Source optimization incorporating margin image average with conjugate gradient method

Jue-Chin Yu[1], Peichen Yu[1]* and Hsueh-Yung Chao[2]

[1]*Department of Photonics and Institute of Electro-Optical Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.*

[2]*ANSYS, Inc., 225 West Station Square Drive, Suite 200, Pittsburgh, PA 15219, USA.*

**Corresponding author:* yup@faculty.nctu.edu.tw

ABSTRACT

Source optimization (SO) becomes increasingly important to resolution enhancement in sub-32 nm lithography nodes because the dense pattern configurations significantly limit the capability of mask correction. A key step in SO is the image formation by Abbe's method, which is a linear operation of integrating all source points' images incoherently to form aerial images. However, the aerial images are usually converted to resist images through the nonlinear *sigmoid* function. Such operation loses the merit of linearity in optimization and leads to slow convergence and time-consuming calculation. In this paper we propose a threshold-based linear resist model to replace the sigmoid model in SO. The effectiveness of our proposed model can be clearly seen from mathematical analysis. We also compare results based on linear and sigmoid models. Highly similar optimal sources are obtained, but the linear model has a significant advantage over the sigmoid in terms of convergence rate and simulation time. Furthermore, the process variations characterized by exposure-defocus (E-D) windows are still in similar trends for optimal sources based on two different resist models.

Keywords: Microlithography, Inverse problems, Computational imaging, Source optimization, SMO

1. INTRODUCTION

In recent years source optimization (SO) [1-3] has attracted great interests among semiconductor foundries and equipment vendors because of its capability for further extending the life of 193 nm optical lithography. With the availability of free-form sources using diffractive optic elements (DOE), SO serves as a new option for achieving higher resolution without increasing the complexity of mask design. The proposal of source mask co-optimization (SMO) further permits the exploration of design spaces for both illuminations and masks [4-8]. However, SO mainly relies on the complexity of computational lithography algorithms to explore all possible illumination shapes. Hence the design of the numerical algorithms and physical models has a significant impact on the quality of developed patterns and the manufacturability of sources.

Among various factors, we have identified that both the cost function and the resist model have a significant impact on the convergence of gradient-based SO. Rigorously speaking, both factors are interrelated to each other and should be designed as a whole. The logarithmic sigmoid function [9-11] has been extensively used to approximate the resist effects in optical microlithography [12-14]. It has the advantage of being differentiable and its parameters are adjustable according to the sensitivity of photoresists. However, the conventional approach by a sigmoid transformation of the aerial image is a nonlinear operation but widely used for mask correction due to its contour aware property. The nonlinearity lengthens the computational time and increases the probability of local minimum traps as seen in mask optimization (MO). To circumvent such problems, Chan et al. have proposed a projection-based active set method to improve the convergence [13].

In this paper, we take a different approach to improve the efficiency and robustness of SO. For Abbe's formulation, image formation via source integration can be viewed as a linear system. If the associated objective functions have a quadratic form, the source optimization (SO) algorithm can be rather efficient and exhibits a global minimum. To reach a compromise between speed and image fidelity, the cost function of SO often involves a quadratic aerial-image objective function with specific weightings for contour pixels (or similar techniques) to account for the photoresist effect. The complexity therefore arises since the result is not exactly the same as the resist-image function in terms of side-lobe suppression. Moreover, tuning of the weighting coefficients is also not straightforward. Hence the applicability of such kinds of cost functions to SO and SMO is rather limited. Instead of using the conventional sigmoid function, our new linear resist model is designed to capture the threshold of images and monitor non-pattern regions for side-lobe

Optical Microlithography XXV, edited by Will Conley, Proc. of SPIE Vol. 8326, 83261W
© 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.916597

suppression. It can be shown that the associated objective functions are quadratic and guarantees the optimal source to be found by a conjugate gradient (CG) method within the number of iterations less than that of optimization parameters.

In Section 2, we will review the image formation in a partially coherent system and introduce the illumination cross coefficient (ICC) [15], which simplifies the forward imaging calculation and serves as a foundation for margin-based cost functions. Through local approximation of the nonlinear sigmoid function, we derive a linear resist model when the image intensity is near the resist threshold. We further analyze why the sigmoid in conjunction with CG may take longer iterations to converge. However, the problem no longer exists in our newly designed quadratic cost functions. The formulations are then applied for source optimization of a line array and a 15-bit SRAM in Section 3. We will demonstrate that our linear resist model in conjunction with CG is capable of achieving 100x speedup over the traditional approach, i.e. sigmoid-based steepest descent, without compromising image fidelity and process windows.

2. METHODOLOGY

2.1 Image formation

Lithography images, or so-called aerial images, can be simulated by Abbe's method [16, 17] which integrates the images formed by all source points incoherently.

$$I(x,y)=\int\int_{-\infty}^{\infty}J(u,v)\left[\left|\int\int_{-\infty}^{\infty}H(u+u',v+v')M(u',v')e^{-i2\pi[u'x+v'y]}du'dv'\right|^{2}\right]dudv, \tag{1}$$

where (x, y) and (u, v) denote the spatial coordinates and spatial frequencies of a mask, respectively. $J(u, v)$ is the strength of the point source located at (u, v) [17], $H(u, v)$ is the optical system transfer function, and $M(u, v)$ is the mask spectrum.

The optical system is band-limited [18], so the transfer function $H(u,v)$ can be described by a low-pass filter

$$\begin{cases} H(u,v)=1, \sqrt{u^{2}+v^{2}}\leq\dfrac{NA}{\lambda}, \\ H(u,v)=0, \mathit{otherwise}, \end{cases} \tag{2}$$

where NA/λ is the cut-off frequency of the optical system, NA is the numerical aperture which limits the largest oblique angle of rays forming the aerial image, and λ is the working wavelength.

In a partially coherent system, the finite source J(u,v) is limited by the coherent factor $\sigma(0<\sigma\leq1)$ [17, 19].

$$\begin{cases} J(u,v)\geq0, \sqrt{u^{2}+v^{2}}\leq\sigma\dfrac{NA}{\lambda}, \\ J(u,v)=0, \mathit{otherwise}. \end{cases} \tag{3}$$

The spectral integral in the bracket of Eq. (1) is the ICC [15]

$$ICC(x,y;u,v)=\left|\int\int_{-\infty}^{\infty}H(u+u',v+v')M(u',v')e^{-i2\pi[u'x+v'y]}du'dv'\right|^{2}. \tag{4}$$

Because ICC represents the image formed by a unit source with spatial frequencies (u,v), Eq. (1) can be interpreted as a linear superposition of images with coefficients $J(u,v)$.

For computing pixelated images, Eq. (1) should be discretized as

$$\mathrm{I}(i,j)=\sum_{k=1}^{S}\sum_{l=1}^{S}\mathrm{J}(k,l)\mathrm{ICC}(i,j;k,l),\quad i,j=1,2,...,N\,. \tag{5}$$

The variables i, j, k, and l denote the indices of discretized x, y, u, and v. N and S are the total sample numbers in spatial and spectral domains, respectively. Likewise the ICC in Eq. (4) can be discretized as

$$\mathrm{ICC}(i,j;k,l)=\left|\sum_{k'=1}^{S}\sum_{l'=1}^{S}\mathrm{H}(k+k',l+l')\mathrm{M}(k',l')e^{-i2\pi[u'(k,l)x(i,j)+v'(k,l)y(i,j)]}\right|^{2}. \tag{6}$$

To simplify the matrix computation, the 2-D discrete source can be converted to a 1-D vector and ICC can be expressed by a 2-D matrix. Consequently the output image is also a 1-D vector that can be converted to a 2-D distribution by rearranging the rank. Thus Eq. (5) can be represented as

$$\mathbf{I} = \mathbf{ICC}\,\mathbf{J}, \tag{7}$$

where the sizes of **I**, **J**, and **ICC** are $N^2\times1$, $S^2\times1$, and $N^2\times S^2$, respectively. Fig. 1 illustrates the aforementioned matrix operations.

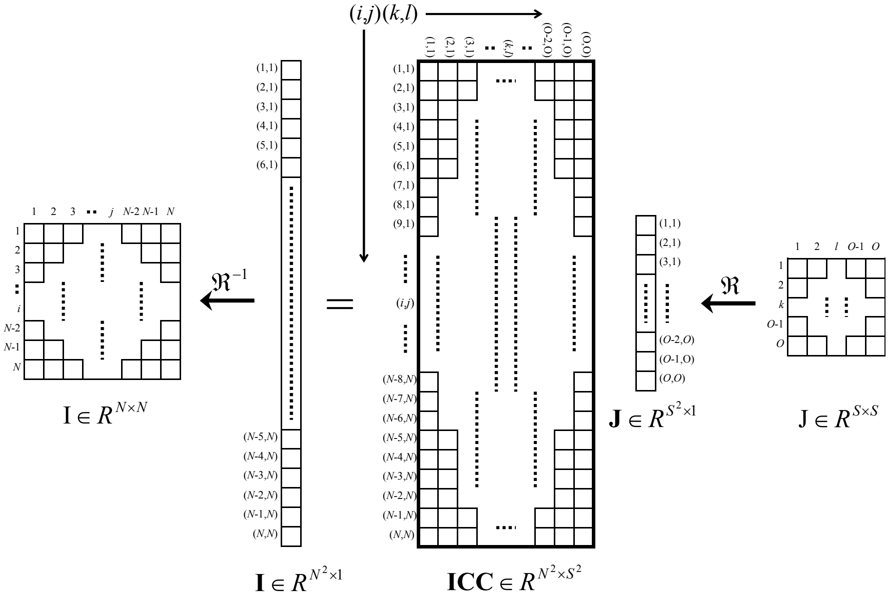


Fig. 1. Illustration of the matrix operations in Eq. (7). $\Re$ denotes the matrix to vector conversion. $\Re^{-1}$ denotes the vector to matrix conversion.

As a result, every row of the **ICC** matrix composes of one spatial image point by summing all row elements with the coefficients in **J**. To obtain the target image points, only the relative rows are required to be extracted from **ICC** for the computation. Thus **ICC** can be partitioned to several parts for different image configurations. For example, the images may be classified into the marginal and face parts. The former preserves the high spatial frequency fidelity and the latter controls the low spatial frequency response. Fig. 2 illustrates the operations of image formation of different parts. As an example, in Fig. 2(a) the inside marginal pixels are painted in light grey, outside marginal pixels in grey, and surrounding face pixels in black. The surrounding face pixels prevent the non-patterned images to be printed. In Fig. 2(b), the combinations of rows extracted from **ICC** form the new sub-**ICCs** for different image configurations. The light grey, grey and black color bands denote the rows corresponding to the pixels in Fig. 2(a). Finally in Fig. 2(c), the matrix operations represent the image formation of the combined images.

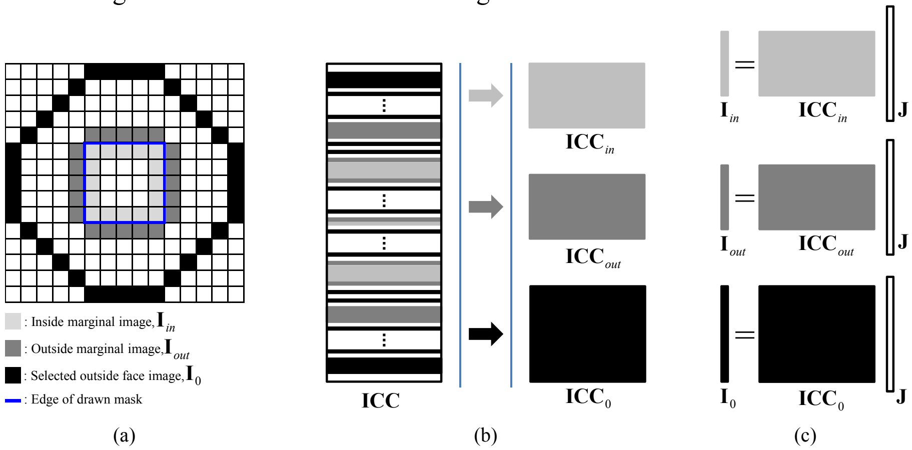


Fig. 2. Partitioned ICC for different parts of image formation. (a) Example of a pixelated square contact mask. (b) Row extractions and sub-ICC generation. (c) Matrix operations of various image formations by using sub-ICCs.

2.2 Resist models

To find the optimal source, the proper objective functions for quantifying the deviation between ideal designs and real simulations should first be defined. In general the final resist images are the uppermost concern. However, the resist images are usually simulated by the nonlinear *sigmoid* function [9-11], which destroys the benefit of linear operations in image formation. The sigmoid function has the following form

$$T(I) = \frac{1}{1+e^{-a(I-tr)}}, \tag{8}$$

where a characterizes the sensitivity of the photoresist and controls the slopes of sidewall profiles. tr is the parameter of the constant threshold level. Here, the value is set and normalized to 0.5. Fig. 3 illustrates such resist model's behavior which acts as a high pass filter.

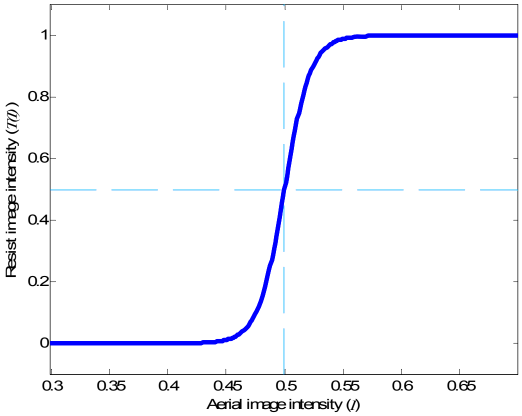


Fig. 3. Sigmoid function in Eq. (8) with a = 90 and tr = 0.5.

Eq. (8) can be expanded into Eq. (9) by a *Taylor* series when I is close to tr. Because such function has odd symmetry, even terms of Eq. (9) are eliminated.

$$\begin{aligned} T(I) &= T(tr) + \sum_{n=1}^{\infty} \left.\frac{T^{(2n-1)}(I)}{(2n-1)!}\right|_{tr} (I-tr)^{(2n-1)} + \sum_{n=1}^{\infty} \left.\frac{T^{(2n)}(I)}{(2n)!}\right|_{tr} (I-tr)^{(2n)} \\ &= T(tr) + \left.\frac{T'(I)}{1!}\right|_{tr} (I-tr) + H.O.T. \\ &\overset{I\to tr}{\cong} T(tr) + \frac{a}{4}(I-tr), \end{aligned} \tag{9}$$

where

$$T'(I) = aT(I)(1-T(I)), \tag{10}$$

$$T''(I) = a(1-2T(I))T'(I), \tag{11}$$

$$T'''(I) = a\left(-2T'(I)^2 + (1-2T(I))T''(I)\right). \tag{12}$$

Eq. (9) shows that $T(I)$ can be approximated by a linear function when I is within $tr\pm\Delta_1$, where $a^2/12(tr\pm\Delta_1 - tr)^2$ is equal to 0.15. Furthermore, $T(I)$ is replaced by 0 and 1 when $I < tr-\Delta_2$ and $I > tr+\Delta_2$, where $T(tr-\Delta_2)$ and $(1\text{-}T(tr+\Delta_2))$ are equal to 0.05. When $T(I)$ is near 0 or 1, the derivatives approach 0, which means $T(I)$ is near constant. Therefore, Eq. (8) can be approximated by the piecewise linear □(I)

$$\dot{T}(I)=\begin{cases} 0, \quad tr-I\geq\Delta_2 \\ \dfrac{T(tr)-\frac{a}{4}\Delta_1}{\Delta_2-\Delta_1}\left(I-(tr-\Delta_2)\right), \quad \Delta_1\leq(tr-I)\leq\Delta_2 \\ T(tr)+\frac{a}{4}(I-tr), \quad |I-tr|\leq\Delta_1 \\ T(tr)+\frac{a}{4}\Delta_1+\dfrac{1-\left(T(tr)+\frac{a}{4}\Delta_1\right)}{\Delta_2-\Delta_1}\left(I-(tr+\Delta_1)\right), \quad \Delta_1\leq(I-tr)\leq\Delta_2 \\ 1, \quad I-tr\geq\Delta_2 \end{cases} \tag{13}$$

where $\Delta_1 = 0.0149$ and $\Delta_2 = 0.0327$. Fig. 4 illustrates the line segments of Eq. (13). The constant equations in Eq. (13) correspond to stable regions which are linear. The linear equations adjacent to constant equations denote the active regions which are nonlinear. The transition region with quasi-linearity between two active regions is approximated by the third linear equation of Eq. (13).

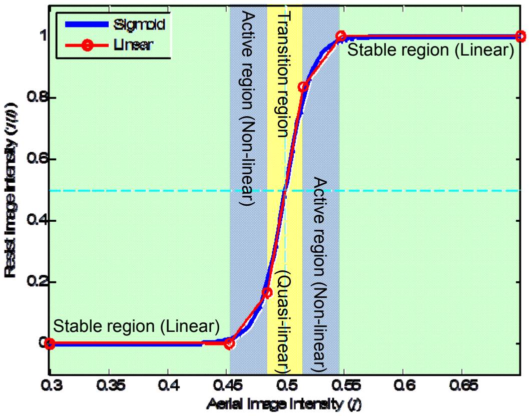


Fig. 4. Resist models of Eq. (8) (Blue) and Eq. (13) (Red+O).

2.3 Objective functions

For any position, we define the cost of resist image as

$$f_R(I)=\left(T(I_t)-T(I)\right)^2. \tag{14}$$

where I_t is the target aerial image which can be elaborately designed according to the geometric shapes of patterns [20-22]. So the overall cost using sigmoid model can be presented by following objective function

$$F_R=\iint f_R(I)dxdy. \tag{15}$$

Similarily, Eq. (14) can be expanded by a *Taylor* series when I is close to tr.

$$\begin{aligned} f_R(I)&=f_R(tr)+\left.\frac{f_R'(I)}{1!}\right|_{tr}(I-tr)+\left.\frac{f_R''(I)}{2!}\right|_{tr}(I-tr)^2+H.O.T. \\ &\overset{I\to tr}{\cong} f_R(tr)-\frac{a}{2}\left(T(I_t)-T(I)\right)(I-tr)+\frac{a^2}{16}(I-tr)^2 \\ &=f_R(tr)+a\left(\frac{a}{16}-\frac{1}{2}\frac{\left(T(I_t)-T(I)\right)}{(I-tr)}\right)(I-tr)^2, \end{aligned} \tag{16}$$

where

$$f_R'(I) = -2\left(T(I_t) - T(I)\right)T'(I), \tag{17}$$

$$f_R''(I) = -2a\left(-T(I)\left(1-T(I)\right) + \left(T(I_t) - T(I)\right)\left(1-T(I)\right) - \left(T(I_t) - T(I)\right)T(I)\right)T'(I). \tag{18}$$

Eq. (16) is not in a purely quadratic form because the coefficient of $(I-tr)^2$ is a function of *I* and *tr*. However, if the relative locations on target are near the drawn edges where I_t is also close to *tr*, *L'Hopital's* rule can be applied to simplify Eq. (16).

$$\lim_{I_t, I \to tr} \frac{\left(T(I_t) - T(I)\right)}{(I - tr)} = \lim_{I_t, I \to tr} \frac{-T'(I)}{1} = -\frac{a}{4}. \tag{19}$$

Therefore, when only marginal places are considered and image intensities in such places are close to *tr,*

$$f_R(I) = f_R(tr) + \frac{3a^2}{16}(I - tr)^2. \tag{20}$$

The above equation has a quadratic form where the minimum is at *I* equals *tr*. Inspired by Eq. (16), (19), and (20), we devise a new objective function

$$F_M = \oint_{\zeta_1} (I - tr)^2\, dxdy, \tag{21}$$

where ζ_1 denotes all contours along margins of drawn patterns. From another point of view, Eq. (21) only monitors the cost in drawn edges regardless of other places. Such formulation is similar to Sayegh's idea [23], except our objective function is second order instead of arbitrary orders. In a sense, Eq. (21) is not sensitive to image slopes for matching the sigmoid characteristic in the transition region. Fig. 5 compares costs of two images with different slopes across drawn edges by evaluating the marginal intensity and the resist image, respectively. The differences in slopes do not contribute to the deviation in costs for both cases. Therefore, the blue and green image profiles have approximate costs no matter the sigmoid or the marginal cost is applied.

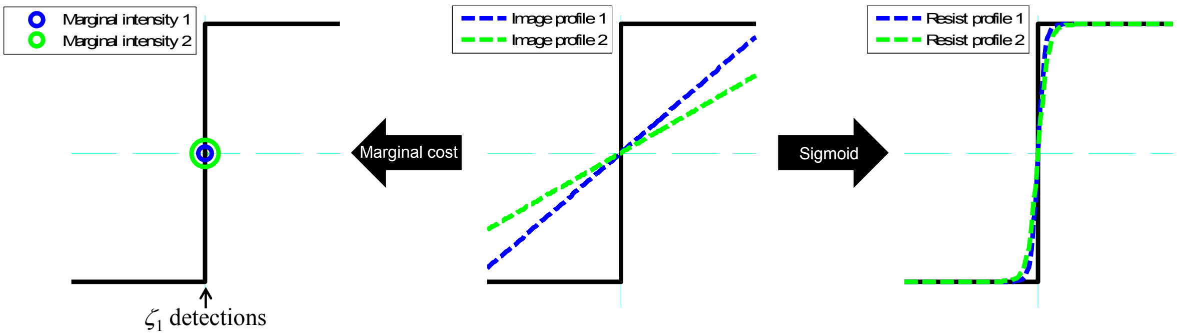
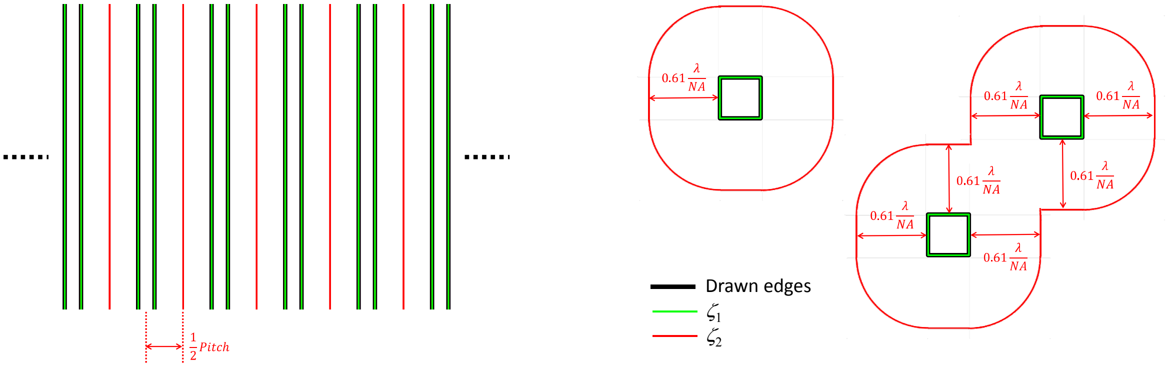


Fig. 5. Cost inspections of images with various slopes using marginal costs and resist images. Black bold lines and cyan dashed lines show ideal target profiles and threshold level, respectively.

However, Eq. (21) only evaluates the deviation of edge images from the threshold by quasi-linear effect in transition region of sigmoid. To be close to the sigmoid resist model, images should be also monitored in non-pattern regions where the over-threshold images lead to extra costs, but under-threshold images have zero costs. Thus to suppress the image in non-pattern regions, the images of closed curves surrounding the drawn features should be incorporated into optimization. A large amount of non-pattern images leads to low yield due to undesired resist images, or so-called side-lobes. Therefore, the objective function for side-lobe printing can be designed as

$$F_0 = \oint_{\zeta_2} (I - \delta)^2\, dxdy, \tag{22}$$

where ζ_2 denotes the contours enclosing all drawn patterns with a distance. Such distance is half a pitch for periodic patterns and $0.61\lambda/NA$ for isolated and semi-isolated patterns. The half-pitch is associated with the minimum intensity of periodic patterns. $0.61\lambda/NA$ is associated with the first minimum of diffractive patterns for a circular aperture. δ is chosen to be as small as possible, but should remain positive. Fig. 6 illustrates the configurations of ζ_1 and ζ_2 for periodic, isolated and semi-isolated patterns.



(a) (b)

Fig. 6. ζ_1 and ζ_2 of (a) periodic patterns, and (b) isolated and semi-isolated patterns.

The sigmoid function acts as a high pass filter where the images will be converted to 0 or 1 beyond the transition region. Such conversion is highly nonlinear. Eq. (22) cannot simulate well, but has a similar trend that prefers to suppress side-lobe printing for minimizing F_0. Fig. 7 shows the cost evaluations by employing non-pattern costs and resist images. While all images over the threshold are being developed, they induce costs to warn the appearance of side-lobe printing. Conversely, all images under the threshold have nearly zero costs. Therefore, various solutions are allowed to have identical threshold contours, but with different image distributions beyond edges.

Comparing to the sigmoid, the non-pattern cost function shows differences no matter the intensities of both curves are over thresholds or not. Eq. (22) checks deviations from δ at the theoretical minimum in aerial image profiles. Therefore, costs vary with different image distributions which are highly sensitive to pattern configurations. This phenomenon causes slight differences in final optimal sources obtained from our approach and the sigmoid model, but it is still practical to use Eqs. (21) and (22) for simulating the resist image costs. However, Eq. (21) alone is not sufficient to reach an optimal sources as good as the sigmoid for dense patterns.

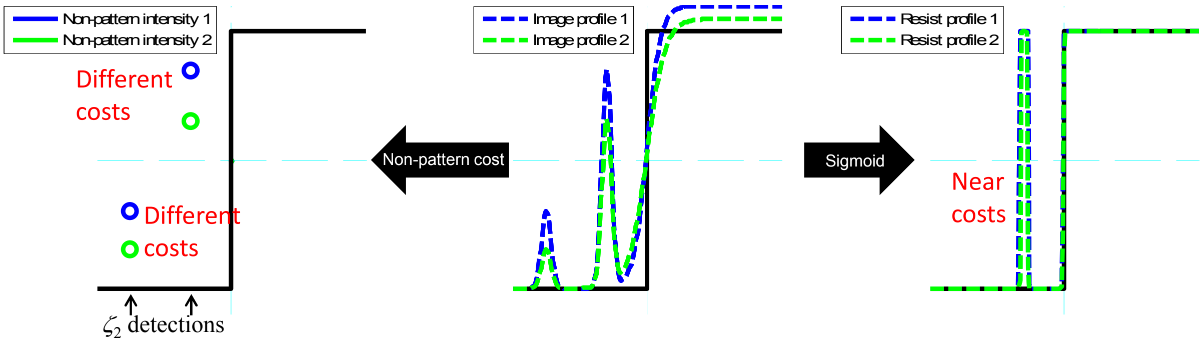


Fig. 7. Cost inspections of images outside drawn patterns with over- and under-threshold intensities by using non-pattern costs and resist images. For brevity, there are two ζ_2 detection results in left sub-plot where one has under-threshold images and the other has over-threshold images. Black bold lines and cyan dashed lines show ideal target profiles and threshold level, respectively.

Finally, in terms of pixelated image calculations, Eq. (8) and (15) can be re-written as

$$T(\mathbf{I}) = \frac{1}{1+e^{-a(\mathbf{I}-tr)}} = \frac{1}{1+e^{-a(\mathbf{ICC}\,\mathbf{J}-tr)}}, \tag{23}$$

and

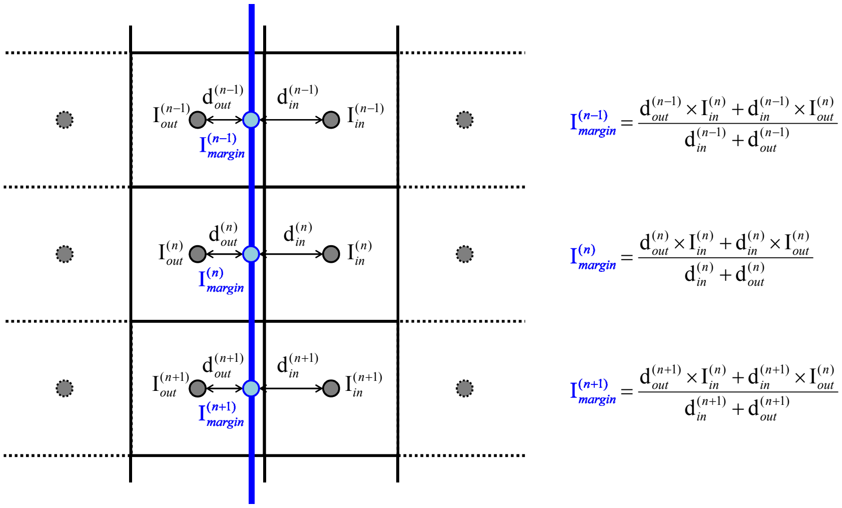
$$F_R = \left\| T(\mathbf{I}_t) - T(\mathbf{I}) \right\|^2, \tag{24}$$

where ||·|| is the operation of *Euclidean* norm. $\mathbf{I}_t$ is the numerical sampling matrix of I_t.

For the same reason, Eq. (21) can be re-written as

$$F_M = \left\| \left((\mathbf{d}_{in} + \mathbf{d}_{out})^{-1} \cdot (\mathbf{d}_{out} \cdot \mathbf{ICC}_{in} + \mathbf{d}_{in} \cdot \mathbf{ICC}_{out}) \right) \mathbf{J} \text{ - } tr \right\|^2, \tag{25}$$

where · denotes pixel-wise multiplication. $(\cdot)^{-1}$ is the operator which replaces the vector elements by their own reciprocals. Both the sizes of $\mathbf{ICC}_{in}$ and $\mathbf{ICC}_{out}$ are $N' \times S^2$, which is the number of pixels on the margins of drawn patterns. $\mathbf{d}_{in}$ ($\mathbf{d}_{out}$) is an array which records the distances from inside (outside) marginal grid points to edges. The marginal images are obtained from weighted interpolations of inner and outer marginal grid images. Fig. 8 illustrates the above operation.



$\mathrm{I}_{out}^{(n-1)}$
$\mathrm{d}_{out}^{(n-1)}$
$\mathrm{d}_{in}^{(n-1)}$
$\mathrm{I}_{in}^{(n-1)}$
$\mathrm{I}_{margin}^{(n-1)}$
$\mathrm{I}_{margin}^{(n-1)} = \frac{\mathrm{d}_{out}^{(n-1)} \times \mathrm{I}_{in}^{(n)} + \mathrm{d}_{in}^{(n-1)} \times \mathrm{I}_{out}^{(n)}}{\mathrm{d}_{in}^{(n-1)} + \mathrm{d}_{out}^{(n-1)}}$
$\mathrm{I}_{out}^{(n)}$
$\mathrm{d}_{out}^{(n)}$
$\mathrm{d}_{in}^{(n)}$
$\mathrm{I}_{in}^{(n)}$
$\mathrm{I}_{margin}^{(n)}$
$\mathrm{I}_{margin}^{(n)} = \frac{\mathrm{d}_{out}^{(n)} \times \mathrm{I}_{in}^{(n)} + \mathrm{d}_{in}^{(n)} \times \mathrm{I}_{out}^{(n)}}{\mathrm{d}_{in}^{(n)} + \mathrm{d}_{out}^{(n)}}$
$\mathrm{I}_{out}^{(n+1)}$
$\mathrm{d}_{out}^{(n+1)}$
$\mathrm{d}_{in}^{(n+1)}$
$\mathrm{I}_{in}^{(n+1)}$
$\mathrm{I}_{margin}^{(n+1)}$
$\mathrm{I}_{margin}^{(n+1)} = \frac{\mathrm{d}_{out}^{(n+1)} \times \mathrm{I}_{in}^{(n)} + \mathrm{d}_{in}^{(n+1)} \times \mathrm{I}_{out}^{(n)}}{\mathrm{d}_{in}^{(n+1)} + \mathrm{d}_{out}^{(n+1)}}$

Fig. 8. Illustration of the marginal image calculation. The black solid lines sketch the inside and outside marginal pixel. The blue solid line shows the one of drawn pattern edges. The back circles painted by gray are grid point positions. The blue circles painted by light blue are sampled marginal images interpolated by using relative marginal images. $\mathrm{I}_{in}^{(n)}$ ($\mathrm{I}_{out}^{(n)}$) denotes *n*th inside (outside) marginal image which is calculated by *n*th row of $\mathbf{ICC}_{\mathrm{in}}$ ($\mathbf{ICC}_{\mathrm{out}}$) multiplying **J**. $\mathrm{d}_{in}^{(n)}$ ($\mathrm{d}_{out}^{(n)}$) denotes the distance of *n*th inner (outer) grid point to the edge. As a result the weightings of $\mathrm{I}_{in}^{(n)}$ and $\mathrm{I}_{out}^{(n)}$ to from $\mathrm{I}_{m\arg in}^{(n)}$ are inversely proportional to the distances of $\mathrm{d}_{in}^{(n)}$ and $\mathrm{d}_{out}^{(n)}$.

Therefore for on-grid patterns, $\mathbf{d}_{in}$ is equal to $\mathbf{d}_{out}$, Eq. (25) can be simplified as

$$F_M = \left\| \frac{1}{2} \left(\mathbf{ICC}_{in} + \mathbf{ICC}_{out} \right) \mathbf{J} \text{ - } tr \right\|^2. \tag{26}$$

Moreover, Eq. (22) can be re-written as

$$F_0 = \left\| \mathbf{ICC}_0 \, \mathbf{J} \text{ - } \delta \right\|^2, \tag{27}$$

where the size of $\mathbf{ICC}_0$ is $N'' \times S^2$ and N'' denotes the number of chosen pixels surrounding the drawn patterns. In Eq. (27), the curves of ζ_2 are discretely sampled by on-grid points. Such curves whose widths are one pixel are like the one composed of the black pixels in Fig. 2(a).

By defining objective functions on critical parts of the image and reformulating the sigmoid as a linear operation, the computational cost of SO is significantly reduced.

2.4 Optimization

In terms of matrix operation, Eqs. (26) and (27) can be rewritten as

$$F_M = \left(\frac{1}{2}\left(\mathbf{ICC}_{in} + \mathbf{ICC}_{out}\right)\mathbf{J} - \mathbf{tr}\right)^{\mathrm{T}}\left(\frac{1}{2}\left(\mathbf{ICC}_{in} + \mathbf{ICC}_{out}\right)\mathbf{J} - \mathbf{tr}\right), \tag{28}$$

$$F_0 = (\mathbf{ICC}_0\ \mathbf{J} - \boldsymbol{\delta})^{\mathrm{T}}(\mathbf{ICC}_0\ \mathbf{J} - \boldsymbol{\delta}), \tag{29}$$

where T denotes transpose operation, $\mathbf{tr} = tr\times[1,\ldots,1]^{\mathrm{T}}$, and $\boldsymbol{\delta} = \delta\times[1,\ldots,1]^{\mathrm{T}}$. The sizes of **tr**, and $\boldsymbol{\delta}$ are $N'\times1$, and $N''\times1$, respectively. Thus the resist image cost by linear model is obtained by linearly superposing F_M, and F_0 with the coefficients c_1, and c_2

$$F_L = c_1 F_M + c_2 F_0. \tag{30}$$

The value of c_1/c_2 can be decided by the reciprocal of average length ratio of ζ_1 and ζ_2.

$$\frac{c_1}{c_2} = \frac{|\zeta_2|_{\mathrm{a}}}{|\zeta_1|_{\mathrm{a}}}, \tag{31}$$

where $|\zeta_1|_{\mathrm{a}}$ and $|\zeta_2|_{\mathrm{a}}$ denote the average length of ζ_1 and ζ_2, respectively. Consequently, the optimal source $\hat{\mathbf{J}}$ can be defined as an argument for minimizing Eq. (30)

$$\hat{\mathbf{J}} = \mathbf{arg}\min_{\mathbf{J}}.\{F_L\}. \tag{32}$$

Moreover, Eq. (30) can be expanded to have a quadratic form

$$F_L = \mathbf{J}^{\mathrm{T}}\mathbf{QJ} - \mathbf{b}^{\mathrm{T}}\mathbf{J} + c, \tag{33}$$

where

$$\mathbf{Q} = \frac{c_1}{4}\left(\mathbf{ICC}_{in} + \mathbf{ICC}_{out}\right)^{\mathrm{T}}\left(\mathbf{ICC}_{in} + \mathbf{ICC}_{out}\right) + c_2\mathbf{ICC}_0^{\mathrm{T}}\mathbf{ICC}_0, \tag{34}$$

$$\mathbf{b} = c_1\left(\mathbf{ICC}_{in} + \mathbf{ICC}_{out}\right)^{\mathrm{T}}\mathbf{tr} + 2c_2\mathbf{ICC}_0^{\mathrm{T}}\boldsymbol{\delta}, \tag{35}$$

$$c = c_1\mathbf{tr}^{\mathrm{T}}\mathbf{tr} + c_2\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\delta}. \tag{36}$$

The sizes of **Q**, **b** and c are $S^2\times S^2$, $S^2\times1$, and 1×1, respectively. Because the overall cost function is quadratic, the optimal source $\hat{\mathbf{J}}$ is guaranteed to be found by conjugate-gradient (CG) method with no more than S^2 iterations [24, 25]. The algorithm can be summarized by the pseudo-code in Table 1.

Table 1. Pseudo-code of CG

```
Algorithm 1. SO by CG
Input:
Load initial source $\mathbf{J}^{(0)}$. Set $k = 0$.
Calculation :
1. $\mathbf{g}^{(0)} = \nabla_{\mathbf{J}}F_L(\mathbf{J}^{(0)})^{\dagger}$. If $\mathbf{g}^{(0)} < \varepsilon^{\ddagger}$, stop; else, set $\mathbf{d}^{(0)} = -\mathbf{g}^{(0)}$.
2. $\alpha_k = -(\mathbf{g}^{(k)\mathrm{T}}\mathbf{d}^{(k)})/(\mathbf{d}^{(k)\mathrm{T}}\mathbf{Q}\mathbf{d}^{(k)})$.
3. $\mathbf{J}^{(k+1)} = \mathbf{J}^{(k)} + \alpha_k\mathbf{d}^{(k)}$.
4. Set all negative entries of $\mathbf{J}^{(k+1)}$ are equal to 0.
5. $\mathbf{g}^{(k+1)} = \nabla_{\mathbf{J}}F_L(\mathbf{J}^{(k+1)})$. If $\|\mathbf{g}^{(k+1)}\| < \varepsilon$, stop; Set $\hat{\mathbf{J}} = \mathbf{J}^{(k+1)}$.
6. $\beta_k = (\mathbf{g}^{(k+1)\mathrm{T}}\mathbf{Q}\mathbf{d}^{(k)})/(\mathbf{d}^{(k)\mathrm{T}}\mathbf{Q}\mathbf{d}^{(k)})$.
7. $\mathbf{d}^{(k+1)} = -\mathbf{g}^{(k+1)} + \beta_k\mathbf{d}^{(k)}$. Set $k = k + 1$; Go to step 2.
Output :
Export optimal source $\hat{\mathbf{J}}$.
```

[†] $\mathbf{g} = 2(\mathbf{QJ} - \mathbf{b})$. $\nabla_{\mathbf{J}} = [\partial/\partial\mathbf{J}(1,1), \partial/\partial\mathbf{J}(2,1), \ldots, \partial/\partial\mathbf{J}(S,S)]^{\mathrm{T}}$.

[‡] ε is an extremely small value, but positive.

Furthermore, to verify the effectiveness of our algorithm, the sigmoid-based SO results are also calculated. Likewise the optimal source by sigmoid model $\hat{\mathbf{J}}'$ can be defined as an argument for minimizing Eq. (24) and like Eq. (31)

$$\hat{\mathbf{J}}' = \mathbf{arg}\min_{\mathbf{J}} .\{F_R\}. \tag{37}$$

However, Eq. (24) is not in a quadratic form and the algorithm presented in Table 1 is not applicable. Although several modified CG methods have been proposed to address non-quadratic problems [25], there are two drawbacks limiting the application of CG in high-order problems. First, the Step 2 in Table 1 that decides α_k is a time-consuming one dimension optimization problem [25, 26]. Second, β_k in Step 6 of Table 1 approximated by a quadratic function is not accurate enough to characterize the nonlinearity of the sigmoid model. That leads to extra iterations for converging.

Thus the steepest-descent (SD) algorithm [25] widely used in inverse mask optimization with sigmoid resist model [11, 21, 27, 28] is performed to compute $\hat{\mathbf{J}}'$. Table 2 summarizes the steps of the SD algorithm.

Table 2. Pseudo-code of SD

```
Algorithm 2. SO by SD
Input :
Load initial source $\mathbf{J}^{(0)}$. Set $k = 0$.
Calculation :
1. $\mathbf{g}'^{(k)} = \nabla_{\mathbf{J}} F_R(\mathbf{J}^{(k)})^{\dagger}$.
2. $\mathbf{J}^{(k+1)} = \mathbf{J}^{(k)} + \gamma \mathbf{g}'^{(k)}$.
3. Set all negative entries of $\mathbf{J}^{(k+1)}$ are equal to 0.
   If $F_R(\mathbf{J}^{(k+1)}) > F_R(\mathbf{J}^{(k)})$, $\gamma = \alpha'\gamma^{\aleph}$; Set $\mathbf{J}^{(k+1)} = \mathbf{J}^{(k)}$.
   Elseif $F_R(\mathbf{J}^{(k)}) - F_R(\mathbf{J}^{(k+1)}) < \varepsilon^{\ddagger}$, stop; Set $\hat{\mathbf{J}}' = \mathbf{J}^{(k+1)}$.
 Else set $k = k + 1$; Go to step 1.
Output :
Export optimal source $\hat{\mathbf{J}}'$.
```

† $\mathbf{g}' = -2a\mathbf{I}\mathbf{C}\mathbf{C}^{\mathrm{T}}[(T(\mathbf{I}_t)-T(\mathbf{I}))\cdot(1-T(\mathbf{I}))\cdot T(\mathbf{I})]$. · is the pixel-wise multiplication. $\nabla_{\mathbf{J}} = [\partial/\partial\mathbf{J}(1,1), \partial/\partial\mathbf{J}(2,1), \ldots, \partial/\partial\mathbf{J}(S,S)]^{\mathrm{T}}$.

ℵ $\alpha' < 1$. $\gamma = 5$ and $\alpha' = e^{-0.5}$ in our work.

‡ ε is an extremely small value, but positive.

The computational complexities of SO by CG and SD incorporating linear and sigmoid models are $O((S^2\times\pi/4)^2\times K_{\mathrm{CG}})$ and $O(N^2\times(S^2\times\pi/4)\times K_{\mathrm{SD}})$, respectively, where K_{CG} and K_{SD} are iteration numbers of CG and SD. Theoretically the speed t^{-1} is inverse proportional to the complexity which is proportional to elapsed time t of computation. Thus the speed ratio of CG comparing to SD is $t_{\mathrm{SD}}/t_{\mathrm{CG}}$ which can be formulated as $\kappa_1\times N^2/(S^2\times\pi/4)\times K_{\mathrm{SD}}/K_{\mathrm{CG}}$. κ_1 is a constant which depends on the programming efficiency of different algorithms. Usually $N^2/(S^2\times\pi/4)$ and $K_{\mathrm{SD}}/K_{\mathrm{CG}}$ are both much larger than one, which implies CG being a more efficient approach.

Finally we define the functions $\mathrm{DoPE}_{\mathrm{J}}$ and $\mathrm{DoPE}_{\mathrm{I}}$ to quantify the difference between sources and aerial images, respectively, where DoPE stands for the degree of pattern error.

$$\mathrm{DoPE}_{\mathrm{J}} = \frac{1}{2}\sum\left|\frac{\mathbf{J}_2}{\sum\mathbf{J}_2} - \frac{\mathbf{J}_1}{\sum\mathbf{J}_1}\right|,\ \ \mathbf{J}_1, \mathbf{J}_2 \geq 0. \tag{38}$$

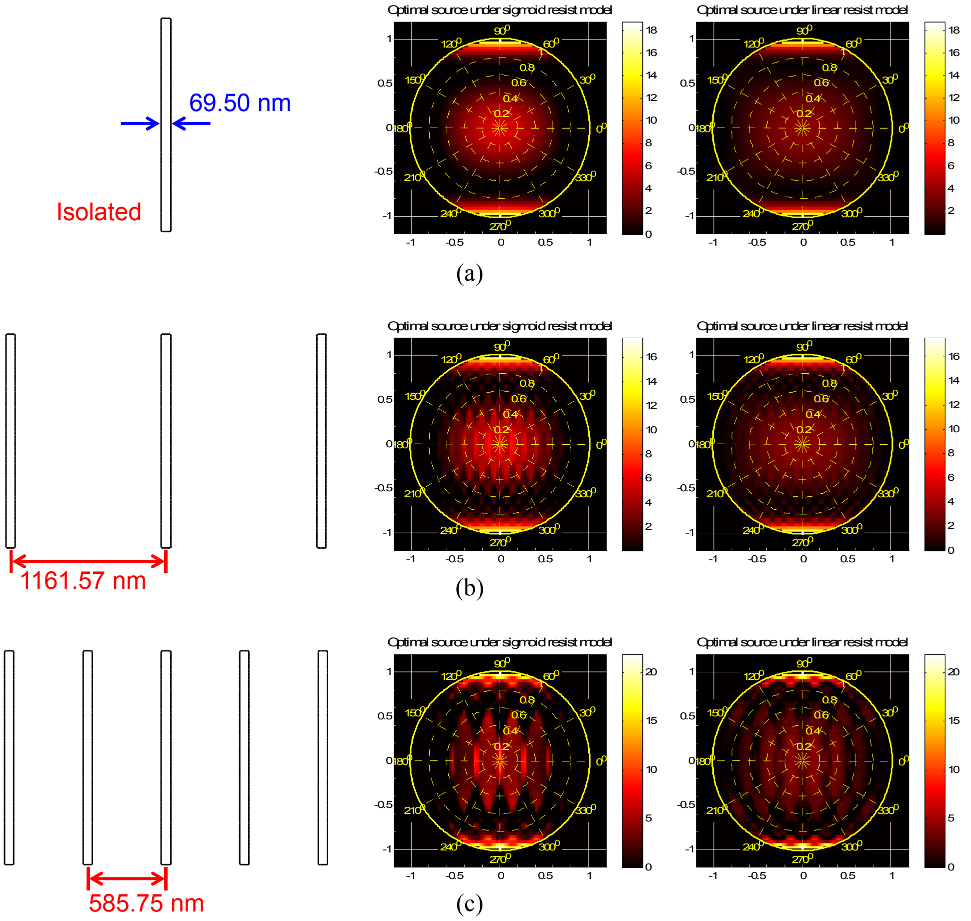
$$\mathrm{DoPE}_{\mathrm{I}} = \frac{1}{2}\sum\left|\frac{\mathbf{I}_2}{\sum\mathbf{I}_2} - \frac{\mathbf{I}_1}{\sum\mathbf{I}_1}\right|,\ \ \mathbf{I}_1, \mathbf{I}_2 \geq 0. \tag{39}$$

By this way, the DoPEs are in the range of [0, 1].

3. RESULTS AND DISCUSSION

The source and drawn mask templates are composed of 65×65 and 256×256 pixels with S and N equal to 65 and 256, respectively. The pixel size is 9.93×9.93 nm^2. a and δ are equal to 90 and 0, respectively. The illumination and projection system have working wavelength λ, numerical aperture NA, and coherent factor σ equal to 193 nm, 1.35, and 0.9, respectively. In order to make a fair comparison, all images are normalized by integration of a full-open source with unit intensity and blank mask. All sources coordinates are normalized by $\sigma NA/\lambda$.

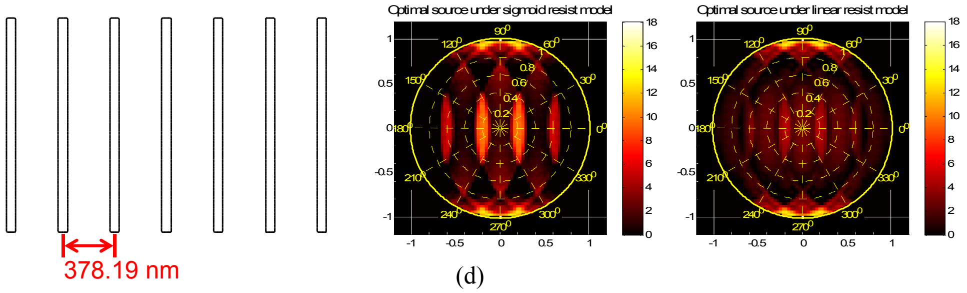
First, we run simulations for line arrays with increasing densities using only Eq. (21) without Eq. (22). The CD of each line is equal to 69.50 nm. Fig. 9 lists the simulation results where optimal sources using two different resist image cost functions are quite similar for large pitches, but suffer from severe deviations for dense patterns. Because the simulations are in finite regions, the dense patterns are not really periodic. Thus the final optimal sources using both resist models are not like conventional horizontal dipole sources used for periodic line arrays. Fig. 10 illustrates the DoPE$_J$ of the optimal sources in Figs. 9(a) – (j). The horizontal axis is reciprocal of pitch whose value is zero for the isolated pattern. As expected, the isolated and sparse configurations have smaller DoPE$_J$ than that of dense configurations. Such results verify that using Eq. (21) alone is not sufficient to approximate the sigmoid function well for dense patterns in SO.



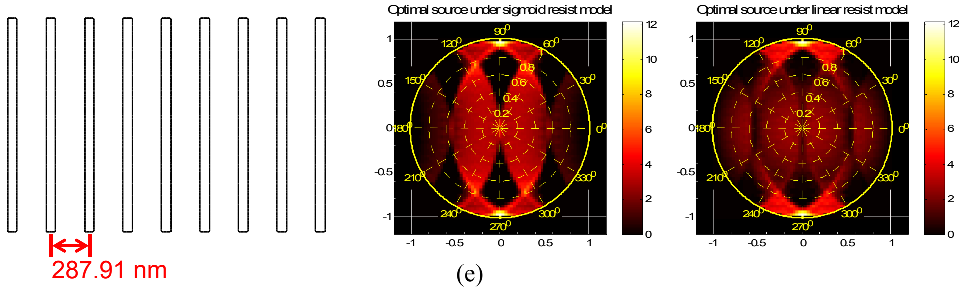
(a)

(b)

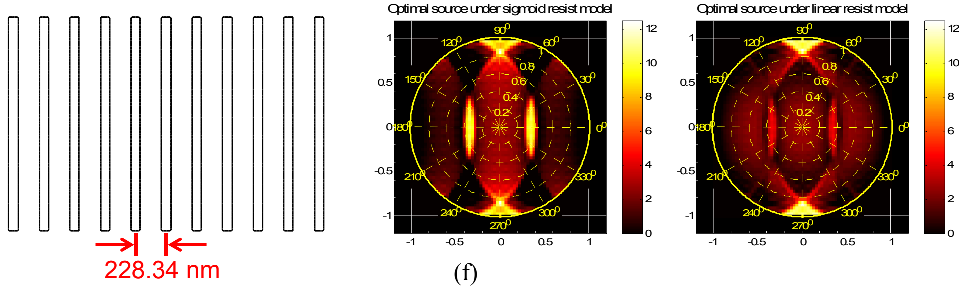
(c)


378.19 nm
Optimal source under sigmoid resist model
Optimal source under linear resist model

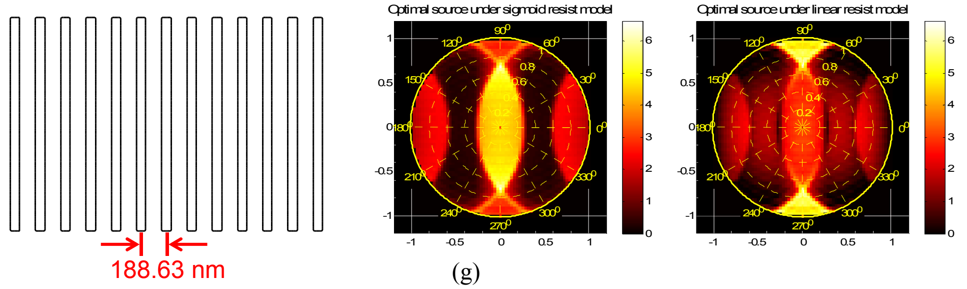
(d)


287.91 nm
Optimal source under sigmoid resist model
Optimal source under linear resist model

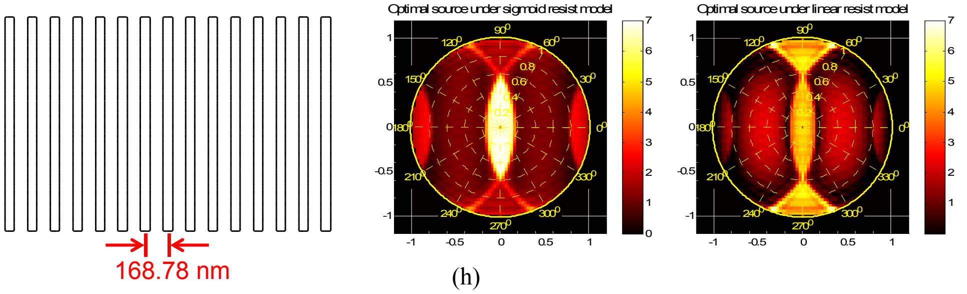
(e)


228.34 nm
Optimal source under sigmoid resist model
Optimal source under linear resist model

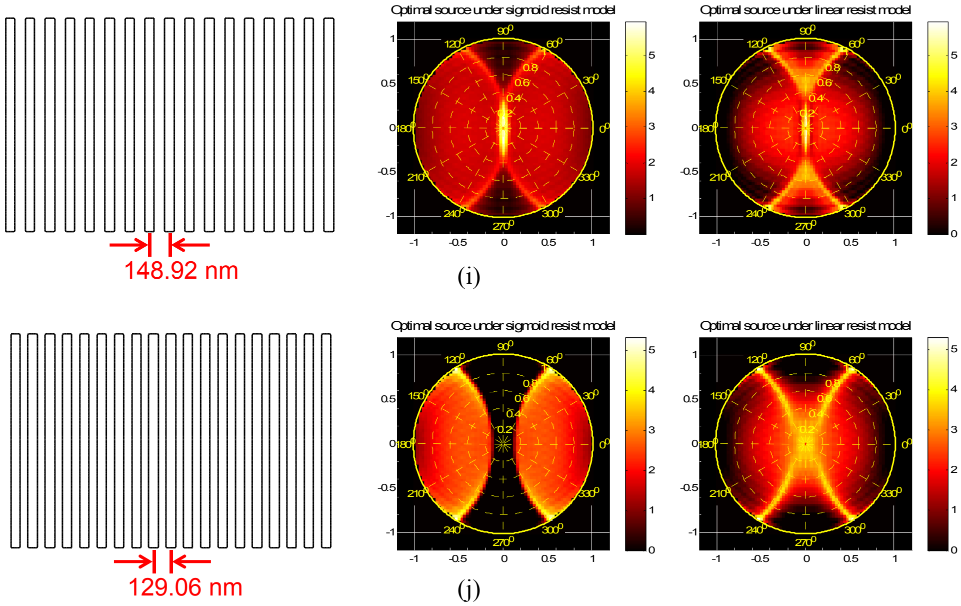
(f)


188.63 nm
Optimal source under sigmoid resist model
Optimal source under linear resist model

(g)


168.78 nm
Optimal source under sigmoid resist model
Optimal source under linear resist model

(h)



(i)

(j)

Fig. 9. Line arrays and SO results for investigating the similarity of optimal sources using only the marginal image cost.

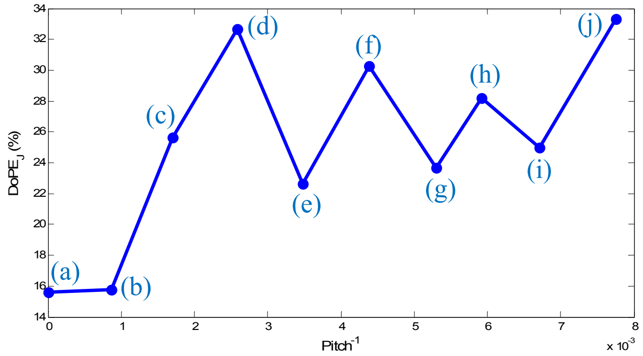
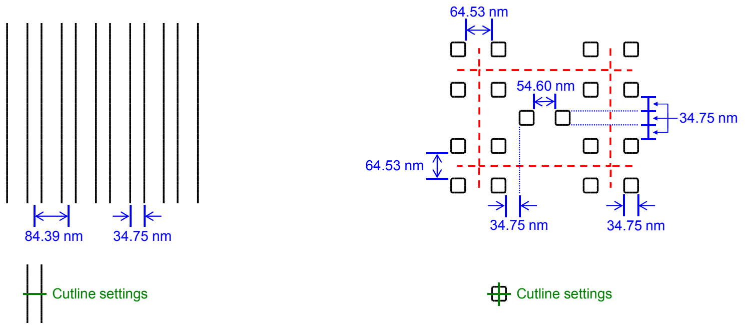


Fig. 10. DoPE$_J$ of optimal sources calculated by different resist models in Figs. 9(a) – (j).

Consequently, we perform detailed analysis for more critical 1-D and 2-D patterns. Two demo mask configurations are poly line array and 15-bit SRAM as shown in Fig. 11, where the pixel size is 4.96×4.96 nm^2. In the following simulations, c1/c2 ratios are set to be 3 for the two masks, respectively. The surrounding pixels for side-lobe checking are placed in the distance of half pitch from the center of every poly line and 0.61λ/NA from the edges of every SRAM's square contact.



(a) (b)

Fig. 11. (a) Poly line array and (b) Unit cell of 15-bit SRAM used for source optimization.

Fig. 12 illustrates the experimental results using the poly line array in Fig. 11(a). The optimal sources of both models are very close. Two energy concentrated arches horizontally locate at $\pm 0.85\sigma NA/\lambda$ along 0°−180° axis within ±30° and 180°±30° in both optimal sources. Such optimal sources match the conventional off-axis illumination (OAI) source that is empirically designed for enhancing the resolution of periodical line array. In a sense, it shows our optimization algorithms match the real physics and practical applications. Similarly, Fig. 13 shows the experimental results using the 15-bit SRAM with the unit cell shown in Fig. 11(b). Both sources are quite similar and have several energy poles where energy is concentrated. Such poles are around the center and circumferences of source pupils. The energy concentration areas generate interfered spatial frequencies which match the spatial distribution of masks. The places besides energy concentration regions are background parts that have a few deviations between both optimal sources. Such parts provide near constant intensity distributions for basis that biases the vibrations interfered by high energy-density places of sources across the threshold *tr*.

As a result, our linear model can model the resist reactions well using Eq. (26) while the image intensity along drawn edges is near the threshold. The threshold-only awareness characteristic of Eq. (26) is sensitive to the average intensity of the two adjacent pixels in and out the drawn edges. Therefore, no matter how sharp the image slops are in drawn edges, the costs are the same when the drawn edge image intensities have no change. That is consistent with sigmoid model in which any image intensity of two adjacent pixels across *tr* will be converted to just 1 or 0. Nevertheless the images away from drawn edge locations have some deviations as using Eq. (27). Such results are predictable because Eq. (27) only forces the intensity of the surrounding rings to δ, that is not like sigmoid as a high pass filter. Moreover, Eq. (24) shows that the cost of any image intensity above (or below) *tr* is equal in one location, but Eq. (27) does not. Furthermore, every drawn pattern has only one side-lobe checking ring whose monitoring ranges are much smaller than the sigmoid model.

In terms of images, they also show similar threshold contours, where the magenta and green curves are associated with sigmoid and linear models, respectively. Moreover, the overall image qualities are excellent after SO although the patterns are still not on the target (black curves), but close. Such results relieve the load of MO and lead to simpler OPC masks.

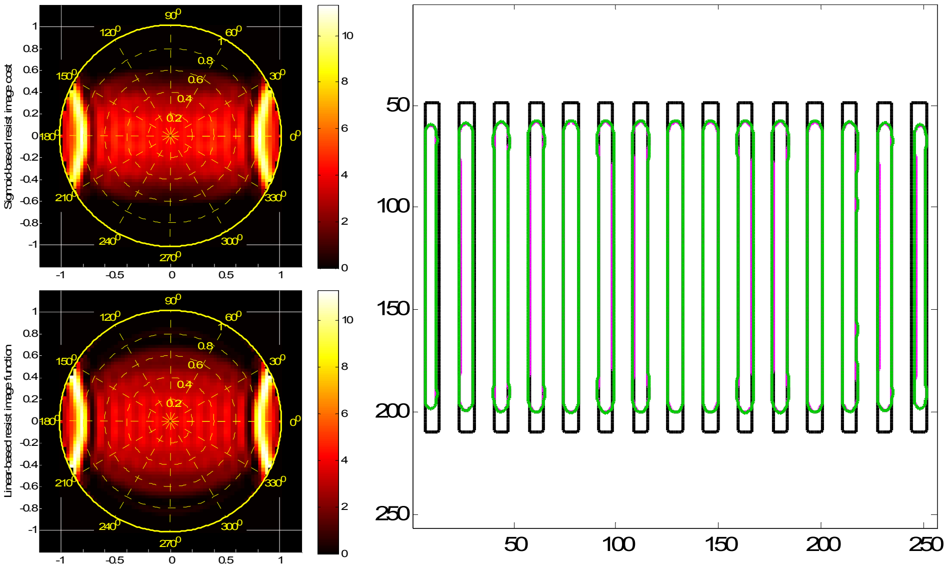


Fig. 12. Optimal sources of the poly line array in Fig. 11(a).

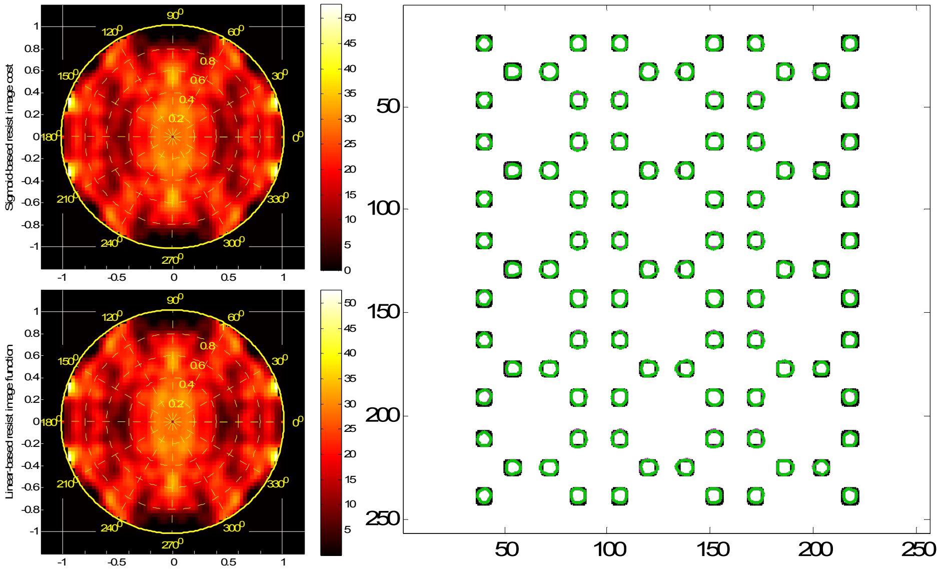


Fig. 13. Optimal sources of the 15-SRAM in Fig. 11(b).

Finally we quantize the similarity of different optimal sources and their images by using Eqs. (38) and (39). The average edge placement errors (EPEs) and normalized image log slopes (NILSs) are also calculated according to cutline settings in Figs. 11(a) and (b).

In Table 3, the average EPEs of both mask structures using two different resist models are close and near the pixel size. Likewise average NILSs are also close. Moreover, $\mathrm{DoPE_J}$ of both masks are 8.86% and 2.88%, which implies the optimal sources using two resist models are highly similar. Although $\mathrm{DoPE_J}$s of both masks are more than 2%, $\mathrm{DoPE_I}$s are less than 0.39%. Such phenomenon indirectly verifies that aerial images are quite insensitive to large defects on sources [29, 30]. The above merit matches one of the characteristics of holograms that 3D images are reproduced well even some parts are damaged [31-33]. In fact the diffraction optical element (DOE), one of the techniques for generating the free form sources, is based on holography [34, 35].

Table 3. Measurements of sources and aerial images.

Measurement Mask; Resist model		EPE (nm)	NILS (AU)	DoPE_{J} (%)	DoPE_{I} (%)
Poly line array	Sigmoid	1.42	3.81	8.86	0.39
	Linear	1.52	3.73		
15-bit SRAM	Sigmoid	2.15	2.83	2.88	0.29
	Linear	1.48	2.49		

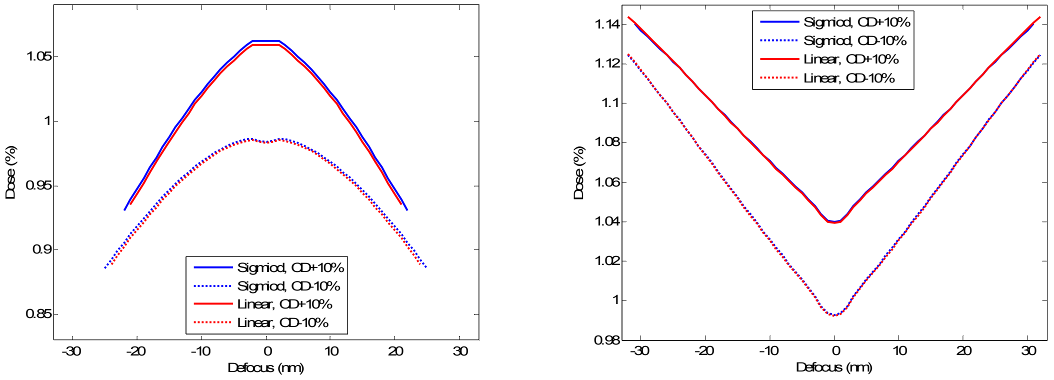
After verifying sources and images obtained from two resist models are the same, the next step is to check their impact on the speed of source optimization. Table 4 lists the relative analyses and measurements.

As a result, our linear resist model with CG takes much less iteration than the sigmoid model with SD. The complexity ratio is evaluated by $N^2/(S^2\times\pi/4)\times K_{\text{SD}}/K_{\text{CG}}$ as mentioned in the previous section and $N^2/(S^2\times\pi/4)$ is 19.75 in our settings. The complexity ratios are up to hundreds. Therefore, the speeds are enhanced by two orders. Moreover, the complexity ratios and speedup ratios of both masks are in the relation of a constant κ_1 whose values are both 0.62 for regular and brick contact arrays, respectively. Such consistent value of κ_1 for both masks implies the parameters for estimating speeds from complexity formulas have been entirely taken into consideration.

Table 4. Evaluation of computational complexity and speed enhancement.

Measurement Mask; Resist model		Iteration number (K)	Complexity Ratio ($19.75\times K_{\text{Sig}}/K_{\text{Lin}}$)	Speed up Ratio $(t^{-1})_{\text{Lin}}/(t^{-1})_{\text{Sig}}$	κ_1
Poly line array	Sigmoid	404	419.95	259.34	0.62
	Linear	19			
15-bit SRAM	Sigmoid	523	295.12	182.46	0.62
	Linear	35			

Moreover, process variations of different optimal sources of each mask are also evaluated. Process variations are important to yields and expected to have close trends if similar sources are used. The exposure-defocus (E-D) process window (PW) is the main metric to characterize the process variations. Fig. 6 shows the average E-D PWs of both masks using sigmoid and our linear resist models. Due to the sub-wavelength CD and pitch, the effective E-D PWs are within *Kirchhoff* diffraction region [19] where the defocus is smaller than half wavelength. In Figs. 14(a) and (b) the average E-D curves match very well, which strongly demonstrate the effectiveness of our linear resist model to approximate the sigmoid function.



(a) (b)

Fig. 14. Average E-D PWs of (a) the poly line array and (b) the 15-bit SRAM.

In Table 5 we list the representative measurements of E-D PWs in Fig. 14. The DoF and ΔDose are measured by finding the optimal ellipse that is tangent with the curves and having the maximum area. Thus the horizontal and vertical axes are DoF and ΔDose, respectively. Such measurements show that using the linear resist model leads to highly similar elliptical E-D PWs to those of the sigmoid model. Finally, small values ($<10^{-2}$) of standard deviations of blue and red curves in Figs. 14 (a) and (b) quantitatively verify the similarity between process variation trends.

Table 5. Measurements of E-D PWs.

Measurement Mask; Resist model		E-D PW (Ellipse)		Standard deviation of E-D curves
		DoF (nm)	ΔDose (%)	
Poly line array	Sigmoid	23.46	4.47	0.001
	Linear	23.07	4.36	
15-bit SRAM	Sigmoid	15.68	2.93	0.001
	Linear	15.68	2.93	

4. CONCLUSION

We propose the quadratic cost functions derived from our linear resist model and successfully demonstrate their effectiveness using two masks, including a 1-D poly line array and a 15-bit SRAM. Our resist image cost functions are capable of achieving similar optimal sources with the sigmoid model. Furthermore, our quadratic cost functions incorporating a conjugate gradient algorithm are much faster than the sigmoid cost function with a steepest descent algorithm. Such speed enhancement is well under expectation because the conjugate gradient algorithm inherently favors quadratic functions.

However, the cost function using sigmoid model is only close to a quadratic form at drawn edges, but not in non-pattern regions. Therefore, the final optimal sources obtained from our approach slightly deviate from sigmoid results, but always within an acceptable tolerance. We also demonstrate that one cannot approximate the sigmoid model well without the non-pattern cost. Finally, the process variations characterized by E-D windows are also in similar trends for two different optimal sources. Such results strongly demonstrate that our approach is favorable to be integrated with SO for simulating resist images efficiently.

REFERENCES

[1] Y. Granik, “Source optimization for image fidelity and throughput,” JM3 **3**, 509-522 (2004).

[2] K. Tian, A. Krasnoperova, D. Melville, A. E. Rosenbluth, D. Gil, J. Tirapu-Azpiroz, K. Lai, S. Bagheri, C.-C. Chen, and B. Morgenfeld, “Benefits and trade-offs of global source optimization in optical lithography,” Proc. SPIE **7274**, 72740C-1-12 (2009).

[3] 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,”Proc. SPIE **8166**, 81662A (2011).

[4] 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,” Proc. SPIE **7640**, 76401J (2010).

[5] 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,” Proc. SPIE **7823**, 78233X (2010).

[6] M. Fakhry, Y. Granik, K. Adam, and K. Lai, “Total source mask optimization: high-capacity, resist modeling, and production-ready mask solution,” Proc. SPIE **8166**, 81663M (2011).

[7] 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. Tritchkov, Y.

Granik, M. Fakhry, K. Adam, G. Berger, M. Lam, A. Dave, N. Cobb, “Demonstrating the benefits of source-mask optimization and enabling technologies through experiment and simulations,” Proc. SPIE **7640**, 764006 (2010).

[8] K. Lai, M. Gabrani, D. Demaris, et al., “Design specific joint optimization of masks and sources on a very large scale,” Proc. SPIE **7973**, 797308 (2011).

[9] A. Poonawala and P. Milanfar, “Prewrapping techniques in imaging: applications in nanotechnology and biotechnology,” Proc. SPIE **5674**, 114-127 (2005).

[10] A. Poonawala and P. Milanfar, “OPC and PSM design using inverse lithography: A non-linear optimization approach,” Proc. SPIE **6154**, 1159-1172 (2006).

[11] A. Poonawala and P. Milanfar, “Mask design for optical microlithography – An inverse imaging problem,” IEEE Trans. Image Process. **16**, 774-788 (2007).

[12] X. Ma and G. R. Arce, “Generalized inverse lithography methods for phase-shifting mask design,” Optics Express **15**, 15066-15079 (2007).

[13] S. H. Chan, A. K. Wong, and E. Y. Lam, “Initialization for robust inverse synthesis of phase-shifting masks in optical projection lithography,” Optical Express **16**, 14746-14760 (2008).

[14] N. Jia and E. Y. Lam, “Pixelated source mask optimization for process robustness in optical lithography,” Optics Express **19**, 19384-19398 (2011).

[15] J. C. Yu and P. Yu, “Gradient-based fast source mask optimization (SMO),” Proc. SPIE **7973**, 787320 (2011).

[16] M. Born and E. Wolf, *Principles of Optics*, 7th(expanded) ed. (Cambridge University Press, 1999).

[17] A. K. Wong, Optical Imaging in Projection Microlithography (SPIE Press, 2005).

[18] J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (McGraw-Hill Science/Engineering/Math, 2005).

[19] C. Mack, Fundamental Principles of Optical Lithography: The Science of Microfabrication, (John Wiley and Sons, 2008).

[20] D. S. Abrams and L. Pang, “Fast inverse lithography technology,” Proc. SPIE **6154**, 534-542 (2006).

[21] J. C. Yu and P. Yu, “Impacts of cost functions on inverse lithography patterning,” Optics Express **18**, 23331–23342 (2010).

[22] J. C. Yu, P. Yu, and H. Y. Chao, “Wavefront-based pixel inversion algorithm for generation of subresolution assist features,” JM3. **10**, 043014-1–043014-12 (2011).

[23] S. I. Sayegh, “Image restoration and image design in non-linear optical systems,” PhD Thesis, Univ. of Wisconsin, Madison (1982).

[24] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*, (SIAM, 1997).

[25] E. K. P. Chong and S. H. Żak, *An Introduction to Optimization*. 3rd ed. (John Wiley and Sons, 2008).

[26] X. Ma and G. R. Arce, “Pixel-based OPC optimization based on conjugate gradients,” Optics Express **19**, 2165-2180 (2011).

[27] X. Ma and G. R. Arce, “Generalized inverse lithography methods for phase-shifting mask design,” Optics Express **15**, 15066–15079 (2007).

[28] Y. Shen, N. Jia, N. Wong, and E. Y. Lam, “Robust level-set-based inverse lithography,” Optics Express **19**, 5511–5521 (2011).

[29] C. Alleaume, E. Yesilada, V. Farys, L. Depre, V. Arnoux, Zhipan Li, Y. Trouiller and A. Serebriakov, "A systematic study of source error in source mask optimization", Proc. SPIE **7823**, 782312 (2010).

[30] T. Matsuyama, N. Kita, T. Nakashima, O. Tanitsu, and S. Owa, "Tolerancing analysis of customized illumination for practical applications of source and mask optimization", Proc. SPIE **7640**, 764007 (2010).

[31] L. Xu, X. Peng, Z. Guo, J. Miao, and A. Asundi, "Imaging analysis of digital holography," Optics Express **13**, 2444-2452 (2005).

[32] I. Moon and B. Javidi, "Shape tolerant three-dimensional recognition of biological microorganisms using digital holography," Optics Express **13**, 9612-9622 (2005).

[33] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive Holography," Optics Express **17**, 13040-13049 (2009).

[34] S. Tamulevicius, A. Guobiene, G. Janusas, A. Palevicius, V. Ostasevicius, and M. Andrulevicius, "Optical characterization of diffractive optical elements replicated in polymers," JM3. **5**, 013004 (2006).

[35] G. D. M. Jeffries, G. Milne, Y. Zhao, C. Lopez-Mariscal, and D. T. Chiu, "Optofluidic generation of Laguerre-Gaussian beams," Optics Express **17**, 17555-17562 (2009).