OPC and PSM design using inverse lithography: A non-linear optimization approach

Amyn Poonawala[a] and Peyman Milanfar[b]

[a]Department of Computer Engineering, University of California, Santa Cruz, USA
[b]Department of Electrical Engineering, University of California, Santa Cruz, USA

ABSTRACT

We propose a novel method for the fast synthesis of low complexity model-based optical proximity correction (OPC) and phase shift masks (PSM) to improve the resolution and pattern fidelity of optical microlithography. We use the pixel-based mask representation, a continuous function formulation, and gradient based iterative optimization techniques to solve the above inverse problem. The continuous function formulation allows analytic calculation of the gradient. Pixel-based parametrization provides tremendous liberty in terms of the features possible in the synthesized masks, but also suffers the inherent disadvantage that the masks are very complex and difficult to manufacture. We therefore introduce the regularization framework; a useful tool which provides the flexibility to promote certain desirable properties in the solution. We employ the above framework to ensure that the estimated masks have only two or three (allowable) transmission values and are also comparatively simple and easy to manufacture. The results demonstrate that we are able to bring the CD on target using OPC masks. Furthermore, we were also able to boost the contrast of the aerial image using attenuated, strong, and 100% transmission phase shift masks. Our algorithm automatically (and optimally) adds assist-bars, dog-ears, serifs, anti-serifs, and other custom structures best suited for printing the desired pattern.

Keywords: inverse lithography, OPC, PSM, pixel-based approach, non-linear programming, optimization, regularization, low-complexity.

1. INTRODUCTION AND BACKGROUND

1.1. Introduction

In the recent years, the semiconductor industry has been finding it increasing difficult to keep pace with Moore's law. This in terms of lithography, demands a thirty percent reduction in the critical dimension (CD) every eighteen months as reflected in the International Road map for Semiconductors (ITRS). Circuit patterns are commonly transferred onto silicon wafer using optical microlithography, a process similar to photographic printing. A typical lithography system consists of two steps; namely the optical image formation and the resist action. The optical imaging system is bandlimited owing to the diffraction effects resulting in severe loss of the higher frequency components in the projected mask image (also known as the aerial image). Thus, the aerial image is a highly distorted (blurred) version of the fed input. The photo-resist responds to the above (projected) image and hence the resulting wafer pattern has poor fidelity due to rounding, line-end shortening, merging, etc. Furthermore, the lithography system is subject to random (uncontrollable) process variations which give rise to variations in the output patterns and the measured CDs. The above factors are the main culprits contributing towards yield loss, a key factor for manufacturing productivity. Thus, the overall requirement is to design a robust lithography system which prints wafers with high pattern fidelity. With Moore's law continuing to demand smaller CDs, this puts very stringent requirements on lithography making it the tightest bottleneck in the semiconductor industry.

A mathematical description of the given imaging system is referred to as the process (or forward) model. The aerial image calculations are based on the underlying optical system model (coherent, incoherent, or partially

[a] E-mail: amyn@ce.ucsc.edu
[b] E-mail: milanfar@ee.ucsc.edu

coherent[1,2]). The resist effects are simulated using Dills model,[3] Mack Model,[4,5] CTR (Constant Threshold Resist) model,[6] VTR (Variable Threshold Resist) model,[7] or other models.[8] We use the CTR model in our analysis.

The resolution(R) of the lithography system is dictated by Rayleigh's criterion,

$$R = \frac{k\lambda}{NA}. \tag{1}$$

The resolution can be improved by increasing the numerical aperture of the imaging system (NA) or decreasing the wavelength (λ) and the above parameters have been actively exploited by researchers in the past (current values are $\lambda = 193$nm and $NA \approx 0.9$). Pushing the wavelength below 193nm is practically deemed infeasible because of the limiting cost requirements, and the physical limit for $NA = 1$ for dry lithography. This leaves out the only option of decreasing the process constant k using resolution enhancement techniques (RETs).[9–11]

RETs are based on exploring three properties of the optical wavefront; namely, its amplitude, phase and direction; and are accordingly classified as optical and process correction, phase-shift methods, and off-axis illumination.[12] Optical and process correction (OPC) consists of carefully changing the sizes of the openings thereby controlling the amount of light let through. This corresponds to adding sub-resolution features to the mask pattern, which pre-compensate for the process losses to come, thereby leading to a general improvement in pattern fidelity. Phase-shift masks (PSM) consists of treating the mask as a three-dimensional structure and inducing phase-shift in the transmitted electric field such that it causes favorable constructive and destructive interference in the desired bright and dark areas respectively. Finally, off-axis illumination consists of modifying the illuminator (source) size and shape (e.g. quasar, annular, quadrapole, dipole, etc), which affects the direction of the incident light and ultimately the diffraction orders captured by the lenses.

1.2. Past Work and Inverse Lithography Techniques

The widely used approach for OPC mask design proposed by Cobb and Zakhor[2] consists of parameterizing the mask using polygons and fragmenting the mask pattern into edges and corners. These geometric elements are then nudged and moved around while simulating the output at specified control sites (using the forward model) until certain criteria are satisfied. The above technique is local in the sense that the changes are made only locally to the edges or corners of the mask to correct the corresponding edge locations at the output. A direct consequence of the above is that assist bars cannot be automatically generated. There is also a danger of printing side-lobes and hence an extra verification step is required.[13] Furthermore, phase assignments cannot be optimally carried out forming another drawback for 65nm and smaller nodes.

Hence, there has been revival of interest in "inverse lithography" or "layout inversion" techniques in recent times which form the thrust of this paper. Inverse lithography is an *image synthesis* or *image design*[14,15] problem, which consists of finding an image that when used as the input to a given imaging system results in the desired output image (to within some prescribed tolerance). The image formation process can be mathematically expressed as,

$$z(x, y) = T\{i(x, y)\}, \tag{2}$$

where $T\{.\}$ is the forward model which maps the input intensity function $i(x, y)$ to the output intensity function $z(x, y)$. Let $z^*(x, y)$ be the desired output intensity function. Therefore, the goal of inverse lithography technique (ILT) is to estimate the input intensity function which will give us an approximation to the desired output $z^*(x, y)$. This is achieved by searching the space of all inputs and choosing $\hat{i}(x, y)$ which minimizes a distance $d(z(x, y), z^*(x, y))$, where $d(\cdot, \cdot)$ is some appropriate distance metric to be defined later.

$$\hat{i}(x, y) = \arg \min_{i(x,y)} d(z^*(x, y), T\{i(x, y)\}). \tag{3}$$

In our case, $T\{.\}$ is the lithography forward (process) model, $z^*(x, y)$ is the desired output wafer pattern, and $\hat{i}(x, y)$ is the estimated optical proximity correction or phase shift mask pattern.

The pioneering work in ILT was by Sherif. et al[16] who employed iterative alternating projections approach to synthesize binary masks. Liu and Zakhor[17] used simulated annealing and Pati-Kailath[18] used POCS (Projection

on Convex Sets) respectively to synthesize phase shift masks. Oh, et al.[19] used random-pixel flipping and Erdmann, et al.[20] proposed genetic algorithms to solve the above problem. More recently Granik[21] used non-linear programming to solve the above problem and Liu, et al[13] demonstrated commercial viability of inverse lithography techniques.

In this article, we discuss the extension of our earlier proposed framework (for image prewarping[22]) to incoherent, partially coherent, and coherent imaging systems, and present extensive "layout inversion" results for the latter case. The goal is to demonstrate automatic generation and optimal placement of assist bars using ILT, and to obtain high contrast aerial images with good pattern fidelity by synthesizing attenuated, strong, and 100% transmission PSM. The key difference in our approach is that we formulate the mask synthesis problem using continuous function optimization and use the *gradient* information to systematically exploit the solution space. The gradient is analytically calculated in $O(MNlog(MN))$ operations for an $M \times N$ discrete pattern. We use the pixel-based parametrization which results in more flexibility in mask representation. However, it also suffers an inherent drawback that the synthesized masks are complex and hence very difficult to manufacture and inspect. Liu and Zakhor addressed this issue in the past using a cell-based approach.[23] The cells are selected and moved around either randomly or using the knowledge from previous moves. We instead, employ the regularization framework and use an L_1 norm based penalty function to curb the mask complexity. The above framework enables us to automatically arrive at low-complexity, high fidelity, and high contrast OPC masks or PSM while searching the solution space.

The process model and the optimization problem are formulated in Section 2. The regularization framework and OPC mask design is discussed in Section 3. The proposed framework is extended to the three cases of phase shift masks, namely, attenuated, strong, and 100% transmission in Section 4. Finally, we provide conclusive remarks in Section 5.

2. PROCESS MODEL AND PROBLEM FORMULATION

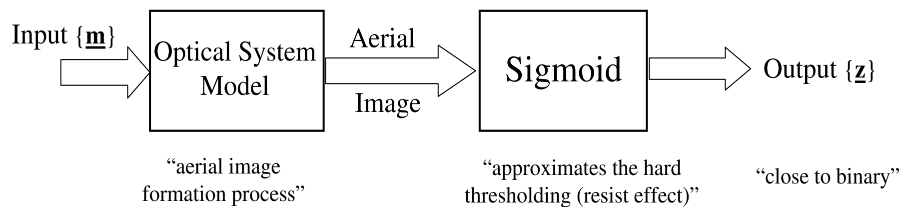

Figure 1. Approximated forward process model

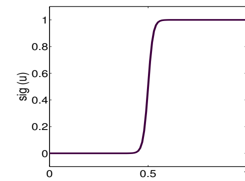

Figure 2. The sigmoid function $\text{sig}(u) = 1/(1 + e^{-80(u-0.5)})$

We employ the approximated forward process model illustrated in Fig. 1 in our analysis. Note that $\underline{\mathbf{m}}$ in Fig. 1 denotes a vector representing the mask which plays the part of the input intensity function $i(x, y)$ used in (2) and (3). The aerial image calculations are performed on the basis of coherence, incoherence, or partial coherence of the underlying imaging system. The resist effect is simulated using the constant threshold resist (CTR) model to bring the contours in the output pattern on target. The resist operation can be described using a Heaviside operator (hard threshold) defined as

$$\Gamma(u) = \begin{cases} 0, & u \leq t_r \\ 1, & u > t_r \end{cases} \tag{4}$$

Employing the Heaviside operator results in a discrete combinatorial optimization problem[14, 17] . Therefore, we approximate the above operator using sigmoid; a smooth, continuous function.[24] We employ the logarithmic sigmoid function,

$$\text{sig}(u) = \frac{1}{1 + e^{-a(u-t_r)}}, \tag{5}$$

where the parameter a dictates the steepness of the sigmoid. A large value of a leads to a very steep sigmoid which closely resembles the hard thresholding operation. The parameter t_r is the threshold parameter of the sigmoid and is set equal to the threshold level of the resist in accordance with the constant threshold resist model. As an example, Fig. 2 illustrates the behavior of a sigmoid with $a = 80$ and $t_r = 0.5$. The above approximation enables us to use gradient-based continuous function optimization techniques like steepest-descent to solve the mask design problem.

Since we employ the pixel based approach, the first step is to represent the input, output, and the desired patterns using 2-D discrete images. We define vectors $\underline{\mathbf{z}}^*,\ \underline{\mathbf{z}},\ \underline{\mathbf{m}} \in \Re^{MN\times 1}$ which are obtained by sampling and lexicographic ordering of $z^*(x, y)$, $z(x, y)$, and $m(x, y)$ respectively. Throughout our discussion, $\underline{\mathbf{z}}^*$ represents the prescribed *binary* pattern, $\underline{\mathbf{z}}$ represents the *gray-level* output pattern, and $\underline{\mathbf{m}}$ represents the input pattern fed to the imaging system (can be *binary* or *gray-level*). The number of samples along the horizontal and vertical directions are given by M and N respectively.

2.1. Modelling the Imaging System

We now discuss the aerial image calculations and process models for the three imaging systems of interest.

2.1.1. Coherent Imaging System

In the case of coherent imaging system, the spatial distribution of the output electric field amplitude $z(x, y)$ is linearly related to the input electric field amplitude generated by the mask $m(x, y)$. This can be mathematically described as,

$$z(x, y) = m(x, y) * h(x, y) \tag{6}$$

where $h(x, y)$ is referred to as the amplitude spread function (ASF) of the given imaging system.[25]

Typical lithography systems employ a circular lens aperture, where the coherent imaging system now acts as an ideal low pass filter with cutoff frequency NA/λ. The higher frequency components of the diffracted mask image are lost by the finite lens aperture stop thereby causing a blurry version of the mask image at the imaging (wafer) plane. The convolution kernel $h(x, y)$ is defined as the Fourier transform of the circular lens aperture with cutoff frequency NA/λ.[1,25] Therefore,

$$h(x, y) = jinc(r) = \frac{J_1(2\pi r\ NA/\lambda)}{2\pi r\ NA/\lambda} \tag{7}$$

where $r = \sqrt{x^2 + y^2}$ and $J_1(\cdot)$ is the Bessel function of first kind.

The photo-resist responds to the intensity of the electric field, where intensity is defined as the square of the complex amplitude. Therefore, the forward model is defined as,

$$\underline{\mathbf{z}} = \mathrm{sig}(|\mathbf{H}\underline{\mathbf{m}}|^2) \tag{8}$$

The sigmoid function simulates the resist behavior and acts on the aerial image $|\mathbf{H}\underline{\mathbf{m}}|^2$ (square of the amplitude), giving the output pattern $\underline{\mathbf{z}}$. Two important points to note; the kernel $\mathbf{H} \in \Re^{MN\times MN}$ is the jinc function $h(x, y)$ sampled using the same interval as $\underline{\mathbf{z}}$, and $|\cdot|^2$ operator here implies element-by-element absolute square of the individual vector entries. Finally, for partially coherent imaging systems, the optical kernel $h(x, y)$ can be instead substituted by the optimal coherent approximation proposed by Pati-Kailath.[18]

2.1.2. Incoherent Imaging System

An incoherent imaging system is linear in intensity (or irradiance) and can be mathematically described as,[1]

$$|z(x, y)|^2 = |m(x, y)|^2 * |h(x, y)|^2. \tag{9}$$

The kernel $h(x, y)$ is defined in (7), and $|\cdot|^2$ once again implies absolute square of the individual elements. Note that the phase of the input electric field does not contribute towards the output. Hence, for incoherent imaging systems we restrict ourselves to only binary masks. The photo-resist directly responds to the above electric field intensity and the forward model is defined as,

$$\underline{\mathbf{z}} = \mathrm{sig}(\tilde{\mathbf{H}}|\underline{\mathbf{m}}|^2) = \mathrm{sig}(\tilde{\mathbf{H}}\underline{\mathbf{m}}) \tag{10}$$

The filter $\tilde{\mathbf{H}}$ in (10) is known as the point spread function of the imaging system and is a jinc-squared function, whereas $|\underline{\mathbf{m}}|^2 = \underline{\mathbf{m}}$ for binary masks.

2.1.3. Partially Coherent Imaging System

Real-world lithography system are partially coherent and can be modelled using Hopkins diffraction model.[1] Pati and Kailath[18] proposed an approximation to the above model called the sum-of-coherent-system (SOCS) by using the singular value decomposition of the transmission cross coefficient matrix. In their approach, the Pth order approximation to the aerial image formulation can be calculated using the weighted sum of P coherent systems. The forward model now becomes,

$$\underline{\mathbf{z}} = \mathrm{sig}\left(\sum_{j=1}^{P} \sigma_j |\mathbf{H}_j \underline{\mathbf{m}}|^2\right) \tag{11}$$

where $\mathbf{H}_j$ for $j = 1, \ldots, P$ are the amplitude spread functions (also referred to as optical system kernels) of the coherent systems, and $\sigma_1, \ldots, \sigma_P$ are the singular values. The singular values quickly decay to zero thereby facilitating an accurate reduced order approximation.

2.2. Optimization Problem Formulation

In this article, we focus on the mask synthesis problem for the case of coherent imaging systems. Every pixel m_j of the mask can be represented as a complex term $m_j = p_j + iq_j$ where $i = \sqrt{-1}$. In our analysis we restrict ourselves only to strong (180 degree) phase shift. Therefore $q_j = 0$ for $j = 1, \ldots, MN$ thereby requiring us to estimate only the real part (p_j) of the mask (which we henceforth refer to as m_j for notational convenience). We formulate the mask design problem as finding the optimized mask layout $\hat{\underline{\mathbf{m}}}$ that minimizes the cost function $F(\underline{\mathbf{m}})$, defined as the L_2 norm of the difference between the desired pattern $\underline{\mathbf{z}}^*$ and the output pattern $\underline{\mathbf{z}}$. That is,

$$\hat{\underline{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}}\{F(\underline{\mathbf{m}})\} = \arg\min_{\underline{\mathbf{m}}} ||\underline{\mathbf{z}}^* - \underline{\mathbf{z}}||_2^2 = \arg\min_{\underline{\mathbf{m}}} \sum_{k=1}^{MN} (z_k^* - z_k)^2 \tag{12}$$

Later in Section 3, we refine this approach by introducing the regularization terms and augmenting the cost function. From (8) we observe that every pixel in a coherent imaging system undergoes a cascade of convolution, squaring, and sigmoidal operation. Therefore, the output pixel z_k in (12) can be represented as,

$$z_k = \frac{1}{1 + \exp\left[-a\left(\sum_{j=1}^{MN} h_{kj} m_j\right)^2 + a t_r\right]} \, , \tag{13}$$

for $k = 1, \ldots, MN$ where $\mathbf{H}$ is the jinc function defined in (7).

The estimated mask is either two or three tone depending on the employed RET. Therefore, the transmission values m_j for $j = 1, \ldots, MN$ should be allowed to take only specific values as summarized in the table below.

RET	Allowable transmission values
OPC	0 or +1
6% Attenuated PSM	-0.2449 or +1
18% Attenuated PSM	-0.4243 or +1
Strong PSM (100% transmission)	-1 or +1
Strong PSM (With chrome)	-1 or 0 or +1

The optimization problem (12) is therefore subject to the constraints given by the allowable transmission values of m_j. This would unfortunately result in a combinatorial optimization problem, thereby preventing the usage of gradient information to explore the search space. To overcome this problem, we relax the parameter values to lie within a *range* $[\underline{m}, \overline{m}]$. Hence, the optimization problem in (12) is subject to the following inequality (bound) constraints

$$\underline{m} \leq m_j \leq \overline{m} \quad \text{for } j = 1, \ldots, MN \; . \tag{14}$$

3. OPC MASK DESIGN

We now discuss the proposed optimization algorithm for synthesizing OPC masks for coherent imaging systems. We also introduce the regularization framework to guarantee near binary results, ease of manufacturing, and good quality aerial image.

3.1. Mask Optimization Algorithm

In the case of OPC masks, the transmission values are restricted to either 0 or 1. For coherent imaging systems, the optimization problem is defined using (12) and (13) subject to the constraints

$$0 \leq m_j \leq 1 \quad \text{for } j = 1, \ldots, MN.$$

The above problem can be solved using constrained optimization algorithms like BFGS (Broyden, Fletcher, Goldfarb, Shanno) or gradient-projection.[26,27] The bound-constrained optimization problem can be further reduced to an unconstrained optimization problem using the following parametric transformation,

$$m_j = \frac{1 + \cos(\theta_j)}{2} \quad \text{for } j = 1, \ldots, MN \tag{15}$$

where $\underline{\boldsymbol{\theta}} = [\theta_1, \ldots, \theta_{MN}]^T$ is the unconstrained parameter vector. The re-parameterized cost function can be formulated in terms of the parameter vector $\underline{\boldsymbol{\theta}}$ as follows,

$$F_1(\underline{\boldsymbol{\theta}}) = \sum_{k=1}^{MN} \left(z_k^* - \frac{1}{1 + \exp\left[-a \left(\sum_{j=1}^{MN} h_{kj} \frac{1 + \cos(\theta_j)}{2} \right)^2 + a t_r \right]} \right)^2 . \tag{16}$$

We can now employ the simple steepest-descent search method to minimize the above cost function. This requires the first order derivatives of (16), and the gradient vector $\nabla F_1(\underline{\boldsymbol{\theta}}) \in \Re^{MN \times 1}$ can be analytically calculated using the following expression,

$$\nabla F_1(\underline{\boldsymbol{\theta}}) = \underline{\mathbf{d}} = a(\mathbf{H^T} \ [(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}) \odot \underline{\mathbf{z}} \odot (\mathbf{1} - \underline{\mathbf{z}}) \odot (\mathbf{Hm})] \ ) \odot \ \sin(\underline{\boldsymbol{\theta}}), \tag{17}$$

where $\odot$ is the element-by-element multiplication operator. The nth iteration of steepest descent is given as,

$$\underline{\boldsymbol{\theta}}^{n+1} = \underline{\boldsymbol{\theta}}^n - s\underline{\mathbf{d}}^n \tag{18}$$

where s is the step-size. Here, we would like to highlight the useful fact that due to the structure of (17), the steepest descent iterations can be quickly and directly carried out on the 2-D image array (matrices) with no need for the raster scanning operation.[27] This saves valuable time and considerably eases the implementation. Finally, we note that the estimated mask $\widehat{\underline{\mathbf{m}}}$ using the above framework has continuous transmission values between 0 and 1. Therefore, we need a post-processing operation to find the optimal threshold t_m to convert $\widehat{\underline{\mathbf{m}}}$ into a physically realizable mask $\widehat{\underline{\mathbf{m}}}_b$.[22]

3.2. Regularization Framework

Inverse lithography problem is an ill-posed problem.[28] The continuous function formulation implies that there can be infinite number of input (gray level) patterns all giving rise to the same binary pattern at the output. However, we are interested in solutions comprising of only two or three tones to achieve physically realizable masks. Moreover, we want our synthesized mask patterns to have low-complexity in order to control the mask manufacturing costs. In general, we may want to inculcate (or promote) certain desirable properties in the

solution. These can be incorporated as prior information about the solution and by employing the regularization framework.[28] In general, the problem formulation can be described as follows,

$$\widehat{\underline{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}} \left[\gamma_{fid} F(\underline{\mathbf{m}}) + \gamma_{reg}\ R(\underline{\mathbf{m}})\right] \tag{19}$$

where $F(\underline{\mathbf{m}})$ is the data-fidelity term, and $R(\underline{\mathbf{m}})$ is the regularization function used to direct the unknown parameter $\underline{\mathbf{m}}$ towards the desired solution space. γ_{fid} and γ_{reg} are user-defined scalars for adequately weighing the first (data fidelity) term against the second (regularization) term. We employ the above framework with the aim of achieving three goals which are introduced in the following sections.

3.2.1. Discretization Penalty Term

The bound constraints in (14) imply that $\widehat{\underline{\mathbf{m}}}$ can have pixels with continuous gray values. Therefore, a post-processing (binarization) step is required to obtain the binary mask $\widehat{\underline{\mathbf{m}}}_b$ from $\widehat{\underline{\mathbf{m}}}$.[22] However, this forms an extra overhead operation, and is sub-optimal because the pattern-error for $\widehat{\underline{\mathbf{m}}}_b$ cannot be guaranteed to be under control. Therefore, the first regularization term is employed to ensure that the estimated mask is near-binary. Every pixel m_j now has an associated penalty given by the quadratic function (see Fig. 3),

$$r(m_j) = 1 - (2m_j - 1)^2.$$

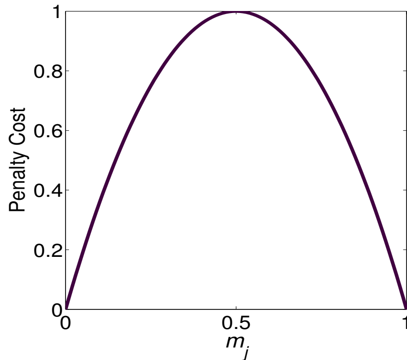

Figure 3. Discretization penalty term for binary masks.

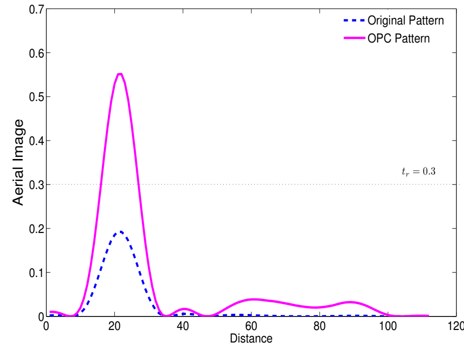

Figure 4. Horizontal slice at row number 97 for the aerial image obtained using the synthesized mask in Fig. 5. Note that the contrast has improved and the side-lobes will not print.

As described earlier, the mask transmission values will be constrained to lie in $[0, 1]$ and hence we are only interested in the behavior of the cost function in that range. The penalty incurred is zero for transmission values 0 or 1 and increases as we move away from binary region in either direction (maximum at $m_j = 0.5$). Thus we favor the estimated pixels to have values closer to 0 and 1 while exploiting the search space. The regularization term is defined as the sum of penalty of all pixels as follows,

$$R_{dis}(\underline{\mathbf{m}}) = \sum_{j=1}^{MN} r(m_j) = \sum_{j=1}^{MN} \left[1 - (2m_j - 1)^2\right] = 4\underline{\mathbf{m}}^T(\underline{\mathbf{1}} - \underline{\mathbf{m}}), \tag{20}$$

where $\underline{\mathbf{1}} = [1, \ldots, 1]^T \in \Re^{MN\times 1}$. The gradient $\nabla R_{dis}(\underline{\mathbf{m}}) \in \Re^{MN\times 1}$ of the discretization penalty term is given by,

$$\nabla R_{dis}(\underline{\mathbf{m}}) = (-8\underline{\mathbf{m}} + 4), \tag{21}$$

which can be used in conjunction with (17) and (18) while carrying out the steepest-descent iterations as before.

3.2.2. Complexity Penalty Term

The pixel-based approach allow tremendous flexibility in representing the mask patterns but also suffers the inherent disadvantage that the masks are rather complex and hence difficult to manufacture and inspect. Researchers in the past have reverted to post-processing operations to simplify the output,[19] but this approach is again sub-optimal. We follow the regularization framework and employ a penalty function to direct our algorithm towards generating low-complexity masks. Isolated perturbations, protrusions, etc are not preferred because they increase the storage and manufacturing cost. Hence we seek a penalty term which suppresses these effects. To achieve this, let us first define an auxiliary variable called *the activation pattern* $\underline{\mathbf{f}}$ where,

$$f_j = |m_j - z_j^*| \quad \text{for} \quad j = 1, \ldots, MN.$$

The *on* pixels in $\underline{\mathbf{f}}$ indicate the positions where pre-warping occurred; so the pre-warped pattern can be obtained by simply flipping the corresponding pixels in $\underline{\mathbf{z}}^*$ from 1 to 0 or 0 to 1.

There are a variety of penalty terms that one can employ depending upon how one defines *mask complexity*. Akin to the idea of Total Variation (TV)[28] penalty, we choose to penalize the mask complexity using the local variation of the activation pattern as follows[29]:

$$||\nabla\underline{\mathbf{f}}||_1 = ||\mathbf{Q_x}\underline{\mathbf{f}}||_1 + ||\mathbf{Q_y}\underline{\mathbf{f}}||_1, \tag{22}$$

where $\mathbf{Q_x}, \mathbf{Q_y} \in \Re^{MN}$ represent the first (directional) derivatives and are defined as $\mathbf{Q_x} = \mathbf{I} - \mathbf{S_x}$ and $\mathbf{Q_y} = \mathbf{I} - \mathbf{S_y}$ where $\mathbf{S_x}$ and $\mathbf{S_y}$ shift $\underline{\mathbf{m}}$ along horizontal (right) and vertical (up) direction by one pixel respectively.

This approach, while relatively simple, enables us to decouple the features of the underlying prescribed pattern $\underline{\mathbf{z}}^*$ from $\underline{\mathbf{m}}$ thereby capturing only the *changes* occurring due to prewarping. Isolated holes, protrusions, and jagged edges have higher associated penalty. The regularization term in (22) suppresses these effects and forces the *changes* to be spatially smoother and less abrupt.

The gradient $\nabla R_{TV}(\underline{\mathbf{m}}) \in \Re^{MN\times 1}$ of the complexity term is given as:

$$\nabla R_{TV}(\underline{\mathbf{m}}) = [\mathbf{Q_x}^T \text{sign}(\mathbf{Q_x}\underline{\mathbf{f}}) + \mathbf{Q_y}^T \text{sign}(\mathbf{Q_y}\underline{\mathbf{f}})] \ \odot \ \text{sign}(\underline{\mathbf{m}} - \underline{\mathbf{z}}^*), \tag{23}$$

which can be used in conjunction with (17) and (18) while carrying out the steepest-descent iterations as before.

3.2.3. Aerial Image Penalty Term

The lithography process needs to be robust to the process errors introduced by undesirable focus and exposure variations. This can be achieved by availing a good quality aerial image; one with sharp contrast or steep transitions along desired edge locations. To achieve the above goal, we employ a penalty term defined as the L_2 norm of the difference between the desired pattern and the aerial image obtained using the input mask. This penalty term will be particularly useful for phase shift masks, where it forces constructive or destructive interference in desirable regions thereby boosting the contrast. For a coherent imaging system, the penalty term is given as,

$$R_{aerial}(\underline{\mathbf{m}}) = ||\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{m}}|^2||_2^2 \tag{24}$$

and the gradient $\nabla R_{aerial}(\underline{\mathbf{m}})$ can be calculated as,

$$\nabla R_{aerial}(\underline{\mathbf{m}}) = -\mathbf{H}\left[(\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{m}}|^{\mathbf{2}}) \odot (\mathbf{H}\underline{\mathbf{m}})\right] \tag{25}$$

It is interesting to note that the optimization problem can also be formulated using (24) as the fidelity term instead of (12). This will guarantee a high quality aerial image, however the contour fidelity will be very poor. The above approach is useful in case of *isolated* contacts where the dose can be varied to bring the CD on target and hence contour fidelity is not an issue of concern.

3.3. Results

The augmented cost function to be minimized is defined as the sum of the fidelity term and the regularization terms:

$$J(\underline{\mathbf{m}}) = \gamma_{fid}F(\underline{\mathbf{m}}) + \gamma_{aerial}R_{aerial}(\underline{\mathbf{m}}) + \gamma_{dis}R_{dis}(\underline{\mathbf{m}}) + \gamma_{TV}R_{TV}(\underline{\mathbf{m}}) \tag{26}$$

Fig. 5 illustrates the estimated masks and the binary output patterns obtained using a coherent imaging system with $\lambda = 193$nm, $NA = 0.85$, and the resist threshold $t_r = 0.3$. The desired pattern is a 100nm random logic contact holes pattern sampled at 10nm. We observe that our optimization algorithm automatically adds assist bars and adjusts the sizes of the centers to bring the output contours on target. The assist features are also shared between adjacent contacts. Fig. 4 compares the aerial image slices and indicates a tremendous improvement in the peak intensity. We also observe that the side-lobes are under control and the assist bars do not print. However, the assist features are broken, irregular, curvy, and complex, making the mask very difficult to manufacture; hence, we repeat the experiment by employing the complexity penalty term ($\lambda_{TV} = 0.001$). This leads to a comparatively simple pattern (see $\widehat{\underline{\mathbf{m}}}_b$ on the right in Fig. 5) and the assist bars tend to become square or rectangular in shape which are more preferable.

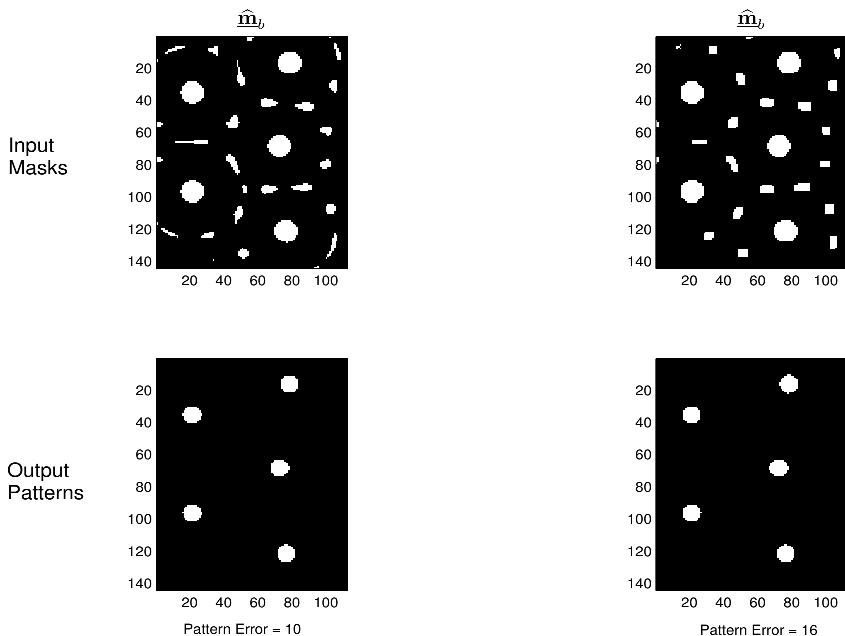

Figure 5. The synthesized masks (top row) and the output binary patterns (bottom row) for 100nm random logic contact patterns. $\widehat{\underline{\mathbf{m}}}_b$ (left) is estimated using $\lambda_{fid} = 0.75$, $\lambda_{aerial} = 0.25$, $\lambda_{dis} = 0.002$, and $\lambda_{TV} = 0$ and $\widehat{\underline{\mathbf{m}}}_b$ (right) is estimated using same parameters except $\lambda_{TV} = 0.001$

4. PSM MASK OPTIMIZATION

We now move our discussion to phase shift mask design and demonstrate the extension of our framework to the cases of attenuated phase shift mask, strong phase shift mask, and 100% transmission phase shift mask.

4.1. Attenuated PSM

Attenuated phase shift masks consist of quartz and molybdenum silicide (MoSi) instead of chrome. MoSi (unlike chrome) allows a small percentage of light intensity (typically 6% or 18%) to pass through it. The thickness of

MoSi is chosen such that light which passes through is 180° out of phase compared to the transmitting quartz regions. In our discussion, we focus on the 6% intensity transmission AttPSM masks. Every pixel m_j can now have only two amplitude transmission values equal to $-\sqrt{0.06} = -0.245$ (the 180 degree phase shift with weak transmission) or 1 (100% transmission with no phase shift). The optimization problem is formulated similar to (12) subject to the constraint that $m_j = -0.245$ or 1. We reduce it to a non-linear programming problem using the bound constraints $-0.245 \leq m_j \leq 1$. The parametric transformation should now map the unconstrained variable θ_j to the above range and is given as,

$$m_j = 0.6225(1 + \cos(\theta_j)) - 0.245 \tag{27}$$

The quadratic penalty term is employed where pixels having values -0.245 and 1 have zero penalty and the cost increases as we move towards the center of the range (also see Fig. 6),

$$r(m_j) = -m_j^2 + 0.755m_j + 0.245 \tag{28}$$

The two-tone AttPSM $\widehat{\underline{\mathbf{m}}}_b$ is simply obtained by thresholding the estimated mask $\widehat{\underline{\mathbf{m}}}$ with $t_m = 0.3775$.

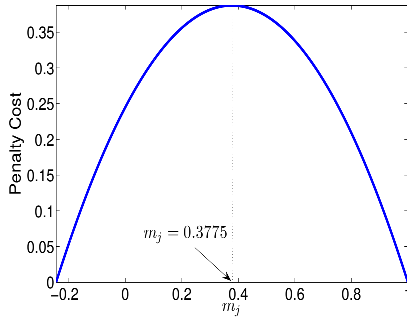

Figure 6. Discretization penalty term for AttPSM (maximum penalty is at $m_j = 0.3775$)

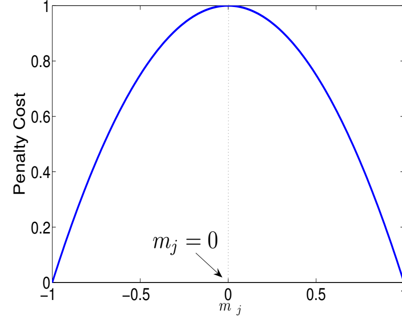

Figure 7. Discretization term for 100% transmissive PSM (maximum penalty is at $m_j = 0$)

4.2. 100% transmission PSM

As the name suggests, 100% transmission PSM is an extreme case of AttPSM and does not use the opaque chrome features at all. It is an all transmissive mask consisting of only zero and 180 degree phase shift features. Thus the synthesized mask can only have values -1 or 1. We formulate the optimization problem similar to (12), subject to the constraint that $m_j = -1$ or 1. Once again we relax the condition and impose the bound constraints $-1 \leq m_j \leq 1$ and perform the parametric transformation $m_j = \cos(\theta_j)$ to reduce the problem to an unconstrained optimization problem. The quadratic penalty term now has zero penalty at -1 and 1 and maximum penalty for $m_j = 0$ (see Fig. 7),

$$r(m_j) = -m_j^2 + 1. \tag{29}$$

Hence, everything is pushed towards -1 and 1 thereby easing the discretization step.

4.3. Strong PSM

The final discussion is for strong PSM where the mask features can have values 0 (chrome) or 1 (quartz with no phase shift) or -1 (quartz etched to provide 180 degree phase shift). The bound constraints and parametric transformation are similar to those employed in 100% transmission PSM. The only difference occurs in the discretization regularization term since we now want a three-tone mask. We employ a fourth order quartic penalty term where each pixel has an associated cost,

$$r(m_j) = -4.5m_j^4 + m_j^2 + 3.5 \tag{30}$$

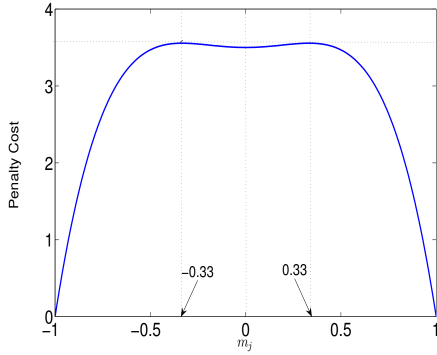

Figure 8. Discretization penalty cost for strong PSM (with chrome). The maxima's are at $m_j = -0.33, +0.33$.

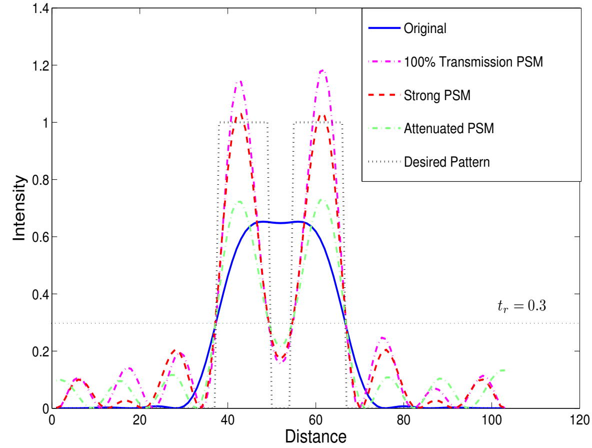

Figure 9. Horizontal slices along the center of the aerial images obtained using the desired pattern $\underline{\mathbf{z}}^*$ and the various synthesized PSM's as input.

Fig. 8 is the plot of the above function. Note that the entire range from $[-1, 1]$ is divided into three equal intervals. The penalty is lower for pixels having values -1, 0, or 1, whereas it is maximum for -1/3 and 1/3. Finally, $\widehat{m}_j \in [-1, -0.33)$, $\widehat{m}_j \in [-0.33, +0.33)$, and $\widehat{m}_j \in [+0.33, 1]$ are quantized to -1, 0, and 1 respectively to obtain the three-tone mask $\underline{\widehat{\mathbf{m}}}_b$.

4.4. Results for PSM

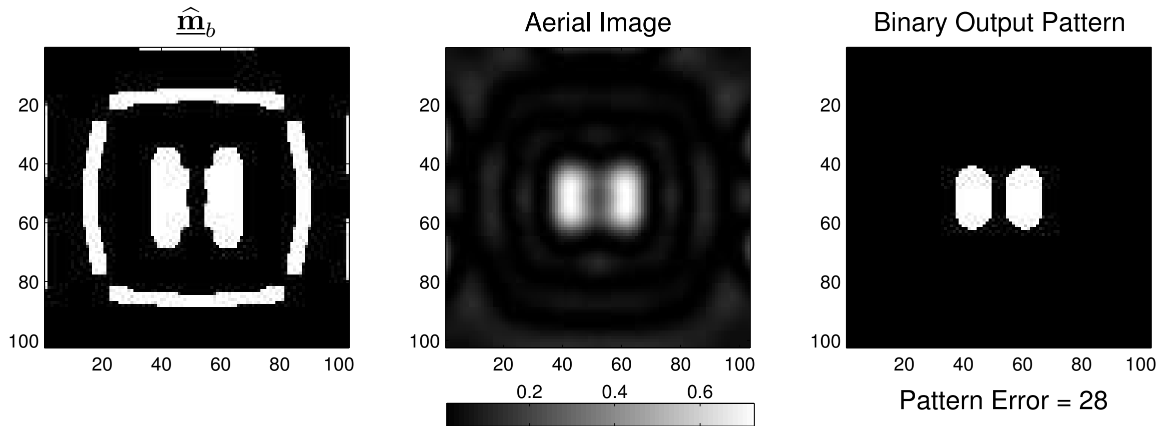

Figure 10. The estimated 6% AttPSM mask (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with λ = 193nm and NA = 0.85. The black and white regions in $\underline{\widehat{\mathbf{m}}}_b$ correspond to -0.245 and 1 respectively.

We now demonstrate results for the above discussed PSM. During our experiments, we observed that a new two step optimization procedure is more efficient than the direct one step optimization technique employed earlier. The new strategy is described below.

Step 1 - To start with, we minimize the cost function defined as,

$$J(\underline{\mathbf{m}}) = R_{aerial}(\underline{\mathbf{m}})$$

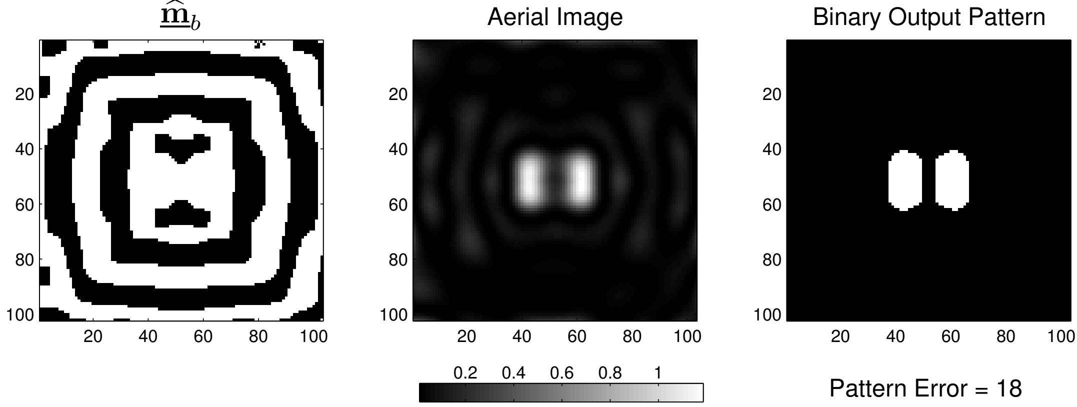

Figure 11. The estimated 100% transmission PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black and white regions in $\widehat{\underline{\mathbf{m}}}_b$ correspond to -1 and 1 respectively.

Since the only term employed is the aerial image penalty, our goal is to estimate a mask which improves the overall contrast of the aerial image. The steepest descent iterations are initialized as $\underline{\mathbf{m}}^0 = \underline{\mathbf{z}}^*$. Note that contour fidelity is not accounted for and hence the resulting contours will not be on target.

Step 2 - We now minimize the augmented cost function defined in (26) which includes the contour fidelity, complexity, and discretization terms. The steepest descent iterations are initialized using the estimated mask pattern obtained from Step 1. Phase shift masks can result in very high contrast aerial images by causing constructive or destructive interference wherever desirable. An advantage of the two step strategy is that the phase assignments are primarily made during the first step, whereas the contour fidelity, tone, and simplicity criteria are addressed during the second round of optimization. Furthermore, the initial guess obtained from Step 1 is often quite far from the initial pattern, and hence the two step approach helps exploit the search space better.

Fig. 10 illustrates the synthesized 6% attenuated PSM for a coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The desired pattern consists of two 120nm thick bars separated by 50nm. The experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\lambda_{fid} = 1$, $\lambda_{aerial} = 0.25$, $\lambda_{TV} = 0.01$, $\lambda_{dis} = 0.01$, and $s = 5$. We observe that the optimization algorithm automatically adds assist bars in all four directions. The width of the assist bars and their placements from the center feature are calculated as part of the optimization procedure. Fig. 9 illustrates the central horizontal slice of the aerial image. If the desired binary pattern is itself fed as the input to the imaging system, the aerial image barely has any modulation and the two bars are not distinguishable. However, the synthesized AttPSM mask causes good modulation and the assist features do not print.

Fig. 11 illustrates the synthesized 100% transmission PSM for the same imaging system. The experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\lambda_{fid} = 0.75$, $\lambda_{aerial} = 0.25$, $\lambda_{TV} = 0.01$, $\lambda_{dis} = 0.01$, and $s = 5$. Note that our goal is to create a horizontal separation between the two bars. However, the corresponding region in the synthesized mask is a zero phase shift fully transmissive (white) contiguous region. The destructive interference is actually created by the two vertically separated 180 degree phase shift (black) features giving a sharp contrast aerial image as observed in Fig. 9. The result also demonstrates that our algorithm can produce synthesized masks which are very different from the desired patterns.

Finally, Fig. 12 illustrates the result using a strong PSM. The experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\lambda_{fid} = 1$, $\lambda_{aerial} = 0.25$, $\lambda_{TV} = 0$, $\lambda_{dis} = 0.01$, and $s = 1.5$. Once again the synthesized mask is quite different from the desired pattern. We also see three assist features around the main pattern which improve the contrast of the aerial image. The aerial image slices in Fig. 9 demonstrate an improvement in the contrast. Furthermore the side-lobes are below the resist threshold t_r and hence they will not print.

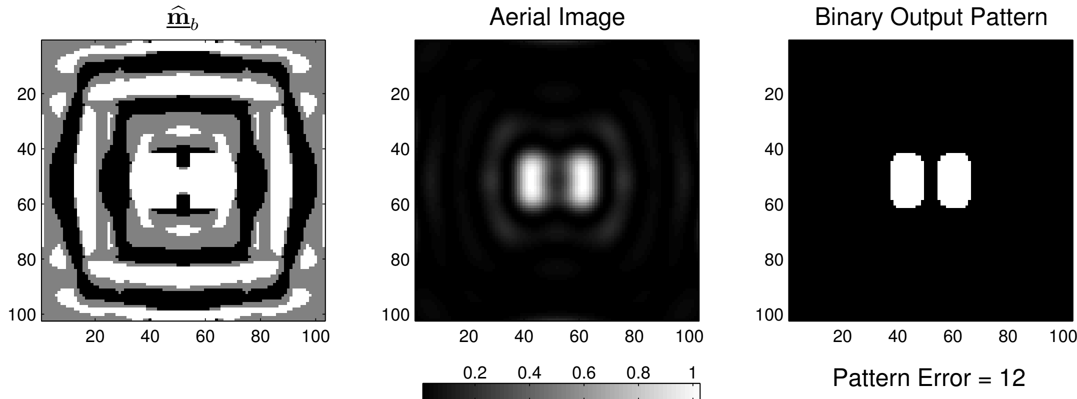

Figure 12. The estimated strong PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black, gray, and white regions in $\widehat{\underline{\mathbf{m}}}_b$ correspond to -1, 0, and 1 respectively.

5. FUTURE WORK AND CONCLUSIONS

We proposed a new framework for fast and efficient pixel-based binary and phase shift mask design. The mask design problem was reduced to an unconstrained continuous function optimization problem, which was iteratively solved using the analytically evaluated gradient information. The regularization framework was effectively used to control the tone and complexity of the resulting masks. The aerial image quality was also successfully boosted for phase shift masks. The proposed technique can be readily extended to employ RET's for optical maskless lithography, proximity correction in e-beam lithography, and DOE source design for off-axis illumination. With the semiconductor industry striving to adhere to Moore's law and the critical dimensions hitting 45nm, there has been tremendous interest in model-based inverse lithography techniques. The algorithms and framework presented in this paper are promising and they have, and will, contribute towards that direction.

ACKNOWLEDGMENTS

This work was supported by Intel Corporation.

REFERENCES

1. M. Born and E. Wolfe, *Principles of Optics*, Cambridge University Press, 1999.
2. N. Cobb and A. Zakhor, "Fast sparse aerial image calculation for OPC," in *BACUS Symposium on Photomask Technology*, *Proc. SPIE* **2621**, pp. 534–545, 1995.
3. P. Choudhury, *Handbook of Microlithography, Micromachining and Microfabrication*, SPIE Press, 1997.
4. C. Mack, "Development of positive photoresist," *Journal of Electrochemical Society* **134**, pp. 148–152, 1987.
5. S. Robertson, C. Mack, and M. Maslow, "Toward a universal resist dissolution model for lithography simulation," in *Lithography for Semiconductor Manufacturing II*, *Proc. SPIE* **4404**, pp. 111–122, 2001.
6. W. Huang, C. Lin, C. Kuo, C. Huang, J. Lin, J. Chen, R. Liu, Y. Ku, and B. Lin, "Two threshold resist models for optical proximity correction," in *Optical Microlithography*, *Proc. SPIE* **5377**, pp. 1536–1543, 2001.
7. J. Randall, K. Ronse, T. Marschner, M. Goethals, and M. Ercken, "Variable threshold resist models for lithography simulation," in *Optical Microlithography*, *Proc. SPIE* **3679**, pp. 176–182, 1999.
8. C. Ahn, H. Kim, and K. Baik, "Novel approximate model for resist process," in *Optical Microlithography*, *Proc. SPIE* **3334**, pp. 752–763, 1998.

9. F. Schellenberg, "Resolution enhancement technology: The past, the present, and extensions for the future," in *Optical Microlithography*, B. W. Smith, ed., *Proc. SPIE* **5377**, pp. 1–20, 2004.
10. F. Schellenberg, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, 2004.
11. L. Liebmann, S. Mansfield, A. Wong, M. Lavin, W. Leipold, and T.Dunham, "TCAD development for lithography resolution enhancement," *IBM Journal of Research and Development* **45**, pp. 651–665, 2001.
12. A. K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, 2001.
13. Y. Liu, D. Abrams, L. Pang, and A. Moore, "Inverse lithography technology principles in practice: Unintuitive patterns," in *BACUS Symposium on Photomask Technology*, *Proc. SPIE* **5992**, pp. 231–238, 2005.
14. S. Sayegh and B. Saleh, "Image design: Generation of a prescribed image at the output of a bandlimited system," *IEEE Transcation on Pattern Analysis and Machine Intelligence* **5**, pp. 441–445, 1983.
15. S. Sayegh, B. Saleh, and K. Nashold, "Image design: Generation of a prescribed image through a diffraction limited system with high-contrast recording," *IEEE Transaction on Acoustics, Speech and Signal Processing* **33**, pp. 460–465, 1985.
16. S. Sherif, B. Saleh, and R. Leone, "Binary image synthesis using mixed linear integer programming," *IEEE Transactions on Image Processing* **4**, pp. 1252–1257, 1995.
17. Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," *IEEE Transactions on Semiconductor Manufacturing* **5**, pp. 138–151, 1992.
18. V. Pati and T. Kailath, "Phase-shifting masks for microlithography : Automated design and mask requirements," *Journal of Optical Society of America A - Optics Image Science and Vision* **9**, pp. 2438–2452, 1994.
19. Y. Oh, J. C. Lee, and S. Lim, "Resolution enhancement through optical proximity correction and stepper parameter optimization for 0.12-μ m mask pattern," in *Optical Microlithography*, *Proc. SPIE* **3679**, pp. 607–613, 1999.
20. A. Erdmann, R. Farkas, T. Fuhner, B. Tollkuhn, and G. Kokai, "Towards automatic mask and source optimization for optical lithography," in *Optical Microlithography*, *Proc. SPIE* **5377**, pp. 646–657, 2004.
21. Y. Granik, "Solving inverse problems of optical microlithography," in *Optical Microlithography*, *Proc. SPIE* **5754**, pp. 506–526, 2005.
22. A. Poonawala and P. Milanfar, "Prewarping techniques in imaging: Applications in nanotechnology and biotechnology," in *Electronic Imaging*, *Proc. SPIE* **5674**, pp. 114–127, 2005.
23. Y. Liu and A. Zakhor, "Computer aided phase shift mask design with reduced complexity," *IEEE Transcation on Semiconductor Manufacturing* **9**, pp. 170–184, 1996.
24. W. Duch and N. Jankowski, "Survey of neural transfer functions," *Computing Surveys* **2**, pp. 163–213, 1999.
25. R. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics*, John Wiley and Sons, 1995.
26. C. Kelly, *Iterative Methods for Optimization*, SIAM, 1999.
27. A. Poonawala and P. Milanfar, "Fast and low-complexity mask design in optical microlithography - an inverse imaging problem," *submitted to IEEE Transactions on Image Processing* .
28. C. Vogel, *Computational Methods for Inverse Problems*, SIAM Press, 2002.
29. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Transcation on Image Processing* **13**, pp. 1327–1344, 2004.