

A pixel-based regularization approach to inverse lithography [☆]

Amy Poonawala ^{a,*}, Peyman Milanfar ^b

^a Computer Engineering Department, University of California, Santa Cruz, CA 95064, United States

^b Electrical Engineering Department, University of California, Santa Cruz, CA 95064, United States

Received 16 June 2006; received in revised form 8 December 2006; accepted 5 February 2007

Available online 21 February 2007

Abstract

Inverse lithography attempts to synthesize the input mask which leads to the desired output wafer pattern by inverting the forward model from mask to wafer. In this article, we extend our earlier framework for *image prewarping* to solve the mask design problem for coherent, incoherent, and partially coherent imaging systems. We also discuss the synthesis of three variants of phase shift masks (PSM); namely, attenuated (or weak) PSM, 100% transmission PSM, and strong PSM with chrome. A new two-step optimization strategy is introduced to promote the generation and placement of assist bar features. The regularization framework is extended to guarantee that the estimated PSM have only two or three (allowable) transmission values, and the aerial-image penalty term is introduced to boost the aerial image contrast and keep the side-lobes under control. Our approach uses the pixel-based mask representation, a continuous function formulation, and gradient-based iterative optimization techniques to solve the inverse problem. The continuous function formulation allows analytic calculation of the gradient in $O(MN \log(MN))$ operations for an $M \times N$ pattern making it practically feasible. We also present some results for coherent and incoherent imaging systems with very low k_1 values to demonstrate the effectiveness of our approach.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Inverse lithography; OPC; PSM; Pixel-based approach; Non-linear programming; Optimization; Regularization; Low-complexity

1. Introduction and background

1.1. Introduction to lithography

Circuit patterns are commonly transferred on to silicon wafer using optical projection lithography, a process similar to photographic printing. Unfortunately, the optical imaging system is bandlimited and the diffraction effects result in severe loss of the higher frequency components in the projected mask image. Furthermore, the lithography system is subject to random (uncontrollable) process errors in the form of dose and focus changes which affect the *repeatability* of the lithography process. These coupled with Moore's law, which demands a 30% reduction in the

critical dimensions of the printed circuit patterns every 18 months [1], make lithography one of the tightest bottlenecks in the semiconductor industry.

The fundamental limit of the resolution of an optical projection lithography system is given by

$$w_{\min} = \frac{k_1 \lambda}{\text{NA}}, \quad (1)$$

and is related to the Rayleigh criterion¹, where w_{\min} is the minimum line-width of the printed feature. The resolution can be improved by increasing the numerical aperture of the imaging system (NA), decreasing the wavelength (λ), or decreasing the process constant (k_1). Current lithography systems employ the ArF laser having wavelength $\lambda = 193$ nm. Extreme Ultra-Violet (EUV) lithography is a

[☆] Supported by Intel Corporation.

* Corresponding author. Tel.: +1 831 459 4929; fax: +1 831 459 4829.

E-mail addresses: amyn@soe.ucsc.edu (A. Poonawala), milanfar@ee.ucsc.edu (P. Milanfar).

¹ The Rayleigh criterion for the resolution of two point sources is that the first diffraction pattern minimum of the image of one source point falls on the central maximum of the other. This corresponds to a separation of $0.61 \frac{\lambda}{\text{NA}}$.

promising next generation lithography technique which uses EUV waves at 13 nm [4]. Unfortunately, it suffers from major technology, infrastructure, and cost challenges making it practically unfavorable. The physical limit for NA is equal to 1 for dry lithography systems. Recently, immersion lithography has been successfully used to push the NA beyond 1 by introducing a transparent fluid between the lens bottom and the wafer [25]. The focus of this paper however, is on the third alternative, namely decreasing the process constant k_1 using resolution enhancement techniques (RETs) [41,40,16].

At the onset we would like to highlight that the resolution limits for pattern periodicity and pattern dimensions are different [46,21]. A pitch is defined as the sum of line and space width pair and determines the packing density of the transistors. The process constant in this context is referred to as k_{pitch} and is lower bounded by 0.5 [5]. On the other hand, the smallest printable feature size is known as the critical dimension (CD). It dictates the gate-length, speed, and the power consumed by the individual transistors. There are no theoretical limits on k_{CD} and it ultimately depends on our ability to control the CD [46]. In general, lower values of k_{pitch} and k_{CD} make the lithography process more sensitive to dose, focus, and other process variations. They indicate a more aggressive lithography strategy and necessitate the use of RETs.

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 [47]. 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 [39,7,33]. Phase-shift masks (PSM) consist 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 [39,18,34]. Finally, off-axis illumination (OAI) consists of modifying the illuminator (source) size and shape (e.g. quasar, annular, quadrupole, dipole, etc.), which affects the direction of the incident light and ultimately the diffraction orders captured by the lenses [47]. We will be studying the first two RET approaches (OPC and PSM) in this paper.

1.2. Inverse lithography techniques (ILT)

The widely used approach for OPC mask design proposed by Cobb and Zakhor [8] 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 [17]. Furthermore, phase assignments cannot be optimally carried out forming another drawback for 65 nm and smaller nodes.

Hence, there has been a 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* [37,38] 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 first step towards solving an *inverse* problem is to define a *forward* (or process) model which is a (possibly approximate) mathematical description of the given imaging system. The lithography imaging system from the mask to the wafer consists of two steps (see Fig. 1). The aerial image calculations are based on the underlying optical system model (coherent, incoherent, or partially coherent [5,8]). The resist effects are simulated using Dills model [6], Mack Model [20,36], constant threshold resist (CTR) model [14], variable threshold resist (VTR) model [35], or other models [2]. We use the CTR model in our analysis (see Section 2 for more details on our imaging model).

The image formation process can be mathematically expressed as

$$z(x, y) = T\{m(x, y)\}, \quad (2)$$

where $T\{\cdot\}$ is the forward model which maps the input intensity function $m(x, y)$ to the output intensity function $z(x, y)$. Let $z^*(x, y)$ be the desired output intensity function. The goal of inverse lithography technique (ILT) is to estimate the input intensity function which will give us a close

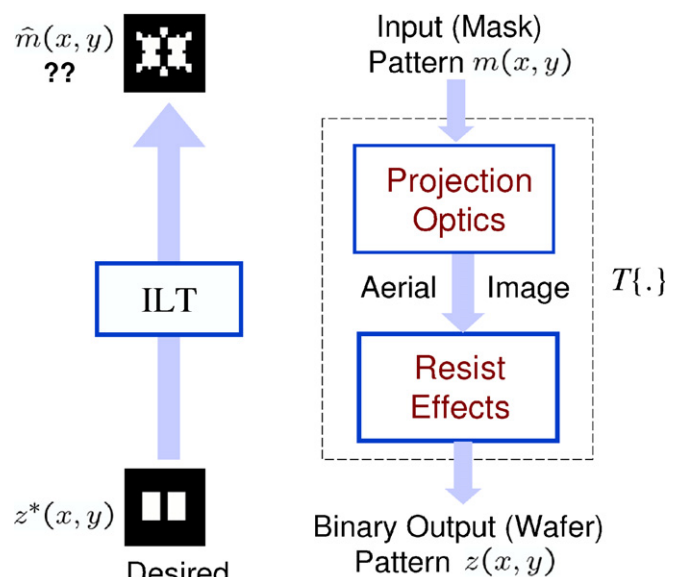


Fig. 1. Forward model and ILT.

approximation to the desired output $z^*(x, y)$ (see Fig. 1). This is achieved by searching the space of all inputs and choosing $\hat{m}(x, y)$ which minimizes a distance $d(z(x, y), z^*(x, y))$, where $d(\cdot, \cdot)$ is some appropriate distance metric to be defined later. Thus,

$$\hat{m}(x, y) = \arg \min_{m(x, y)} d[z^*(x, y), T\{m(x, y)\}]. \quad (3)$$

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

The pioneering work in ILT was by Saleh and his students at the University of Wisconsin in the early 1980s. Sayegh et al. [38] used linear-programming technique to do OPC for a system modeled as a band-limited linear system followed by hard threshold operation. Nashold and Saleh [23] employed iterative alternating projections and Sherif et al. [42] used mixed linear integer programming to synthesize binary masks for the above imaging system. Pati and Kailath [27] approximated the more prevalent partially coherent imaging to a coherent imaging system using optimal coherent approximations and were able to use projection on convex sets (POCS) to synthesize phase shift masks. Liu and Zakhor [18] formulated the mask design problem as the minimization of the L_2 norm of the difference between the ideal and the actual wafer images, and employed branch-and-bound and simulated annealing algorithms to synthesize binary and phase shift masks. Peckerar et al. [28,30,29] also employed L_2 norm based cost function and proposed gradient-based optimization technique to solve the proximity effects arising in the related e-beam lithography problem. Oh et al. [24] used random-pixel flipping and Erdmann et al. [10] proposed genetic algorithms to solve the joint mask and source optimization problem. Granik [13] did a comprehensive review of the past ILT work and discussed the reduction of the mask design problem to linear, quadratic, and general non-linear programming problems. Also Liu et al. [17] demonstrated the commercial viability of inverse lithography techniques.

In [33], we (the authors) proposed a new framework and employed gradient descent algorithm to design a binary input image which reproduces the desired binary image at the output of an imaging system. The above system was modeled as a cascade of low-pass filtering using a Gaussian kernel followed by a hard threshold type operation. We had also introduced a regularization framework to guarantee that the estimated masks were close-to-binary, with low manufacturing complexity.

In this work, we generalize our earlier framework to perform inverse lithography for coherent, incoherent, and partially coherent imaging systems. We have extended our framework to incorporate the mask design for attenuated PSM, 100% transmission PSM, and strong PSM with allowable chrome features. We also discuss a new two-step optimization strategy which favors the generation of assist bars. The sizes of the assist bars and their placements are automati-

cally performed as part of the optimization process. The key difference in our ILT approach (compared to our predecessors) is that we model the mask-to-wafer process using a continuous transfer function. This enables us to formulate the mask synthesis problem using continuous function optimization and use the *gradient* information to systematically exploit the solution space described in (3). In this article, we will discuss the analytic gradient calculations for coherent and incoherent imaging systems. Furthermore, we demonstrate the ability of the regularization framework (introduced in [33]) to control the tone and complexity of the estimated OPC and PSM masks. We also introduce the aerial image penalty term to improve the aerial image contrast and keep the side-lobes under control. The extended framework enables us to automatically arrive at low-complexity, high fidelity, high contrast, discrete-toned OPC masks or PSMs by effectively searching the solution space. The preliminary analysis along with some results for coherent imaging system were reported earlier in [34]. Here, we will present results for the cases of both coherent and incoherent imaging systems for very low k_1 values. The OPC masks for incoherent imaging systems will indicate very interesting *feature splitting* in order to bring the contours on target. The OPC and attenuated PSM examples for coherent imaging systems lead to automatic generation of assist features. Finally, the strong PSM significantly improve the contrast while also achieving good pattern fidelity.

The process model and the optimization problem for different types of imaging systems are formulated in Section 2. The regularization framework and the OPC mask design algorithm are discussed in Section 3. The 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

In this section we discuss the forward (process) model of the lithography system from the (input) mask to the (output) wafer. A typical lithography system can be seen as comprising of two stages; namely the optical (aerial) image formation, and the resist action (see Fig. 1). The aerial image calculations are performed on the basis of coherence, incoherence, or partial coherence of the underlying imaging system and will be discussed in Section 2.1. The simplest way to simulate the resist effect is using the CTR model [14]. Thus, for positive resists, the areas having aerial image intensity higher than threshold t_r are completely removed leaving behind a *space* in the wafer. The above operation can be described using a Heaviside operator (hard threshold) defined as

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

The Heaviside operator brings us into the discrete domain and necessitates the use of branch and bound or other integer

optimization algorithms (like our predecessors [37,18]) to solve the mask design problem. However, the focus of our work is to solve the inverse problem in the continuous domain. With this in mind, we approximate the heaviside operator using a sigmoid: a smooth, continuous function [9]. The approximated forward process model is illustrated in Fig. 2. We employ the logarithmic sigmoid function,

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

where the parameter a dictates the steepness of the sigmoid. 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. Fig. 3 illustrates the behavior of the sigmoid for different values of a with $t_r = 0.5$. A large value of a leads to a very steep sigmoid which closely resembles the hard thresholding operation. Owing to the above approximation, the output pattern \underline{z} will not be binary, but a (continuous tone) “close-to-binary” pattern. 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 2D discrete images. We define vectors $\underline{z}^*, \underline{z}, \underline{m} \in \mathfrak{R}^{MN \times 1}$ which are obtained by sampling and lexicographic ordering of $z^*(x, y)$, $z(x, y)$, and $m(x, y)$, respectively. The number of samples along the horizontal and vertical directions are given by M and N , respectively. Note that the edge placement error is related to the sampling interval δ_s nm and is upper-bounded by $(\delta_s/2)$ nm. Smaller pixels enable better edge placement but also increase the number of samples M and N which are related to the algorithmic complexity by $O(MN \log(MN))$. Thus, there is a trade-off between speed and accuracy. Throughout our discussion, \underline{z}^* represents the prescribed *binary* pattern, \underline{z} represents the *gray-level* output pattern, and \underline{m} represents the

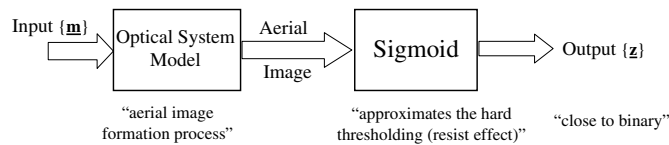


Fig. 2. Approximated forward process model.

input pattern fed to the imaging system (which can be *binary* or *gray-level*).

2.1. Modelling the imaging system

We now discuss the individual aerial image calculations and the forward 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 $e(x, y)$ is linearly related to the input electric field amplitude generated by the mask $m(x, y)$. This can be mathematically described as

$$e(x, y) = m(x, y) * h(x, y) \quad (6)$$

where $h(x, y)$ is referred to as the amplitude spread function (ASF) of the given imaging system [45].

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 [5,45]. Therefore,

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

where $r = \sqrt{x^2 + y^2}$ and $J_1(\cdot)$ is the first-order 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 $e(x, y)$. Therefore, the forward model is defined as

$$\underline{z}^c = \text{sig}(|\underline{e}|^2) = \text{sig}(|\underline{H}\underline{m}|^2), \quad (8)$$

where \underline{z}^c represents the output pattern for coherent imaging systems. The sigmoid function simulates the resist behavior and acts on the aerial image $|\underline{H}\underline{m}|^2$ (square of the amplitude), giving the output pattern \underline{z}^c . Two important points to note: the kernel $\underline{H} \in \mathfrak{R}^{MN \times MN}$ is the jinc function $h(x, y)$ sampled using the same sampling rate as \underline{z}^c , and $|\cdot|^2$ operator here implies element-by-element absolute

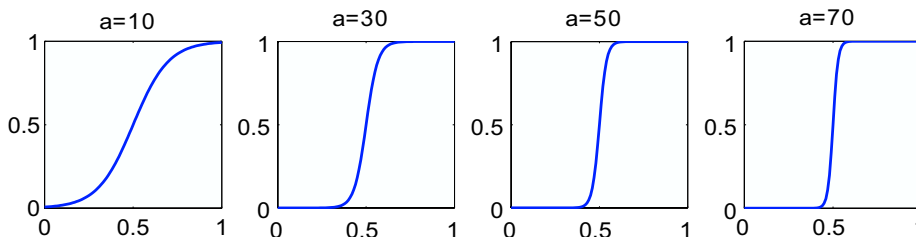


Fig. 3. The effect of the steepness parameter a on the sigmoid function $\text{sig}(u) = 1/(1 + e^{-a(u-0.5)})$.

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 and Kailath [27].

2.1.2. Incoherent imaging system

An incoherent imaging system is linear in intensity (or irradiance) and can be mathematically described as [5],

$$|e(x, y)|^2 = |m(x, y)|^2 * |h(x, y)|^2 \quad (9)$$

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 our attention to only binary masks. The photo-resist directly responds to the above electric field intensity and the forward model for a binary mask is defined as

$$\underline{\mathbf{z}}^i = \text{sig}(\tilde{\mathbf{H}}|\underline{\mathbf{m}}|^2) = \text{sig}(\tilde{\mathbf{H}}\underline{\mathbf{m}}), \quad (10)$$

where $\underline{\mathbf{z}}^i$ represents the output pattern for incoherent imaging systems. The filter $\tilde{\mathbf{H}}$ in (10) is known as the point spread function of the imaging system and is a jinc-squared function. It is defined as the square of the PSF shown in (7). Also, since the mask is binary, $|\underline{\mathbf{m}}|^2 = \underline{\mathbf{m}}$.

2.1.3. Partially coherent imaging system

Real-world lithography systems are partially coherent and can be modelled using the Hopkins diffraction model [5]. Pati and Kailath [27] 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 P th order approximation to the aerial image formulation can be calculated using the weighted sum of P coherent systems. The forward model now becomes,

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

where \mathbf{H}_j for $j = 1, \dots, P$ are the amplitude spread functions (also referred to as optical system kernels) of the coherent systems, and $\sigma_1, \dots, \sigma_P$ are the corresponding 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 fully coherent and incoherent 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°) phase shift. Therefore, $q_j = 0$ for $j = 1, \dots, MN$ 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,

$$\begin{aligned} \hat{\underline{\mathbf{m}}} &= \arg \min_{\underline{\mathbf{m}}} \{F(\underline{\mathbf{m}})\} = \arg \min_{\underline{\mathbf{m}}} \|\underline{\mathbf{z}}^* - \underline{\mathbf{z}}\|_2^2 \\ &= \arg \min_{\underline{\mathbf{m}}} \sum_{k=1}^{MN} (z_k^* - z_k)^2. \end{aligned} \quad (12)$$

Later in Section 3, we refine this approach by introducing the regularization terms and augmenting the cost function.

2.2.1. Coherent imaging system

For coherent imaging system $\underline{\mathbf{z}} = \underline{\mathbf{z}}^c$ and the cost function can be formulated using (12) as follows:

$$\begin{aligned} F(\underline{\mathbf{m}}) &= F^c(\underline{\mathbf{m}}) = \arg \min_{\underline{\mathbf{m}}} \|\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^c\|_2^2 \\ &= \arg \min_{\underline{\mathbf{m}}} \sum_{k=1}^{MN} (z_k^* - z_k^c)^2. \end{aligned} \quad (13)$$

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^c in (13) can be represented as

$$z_k^c = \frac{1}{1 + \exp\left[-a\left(\sum_{j=1}^{MN} h_{kj} m_j\right)^2 + at_r\right]}, \quad (14)$$

for $k = 1, \dots, MN$ where h_{kj} are the elements of the k th row of \mathbf{H} , which 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 $j = 1, \dots, MN$ 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 (13) is therefore subject to the constraints given by the allowable transmission values of m_j . This unfortunately makes the search space discrete thereby again reducing our problem into an integer optimization one. To overcome this issue and move back into the continuous domain, we relax the parameter values to lie within a range $[\underline{m}, \bar{m}]$. Thus, the discrete equality constraints are substituted by the inequality (bound) constraints and the optimization problem in (13) is made subject to

$$\underline{m} \leq m_j \leq \bar{m} \quad \text{for } j = 1, \dots, MN. \quad (15)$$

The above procedure reduces the mask design problem into a bound constrained optimization problem which will be henceforth employed in our discussion.

2.2.2. Incoherent imaging system

For an incoherent imaging system $\underline{z} = \underline{z}^i$ and the cost function can be formulated using (12) as follows:

$$\begin{aligned} F(\underline{m}) &= F^i(\underline{m}) = \arg \min_{\underline{m}} \|\underline{z}^* - \underline{z}^i\|_2^2 \\ &= \arg \min_{\underline{m}} \sum_{k=1}^{MN} (z_k^* - z_k^i)^2 \end{aligned} \quad (16)$$

From (10) we observe that every pixel in an incoherent imaging system undergoes a cascade of convolution followed by a sigmoidal operation. Therefore, the z_k^i in (16) is given as,

$$z_k^i = \frac{1}{1 + \exp\left(-a \sum_{j=1}^{MN} \tilde{h}_{kj} m_j + at_r\right)}, \quad (17)$$

for $k = 1, \dots, MN$ where \tilde{h}_{kj} are the elements of the k th row of $\tilde{\mathbf{H}}$, which is the jinc-squared function.

As discussed earlier, we are only interested in binary masks, and the optimization problem in (16) is subject to the bound constraints

$$0 \leq m_j \leq 1 \quad \text{for } j = 1, \dots, MN. \quad (18)$$

3. OPC mask design algorithm and regularization framework

In this section, we discuss the proposed optimization algorithm for synthesizing OPC masks for coherent and incoherent imaging systems. We also discuss the regularization framework to guarantee near binary results, ease of manufacturing, and good quality aerial images. Finally, we will present results for OPC mask synthesis for the above cases.

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 in (13) subject to the constraints in (18). The above problem can be solved using constrained optimization algorithms like BFGS (Broyden, Fletcher, Goldfarb, and Shanno) or gradient-projection [15,33]. The bound-constrained optimization problem can be further reduced to an unconstrained optimization problem using the following parametric transformation,

$$m_j = \frac{1 + \cos(\theta_j)}{2} \quad \text{for } j = 1, \dots, MN \quad (19)$$

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

$$F_1^c(\underline{\theta}) = \sum_{k=1}^{MN} \left(z_k^* - \frac{1}{1 + \exp\left[-a \left(\sum_{j=1}^{MN} h_{kj} \frac{1 + \cos(\theta_j)}{2} \right)^2 + at_r\right]} \right)^2. \quad (20)$$

We can now employ steepest-descent search to minimize the above cost function. This requires the first-order derivatives of (20), and the gradient vector $\underline{d}^c = \nabla F_1^c(\underline{\theta}) \in \mathfrak{R}^{MN \times 1}$ can be analytically calculated using the following expression:

$$\begin{aligned} \nabla F_1^c(\underline{\theta}) &= \underline{d}^c \\ &= a(\mathbf{H}^T[(\underline{z}^* - \underline{z}^c) \odot \underline{z}^c \odot (\underline{\mathbf{1}} - \underline{z}^c) \odot (\mathbf{H}\underline{m})]) \\ &\quad \odot \sin(\underline{\theta}), \end{aligned} \quad (21)$$

where \odot is the Hadamard product (element-by-element multiplication) of the two vectors, $\underline{\mathbf{1}} = [1, \dots, 1]^T$, and \underline{z}^c is defined in (8). Note that the gradient calculation involves two convolution operations which dictate the complexity of our algorithm. Thus, the algorithmic complexity is $O(MN \log(MN))$.

The cost-function defined in (20) is a quartic function and is non-convex with multiple local-minima (as also noted by Granik [13]). Since, we are using local gradient-based search technique; there is no guarantee of reaching the global minimum. However, ILT is an ill-posed problem and it is often not necessary to arrive at the global optima (see [26]). Any *good local minimum* (where the *goodness* is defined using data-fidelity and user-desired properties), can suffice as an acceptable solution. In the next section, we introduce the regularization framework to incorporate the above requirement.

Returning to the optimization problem at hand, the n^{th} iteration of steepest descent algorithm is given as

$$\underline{\theta}_{n+1} = \underline{\theta}_n - s \underline{d}_n^c, \quad (22)$$

where s is the step-size. The algorithm is initialized at $\underline{\theta}^0 = \cos^{-1}(2\underline{z}^* - \underline{\mathbf{1}})$. Here, we would like to highlight the useful fact that due to the structure of (21), the steepest descent iterations can be quickly and directly carried out on the 2D image array (matrices) with no need for the raster scanning operation [33]. This saves valuable time and considerably eases the implementation.

For incoherent imaging systems, the optimization problem is defined in (16) subject to the constraints given by (18). We follow the parametric transformation approach discussed above to obtain $F_1^i(\underline{\theta})$ and the gradient vector in this case is given as

$$\begin{aligned} \nabla F_1^i(\underline{\theta}) &= \underline{d}^i \\ &= a(\mathbf{H}^T[(\underline{z}^* - \underline{z}^i) \odot \underline{z}^i \odot (1 - \underline{z}^i)]) \odot \sin(\underline{\theta}). \end{aligned} \quad (23)$$

Finally, we would like to highlight that the estimated mask \underline{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 $\hat{\mathbf{m}}$ into a physically realizable mask $\hat{\mathbf{m}}_b$ [33]. In the next section, we propose an alternative to further simplify the above process.

3.2. Regularization framework

As we noted earlier, the inverse lithography problem is an ill-posed problem [44]. The continuous function formulation implies that there can be infinitely many input (gray level) patterns all giving rise to the same binary pattern at the output. Similarly, there may be many discrete-tone masks all capable of providing good contour fidelity for a given pattern. Since multiple solutions exist, our goal is to choose a solution which is more favorable to us compared to others. For example, continuous tone masks are physically not realizable and we are only interested in solutions comprising of two or three tones. Furthermore, we want the synthesized mask patterns to have low-complexity in order to control the mask manufacturing costs. The user may also have other requirements like low MEEF (mask error enhancement factor), large process-window, minimum feature spacing, etc. [26]. In general, we may want to inculcate (or promote) certain desirable properties in the solution. The regularization framework [44] incorporates these requirements as *prior* information about the solution, and helps us arrive at the *preferred* solution. The “regularized” problem formulation can be described as follows:

$$\hat{\mathbf{m}} = \arg \min_{\mathbf{m}} [\gamma_{\text{fid}} F(\mathbf{m}) + \gamma_{\text{reg}} R(\mathbf{m})], \quad (24)$$

where $F(\mathbf{m})$ is the data-fidelity term, and $R(\mathbf{m})$ is the regularization function (or the penalty term) used to direct the unknown parameter \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. The prior knowledge is contained in the penalty term $R(\mathbf{m})$ and solutions in closer agreement with the prior are penalized less compared to others.

The regularization framework was first employed in the context of lithography by Peckerar et al. [30,29] to solve the proximity effect problem arising in e-beam lithography. The estimated dose was obtained by solving an unconstrained continuous function optimization problem using gradient-descent. This led to impractical negative dose values which was overcome by employing a regularization framework similar to (24). The above framework was also successfully used by the authors of the present article for designing low-complexity binary masks for optical microlithography in [33]. Here, we provide a brief review of the earlier employed penalty terms, and also introduce a new aerial image penalty term to improve the robustness of the lithography process.

3.2.1. Discretization penalty term

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. 4)

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

The penalty incurred is zero for transmission values 0 or 1 and increases as we move away from these values 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:

$$\begin{aligned} R_{\text{dis}}(\mathbf{m}) &= \sum_{j=1}^{MN} r(m_j) = \sum_{j=1}^{MN} [1 - (2m_j - 1)^2] \\ &= 4\mathbf{m}^T(\mathbf{1} - \mathbf{m}). \end{aligned} \quad (25)$$

3.2.2. Complexity penalty term

The pixel-based approach allows 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. Liu and Zakhor addressed this issue in the past using a cell-based approach [19]. The cells are selected and moved around either randomly or using the knowledge from previous moves. Researchers in the past have also reverted to post-processing operations to simplify the output [24], but this approach is 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. Thus, we integrate a mask simplicity criterion into the optimization objective and inherently favor low-complexity masks while exploring the search space.

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) [44] penalty, we choose to penalize the mask complexity using the local variation of the mask as follows [11]:

$$\|\nabla \mathbf{m}\|_1 = \|\mathbf{Q}_x \mathbf{m}\|_1 + \|\mathbf{Q}_y \mathbf{m}\|_1, \quad (26)$$

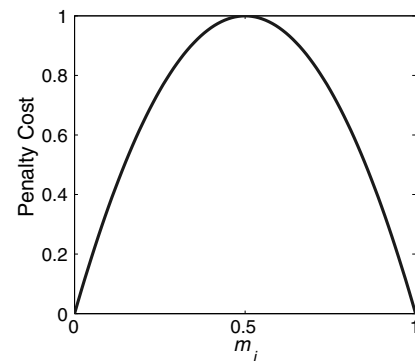


Fig. 4. Discretization penalty term for binary masks.

where $\mathbf{Q}_x, \mathbf{Q}_y \in \mathfrak{R}^{MN}$ 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 the image represented by $\underline{\mathbf{m}}$ along horizontal (right) and vertical (up) direction by one pixel, respectively.

We penalize the L_1 norm² of the gradient of the synthesized mask. Isolated holes, protrusions, and jagged edges will contribute more towards the gradient and as such will have higher associated penalty. The regularization term in (26) suppresses these effects and forces the changes to be spatially smoother and less abrupt. The L_1 norm based penalty term inherently favors piece-wise constant features thereby making the masks easier to print using e-beam and laser pattern generators.

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. Similarly, slight variations in the lithography process should not result in printing the side-lobes. To achieve the above goal, we employ a penalty term defined as the L_2 norm of the difference between the desired (binary) pattern and the aerial image obtained using the input mask. For a coherent imaging system, the penalty term is given as

$$R_{\text{aerial}}(\underline{\mathbf{m}}) = \|\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{m}}|^2\|_2^2 \quad (27)$$

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

$$\nabla R_{\text{aerial}}(\underline{\mathbf{m}}) = -\mathbf{H}[(\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{m}}|^2) \odot (\mathbf{H}\underline{\mathbf{m}})]. \quad (28)$$

Note that the optimization problem (13) can also be formulated using (27) as the data fidelity term, that is, $F(\underline{\mathbf{m}}) = F^c(\underline{\mathbf{m}}) = R_{\text{aerial}}(\underline{\mathbf{m}})$. 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.

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

$$J(\underline{\mathbf{m}}) = \gamma_{\text{fid}}F(\underline{\mathbf{m}}) + \gamma_{\text{aerial}}R_{\text{aerial}}(\underline{\mathbf{m}}) + \gamma_{\text{dis}}R_{\text{dis}}(\underline{\mathbf{m}}) + \gamma_{\text{TV}}R_{\text{TV}}(\underline{\mathbf{m}}), \quad (29)$$

where γ_{aerial} , γ_{dis} , and γ_{TV} are the regularization weights corresponding to their respective penalty terms.

3.3. An alternative two-step strategy for optimization

A closer look at the above regularization terms indicates that on some occasions they tend to conflict and suppress certain type of features in the estimated mask. For exam-

ple, assist bars have been long known to improve the contrast of the aerial image and making the lithography process more robust [31]. The gradient descent optimization is usually initialized at $\underline{\mathbf{m}}^0 = \underline{\mathbf{z}}^*$. In such cases, the assist bar generation implies a switch from 0 to 1 in regions away from the main feature in the estimated mask. However, in the continuous domain, any movement away from 0 is penalized by the quadratic penalty term $R_{\text{dis}}(\underline{\mathbf{m}})$. This tends to counteract the generation of assist-bars. We also observe from our experiments (see Fig. 8) that often the best contour fidelity and aerial image contrast is obtained by *breaking* the feature into two disjoint ones. This effect is again suppressed by both the complexity and discretization penalty terms.

Hence, we propose an alternative optimization strategy which involves distributing the optimization objectives into two steps as described below.

Step 1 In the first step, the data fidelity are defined as the aerial image fidelity and we aim to minimize the cost function,

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

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 since the data fidelity does not account for the contour fidelity, the resulting contours will not be on target yet.

Step 2 We next minimize the augmented cost function defined in (29) which consists of the contour fidelity, complexity, discretization, and aerial image penalty terms. The key point is that the steepest descent algorithm for step two is initialized using the estimated mask pattern obtained from step one. As stated earlier, the phase assignments for strong PSM and generation of assist-bars for binary and AttPSM masks occurs during the first step. In step two, we start exploring the search space from the above solution point and incorporate the other objectives.

3.4. OPC results

We now demonstrate some results for the case of binary masks for both coherent and incoherent imaging systems.

3.4.1. Coherent imaging

Fig. 5 illustrates the estimated masks and the binary output patterns obtained using a coherent imaging system with $\sigma = 0$, $\lambda = 193$ nm, $\text{NA} = 0.85$, and the resist threshold $t_r = 0.3$. The desired pattern consists of 90 nm random logic contact holes sampled at 10 nm ($k_1 = 0.39$). We follow the two step optimization approach and Fig. 5 illustrates the estimated mask obtained at the end of step one. The result was obtained in 60 iterations with $s = 20$ and the run-time was 22 s (all run-times reported were

² L_1 norm of a vector is defined as the sum of the absolute values of the vector elements.

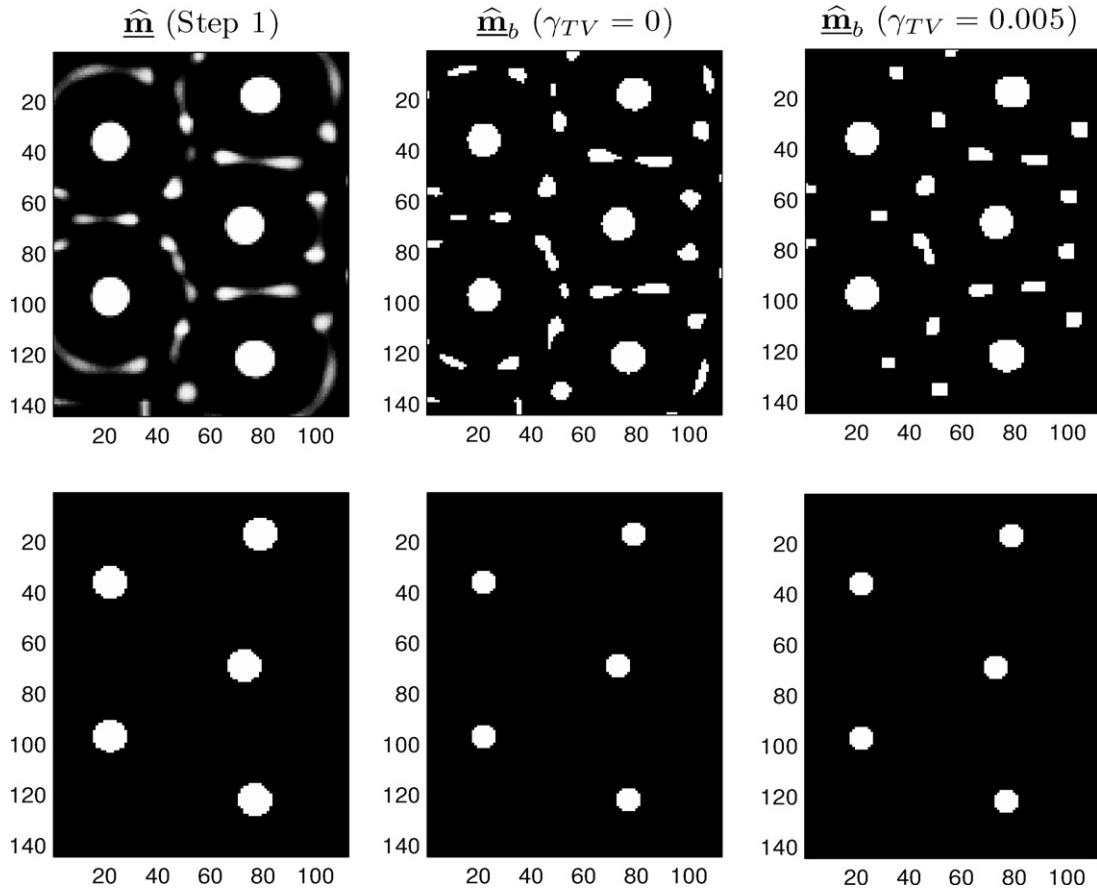


Fig. 5. The synthesized masks (top row) and the output binary patterns (bottom row) for 90 nm random logic contact patterns. $\hat{\mathbf{m}}_b$ (left) is the estimated mask at the end of Step 1 ($\gamma_{\text{aerial}} = 1$), $\hat{\mathbf{m}}_b$ (center) is estimated at the end of Step 2 ($\gamma_{\text{fid}} = 1$, $\gamma_{\text{aerial}} = 0.25$, $\gamma_{\text{dis}} = 0.005$, $\gamma_{\text{TV}} = 0$) and $\hat{\mathbf{m}}_b$ (right) is estimated using the above parameters with $\gamma_{\text{TV}} = 0.005$.

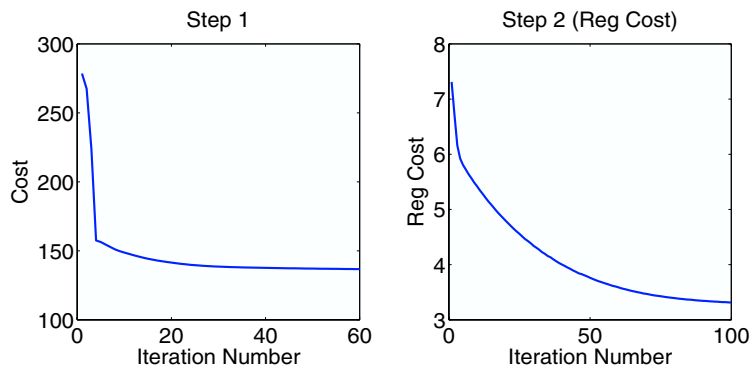


Fig. 6. Convergence curves for Fig. 5. The left curve indicates the cost function behavior for Step 1 and the right curve indicates the regularization cost ($\gamma_{\text{dis}} R_{\text{dis}}(\mathbf{m}) + \gamma_{\text{TV}} R_{\text{TV}}(\mathbf{m})$) versus iteration number for Step 2 of the optimization.

calculated on a 1.3 GHz Pentium-M machine using Matlab). The cost function behavior is illustrated in Fig. 6 and indicates quick convergence. We have observed that the algorithm scales the center feature during the initial iterations and adds assist bars during the remaining ones.³

³ The above effect of scaling of the main feature was also recently reported in the ILT-based assist bar generation work in [22].

Note that the assist features are also shared between adjacent contacts and the mask has continuous tone. Furthermore, the output wafer pattern (bottom row) obtained using the above mask does not have the contours on target. This is not surprising since the contour fidelity was not considered as an optimization objective.

The estimated mask in the center in Fig. 5 illustrates the result at the end of the second step. The parameters are: $\gamma_{\text{fid}} = 1$, $\gamma_{\text{aerial}} = 0.25$, and $\gamma_{\text{dis}} = 0.005$. We observe that

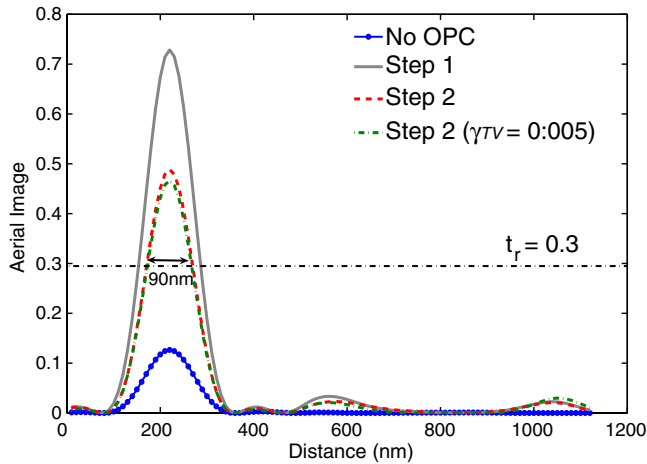


Fig. 7. Horizontal slice at row number 35 for the aerial image obtained using the synthesized mask in Fig. 5. Note that the contrast has improved and the side-lobes will not print.

the mask is having discrete tone and the output contours for all the contacts are on target. Fig. 7 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 second step by employing the complexity penalty term ($\gamma_{TV} = 0.005$). This leads to a comparatively simple pattern (see $\hat{\mathbf{m}}_b$ on the right in Fig. 5) and the assist bars tend to become square or rectangular in shape, which are more preferable. The above result was obtained by performing 100 iterations (see Fig. 6) with $s = 8$ and the run-time was 60 s.

3.4.2. Incoherent imaging

Fig. 8 illustrates the estimated masks and the binary output patterns obtained using an incoherent imaging system with $\sigma = 1$, $\lambda = 193$ nm, $NA = 0.95$, and the resist threshold $t_r = 0.5$. The desired pattern is a more complicated pattern sampled at 5 nm with features as small as 50 nm ($k_1 = 0.25$). The top row indicates the input patterns and the bottom row indicates the corresponding output wafer patterns. The center mask was obtained without any regularization with only $\gamma_{fid} = 1$. We observe that although the contours are on target, the mask itself is very choppy and irregular, making it very hard to manufacture. Hence, we employ the two step procedure outlined earlier and optimize using the augmented cost function (29) with $\gamma_{fid} = 1$, $\gamma_{TV} = 0.1$ and $\gamma_{dis} = 0.01$. This results in an estimated mask which is comparatively much smoother with little loss in performance. It is also interesting to note that the algorithm automatically decided to *break* some features into two disjoint sections.⁴ For example, the elbow on the

top-left region is split into two parts in $\hat{\mathbf{m}}_b$, but is reproduced accurately at the output. Such counter-intuitive results are hard to obtain using edge-based parametrization and would require extensive and tedious segmentation scripts. The results were obtained using 100 iterations (run-time 55 s) for each step.

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 PSM, 100% transmission PSM, and strong PSM with chrome.

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° phase shift with weak transmission) or 1 (100% transmission with no phase shift). The optimization problem is formulated in (13) subject to the constraints that $m_j = -0.245$ or 1. We reduce it to a non-linear optimization problem with 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

$$m_j = 0.6225(1 + \cos(\theta_j)) - 0.245. \quad (30)$$

The steepest descent algorithm is initialized using the target pattern (similar to the OPC case).

The quadratic penalty term is again 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. 9),

$$r(m_j) = -m_j^2 + 0.755m_j + 0.245. \quad (31)$$

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

4.2. 100% transmission PSM

As the name suggests, 100% transmission PSM does not use the opaque chrome features at all and is an extreme case of AttPSM. 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. The optimization problem is formulated in (13) subject to the constraints 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

⁴ The above effect of *breaking* of features was also observed for partially coherent imaging systems in [17].

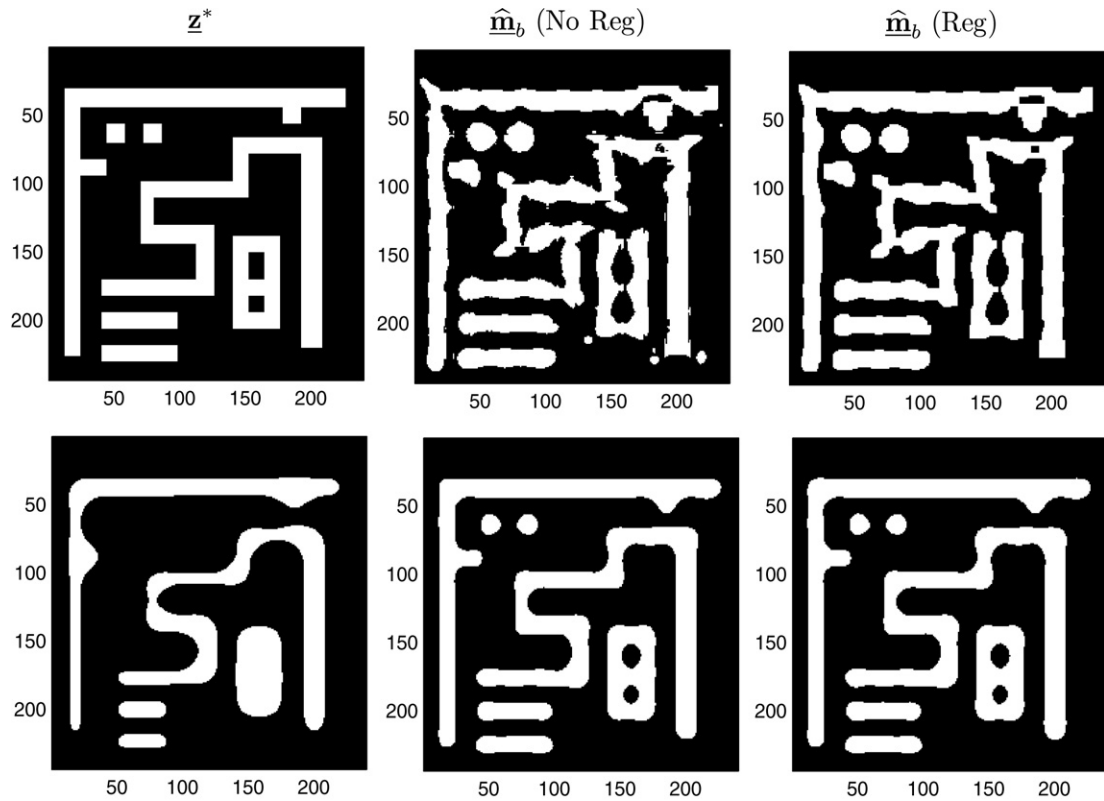


Fig. 8. The top row consists of the original pattern and the synthesized masks before and after regularization. The bottom row indicates the corresponding output wafer patterns. \hat{m}_b (center) is estimated using $\gamma_{fid} = 1$ and \hat{m}_b (right) is obtained using $\gamma_{fid} = 1$, $\gamma_{dis} = 0.01$, and $\gamma_{TV} = 0.1$.

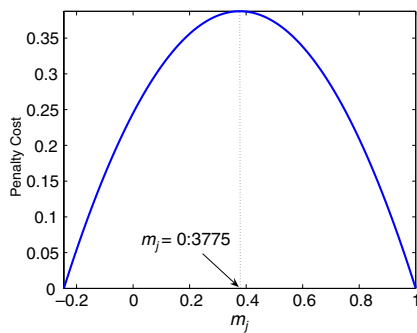


Fig. 9. Discretization penalty term for AttPSM (maximum penalty is at $m_j = 0.3775$).

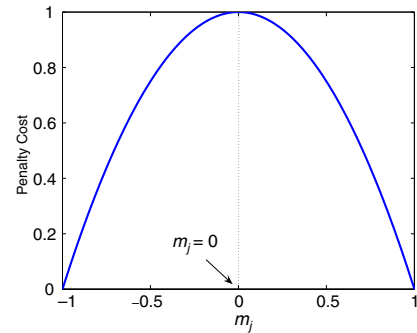


Fig. 10. Discretization term for 100% transmissive PSM (maximum penalty is at $m_j = 0$).

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. 10)

$$r(m_j) = -m_j^2 + 1. \tag{32}$$

Hence, the transmission values are pushed towards -1 and 1 thereby easing the discretization step. The two tone 100% transmission mask \hat{m}_b is obtained by thresholding the estimated mask \hat{m} with $t_m = 0$. It is important to note that in case of strong PSM the estimated mask seldom resembles the target (see Figs. 13, 16, and 17). Hence, the steepest descent algorithm was initialized to all zeros, thereby allowing the algorithm to automatically perform the phase assignments for all the regions.

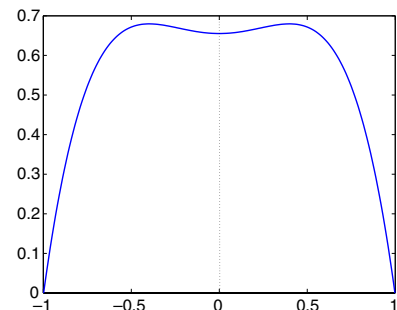


Fig. 11. Discretization penalty cost for strong PSM (with chrome). The minima are at -1 , 0 , and 1 .

4.3. Strong PSM with chrome

The final case we consider 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. There can be different ways of formulating the above regularization term. One possible approach is to divide the interval $[-1\ 1]$ into three equal parts and

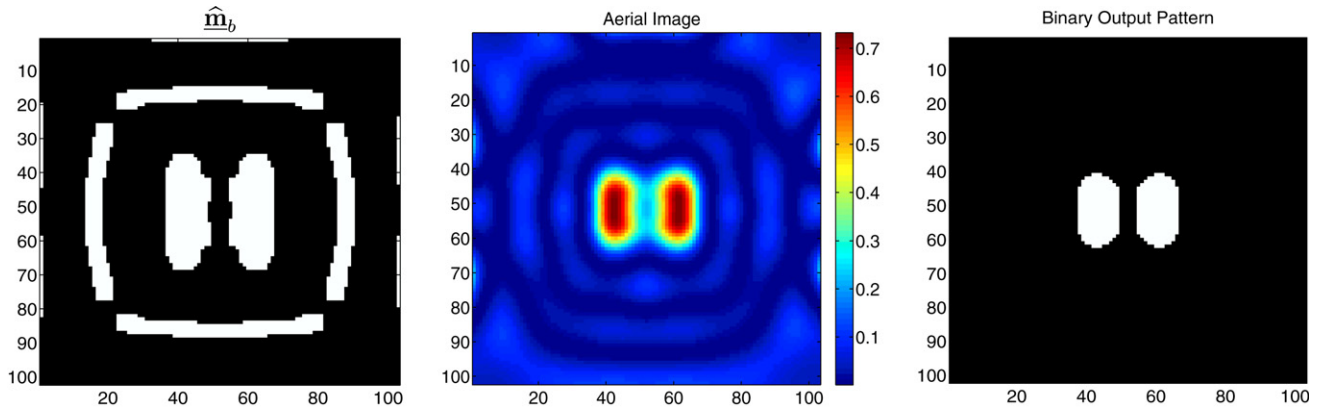


Fig. 12. The estimated 6% AttPSM mask (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193\text{ nm}$ and $\text{NA} = 0.85$. The black and white regions in $\hat{\mathbf{m}}_b$ correspond to -0.245 and 1 , respectively.

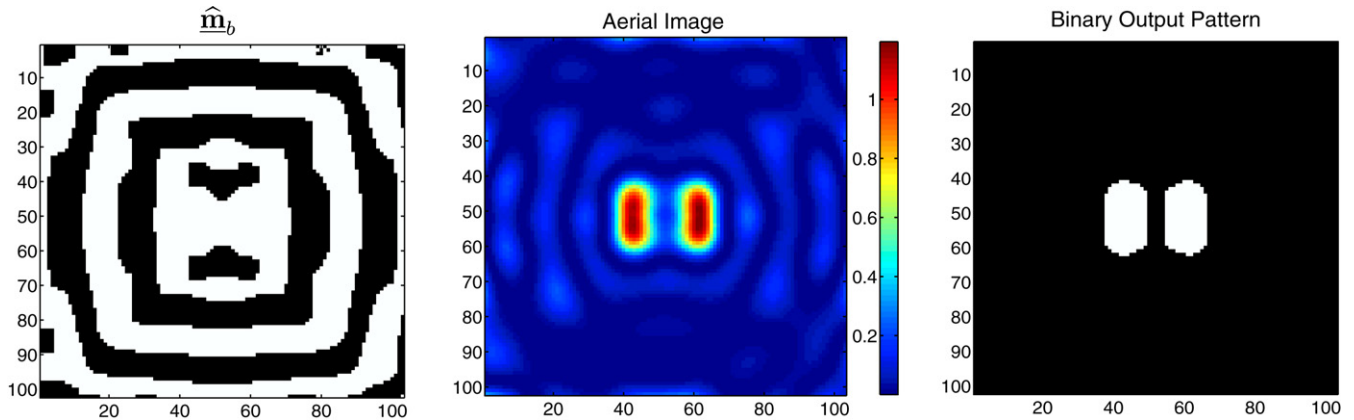


Fig. 13. The estimated 100% transmission PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193\text{ nm}$ and $\text{NA} = 0.85$. The black and white regions in $\hat{\mathbf{m}}_b$ correspond to -1 and 1 , respectively.

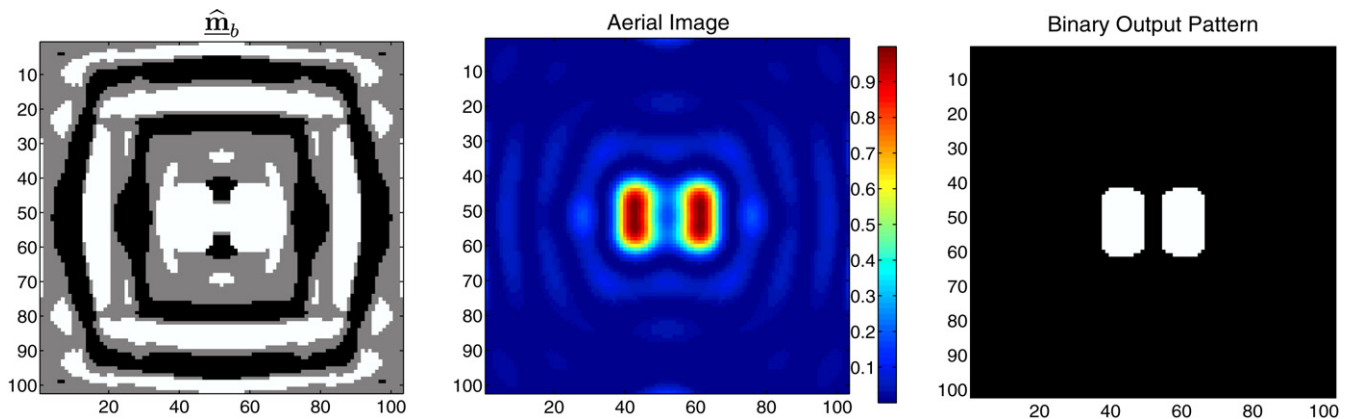


Fig. 14. The estimated strong PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for a coherent imaging system with $\lambda = 193\text{ nm}$ and $\text{NA} = 0.85$. The black, gray, and white regions in $\hat{\mathbf{m}}_b$ correspond to -1 , 0 , and 1 , respectively.

employ a fourth order quartic penalty term where each pixel has an associated cost,

$$r(m_j) = -0.967m_j^4 + 0.307m_j^2 + 0.655. \quad (33)$$

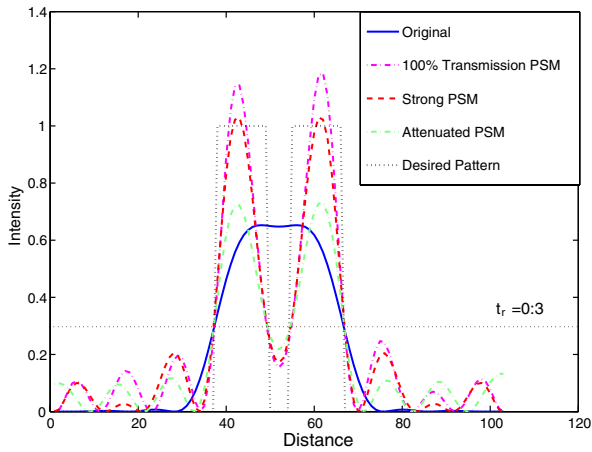


Fig. 15. Horizontal slices along the center of the aerial images obtained using the desired pattern \underline{z}^* and the various synthesized PSMs in Figs. 12–14 as inputs.

Fig. 11 is the plot of the above function. The above curve was obtained by fitting a fourth order polynomial to an over-determined linear system. The latter was formulated to obtain stationary points at $m_j = -0.33, 0,$ and 0.33 , zero penalty at $m_j = -1, 0,$ and 1 , and high penalty at $m_j = -0.33$ and 0.33 . The required three-tone mask $\hat{\underline{m}}_b$ can be finally obtained by quantizing $\hat{m}_j \in [-1, -0.33)$, $\hat{m}_j \in [-0.33, +0.33)$, and $\hat{m}_j \in [+0.33, 1]$ to $-1, 0,$ and 1 , respectively.

4.4. Results for PSM

We now demonstrate some results for the above discussed PSM design framework.

4.4.1. Experiment 1

In the first experiment, our goal is to print two 120 nm thick bars separated by 50 nm ($k_1 = 0.22$) with high contour fidelity. Fig. 12 illustrates the synthesized 6% attenuated PSM for a coherent imaging system with $\sigma = 0$, $\lambda = 193$ nm and $NA = 0.85$. We employ the two-step optimization strategy outlined in Section 3.3 and the experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\gamma_{fid} = 1$, $\gamma_{aerial} = 0.25$, $\gamma_{TV} = 0.01$, $\gamma_{dis} = 0.01$, and $s = 5$.

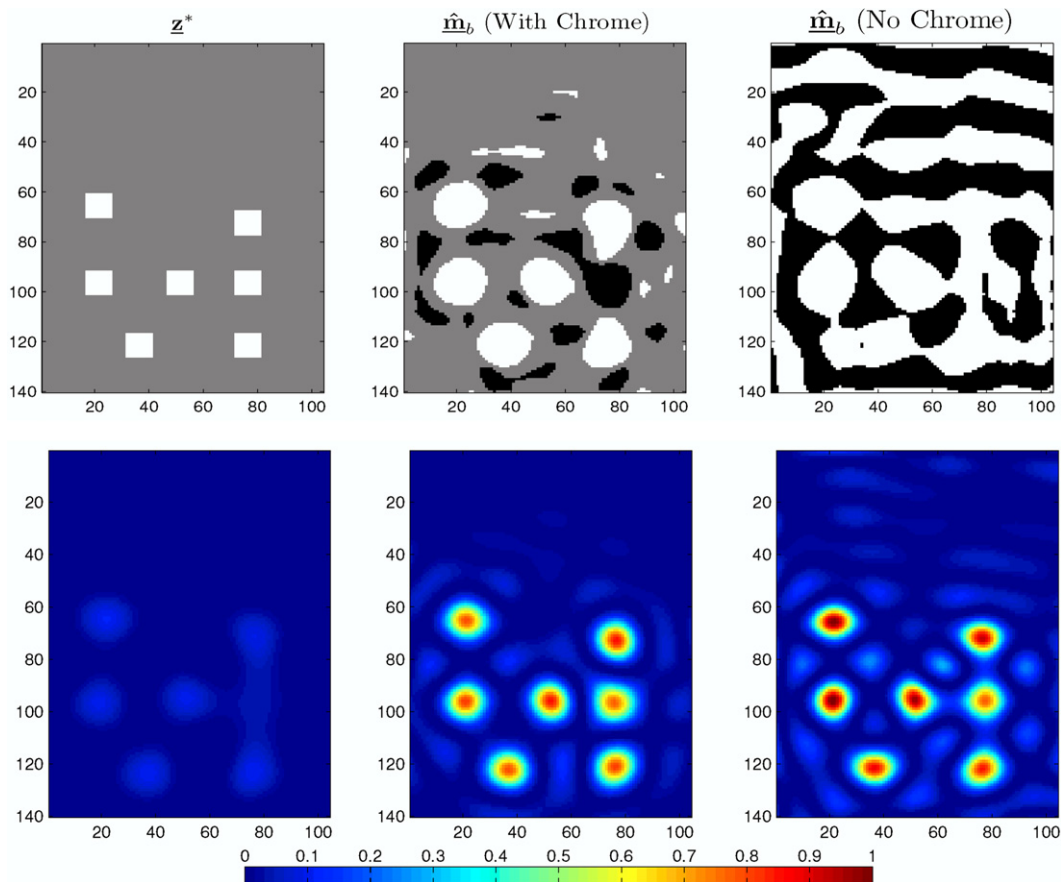


Fig. 16. The top row consists of the desired pattern, the estimated strong PSM with chrome, and the 100% transmission PSM. The black, gray, and white regions correspond to transmission values of $-1, 0,$ and 1 , respectively. The bottom row consists of the aerial images corresponding to the masks in the top row. Here $\lambda = 193$ nm, $NA = 0.7$, and $\sigma = 0$.

The number of iterations for the two steps were 80 and 160, and the run-times were 30 and 60 s, respectively. 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. 15 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. 13 illustrates the synthesized 100% transmission PSM for the same imaging system. The experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\gamma_{\text{fid}} = 0.75$, $\gamma_{\text{aerial}} = 0.25$, $\gamma_{\text{TV}} = 0.01$, $\gamma_{\text{dis}} = 0.01$, and $s = 5$. The number of iterations for the two steps were 120 and 240, and the run-times were 45 and 90 s, respectively. 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. 15. The result also demonstrates that our algorithm can produce synthesized masks which can be very different from the desired patterns.

Finally, Fig. 14 illustrates the result using a strong PSM with chrome. The experimental parameters are as follows:

$a = 25$, $t_r = 0.3$, $\gamma_{\text{fid}} = 1$, $\gamma_{\text{aerial}} = 0.25$, $\gamma_{\text{TV}} = 0$, $\gamma_{\text{dis}} = 0.01$, and $s = 1.5$. The number of iterations for the two steps were 80 and 160, and the run-times were 30 and 60 s, respectively. 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. 15 demonstrate an improvement in the contrast. Furthermore the side-lobes are below the resist threshold t_r and hence they will not print.

4.4.2. Experiment 2

In the second experiment, our goal is to improve the contrast of the aerial image and distinguish the 100 nm random logic contacts (see Fig. 16) using strong PSMs. The lithography system parameters are: $\sigma = 0$, $\lambda = 193$ nm, and $\text{NA} = 0.7$ ($k_1 = 0.36$). The first row in Fig. 16 indicates the desired pattern (\underline{z}^*) and the estimated discrete tone masks ($\underline{\hat{m}}_c$) for strong PSM with chrome and 100% transmission PSM. The colors black, gray, and, white correspond to transmission values -1 , 0 , and $+1$, respectively. Note that in this case we are only concerned about matching the aerial image without worrying about the resist effects. Hence, we set the contour fidelity term $\gamma_{\text{fid}} = 0$. The remaining experimental parameters for strong PSM with chrome are $\gamma_{\text{aerial}} = 1$, $\gamma_{\text{dis}} = 0.03$, $\gamma_{\text{TV}} = 0$, $s = 7$, and number of iterations = 100. The parameters for 100% transmission PSM are $\gamma_{\text{aerial}} = 1$, $\gamma_{\text{dis}} = 0.01$,

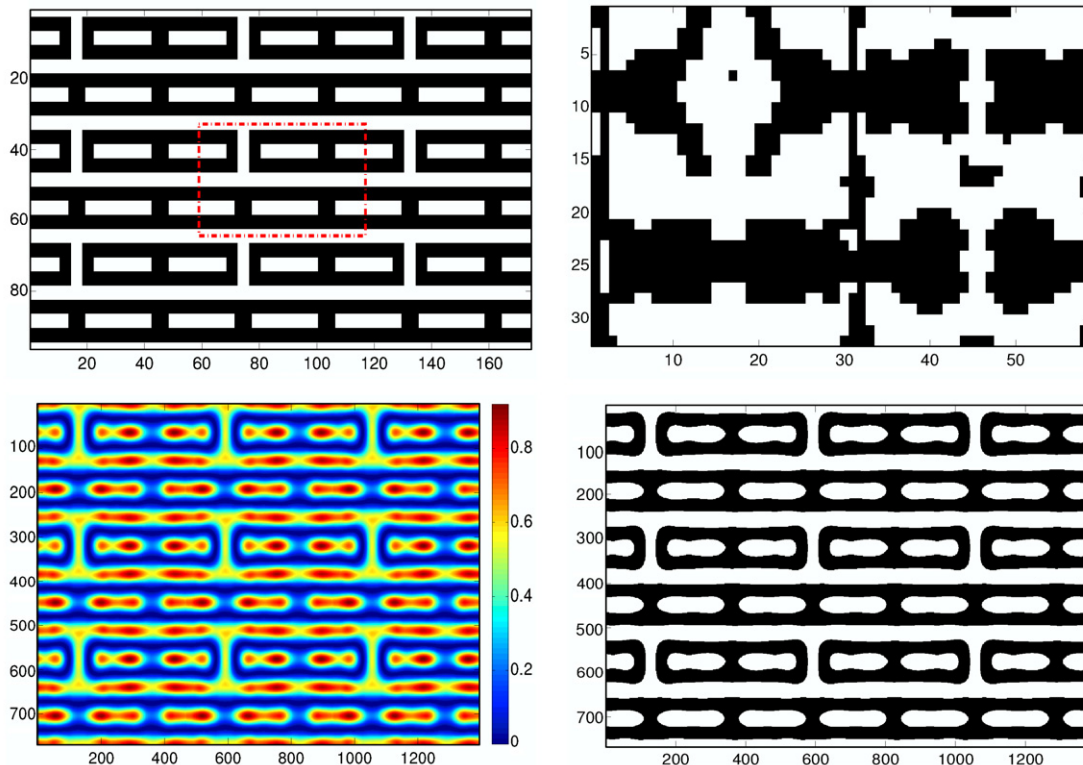


Fig. 17. The top row indicates the periodic target pattern and the estimated 100% transmission PSM. The bottom row indicates the aerial image and its contour at $t_r = 0.3$. Here $\text{NA} = 0.85$, $\sigma = 0$, and $k_1 = 0.35$.

$\gamma_{TV} = 0$, $s = 7$, and number of iterations = 100. The optimization was carried out in a single step for both cases and each had a runtime of 50 s.

The aerial images corresponding to the input masks are illustrated in the bottom row of Fig. 16. We observe that the PSMs bring tremendous improvement in the aerial image contrast thereby making the contacts distinguishable. Note that the upper half of the desired pattern is a contiguous zero region. The 100% transmission PSM deals with such regions by placing the +1 and -1 assist bars in a manner which destructively interfere leading to very little energy deposition. The PSM with chrome blocks the energy deposition by simply using chrome features.

4.4.3. Experiment 3

The final experiment is for a periodic dense pattern consisting of phase conflicts as indicated in Fig. 17. The grating pitch is 160 nm, $NA = 0.85$, $k_1 = 0.35$, and the pattern is sampled using 20 nm resolution. The target pattern consists of T-joints and line-ends, two commonly arising phase-conflict problems. In the past, researchers have proposed using alternating phase shift mask with trim mask (double exposure) [12] or layout modification using graph-cut methods [3] to resolve the above problem. We employ our ILT approach and solve for one period of the pattern (see the dotted red box marked in the target pattern in Fig. 17).⁵ The boundary conditions are simulated using circularly symmetric padding for the convolution operations. The experimental parameters are $\gamma_{fid} = 1$, $\gamma_{dis} = 0.01$, $s = 0.5$, number of iterations = 1500, and the runtime was 50 s. The estimated 100% transmission mask corresponding to one period of the pattern is indicated in Fig. 17. The results indicate that the aerial image has low peak-intensity value in the phase-conflict regions, but the overall contour fidelity (at $t_r = 0.3$) is good. As part of future work, we are working on double exposure inverse lithography technique to further improve the aerial image quality (particularly for low k_1 values with phase conflicts) [32,48].

5. Future work and conclusions

In this article, we proposed a new framework for fast and efficient pixel-based binary, attenuated, and strong phase shift mask design using inverse lithography. Due to public unavailability of the optical kernels used in practice, we have only presented results for the case of coherent and incoherent imaging systems. However, our framework is also applicable to the commonly employed partially coherent imaging systems.

Our work primarily focuses on exploring feasibility and limitation of ILT solutions that provide aerial image contrast necessary for patterning at $k_1 < 0.35$. Practical reduc-

tion of suggested ILT methods to synthesis of actual masks would require computationally efficient treatment of thick mask effects with partially coherent illumination as well as use of robust resist models to account for complex effects accompanying transformation of image intensity into 3D distributions of dissolution rates of actual resists, both of which are outside of the scope of this paper. The above effects, if ignored from the modeling steps, may lead to ILT solutions resulting in catastrophic patterning failures such as bridging and pinching of features of interest, patterning of side-lobes, and unacceptably high sensitivity to mask errors as well as incorrectly assessed sensitivities to changes in focus and exposure. Yet, we believe an improvement in the aerial image modulation (contrast) provided through the ILT path will point to a solution path that leads to an overall improvement in process latitude, resist side-wall angles and give better resist profiles as observed in [43], with possible exception regarding sensitivity to side-lobe printing.

The integer RET constraints were substituted by bound constraints thereby reducing the mask design problem to an unconstrained non-linear optimization problem. We currently employ the steepest descent algorithm to solve the above problem. The convergence behavior can be further improved by exploring advanced optimization techniques like conjugate gradient method, quasi-convex method, etc., which search the solution space more efficiently.

Note that we are employing a local gradient-based method to optimize a non-convex function. We have observed that for low k_1 dense patterns with several *phase-conflicts*, the algorithm sometimes lands in bad local minima. One possible approach to address this problem is to inspect the local minima of a simplified QP problem, and choose a *good minimum* as the initialization guess to the quartic cost function optimization routine (as suggested in [13]). Another alternative is to use double exposure ILT which automatically splits the target pattern into two parts and resolves the phase-conflicts. This forms part of our ongoing research and the preliminary results are very encouraging [32,48].

We introduced the regularization framework and effectively used it to control the tone and complexity of the estimated masks. The regularization framework is a powerful tool which can be further extended to incorporate other user-defined criteria like mask error enhancement factor control, process window optimization, minimum mask feature size control, alignment error control, etc. The proposed technique can be readily extended to employ RETs for optical maskless lithography, proximity correction in e-beam lithography, and diffractive optical elements (DOE) source design for off-axis illumination.

The ILT approach presented in this paper represents a very aggressive RET which will help enable 65 nm and 45 nm nodes using 193 nm exposure tools.

Acknowledgements

The authors acknowledge Dr. Yan Borodovsky from Intel Corporation for his advice on practical aspects of ILT.

⁵ For interpretation of the references in colour, the reader is referred to the web version of this article.

References

- [1] International technology roadmap for semiconductors, <<http://public.itrs.net/>> (2003).
- [2] C. Ahn, H. Kim, K. Baik, *Optical Microlithography*, Proc. SPIE, vol. 3334, 1998, pp. 752–763.
- [3] P. Berman, A.B. Kahng, D. Vidhani, E.H. Wang, F.A. Zelikovsky, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 19 (2000) 175–187.
- [4] J. Bjorkholm, *Intel Technology Journal* Q3 (1998) 8.
- [5] M. Born, E. Wolfe, *Principles of Optics*, Cambridge University Press, 1999.
- [6] P. Choudhury, *Handbook of Microlithography, Micromachining and Microfabrication*, SPIE Press, 1997.
- [7] N. Cobb, A. Zakhor, *Optical Microlithography*, Proc. SPIE, vol. 2440, 1995, pp. 313–327.
- [8] N. Cobb, A. Zakhor, *BACUS Symposium on Photomask Technology*, Proc. SPIE, vol. 2621, 1995, pp. 534–545.
- [9] W. Duch, N. Jankowski, *Computing Surveys* 2 (1999) 163–213.
- [10] A. Erdmann, R. Farkas, T. Fuhner, B. Tollkuhn, G. Kokai, *Optical Microlithography*, Proc. SPIE, vol. 5377, 2004, pp. 646–657.
- [11] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, *IEEE Transaction on Image Processing* 13 (2004) 1327–1344.
- [12] M. Fritze, B. Tyrrell, D. Astolfi, R. Lambert, D. Yost, A. Forte, S. Cann, B. Wheeler, *Lincoln Laboratory Journal* 14 (2003) 237–250.
- [13] Y. Granik, *Optical Microlithography*, Proc. SPIE, vol. 5754, 2005, pp. 506–526.
- [14] W. Huang, C. Lin, C. Kuo, C. Huang, J. Lin, J. Chen, R. Liu, Y. Ku, B. Lin, *Optical Microlithography*, Proc. SPIE, vol. 5377, 2001, pp. 1536–1543.
- [15] C.T. Kelly, *Iterative Methods of Optimization*, SIAM, (1999).
- [16] L. Liebmann, S. Mansfield, A. Wong, M. Lavin, W. Leipold, T. Dunham, *IBM Journal of Research and Development* 45 (2001) 651–665.
- [17] Y. Liu, D. Abrams, L. Pang, Andrew Moore, *BACUS Symposium on Photomask Technology*, Proc. SPIE, vol. 5992, 2005, pp. 231–238.
- [18] Y. Liu, A. Zakhor, *IEEE Transactions on Semiconductor Manufacturing* 5 (1992) 138–151.
- [19] Y. Liu, A. Zakhor, *IEEE Transaction on Semiconductor Manufacturing* 9 (1996) 170–184.
- [20] C. Mack, *Journal of Electrochemical Society* 134 (1987) 148–152.
- [21] C. Mack, *Optical Microlithography*, Proc. SPIE, vol. 5374, 2004, pp. 1–8.
- [22] A. Moore, T. Lin, Y. Liu, G. Russell, L. Pang, D. Abrams, *BACUS Symposium on Photomask Technology*, Proc. SPIE, vol. 6349, 2006.
- [23] K. Nashold, B. Saleh, *Journal of Optical Society of America A – Optics Image Science and Vision* 2 (1985) 635–643.
- [24] Y. Oh, J.C. Lee, S. Lim, *Optical Microlithography*, Proc. SPIE, vol. 3679, 1999, pp. 607–613.
- [25] S. Owa and H. Nagasaka, *Optical Microlithography*, Proc. SPIE, vol. 5040, 2003, pp. 724–733.
- [26] L. Pang, Y. Liu, D. Abrams, in: *Symposium on Photomask and Next Generation Mask Technology VII*, Proc. SPIE, vol. 6283, 2006.
- [27] V.C. Pati, T. Kailath, *Journal of Optical Society of America A – Optics Image Science and Vision* 9 (1994) 2438–2452.
- [28] Y. Pati, A. Teolis, D. Park, R. Bass, K.-W. Rhee, B. Bradie, M.C. Peckerar, *Journal of Vacuum Science and Technology B* 8 (1990) 1882–1888.
- [29] M.C. Peckerar, S. Chang, C.R.K. Marrian, *Journal of Vacuum Science and Technology B* 13 (1995) 2518–2525.
- [30] M.C. Peckerar and C.R.K. Marrian, *Electron-Beam, X-Ray, EUV and Ion-Beam Submicrometer Lithographies for Manufacturing*, Proc. SPIE, vol. 2437, 1995, pp. 222–238.
- [31] J. Peterson, *Optical Microlithography*, Proc. SPIE, vol. 4000, 2000, pp. 77–89.
- [32] A. Poonawala and P. Milanfar, *Journal of Microlithography, Microfabrication and Microsystems* (Preprint available online).
- [33] A. Poonawala, P. Milanfar, *IEEE Transactions on Image Processing* 16 (2007) 774–788.
- [34] A. Poonawala and P. Milanfar, *Optical Microlithography*, Proc. SPIE, vol. 6154, 2006, pp. 114–127.
- [35] J. Randall, K. Ronse, T. Marschner, M. Goethals, M. Ercken, *Optical Microlithography*, Proc. SPIE, vol. 3679, 1999, pp. 176–182.
- [36] S. Robertson, C. Mack, M. Maslow, *Lithography for Semiconductor Manufacturing II*, Proc. SPIE, vol. 4404, 2001, pp. 111–122.
- [37] S. Sayegh, B. Saleh, *IEEE Transaction on Pattern Analysis and Machine Intelligence* 5 (1983) 441–445.
- [38] S. Sayegh, B. Saleh, K. Nashold, *IEEE Transaction on Acoustics, Speech and Signal Processing* 33 (1985) 460–465.
- [39] F. Schellenberg, *Future Fab. Intl.* 9 (2000).
- [40] F. Schellenberg, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, 2004.
- [41] F. Schellenberg, *Resolution enhancement technology: the past, the present, and extensions for the future*, in: Bruce W. Smith(Ed.) *Optical Microlithography*, Proc. SPIE, vol. 5377, 2004, pp. 1–20.
- [42] S. Sherif, B. Saleh, R. Leone, *IEEE Transactions on Image Processing* 4 (1995) 1252–1257.
- [43] S. Shin, G. Han, Y. Ma, K. Moloni, F. Cerrina, *Journal of Vacuum Science and Technology B* 19 (2001) 2890–2895.
- [44] C. Vogel, *Computational Methods for Inverse Problems*, SIAM Press, 2002.
- [45] R.G. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics*, Wiley, 1995.
- [46] A. Wong, *IEEE Micro* 23 (2003) 12–21.
- [47] Alfred Kwok-Kit Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, 2001.
- [48] A. Poonawala, Y. Borodovsky, P. Milanfar, *ILT for Double Exposure Lithography with Conventional and Novel Materials*, *Proceedings of the SPIE Advanced Lithography Symposium*, Feb 2007.