., Proc. SPIE **7520**, 75200D (2009).
31. J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).
32. N. Jia and E. Y. Lam, "Machine learning for inverse lithography: using stochastic gradient descent for robust photomask synthesis," *J. Opt.* **12**, 045601 (2010).
33. T. H. Coskun, H. Dai, V. Kamat, C.-M. Hsu, G. Santoro, C. Ngai, M. Reybrouck, G. Grozev, and H.-T. Huang, "Free form source and mask optimization for negative tone resist development for 22nm node contact holes," in *Optical Microlithography XXV*, W. Conley, ed., Proc. SPIE **8326**, 83260V (2012).

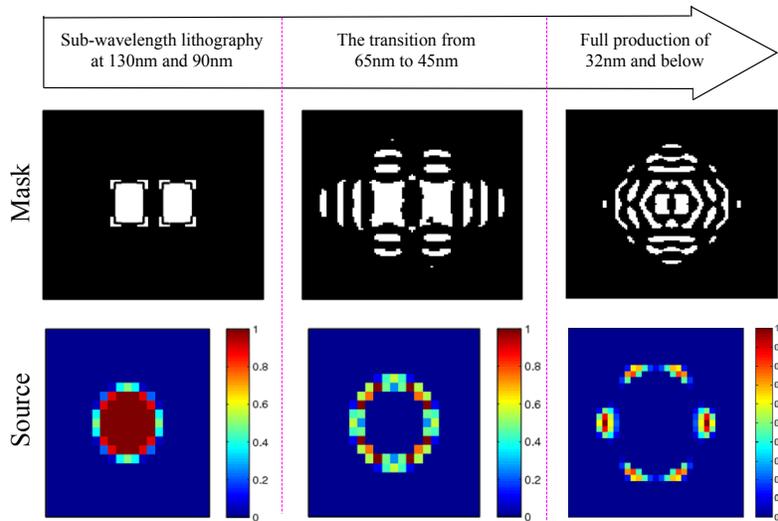


Fig. 1. Typical masks and sources are used to print a sparse pattern consisting of two rectangle shapes at different development stages of lithography.

1. Introduction

The development and mass production of three-dimensional transistors at the 22nm node impose severe constraints on optical lithography to achieve sufficient pattern fidelity and a reasonable process window (PW). As illustrated in Fig. 1, with continuously shrinking integrated circuit (IC) feature size, various resolution enhancement techniques (RETs) have to be used. At sub-wavelength lithography such as the 130nm and 90nm nodes, relatively simple rule-based optical proximity correction (OPC) and phase-shifting masks (PSMs) sufficed [1, 2]. As technology migrated from 65 nm to 45 nm, inverse lithography was seen by many as the new solutions. It poses the mask synthesis as an inverse problem solved by optimizing an objective function using either the thin mask model [3–6] or the thick mask model [7], and as a result, it is capable of exploiting a larger solution space to achieve a better pattern fidelity. Nevertheless, at 32nm and below, even inverse lithography by itself becomes inadequate for the task. Forbidden pitches are more pronounced and design rules become very complicated [8]. It is now necessary to combine the mask design with illumination configuration adjustment [9], a process known as source-mask optimization (SMO).

Indeed, source profile also experiences a series of development, from a traditional circular shape at the beginning to off-axis illumination (OAI), to the realization of the pixelated source using diffractive optical elements (DOE). Consequently, source optimization (SO) provides more degrees of freedom on both the source profile and its intensity [10–12]. Together with the pixelated mask design in inverse lithography, SMO promises a higher resolution and improved performance on aberration robustness without resorting to extremely expensive optical systems, extending the lifetime of the current immersion lithography [13–19].

Unfortunately, although various mathematical algorithms provide rich techniques for solving SMO, the heavy computation still limits its adoption for practical use in full-chip simulation [20]. In addition, smaller dimensions are more sensitive to process variations, necessitating algorithms that incorporate such variations in the optimization process [21]. In this paper, we aim to develop an efficient and robust SMO algorithm using a hotspot-aware inverse synthesis technique. We first propose a hotspot selection criterion based on the target image and

its “smallest image”. In accordance with the selected hotspots, we then form a corresponding weight matrix and use it in the optimization procedure. This enables a heavier correction of the hotspot regions, and as a result we need fewer number of iterations, thereby shortening the overall execution time. The weight matrix also enables an expansion of the process window by simultaneously taking into account the focus and dose variations. Experimental results show that our method leads to faster convergence and an improved performance on process robustness against aberrations.

2. Forward imaging model

The lithography modeling is often divided into two parts, namely the aerial image formation and the photo-resist development. Under a partially coherent illumination (PCI) system, the aerial image $I_a(x, y)$ can be described as [22]

$$\begin{aligned} I_a(x, y) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} J(f, g) \hat{M}(f_1, g_1) \hat{M}^\dagger(f_2, g_2) \hat{H}(f + f_1, g + g_1) \hat{H}^\dagger(f + f_2, g + g_2) \\ &\quad \times e^{-i2\pi[(f_1 - f_2)x + (g_1 - g_2)y]} df dg df_1 dg_1 df_2 dg_2 \\ &= \iint_{-\infty}^{\infty} J(f, g) |M(x, y) * \tilde{H}(x, y)|^2 df dg, \end{aligned} \quad (1)$$

where

$$\tilde{H}(x, y) = H(x, y) e^{-i2\pi(fx + gy)}, \quad (2)$$

In Eq.(1), $M(x, y)$ is the input mask pattern, with its spectrum denoted by \hat{M} . The pupil function \hat{H} , where its inverse Fourier transform is $H(x, y)$, is called the point spread function. The symbol \dagger and $*$ are complex conjugate and convolution operators, respectively. The function $J(f, g) \geq 0$ represents the effective source, which is normalized by its total energy [23, 24], i.e.,

$$J(f, g) = \frac{J'(f, g)}{\iint_{-\infty}^{\infty} J'(f, g) df dg}, \quad (3)$$

where $J'(f, g)$ is an arbitrary source point.

For image computation, we use the discrete form of Eq. (1), given approximated by

$$I_a(x, y) \approx \frac{\sum_{f, g} \left\{ J'(f, g) |M(x, y) * \tilde{H}(x, y)|^2 \right\}}{\sum_{f, g} J'(f, g)}. \quad (4)$$

The aerial image $I_a(x, y)$ then goes through the photoresist development to form the printed image $I(x, y)$. Approximating the resist effect with a sigmoid function due to its differentiability [25], the output $I(x, y)$ is given by

$$I(x, y) = \text{sig} \{I_a(x, y)\} = \frac{1}{1 + e^{-\alpha[I_a(x, y) - t_r]}}, \quad (5)$$

in which t_r is the threshold and α indicates the steepness of the sigmoid function.

3. Weighted source mask optimization algorithm

3.1. Hotspots selection and weight matrix formation

The objective of our work is to devise a method that incorporates different degrees of source and mask corrections at various regions. Given a mask pattern, our algorithm needs to detect the

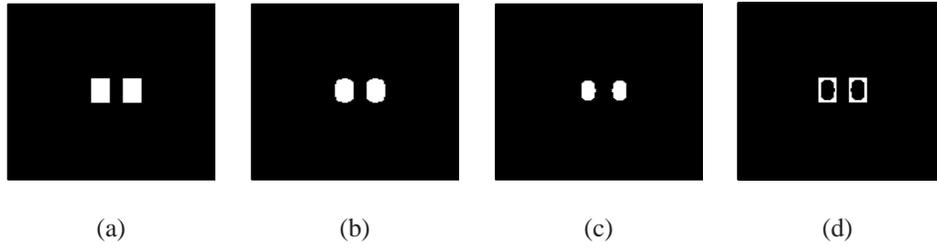


Fig. 2. Generation of images at multiple process conditions: (a) is an input pattern, (b) is its printed image at the nominal focus and dose, and (c) is the smallest image created at 70nm and 10% exposure dose increase. The difference between (c) and the input pattern (a) is given in (d).

hotspot regions automatically, which is then embedded in the cost function of a weighted SMO algorithm through the use of a weight matrix. In addition, process conditions such as focus and exposure dose have a significant influence on the quality of the output patterns. Variations in these process conditions degrade the printing performance, leading to hotspot regions which are prone to bridging or pinching [26].

In our approach, we make use of the concept called the “smallest image” to show how printing responds to such variations. Ref. [27] defines it to be the printed image that is out of focus with a small exposure; note, however, that for a positive photoresist, a higher exposure dose results in a smaller feature size and a decrease in image sharpness. Accordingly, in Fig. 2, (a) is an input pattern, (b) is its printed image at the nominal focus and dose, and (c) is the smallest image created at 70nm and 10% exposure dose increase. The difference between (c) and the input pattern (a) is given in (d). Mathematically, assuming that the desired printed image is $I_0(x,y)$ and $I_s(x,y)$ is the smallest image, the difference image, denoted $\Delta I(x,y)$, is therefore

$$\Delta I(x,y) = |I_0(x,y) - I_s(x,y)|. \quad (6)$$

For the purpose of this work, “hotspot regions” are places where $\Delta I(x,y)$ take on larger values, showing greater sensitivity to focus and dose variations. They represent the critical locations that are often difficult to print, thus requiring more severe corrections in the inverse imaging process. To allow for a smooth transition from such hotspot regions to other areas, we convolve $\Delta I(x,y)$ (at places where it is non-zero) with a lowpass filter $G(x,y)$, such as a 3×3 Gaussian kernel, i.e., we define

$$I_m(x,y) = \begin{cases} 0 & , \Delta I(x,y) = 0 \\ \Delta I(x,y) * G(x,y) & , \Delta I(x,y) \neq 0. \end{cases} \quad (7)$$

We then design the weight matrix $W(x,y)$ that corrects a certain location more severely when the corresponding $I_m(x,y)$ is large. When $I_m(x,y)$ is small compared with a threshold t_w , empirically set to be half of the maximum value in I_m , we put in 1 in the corresponding weight matrix position, which is to be the smallest value in the matrix. When it exceeds the threshold, $W(x,y)$ is set to be proportional to $I_m(x,y)$, with the constant of proportionality, λ_1 , a user-defined parameter. Mathematically, we therefore have

$$W(x,y) = \begin{cases} 1 & , I_m(x,y) < t_w \\ \lambda_1 I_m(x,y) & , I_m(x,y) \geq t_w. \end{cases} \quad (8)$$

The threshold t_w has an effect on the hotspot selection. Specifically, a larger t_w will miss the detection of some potential hotspot regions, while a smaller t_w will include extra pattern regions,

in both cases leading to unnecessary computation. The parameter λ_1 plays an important role in controlling the optimal step size. A smaller λ_1 cannot sufficiently penalize the hotspot regions, while a larger one may exert excessive penalty. In our experiments, $t_w = 0.5$ and $\lambda_1 = 3$ are selected empirically.

3.2. The cost function

Just like other inverse imaging methods, the design of a proper cost function is crucial for computational lithography. Generally speaking, the cost function consists of a pattern fidelity term and one or more regularization terms for specific purposes [28]. For the former, the weight matrix $W(x, y)$ enters via a pixel-by-pixel multiplication with the difference between the simulated circuit image $I(x, y)$ and the desired image, $I_0(x, y)$. In this work, such difference is measured by the ℓ_2 norm. The weight matrix has the effect of varying the step length during the iteration process, resulting in faster convergence with fewer iterations. Mathematically, the pattern fidelity term is given by

$$\mathcal{F}\{I(x, y), I_0(x, y)\} = \sum_{x, y} W(x, y) \odot \|I(x, y) - I_0(x, y)\|_2^2, \quad (9)$$

where \odot indicates pixel-by-pixel multiplication.

Gradient methods are frequently used to tackle the optimization problem iteratively. In SMO, we need two gradients: one is the derivative of \mathcal{F} with respect to an arbitrary source point $J'(f, g)$, and the other one is the derivative of \mathcal{F} with respect to the mask $M(x, y)$. Here we define the differential operator $\nabla F(a)$ to evaluate the gradient of a function F with respect to its argument a in the discrete domain, due to the discrete nature of the mask and source. As shown in the Appendix, these are given by

$$\nabla \mathcal{F}(M) = \sum_{f, g} J(f, g) \odot \text{Re} \left\{ W \odot [2\alpha(I - I_0) \odot I \odot (1 - I) \odot (M * \tilde{H})^\dagger] * \tilde{H}(-x, -y) \right\}, \quad (10)$$

$$\nabla \mathcal{F}(J') = \sum_{x, y} W \odot 2\alpha(I - I_0) \odot I \odot (1 - I) \odot \frac{|M * \tilde{H}|^2 - I_a}{\sum_{f, g} J'(f, g)}. \quad (11)$$

As for the regularization terms, we use three in the present work. The first one is designed to improve the image contrast by driving the intensity values away from the threshold t_r as much as possible. To achieve this, we minimize $\|I_a(x, y) - 2t_r I_0(x, y)\|_2^2$, such that $I_a \approx 0$ at locations where the desired pattern $I_0(x, y) = 0$, and force $I_a \approx 2t_r$ where $I_0(x, y) = 1$. This enables the intensity both above and below the nominal threshold t_r to be equally penalized to reduce their sensitivity to dose changes [23]. Hence, the first regularization term, as used in our previous work, is given by

$$\mathcal{R}_a\{I_a(x, y), I_0(x, y)\} = \sum_{x, y} \|I_a(x, y) - 2t_r I_0(x, y)\|_2^2. \quad (12)$$

The second regularization term relates to the illumination source. Since the experimental confirmation of the feasibility of using pixelated illumination in SMO technology [29], research has shown that the pixelated source consistently gives better performance in comparison with traditional source types [23, 30]. Nevertheless, from the viewpoint of manufacturing, the sensitivity of the illumination pupil is also higher than that of a conventional simple source due to the smaller and isolated pixel size [29]. Accordingly, we aim to design a light source that is not too complex, i.e., it would contain as few isolated pixels as possible. This is a design tradeoff; thus, we achieve this through devising a regularization term and allow the regularization parameter

to control the extent to which this requirement needs to be enforced. Unlike in Ref. [24], which seeks to limit the total perimeter of all the isolated patterns, we constrain on the local variation of the source, which can suppress the small-scale protrusion while preserving the large-scale features. This gives rise to the regularization term

$$\mathcal{R}_s\{J(f, g)\} = \sum_{f, g} \left\{ \|\nabla_f[J(f, g)]\|_1 + \|\nabla_g[J(f, g)]\|_1 \right\}, \quad (13)$$

where $\nabla_f[J(f, g)]$ and $\nabla_g[J(f, g)]$ are denotations of finite difference in numerical implementations, defined by

$$\nabla_f[J(f, g)] = J(f+1, g) - J(f, g) \text{ and } \nabla_g[J(f, g)] = J(f, g+1) - J(f, g). \quad (14)$$

The third regularization term is needed to obtain a near binary mask. As in Ref. [25], we first relax the mask pixel values to take on values between 0 and 1 inclusive, and penalize as they depart from these two end points. The resulting binary mask is then obtained by thresholding at 0.5. Thus, the third regularization term is

$$\mathcal{R}_m\{M(x, y)\} = \sum_{x, y} M(x, y) [1 - M(x, y)]. \quad (15)$$

As with the pattern fidelity term, using the gradient method requires the first derivatives of these regularization terms with respect to the source and the mask. Note that $\mathcal{R}_s\{J(f, g)\}$ does not involve the mask, and $\mathcal{R}_m\{M(x, y)\}$ does not involve the source. Thus, the two gradients we need are

$$\begin{aligned} \nabla \mathcal{R}_a(M) + \nabla \mathcal{R}_m(M) = \\ \sum_{f, g} J(f, g) \odot \text{Re} \left\{ [2(I_a - 2t_r I_0) \odot (M * \tilde{H})^\dagger] * \tilde{H}(-x, -y) \right\} + (-2M + 1) \end{aligned} \quad (16)$$

and

$$\begin{aligned} \nabla \mathcal{R}_a(J') + \nabla \mathcal{R}_s(J) = \\ \sum_{x, y} 2(I_a - 2t_r I_0) \odot \frac{|M * \tilde{H}|^2 - I_a}{\sum_{f, g} J'(f, g)} + \text{sgn}\{\nabla_f[J(f, g)]\} + \text{sgn}\{\nabla_g[J(f, g)]\}, \end{aligned} \quad (17)$$

The derivations are detailed in the Appendix.

Combining Eqs. (9), (12), (13) and (15), the overall cost function is

$$\mathcal{C} = \mathcal{F}\{I(x, y), I_0(x, y)\} + \gamma_1 \mathcal{R}_a\{I_a(x, y), I_0(x, y)\} + \gamma_2 \mathcal{R}_s\{J(f, g)\} + \gamma_3 \mathcal{R}_m\{M(x, y)\}, \quad (18)$$

where γ_1 , γ_2 and γ_3 are parameters for the tradeoff between the pattern fidelity term and the regularization terms.

3.3. Optimization flow

By minimizing the above cost function, our weighted SMO algorithm leads to an optimized mask pattern $M_{\text{opt}}(x, y)$ and the corresponding source $J_{\text{opt}}(f, g)$, i.e.

$$\{M_{\text{opt}}(x, y), J_{\text{opt}}(f, g)\} = \arg \min_{\{M, J\}} \mathcal{C}\{I(x, y), I_0(x, y)\}. \quad (19)$$

The optimization is solved iteratively using gradient methods, such as conjugate-gradient (CG) [31]. We can initialize the source with the conventional annular illumination and the

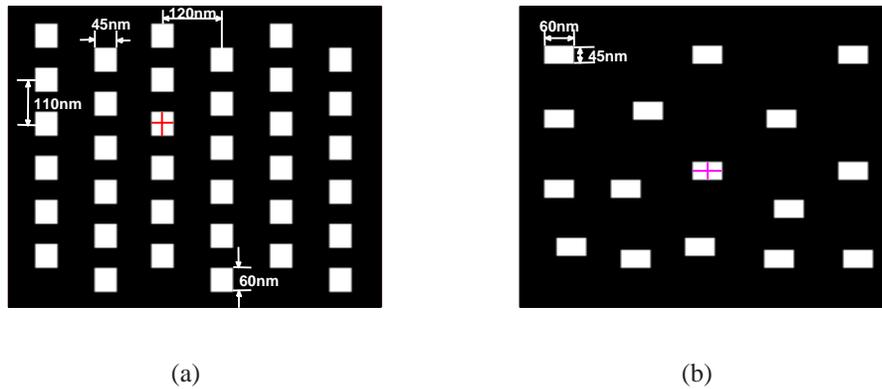


Fig. 3. Two test patterns used in experiments: (a) brick contact holes and (b) random contact holes.

mask with the target pattern. Then, we apply a statistical technique similar to our previous work in Ref. [32] that caters for not only the nominal condition but optimizes over a range of focus and dose variations. This is achieved by treating the process variations as normally distributed random variables [32]. Specifically, mathematical defocus model adds a phase term to the optical transfer function, as described in Ref. [6]; dose variation can be accounted for by varying the threshold t_r , and then the cost function is optimized by a weighted sum of the nominal and the aberration terms, in which the weighting coefficients are determined by the variable probabilities. The mask and source updates are performed alternately until the termination criterion is reached. The gradients of the objective function with respect to the mask and source are given by

$$\nabla \mathcal{L}(M) = \nabla \mathcal{F}(M) + \gamma_1 \nabla \mathcal{R}_a(M) + \gamma_3 \nabla \mathcal{R}_m(M). \quad (20)$$

and

$$\nabla \mathcal{L}(J') = \nabla \mathcal{F}(J') + \gamma_1 \nabla \mathcal{R}_a(J') + \gamma_2 \nabla \mathcal{R}_s(J). \quad (21)$$

The pseudo-code in Table 1 explains the general procedure of this proposed algorithm. The k th iteration is denoted by the superscript k with brackets.

4. Results

Contact arrays patterning has been one of the most challenging features in semiconductor device manufacturing due to their smaller process window [33]. Here, we employ the weighted SMO algorithm described above to analyze two target patterns, namely, brick and random contact holes, as shown in Fig. 3(a) and (b) respectively. Both are represented by 151×151 matrix with a pixel size of $5 \text{ nm} \times 5 \text{ nm}$ and critical dimension (CD) of 45 nm. An annular illumination composed of 21×21 pixels with $\sigma_{in} = 0.7$ and $\sigma_{out} = 0.9$ is adopted as the reference source, as well as the initial value for our source optimization. The parameters of the projection system are set to be $\lambda = 193 \text{ nm}$ and $NA = 1.35$. In the sigmoid function, t_r and α are equal to 0.3 and 85, respectively. It is noted that effective source $J(f, g)$ is applied to aerial image computation throughout the optimization flow to normalize the aerial image in Eq.(4) such that in spite of changing source integration in the iterations, identical total source energy justifies a constant threshold t_r in the photoresist approximation. In the hotspots selection step, the smallest image is generated at 70nm defocus and 10% dose increase.

Table 1. Pseudo-code of SMO

Mask optimization procedure

1. Initialization: Assign $d_m^{(0)} = 0$, $g_m^{(0)} = 0$, the iteration number P , a step size $s_1 = 3$ and the starting mask $M^{(1)}$.
2. Iteratively update the mask:
 - for $k = 1 : P$
 - $d_m^{(k)} = \nabla \mathcal{C}(M^{(k)})$, calculate the gradient with respect to $M^{(k)}$, as in Eq.(20);
 - Update the scalar $\alpha_m^{(k)} = [d_m^{(k)} (d_m^{(k)} - d_m^{(k-1)})] / [(d_m^{(k)} - d_m^{(k-1)}) g_m^{(k-1)}]$;
 - $g_m^{(k)} = -d_m^{(k)} + \alpha_m^{(k)} g_m^{(k-1)}$;
 - Set $M^{(k+1)} = M^{(k)} + s_1 g_m^{(k)}$;
 - $k \leftarrow k + 1$;
 - end
3. Output the optimal mask for source optimization.

Source optimization procedure

1. Initialization: Assign $d_s^{(0)} = 0$, $g_s^{(0)} = 0$, the iteration number Q , a step size $s_2 = 2$ and the starting source $J^{(1)}$.
2. Iteratively update the source:
 - for $k = 1 : Q$
 - $d_s^{(k)} = \nabla \mathcal{C}(J^{(k)})$, calculate the gradient with respect to $J^{(k)}$, as in Eq.(21);
 - Update the scalar $\alpha_s^{(k)} = [d_s^{(k)} (d_s^{(k)} - d_s^{(k-1)})] / [(d_s^{(k)} - d_s^{(k-1)}) g_s^{(k-1)}]$;
 - $g_s^{(k)} = -d_s^{(k)} + \alpha_s^{(k)} g_s^{(k-1)}$;
 - Set $J^{(k+1)} = J^{(k)} + s_2 g_s^{(k)}$;
 - $k \leftarrow k + 1$;
 - end
3. Output the optimal source to mask computation.

Overall stop condition:

If $\mathcal{C}^{k+1} - \mathcal{C}^k < \varepsilon$ is satisfied, a final mask optimization is performed; otherwise, run mask optimization procedure and source optimization procedure again.

In order to validate the performance of hotspot-aware SMO presented in this paper, Fig. 4 shows the comparison of experimental results using the proposed weighted SMO and conventional mask optimization illuminated by the reference annular source, with the contact pattern in Fig. 3(a) as input. Figure 4(a) to (d) illustrate the resulting source and the optimized mask, as well as the printed image at nominal conditions and magnified output at a defocus of 60 nm, respectively, using the proposed weighted SMO. We can compare with corresponding results from mask optimization (MO) under a reference source in terms of pattern error, which are given in the following row with the same structure. It is observed that both methods generate similar output at best focus, however, under the circumstance of defocus, printed image quality using the proposed weighted SMO is greatly improved, especially that the regions around corners of the contact holes are better printed in Fig. 4(d), compared with that in Fig. 4(h) using conventional mask optimization. The annular source and all the optimized sources are normalized by the maximum pixel intensity for better visualization.

Another experiment is conducted with the contact pattern in Fig. 3(b) as input, using our

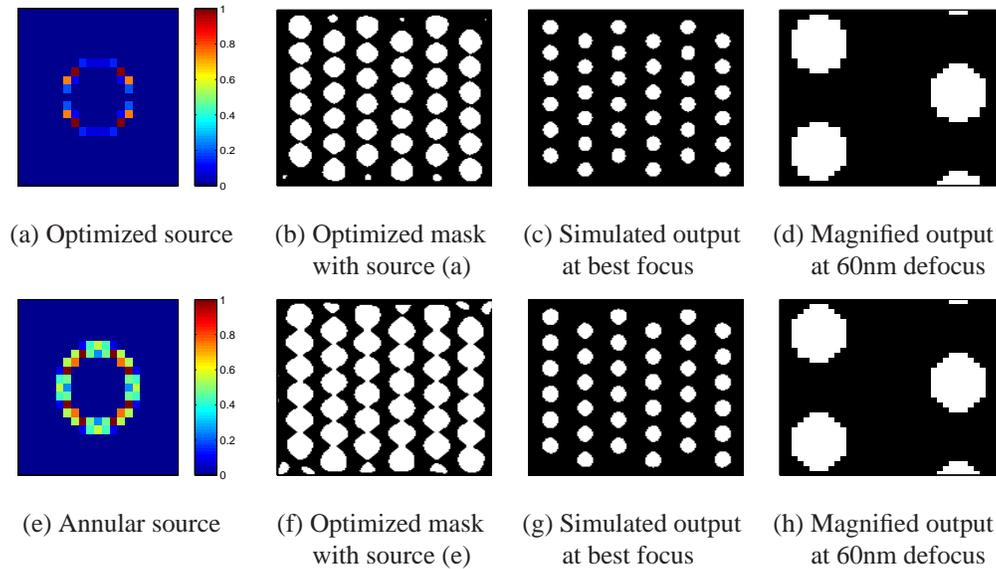


Fig. 4. Simulation results of the first test pattern.

proposed SMO framework and conventional SMO without using the hotspot-aware weight matrix. Figure 5 presents the corresponding results. We notice that in Fig. 5(d) of the magnified printed image in 60nm defocus condition, conventional SMO fails to print the contact hole marked by a magenta cutlines in Fig. 3(b). In contrast, our algorithm still produces the output image with high pattern fidelity. It is duly noted that the optimized mask in Fig. 5(f) contains very small features which might lead to mask manufacturing problems. To investigate the effect of the removal of the small features, simulations are conducted with inputs removing several smallest features and regularizing some assist feature shapes, showing pattern error results of 290 and 1017 under best focus and 60nm defocus, respectively, which is similar to the results in Fig. 5(g) and (h) (with pattern errors of 288 and 1039, respectively). Thus, the effect of removal of the smallest features is small. In the following analysis, one can see that the optimal source and mask in Fig. 5(e) and (f) can be found with fewer number of iterations than that in a uniform SMO method.

After evaluating the performance of different algorithms, we can now assess the impact of the proposed weighted SMO in terms of convergence rate. Since the hotspots selection step is conducted only once and the execution time of computing the source optimization and mask optimization is similar, fewer iterations spent indicates less overall computing time.

Table 2 summarizes the measurements of pattern error, cost function value and speed on two test patterns using three different methods. For both test patterns, when three methods all produce similar cost function value, comparing with mask optimization only and conventional SMO, our method takes fewer iterations to achieve a better performance, exhibiting a 30% to 200% convergence improvement.

It can be seen that weighted SMO converges faster than mask optimization only. For instance, with the brick contacts pattern as input, the weighted SMO has reached the optimal solution after 87 iterations, with a pattern error of 675, while MO, during the optimization process, we observe that after undergoing the same numbers of iterations, prints the image with a pattern error of 930, and another 80 iterations are required to obtain the optimal solution, namely the cost function change little. In addition, we observe that MO is not capable of acquiring the best

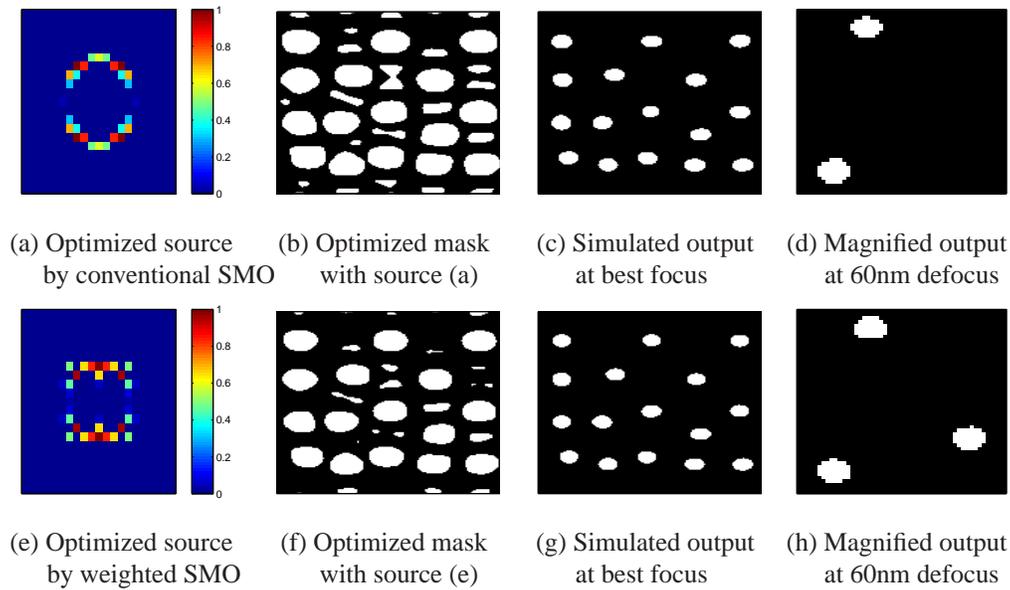


Fig. 5. Simulation results of the second test pattern.

Table 2. Comparison of performance and convergence rate

Test patterns	Methods	Pattern error	Cost function	Iteration number
Brick contacts	MO	878	880.4	167
	Conventional SMO	770	832.8	129
	Weighted SMO	675	833.8	87
Random contacts	MO	353	355.6	256
	Conventional SMO	348	353.2	272
	Weighted SMO	288	354.9	181

pattern fidelity in terms of pattern error achieved by SMO even if it continues the iterations. This is reasonable because a more flexible source enlarges the solution space of the inverse problems as compared to MO.

It should also be noted that although conventional SMO and weighted SMO explore the same solution space, the latter penalizes the hotspot regions of the mask pattern subject to process variations and system distortions with heavier corrections, resulting in different shapes of the optimal source and the arrangement of the assist features, and contributing greatly to pattern fidelity in hotspot regions. This is consistent with our observation in the corner areas of Fig. 4(d) and (h), as well as Fig. 5(d) and (h). Moreover, this result is related to the level-set based lithography framework [6] in that penalizing pattern regions distinctively with different corrections drives the contour of the mask pattern, namely the zero-level-set of the image function, with bigger nominal speed in the hotspot regions. This minimizes the distance functions with the optimal contour, easing the contour movement.

The process window describes range of process variations that still permit an acceptable yield. In Fig. 6, the average focus-exposure window of the width of the printed lines is depicted. Although there are other characteristic parameters one can use to evaluate the process window,

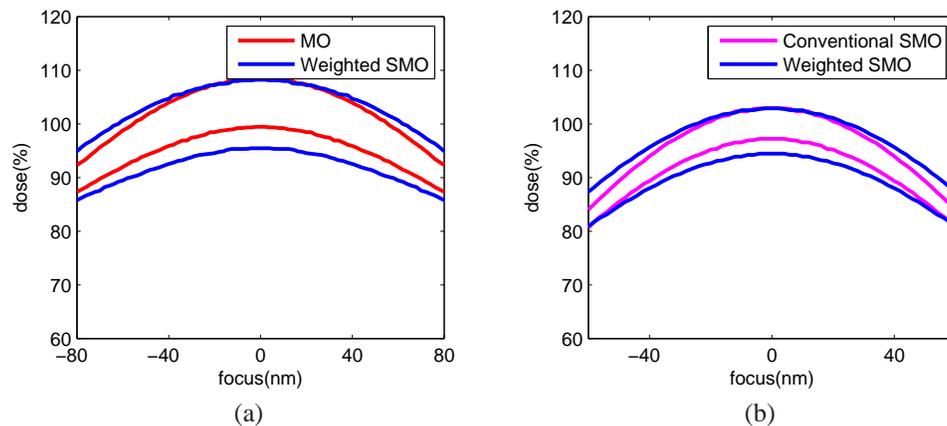


Fig. 6. Comparison of average process window of (a) brick contact holes and (b) random contact holes.

linewidth remains to be the most useful metric. In this figure, the sensitivity of pattern CD to defocus and exposure dose is measured at all contacts width and length, as marked outline settings in Fig. 3(a) and (b) by calculating how large the depth of focus (DOF) can be when exposure latitude (EL) is fixed. Allowable maximum and minimum doses of the corresponding locations with linewidth change within 10% are plotted as different color curve pairs, indicating the proposed weighted SMO, the conventional SMO, and MO, respectively. DOF is evaluated by checking the largest acceptable defocus range of an ellipse tangent with the color curve pairs at a particular dose. If we fix the EL condition at 10%, from Fig. 6(a), the proposed weighted SMO exhibits a larger average process window than conventional MO illuminated by reference annular source, and enlarges the DOF by 15 nm, demonstrating an enhanced variation robustness. Similar improvements of PW can be observed in Fig. 6(b), where the proposed hotspot-aware SMO algorithm increases the process capability by producing a larger average PW and a 22 nm larger defocus range than conventional SMO.

5. Conclusions

In this paper, a weighted SMO algorithm is developed for efficient and robust source and mask design in optical lithography. The proposed scheme solves an optimization problem incorporating hotspots detection and statistical process variation models, providing not only better performance in terms of pattern error quality, convergence rate and robustness against defocus and dose variations justified by experimental results, but also algorithmic insights of how weighting of the cost function improves the convergence in the optimization process.

A. Appendix: Gradients derivation

In the following we present how to compute the derivatives of the cost function in Eqs. (10), (11), (16) and (17). Because of the discrete nature of the source and mask, the differential operator $\partial/\partial M$ and $\partial/\partial J'$ are approximated by numerical differences.

The first gradients of the pattern fidelity term (Eq. (9)) with respect to mask pattern and

illumination source are given by

$$\begin{aligned}
\nabla_{\mathcal{F}}(M) &= \frac{\partial \sum_{x,y} W \odot \|I - I_0\|_2^2}{\partial M} \\
&= W \odot [2\alpha(I - I_0) \odot I \odot (1 - I) \odot \frac{\partial I_a}{\partial M}] \\
&= \sum_{f,g} J(f, g) \odot \text{Re} \left\{ W \odot [2\alpha(I - I_0) \odot I \odot (1 - I) \odot (M * \tilde{H})^\dagger] * \tilde{H}(-x, -y) \right\} \quad (22) \\
\nabla_{\mathcal{F}}(J') &= \frac{\partial \sum_{x,y} W \odot \|I - I_0\|_2^2}{\partial J'} \\
&= \sum_{x,y} W \odot 2\alpha(I - I_0) \odot I \odot (1 - I) \odot \frac{\partial I_a}{\partial J'} \\
&= \sum_{x,y} W \odot 2\alpha(I - I_0) \odot I \odot (1 - I) \odot \frac{|M * \tilde{H}|^2 \sum_{f,g} J'(f, g) - \sum_{f,g} J'(f, g) |M * \tilde{H}|^2}{[\sum_{f,g} J'(f, g)]^2} \\
&= \sum_{x,y} W \odot 2\alpha(I - I_0) \odot I \odot (1 - I) \odot \frac{|M * \tilde{H}|^2 - I_a}{\sum_{f,g} J'(f, g)}. \quad (23)
\end{aligned}$$

The analytical form of the partial gradients of the regularization terms in Eqs. (16) and (17) are derived by

$$\begin{aligned}
\nabla_{\mathcal{R}_a}(M) + \nabla_{\mathcal{R}_m}(M) &= \frac{\partial \sum_{x,y} \|I_a - 2t_r I_0\|_2^2}{\partial M} + \frac{\partial \sum_{x,y} M(1 - M)}{\partial M} \\
&= 2(I_a - 2t_r I_0) \odot \frac{\partial I_a}{\partial M} + (-2M + 1) \\
&= \sum_{f,g} J(f, g) \odot \text{Re} \left\{ [2(I_a - 2t_r I_0) \odot (M * \tilde{H})^\dagger] * \tilde{H}(-x, -y) \right\} + (-2M + 1) \quad (24) \\
\nabla_{\mathcal{R}_a}(J') + \nabla_{\mathcal{R}_s}(J) &= \frac{\partial \sum_{x,y} \|I_a - 2t_r I_0\|_2^2}{\partial J'} + \frac{\partial \sum_{f,g} \left\{ \|\nabla_f[J(f, g)]\|_1 + \|\nabla_g[J(f, g)]\|_1 \right\}}{\partial J} \\
&= 2(I_a - 2t_r I_0) \odot \frac{\partial I_a}{\partial J'} + \frac{\nabla_f[J(f, g)]}{\sqrt{\{\nabla_f[J(f, g)]\}^2}} + \frac{\nabla_g[J(f, g)]}{\sqrt{\{\nabla_g[J(f, g)]\}^2}} \\
&= \sum_{x,y} 2(I_a - 2t_r I_0) \odot \frac{|M * \tilde{H}|^2 - I_a}{\sum_{f,g} J'(f, g)} + \text{sgn}\{\nabla_f[J(f, g)]\} + \text{sgn}\{\nabla_g[J(f, g)]\}. \quad (25)
\end{aligned}$$

Acknowledgments

This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project HKU 7134/08E, and by the UGC Areas of Excellence project Theory, Modeling, and Simulation of Emerging Electronics.