Scanning Laser Lithography with Constrained Quadratic Exposure Optimization

Andrew J. Fleming, *Member, IEEE,* Omid T. Ghalehbeygi, Ben S. Routley, and Adrian G. Wills, *Member, IEEE.*

***Abstract*—Scanning laser lithography is a maskless lithography method for selectively exposing features on a film of photoresist. A set of exposure positions and beam energies are required to optimally reproduce the desired feature pattern. The task of determining the exposure energies is inherently non-linear due to the photoresist model and the requirement for only positive energy. In this article, a nonlinear programming approach is employed to find an optimal exposure profile that minimizes the feature error and total exposure energy. This method is demonstrated experimentally to create a features with sub-wavelength geometry.**

I. Introduction

Lithography is the process of selectively exposing optically sensitive materials during semiconductor fabrication [1]–[3]. One issue with standard processes is the high cost of infrastructure and mask sets. In order to bypass the cost of mask production, a number of maskless lithography processes have been developed [4]. Machines for scanning Electron Beam Lithography (EBL) are already commercially available [5].

In addition to electron-beam [5] and ion-beam lithography [6], maskless optical lithography is also developing. In its simplest form, a laser beam is focused to a spot size of approximately 500 nm and scanned over the surface [7]. To improve the throughput, zone-plate arrays create a large array of focused spots [8], [9]. Photon sieve lithography [10], [11] is a similar method for achieving the same result.

Rather than focusing light with diffractive or refractive optics, it can also be directed through a sharpened optical fiber or probe [12]–[14]. Below one wavelength from the tip, the emitted light forms an evanescent field with highly localized intensity. Enhanced apertures have achieved a resolution of 22 nm [15], [16]. By employing an array of such probes, the throughput can be significantly increased [17]–[19]. Maskless thermal-probe [20] and mechanical-probe lithography [21] have also been demonstrated [22].

A number of challenges exist with probe-based and maskless optical lithography. Firstly, the throughput is extremely low compared to mask based methods. However, advances in nanopositioning systems have allowed scan rates to exceed 1000 Hz, which may allow thousands or millions of features to be written per second [23]–[25]. Another difficulty is the problem of finding an exposure pattern which optimizes the fidelity of developed features. In scanning laser lithography, this equates to finding a set of exposure locations and laser pulse energies [26]. The exposure optimization problem is associated with many scanning and probe-based lithography systems despite the differences in physical processes. For example, in electron beam lithography, the exposure variable is electron dosage rather than optical energy [27]–[29]. Other examples where a 2D exposure profile is required include near-field lithography [13], [14], zone-plate array lithography [8], [9], and photon sieve lithography [10], [11]. Methods that don't require exposure optimization include thermal-probe [20] and mechanical-probe [21] based methods.

In electron beam lithography and scanning laser lithography the dosage energy can be precisely controlled for each location, which is a property unique to scanning beam methods. The point spread function (spatial distribution) of the beam can be measured [30] and corrected for. Early methods used rules [31], [32] which were similar to proximity correction methods in mask-based lithography. Later methods employed linear programming to determine optimal exposure patterns [29]. This approach resulted in the release of commercially available software for proximity correction [33]–[35]; however, the details are proprietary.

The problem of exposure optimization also exists in mask based lithography [36], [37]. However, in this case, the optimization variables are the mask and source pattern [38], [39]. Resolution enhancement methods were primarily rule-based before 2000, then either model-based or hybrid thereafter [38], [39]. Between 2000 to 2010, the primary advance was the simultaneous optimization of the source pattern, or pupil, and the exposure mask [40], [41]. Pixelated source patterns provide the greatest flexibility and have received the most attention [42]–[45].

Since 2013, source and mask optimization methods have also been applied to immersion lithography [45], [46] and EUV systems [47]. Due to the increasing density of integrated circuits, the numerical efficiency has become a major issue. Computational improvements have been achieved with basis functions [48] and augmented Lagrangian methods [49].

In regards to the optimization method used for mask-based resolution enhancement, early work included simulated annealing [50], mixed integer programming [51], random pixel flipping [52], and genetic algorithms [53]. However, from 2000, the majority of methods aimed to solve the inverse lithography problem, for example, by solving a non-linear program [54] or by a level set method [55]. Since 2007, gradient-based methods have been preferred due to their

This work was supported by the Australian Research Council Grants DP150103521, FT130100543, and DP120100487.

A. J. Fleming, O. T. Ghalehbeygi, and B. S. Routley are with the School of Electrical Engineering and Computing, University of Newcastle, Callaghan, NSW 2308, Australia (e-mails: andrew.fleming@newcastle.edu.au, omid.tayefeh@gmail.com, ben.s.routley@gmail.com).

A. G. Wills is with the School of Engineering, University of Newcastle, Callaghan, NSW 2308, Australia (e-mail: Adrian.Wills@newcastle.edu.au).

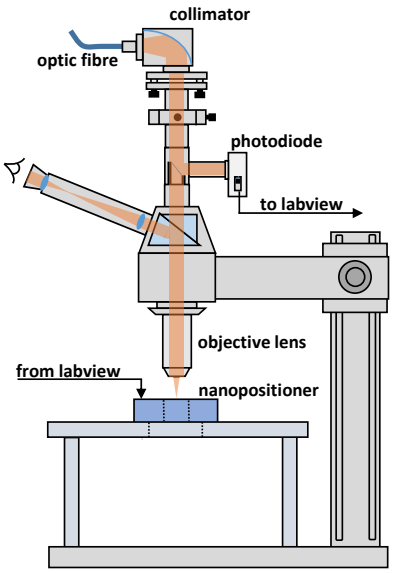

Fig. 1. The optical exposure system which focuses light from the fiber onto the scanning platform.

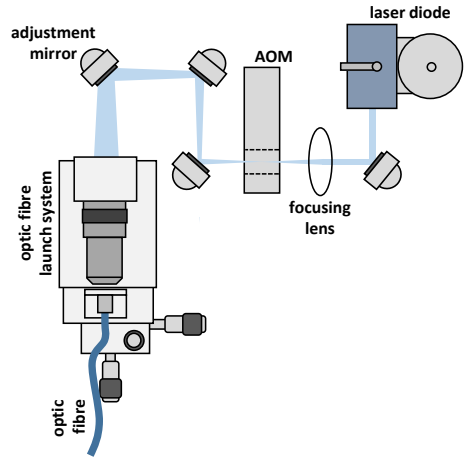

Fig. 2. The laser source, modulation system, and fiber coupling.

favorable convergence properties and computational cost [42], [56]–[61].

A. Contributions

This article describes a new method for optimizing the exposure profile in maskless scanning laser lithography. This method is also directly applicable to electron beam lithography. Unlike mask-based optimization methods, the energy at each exposure point is freely variable. which improves the flexibility, but may also increase complexity.

The majority of optimization methods for scanning beam lithography are aimed at achieving a target dosage. However, in this work, the target is the developed feature. This approach requires the inclusion of the photoresist model which transforms the optimization from a quadratic program to a constrained non-linear optimization. This problem is solved sequentially by computing an analytical gradient and employing a log-barrier method. An analytical Hessian approximation is also derived to minimize the number of required iterations.

In previous work, the process was successfully simulated [26]; however, the experimental work in this article required the definition of a new cost function. In reference [26], the squared sum of exposure energy was penalized. Conversely, in this article, the sum of energy and the squared sum of dosage are penalized. This cost function more closely represents the desired physical outcomes, which include minimizing scatter and over-exposure.

The process model, optimization method, and experimental results are described in the following sections.

II. Experimental Setup and Process Flow

As illustrated in Figure 1 the exposure optics are based on a trinocular microscope modified so that the primary beam path is infinity corrected. 405-nm laser light is introduced via a single mode optical fiber and off-axis parabolic reflector which results in a Gaussian TEM_{00} beam of sufficient width to fill the back aperture of the Nikon 40x/0.75 objective lens. The focused beam is then directed at the sample which is positioned by an N-point LC402 nanopositioner. The beam is also directed to a photodiode by a 50:50 beam splitter. The measured power is used in a feedback system to precisely control the dosage. As shown in Figure 2, the laser source is modulated by an acoustic optical modulator (AOM) which provides power control and shuttering.

The glass substrates are initially washed in methanol and acetone to remove debris. A Laurell WS-400A spin-coater is then used to deposit AZ ECI3007 photoresist onto the substrate. As per the manufacture's specifications, the speed was 4000 rpm for one minute which resulted in a film thickness of approximately 700 nm. After the coating step, the photoresist was baked at 90 °C for one minute to improve the substrate adhesion and minimize dark erosion during development. After the exposure process, the sample is immersed in AZ-726MIF developer for one minute removing the exposed pattern. Finally the sample is rinsed in distilled water and dried using nitrogen gas.

III. Process Modeling

This section develops a model of the lithography process described in Section II. The model assumes that the photoresist layer is thin and that the beam profile remains constant throughout its depth. The optical properties of the film, which are a function of the exposure state, are also assumed to be constant. Other optical effects such as scattering and cavity formation are also ignored.

A. Beam Profile

In the experimental setup, a single-mode fiber is utilized to create an ideal Gaussian beam profile. The light intensity (in W/m^2) at the focal point of the objective lens can be analytically expressed as:

$$B(x,y) = \alpha e^{-\beta(x^2+y^2)} \ , \tag{1}$$

$$\text{where } \alpha = \frac{2P}{\pi w_0^2} \text{ and } \beta = \frac{2}{w_0^2} \ ,$$

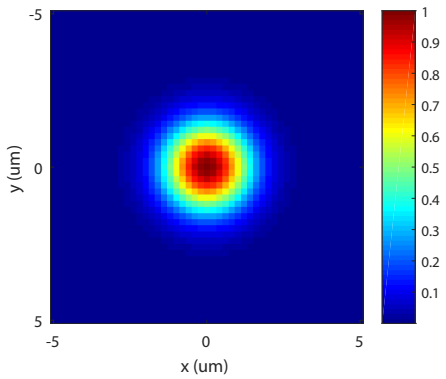

Fig. 3. The normalized beam power for an exposure at $x = 0$ and $y = 0$. The beam width is $w_0 = 410$ nm.

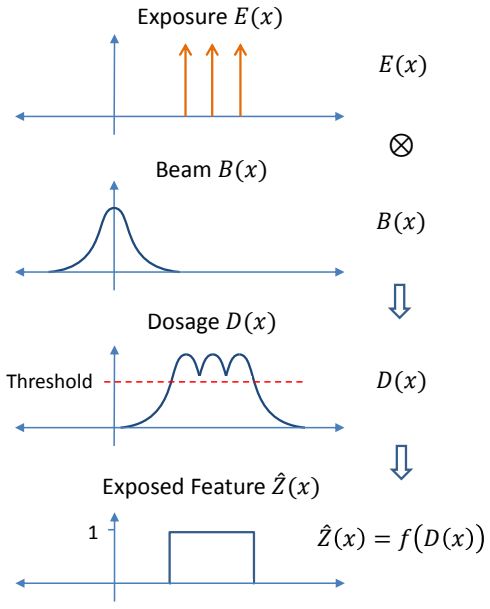

Fig. 4. A simplified one-dimensional model of scanning laser lithography. In this example, the exposure pattern $E(x)$ is three discrete exposures of equal energy. The resulting dosage $D(x)$ is the sum of each exposure point convolved with the beam profile $B(x)$. Finally, the photoresist function $f(\sigma)$ maps the cumulative dosage $D(x)$ to the predicted feature $\widehat{Z}(x)$.

where x and y indicate the transverse axes of the beam at focal point w_0, and P is the total power in the beam. An example of this function is plotted in Figure 3.

B. Continuous Exposure Modeling

A one-dimensional model of the exposure process along the x-axis (i.e. $y = 0$) is illustrated in Figure 4. The exposure profile $E(x)$ represents the energy delivered at a position x. In this work, the exposure energy is modulated by controlling the time interval for which the laser shutter is open. Since the beam power is constant, the time interval is proportional to the resulting dosage. Other possibilities include modulating the beam power or the scanning speed.

The light intensity (in W/m^2) is a Guassian function described in Equation 1. To calculate the dosage $D(x)$ (in J/m^2) at a single point, the intensity is multiplied by the exposure time, that is $D(x) = t_{on}B(x, 0)$. Where multiple exposures $\boldsymbol{t}_i$ are involved at arbitary locations $\boldsymbol{x}_i$, the total dosage is

$$D(x) = \sum_{i=1}^{N} \boldsymbol{t}_i B(x - \boldsymbol{x}_i, 0). \tag{2}$$

The above equation is a convolution operation which can be generalized to discrete or continuous exposures in one or more dimensions. That is, in general

$$D(x, y) = E(x, y) \otimes B(x, y). \tag{3}$$

where $\otimes$ is the convolution operator. When the exposure function is discrete, the dosage can be expressed as

$$D(x, y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \boldsymbol{E}_{i,j}\ B\left(x - \boldsymbol{x}_i, y - \boldsymbol{y}_j\right)\ , \tag{4}$$

where the i, j element of the matrix $\boldsymbol{E} \in \mathbb{R}^{N_x \times N_y}$ represents the exposure energy at a location $x = \boldsymbol{x}_i$ and $y = \boldsymbol{y}_j$, where $\boldsymbol{x} \in \mathbb{R}^{N_x}$ and $\boldsymbol{y} \in \mathbb{R}^{N_y}$.

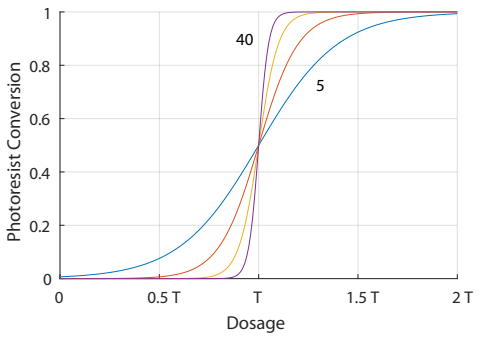

Fig. 5. Sigmoid threshold function with varying steepness parameter $\gamma = 5, 10, 20, 40$.

C. Photoresist Development Model

The photoresist model quantifies the chemical composition of the photoresist based on the dosage energy received. The simplest model is a threshold function which indicates 100% conversion when the dosage is above a threshold. For example,

$$\widehat{Z}(x, y) = \begin{cases} 1 & D(x, y) \geq T \\ 0 & D(x, y) < T \end{cases} \tag{5}$$

where $\widehat{Z}(x, y)$ is the fraction of converted photoresist and T is the threshold energy.

A more realistic model is a sigmoid function which relates the dosage energy to the fraction of converted photoresist:

$$\widehat{Z}(x, y) = f\left(D(x, y)\right) = \frac{1}{1 + e^{-\gamma(D(x,y) - T)}}, \tag{6}$$

where $\widehat{Z}(x, y)$ is the fraction of converted photoresist, T is the threshold energy, and the parameter γ dictates the steepness of the sigmoid. When this parameter is large, the function resembles a binary exposure model. In Figure 5, the sigmoid function is plotted for several values of the parameter γ.

D. Discrete Exposure Modelling

To facilitate optimization, the functions for exposure, beam profile and dosage will be replaced by matrices which represent these functions at discrete locations in a workspace. The workspace is discretized into N locations along the x and y axes,

$$\boldsymbol{x} = \boldsymbol{y} = [0, \Delta, 2\Delta, \cdots, (N-1)\Delta)] \ , \tag{7}$$

where Δ is the resolution.

Using this approach, the exposure matrix $\boldsymbol{E} \in \mathbb{R}^{N\times N}$ is the exposure energy at each grid location. That is, the element $\boldsymbol{E}_{i,j}$ represents the exposure energy at location $(\boldsymbol{x}_i, \boldsymbol{y}_j)$, where $\boldsymbol{E}_{i,j}$ refers to the i^{th} row and j^{th} column of $\boldsymbol{E}$. Similar matrices will be used for the dosage $\boldsymbol{D}$ and predicted feature $\widehat{\boldsymbol{Z}}$.

The beam profile matrix $\boldsymbol{B}^{k,l} \in \mathbb{R}^{N\times N}$ is the beam power over the workspace for a focal point located at $\boldsymbol{x}_k, \boldsymbol{y}_l$. That is, the array of beam profile matrices are

$$\boldsymbol{B}^{k,l}_{i,j} = \alpha e^{-\beta(\boldsymbol{x}_i - \boldsymbol{x}_k)^2 - \beta(\boldsymbol{y}_j - \boldsymbol{y}_l)^2} \tag{8}$$

$$i, j = 1, \cdots, N \quad \text{and} \quad k, l = 1, \cdots, N \tag{9}$$

Where $\boldsymbol{B}^{k,l} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{B}$ is an $N \times N$ array of matrices. Using this definition for $\boldsymbol{B}$, the dosage matrix $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ can be written

$$\boldsymbol{D} = \sum_{k=1}^{N}\sum_{l=1}^{N} \boldsymbol{E}_{k,l} \ \boldsymbol{B}^{k,l} \ , \tag{10}$$

where an individual element is

$$\boldsymbol{D}_{i,j} = \sum_{k=1}^{N}\sum_{l=1}^{N} \boldsymbol{E}_{k,l} \ \boldsymbol{B}^{k,l}_{i,j} \ . \tag{11}$$

For compatibility with standard optimization methods, it is convenient to vectorize the matrices by stacking the rows. That is, We define the "vec" operator,

$$\text{vec}\left\{\boldsymbol{E}\right\} \triangleq \begin{bmatrix} \boldsymbol{E}_{:,1} \\ \boldsymbol{E}_{:,2} \\ \vdots \\ \boldsymbol{E}_{:,N} \end{bmatrix} , \tag{12}$$

where Matlab notation is used and $\boldsymbol{E}_{:,k}$ refers to column k of the matrix $\boldsymbol{E}$. As a vector, the exposure matrix becomes

$$\boldsymbol{e} \triangleq \text{vec}\left\{\boldsymbol{E}\right\} \ . \tag{13}$$

The dosage matrix can also be vectorized $\boldsymbol{d} \triangleq \text{vec}\left\{\boldsymbol{D}\right\}$ so that Equation (10) can be rewritten as the multiplication

$$\boldsymbol{d} = \boldsymbol{\Omega} \ \boldsymbol{e} \ , \tag{14}$$

where the columns of $\boldsymbol{\Omega}$ are the vectorized versions of $\boldsymbol{B}^{k,l}$, that is

$$\boldsymbol{\Omega} = \left[\text{vec}\left\{\boldsymbol{B}^{1,1}\right\}, \cdots, \text{vec}\left\{\boldsymbol{B}^{N,1}\right\}, \text{vec}\left\{\boldsymbol{B}^{1,2}\right\}, \cdots, \text{vec}\left\{\boldsymbol{B}^{N,N}\right\}\right] . \tag{15}$$

In this form $\boldsymbol{\Omega} \in \mathbb{R}^{N^2\times N^2}$ and $\boldsymbol{d}, \boldsymbol{e} \in \mathbb{R}^{N^2}$.

The vectorized predicted feature $\widehat{\boldsymbol{z}}$ can be estimated by applying the thresholding function (6) element wise to $\boldsymbol{d}$.

$$\widehat{\boldsymbol{z}}_i = f(\boldsymbol{d}_i) \ . \tag{16}$$

Finally, the original form of the matrices $\boldsymbol{E}$, $\boldsymbol{D}$, and $\widehat{\boldsymbol{Z}}$ can be reconstructed by reshaping the vectors $\boldsymbol{e}$, $\boldsymbol{d}$, and $\widehat{\boldsymbol{z}}$ respectively.

IV. Optimization Approach

The aim of the optimization is to compute an exposure matrix $\boldsymbol{E}$ which minimizes the difference between the desired feature $\boldsymbol{Z}$ and the predicted feature $\widehat{\boldsymbol{Z}}$. That is, the goal is to minimize

$$V_1(\boldsymbol{e}) \triangleq \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\boldsymbol{Z}_{i,j} - \widehat{\boldsymbol{Z}}_{i,j})^2 = \frac{1}{N^2} \ \boldsymbol{r}^T\boldsymbol{r} \ , \tag{17}$$

where the residual $\boldsymbol{r} = \boldsymbol{z} - \widehat{\boldsymbol{z}}$, and $\boldsymbol{z} \triangleq \text{vec}\left\{\boldsymbol{Z}\right\}$. It is also desirable to minimize the total exposure energy,

$$V_2(\boldsymbol{e}) \triangleq \frac{1}{N^2}\sum_{i=1}^{N^2} \boldsymbol{e}_i = \frac{1}{N^2} \ \boldsymbol{J}_{1,N^2} \ \boldsymbol{e} \ , \tag{18}$$

where $\boldsymbol{J}_{1,M}$ is a unitary row vector of length M. It is also desirable to control the total dosage

$$V_3(\boldsymbol{e}) \triangleq \frac{1}{N^2}\sum_{i=1}^{N^2} \boldsymbol{d}_i^2 = \frac{1}{N^2} \ \boldsymbol{d}^T\boldsymbol{d} \ . \tag{19}$$

These three cost components can be combined in a weighted manner with a scalar weighting $\lambda_2 \geq 0$ and $\lambda_3 > 0$ to define the overall cost as

$$V(\boldsymbol{e}) \triangleq V_1(\boldsymbol{e}) + \lambda_2 V_2(\boldsymbol{e}) + \lambda_3 V_3(\boldsymbol{e}). \tag{20}$$

The exposure values $\boldsymbol{e}$ must be individually non-negative since the dosage can only be positive. Therefore, the optimal exposure pattern may be expressed via the following problem.

$$\boldsymbol{e}^* = \arg\min_{\boldsymbol{e}} V(\boldsymbol{e}), \quad \text{s.t.} \quad \boldsymbol{e}_k \geq 0, \ \forall k, \tag{21}$$

A. Problem Solution

The optimisation problem expressed in (21) is a nonlinear, and importantly non-convex, programming problem. In the absence of the thresholding function $f(\cdot)$, the problem reduces to a quadratic program (QP) with simple positivity bound constraints. However, the sigmoid thresholding function, while smooth, is neither convex nor concave and renders the problem more difficult to solve. Similar problems are encountered in model predictive control [62].

Nevertheless, problem (21) will be solved in this paper by employing a barrier function approach where the inequality constraints are replaced with a weighted logarithmic barrier function [63]. More specifically, the barrier problem is defined as

$$\boldsymbol{e}(\mu) \triangleq \arg\min_{\boldsymbol{e}} V_\mu(\boldsymbol{e}), \tag{22}$$

$$V_\mu(\boldsymbol{e}) \triangleq V(\boldsymbol{e}) - \frac{\mu}{N^2}\sum_{k=1}^{N^2}\log(\boldsymbol{e}_k)$$

The above problem is well defined on the interior of the constraint set where $\boldsymbol{e}_k > 0$ for all k. The barrier method approach solves a sequence of problems in the form of (22) where the the barrier function weighting is gradually reduced toward zero and it can be shown that (see e.g. [63])

$$\lim_{\mu\to 0} \boldsymbol{e}(\mu) = \boldsymbol{e}^* \tag{23}$$

The main attraction of this approach is that (22) is directly amenable to Newton's method since $V_\mu(\boldsymbol{e})$ has smooth first and secord order derivatives on the interior of the constraint set. The algorithm is summarized below.

Algorithm 1 Solve (21) using the barrier method

Require: $\epsilon_\mu > 0$, $\rho > 0$, $\kappa > 0$, $\mu > 0$ and $\boldsymbol{e}_k > 0,\ \forall k$
  Compute the gradient vector $\boldsymbol{g} \triangleq \nabla_{\boldsymbol{e}} V_\mu(\boldsymbol{e})$
  Compute a positive definite scaling matrix $\boldsymbol{H}$
  Compute the weighted gradient norm $\delta = (\boldsymbol{g}^T \boldsymbol{H}^{-1} \boldsymbol{g})^{1/2}$
  while $\mu > \epsilon_\mu$ or $\delta > \rho$ **do**
    Compute the search direction $\boldsymbol{p} \triangleq -\boldsymbol{H}^{-1}\boldsymbol{g}$
    Compute step length $\eta \in (0, 1]$ such that

$$V_\mu(\boldsymbol{e} + \eta\boldsymbol{p}) < V_\mu(\boldsymbol{e}), \qquad (\boldsymbol{e} + \eta\boldsymbol{p})_k > 0, \forall k \tag{24}$$

    Update $\boldsymbol{e} \leftarrow \boldsymbol{e} + \eta\boldsymbol{p}$
    Update the gradient $\boldsymbol{g} \leftarrow \nabla_{\boldsymbol{e}} V_\mu(\boldsymbol{e})$
    Update the scaling matrix $\boldsymbol{H}$
    Update the weighted gradient norm $\delta \leftarrow (\boldsymbol{g}^T \boldsymbol{H}^{-1} \boldsymbol{g})^{1/2}$
    if $\delta \leq \rho$ **then**
      $\mu \leftarrow \kappa\mu$
    end if
  end while

It remains to explain how to compute the gradient vector $\boldsymbol{g}$, the positive definite scaling matrix $\boldsymbol{H}$, and to define suitable values for ϵ_μ, ρ and κ that are all used within the algorithm. These will be outlined in the sub-sections IV-B and IV-C. Section IV-D provides some comments on a suitable stopping criteria.

B. Gradient Calculation

The gradient vector $\boldsymbol{g}$ is defined as

$$\begin{aligned}
\boldsymbol{g} &\triangleq \nabla_{\boldsymbol{e}} V_\mu(\boldsymbol{e}), \\
&= \nabla_{\boldsymbol{e}} \left[V_1(\boldsymbol{e}) + \lambda_2 V_2(\boldsymbol{e}) + \lambda_3 V_3(\boldsymbol{e}) - \frac{\mu}{N^2} \sum_{k=1}^{N^2} \log(\boldsymbol{e}_k) \right] && (25) \\
&= \frac{2}{N^2}\Phi^T \boldsymbol{r} \ + \ \frac{2\lambda_2}{N^2}\boldsymbol{I} \ + \ \frac{2\lambda_3}{N^2}\boldsymbol{\Omega}^T\boldsymbol{d} \ - \ \frac{\mu}{N^2} \begin{bmatrix} \frac{1}{\boldsymbol{e}_1} \\ \vdots \\ \frac{1}{\boldsymbol{e}_{N^2}} \end{bmatrix} && (26)
\end{aligned}$$

where Φ is the Jacobian matrix

$$\begin{aligned}
\Phi &\triangleq \nabla_{\boldsymbol{e}} \boldsymbol{r}(\boldsymbol{e}), \\
&= \nabla_{\boldsymbol{e}} (\boldsymbol{z} - \hat{\boldsymbol{z}}(\boldsymbol{e})) = -\nabla_{\boldsymbol{e}} \text{vec}\left\{ f(\boldsymbol{D}(\boldsymbol{e})) \right\} && (27) \\
&= -\nabla_{\boldsymbol{e}} f(\text{vec}\left\{ \boldsymbol{D}(\boldsymbol{e}) \right\}) = -\nabla_{\boldsymbol{d}(\boldsymbol{e})} f(\boldsymbol{d}(\boldsymbol{e}))\ \nabla_{\boldsymbol{e}} \boldsymbol{d}(\boldsymbol{e}) && (28) \\
&= -\boldsymbol{F}\ \nabla_{\boldsymbol{e}} \boldsymbol{\Omega} \boldsymbol{e} = -\boldsymbol{F}\ \boldsymbol{\Omega}, && (29)
\end{aligned}$$

where $\boldsymbol{F} \in \mathbb{R}^{N^2 \times N^2}$ is a diagonal matrix whose diagonal elements are given by

$$\boldsymbol{F}_{i,i} = \frac{\partial f(\boldsymbol{d})}{\partial \boldsymbol{d}_i} \qquad \forall i \in [1, N^2], \tag{30}$$

The specific threshold function and its derivative are

$$f(\boldsymbol{d}) = \begin{bmatrix} \frac{1}{1+e^{-\gamma(\boldsymbol{d}_1 - T)}} \\ \vdots \\ \frac{1}{1+e^{-\gamma\left(\boldsymbol{d}_{N^2} - T\right)}} \end{bmatrix}, \ \frac{\partial f(\boldsymbol{d})}{\partial \boldsymbol{d}} = \begin{bmatrix} \frac{\gamma e^{-\gamma(\boldsymbol{d}_1 - T)}}{\left(1+e^{-\gamma(\boldsymbol{d}_1 - T)}\right)^2} \\ \vdots \\ \frac{\gamma e^{-\gamma\left(\boldsymbol{d}_{N^2} - T\right)}}{\left(1+e^{-\gamma\left(\boldsymbol{d}_{N^2} - T\right)}\right)^2} \end{bmatrix} \tag{31}$$

Note that the transition from equation (27) to (28) employs the fact that $f(\cdot)$ operates element-wise, so that the "vec" operator can be mapped through to the argument, and secondly the product rule is used.

C. Scaling Matrix $\boldsymbol{H}$

The scaling matrix employed here is based on the standard sum-of-squares Hessian approximation used in the Gauss-Newton approach to unconstrained optimisation, combined with the barrier term. The Gauss-Newton approximation can be motivated by noticing that [64],

$$\nabla^2_{\boldsymbol{e}} V_\mu(\boldsymbol{e}) = \underbrace{\sum_{i=1}^{N^2}\sum_{j=1}^{N^2} \frac{\partial^2 \boldsymbol{r}(\boldsymbol{e})}{\partial \boldsymbol{e}_i \partial \boldsymbol{e}_j} \boldsymbol{r}_i(\boldsymbol{e})}_{\tilde{\boldsymbol{H}}} \tag{32}$$

$$+ \underbrace{\frac{2}{N^2}\Phi^T\Phi + \frac{2\lambda_3}{N^2}\boldsymbol{\Omega}^T\boldsymbol{\Omega} + \frac{\mu}{N^2} \begin{bmatrix} \frac{1}{\boldsymbol{e}_1^2} & & \\ & \ddots & \\ & & \frac{1}{\boldsymbol{e}_{N^2}^2} \end{bmatrix}}_{\boldsymbol{H}} \tag{33}$$

and that $\tilde{\boldsymbol{H}}$ contains all the components that might contribute to directions of negative curvature since $\boldsymbol{H}$ is positive definite by construction. At the same time, it is desired that the error term $\boldsymbol{r}(\boldsymbol{e})$ tends to zero so that $\tilde{\boldsymbol{H}}$ diminishes as the solution is approached (while this is desired, it is rarely achieved in practice). Therefore, the Hessian approximation $\boldsymbol{H}$ is used in this paper, i.e. for reference

$$\boldsymbol{H} \triangleq \frac{2}{N^2}\Phi^T\Phi + \frac{2\lambda_3}{N^2}\boldsymbol{\Omega}^T\boldsymbol{\Omega} + \frac{\mu}{N^2} \begin{bmatrix} \frac{1}{\boldsymbol{e}_1^2} & & \\ & \ddots & \\ & & \frac{1}{\boldsymbol{e}_{N^2}^2} \end{bmatrix} \tag{34}$$

D. Algorithm Parameter Values

Algorithm 1 is designed to aim for a local minima of $V_\mu(\boldsymbol{e})$ for a fixed value of μ, while gradually reducing μ so that $\boldsymbol{e}^*_\mu$ coincides with $\boldsymbol{e}^*$ in the limit as $\mu \to 0$. Therefore, it will not terminate until the barrier weighting μ is below some threshold value ϵ_μ. In addition, it is also required that the weighted gradient norm

$$\delta \triangleq \left(\boldsymbol{g}^T \boldsymbol{H}^{-1} \boldsymbol{g}\right)^{1/2} \tag{35}$$

is below the threshold defined by some fixed value ρ. This requirement ensures that the gradient is approaching zero at the solution (a first-order necessary condition of optimality). Note that since this paper employs a Hessian approximation

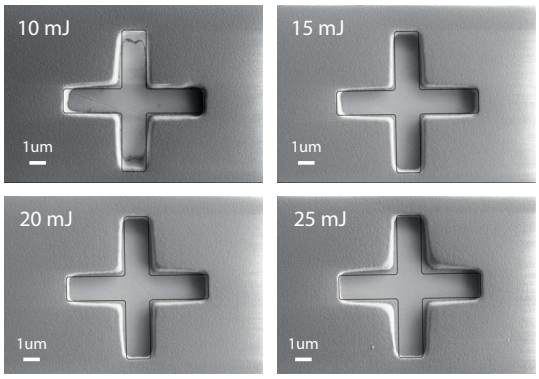

Fig. 6. Optimal exposures based on a threshold energy ranging from 10 mJ to 25 mJ. The black outline illustrates the desired feature.

$\boldsymbol{H}$ as the weighting matrix, then (35) may be likened to the Newton-decrement norm defined in, for example, Section 2.2.1 in [65]).

For the optimization in this article, the following parameter value choices were made: $\epsilon_\mu = 10^{-16},\ \rho = 10^{-3},\ \kappa = 10^{-1}$.

V. Experimental Results

Before optimization, the photoresist parameters were identified by performing a number of optimizations and exposures for the cross feature shown in Figure 6. The 15-mJ threshold results in the closest match between the desired and experimental feature. The beam width was measured to be $w_0 = 413$ nm and the photoresist steepness parameter was estimated to be $\gamma = 5$.

With the threshold dosage identified, an optimal exposure pattern was determined for the feature shown in Figure 9. The optimization objectives were $\lambda_2 = 0.001$ and $\lambda_3 = 0.002$. The initial conditions for the exposure function were obtained by exposing at every point where the feature is positive, which is shown on the top left of Figure 7. The initial conditions result in a gross over-exposure which is evident in the dosage and feature geometry plotted in the top row of Figure 7. After 10 iterations (middle row), the exposure function and feature geometry are observed to show significant improvements. After 35 iterations, the algorithm converges to an optimal solution with excellent correlation between the desired and predicted exposures.

On a DELL XPS 15 9550, with an Intel i7-6700HQ CPU running at 2.6 GHz and 32GB RAM, on Windows 10 64-Bit with Matlab 2016a, the average time per iteration was 1.56 s. The cost function and feature error are plotted against iteration number in Figure 8. It can be observed that the first ten iterations primarily reduce the exposure energy, while the second ten iterations primarily reduce the feature error.

The optimal exposure pattern was implemented experimentally which resulted in the developed feature shown in Figure 9. Also shown is a comparison of the predicted and experimental feature area. This comparison shows under exposure of sharp corners, and over exposure between areas in close proximity. These results suggest that the accuracy of the process model could be improved. In particular, these aspects include the beam profile, scatter induced proximity effects, and background exposure due to finite contrast in the laser modulator.

VI. Conclusions

This article describes optimal exposure planning for a scanning laser lithography system. The problem is cast as a non-linear program and solved using Newtons method by calculating an analytical gradient and Hessian approximation. The proposed method is demonstrated experimentally by exposing a test feature.

In prototyping and low-volume production applications, where a small exposure area is required, the proposed method has the potential to overcome the high cost of mask-based lithography.

To improve the feature quality, current research includes improving the accuracy of the process model by developing optimized experiments that identify the beam profile and photoresist parameters. Other processes such as scattering and background exposure due to the finite contrast of the laser modulator are also under investigation.

Methods for improving the numerical efficiency are also under investigation. This includes the exploitation of sparsity which arises from the finite diameter of the beam profile. To improve the process speed, motion planning for the nanopositioner is also being investigated. This includes limiting the number of exposure points, and continuous rather than pointwise scanning.

References

[1] H. J. Levinson, *Principles of lithography*, 3rd ed., ser. Press monograph. Bellingham, Wash: SPIE Press, 2010, no. 198.
[2] H. Butler, "Position control in lithographic equipment [applications of control]," *IEEE Control Systems*, vol. 31, no. 5, pp. 28–47, Oct 2011.
[3] V. M. Martinez and T. F. Edgar, "Control of lithography in semiconductor manufacturing," *IEEE Control Systems*, vol. 26, no. 6, pp. 46–55, Dec 2006.
[4] B. J. Lin, "Marching of the microlithography horses: electron, ion, and photon: past, present, and future," in *Proc. SPIE Optical Microlithography*, vol. 6520, March, 2007.
[5] M. Altissimo, "E-beam lithography for micro-/nanofabrication," *Biomicrofluidics*, vol. 4, no. 2, p. 026503, 2010.
[6] A. Joshi-Imre and S. Bauerdick, "Direct-Write Ion Beam Lithography," *Journal of Nanotechnology*, vol. 2014, p. e170415, feb 2014.
[7] E. S. Park, D. Jang, J. Lee, Y. J. Kim, J. Na, H. Ji, J. W. Choi, and G.-T. Kim, "Maskless optical microscope lithography system," *Review of Scientific Instruments*, vol. 80, no. 12, p. 126101, 2009.
[8] D. Gil, R. Menon, and H. I. Smith, "The case for diffractive optics in maskless lithography," *Journal of Vacuum Science & Technology B*, vol. 21, no. 6, pp. 2810–2814, 2003.
[9] H. I. Smith, R. Menon, A. Patel, D. Chao, M. Walsh, and G. Barbastathis, "Zone-plate-array lithography: a low-cost complement or competitor to scanning-electron-beam lithography," *Microelectronic engineering*, vol. 83, no. 4, pp. 956–961, 2006.
[10] L. Kipp, M. Skibowski, R. Johnson, R. Berndt *et al.*, "Sharper images by focusing soft x-rays with photon sieves," *Nature*, vol. 414, no. 6860, p. 184, 2001.
[11] R. Menon, D. Gil, G. Barbastathis, and H. I. Smith, "Photon-sieve lithography," *JOSA A*, vol. 22, no. 2, pp. 342–345, 2005.
[12] R. Garcia, A. W. Knoll, and E. Riedo, "Advanced scanning probe lithography," *Nature nanotechnology*, vol. 9, no. 8, pp. 577–587, 2014.
[13] D. Credgington, O. Fenwick, A. Charas, J. Morgado, K. Suhling, and F. Cacialli, "High-resolution scanning near-field optical lithography of conjugated polymers," *Advanced Functional Materials*, vol. 20, no. 17, pp. 2842–2847, 2010.
[14] B. S. Routley, J. L. Holdsworth, and A. J. Fleming, "Optimization of near-field scanning optical lithography," in *SPIE Advanced Lithography*, San Jose, CA, March 2015, p. 94230F.

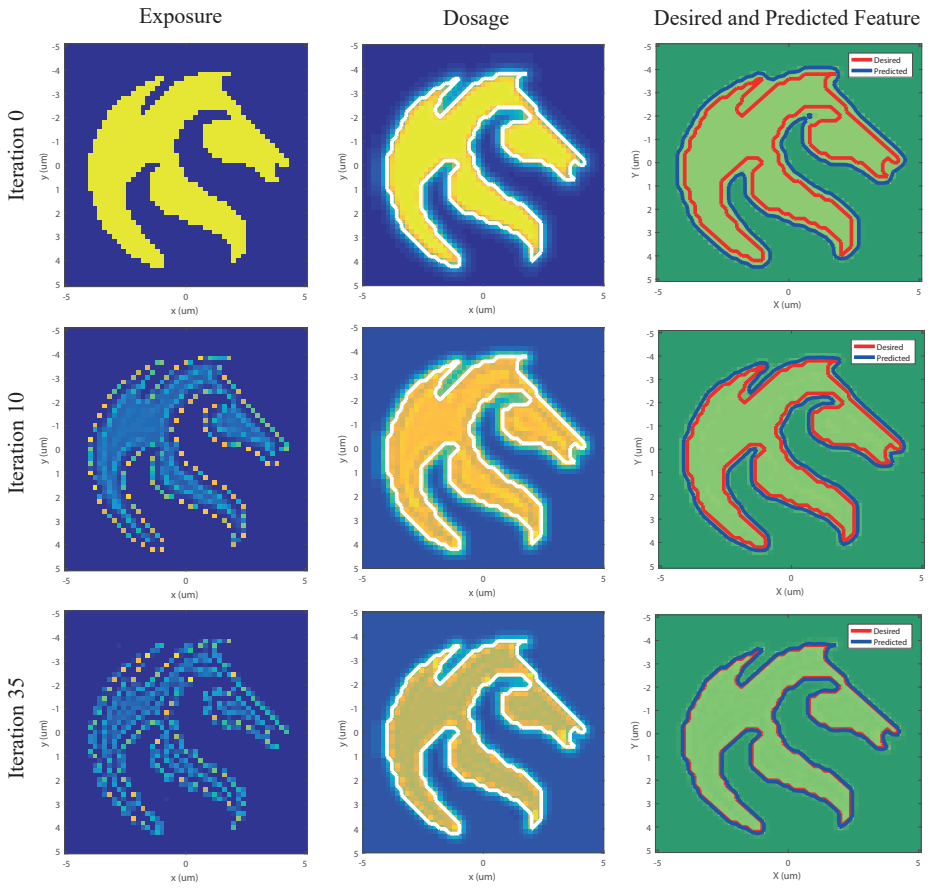

Fig. 7. The optimization results with the initial conditions, 10 iterations, and the final result. The exposure, resulting dosage, and developed feature are plotted in the left, middle, and right columns. The optimized feature is observed to closely match the desired feature in the right column.

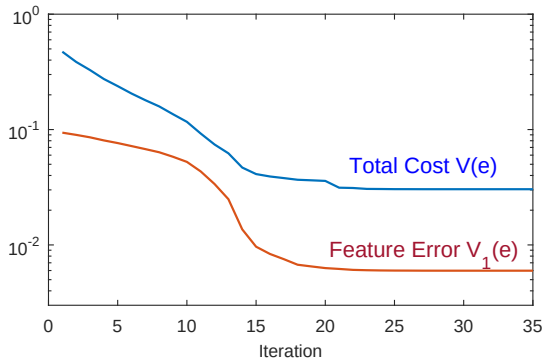

Fig. 8. The cost function $V(e)$ and feature error $V_1(e)$ versus iteration.

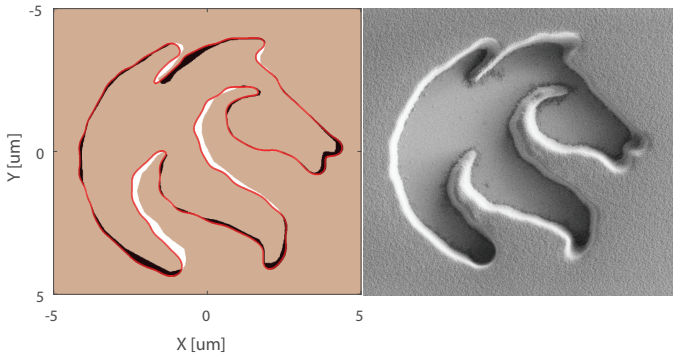

Fig. 9. Scanning electron micrograph of the developed feature (right). The measured feature area is compared to the prediction (left). The red outline is the predicted feature, areas of under exposure are shaded black, and areas of over exposure are shaded white. The total area of over and under exposure is 8.2% of the predicted area.

[15] L. Pan, Y. Park, Y. Xiong, E. Ulin-Avila, Y. Wang, L. Zeng, S. Xiong, J. Rho, C. Sun, D. B. Bogy *et al.*, "Maskless plasmonic lithography at 22 nm resolution," *Scientific reports*, vol. 1, p. 175, 2011.

[16] Y. Lin, M. H. Hong, W. J. Wang, Y. Z. Law, and T. C. Chong, "Sub-30 nm lithography with near-field scanning optical microscope combined with femtosecond laser," *Applied Physics A*, vol. 80, no. 3, pp. 461–465, Feb 2005.

[17] E. u. Haq, Z. Liu, Y. Zhang, S. A. A. Ahmad, L.-S. Wong, S. P. Armes, J. K. Hobbs, G. J. Leggett, J. Micklefield, C. J. Roberts *et al.*, "Parallel scanning near-field photolithography: the snomipede," *Nano letters*, vol. 10, no. 11, pp. 4375–4380, 2010.

[18] X. Liao, K. A. Brown, A. L. Schmucker, G. Liu, S. He, W. Shim, and C. A. Mirkin, "Desktop nanofabrication with massively multiplexed beam pen lithography," *Nature communications*, vol. 4, p. 2103, 2013.

[19] S. Bian, S. B. Zieba, W. Morris, X. Han, D. C. Richter, K. A. Brown, C. A. Mirkin, and A. B. Braunschweig, "Beam pen lithography as a new tool for spatially controlled photochemistry, and its utilization in the synthesis of multivalent glycan arrays," *Chemical Science*, vol. 5, no. 5, pp. 2023–2030, 2014.

[20] L. L. Cheong, P. Paul, F. Holzner, M. Despont, D. J. Coady, J. L. Hedrick, R. Allen, A. W. Knoll, and U. Duerig, "Thermal probe maskless lithography for 27.5 nm half-pitch si technology," *Nano letters*, vol. 13, no. 9, pp. 4485–4491, 2013.

[21] M. Malekian, S. Park, D. Strathearn, M. G. Mostofa, and M. Jun, "Atomic force microscope probe-based nanometric scribing," *Journal of Micromechanics and Microengineering*, vol. 20, no. 11, p. 115016,

2010.

[22] H. T. Soh, K. W. Guarini, and C. F. Quate, *Scanning probe lithography*. Springer Science & Business Media, 2013, vol. 7.

[23] A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*. London, UK: Springer, 2014.

[24] M. Fairbairn, S. O. R. Moheimani, and A. J. Fleming, "Q control of an atomic force microscope micro-cantilever: a sensor-less approach," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 20, no. 6, pp. 1372–1381, Dec. 2011.

[25] A. J. Fleming, "Measuring and predicting resolution in nanopositioning systems," *Mechatronics*, vol. 24, no. 6, pp. 605–618, Sep. 2014.

[26] A. J. Fleming, A. Wills, O. T. Ghalehbeygi, B. Routley, and B. Ninness, "A nonlinear programming approach to exposure optimization in scanning laser lithography," in *American Control Conference*, Boston, MA, July 2016.

[27] M. A. Mohammad, M. Muhammad, S. K. Dew, and M. Stepanova, "Fundamentals of electron beam exposure and development," in *Nanofabrication*. Springer, 2012, pp. 11–41.

[28] K. Yuan, B. Yu, and D. Z. Pan, "E-beam lithography stencil planning and optimization with overlapped characters," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 2, pp. 167–179, 2012.

[29] F. Yesilkoy, K. Choi, M. Dagenais, and M. Peckerar, "Implementation of e-beam proximity effect correction using linear programming techniques for the fabrication of asymmetric bow-tie antennas," *Solid-State Electronics*, vol. 54, no. 10, pp. 1211–1215, 2010.

[30] V. R. Manfrinato, J. Wen, L. Zhang, Y. Yang, R. G. Hobbs, B. Baker, D. Su, D. Zakharov, N. J. Zaluzec, D. J. Miller *et al.*, "Determining the resolution limits of electron-beam lithography: direct measurement of the point-spread function," *Nano letters*, vol. 14, no. 8, pp. 4406–4412, 2014.

[31] B. D. Cook and S.-Y. Lee, "PYRAMID-a hierarchical, rule-based approach toward proximity effect correction. II. Correction," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 1, pp. 117–128, Feb 1998.

[32] S.-Y. Lee and B. D. Cook, "PYRAMID-a hierarchical, rule-based approach toward proximity effect correction. I. Exposure estimation," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 1, pp. 108–116, Feb 1998.

[33] J. Bolten, T. Wahlbrink, N. Koo, H. Kurz, S. Stammberger, U. Hofmann, and N. Ünal, "Improved cd control and line edge roughness in e-beam lithography through combining proximity effect correction with gray scale techniques," *Microelectronic Engineering*, vol. 87, no. 5, pp. 1041–1043, 2010.

[34] J. Bolten, T. Wahlbrink, M. Schmidt, H. D. Gottlob, and H. Kurz, "Implementation of electron beam grey scale lithography and proximity effect correction for silicon nanowire device fabrication," *Microelectronic Engineering*, vol. 88, no. 8, pp. 1910–1912, 2011.

[35] L. E. Ocola, D. J. Gosztola, D. Rosenmann, and G. Lopez, "Automated geometry assisted proximity effect correction for electron beam direct write nanolithography," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 33, no. 6, p. 06FD02, Sep 2015.

[36] H. J. Levinson, *Principles of Lithography*. Washington: The Society of Photo-Optical Instrumentation Engineers, 2005.

[37] B. W. Smith and K. Suzuki, *Microlithography: science and technology*. CRC press, 2007, vol. 126.

[38] A. K.-K. Wong, *Resolution enhancement techniques in optical lithography*. SPIE press, 2001, vol. 47.

[39] L. Capodieci, "From optical proximity correction to lithography-driven physical design (1996-2006): 10 years of resolution enhancement technology and the roadmap enablers for the next decade," in *Proc. SPIE*, vol. 6154, 2006, p. 615401.

[40] R. Sochaa, X. Shib, and D. LeHotyb, "Simultaneous source mask optimization (smo)," in *Proc. of SPIE Vol*, vol. 5853, 2005, p. 181.

[41] D. Melville, A. E. Rosenbluth, K. Tian, K. Lai, S. Bagheri, J. Tirapu-Azpiroz, J. Meiring, S. Halle, G. McIntyre, T. Faure *et al.*, "Demonstrating the benefits of source-mask optimization and enabling technologies through experiment and simulations," in *SPIE Advanced Lithography*, 2010, pp. 764 006–764 006.

[42] X. Ma and G. R. Arce, "Pixel-based simultaneous source and mask optimization for resolution enhancement in optical lithography," *Optics express*, vol. 17, no. 7, pp. 5783–5793, 2009.

[43] N. Jia and E. Y. Lam, "Pixelated source mask optimization for process robustness in optical lithography," *Optics express*, vol. 19, no. 20, pp. 19 384–19 398, 2011.

[44] S. Li, X. Wang, and Y. Bu, "Robust pixel-based source and mask optimization for inverse lithography," *Optics & Laser Technology*, vol. 45, pp. 285–293, 2013.

[45] X. Ma, C. Han, Y. Li, L. Dong, and G. R. Arce, "Pixelated source and mask optimization for immersion lithography," *JOSA A*, vol. 30, no. 1, pp. 112–123, 2013.

[46] X. Ma, Y. Li, and L. Dong, "Mask optimization approaches in optical lithography based on a vector imaging model," *JOSA A*, vol. 29, no. 7, pp. 1300–1312, 2012.

[47] X. Liu, R. Howell, S. Hsu, K. Yang, K. Gronlund, F. Driessen, H.-Y. Liu, S. Hansen, K. van Ingen Schenau, T. Hollink *et al.*, "Euv source-mask optimization for 7nm node and beyond," in *SPIE Advanced Lithography*, 2014, pp. 90 480Q–90 480Q.

[48] X. Wu, S. Liu, J. Li, and E. Y. Lam, "Efficient source mask optimization with zernike polynomial functions for source representation," *Optics express*, vol. 22, no. 4, pp. 3924–3937, 2014.

[49] J. Li, S. Liu, and E. Y. Lam, "Efficient source and mask optimization with augmented lagrangian methods in optical lithography," *Optics express*, vol. 21, no. 7, pp. 8076–8090, 2013.

[50] Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," *IEEE Transactions on Semiconductor Manufacturing*, vol. 5, no. 2, pp. 138–152, May 1992.

[51] S. Sherif, B. Saleh, and R. De Leone, "Binary image synthesis using mixed linear integer programming," *IEEE transactions on image processing*, vol. 4, no. 9, pp. 1252–1257, 1995.

[52] Y. Oh, J.-C. Lee, and S. Lim, "Resolution enhancement through optical proximity correction and stepper parameter optimization for 0.12-m mask pattern," in *Proc. SPIE Optical Microlithography*, vol. 3679, 1999, pp. 607–613.

[53] T. Matsunawa, H. Nosato, H. Sakanashi, M. Murakawa, E. Takahashi, T. Terasawa, T. Tanaka, O. Suga, and T. Higuchi, "Adaptive Optical Proximity Correction Using an Optimization Method," in *7th IEEE International Conference on Computer and Information Technology (CIT 2007)*, Oct 2007, pp. 853–860.

[54] Y. Granik, "Fast pixel-based mask optimization for inverse lithography," *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 5, no. 4, pp. 043 002–043 002–13, 2006.

[55] L. Pang, G. Dai, T. Cecil, T. Dam, Y. Cui, P. Hu, D. Chen, K.-H. Baik, and D. Peng, "Validation of inverse lithography technology (ilt) and its adaptive sraf at advanced technology nodes," in *Optical Microlithography XXI*, vol. 6924, San Jose, CA, March 2008, pp. 69 240T, 12 Pages.

[56] A. Poonawala and P. Milanfar, "Mask design for optical microlithography; an inverse imaging problem," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 774–788, March 2007.

[57] X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," *Optics Express*, vol. 15, no. 23, pp. 15 066–15 079, 2007.

[58] X. Ma and G. Arce, "Binary mask optimization for inverse lithography with partially coherent illumination," *JOSA A*, vol. 25, no. 12, pp. 2960–2970, 2008.

[59] K. Lai, A. E. Rosenbluth, S. Bagheri, J. Hoffnagle, K. Tian, D. Melville, J. Tirapu-Azpiroz, M. Fakhry, Y. Kim, S. Halle *et al.*, "Experimental result and simulation analysis for the use of pixelated illumination from source mask optimization for 22nm logic lithography process," in *SPIE Advanced Lithography*, 2009, pp. 72 740A–72 740A.

[60] Y. Peng, J. Zhang, Y. Wang, and Z. Yu, "Gradient-based source and mask optimization in optical lithography," *IEEE Transactions on image processing*, vol. 20, no. 10, pp. 2856–2864, 2011.

[61] O. T. Ghalehbeygi, A. G. Wills, B. S. Routley, and A. J. Fleming, "Gradient-based optimization for efficient exposure planning in maskless lithography," *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 16, no. 3, p. 033507, Sep. 2017.

[62] A. Wills, D. Bates, A. J. Fleming, B. Ninness, and S. O. R. Moheimani, "Application of MPC to an active structure using sampling rates up to 25khz," in *IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, Dec. 2005, pp. 3176–3181.

[63] A. V. Fiacco and G. P. McCormick, *Nonlinear programming; sequential unconstrained minimization techniques*. Wiley New York, 1968.

[64] J. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1996.

[65] Y. Nesterov and A. Nemirovsky, *Interior-point polynomial methods in convex programming, volume 13 of*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1994.