Prewarping Techniques in Imaging: Applications in Nanotechnology and Biotechnology

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

In all imaging systems, the underlying process introduces undesirable distortions that cause the output signal to be a warped version of the input. When the input to such systems can be controlled, pre-warping techniques can be employed which consist of systematically modifying the input such that it cancels out (or compensates for) the process losses. In this paper, we focus on the mask (reticle) design problem for 'optical micro-lithography', a process similar to photographic printing used for transferring binary circuit patterns onto silicon wafers. We use a pixel-based mask representation and model the above process as a cascade of convolution (aerial image formation) and thresholding (high-contrast recording) operations. The pre-distorted mask is obtained by minimizing the norm of the difference between the *desired* output image and the *reproduced* output image. We employ the regularization framework to ensure that the resulting masks are close-to-binary as well as simple and easy to fabricate.

Finally, we provide insight into two additional applications of pre-warping techniques. First is 'e-beam lithography', used for fabricating nano-scale structures, and second is 'electronic visual prosthesis' which aims at providing limited vision to the blind by using a prosthetic retinally implanted chip capable of electrically stimulating the retinal neuron cells.

Keywords: Optical micro-lithography, mask design, OPC, image synthesis, inverse problems, sigmoid, regularization, retinal prosthesis, e-beam lithography

1. INTRODUCTION

In all imaging systems, the underlying process introduces undesirable distortions which causes the output signal to be a warped version of the input. To overcome this, when the input is controllable, pre-warping techniques can be employed which consist of systematically modifying the input such that it will cancel out the process losses. Thus, in effect, we are pre-compensating for the process distortions to come. This is an *image synthesis*[1] problem which consists of finding an image that when used as the input to a given imaging system results in the prescribed desired output image (to within some prescribed tolerance).

The problems of *image restoration* (reconstruction) and *image synthesis* (design) are similar, in both cases the output and imaging system are known but the input is unknown.[1] In the restoration case, the output image results from an actual but unknown input image,[2] and therefore at least one solution must exist in the absence of measurement error or noise. However, the existence of solution is an issue for image synthesis, and there may be no input capable of producing the prescribed output.[3]

The image formation process can be mathematically expressed as,

$$z(x,y) = T\{i(x,y)\}, \tag{1}$$

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. We seek to find a pre-distorted input intensity

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

function which will give us 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 distance metric.

$$\hat{i}(x,y) = \arg\min_{i(x,y)} d(z^*(x,y), T\{i(x,y)\}). \tag{2}$$

In this article, we focus on studying prewarping techniques to design optical and process correction (OPC) masks for microlithography. Section 2 discusses the optical lithography procedure and introduces the forward model employed to simulate the distortion effects of the imaging system. The optimization procedure is discussed in Section 3 and in Section 4 we introduce two regularization functions to obtain near-binary and easy to manufacture masks, respectively. Finally, we discuss the extensions and applications of prewarping techniques to nanolithography and visual prosthesis along with some conclusive remarks in Section 5.

2. OPTICAL MICROLITHOGRAPHY AND OPC

2.1. Optical Microlithography

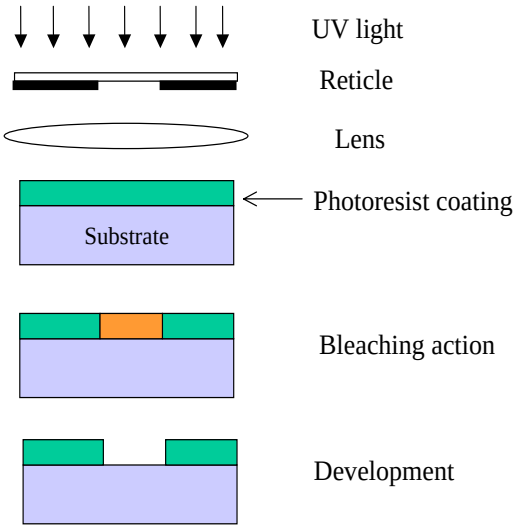

Figure 1. Steps involved in optical microlithography

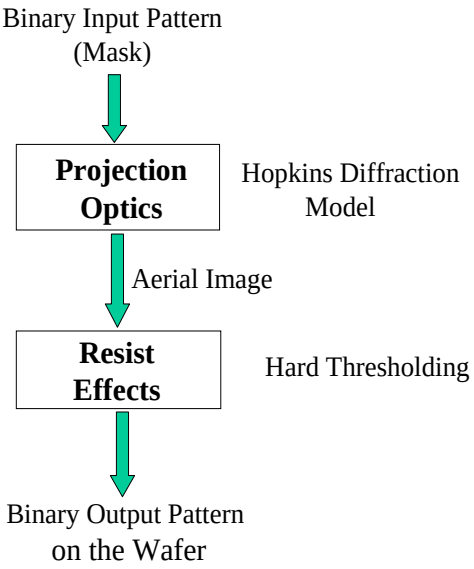

Figure 2. A simplified view of the lithography process

Optical microlithography, a process similar to photographic printing, is used for transferring circuit patterns onto silicon wafers and forms a very critical step in the IC fabrication flowchart.[4] As illustrated in Fig. 1, the pattern to be replicated on the wafer is first carved on a reticle (mask). An illuminator (UV source) is shone through this mask producing an image of the pattern through the lens system, which is eventually projected down onto a photoresist-coated silicon wafer using a projection system (typical aspect ratio 4:1). The photoresist is chemically sensitive to light, and hence only the exposed regions are washed away leaving behind a replica of the mask pattern on the substrate (wafer).

In an ideal world, the output circuit pattern on the wafer will be exactly similar to the mask pattern. Unfortunately, the band-limited imaging system introduces distortions caused by the diffraction effects of the lens-projection system, particularly at higher resolutions. These undesirable distortions lead to a loss of pattern fidelity and hence the output pattern obtained on the wafer is a distorted (blurred) version of the input.[5] The semiconductor industry strives to adhere to the ITRS[6] (International Technology Roadmap for Semiconductors) road-map which is driven by Moore's law for the past three decades. Moore's law with regard to lithography requires the critical dimensions (CD) to shrink by thirty percent every two years (currently CD = 65nm). This

puts very stringent requirements on lithography thereby making it one of the tightest bottlenecks faced by the semiconductor industry.

The resolution(R) of the lithography system in Fig. 1 can be expressed using Rayleigh's criterion as follows,

$$R = \frac{k\lambda}{NA}. \tag{3}$$

Obviously the resolution can be improved by increasing the numerical aperture of the imaging system (NA) or decreasing the wavelength (λ) and these alternatives have been actively explored by researchers in the lithography community (current values are $NA = 0.9$ and $\lambda = 193nm$). Pushing the above parameters values beyond this limit is very costly, risky and time consuming. Therefore, it has been established that the most important, feasible, and promising method to achieve the goal of enhanced resolution is by decreasing the process constant k using resolution enhancement techniques (RETs).[7,8]

RETs are based on exploring three properties of the wavefront, namely, its amplitude, phase and direction and are accordingly classified as optical and process correction, phase-shift method and off-axis illumination.[9] Phase-shift method and optical process correction are implemented by making physical changes to the mask (reticle) and the mask pattern. Unfortunately, these changes increase the complexity, storage requirements and cost of the masks (currently in millions) and hence the challenge is to implement RETs constrained by the above factors.

2.2. Optical and Process Correction

In this paper, we focus on designing OPC masks to overcome the distortions occurring in optical microlithography. OPC consists of adding sub-resolution features to the original layout, a process also known as micromanipulation. Thus we pre-compensate for the process losses by modifying the original layout, which eventually leads to a better pattern fidelity and improved resolution (see Fig. 5).[10]

OPC has been carried out using two approaches; namely rule-based, or model-based. As the name suggests, in the rule-based OPC scheme, empirical rules are developed to counteract the commonly occurring problems around corners, edges, local interactions, etc. These are then applied throughout the pattern to provide a general improvement in pattern fidelity. Rule-based OPC is very simple to implement but it only compensates for the local features and does not optimize the global performance depending on the overall layout. Hence it has limited scope, effectiveness and applications.[9]

Model-based methods on the other hand use a mathematical description to represent the warping process (forward model). As such, they are more universal and represent a more aggressive OPC strategy. The success of these methods relies heavily on accurate modelling of the distortion process which has been studied carefully by physicists, chemists, and other researchers in the lithography community.[11] Fig. 2 illustrates a simplified view of the lithography process. It consists of two functional blocks; namely the projection optics effects (aerial image formation), and the resist effects. The former is simulated using Hopkins scalar (or vector) model for partially coherent imaging systems[12] and the later is simulated using Dills model[4] or in the simplest case using a constant threshold model. Model-based OPC itself has two flavors. The forward model-based OPC techniques suggested by Cobb and Zakhor[13] parameterize the pattern using edges and corners and proceed by nudging these geometric elements while simulating the output pattern (using the forward model) until certain criteria are satisfied. Backward model based techniques invert the mathematical model and attempt to directly synthesize the optimized pattern.[9] Sherif, et al.[1] and Liu-Zakhor[14] employed branch and bound algorithms to design OPC and phase-shift masks following the backward model-based technique.

2.3. Proposed Method for Mask Synthesis

In this paper, we treat the problem as an "inverse" problem, and focus on developing fast and efficient methods for backward (or inverse) model-based OPC. We use a pixel-based mask representation, hence the input, output and desired patterns are all represented using discrete 2-D images. We employ the approximated forward process model illustrated in Fig. 3 in our analysis. Note that the aerial image formation step from Fig. 2 is approximated using the convolution of the input pattern with a 2-D Gaussian kernel for the sake of simplicity. Secondly, we

employ the sigmoid type transfer function instead of the hard-thresholding (heaviside) operation to simulate the resist effect. A heaviside step function is defined as,

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

Using a Heaviside operator (hard threshold) would result in a discrete combinatorial optimization problem. A sigmoid is a smooth, continuous function which closely approximates the Heaviside function.[15] With this choice we can use gradient-based continuous function optimization techniques like steepest-descent to solve the mask design problem.

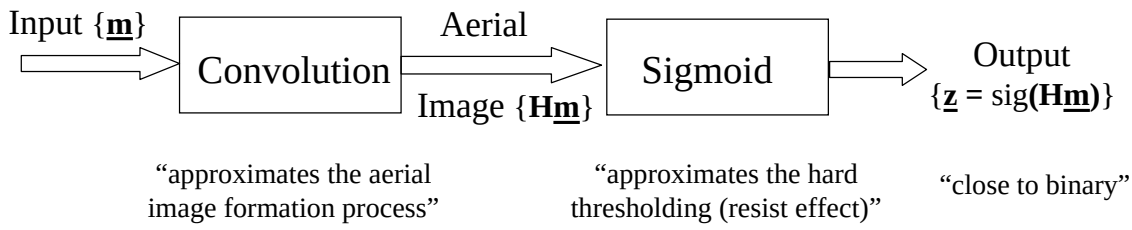

Figure 3. Approximated forward process model

In particular, 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. Fig. 4 illustrates the behavior of a sigmoid with $a = 80$ and $t_r = 0.5$.

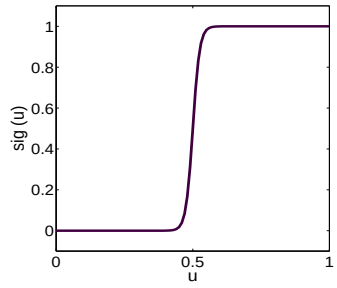

Figure 4. The sigmoid function $\text{sig}(u) = 1/(1 + e^{-80(u-0.5)})$

3. MODEL-BASED OPTIMIZATION

We assume that $\mathbf{M}$, $\mathbf{Z}$, and $\mathbf{Z}^*$ are image arrays of size $N \times N$ such that $[\mathbf{M}]_{ij} = m(x_i, y_j)$, $[\mathbf{Z}]_{ij} = z(x_i, y_j)$, and $[\mathbf{Z}^*]_{ij} = z^*(x_i, y_j)$. Throughout our discussion, $\mathbf{Z}^*$ represents the prescribed *binary* pattern, $\mathbf{Z}$ represents the *gray-level* output pattern, and $\mathbf{M}$ represents the input pattern fed to the imaging system (can be *binary* or *gray-level*). Let $\underline{\mathbf{z}}$, $\underline{\mathbf{m}}$, $\underline{\mathbf{z}}^* \in \Re^{N^2 \times 1}$ represent the raster scanned (lexicographic ordering) vectors of the respective matrices. The forward model in Fig. 3 can be mathematically represented as,

$$\underline{\mathbf{z}} = \text{sig}(\mathbf{H}\underline{\mathbf{m}}), \tag{6}$$

where $\mathbf{H} \in \Re^{N^2 \times N^2}$ is the blur matrix representing the point spread function (PSF) of the imaging system, $\underline{\mathbf{z}} = [\underline{z}_1, \ldots, \underline{z}_{N^2}]^T$, and $\underline{\mathbf{m}} = [\underline{m}_1, \ldots, \underline{m}_{N^2}]^T$. Thus, every pixel undergoes a cascade of convolution followed by the sigmoidal transformation

$$\underline{z}_i = \frac{1}{1 + \exp\left[-a\left(\sum_{j=1}^{N^2} h_{ij}\underline{m}_j\right) + at_r\right]}\ , \tag{7}$$

for $i = 1, \ldots, N^2$.

3.1. Optimization Problem

We formulate the OPC-mask design problem as finding the optimized layout $\underline{\hat{\mathbf{m}}}$ 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 $\underline{\mathbf{z}}$,

$$\underline{\hat{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}}\{F(\underline{\mathbf{m}})\} = \arg\min_{\underline{\mathbf{m}}} ||\underline{\mathbf{z}}^* - \underline{\mathbf{z}}||_2^2 = \arg\min_{\underline{\mathbf{m}}} ||\underline{\mathbf{z}}^* - \mathrm{sig}(\mathbf{H}\underline{\mathbf{m}})||_2^2 \tag{8}$$

Therefore,

$$\underline{\hat{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}} \sum_{i=1}^{N^2} \left(\underline{z}_i^* - \underline{z}_i\right)^2 \tag{9}$$

where $\underline{z}_i$ is defined using (7) *.

Note that $\underline{\mathbf{m}}$ consists of the transmission values of a binary mask which can only take values 0 or 1, resulting in a combinatorial optimization problem. However, to make the problem analytically tractable, we relax the parameter values to lie in the *range* [0,1]. This is achieved by imposing the following inequality constraints on the optimization problem given in (8),

$$0 \leq \underline{m}_j \leq 1 \quad \text{for } j = 1, \ldots, N^2\ . \tag{10}$$

The bound-constrained optimization problem can be reduced to an unconstrained optimization problem using the following parametric transformation,

$$\underline{m}_j = \frac{1 + \cos(\underline{\theta}_j)}{2} \quad \text{for } j = 1, \ldots, N^2 \tag{11}$$

where $\underline{\boldsymbol{\theta}} = [\underline{\theta}_1, \ldots, \underline{\theta}_{N^2}]^T$ is the unconstrained parameter vector. The reparameterized cost function can be formulated in terms of the parameter vector $\underline{\boldsymbol{\theta}}$ as follows,

$$F_1(\underline{\boldsymbol{\theta}}) = \sum_{i=1}^{N^2} \left(\underline{z}_i^* - \frac{1}{1 + \exp\left[-a\left(\sum_{j=1}^{N^2} h_{ij}\frac{1 + \cos(\underline{\theta}_j)}{2}\right) + at_r\right]} \right)^2 . \tag{12}$$

We now employ the steepest-descent method to find the optimum solution of the above problem which involves finding the first order derivatives of (12). The gradient vector $\nabla F_1(\underline{\boldsymbol{\theta}}) \in \Re^{N^2 \times 1}$ can be calculated using the following expression,

$$\nabla F_1(\underline{\boldsymbol{\theta}}) = \underline{\mathbf{d}} = a(\mathbf{H^T}\ [(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}) \odot \underline{\mathbf{z}} \odot (\mathbf{1} - \underline{\mathbf{z}})]\ ) \odot\ \sin(\underline{\boldsymbol{\theta}}), \tag{13}$$

*As an alternative we have also considered the weighted least squares where weights are chosen according to the importance of different parts of the pattern. This may be viewed as a mix of model-based and rule-based approaches, where the rules determine weights in the model

where $\odot$ is the element-by-element multiplication operator. The k^{th} iteration of steepest descent is given as,

$$\underline{\boldsymbol{\theta}}^{k+1} = \underline{\boldsymbol{\theta}}^{k} - s\underline{\mathbf{d}}^{k} \tag{14}$$

where s is the step-size. We would like to highlight the useful fact that due to the structure of (13), the steepest descent iterations can be quickly and directly carried out on the 2-D matrices with no need for the raster scanning operation. More importantly, since we know the solution to be a perturbation of the prescribed pattern, the iteration can be initialized with the prescribed pattern, leading to quick convergence to a global solution. The optimized pattern $\widehat{\mathbf{M}}$ can finally be obtained from $\boldsymbol{\Theta}$ using (11).

3.2. Binarization of the gray-mask

The pattern $\widehat{\mathbf{M}}$ obtained using the above method is not binary. Instead each pixel can have gray values anywhere in [0,1]. This makes the resulting mask physically unrealizable and hence we need a post-processing step to obtain the synthesized *binary* OPC mask $\widehat{\mathbf{M}}_b$. The simplest way to obtain $\widehat{\mathbf{M}}_b$ from $\widehat{\mathbf{M}}$ is using a global threshold parameter t_m, such that the error between $\mathbf{Z}^*$ and the output binary pattern obtained using $\widehat{\mathbf{M}}_b$ (as the input) is minimized. The optimum value of t_m can be obtained using a simple line search operation.[16] However, the above approach for binarization is sub-optimal and in Section 3 we propose an alternative method for this step.

3.3. Results

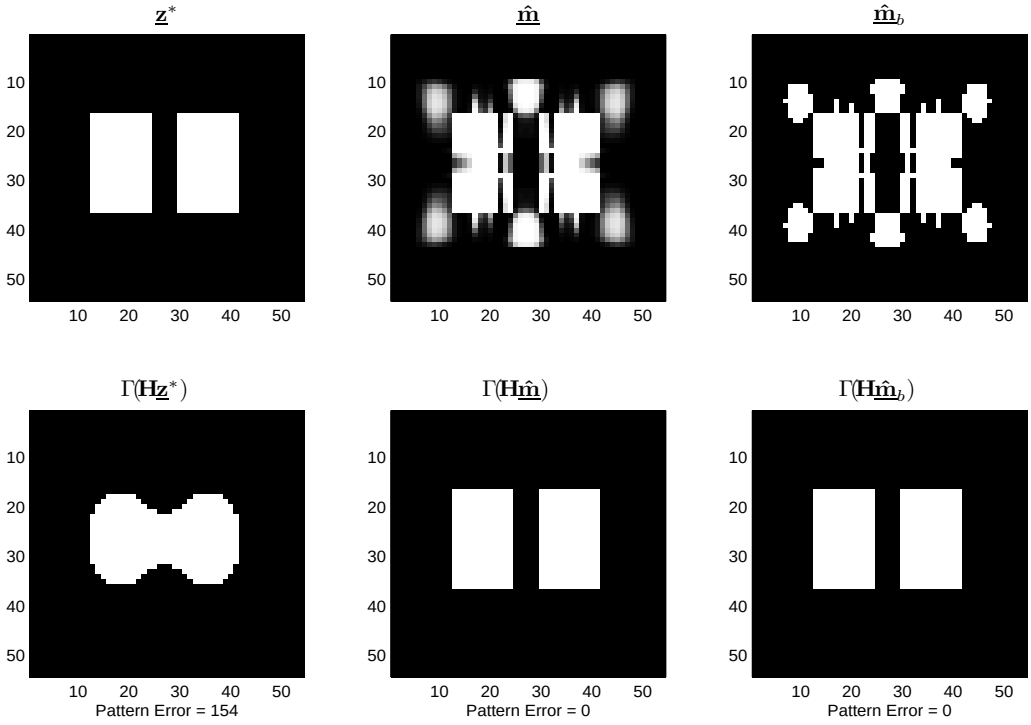

Figure 5. Top row, left to right: original pattern, optimized gray pattern and binary pattern obtained using an optimum global threshold t_m. The bottom row indicates the corresponding binary output patterns. The parameters are $a = 90$, $t_r = 0.5$, 15×15 filter with $\sigma = 5$, and $s = 0.4$

We now present two experimental results for the OPC masks synthesized using the aforementioned method. The bottom row in Fig. 5 illustrates the binary output patterns obtained using the original pattern (i.e. $\mathbf{M} = \mathbf{Z}^*$), the optimized gray pattern ($\mathbf{M} = \widehat{\mathbf{M}}$) and the binary pattern ($\mathbf{M} = \widehat{\mathbf{M}}_b$) as inputs. The PSF effect was simulated

using a Gaussian filter of size 15×15 with $\sigma = 5$ pixels, with the sigmoid parameters $a = 90$, $t_r = 0.5$ and the step size $s = 0.4$. The final output binary patterns were obtained by thresholding the aerial images using the heaviside operator given in (4) with $t_r = 0.5$.

The error $F(\underline{\mathbf{m}})$ associated with the above three input patterns can be calculated using (8); $F(\underline{\mathbf{z}}^*) = 142.76$, $F(\underline{\hat{\mathbf{m}}}) = 4.09$, and $F(\underline{\hat{\mathbf{m}}}_b) = 8.46$. However for the given imaging system (see Fig. 2), the aerial image is actually subjected to the heaviside operation to obtain the binary output image. Hence, it is more appropriate to compute the error using the binary output image instead of the sigmoidal transformed (gray level) output image. We employ a new metric *pattern error*, defined as the total number of pixels which are not faithfully reproduced in the binary output pattern,

$$\text{error}(\underline{\mathbf{m}}) = ||\underline{\mathbf{z}}^* - \Gamma(\mathbf{H}\underline{\mathbf{m}})||_1. \tag{15}$$

From Fig. 5, we observe that $\text{error}(\underline{\mathbf{m}}) = 154$, $\text{error}(\underline{\hat{\mathbf{m}}}) = 0$, and $\text{error}(\underline{\hat{\mathbf{m}}}_b) = 0$.

Note that the imaging system was incapable of rendering the two bars distinguishable if the desired image is itself used as the input (i.e. $\mathbf{M} = \mathbf{Z}^*$). However, the patterns are reproduced very faithfully using the synthesized OPC mask. Fig. 6 illustrates the cost function behavior for the first 200 iterations of steepest descent, indicating quick convergence.

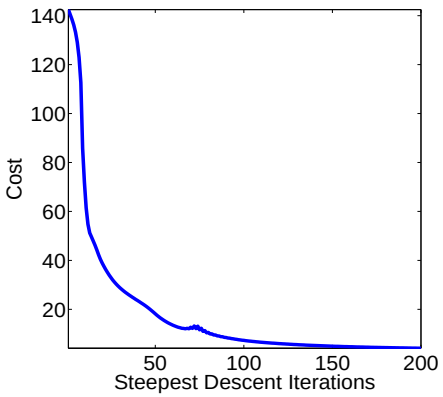

Figure 6. Cost function versus steepest-descent iterations for Fig. 5

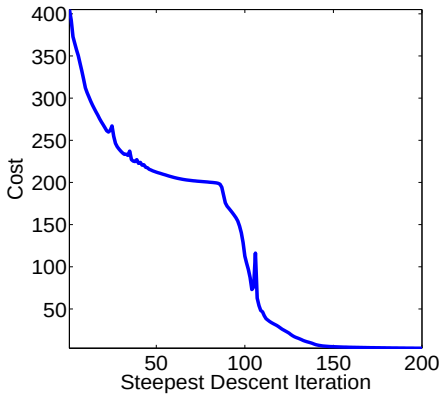

Figure 7. Cost function versus steepest-descent iterations for Fig. 8.

Fig. 8 illustrates the input and output patterns for a more complicated circuit pattern. The Gaussian filter employed has size 11×11 with $\sigma = 14$ pixels (the high value of σ gives the effect of an averaging filter), while the sigmoid parameters are $a = 80$, $t_r = 0.5$, and $s = 0.5$. We observe that our algorithm performs an excellent job in prewarping and the output is extremely faithful to the desired pattern (*pattern error* 447 vs 1 pixels). Note that the binarization step employed from Section 3.2 increases the error from 1 to 18. Fig. 7 illustrates the cost function behavior for the first 200 iterations of the steepest-descent procedure. Note that the long vertical bar (on the extreme left) in the prescribed pattern in Fig. 8 is completely missing when reproduced using $\mathbf{M} = \mathbf{Z}^*$. However, our optimization procedure starts adding prewarping elements to correct it at around 85^{th} iteration, which results in a steep decrease in the cost function observed in Fig. 7. The kinks in the curve can be eliminated by choosing a smaller step size.

The imaging system given in Fig. 2 is a binary-in-binary-out (BIBO) system. The binary output part is approximated by using the sigmoid based transfer function in the process model (see Fig. 3) which guarantees that the output pattern is always close to binary. However, the current setup does not incorporate the fact that the input pattern should also be binary. The disadvantage is the need for an extra post-processing step (binarization) which is sub-optimal with no guarantee that the *pattern error* will be under control (see Fig. 8). There is also a possibility to jointly optimize the binarizing threshold and the gray pattern. In the next section, we use the regularization framework to overcome this problem and complete the binary-to-binary loop.

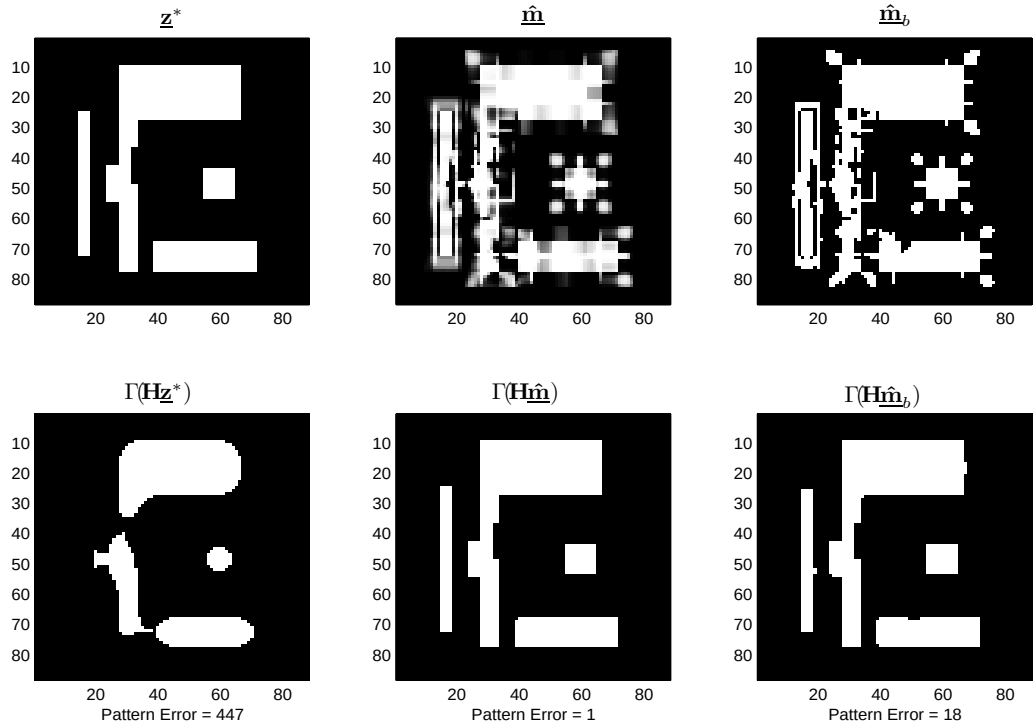

Figure 8. Top row, left to right: original pattern, optimized gray pattern and binary pattern obtained using an optimum global threshold t_m. The bottom row indicates the corresponding binary output patterns. The parameters are $a = 80$, $t_r = 0.5$, 11×11 filter with $\sigma = 14$, and $s = 0.5$

4. REGULARIZATION

If we look back again at Fig. 5, we observe that both the gray-level and binary input patterns give rise to the same (desired) binary pattern at the output. The BIBO mask design problem can itself have multiple solutions and the continuous domain formulation implies that now there can be infinitely many different input patterns, all giving rise to the same binary pattern at the output. However, we want our estimated pattern to satisfy certain properties which can be incorporated as prior knowledge about the solution. We can constrain the space of solutions to obtain a general desirable solution using an appropriate regularization term.[17] In general, the formulation can be described as follows,

$$\underline{\hat{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}} [F(\underline{\mathbf{m}}) + \gamma \; R(\underline{\mathbf{m}})] \tag{16}$$

where $F(\underline{\mathbf{m}})$ is the data-fidelity term, $R(\underline{\mathbf{m}})$ is the regularization function used to direct the unknown parameter $\underline{\mathbf{m}}$ towards the desired solution space, and γ is the user-defined scalar for adequately weighing the first(fidelity) term against the second(regularization) term.

In this section we discuss two regularization terms corresponding to two desirable properties we chose to promote in our estimated solution. The first property discussed earlier is that our estimated mask should be (close to) binary, the second is that our masks should be simple, and therefore cheap and easy to manufacture.

4.1. Quadratic Penalty Term

The first penalty term employed to obtain near binary patterns is as follows,

$$R_{\text{bin}}(\underline{\mathbf{m}}) = \sum_{j=1}^{N^2} \left[1 - (2\underline{m}_j - 1)^2\right] = 4\underline{\mathbf{m}}^T(\underline{\mathbf{1}} - \underline{\mathbf{m}}), \tag{17}$$

where $\underline{\mathbf{1}} = [1, \ldots, 1]^T \in \Re^{N^2}$. Thus every pixel $\underline{m}_j$ has an associated penalty given by the quadratic function (see Fig. 9),

$$r(\underline{m}_j) = 1 - (2\underline{m}_j - 1)^2.$$

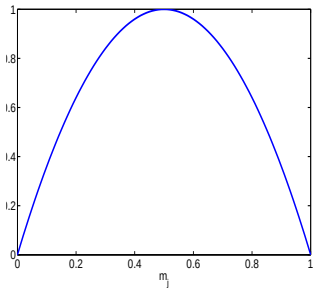

Figure 9. Quadratic penalty cost function

The mask transmission values are 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 $m_j = 0.5$).

The gradient $\nabla R_{\text{bin}}(\underline{\mathbf{m}}) \in \Re^{N^2 \times 1}$ of the quadratic penalty term is given by,

$$\nabla R_{\text{bin}}(\underline{\mathbf{m}}) = (-8\underline{\mathbf{m}} + 4), \tag{18}$$

which can be used in conjunction with (13) and (14) while carrying out the steepest-descent iterations as before.

Fig. 10 illustrates the results with the same prescribed pattern, PSF and sigmoidal parameters used in Fig. 8 (Section 3.3). However, here we employ the quadratic regularization term and minimize the overall cost function

$$J(\underline{\mathbf{m}}) = F(\underline{\mathbf{m}}) + \gamma_{\text{bin}} R_{\text{bin}}(\underline{\mathbf{m}}),$$

where $F(\underline{\mathbf{m}})$ is defined in (8) and $R_{\text{bin}}(\underline{\mathbf{m}})$ is defined in (17), with $\gamma_{\text{bin}} = 0.015$ and $s = 1.5$. Note that unlike Fig. 8, the optimized gray-pattern is very close to binary. Hence, there is no need for the line search operation discussed in Section 3.2 to find the optimal t_m. In fact, we can simply obtain the synthesized binary pattern $\widehat{\mathbf{M}}_b$ by thresholding $\widehat{\mathbf{M}}$ with $t_m = 0.5$. Note that the *pattern error* only increases from 0 to 1 pixel due to the above step. We also note that the inclusion of the regularization penalty that promotes binary results did not introduce any new error when applied to the pattern in Fig. 5, and therefore, due to space constraints, we do not show the results for that case.

4.2. Total Variation Penalty Term

A second penalty term we will incorporate is to ensure that the resulting OPC mask is less complex and therefore cheap and easy to fabricate, and inspect. 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 a new pattern called *the flip pattern* $\underline{\mathbf{f}}$ where,

$$\underline{f}_j = |\underline{m}_j - \underline{z}_j^*| \quad \text{for} \quad j = 1, \ldots, N^2.$$

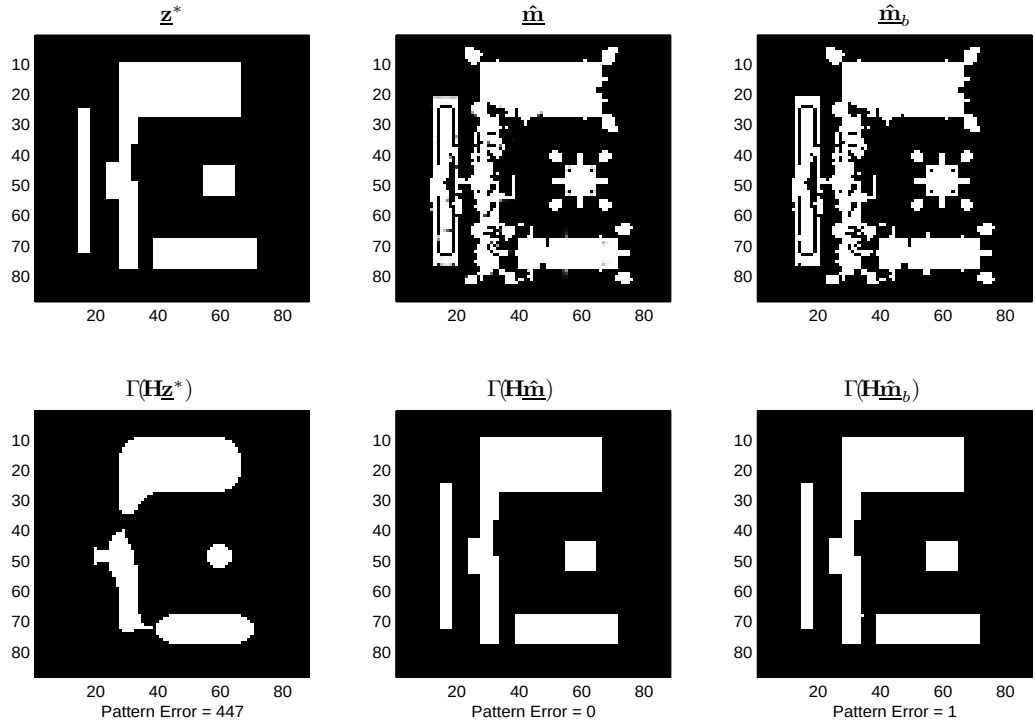

Figure 10. Left to right: original pattern, optimized gray pattern and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}}) + \gamma_{\text{bin}} R_{\text{bin}}(\underline{\mathbf{m}})$ and the parameters are $a = 80,\ t_r = 0.5$, 11×11 filter with $\sigma = 14$, $s = 1.5$, and $\gamma_{\text{bin}} = 0.015$

The *on* pixels in $\underline{\mathbf{f}}$ indicate the positions where pre-warping occurred; 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*. Our experience indicates that the Total Variation (TV)[17] penalty term is an appropriate choice, leading to desirable results, hence that is what we use for now. However, we are also exploring other possibilities. TV penalty term is defined as the L_1 norm of the magnitude of the gradient of the *flip pattern* $\underline{\mathbf{f}}$,

$$R_{\text{TV}}(\underline{\mathbf{m}}) = ||\nabla \underline{\mathbf{f}}||_1^1. \tag{19}$$

The *flip pattern* $\underline{\mathbf{f}}$ 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 (19) suppresses these effects and forces the *changes* to be spatially smoother and less abrupt. This leads to simple, and easy to manufacture OPC masks.

The TV equation in (19) can also be approximated using the following equation,[18]

$$||\nabla \underline{\mathbf{f}}||_1^1 = ||\mathbf{Q_x}\underline{\mathbf{f}}||_1^1 + ||\mathbf{Q_y}\underline{\mathbf{f}}||_1^1, \tag{20}$$

where $\mathbf{Q_x}, \mathbf{Q_y} \in \Re^{N^2}$ 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 $\mathbf{M}$ along horizontal (right) and vertical (up) direction by one pixel respectively. The gradient $\nabla R_{\text{TV}}(\underline{\mathbf{m}}) \in \Re^{N^2 \times 1}$ of the TV penalty term is given as,

$$\nabla R_{\text{TV}}(\underline{\mathbf{m}}) = [\mathbf{Q_x}^T \text{sign}(\mathbf{Q_x}\underline{\mathbf{f}}) + \mathbf{Q_y}^T \text{sign}(\mathbf{Q_y}\underline{\mathbf{f}})] \ \odot \ \text{sign}(\underline{\mathbf{m}} - \underline{\mathbf{z}}^*), \tag{21}$$

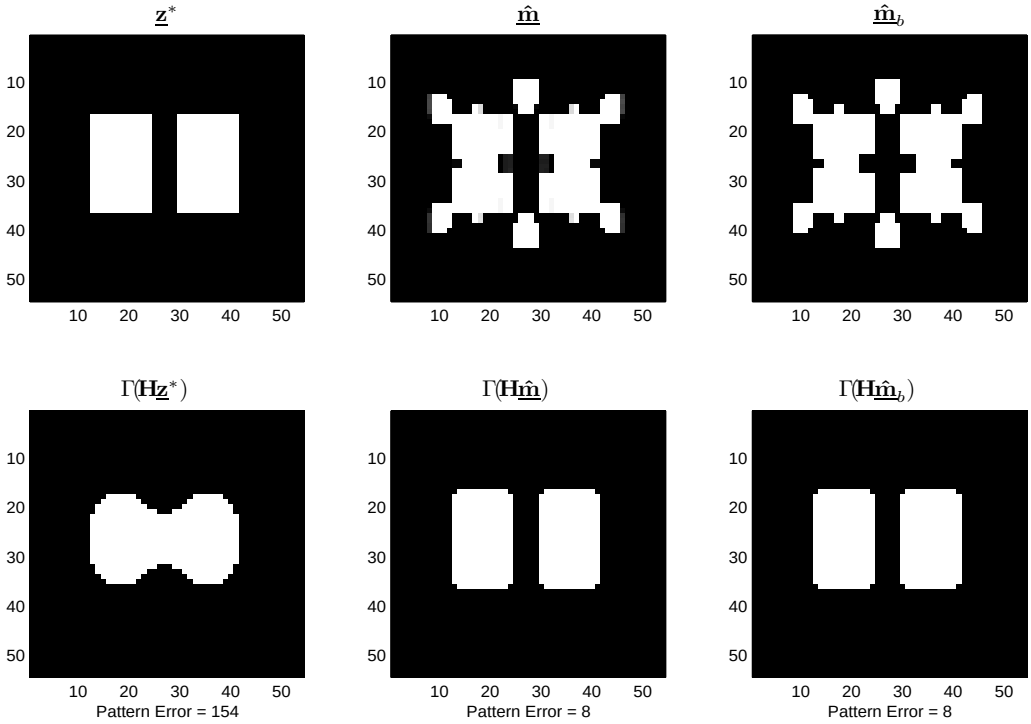

Figure 11. Left to right: original pattern, optimized gray pattern and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}}) + \gamma_{\text{bin}} R_{\text{bin}}(\underline{\mathbf{m}}) + \gamma_{\text{TV}} R_{\text{TV}}(\underline{\mathbf{m}})$ and the parameters are $a = 90$, $t_r = 0.5$, 15×15 filter with $\sigma = 5$, $s = 1$, $\gamma_{\text{bin}} = 0.025$ and $\gamma_{\text{TV}} = 0.045$

which can be used in conjunction with (13) and (14) while carrying out the steepest-descent iterations as before.

Fig. 11 illustrates the results with the same prescribed pattern, PSF, and sigmoidal parameters used in Fig. 5 (Section 3.3). However, here we employ both the quadratic and TV regularization terms and minimize the overall cost function

$$J(\underline{\mathbf{m}}) = F(\underline{\mathbf{m}}) + \gamma_{\text{bin}} R_{\text{bin}}(\underline{\mathbf{m}}) + \gamma_{\text{TV}} R_{\text{TV}}(\underline{\mathbf{m}}),$$

where $F(\underline{\mathbf{m}})$ is defined in (8), $R_{\text{bin}}(\underline{\mathbf{m}})$ is defined in (17), and $R_{\text{TV}}(\underline{\mathbf{m}})$ is defined in (19) with $\gamma_{\text{bin}} = 0.025$, $\gamma_{\text{TV}} = 0.045$, and $s = 1$. We observe that the synthesized OPC pattern is much simpler than the one in Fig. 5.

Similarly, Fig. 12 illustrates the results obtained using both regularization terms with the same prescribed pattern, PSF, and sigmoidal parameters used in Fig. 10 (Section 4.1) with $\gamma_{\text{bin}} = 0.01$, $\gamma_{\text{TV}} = 0.02$, and $s = 1$. In both the above cases, we observe that the isolated holes, protrusions, and jagged edges are suppressed and hence the resulting mask is faster, easier, and cheaper to fabricate and inspect.

5. EXTENSIONS AND APPLICATIONS OF PRE-WARPING TECHNIQUES

In this section we suggest prewarping techniques to solve related problems arising in microlithography, nanolithography, and the development of retinal visual prostheses. We also provide some conclusive remarks towards the end of this section.

5.1. Extensions in Microlithography

In this paper, we discussed OPC mask synthesis problem for microlithography. However, we use a very simplified view of the forward model and future work will comprise of incorporating more sophisticated models. An

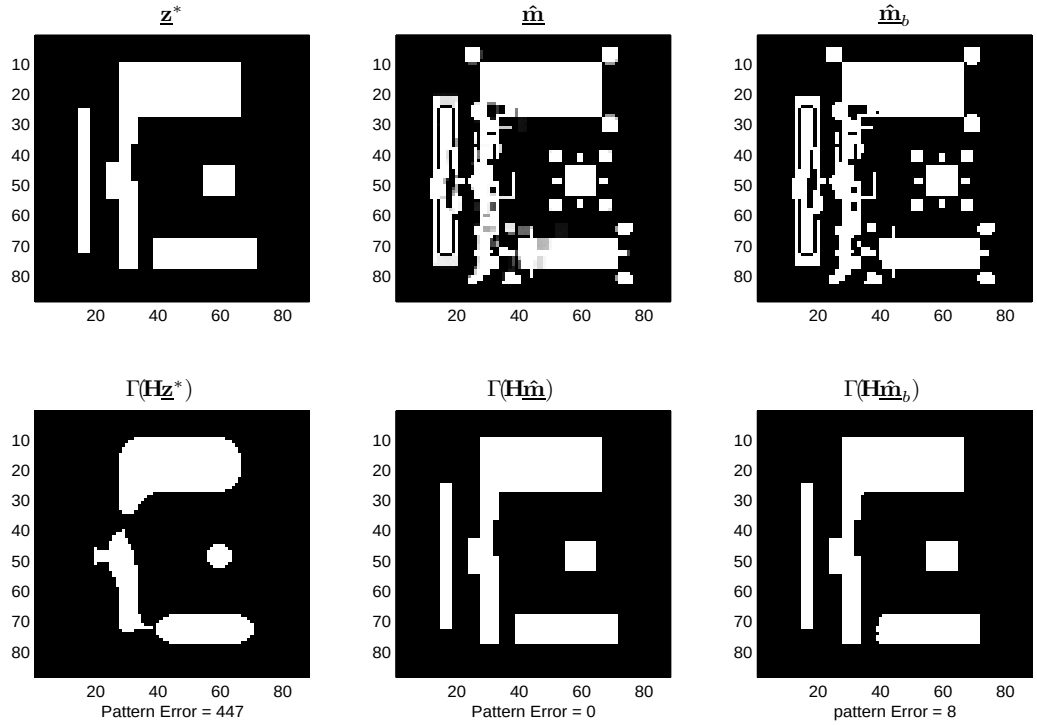

Figure 12. Left to right: original pattern, optimized gray pattern and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}}) + \gamma_{\text{bin}} R_{\text{bin}}(\underline{\mathbf{m}}) + \gamma_{\text{TV}} R_{\text{TV}}(\underline{\mathbf{m}})$ and the parameters are $a = 80$, $t_r = 0.5$, 11×11 filter with $\sigma = 14$, $s = 1$, $\gamma_{\text{bin}} = 0.01$ and $\gamma_{\text{TV}} = 0.02$

immediate extension will be to use the sum-of-coherent-system approximation to Hopkins partial coherence model for simulating the aerial image formation.[12] The resist effects are also dependent on the time of the exposure and the dissolution rate which need to be taken into account for higher accuracy. We can also employ the variable threshold resist model[19] where our goal now is to jointly estimate the prewarping along with the best threshold parameter. Suitable weight functions can be used to ensure that critical regions are faithfully reproduced. The optimization process can be sped up using constrained pre-warping which limits the pre-distortion to only the user-defined regions. A very promising future direction to pursue is synthesizing phase shift masks where the estimated pixels now have values -1, 0 or 1.[14] The above framework will enable us to combine phase shift mask and OPC mask design, a powerful combination of RETs.[7] We can also enlist the features and feature types totally undesirable in our optimized mask and propose new regularization terms specifically tailored to combat those effects.

5.2. Applications in Nanotechnology

The fabrication of devices at nanometer scale is performed using electron-beam lithography, which consists of directly firing a focused beam of electrons onto a photoresist coated substrate. Unlike microlithography, a mask is not used. Instead, the pattern is directly carved on the substrate using the electron beam. Unfortunately, once the electrons enter the resist, they tend to scatter before depositing their energy, giving rise to proximity effects. That is, the electron scattering causes the energy to spread throughout the substrate which results in a blurring effect. The point spread function of the above system is commonly modelled as sum of two Gaussian kernels accounting for the forward and backscattering of the electrons. The above effect is counteracted using electron-beam proximity correction, a predistortion technique commonly adopted by the nanolithography community.[20]

The e-beam lithography process can be modelled similarly to microlithography, as shown in Fig. 3. The

pattern is carved one pixel at a time by the e-beam writer, and hence pixel-based pattern representation is most suited for this purpose. E-beam proximity effect correction techniques fall into two sub-categories of dose modification and shape modification. Unlike before, here the dose (or energy) of the electrons can also be modified and hence the estimated pattern is not restricted to be binary anymore. Each pixel of $\mathbf{M}$ now represents the dose given to the corresponding location. The allowable dose values may be continuous within a finite range, or fixed to some discrete levels, depending on the properties (switching time, power, resolution, etc) of the e-beam writer. We can adopt our proposed algorithms to fit the above framework and solve for dose precompensation or shape precompensation or the powerful combination of both.

It is worth noting that the reticles used for microlithography are themselves fabricated using e-beam lithography. Therefore, one needs to run one more round of prewarping techniques to solve for the distortions introduced by the e-beam lithography process. Instead, we can cascade the e-beam lithography and optical lithography process to form a lumped process (distortion) model. We can then directly solve for the optimum e-beam lithography pattern required to obtain the desired pattern $\mathbf{Z}^*$ on the silicon wafer. The actual reticle is just treated as an intermediate by-product. The penalty on reticle costs, storage, complexity, time, etc are actually incurred at the mask fabrication level and hence this approach gives us a more direct control on the above factors.

5.3. Application of Prewarping to Retinal Visual Prosthesis

Retinal Pigmentosa and Age-Related Macular Degeneration are incurable retinal diseases which result in profound loss of vision due to non-functional photoreceptor cells. Retinal prosthesis[21] aims at providing limited vision capability using an implanted electrode array which directly stimulates the retinal ganglion cells thereby giving a limited sense of visual perception. An externally mounted camera captures the real world data and transmits an encoded and severely downsampled stream to this electrode array via a wireless link. Recently, a 4x4 electrode array has been successfully implanted[22] and the eventual goal is to have a 32x32 implant which will hopefully enable the tasks of face and character recognition. Furthermore, the system only permits 8 to 16 excitation levels and color cannot be perceived.

Unfortunately, the final image perceived by the patient will be a distorted or blurred version of the input owing to the proximity of the densely packed electrode array, superficial ganglion cell axons, etc. To overcome this problem, we have considered pre-distortion techniques by systematically modifying the input fed to the electrode array such that the *actual perceived image* is closer to the captured real world image. A prime challenge is that it is very hard to generalize and characterize the above imaging system, hence the model should be able to incorporate user inputs and dynamically tune itself to cause the best prewarping.

5.4. Conclusion

We studied the distortions arising in optical microlithography (a binary-to-binary imaging system) and successfully studied predistortion techniques to counteract them. The binary output part was modelled using the log-sigmoid transfer function and the synthesized input pattern was forced to be binary using a quadratic penalty regularization term. We demonstrated that the complexity of the resulting patterns can be curbed by employing suitable penalty terms. Finally, we shed some light on application of prewarping techniques to new avenues in nanotechnology and biotechnology.

ACKNOWLEDGMENTS

This work was supported in part by Intel Corporation.

REFERENCES

1. 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.
2. R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2002.
3. S. Sayegh and A. Saleh, "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.

4. P. Choudhury, *Handbook of Microlithography, Micromachining and Microfabrication*, SPIE PRESS Monograph, 1997.
5. F. Schellenberg, "A little light magic," *IEEE Spectrum* **40**, pp. 34–39, 2003.
6. "International technology roadmap for semiconductors," *http://public.itrs.net/* , 2003.
7. 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.
8. F. Schellenberg, *Resolution Enhancement Techniques in Optical Lithography*, SPIE, 2004.
9. A. K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE, 2001.
10. F. Schellenberg, "Resolution enhancement with OPC/PSM," *Future Fab Intl* , 2001.
11. A. Erdmann and W. Henke, "Simulation of optical lithography," in *Optical Microlithography*, *Proc. SPIE* **3729**, pp. 480–494, 1999.
12. 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.
13. N. Cobb and A. Zakhor, "Fast, low-complexty mask design," in *Optical Microlithography*, *Proc. SPIE* **2440**, pp. 313–327, 1995.
14. Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," *IEEE Transactions on Semiconductor Manufacturing* **5**, pp. 138–151, 1992.
15. W. Duch and N. Jankowski, "Survey of neural transfer functions," *Computing Surveys* **2**, pp. 163–213, 1999.
16. M. Bazaraa and C. Shetty, *NonLinear Programming Theory and Algorithms*, John Wiley and Sons, 1979.
17. C. Vogel, *Computational Methods for Inverse Problems*, SIAM, 2002.
18. Y. Li and F. Santosa, "A computational algorithm for minimzing total variation in image restoration," *IEEE Transactions on Image Processing* **5**, pp. 987–995, 1996.
19. 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.
20. B. Cook, *PYRAMID - A Hierarchical Rule-based Proximity Effect Correction Scheme for Electron-beam Lithography: Generalization and Optimization for Homogeneouse Substrates*, PhD Dissertation, Cornell University, 1996.
21. W. Liu and M. Humayun, "Retinal prosthesis," *IEEE International Solid-State Circuits Conference* **1**, pp. 218 – 219, 2004.
22. M. Humayun et al, "Visual perception in a blind subject with a chronic microelectronic retinal prosthesis," *Vision Research* **43**, pp. 2573–2581, 2003.