Gradient-based optimization for efficient exposure planning in maskless lithography

Omid Tayefeh Ghalehbeygi,* Adrian G. Wills, Ben S. Routley, and Andrew J. Fleming
The University of Newcastle, Faculty of Engineering and Built Environment, Department of Electrical and Engineering and Computing, Callaghan, Newcastle, Australia

Abstract. Scanning laser lithography is a maskless method for exposing photoresist during semiconductor manufacturing. In this method, the energy of a focused beam is controlled while scanning the beam or substrate. With a positive photoresist material, areas that receive an exposure dosage over the threshold energy are dissolved during development. The surface dosage is related to the exposure profile by a convolution and nonlinear function, so the optimal exposure profile is nontrivial. A gradient-based optimization method for determining an optimal exposure profile, given the desired pattern and models of the beam profile and photochemistry, is described. This approach is more numerically efficient than optimal barrier-function-based methods but provides near-identical results. This is demonstrated through simulation and experimental lithography. [DOI: 10.1117/1.JMM.16.3.033507]

Keywords: maskless lithography; optimization; gradient-descent method; laser scanning lithography.

Paper 17079 received May 22, 2017; accepted for publication Aug. 9, 2017; published online Sep. 8, 2017.

1 Introduction

Lithography is a critical process in semiconductor manufacturing and is responsible for half of the production cost.[1] To address the high cost of mask-set production, maskless lithography methods have been developed for low-volume and prototyping applications.[2] These methods include electron beam lithography[3] and ion beam lithography.[4] Electron beam methods can provide a high resolution but suffer from charging and proximity effects. Ion beam methods are still in an early stage of development and have yet to reach the resolution of electron beam methods. Both of these techniques require specialized resist chemistry and suffer from a high cost of equipment maintenance.[3–6]

Scanning laser lithography is a cost-effective maskless method, which has the advantage of using standard photoresist and process chemistry. In this method, a focused laser is scanned over the photoresist while modulating the beam power.

The foremost difficulties associated with scanning laser lithography are the resolution, throughput, and complicated exposure planning. However, the resolution has recently been improved to 30 nm[7] by using optimized near-field probes.[8] The speed of nanopositioning systems has also increased to allow 1-kHz scan rates[9] and the exposure of thousands of features per second.

The problem of optimal exposure planning is common to all of the maskless lithography approaches that involve a scanning beam. The exposure pattern defines the position and intensity of the laser, electron, or ion beam. Without a modified exposure pattern, the resultant image is a nonlinear low-pass filtered version of the desired feature, due to the finite resolution of the beam profile. Rule- and model-based methods have been proposed for optimizing the exposure profile. Rule-based methods use a predistortion pattern derived from previous simulation and experimental outcomes.[10–13] The efficacy of these methods becomes doubtful in large-scale application, which requires a more robust solution. On the other hand, model-based approaches seek a solution by attempting to invert a model of the lithography process.

In the model-based approach, an optimization problem is defined that minimizes the predicted error in a developed feature. This optimization is complicated by the nonlinear behavior of the photoresist, the dosage constraints, and the large problem size, which may range from 10^3 to $>10^9$ variables.

The barrier method, introduced in Ref. 14, employs a nonlinear programming approach to find an optimal exposure pattern. This method is computationally demanding and becomes impractical beyond 10,000 variables. The computational requirements are primarily the calculation of first and second derivatives of the cost function. Thus, an alternative method with less computational time and complexity is necessary.

1.1 *Motivation and Contribution*

The problem of mask optimization in standard projection lithography has received significant attention in the literature. The first method for synthesizing a binary mask was published in 1981 using the so-called iterative altering projection method.[15] Many other optimization algorithms were proposed over the following years, including simulated annealing,[16] mixed integer programming,[17] random pixel flipping,[18] genetic algorithms,[19] and the Lagrangian method,[20] but none of these methods are suited to large-scale problems. However, in the last decade, inverse lithography methods have become popular for dealing with large-scale problems. In this technique, nonlinear programming,[21] level-set,[22] or gradient-based[23] methods are employed. The

*Address all correspondence to: Omid Tayefeh Ghalehbeygi, E-mail: omid.tayefehghalehbeygi@uon.edu.au

1932-5150/2017/$25.00 © 2017 SPIE

gradient-based methods have a lower computational cost compared to the other methods.

In this article, the gradient-based method from mask-based lithography is adapted for exposure optimization in scanning beam lithography. In previous work, the simulation of simple features in a workspace with 10^6 variables demonstrated how quickly a solution can be achieved for a large problem.[24] In this article, experimental results for complex features are reported. The gradient-based method requires less memory and provides a faster solution than other techniques.[14,17,21] In the following sections, the process flow and model are described, followed by a description of the optimization method and experimental results.

2 Experimental Setup and Process Flow

As shown in Fig. 1, the exposure optics is 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 $40\times/0.75$ objective lens. The focused beam is then directed at the sample that 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 Fig. 2, the laser source is modulated by an acoustic optical modulator that 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 1 min, which resulted in a film thickness of ~700 nm. After the coating step, the photoresist was baked at 90°C for 1 min to improve the substrate adhesion and minimize dark erosion during development. After the exposure process, the sample is immersed in AZ-726MIF developer for 1 min removing the exposed pattern.

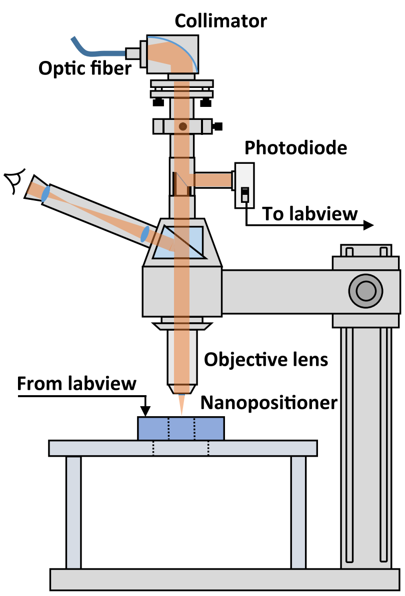

Fig. 1 The optical exposure system that focuses light from the fiber onto the scanning platform.

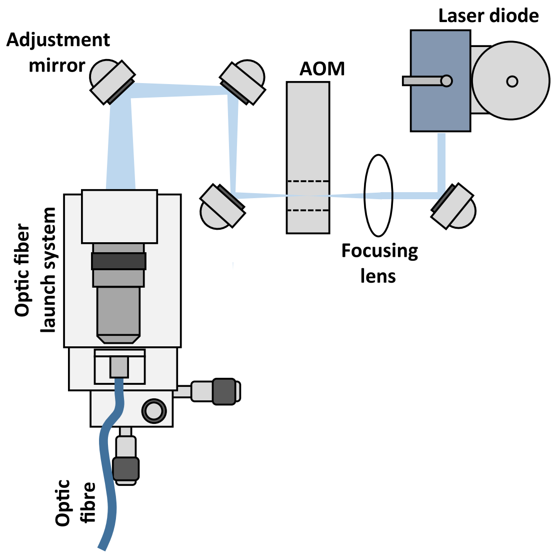

Fig. 2 Schematic of the laser source, modulation system, and fiber coupling.

Finally, the sample is rinsed in distilled water and dried using nitrogen gas.

3 Process Modeling

This section develops a model for the lithography process described in Sec. 2. The model assumes that the photoresist layer 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. Moreover, all functions for the exposure pattern, beam profile, and dosage are defined as matrices, which represent these functions at discrete locations in a workspace. The workspace is modeled in a $N\times N$ square matrix discretized into N pixels along x and y axes

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

where Δ is the grid resolution.

3.1 *Beam Profile*

The light intensity (in W/m^2) at the focal point of the objective lens can be analytically expressed as[25]

$$B(x,y) = \frac{2P}{\pi w_0^2} e^{-\frac{2(x^2+y^2)}{w_0^2}}, \tag{2}$$

where x and y represent the transverse axes of the beam at focal point w_0, and P is the power.

To replace a convolution operation by a matrix multiplication, it is convenient to define an $N\times N$ array of $N\times N$ matrices B, where each matrix in the array $B^{k,l}$ represents the spatial beam intensity with a focal point centered at (x_k, y_l), and $B^{k,l}_{i,j}$ represents the intensity at the location (x_i, y_j), that is

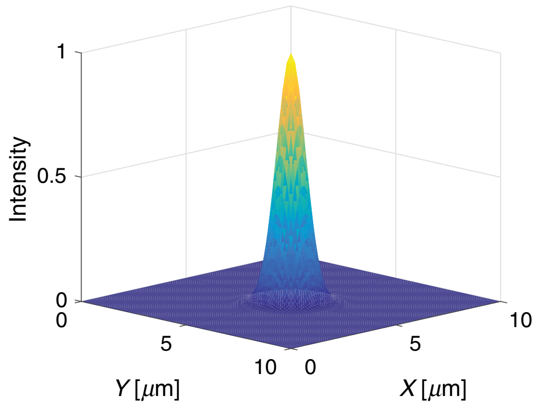

Fig. 3 The normalized intensity of a Gaussian beam, where the focal point is located in the middle of the workspace ($x_k = 5\ \mu$m and $y_l = 5\ \mu$m). The beam width is $w_0 = 450$ nm.

$$B_{i,j}^{k,l} = \frac{2P}{\pi w_0^2}\ \exp\left[-\frac{2(x_i - x_k)^2 + (y_j - y_l)^2}{w_0^2}\right]$$

$$\text{for } i, j = 1, \ldots, N, \quad \text{and,} \quad k, l = 1, \ldots, N. \tag{3}$$

This normalized intensity is shown in Fig. 3, where the focal point is located in the center of the workspace (B_ϕ).

3.2 *Exposure Modeling*

The exposure and dosage matrices are defined by E and $D \in \mathbb{R}^{N\times N}$, respectively. That is, the element $E_{k,l}$ represents the exposure energy at location (x_k, y_l), where $E_{k,l}$ refers to the k'th row and l'th column. The dosage D represents the cumulative energy per unit area, calculated by summing the dosage from each exposure. Since the dosage from each exposure is the product of energy and the associated beam profile $B^{k,l}$. The dosage matrix D is

$$D = \sum_{k=1}^{N}\sum_{l=1}^{N} E_{k,l}B^{k,l}. \tag{4}$$

Figure 4 shows the matrix D, when E contains a single nonzero entry at (x_k, y_l).

3.3 *Photoresist Modeling*

The photoresist model describes the fraction of chemical conversion as a function of dosage. The photoresist effects can be precisely modeled by Dill's model,[26] Mack's model,[27] or other models.[28,29] In large optimization problems, computation time is an important consideration that can be reduced by using a simplified model, for example, the variable threshold resist model[30] and the constant threshold resist model.[31] The latter model indicates 100% conversion when the pixel's dose is above a threshold. That is

$$\hat{Z}_{i,j} = \begin{cases} 1 & D_{i,j} \geq T \\ 0 & D_{i,j} \end{cases}, \quad \text{for } i, j = 1, \ldots, N, \tag{5}$$

where $\widehat{Z}$ is the predicted exposure result and T is the threshold.

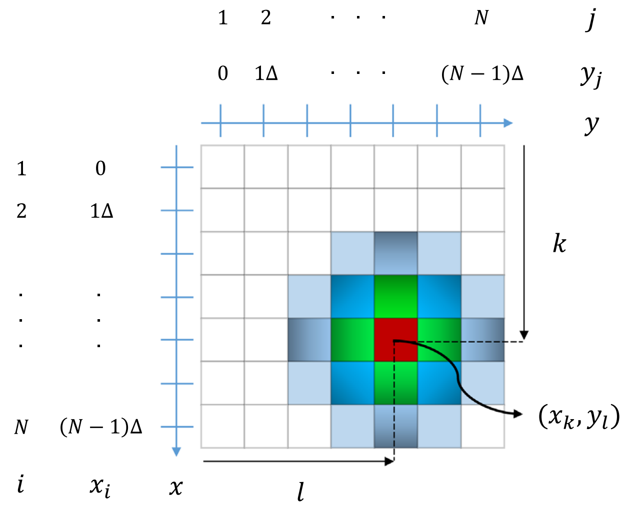

Fig. 4 The resulting D matrix which is the product of multiplying the only nonzero array of the E matrix, $E_{k,l}$ as a scalar, by the coordinated beam matrix $B^{k,l}$.

In practice, the photoresist conversion is a continuous function of dosage. The sigmoid function is an approximation of this process

$$\hat{Z}_{i,j} = f(D_{i,j}) = \frac{1}{1 + e^{-\alpha(D_{i,j} - D_{50\%})}} \quad \text{for } i, j = 1, \ldots, N, \tag{6}$$

where the parameter α dictates the steepness of the sigmoid and $D_{50\%}$ is the dosage where half of the photoresist is converted. When α is large, the sigmoid approaches the threshold model. Figure 5 shows the behavior of the sigmoid function with different values of α.

By combining the exposure and photoresist model, the resulting feature can be predicted from an arbitrary exposure pattern. This model is schematically shown in Fig. 6.

3.4 *Resolution Limit*

The resolution of both maskless and traditional optical lithography is fundamentally limited by diffraction. The minimum resolution is described as[32]

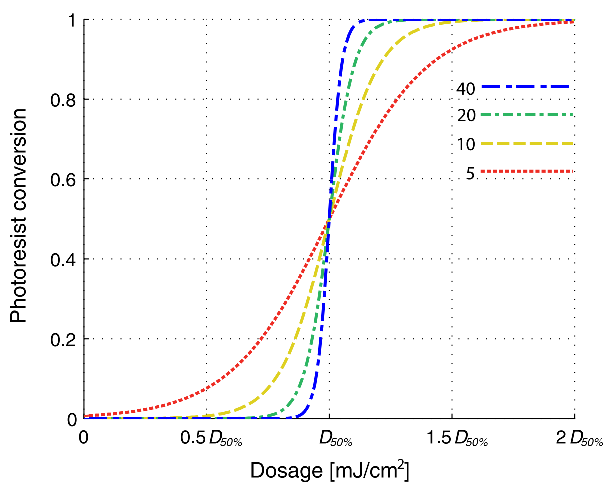

Fig. 5 Sigmoid function for different steepness parameter varying $\alpha = 5$, 10, 20, 40 with the same $D_{50\%}$.

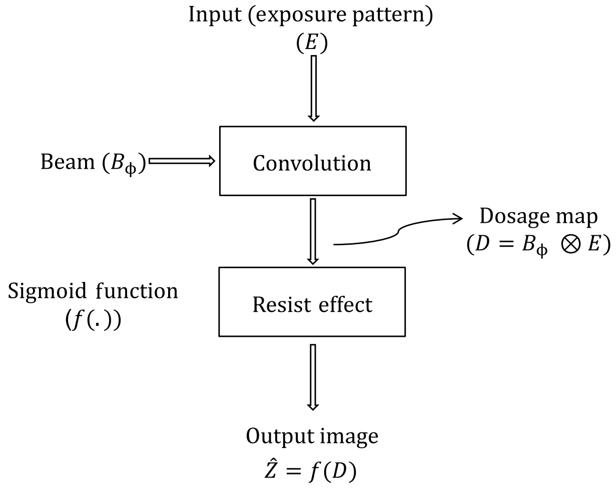

Fig. 6 The forward process model, which maps an exposure pattern (E) onto the output features ($\hat{Z}$).

$$\text{Minimum resolution} = k_1 \frac{\lambda}{\text{NA}}, \tag{7}$$

where λ is the wavelength, NA is the numerical aperture of the objective lens, and k_1 is a process parameter, which is dependent on variables such as the photoresist thickness and dispersion. For mask-based lithography, with a phase-shifted mask and resolution enhancement, the value of k_1 is typically 0.35.[33] Similar values for k_1 have been reported for maskless optical lithography.[2]

4 Model-Based Optimization

There are several ways to formulate the optimization problem.[20,34] In this article, the cost function is the 2-norm of the difference between the desired and predicted output. That is

$$J_1(E) = \sum_{i=1}^{N} \sum_{j=1}^{N} (Z_{i,j} - \hat{Z}_{i,j})^2, \tag{8}$$

where $\hat{Z}$ represents the output image and Z indicates the desired image. In addition, a regularization term is required to suppress excessive dosage, which results in light scattering and reduced resolution. The regularization term is

$$J_2(E) = \gamma \sum_{i=1}^{N} \sum_{j=1}^{N} D_{i,j}^2, \tag{9}$$

where γ is a user-defined scalar. The overall cost function is

$$J(E) = J_1(E) + J_2(E). \tag{10}$$

Since the exposure power can only be positive, each element in E must be a nonnegative. Therefore, optimal exposure pattern E^* is then achieved by solving the following expression:

$$E^* = \arg \min_E J(E), \quad \text{s.t } E \geq 0. \tag{11}$$

4.1 *Parameter Transformation*

As described in Eq. (11), the feasible domain is nonnegative values. In addition, a dosage exceeding 1.5 times the threshold generally causes undesired results due to overexposure and scattering. Therefore, the maximum value of each exposure spot can be restricted to 2, that is

$$0 \leq E_{i,j} \leq 2 \quad \text{for } i, j = 1, \ldots, N. \tag{12}$$

The bounded-constraint problem generally adds more complexity; however, the problem can be converted to unconstrained by using the following transformation:[35]

$$E = 1 + \cos(\Theta), \tag{13}$$

$$\Theta = \arccos(E - 1), \tag{14}$$

where Θ is a fictitious matrix with unconstrained elements

$$\Theta_{i,j} \in \mathbb{R} \quad \text{for } i, j = 1, \ldots, N.$$

By replacing the parameters E with Θ, Eq. (8) becomes

$$J_1(\Theta) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(Z_{i,j} - \frac{1}{1 + e^{-\alpha\left(\sum_{k=1}^{N} \sum_{l=1}^{N} \{[1+\cos(\Theta_{k,l})] B_{i,j}^{k,l}\} - D_{50\%}\right)}} \right)^2, \tag{15}$$

and Eq. (9) becomes

$$J_2(\Theta) = \gamma \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\sum_{k=1}^{N} \sum_{l=1}^{N} \{[1 + \cos(\Theta_{k,l})] B_{i,j}^{k,l}\} \right)^2. \tag{16}$$

With the transformation of variables, the optimization problem becomes

$$J(\Theta) = J_1(\Theta) + J_2(\Theta), \tag{17}$$

$$\Theta^* = \arg \min_\Theta J(\Theta), \quad \text{s.t } \Theta \in \mathbb{R}, \tag{18}$$

where Θ^* represents the optimal exposure pattern.

4.2 *Problem Solution*

The optimization problem expressed in Eq. (18) is a nonlinear and importantly, nonconvex problem due to the thresholding function $f(\cdot)$ and the cosine function. However, the gradient-descent method can be applied directly as it does not require a second derivative. To apply this method, the gradient of the cost function needs to be calculated analytically.

4.2.1 *Gradient calculation*

Rather than vectorization of matrices to derive the first derivative, all matrices are kept in two-dimensional (2-D) form. Thus, the gradient of the cost function is

$$g \triangleq \nabla_\Theta J = \nabla_\Theta J_1 + \nabla_\Theta J_2. \tag{19}$$

The gradients of the two terms in the cost function are

$$\nabla_\Theta J_1 = \frac{\partial J_1}{\partial \Theta}, \tag{20}$$

and

$$\nabla_\Theta J_2 = \frac{\partial J_2}{\partial \Theta}. \tag{21}$$

Analytical expressions for the gradients are derived in the Appendix. It is shown that these expressions simplify to

$$\nabla_\Theta J_1 = 2\alpha\Big\{B \otimes \Big[(Z - \hat{Z})\odot\hat{Z}\odot(1 - \hat{Z})\Big]\Big\}\odot \sin(\Theta), \tag{22}$$

and

$$\nabla_\Theta J_2 = -2\gamma(B \otimes D)\odot \sin(\Theta), \tag{23}$$

where $\odot$ is the element-by-element multiplication operator and $\otimes$ is the convolution operator.

To expedite the calculation of the discrete convolution in Eqs. (22) and (23), the fast Fourier transform (FFT) is utilized. MATLAB's® built-in multithreaded FFT algorithm provides a significant improvement in numerical efficiency. By applying the Fourier transform to Eqs. (22) and (23)

$$\nabla_\Theta J_1 = 2\alpha\mathcal{F}^{-1}\{\mathcal{F}(B)\odot\mathcal{F}[(Z-\hat{Z})\odot\hat{Z}\odot(1-\hat{Z})]\}\odot\sin(\Theta), \tag{24}$$

and

$$\nabla_\Theta J_2 = -2\gamma\mathcal{F}^{-1}\{\mathcal{F}(B)\odot\mathcal{F}(E)\odot\mathcal{F}(B_\phi)\}\odot \sin(\Theta), \tag{25}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and its inverse, respectively.

Another parameter affecting the performance of the gradient-descent method is step size. For fixed values, the optimization problem has difficulty converging to a solution. Thus, this value is dynamically selected in each iteration based on the Barzilai–Borwein approximation described in Sec. 4.2.2.

4.2.2 *Step-size computation*

The Barzilai–Borwein method provides an approximate value for the Newton step size without a Hessian calculation. The step size β can be calculated by solving

$$\beta_k = \arg\min_\beta \frac{1}{2}\|\Delta x - \beta\Delta g(x)\|^2, \tag{26}$$

where $\Delta x = x_k - x_{k-1}$ and $\Delta g(x) = \nabla f(x_k) - \nabla f(x_{k-1})$. Differentiating Eq. (26) with respect to β and equating it to zero provides an approximation of β. That is

$$\frac{d}{d\beta}\frac{1}{2}\|\Delta x - \beta\Delta g(x)\|^2 = 0 \Rightarrow \quad \beta = \frac{\Delta g(x)^T\Delta x}{\Delta g(x)^T\Delta g(x)}. \tag{27}$$

This approximation significantly improves the convergence speed of gradient-based methods. To implement the above equation, the following replacements are required

$$\mathrm{vec}\{\Delta\Theta\}\leftarrow\Delta x, \quad \mathrm{vec}\{\Delta g(\Theta)\}\leftarrow\Delta g(x), \tag{28}$$

where the vec operator stacks the columns of a matrix into a long vector.

By calculating the cost function gradient and a dynamic step size into each iteration, the steepest descent algorithm (Algorithm 1) is outlined below

4.3 *Complexity Analysis Per Iteration*

For the algorithm under consideration, the convolution operation is the most time-consuming. Each iteration involves three 2-D convolutions of $N \times N$ matrices, which has a complexity of $O(N^4)$. Using the FFT function in MATLAB®, the overall complexity reduces to $O(N^2 \log N^2)$. Thus, for $N = 100$, the complexity is equal to 4×10^4 operations. The largest matrix stored in memory during each iteration is $N \times N$.

In the barrier method, the most time-consuming calculation involves replacing the convolution with a multiplication operation,[14] which leads to the creation of an $N^2 \times N^2$ matrix. As this matrix is multiplied by itself twice in each iteration, the overall complexity is $O(N^6)$. However, by using the FFT to compute this convolution, the complexity can be decreased to $O(N^4 \log N^4)$, which leads to a complexity of 8×10^8 when $N = 100$. This is four orders-of-magnitude more operations than the gradient algorithm proposed in this work. In addition, the memory required to construct an $N^2 \times N^2$ matrix is a significant problem.

Due to the significantly reduced complexity and memory requirements, the proposed algorithm is significantly more efficient, per iteration, than the log-barrier method described in Ref. 14. However, it should be noted that the algorithm in Ref. 14 utilizes a Hessian approximation that may result in fewer iterations before a termination condition is achieved. In practice, the most significant limitation is memory, not the number of iterations. In this regard, the proposed gradient method requires an $N \times N$ matrix to be constructed per iteration, which is $1/N^2$ fewer elements than the log-barrier method in Ref. 14.

Algorithm 1 Solve Eq. (18) using the steepest-descent method.

Require: $\gamma > 0$, and $\beta \in \mathbb{R}$

while Termination condition is not satisfied **do**

Compute the gradient vector $g \triangleq \nabla J(\Theta)$

Compute the step size by Barzilai–Borwein approximation β_k

Update $\Theta^{k+1}\leftarrow\Theta^k - |\beta_k|g^k$

end while

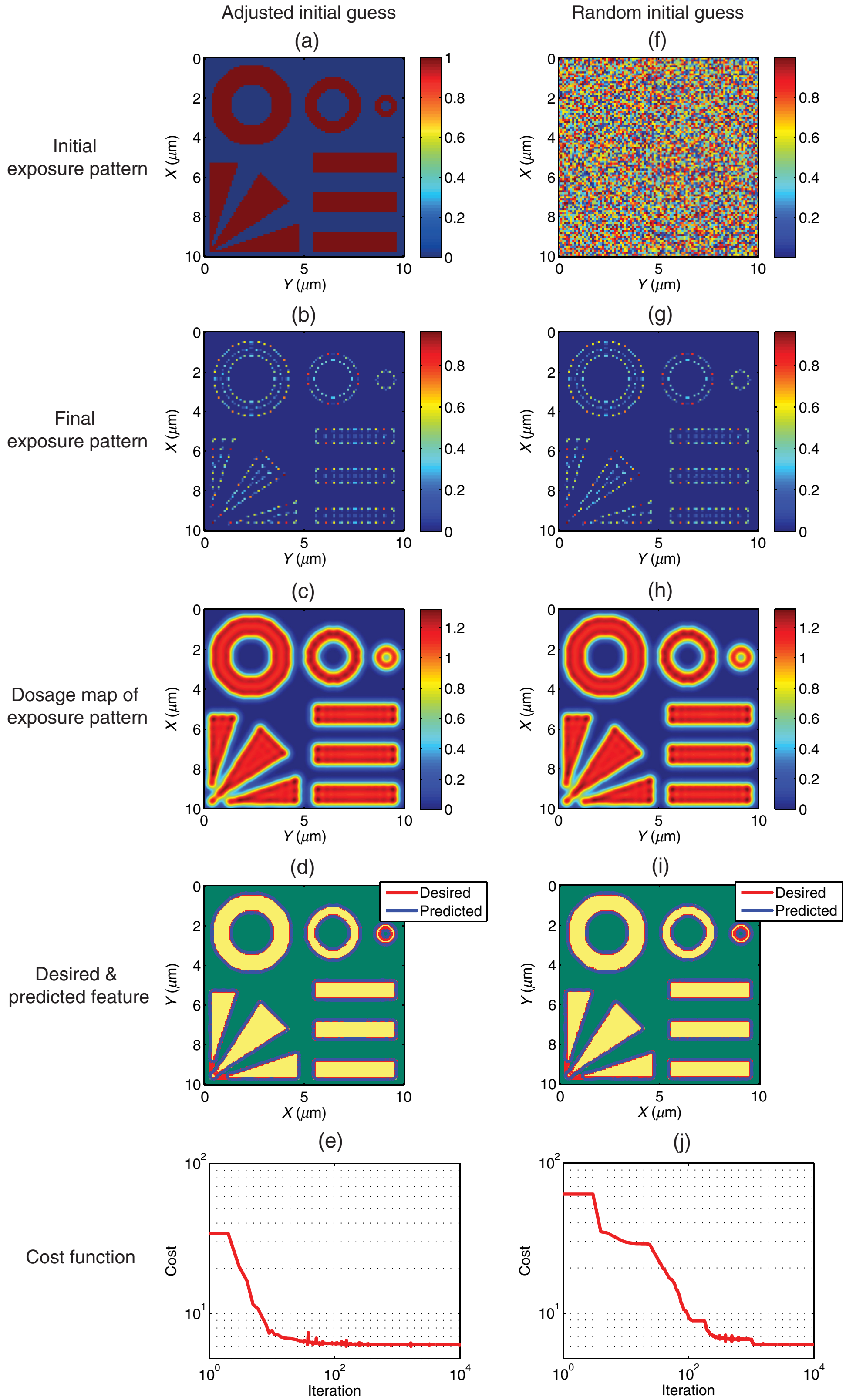

Fig. 7 A comparison of two optimization procedures where the left column starts with an educated guess and the right column starts with a random guess. In this example, the random initial guess requires a period of optimization and then follows a similar trajectory to the educated initial guess.

5 Simulation and Experimental Results

5.1 *Simulation Results*

The following simulations are based on a Gaussian 450-nm beam width, 100-nm grid resolution, and a workspace of $10\ \mu\text{m} \times 10\ \mu\text{m}$, which results in a 100×100 array. The user-defined constants γ and α are 0.3 and 5, respectively.

The test pattern for evaluating the proposed algorithm is shown in Fig. 7(a). This pattern has a variety of features, which examines the resolution for parallel, circular, and irregularly geometries. The final exposure pattern after 10^4 iterations is shown in Fig. 7(b). This comparison shows a strong correlation between the desired and predicted features. However, where the features become significantly smaller than the beam width, the contrast becomes poor. For example, the sharp ends of triangles are not resolved.

In the example considered, the final solution was independent of the initial guess, as observed by comparing the left and right column of Fig. 7.

5.2 *Experimental Results*

The first step is to determine the threshold value of the photoresist. Predicting this value from specifications is inaccurate as the threshold dosage is related to the thickness of the film.[36] Based on the specifications, the photoresist is fully exposed by a dosage of $40\ \text{mJ/cm}^2$. When $\alpha = 5$, this is equivalent to $D_{50\%} = 20\ \text{mJ/cm}^2$. The exact value is found empirically by performing a series of experimental exposures with $D_{50\%}$ ranging from 15 to $25\ \text{mJ/cm}^2$. Incidentally, the experimentally determined value for $D_{50\%}$ was $20\ \text{mJ/cm}^2$, which is identical to that predicted from the specifications.

The exposure and development process proceeded according to Sec. 2. An electron micrograph of the developed features is shown in Fig. 8. A close correlation between the predicted and experimental features can be observed. Some overexposure can be observed in the horizontal rectangles due to proximity effects, which have led to a microloading effect after developing the resist. This issue was alleviated to some extent by employing a regularization term in the cost function, which minimizes the total exposure energy and indirectly reduces scattering and proximity effects.

The background exposure dosage is thought to be the most significant source of error in the experimental results. Since the laser on-time is microseconds and the nanopositioner settling time is milliseconds, a large portion of the total time is consumed waiting for the nanopositioner to settle with the laser off. While the laser is assumed to be off, the beam power is nonzero due to the finite contrast ratio of the acousto-optic modulator. In the experimental setup, the measured contrast ratio was 1∶6000. Although this background exposure energy could potentially be modeled, it is simpler to improve the contrast ratio by optical means, which is a topic of current research.

Other potential error sources include a nonideal beam profile and varying photoresist thickness. Despite the use of a single-mode fiber, the beam profile may not be perfectly Gaussian. To eliminate this potential error source in the future, a 100-nm pin-hole will be used with an XY scanner to experimentally identify the beam profile. If this varies significantly from that theoretically predicted, then the experimental data could be used directly in the optimization.

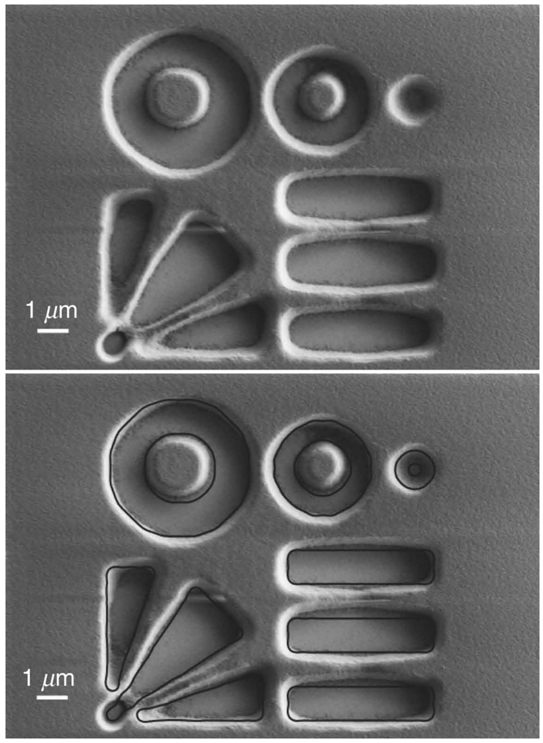

Fig. 8 Scanning electron micrograph of the developed test pattern. The bottom image includes an overlay of the predicted feature geometry.

In this work, the photoresist thickness was 700 nm. This is an extremely robust thickness that could withstand a range of deposition and etching processes. However, by diluting the photoresist with an organic solvent, the final thickness and, hence, variability could be reduced at the expense of robustness. The use of thinner films is presently being explored.

6 Conclusion

This article describes a gradient-based method for optimizing the exposure profile of scanning beam lithography processes. The proposed method is significantly more efficient than other methods as a second derivative of the cost function is not required. Simulation and experimental results show robust convergence even with complex geometries.

The experimental exposure and development of a test pattern demonstrated a minimum feature size of $\sim 1\ \mu\text{m}$. The foremost limitation at present is thought to be background dosage due to finite contrast ratio in the exposure system.

Present optimization research is focused on the convergence properties of the algorithm, including whether optimality can be guaranteed under any particular conditions. Other optimization research is aimed at increasing the exposure speed by minimizing the number of nonzero exposures and taking the dynamics of the nanopositioner system into consideration.

Future improvements to the optical system include an increased contrast ratio and direct beam profile measurement. The use of thinner photoresist films is also under investigation. In regards to the optimization, alternative cost functions are under consideration, which directly penalize proximity and microloading effects, rather than the broader measure of total exposure energy.

Appendix: Gradient Derivation

In the following, gradient of the cost function [Eq. (17)] is derived. The first term in the cost function can be rewritten

$$J_1(\Theta) = \sum_{i=1}^{N}\sum_{j=1}^{N}(Z_{i,j} - \hat{Z}_{i,j})^2, \tag{29}$$

where

$$\hat{Z}_{i,j}=\frac{1}{1+\exp\left[-\alpha\left(\sum_{k=1}^{N}\sum_{l=1}^{N}\{[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l}\}-D_{50\%}\right)\right]}. \tag{30}$$

The partial derivatives of $J_1(\Theta)$ with respect to $\Theta_{m,n}$ are

$$\frac{\partial J_1(\Theta)}{\partial \Theta_{m,n}} = 2\sum_{i=1}^{N}\sum_{j=1}^{N}(Z_{i,j} - \hat{Z}_{i,j})\left(-\frac{\partial \hat{Z}_{i,j}}{\partial \Theta_{m,n}}\right). \tag{31}$$

The partial derivatives of $\hat{Z}_{i,j}$ with respect to $\Theta_{m,n}$ are

$$\left(-\frac{\partial \hat{Z}_{i,j}}{\partial \Theta_{m,n}}\right) = \frac{\frac{\partial}{\partial \Theta_{m,n}}\exp\left(-\alpha\left\{\sum_{k=1}^{N}\sum_{l=1}^{N}[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l}-D_{50\%}\right\}\right)}{\left(1+\exp\left[-\alpha\left(\sum_{k=1}^{N}\sum_{l=1}^{N}\{[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l}\}-D_{50\%}\right)\right]\right)^2}, \tag{32}$$

$$=(\hat{Z}_{i,j})^2\frac{\partial}{\partial \Theta_{m,n}}\exp\left[-\alpha\left\{\sum_{k=1}^{N}\sum_{l=1}^{N}[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l}-D_{50\%}\right\}\right], \tag{33}$$

$$= (\hat{Z}_{i,j})^2 \exp\left[-\alpha\left\{\sum_{k=1}^{N}\sum_{l=1}^{N}[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l} - D_{50\%}\right\}\right],$$
$$\frac{\partial}{\partial \Theta_{m,n}}\left[-\alpha\left\{\sum_{k=1}^{N}\sum_{l=1}^{N}[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l} - D_{50\%}\right\}\right], \tag{34}$$

$$= \hat{Z}_{i,j}(1-\hat{Z}_{i,j})\frac{\partial}{\partial \Theta_{m,n}} \times \left[-\alpha\left\{\sum_{k=1}^{N}\sum_{l=1}^{N}[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l} - D_{50\%}\right\}\right], \tag{35}$$

$$= \alpha\hat{Z}_{i,j}(1-\hat{Z}_{i,j})\sin(\Theta_{m,n})B_{i,j}^{m,n}. \tag{36}$$

Then

$$\frac{\partial J_1(\Theta)}{\partial \Theta_{m,n}} = 2\alpha \sin(\Theta_{m,n})\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Z}_{i,j}(1-\hat{Z}_{i,j})(Z_{i,j}-\hat{Z}_{i,j})B_{i,j}^{m,n}. \tag{37}$$

Therefore, the gradient matrix of the first term is

$$\nabla_\Theta J_1 = \begin{bmatrix} \frac{\partial J_1(\Theta)}{\partial \Theta_{1,1}} & \frac{\partial J_1(\Theta)}{\partial \Theta_{1,2}} & \cdots & \frac{\partial J_1(\Theta)}{\partial \Theta_{1,N}} \\ \frac{\partial J_1(\Theta)}{\partial \Theta_{2,1}} & \frac{\partial J_1(\Theta)}{\partial \Theta_{2,2}} & \cdots & \frac{\partial J_1(\Theta)}{\partial \Theta_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial J_1(\Theta)}{\partial \Theta_{N,1}} & \frac{\partial J_1(\Theta)}{\partial \Theta_{N,2}} & \cdots & \frac{\partial J_1(\Theta)}{\partial \Theta_{N,N}} \end{bmatrix}, \tag{38}$$

$$= 2\alpha \sin(\Theta)\odot\left\{B \otimes \left[(Z-\widehat{Z})\odot\widehat{Z}\odot(1-\widehat{Z})\right]\right\}. \tag{39}$$

By using FFT function instead of convolution

$$\nabla_\Theta J_1 = 2\alpha\mathcal{F}^{-1}\left\{\mathcal{F}(B)\odot\mathcal{F}\left[(Z-\hat{Z})\odot\hat{Z}\odot(1-\hat{Z})\right]\right\}\odot\sin(\Theta). \tag{40}$$

Moving on to the second term in the cost function, the regularization term can be rewritten as

$$J_2(\Theta) = \gamma\sum_{i=1}^{N}\sum_{j=1}^{N}(D_{i,j})^2, \tag{41}$$

where

$$D_{i,j} = \sum_{k=1}^{N}\sum_{l=1}^{N}\{[1+\cos(\Theta_{k,l})]B_{i,j}^{k,l}\}. \tag{42}$$

The partial derivative of $J_2(\Theta)$ with respect to $\Theta_{m,n}$ is

$$\frac{\partial J_2(\Theta)}{\partial \Theta_{m,n}} = 2\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}D_{i,j}\left(\frac{\partial D_{i,j}}{\partial \Theta_{m,n}}\right), \tag{43}$$

$$= 2\gamma\sum_{i=1}^{N}\sum_{j=1}^{N}D_{i,j}\frac{\partial}{\partial \Theta_{m,n}}\left\{\sum_{k=1}^{N}\sum_{l=1}^{N}\left[1+\cos(\Theta_{k,l})B_{i,j}^{k,l}\right]\right\}, \tag{44}$$

$$= -2\gamma \sin(\Theta_{m,n})\sum_{i=1}^{N}\sum_{j=1}^{N}D_{i,j}B_{i,j}^{m,n}. \tag{45}$$

Thus, the gradient matrix of the second term is

$$\nabla_\Theta J_2 = -2\gamma \sin(\Theta)\odot(B\otimes D). \tag{46}$$

By using FFT function instead of convolution

$$\nabla_\Theta J_2 = -2\gamma\mathcal{F}^{-1}\{\mathcal{F}\{B\}\odot\mathcal{F}\{E\}\odot\mathcal{F}\{B_\phi\}\}\odot\sin(\Theta). \tag{47}$$

Acknowledgments

We acknowledge the contribution of the electron microscope and x-ray unit of the University of Newcastle. This work was supported by the Australian Research Council Grants DP150103521, FT130100543, and DP120100487.

References

1. L. R. Harriott, "Limits of lithography," *Proc. IEEE* **89**, 366–374 (2001).
2. R. Menon et al., "Maskless lithography," *Mater. Today* **8**(2), 26–33 (2005).
3. M. Altissimo, "E-beam lithography for micro-nanofabrication," *Biomicrofluidics* **4**, 026503 (2010).
4. A. Joshi-Imre and S. Bauerdick, "Direct-write ion beam lithography," *J. Nanotechnol.* **2014**, 170415 (2014).
5. B. J. Lin, "Marching of the microlithography horses: electron, ion, and photon: past, present, and future," *Proc. SPIE* **6520**, 652002 (2007).
6. Y. Chen, "Nanofabrication by electron beam lithography and its applications: a review," *Microelectron. Eng.* **135**, 57–72 (2015).
7. Y. Lin et al., "Sub-30 nm lithography with near-field scanning optical microscope combined with femtosecond laser," *Appl. Phys. A* **80**, 461–465 (2005).
8. B. S. Routley, J. L. Holdsworth, and A. J. Fleming, "Optimization of near-field scanning optical lithography," *Proc. SPIE* **9423**, 94230F (2015).
9. A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Springer, London (2014).
10. B. D. Cook and S.-Y. Lee, "PYRAMID-a hierarchical, rule-based approach toward proximity effect correction. II. Correction," *IEEE Trans. Semicond. Manuf.* **11**, 117–128 (1998).
11. S.-Y. Lee and B. D. Cook, "PYRAMID-a hierarchical, rule-based approach toward proximity effect correction. I. Exposure estimation," *IEEE Trans. Semicond. Manuf.* **11**, 108–116 (1998).
12. B. D. Cook and S.-Y. Lee, "Dose, shape, and hybrid modifications for PYRAMID in electron beam proximity effect correction," *J. Vac. Sci. Technol. B* **15**, 2303–2308 (1997).
13. L. E. Ocola et al., "Automated geometry assisted proximity effect correction for electron beam direct write nanolithography," *J. Vac. Sci. Technol. B Nanotechnol. Microelectron.: Mater. Process. Meas. Phenom.* **33**, 06FD02 (2015).
14. A. J. Fleming et al., "A nonlinear programming approach to exposure optimization in scanning laser lithography," in *American Control Conf.*, Boston, Massachusetts (2016).
15. B. E. Saleh and S. I. Sayegh, "Reduction of errors of microphotographic reproductions by optimal corrections of original masks," *Opt. Eng.* **20**(5), 205781 (1981).
16. Y. Liu and A. Zakhor, "Binary and phase shifting mask design for optical lithography," *IEEE Trans. Semicond. Manuf.* **5**, 138–152 (1992).
17. S. Sherif, B. Saleh, and R. De Leone, "Binary image synthesis using mixed linear integer programming," *IEEE Trans. Image Process.* **4**(9), 1252–1257 (1995).
18. Y.-H. Oh, J.-C. Lee, and S. Lim, "Resolution enhancement through optical proximity correction and stepper parameter optimization for 0.12-m mask pattern," *Proc. SPIE* **3679**, 607 (1999).
19. T. Matsunawa et al., "Adaptive optical proximity correction using an optimization method," in *7th IEEE Int. Conf. on Computer and Information Technology (CIT '07)*, pp. 853–860 (2007).
20. J. Li, S. Liu, and E. Y. Lam, "Efficient source and mask optimization with augmented Lagrangian methods in optical lithography," *Opt. Express* **21**, 8076–8090 (2013).
21. Y. Granik, "Fast pixel-based mask optimization for inverse lithography," *J. Micro/Nanolith. MEMS MOEMS* **5**(4), 043002 (2006).
22. L. Pang et al., "Validation of inverse lithography technology (ILT) and its adaptive SRAF at advanced technology nodes," *Proc. SPIE* **6924**, 69240T (2008).
23. A. Poonawala and P. Milanfar, "Mask design for optical microlithography-an inverse imaging problem," *IEEE Trans. Image Process.* **16**, 774–788 (2007).
24. O. T. Ghalehbeygi et al., "Optimization and simulation of exposure pattern for scanning laser lithography," in *IEEE Conf. on Control Applications (CCA)* (2015).
25. NEWPORT, "Gaussian beam optics," 2017, www.newport.com/n/gaussian-beam-optics (20 March 2017).
26. A. R. Neureuther and C. A. Mack, "Optical lithography modeling," in *Handbook of Microlithography, Micromachining, and Microfabrication*, Microlithography, Vol. **1**, SPIE Press, Bellingham, Washington (1997).
27. C. A. Mack, "Development of positive photoresists," *J. Electrochem. Soc.* **134**(1), 148–152 (1987).
28. C.-N. Ahn, H.-B. Kim, and K.-H. Baik, "Novel approximate model for resist process," *Proc. SPIE* **3334**, 752 (1998).
29. S. A. Robertson, C. A. Mack, and M. J. Maslow, "Toward a universal resist dissolution model for lithography simulation," *Proc. SPIE* **4404**, 111 (2001).
30. J. Randall et al., "Variable-threshold resist models for lithography simulation," *Proc. SPIE* **3679**, 176 (1999).
31. W.-C. Huang et al., "Two threshold resist models for optical proximity correction," *Proc. SPIE* **5377**, 1536 (2004).
32. M. Born et al., *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed., Cambridge University Press, Cambridge, New York (1999).
33. D. Gil, R. Menon, and H. I. Smith, "The case for diffractive optics in maskless lithography," *J. Vac. Sci. Technol. B* **21**, 2810–2814 (2003).
34. F. Liu and X. Shi, "An efficient mask optimization method based on homotopy continuation technique," in *Design, Automation and Test in Europe Conf. and Exhibition (DATE '11)*, pp. 1–6 (2011).
35. C. Kelley, *Iterative Methods for Optimization*, Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania (1999).
36. Microchemicals, "AZ ECI series; high resolution with broad process window," 2017, www.microchemicals.com/products/photoresists.html (20 March 2017).

Omid Tayefeh Ghalehbeygi is a PhD student in the School of Electrical Engineering and Computer Science, the University of Newcastle and his dissertation research involves study of the physics and photonic properties of structures created from subwavelength optical phenomena. He is a member of the Precision Mechatronics Lab (PML), and he is interested in laser scanning lithography, probe-based nanofabrication, computer modeling, and optimization.

Adrian G. Wills earned his BE (Elec.) and PhD degrees from The University of Newcastle, Australia, (Callaghan Campus) in May 1999 and May 2003, respectively. Since then, he has held postdoctoral research positions at Newcastle and spent three years working in industry. The focus of his research has been in the areas of system identification and model predictive control. In July 2015, he returned to The University of Newcastle to lead the mechatronics engineering program.

Ben S. Routley is a research assistant in the School of Electrical Engineering and Computer Science, The University of Newcastle, with a specialization in near-field optics. He is interested in probe-based lithography and computer modeling. Currently, he is completing his PhD in the School of Mathematical and Physical Sciences with a specialization in organic photovoltaic devices.

Andrew J. Fleming graduated from The University of Newcastle, Australia (Callaghan campus) with a bachelor of electrical engineering in 2000 and PhD in 2004. He is presently an Australian research council future fellow and director of the PML, The University of Newcastle, Australia. His research awards include the ATSE Batterham Medal, the IEEE Transactions on Control Systems Technology Outstanding Paper Award, and The University of Newcastle Researcher of the Year Award.