Research Article Vol. 30, No. 12 / 6 Jun 2022 / Optics Express 20564
Optics EXPRESS

Memory efficient constrained optimization of scanning-beam lithography

CARL JIDLING,[1,*] ANDREW J. FLEMING,[2] ADRIAN G. WILLS,[2] AND THOMAS B. SCHÖN[1]

[1]Department of Information Technology, Uppsala University, Lägerhyddsvägen 1 Uppsala, Sweden
[2]School of Engineering, University of Newcastle, Callaghan NSW 2308, Australia
[*]carl.jidling@it.uu.se

Abstract: This article describes a memory efficient method for solving large-scale optimization problems that arise when planning scanning-beam lithography processes. These processes require the identification of an exposure pattern that minimizes the difference between a desired and predicted output image, subject to constraints. The number of free variables is equal to the number of pixels, which can be on the order of millions or billions in practical applications. The proposed method splits the problem domain into a number of smaller overlapping subdomains with constrained boundary conditions, which are then solved sequentially using a constrained gradient search method (L-BFGS-B). Computational time is reduced by exploiting natural sparsity in the problem and employing the fast Fourier transform for efficient gradient calculation. When it comes to the trade-off between memory usage and computational time we can make a different trade-off compared to previous methods, where the required memory is reduced by approximately the number of subdomains at the cost of more computations. In an example problem with 30 million variables, the proposed method reduces memory requirements by 67% but increases computation time by 27%. Variations of the proposed method are expected to find applications in the planning of processes such as scanning laser lithography, scanning electron beam lithography, and focused ion beam deposition, for example.

1. Introduction

During integrated circuit fabrication, materials are selectively added or subtracted by depositing a layer of resist material, then modifying certain areas using a lithography process [1,2]. The lithography process involves exposing the resist to light through a complex mask or reticle. The exposed resist is then removed (or retained) through a development process. Many mask sets are required to produce a circuit, which can be prohibitively expensive for low-volume or prototype applications.

To eliminate the need for mask sets, maskless lithography methods have evolved for low-volume or quick turnaround applications. These processes use scanning laser beams [3,4], electron beams [5,6], or ion-beams [7–9] to directly modify the substrate or expose resist. Scanning laser and electron beam lithography is also used in many other low-volume 2D micro-fabrication applications, such as microfluidics [10], meta-materials [11], and reticle fabrication for projection lithography [12]. The processing speed can be improved using methods such as zone-plate arrays [13,14] or photon sieve lithography [15,16] which generate an array of focused spots. Other serial processes include near-field optical probe lithography [17–22], thermal-probe lithography [23,24] and mechanical-probe lithography [25,26].

Although the exposure source and physical mechanism of serial maskless lithography methods are varied, many can be described as a controllable source with a known spatial distribution. For example, the spatial distribution of scanning optical and electron beam lithography is limited by the wavelength and numerical aperture. Since the dimension of the desired features are similar in scale to the spatial distribution of the beam, an optimization problem arises. In scanning beam

#457334 https://doi.org/10.1364/OE.457334
Journal © 2022 Received 13 Mar 2022; revised 13 May 2022; accepted 16 May 2022; published 24 May 2022

Optics EXPRESS

lithography, a set of exposure locations and beam power settings must be found that optimize the resolution of developed features [27–29]. A similar planning step is required by any method where the spatial distribution of the source cannot be neglected.

The first methods for exposure planning used rules [30,31] that were similar to those employed in projection lithography. These were improved by linear programming methods [29] that were commercialized for proximity correction [32–34]. In 2017, quadratic cost functions were introduced to optimize feature geometry with regularization of exposed power [35]. A non-linear programming approach with interior-point optimization was described in [36]. Although this method was quick to converge due to the calculation of first and second order derivatives, it required the storage of an $N^2 \times N^2$ matrix for an $N \times N$ problem, which is only suitable for small problems (300×300). The numerical efficiency was later improved by exploiting sparsity and approximating the first derivative [4,35]; however, the required memory was still on the order of $N^2 \times N^2$. An alternative to optimization is deconvolution based on the Fredholm integral [37], which reduces the largest matrix to $N \times N$; however, this method requires an order of magnitude more iterations than gradient based methods and does not minimize a known cost function so optimality is neither guaranteed nor expected. These three methods are directly compared in [38]. In summary, methods for planning serial maskless lithography processes are either limited by memory or do not guarantee optimality and require significant iterations.

Although this work focuses on maskless lithography, it should be noted that exposure optimization problems also exists in projection lithography [39–42]. However, these methods are fundamentally different since the optimization variables are the binary mask and source pattern [43–45]. Due to the large scale of mask optimization problems, considerable efforts have been made for improving the numerical efficiency; for example, through the use of basis functions [46], Lagrangian methods [47], set-based methods [48,49], and neural networks [50–52].

The *contribution* of this article is to reduce the memory requirements of gradient-based optimization methods for scanning-beam lithography planning. The proposed method splits the problem domain into a number of smaller overlapping subdomains with constrained boundary conditions, which are then solved sequentially using a constrained gradient search method (L-BFGS-B). Computational time is reduced by exploiting natural sparsity in the problem and employing the fast Fourier transform for efficient gradient calculation. Compared to previous methods [4,35–37], the proposed method is slower but the required memory is reduced by approximately the number of subdomains.

The following three sections describe the forward process model, define the optimization problem, and derive the analytical gradients of the cost function. The problem space is them subdivided in Section 5, followed by memory efficient optimization in Section 6, and an example application in Section 7. A complexity analysis and an investigation of accuracy is presented in Section 8, then the article is concluded in Section 9.

2. Process model

This section describes a general model of scanning-beam lithography processes. Given an input exposure pattern, the model predicts the photoresist conversion fraction, which is directly related to developed features. The model assumes that the photoresist is sufficiently thin so that the beam profile remains approximately constant throughout the 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 ignored.

An introduction to the scanning-beam lithography process is illustrated in Fig. 1. This figure shows how a set of discrete exposures can be used to create features with a resolution comparable to the beam width. However, the middle row demonstrates that simple choices of exposure energy and location result in sub-optimal developed features. The bottom row illustrates how an

Optics EXPRESS

optimized exposure pattern yields the best possible feature resolution, and it is also a numerical example of the proposed method.

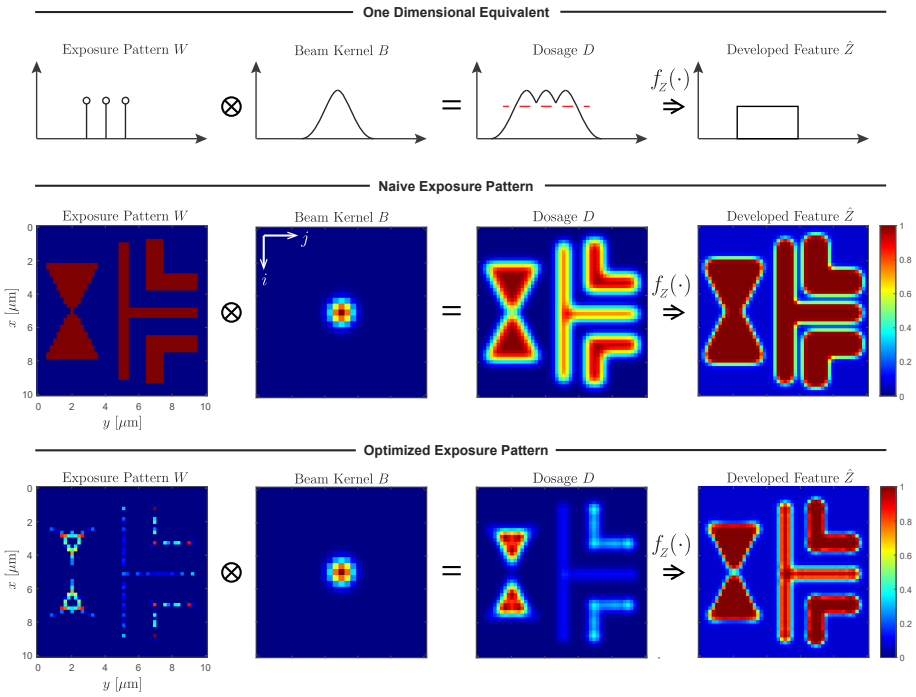


Fig. 1. Illustration of the scanning-beam lithography process. The top row shows a one-dimensional interpretation, where a discrete exposure pattern W represents the beam exposure time or power setting (vertical axis) at discrete locations (horizontal axis). The resulting dosage D (exposure energy per unit area) is the sum of beam kernels B, scaled and shifted by the exposure pattern W. The output feature $\hat{Z}$ is predicted by the threshold function $f_Z(\cdot)$ (Eq. (7)) which models the development process. In the middle row, a two dimensional exposure problem is considered. The 'naive' exposure pattern is a scaled version of the desired feature; however, the beam kernel B can be observed to significantly expand the desired feature due to over exposure. The middle and bottom row use a realistic development function $f_Z(\cdot)$ plotted in Fig. 2. To minimize the achievable difference between the desired and developed features, the bottom row shows an example of an optimized exposure pattern, which results in the best possible developed feature given the beam kernel and development model.

The exposure pattern W, beam kernel B, dosage D, and predicted feature $\hat{Z}$ are matrices which contain values at discrete locations in a workspace. The workspace is represented by a uniformly spaced rectangular grid

$$X = \begin{bmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_{1N_2} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{N_1 1} & \cdots & \mathbf{x}_{N_1 N_2} \end{bmatrix}, \tag{1}$$

where $\mathbf{x}_{ij} = [x_i^{(1)} \;\; x_j^{(2)}]^\mathsf{T}$. We let $\delta_p = x_{i+1}^{(p)} - x_i^{(p)}$ denote the grid resolution in direction p, whereas $N = N_1 N_2$ is the grid size (number of grid points). Each grid point is associated with a desired feature value stored in the binary matrix $Z = [z_{ij}]$.

Optics EXPRESS

The grid is exposed to a scanned beam with intensity modelled by an exponential function, which represents a two-dimensional Gaussian beam with arbitrary x and y axis width and rotation, as described in [35]. That is,

$$B_H(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^\mathsf{T} H^{-1}\mathbf{x}\right). \tag{2}$$

We refer to $H = [h_{ij}]$ as the *bandwidth*, with entries defined in terms of its inverse according to

$$[H^{-1}]_{11} = \frac{4\cos^2\phi}{w_{x0}^2} + \frac{4\sin^2\phi}{w_{y0}^2}, \tag{3}$$

$$[H^{-1}]_{12} = [H^{-1}]_{21} = -\frac{2\sin 2\phi}{w_{x0}^2} + \frac{2\sin 2\phi}{w_{y0}^2}, \tag{4}$$

$$[H^{-1}]_{22} = \frac{4\sin^2\phi}{w_{x0}^2} + \frac{4\cos^2\phi}{w_{y0}^2}. \tag{5}$$

In this work, the UV scanning laser lithography system described in [35] will be used as an example. The beam parameters for this system were measured to be $w_{x0} = 570$ nm, $w_{y0} = 560$ nm and $\phi = 2.2°$ [35].

The dosage D represents the absorbed energy per unit area, which is determined by the exposure pattern W and beam kernel B. The exposure W can represent any variable that is proportional to energy, e.g. beam power for a constant exposure time, or exposure time with a constant beam power. The dosages and exposure values are stored in the matrices $D = [d_{qk}]$ and $W = [w_{ij}]$, respectively. At a given point, the total dosage is found by summing the contributions from the entire grid

$$d_{qk}(\mathbf{w}) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} w_{ij} B_H(\mathbf{x}_{ij} - \mathbf{x}_{qk}), \tag{6}$$

where $\mathbf{w} = \operatorname{vec} W$.

As the photoresist is exposed by the beam, the increasing dosage leads to a chemical change due to photon absorption or photocatalysis. A negative photoresist becomes less sensitive to a developing agent after exposure and the developed photoresist features are similar to the dosage pattern. Conversely, a positive photoresist becomes more sensitive to the developing agent and results in subtractive features that are similar to the dosage pattern. Regardless of the photoresist polarity, the development process is dependent on the fraction of photoresist that is converted; therefore, this quantity is used to define the desired feature. To complete the process model, a function is required that maps the dosage to the predicted conversion fraction.

The simplest model of the development process is a threshold function, where the photoresist is assumed to be 100% converted above a certain dosage threshold, as depicted in the top row of Fig. 1. However, in general the conversion function can be any increasing function of dosage. In this work, the sigmoid threshold function plotted in Fig. 2 is used to model the development process [35–37]. The sigmoid mapping function $f_Z(\cdot)$ from the dosage to the predicted feature $\hat{Z} = [z_{qk}]$ is

$$\hat{z}_{qk}(\mathbf{w}) = f_Z\left(d_{qk}(\mathbf{w})\right) = \left[1 + \exp\left[-\alpha(d_{qk}(\mathbf{w}) - d_{50\%})\right]\right]^{-1}, \tag{7}$$

where α is the steepness of the curve, and $d_{50\%}$ is the dosage where half of the photoresist is converted. The steepness is plotted for $\alpha = 5, 10, 20, 40$ in Fig. 2. In Section 7, the proposed method is applied to the optimization of ultraviolet scanning laser lithography with a positive photoresist (AZ ECI3007, MicroChemicals, Germany). The conversion fraction of this photoresist was adequately modeled using $\alpha = 5$ in Ref. [35–37]. It is convenient to normalize the fully converted photoresist to 1 and therefore $d_{50\%} = 0.5$.

Optics EXPRESS

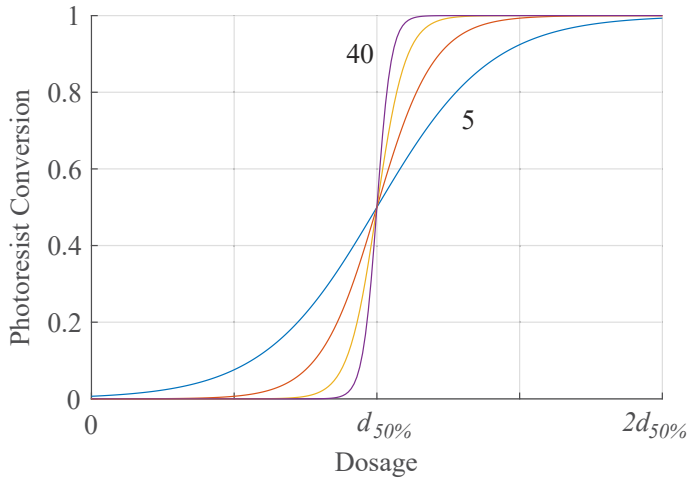


Fig. 2. The sigmoid threshold function $f_Z(\cdot)$ defined in Eq. (7) models the development process. In scanning laser lithography, this function relates the exposure energy to the fraction of converted photoresist. It is parameterized by the steepness α and the dosage $d_{50\%}$ where half of the photoresist is converted. The steepness is plotted for $\alpha = 5,\ 10,\ 20,\ 40$.

3. Optimization problem

The optimization objective is to find an exposure pattern $w_{ij} \geq 0$ that minimizes the difference between the desired feature Z and the predicted feature $\hat{Z}$. The non-negative constraint on w_{ij} is due to the nature of exposure energy, which cannot be less than zero. It also desirable to minimize the total dosage, since this minimizes negative effects such as heating, scattering, and other background exposure processes such as reflection to and from the substrate and objective lens. These optimization objectives are summarized by the following problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}) = \sum_{q=1}^{N_1}\sum_{k=1}^{N_2}[z_{qk} - \hat{z}_{qk}(\mathbf{w})]^2 + \gamma d_{qk}^2(\mathbf{w}), \\ \text{subject to} \quad & w_{ij} \geq 0. \end{aligned} \tag{8}$$

The penalty parameter γ>0 balances the trade-off between feature matching and total dosage. Penalizing the L_2 norm is preferred in this application as the most significant sources of background exposure are scatter within the substrate, and reflection between the substrate and objective. Both of these sources are proportional to cumulative power, which is proportional to the L_2 norm. In this work, the optimal value of γ was found to be 10^{-4}, which is recommended as a starting point in other applications. Some experimentation may be required as the regularization is likely to be affected by other factors such as photoresist properties.

4. Gradient calculation

Following the procedure described in [53], the fast Fourier transform (FFT) is utilized for efficient computation of the cost function and its gradient. First, Eq. (6) is rewritten as a convolution

$$d_{qk} = \sum_{i=-L_1}^{L_1}\sum_{j=-L_2}^{L_2} w_{q-i,k-j} B_H^{i,j}, \tag{9}$$

where $B_H^{i,j} = B_H([i\delta_1, j\delta_2]^\mathsf{T})$ and $w_{q-i,k-j} = 0$ if the indexed point lies outside the grid. The dosage in Eq. (6) is recovered if $L_1 = N_1 - 1$ and $L_2 = N_2 - 1$; however, as the beam intensity of Eq. (2)

Optics EXPRESS

decays exponentially and therefore in practice lacks support outside a finite region, Eq. (6) can be approximated by shrinking L_1 and L_2 for faster computations. A suitable choice is

$$L_p = \min\left(N_p - 1, \lceil \frac{\tau\sqrt{\lambda}}{\delta_p} \rceil\right), \tag{10}$$

where τ is a user-defined scaling parameter, λ is the maximum eigenvalue of H and $\lceil\cdot\rceil$ denotes the ceiling function. The authors of [53] suggest using $\tau = 3.7$ in the context of bivariate kernel density estimation; however, to safe-guard against numerical inaccuracies, a slightly higher value of $\tau = 10$ is used here.

The gradient of the cost function in Eq. (8) can also be expressed as a convolution [35]. Specifically, $\nabla f = \operatorname{vec} G$ where $G = [g_{qk}]$ and

$$g_{qk} = \sum_{i=-L_1}^{L_1}\sum_{j=-L_2}^{L_2}\left[-2\alpha\hat{z}_{q-i,k-j}(1-\hat{z}_{q-i,k-j})(z_{q-i,k-j}-\hat{z}_{q-i,k-j}) + 2\gamma d_{q-i,k-j}\right]B_H^{i,j}. \tag{11}$$

Before applying the FFT, adequate zero-padding is required as the bandwidth H is *unconstrained* (non-diagonal) [53]. The matrices are

$$\mathbf{B} = \begin{bmatrix} B_H^{-L_1,-L_2} & B_H^{-L_1,-(L_2-1)} & \cdots & B_H^{-L_1,L_2} & \mathbf{0} \\ B_H^{-(L_1-1),-L_2} & \ddots & \cdots & \vdots & \vdots \\ \vdots & \cdots & \ddots & \vdots & \vdots \\ B_H^{L_1,-L_2} & \cdots & \cdots & B_H^{L_1,L_2} & \vdots \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \end{bmatrix}, \tag{12}$$

and

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \vdots & w_{11} & \cdots & w_{1N_2} & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & w_{N_1 1} & \cdots & w_{N_1 N_2} & \vdots \\ \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \end{bmatrix}, \tag{13}$$

where the entry w_{11} is placed at index (L_1+1, L_2+1). The matrices $\mathbf{Z}$, $\hat{\mathbf{Z}}$ and $\mathbf{D}$ are formed with the same structure as in $\mathbf{W}$. The $\mathbf{0}$-blocks are built to match the dimension $P_1 \times P_2$, where P_1 and P_2 are chosen as a suitable composite integers. Next, the dosage and gradient are computed

$$\mathbf{D} = \mathscr{F}^{-1}\left[\mathscr{F}[\mathbf{W}] \circ \mathscr{F}[\mathbf{B}]\right], \tag{14}$$

$$\mathbf{G} = \mathscr{F}^{-1}\left[\mathscr{F}[-2\alpha\hat{\mathbf{Z}} \circ (1-\hat{\mathbf{Z}}) \circ (\mathbf{Z}-\hat{\mathbf{Z}}) + 2\gamma\mathbf{D}] \circ \mathscr{F}[\mathbf{B}]\right], \tag{15}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ denote the FFT and its inverse, respectively, while $\circ$ denotes the Hadamard (element-wise) product of matrices. The desired matrices D and G are found in $\mathbf{D}$ and $\mathbf{G}$ as the blocks defined by the rows $2L_1+1$ to $2L_1+1+N_1$ and columns $2L_2+1$ to $2L_2+1+N_2$.

Optics EXPRESS

5. Subdomain division

For large problems sizes when the number of variables N renders the problem computationally infeasible, the optimization can be performed in different subdomains. The total solution is then obtained by concatenation of the local solutions.

Let $\mathbf{w}_*$ denote the optimal solution on the grid X, and let $\mathbf{w}_*^\Pi \in \mathbf{w}_*$ denote the optimal solution on the subdomain $\Pi \in X$. The limited support of the beam intensity (Eq. (2)) assures that a variable w_{ij} affects the value of f only in a limited region. That being said, there is a possibility of a chain effect propagating through the grid since nearby variables are affecting each other. However, the binary structure of Z significantly reduces this effect. Hence, by solving the problem on the subdomain, an estimate $\hat{\mathbf{w}}_*^\Pi$ of $\mathbf{w}_*^\Pi$ is obtained. However, solving the problem only on the domain Π is insufficient as the boundary components may be incorrect. The trade-off is to use another region Γ such that $\Pi \in \Gamma \in X$. Then $\hat{\mathbf{w}}_*^\Pi$ can be obtained by extracting the correct components from the solution $\hat{\mathbf{w}}_*^\Gamma$ computed on Γ. Thus, if Γ is sufficiently larger than Π, the difference between $\hat{\mathbf{w}}_*^\Pi$ and $\mathbf{w}_*^\Pi$ is negligible. In practice, for the class of problems under consideration, it was found to be reasonable to form Γ by adding L_p points to Π in direction p. A simple illustration of this subdomain division is given in Fig. 3.

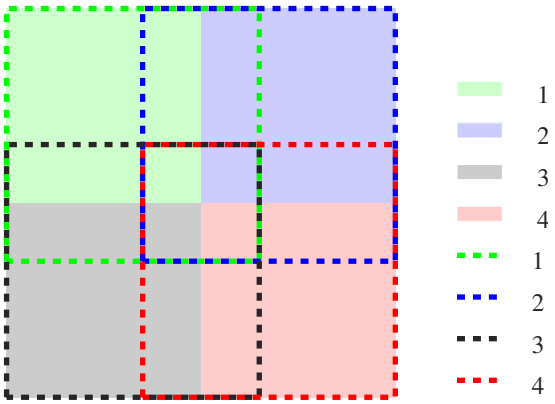


Fig. 3. Illustration of subdomain division. The problem is solved on the Γ-regions, and then the Π-regions are concatenated.

Although this approach is novel for the application under consideration, it should be stressed that subdomain division and local optimization do not constitute a novel idea in general. In a broader context, so called domain decomposition methods [54,55] are well-established tools for solving differential equations through subdomain division, and originate from the work by Hermann Schwarz in 1870 [54]. Due to the iterative procedure of these methods, they are in essence comprising local optimization problems. Example applications include finite element solvers [56], total variation de-noising [57] and plasticity with hardening [58]. Moreover, the same principle has for example been used for the ptychography problem [59], including efficient parallel GPU implementations [60,61]. In [62] a similar approach is used to reduce the computational time within composite pile foundation.

It can observed that the subproblems are independent of each other. In principle, this makes the optimization approach straight-forward to solve in parallel by distributing the subproblems on different CPU cores. However, this conflicts with the aim of reducing the memory requirements since the different processes compete for the same RAM memory. In the case where the RAM memory is sufficiently large, solving the subproblems in parallel will reduce the computation time by approximately the number of subdomains. Future improvements to computation time could also be achieved by utilizing a GPU for FFT computation.

Optics EXPRESS

6. Memory efficient optimization

The limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints (L-BFGS-B) [63] is employed to solve the optimization problem of Eq. (8). This relies upon the BFGS method [64], which is a well-established quasi-Newton algorithm that approximates the inverse Hessian matrix and provides significantly faster convergence.

The limited-memory BFGS (L-BFGS) [65] method is a modification of BFGS that avoids building an entire matrix, which is beneficial for high-dimensional problems. The L-BFGS method expresses the BFGS update recursively and disregards all but the m latest Hessian approximations. That way the computation of the search direction can be carried out in memory-efficient for-loops.

The L-BFGS-B method [63] is a further extension that allows for bounded input constraints of the form $l_{ij}<w_{ij}<u_{ij}$. At each iteration k, it minimizes a quadratic model of the cost function along the path $w_{ij}(t) = p(w_{ij}^{(k)} - tg_{ij}^{(k)}, l_{ij}, u_{ij})$, where

$$p(w, l, u) = \begin{cases} l & \text{if } w<l, \\ w & \text{if } w \in [l,\ u], \\ u & \text{if } w>u. \end{cases} \tag{16}$$

In the problem under consideration, $l_{ij} = 0$ and $u_{ij} = \infty$ for all inputs w_{ij}. The optimization is initialized by setting $w_{ij} = 1$ if $z_{ij} = 1$, which corresponds to all grid points that lie inside the desired feature pattern. The remaining variables are initialized to 0. Judging from previous experiments [35–37] and by intuition, grid points outside the pattern are unlikely to be exposed at the optimum.

To speed up the optimization, the sparsity of the problem is exploited. The finite region of the desired feature combined with the limited support of the kernel function allows for an *a priori* identification of grid points that will, in practice, be un-exposed at the optimum. For instance, variables at grid points that lie outside the desired pattern can be disregarded. Hence, the corresponding entries are removed from the gradient before it is returned to the L-BFGS-B routine.

In the implementation, we are using the MATLAB wrapper [66] of the L-BFGS-B implementation described in [67]. The proposed procedure is summarized in Algorithm 1 and illustrated in Fig. 4.

Note that the L-BFGS-B optimizer is not a crucial choice. There are several other optimizers suitable for the large-scale problems, such as gradient descent as used in [35]. The constraint $w_{ij} \geq 0$ can also be achieved through a variable transformation rather than by the optimizer, for instance by setting $w_{ij} = e^{p_{ij}}$ and then optimizing with respect to p_{ij}. A pure gradient based routine is more memory efficient than L-BFGS-B, which keeps track of the m latest pairs of input points and gradients. The experiment below uses $m = 5$, which adds a relatively small memory requirement compared to the matrix formations and operations involved when computing the cost function and its gradient. However, a smaller value of m can be used as well (even $m = 1$, which resembles a gradient descent optimizer). L-BFGS-B is used in this work as it is a well-established method with an efficient and readily available implementation, and it is appealing due to the Hessian approximation and the built in support for constrained variables. Finally, L-BFGS-B worked well in practice for a range of problems.

Optics EXPRESS

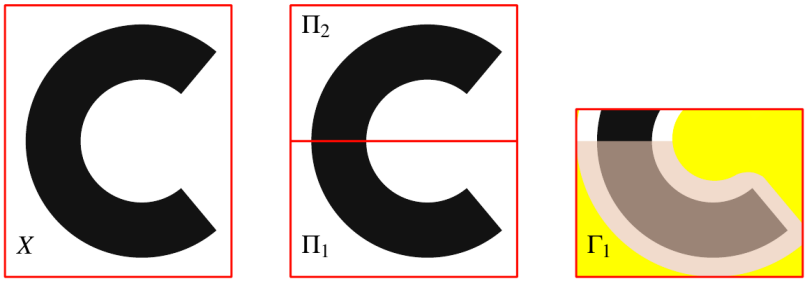


Fig. 4. Illustration of Algorithm 1. The desired feature is the C-shaped black region (left). Assuming that the problem is too large to be solved directly, the grid X is split horizontally into two equally-sized subdomains Π_1 and Π_2 (middle). To avoid undesired boundary effects in the solutions, the extended regions Γ_1 and Γ_2 are constructed, of which one is shown above (right). The yellow parts indicate variables that with great certainty will be zero at the optimum and therefore are excluded from the optimization. Hence, we are interested in the variables indicated by the beige overlay. These are then concatenated with the equivalent part of Γ_2 to form the solution over X.

7. Numerical example

This section demonstrates the proposed method on a medium scale lithography test pattern shown in Fig. 5(a). The test pattern represents a 100 μm × 100 μm area with 5486 × 5488 pixels and a resolution of 18.2 nm; that is, there are approximately 30 million variables to optimize. The beam parameters that represent the laser and optical system were experimentally identified in previous work as $w_{x0} = 570$ nm, $w_{y0} = 560$ nm and $\phi = 2.2°$ [35]. The photoresist fully exposed state is defined to be 1 and $d_{50\%} = 0.5$. The steepness of the photoresist exposure curve (Eq. (7)) is $\alpha = 5$, and the penalty parameter $\gamma = 10^{-4}$.

The test pattern shown in Fig. 5(a) examines the limits of a lithography process and is ideal for evaluating two important aspects of the optimization procedure. First, the optimization should result in consistent lines and features that do not vary along the length, or at different locations in the pattern, unless there are nearby features that require consideration. In other words, there should be no visible artefacts or asymmetry in geometric features. Secondly, the optimized features should degrade predictably as the objective becomes unachievable. The minimum feature size in scanning beam lithography is determined by the steepness of the photoresist exposure curve and the beam width. In this example where $\alpha = 5$, the minimum size of an exposed feature is approximately equal to the beam width, which is shown for reference in the close-up views in Fig. 5.

The optimization procedure was completed for two cases, one where the entire pattern was optimized in one step, and another where the area is split into a 2 × 2 subdomain array. Both optimizations resulted in numerically identical solutions but the 2 × 2 subdomain required 67% less memory at the expense of 27% longer optimization time. The longer optimization time is due to additional computations required by the overlap area. The overlap area is also the reason why memory usuage is not reduced by 75%. The numerical results are summarized in Table 1.

Since the optimization results were numerically identical, only the results for the 2 × 2 solution are reported in Fig. 5. The top row of Fig. 5 shows the initial guess of the exposure pattern (Fig. 5(a)), which results in the dosage (Fig. 5(b)) and the resulting feature (Fig. 5(c)), which is grossly over exposed. The pixels of the initial guess (Fig. 5(a)) are set to 1 where the desired output feature is non-zero; therefore, (Fig. 5(a)) represents both the initial guess and the desired output feature. The middle row shows the optimized exposure pattern (Fig. 5(d)) and the resulting dosage (Fig. 5(e)) and output feature (Fig. 5(f)). In the close up view of the corner features

Optics EXPRESS

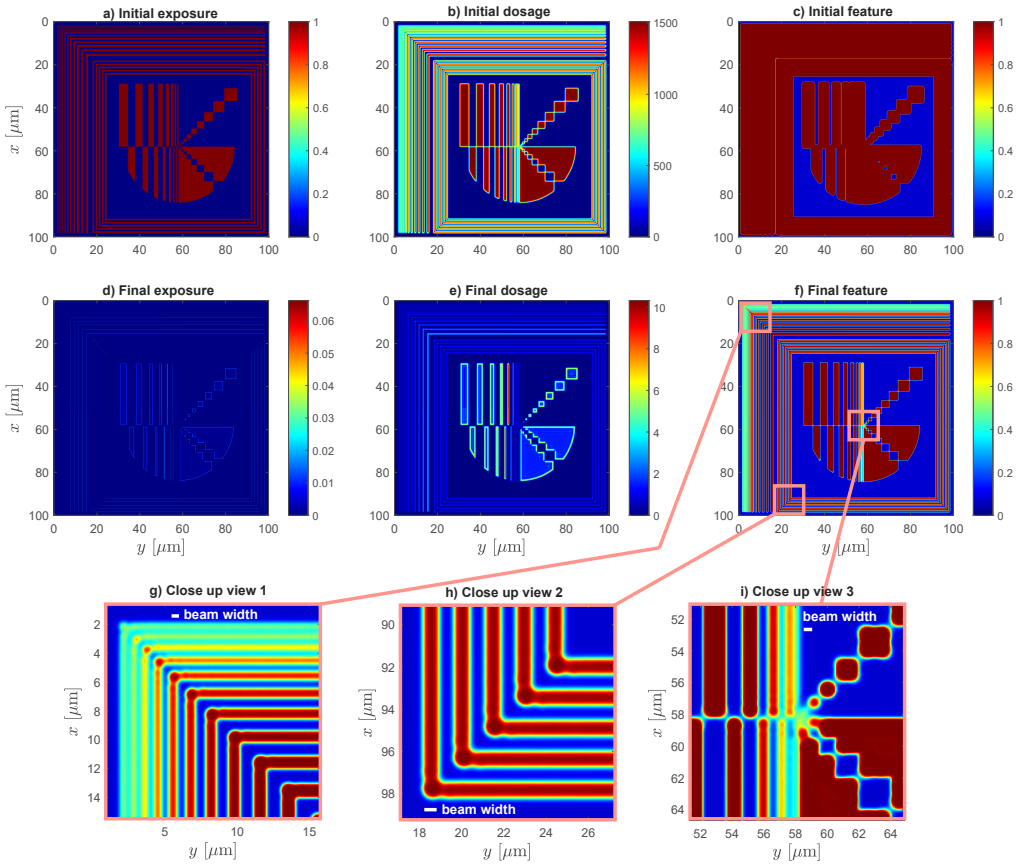


Fig. 5. Example application of the proposed method to a lithography test pattern. The top row shows the desired output feature (a) which is also used as the initial guess of the exposure pattern. The exposure (a) results in the dosage (b) and the resulting feature (c), which is grossly over exposed. The middle row shows the optimized exposure pattern (d) and the resulting dosage (e) and output feature (f). In the close up view of the corner features (g), the optimization result is observed to degrade predictably as the line width and spacing approach the beam-width, which is illustrated by a white bar. In the close up views (h) and (i), the simulated exposure is observed to produce parallel lines and geometric features with no visible artefacts and degrades predictably as the feature size approaches the beam-width.

Table 1. Optimization time and memory usage for the example in Fig. 5. When the area is split into 4 overlapping subdomains, the memory usage is reduced by 67% at the expense of a 27% increase in optimization time. The optimization was implemented in MATLAB on a Windows 10 PC with an i5-6300HQ processor and 32 GB RAM.

Subdomains	Subdomain Size	Optimization Time	Memory Usage
1×1	5486×5486	1446 s	25.6 GB
2×2	1372×1372	1843 s	8.3 GB

Optics EXPRESS

Algorithm 1 Exposure optimization

Input: Grid X, desired feature Z
Divide X into $n \geq 1$ subdomains $\{\Pi_i\}$
for $i = 1 \ldots n$ **do**
    Form Γ_i such that $\Pi_i \in \Gamma_i$
    Identify sparse entries in Γ_i and keep them constant 0
    Compute $\hat{\mathbf{w}}_*^{\Gamma_i} = \underset{\mathbf{w}}{\operatorname{argmin}} f(\mathbf{w})$ on Γ_i using L-BFGS-B
    Extract $\hat{\mathbf{w}}_*^{\Pi_i} \in \hat{\mathbf{w}}_*^{\Gamma_i}$
end for
Concatenate the solutions $\{\hat{\mathbf{w}}_*^{\Pi_i}\}$ to form $\hat{\mathbf{w}}_*$

(Fig. 5(g)), the optimization result can be observed to degrade predictably as the line width and spacing approach the beam width, which is illustrated by a white bar. In the close up views (Fig. 5(h) and 5(i)), the simulated exposure is observed to produce parallel lines and geometric features with predictable degradation as the feature size approaches the beam width.

In summary, the proposed subdomain method optimizes an exposure pattern with approximately 30 million variables subject to positivity constraints in 30 minutes on a modest PC. Memory requirements were reduced by nearly three-quarters compared to a direct solution using the same underlying method. This example demonstrates the utility of the proposed method for increasing the scale of scanning beam lithography optimization, or significantly decreasing the memory requirements.

8. Complexity and accuracy analysis

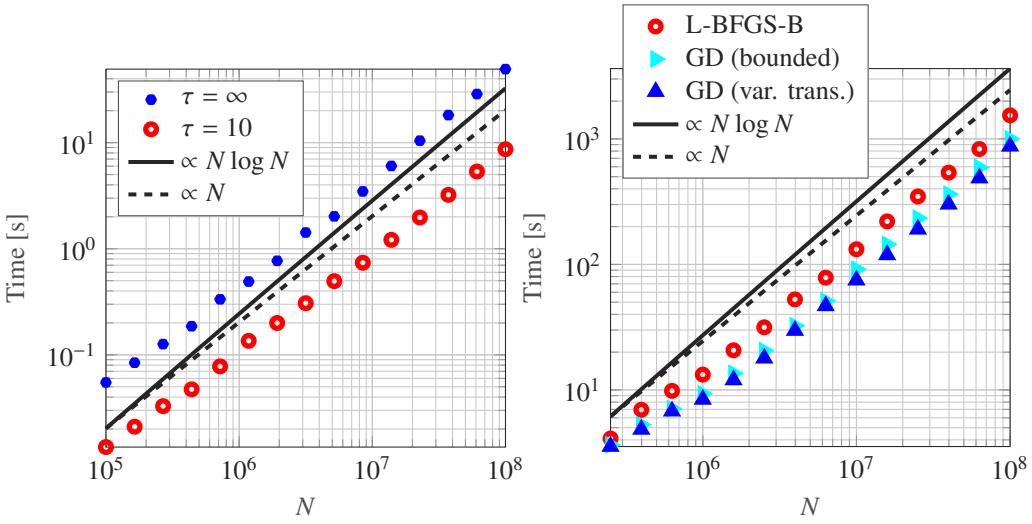
As the memory requirement is directly proportional to the grid size, it scales with $O(N)$. The complexity of the cost function and gradient computations are dictated by the convolutional operations, which are of order $O(N \log N)$ when using the FFT. The approximative FFT computations should improve this scaling, as the reduced convolutional limits in Eq. (10) are independent of N. As the L-BFGS-B routine has a computational cost of approximately $O(N)$ per iteration [63], the overall complexity is expected to be (at most) of order $O(N \log N)$.

To confirm the above estimates, an empirical study was performed on the example problem described in Section 7. The results of the analysis are plotted in Fig. 6. In Fig. 6(a), the computation time of the cost function and gradient is plotted against the number of variables, averaged over 10 trials. The plot includes results for exact FFT computations ($\tau = \infty$) and the approximate version with $\tau = 10$. Reference lines corresponding to scaling rates of $O(N \log N)$ and $O(N)$ are also plotted. The results show that computation time scales with $O(N)$ rather than $O(N \log N)$, especially when $\tau = 10$. Furthermore, the approximation reduces the computation time with a factor of 4-6 on this interval (increasing with N).

The time complexity of the complete optimization procedure (here comprising 50 iterations) is plotted against N in Fig. 6(b). The results also show a scaling of rate $O(N)$, which confirms that the L-BFGS-B routine does not degrade the overall complexity. In addition to L-BFGS-B with memory limit $m = 5$, this plot includes two versions of gradient descent (GD) obtained by setting $m = 1$: one using the same bound constraint treatment as in L-BFGS-B, and the other using the variable transformation $w_{ij} = e^{p_{ij}}$ mentioned in Section 6. It is seen that the time consumption per iteration is lower in these cases (roughly 40% reduction).

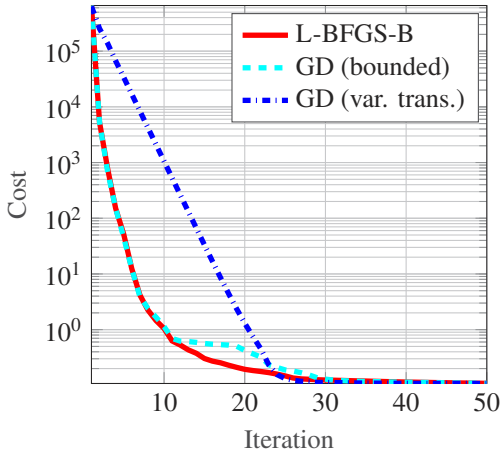
Also, Fig. 6(c) shows the evaluation of the cost function for the largest number of variables ($N = 10^8$) used in the experiments of Fig. 6(b). The bound constrained versions converge faster initially. However, the results are very similar for the memory limits 5 and 1 – as noted in [64], the optimal choice of memory limit is highly problem dependent.

Optics EXPRESS



(a) Cost function and gradient computation.

(b) Complete optimization (50 iterations).



(c) Cost evaluation ($N = 10^8$).

Fig. 6. Empirical analysis of time complexity and cost evaluation.

Optics EXPRESS

A study was also performed on the effect of subdomain division on the accuracy of results and the computation time. Figure 7 plots RMS difference between the desired and predicted feature, and the computation time for different subdomain divisions. The results are reported relative to the full grid solution with exact FFT computations (1×1, $\tau = \infty$). It is seen that the accuracy is not affected by the FFT approximation, or by the subdomain division (maximum error less than 0.02%). Although there is a computational overhead from the subdomain division, it is small in comparison to the time we gain from the approximation. These studies were conducted on a Scientific Linux 6.10 server with an AMD Opteron (Bulldozer) 6282SE processor and 128 GB RAM.

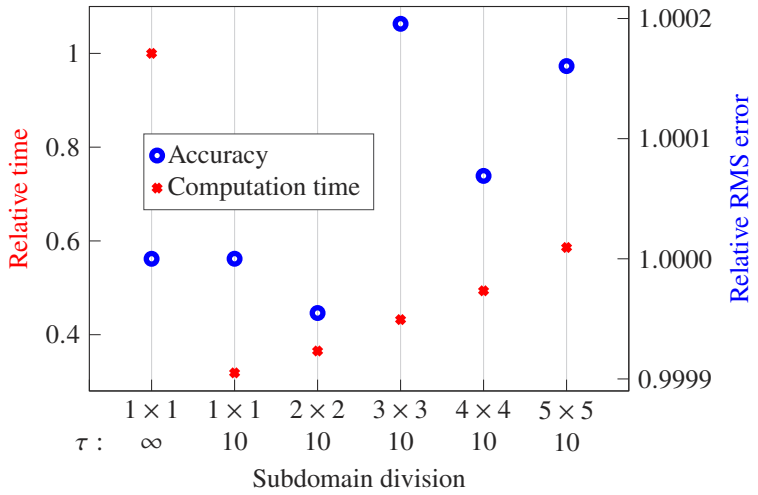


Fig. 7. The impact of subdomain division on accuracy and computation time. The results reported are relative to the full grid with exact FFT computations (1×1, $\tau = \infty$).

9. Conclusion

The article extends previous work on scanning beam lithography processes by proposing a memory efficient optimization procedure. The benefits of reduced memory requirements include lower minimum hardware specifications, or direct solution of larger scale problems, or a finer pixel resolution.

Instead of optimizing for the full grid directly, the problem is divided into smaller subdomains that are partly overlapping to avoid inaccurate boundary effects. The sub-problems are solved one at a time using the L-BFGS-B gradient search algorithm, and are then concatenated into a global solution. Furthermore, efficient computations are obtained through local application of the fast Fourier transform and utilization of natural sparsity. The overlaps between the subdomains imply a computational overhead resulting in a longer run-time; the upside is that the memory requirements are reduced by a factor almost equal to the number of subdomains, which opens up for problem sizes much larger than otherwise possible.

Empirical experiments demonstrate the performance of the proposed method; for instance, by splitting a problem with 30 million variables into 4 subdomains, it is solved with 67% less memory and a time increase of merely 27%. This trade-off between reduced memory and longer computation time is expected to be desirable in large-scale applications that tend to be memory limited. Moreover, it is shown that the time complexity scales linearly with the problem size and that the subdomain division has a negligible impact on the accuracy of the solution.

Optics EXPRESS

This work focuses on applications in scanning laser lithography. Future work will consider modified process models and cost functions that are suited to other point-wise lithography and direct fabrication processes such as scanning electron beam lithography, focused ion beam deposition, and plasma etching.

Funding. Kjell och Märta Beijers Stiftelse; Stiftelsen för Strategisk Forskning (RIT15-0012); Australian Research Council (DP210103383).

Disclosures. The authors declare no conflicts of interest.

Data availability. No data were generated or analyzed in the presented research. Source code for reproduction of experimental results is available online [68].

References

1. H. J. Levinson and T. A. Brunner, "Current challenges and opportunities for EUV lithography," in *International Conference on Extreme Ultraviolet Lithography 2018*, vol. 10809 (SPIE, 2018), pp. 5–11.
2. M. Van de Kerkhof, J. Benschop, and V. Banine, "Lithography for now and the future," Solid-State Electron. **155**, 20–26 (2019).
3. S. Srikanth, S. Dudala, S. Raut, S. K. Dubey, I. Ishii, A. Javed, and S. Goel, "Optimization and characterization of direct UV laser writing system for microscale applications," J. Micromech. Microeng. **30**(9), 095003 (2020).
4. F. Peng, Z. Yang, and Y. Song, "3D grayscale lithography based on exposure optimization," in *International Workshop on Advanced Patterning Solutions (IWAPS)*, (IEEE, 2021), pp. 1–3.
5. J. N. Randall, J. H. Owen, J. Lake, and E. Fuchs, "Next generation of extreme-resolution electron beam lithography," J. Vac. Sci. Technol., B: Nanotechnol. Microelectron.: Mater., Process., Meas., Phenom. **37**(6), 061605 (2019).
6. V. R. Manfrinato, F. E. Camino, A. Stein, L. Zhang, M. Lu, E. A. Stach, and C. T. Black, "Patterning Si at the 1 nm length scale with aberration-corrected electron-beam lithography: tuning of plasmonic properties by design," Adv. Funct. Mater. **29**(52), 1903429 (2019).
7. A. Joshi-Imre and S. Bauerdick, "Direct-Write Ion Beam Lithography," J. Nanotechnol. **2014**, 1–26 (2014).
8. M. Horak, K. Bukvisova, V. Svarc, J. Jaskowiec, V. Krapek, and T. Sikola, "Comparative study of plasmonic antennas fabricated by electron beam and focused ion beam lithography," Sci. Rep. **8**(1), 9640 (2018).
9. A. Cattoni, D. Mailly, O. Dalstein, M. Faustini, G. Seniutinas, B. Rösner, and C. David, "Sub-10 nm electron and helium ion beam lithography using a recently developed alumina resist," Microelectron. Eng. **193**, 18–22 (2018).
10. T. Trantidou, M. S. Friddin, K. B. Gan, L. Han, G. Bolognesi, N. J. Brooks, and O. Ces, "Mask-free laser lithography for rapid and low-cost microfluidic device fabrication," Anal. Chem. **90**(23), 13915–13921 (2018).
11. A. Melnikov, S. Köble, S. Schweiger, Y. K. Chiang, S. Marburg, and D. A. Powell, "Microacoustic metagratings at ultra-high frequencies fabricated by two-photon lithography," arXiv preprint arXiv:2202.03490 (2022).
12. S. Achenbach, S. Hengsbach, J. Schulz, and J. Mohr, "Optimization of laser writer-based UV lithography with high magnification optics to pattern X-ray lithography mask templates," Microsyst. Technol. **25**(8), 2975–2983 (2019).
13. K. Keskinbora, U. T. Sanli, M. Baluktsian, C. Grévent, M. Weigand, and G. Schütz, "High-throughput synthesis of modified Fresnel zone plate arrays via ion beam lithography," Beilstein J. Nanotechnol. **9**, 2049–2056 (2018).
14. K. Xu, J. Qin, and L. Wang, "Sub-micrometer direct laser writing using an optimized binary-amplitude zone plate lens," Opt. Lett. **46**(20), 5185–5188 (2021).
15. R. Menon, D. Gil, G. Barbastathis, and H. I. Smith, "Photon-sieve lithography," J. Opt. Soc. Am. A **22**(2), 342–345 (2005).
16. M. N. Julian, D. G. MacDonnell, and M. C. Gupta, "High-efficiency flexible multilevel photon sieves by single-step laser-based fabrication and optical analysis," Appl. Opt. **58**(1), 109–114 (2019).
17. R. Garcia, A. W. Knoll, and E. Riedo, "Advanced scanning probe lithography," Nat. Nanotechnol. **9**(8), 577–587 (2014).
18. L. Pan, Y. Park, Y. Xiong, E. Ulin-Avila, Y. Wang, L. Zeng, S. Xiong, J. Rho, C. Sun, D. B. Bogy, and X. Zhang, "Maskless plasmonic lithography at 22 nm resolution," Sci. Rep. **1**(1), 175 (2011).
19. 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," Nat. Commun. **4**(1), 2103 (2013).
20. 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," Chem. Sci. **5**(5), 2023–2030 (2014).
21. L. R. McCourt, M. G. Ruppert, B. S. Routley, S. C. Indirathankam, and A. F. Fleming, "A comparison of gold and silver nanocones and geometry optimisation for tip-enhanced microscopy," J. Raman Spectrosc. **51**(11), 2208–2216 (2020).
22. Y. Hu and Y. Meng, "Numerical modeling and analysis of plasmonic flying head for rotary near-field lithography technology," Friction **6**(4), 443–456 (2018).
23. 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 Lett. **13**(9), 4485–4491 (2013).

Optics EXPRESS

24. S. W. Tang, M. H. Uddin, W. Y. Tong, P. Pasic, W. Yuen, H. Thissen, Y. W. Lam, and N. H. Voelcker, "Replication of a tissue microenvironment by thermal scanning probe lithography," ACS Appl. Mater. Interfaces **11**(21), 18988–18994 (2019).
25. H. T. Soh, K. W. Guarini, and C. F. Quate, *Scanning probe lithography*, vol. 7 (Springer Science & Business Media, 2013).
26. Y. Yan, Y. He, G. Xiao, Y. Geng, and M. Ren, "Effects of diamond tip orientation on the dynamic ploughing lithography of single crystal copper," Precis. Eng. **57**, 127–136 (2019).
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 Trans. Comput.-Aided Des. Integr. Circuits Syst. **31**(2), 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 Electron. **54**(10), 1211–1215 (2010).
30. B. D. Cook and S.-Y. Lee, "PYRAMID-a hierarchical, rule-based approach toward proximity effect correction. II. Correction," IEEE Trans. Semicond. Manufact. **11**(1), 117–128 (1998).
31. S.-Y. Lee and B. D. Cook, "PYRAMID-a hierarchical, rule-based approach toward proximity effect correction. I. Exposure estimation," IEEE Trans. Semicond. Manufact. **11**(1), 108–116 (1998).
32. 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," Microelectron. Eng. **87**(5-8), 1041–1043 (2010).
33. 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," Microelectron. Eng. **88**(8), 1910–1912 (2011).
34. L. E. Ocola, D. J. Gosztola, D. Rosenmann, and G. Lopez, "Automated geometry assisted proximity effect correction for electron beam direct write nanolithography," J. Vac. Sci. Technol., B: Nanotechnol. Microelectron.: Mater., Process., Meas., Phenom. **33**(6), 06FD02 (2015).
35. O. T. Ghalehbeygi, A. G. Wills, B. S. Routley, and A. J. Fleming, "Gradient-based optimization for efficient exposure planning in maskless lithography," J. Micro/Nanolithogr., MEMS, MOEMS **16**(03), 1 (2017).
36. A. J. Fleming, O. T. Ghalehbeygi, B. S. Routley, and A. G. Wills, "Scanning laser lithography with constrained quadratic exposure optimization," IEEE Trans. Contr. Syst. Technol. **27**(5), 2221–2228 (2019).
37. O. T. Ghalehbeygi, J. O'Connor, B. S. Routley, and A. J. Fleming, "Iterative deconvolution for exposure planning in scanning laser lithography," in *American Control Conference*, (Milwaukee, WI, 2018).
38. O. T. Ghalehbeygi, "Exposure planning for scanning laser lithography," Ph.D. thesis, University of Newcastle, Newcastle, Australia (2018).
39. Y. Ma, W. Zhong, S. Hu, J.-R. Gao, J. Kuang, J. Miao, and B. Yu, "A unified framework for simultaneous layout decomposition and mask optimization," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **39**(12), 5069–5082 (2020).
40. E. van Setten, K. Rook, H. Mesilhy, G. Bottiglieri, F. Timmermans, M. Lee, A. Erdmann, and T. Brunner, "Multilayer optimization for high-NA EUV mask3D suppression," in *Extreme Ultraviolet Lithography 2020*, vol. 11517 (International Society for Optics and Photonics, 2020), p. 115170Y.
41. Y. Shen, F. Peng, X. Huang, and Z. Zhang, "Adaptive gradient-based source and mask co-optimization with process awareness," Chin. Opt. Lett. **17**(12), 121102 (2019).
42. H. Mesilhy, P. Evanschitzky, G. Bottiglieri, E. Van Setten, T. Fliervoet, and A. Erdmann, "Pathfinding the perfect EUV mask: the role of the multilayer," in *Extreme Ultraviolet (EUV) Lithography XI*, vol. 11323 (International Society for Optics and Photonics, 2020), p. 1132316.
43. X. Ma, C. Han, Y. Li, L. Dong, and G. R. Arce, "Pixelated source and mask optimization for immersion lithography," J. Opt. Soc. Am. A **30**(1), 112–123 (2013).
44. X. Liu, R. Howell, S. Hsu, K. Yang, K. Gronlund, F. Driessen, H.-Y. Liu, S. Hansen, K. van Ingen Schenau, T. Hollink, P. van Adrichem, K. Troost, J. Zimmermann, O. Schumann, C. Hennerkes, and P. Gräupner, "EUV source-mask optimization for 7nm node and beyond," in *Extreme Ultraviolet (EUV) Lithography V*, vol. 9048 International Society for Optics and Photonics (SPIE, 2014), pp. 171–181.
45. T. Li, Y. Sun, E. Li, N. Sheng, Y. Li, P. Wei, and Y. Liu, "Multi-objective lithographic source mask optimization to reduce the uneven impact of polarization aberration at full exposure field," Opt. Express **27**(11), 15604–15616 (2019).
46. X. Wu, S. Liu, J. Li, and E. Y. Lam, "Efficient source mask optimization with zernike polynomial functions for source representation," Opt. Express **22**(4), 3924–3937 (2014).
47. J. Li, S. Liu, and E. Y. Lam, "Efficient source and mask optimization with augmented Lagrangian methods in optical lithography," Opt. Express **21**(7), 8076–8090 (2013).
48. Y. Shen, "Lithographic source and mask optimization with narrow-band level-set method," Opt. Express **26**(8), 10065–10078 (2018).
49. Y. Shen, F. Peng, and Z. Zhang, "Semi-implicit level set formulation for lithographic source and mask optimization," Opt. Express **27**(21), 29659–29668 (2019).

Optics EXPRESS

50. X. Ma, Q. Zhao, H. Zhang, Z. Wang, and G. R. Arce, "Model-driven convolution neural network for inverse lithography," Opt. Express **26**(25), 32565–32584 (2018).
51. S. Lan, J. Liu, Y. Wang, K. Zhao, and J. Li, "Deep learning assisted fast mask optimization," in *Optical Microlithography XXXI*, vol. 10587 (International Society for Optics and Photonics, 2018), p. 105870H.
52. H. Yang, W. Zhong, Y. Ma, H. Geng, R. Chen, W. Chen, and B. Yu, "VLSI mask optimization: From shallow to deep learning," Integration **77**, 96–103 (2021).
53. A. Gramacki and J. Gramacki, "FFT-based fast computation of multivariate kernel density estimators with unconstrained bandwidth matrices," J. Comput. Graph. Stat. **26**(2), 459–462 (2017).
54. T. F. Chan and T. P. Mathew, "Domain decomposition algorithms," Acta Numer. **3**, 61–143 (1994).
55. J. Xu, "Iterative methods by space decomposition and subspace correction," SIAM Rev. **34**(4), 581–613 (1992).
56. J. Galtier and S. Lanteri, "On overlapping partitions," in *Proceedings of the 2000 International Conference on Parallel Processing*, (IEEE Computer Society, USA, 2000), p. 461.
57. A. Langer and F. Gaspoz, "Overlapping domain decomposition methods for total variation denoising," SIAM J. Numer. Anal. **57**(3), 1411–1444 (2019).
58. C. Carstensen, "Domain decomposition for a non-smooth convex minimization problem and its application to plasticity," Numer. Linear Algebra Appl. **4**(3), 177–190 (1997).
59. M. Guizar-Sicairos, I. Johnson, A. Diaz, M. Holler, P. Karvinen, H.-C. Stadler, R. Dinapoli, O. Bunk, and A. Menzel, "High-throughput ptychography using Eiger: scanning X-ray nano-imaging of extended regions," Opt. Express **22**(12), 14859–14870 (2014).
60. Y. S. G. Nashed, D. J. Vine, T. Peterka, J. Deng, R. Ross, and C. Jacobsen, "Parallel ptychographic reconstruction," Opt. Express **22**(26), 32082–32097 (2014).
61. S. Marchesini, H. Krishnan, B. J. Daurer, D. A. Shapiro, T. Perciano, J. A. Sethian, and F. R. N. C. Maia, "*SHARP*: a distributed GPU-based ptychographic solver," J. Appl. Crystallogr. **49**(4), 1245–1252 (2016).
62. J. Shi, P. Wang, and K. Cai, "Subdomain method for layout optimization of piles in a composite pile foundation," Current Trends in Civil & Structural Engineering **4**, 000576 (2019).
63. R. Byrd, P. Lu, and J. Nocedal, "A limited memory algorithm for bound constrained optimization," SIAM J. Sci. Comput. **16**(5), 1190–1208 (1995).
64. J. Nocedal and S. J. Wright, *Numerical Optimization, Springer Series in Operations Research and Financial Engineering* (Springer, New York, 2000).
65. J. Nocedal, "Updating quasi-newton matrices with limited storage," Math. Comp. **35**(151), 773–782 (1980).
66. S. Becker, "L-BFGS-B-C," https://github.com/stephenbeckr/L-BFGS-B-C. Accessed: 2018-06-01.
67. C. Zhu, R. H. Byrd, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," ACM Trans. Math. Softw. **23**(4), 550–560 (1997).
68. C. Jidling, "Source code for 'Memory efficient constrained optimization of scanning-beam lithography'," https://github.com/carji475/scanning-beam-lithography (2022).