RESEARCH ARTICLE | JUNE 26 2023

Generalized performance optimization for massively-parallel electron-beam systems

Soo-Young Lee; Md Nabid Hasan

Journal of Vacuum Science & Technology B 41, 042603 (2023)
https://doi.org/10.1116/6.0002653

CrossMark

26 June 2023 20:00:27

JVST B
Journal of Vacuum Science & Technology B
ARTICLE
avs.scitation.org/journal/jvb

Generalized performance optimization for massively-parallel electron-beam systems

Cite as: J. Vac. Sci. Technol. B **41**, 042603 (2023); doi: 10.1116/6.0002653
Submitted: 8 March 2023 · Accepted: 17 May 2023 ·
Published Online: 26 June 2023

View Online

Export Citation

CrossMark

Soo-Young Lee[a)] and Md Nabid Hasan

AFFILIATIONS

Department of Electrical and Computer Engineering, Auburn University, Auburn, Alabama 36849

[a)]**Electronic mail:** leesooy@eng.auburn.edu

ABSTRACT

The massively-parallel e-beam system (MPES) provides a large number of programmable beams of which the optimal use is critical to maximizing the efficiency of the system. In our previous study of proximity effect correction (PEC) under the constraints of the MPES, the critical dimension error and the line edge roughness were minimized by reducing the feature area to be exposed and varying the dose spatially. In another study, a method to reduce the exposing time while still ensuring a near-optimal PEC result was developed. In this method, the maximum dose difference between two regions of a feature was carefully decreased after first obtaining the optimal linewidth reduction and the spatial dose distribution for the PEC. However, this method designed with an emphasis on the PEC and simplicity may miss the optimal result due to the fixed order of performance metrics considered and was developed for a single feature. To address these limitations, an adaptive optimization method that can handle any combination of performance metrics in a cost function is proposed. It allows for more flexible optimization and can achieve better results than the old method. Additionally, the optimization method is extended for large-scale patterns with uniform features, such as line-space patterns. Also, an effective way to handle the recursive effect among critical locations in a large-scale pattern is described. The simulation results show that the proposed optimization method outperforms the old method in terms of the cost function and linewidth uniformity among the features.

26 June 2023 20:00:27

Published under an exclusive license by the AVS. https://doi.org/10.1116/6.0002653

I. INTRODUCTION

Electron-beam (e-beam) lithography plays an important role in nanofabrication, being able to transfer high-resolution patterns directly onto the resist.[1–5] Its main disadvantage is low throughput due to the pixel-by-pixel or feature-by-feature writing. The recently developed massively-parallel electron-beam systems (MPESs), e.g., the eMET,[6] MBM-1000,[7] and MAPPER Lithography FLX-1200 tool,[8] are equipped with multiple beams, e.g., 512×512, and achieve improved productivity compared to a conventional single-beam system. The write time of a mask sized $100 \times 130\,\mathrm{mm}^2$ is 30+ h on the single-beam system while less than 10 h on the MPES.[9] However, the MPES still faces challenges due to their unique constraints, such as a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams.

One of the major issues in e-beam lithography is the proximity effect due to electron scattering, which limits the minimum feature size and the maximum circuit density realizable in a circuit pattern of nanoscale features. Therefore, the proximity effect correction (PEC) is critical to achieving the highest feature resolution possible with minimal critical dimension (CD) error and line edge roughness (LER). Several effective PEC schemes have been developed,[10–13] but these methods do not take into account the unique constraints of MPES. Therefore, minimizing the CD error and LER while ensuring minimal exposing time for maximum throughput in the MPES remains a challenge.

In our previous work,[14] a shape+dose correction procedure was implemented for PEC in a single feature under the constraints of MPES (e.g., a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams). A cost function consisting of the CD error, LER, and total dose was minimized by iteratively adjusting the linewidth reduction, ΔW, and the spatial dose distribution. Four different spatial dose distributions were considered, i.e., uniform, A-type, M-type, and V-type.[5] However, the dose difference between two regions in a feature can be significant in the nonuniform dose distributions, which reduces the beam utilization and lengthens the exposing time in the MPES.[14] An optimization method was proposed to overcome this problem in our previous work,[15] where the performance metrics for the PEC and exposing

Published under an exclusive license by the AVS

time were both included in the cost function. In this method, the optimal ΔW and the spatial dose distribution were determined for the PEC at first. Then, the dose difference among the regions of a feature was carefully reduced iteratively to increase the beam utilization while still ensuring a near-optimal PEC result. However, this method may miss the optimal result due to the fixed order of performance metrics considered. This paper proposes a new optimization method where all the performance metrics, i.e., CD error, LER, and exposing time, are considered in each iteration of the algorithmic process to adjust the control parameters. After an iteration, the contribution of each metric to the cost function is calculated. Then, in the next iteration, the dose distribution is adjusted to reduce the metric with the highest contribution.

Another issue in our previous optimization method is that it was developed for the case of a single feature. Applying the optimization method for a single feature to a large circuit pattern is not straightforward since the exposure (energy deposited in the resist) level varies with the location within the pattern. Also, due to the resist development simulation, the feature-by-feature optimization would be very time-consuming. Nevertheless, for a large uniform pattern, such as line-space (L/S) patterns where the same feature is replicated uniformly, the exposure level varies gradually in space. Note that the spatial variation of exposure level mainly depends on the global exposure due to the backscattering of electrons. Therefore, the optimization method may be explicitly applied to obtain the optimal dose distributions only at the critical locations in a circuit pattern, and the dose distributions at other locations can be obtained by exploiting the gradual spatial variation of exposure level. However, optimizing the dose distribution for one critical location in a pattern may have an unintended effect on the dose distributions at other critical locations, resulting in a recursive effect. To compensate for the recursive effect among the critical locations, the adjustment of the dose distribution needed for each critical location is underestimated with the use of an adjustment factor ($a.f.$).

The main contributions of this work are (i) addressing the unique constraints of MPES and enabling sub-beam-size adjustment of feature size and spatial dose modulation, (ii) taking into account all major performance metrics simultaneously during optimization to provide a more flexible and balanced approach, and (iii) achieving a uniform resist profile throughout large-scale patterns with uniform features under the MPES constraints.

The rest of the paper is organized as follows. The simulation model is depicted in Sec. II. In Sec. III, the writing method for the MPES used in this study is briefly reviewed. In Sec. IV, the new optimization method for a single feature is described in detail along with a brief review of the old optimization method. The optimization method for large-scale patterns is presented in Sec. V. In Sec. VI, the simulation setup is described. In Sec. VII, the simulation results are discussed, followed by a summary in Sec. VIII.

II. MODEL

The model of MPES employed in this study is derived from the eMET[6] and MBM-1000[7] where the beam array size is 512×512, the cross section of a beam is a square of $10 \times 10\,\text{nm}^2$, and the beam energy is 50 KeV. Beams can be turned on/off individually, but all beams must be deflected in a synchronized manner, i.e., at the same angle and direction.

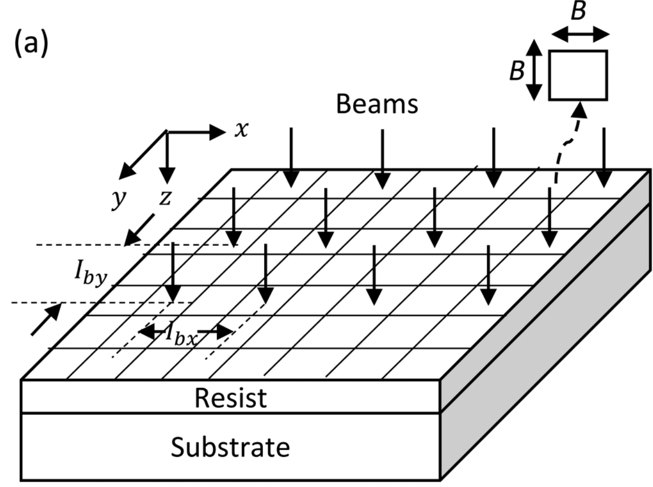


FIG. 1. (a) 3D model of a substrate system where the substrate moves in the X direction exposed by the parallel beams, the beam size is $B \times B$, and the beam intervals are I_{bx} and I_{by} in the x and y dimensions, respectively, and (b) a large-scale uniform pattern with line features where the three critical locations (corner, edge, and center) and test locations (A, B, and C) are shown.

It is assumed in this study that the resist layer of a substrate system is on the X-Y plane and parallel beams are in the Z direction, i.e., normal to the X-Y plane, as illustrated in Fig. 1(a). The resist layer to be exposed can be modeled as an array of square pixels with each pixel of size $B \times B$, the same as the beam size at the resist

26 June 2023 20:00:27

Published under an exclusive license by the AVS

surface. The exposing interval, I_{ex}, can be larger, smaller, or equal to B. For simplicity, it is assumed that $I_{ex} = B$ throughout this paper. A large-scale uniform pattern considered in this study is illustrated in Fig. 1(b), where the critical and test locations are shown.

The 3D exposure distribution in the resist, $e(x, y, z)$, can be computed through the following convolution:

$$e(x, y, z) = \iint d(x - x', y - y')TF(x', y', z)dx'dy', \quad (1)$$

where $d(x, y)$ specifies the spatial distribution of the dose given by all beams, and $TF(x, y, z)$ denotes the 3D transfer function describing the exposure distribution in the resist when a point (pixel) is exposed by a beam.

The TF (refer to Fig. 2) is modeled by the convolution of an ideal TF (constant within the area of $B \times B$ and zero outside for all layers of resist) and the sum of two Gaussian functions representing the forward and backward scattering of electrons. The sum of two Gaussian functions is usually expressed as follows:[16]

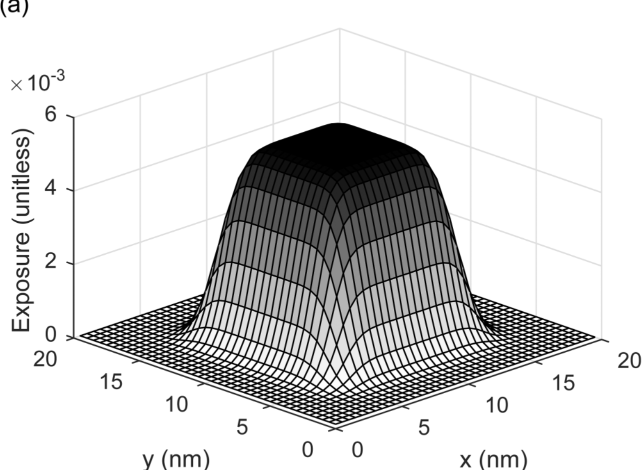
$$f(r) = \frac{1}{\pi(1+\eta)}\left\{\frac{1}{\sigma_1^2} exp\left(-\frac{r^2}{\sigma_1^2}\right) + \frac{\eta}{\sigma_2^2} exp\left(-\frac{r^2}{\sigma_2^2}\right)\right\}, \quad (2)$$

where η is the ratio of the backscattered energy to the forward-scattered energy, r is the distance from the point of electron incidence, and σ_1 and σ_2 are the forward and backward scattering ranges, respectively.

The exposure distribution in the case of a single feature is mainly of local exposure, i.e., the exposure contribution from the forward scattering of electrons. The global exposure, i.e., the backscattering component of exposure, has a relatively low amplitude and stays almost constant within a single feature. On the other hand, in the case of a large-scale pattern, the global exposure varies significantly with the location in the pattern. Therefore, the forward and backward scattering components are both included while modeling the TF for this study. Also, a certain level of Gaussian noise is added to the (deterministic) transfer function since the exposure in the resist is stochastic in nature.[14]

III. WRITING METHOD

In our previous work,[15] a writing method was proposed to reduce the exposing time in the MPES. This writing method is applicable to realizing both uniform and nonuniform dose distributions. In this writing method, the dose given to a pixel in a *step* is denoted by d, and the duration of n_s steps to give the target dose of $D = n_s d$ to each pixel is referred to as a *cycle*. The beam interval I_{bx}, the distance between two adjacent beams, should be equal to $(n_s + 1)B$ so that a uniform dose distribution can be achieved without any hole (dose lower than D) in the exposed region. As the substrate moves continuously underneath the array of beams, each beam follows a pixel by being deflected in each step of a cycle.[17] Then, the beam is reset back to its vertical orientation and exposes another pixel in the next cycle. It is possible that all the beams may fall onto spaces between features in a L/S pattern or pixels that do not require a dose in some cycles, referred to as the "empty cycle." Especially when $I_{bx} = l + s$ where l and s are the widths of the feature and space,



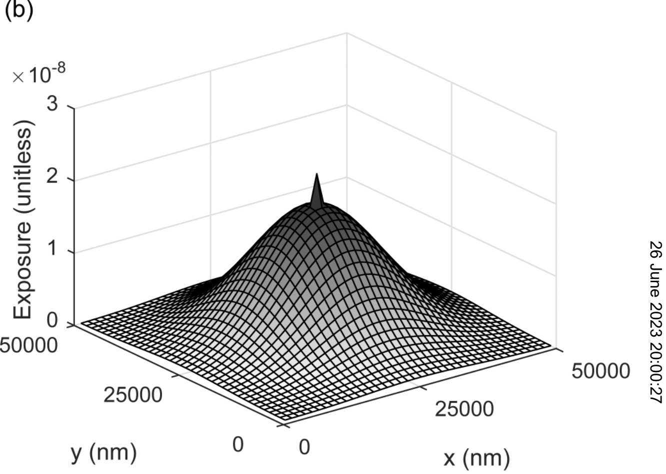


FIG. 2. Deterministic transfer function for the beam aperture of size $10 \times 10\,nm^2$ where (a) $\sigma_1 = 2\,nm$ and $\sigma_2 = 0$ and (b) $\sigma_1 = 4\,nm$ and $\sigma_2 = 9.5\,\mu m$.

26 June 2023 20:00:27

respectively, $\frac{s}{I_{bx}}$ of the total cycles are empty cycles. These empty cycles need to be removed to improve the beam utilization and reduce the exposing time. In our writing method, the deflection angle of beams is adjusted such that the beams skip to the next cycle in which at least one beam falls on a pixel requiring a dose.

Figure 3 illustrates the idea of the writing method used in this study. In the fourth cycle, all the beams fall onto spaces between features. Hence, the deflection angle of beams is adjusted to expose the next available pixels in the pattern.

IV. OPTIMIZATION METHOD FOR A SINGLE FEATURE

In this section, our previous work[15] on the optimization method for a single feature is briefly reviewed, and then the proposed optimization method is presented.

Published under an exclusive license by the AVS

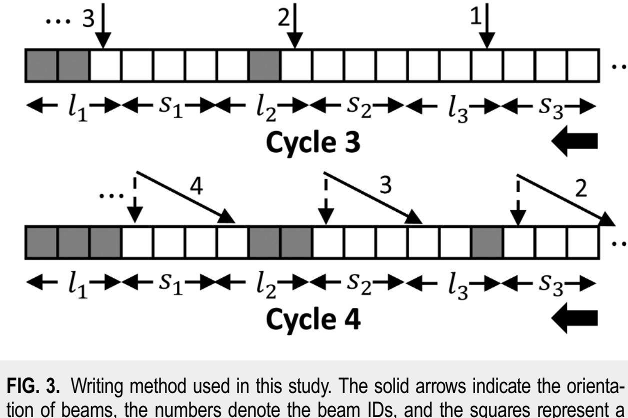


FIG. 3. Writing method used in this study. The solid arrows indicate the orientation of beams, the numbers denote the beam IDs, and the squares represent a row of pixels. The substrate is moving to the left as more beams are visible exposing the pixels. The gray and white squares represent the exposed and unexposed pixels, respectively. In the figure, l_i is the line i in an L/S pattern, s_i is the space between two adjacent lines l_i and l_{i+1}, $l_{bx} = l + s$, and $l = s = 3B$.

FIG. 4. Top-down view of the remaining resist profile where the solid and dashed lines represent the actual and target resist profiles, respectively. The CD error is calculated as the absolute difference between the actual and target CD. The LER is computed as the standard deviation of edge location $x_e(y)$ from the average edge location.

A. Cost function

The cost function employed in this study, C, consists of three metrics, i.e., the CD error, LER, and exposing time t_{exp},

$$C = a_1 CD_error_{norm} + a_2 LER_{norm} + a_3 t_{exp_{norm}}, \tag{3}$$

where a_1, a_2, and a_3 are the weights given to the normalized CD_error, LER, and t_{exp}, respectively.

The CD error is defined as the absolute difference between the actual and target CDs and the LER as the standard deviation of edge location, averaged over the top, middle, and bottom layers as illustrated in Fig. 4. The exposing time t_{exp} is calculated in cycles defined in Sec. III. The ranges and units of the metrics in the cost function are different. Therefore, the metrics are normalized so that they are on the same scale as follows:

$$\tilde{X} = \frac{X - X_{\min}}{X_{range}}, \tag{4}$$

where X is a metric, $X_{\min}$ is the minimum value of X, X_{range} is the difference between the maximum and minimum values of X, and $\tilde{X}$ is the normalized metric.

In the optimization method, the optimal dose in each region i is searched between $d_{\min_allowed}$ and $d_{\max_allowed}$, where $d_{\min_allowed}$ and $d_{\max_allowed}$ are the minimum and maximum doses allowed for a region, respectively. The $d_{\min_allowed}$ is selected such that when all the regions of a feature receive that dose, the CD is smaller than the target CD in all of the top, middle, and bottom layers and the CD error in the bottom layer is a certain percentage of the target CD (i.e., 10%). The $d_{\max_allowed}$ is selected to be an integer multiple of $d_{\min_allowed}$, i.e., $d_{\max_allowed} = P d_{\min_allowed}$, where P is a positive integer. When all the regions of a feature receive a relatively low dose $d_{\min_allowed}$ and therefore, the edge location of the developed feature is inside the exposed area, the LER is relatively large.[18] For the normalization, the CD error and LER obtained in the bottom layer when all the regions of a feature get the same low dose $d_{\min_allowed}$ are set to be the maximum CD error and LER, respectively. In either a uniform or nonuniform dose distribution, the CD error and LER cannot become larger than the maximum CD error and LER, respectively, when the minimum dose allowed for a region is $d_{\min_allowed}$. The minimum CD error and LER are both 0. The maximum exposing time for the normalization is obtained considering the case when all the regions of a feature receive the maximum dose $d_{\max_allowed}$. Similarly, the minimum exposing time is obtained when all the regions of a feature receive the minimum dose $d_{\min_allowed}$.

26 June 2023 20:00:27

B. Old optimization method

In our previous work,[14] the shape+dose correction was implemented through an iterative procedure to find the optimal linewidth reduction and the optimal spatial-dose-distribution ratio for a single feature. The feature is partitioned into five regions along its length dimension, and a dose is determined for each region using the 3D resist profile correction. The dose adjustment procedure involves selecting a region in the feature based on the CD errors in

Published under an exclusive license by the AVS

the top, middle, and bottom layers of resist and determining the amount of dose adjustment in that region based on the resist profile while maintaining the relationship among region doses for the given type of dose distribution, i.e., uniform, A-type, M-type, and V-type. The region for the dose adjustment is selected such that the CD errors are balanced among layers. In the optimal dose distribution for the PEC, the dose difference between any two regions can be significant causing many beams to be turned off in multiple cycles when they fall over the pixels requiring lower doses than others. The larger the dose difference among the regions, the lower the beam utilization and, therefore, the longer the exposing time.

In another work,[15] an optimization method to reduce the exposing time while still ensuring a near-optimal PEC result was proposed. In this method, the performance metrics are considered in a fixed order, i.e., CD error (and LER) and the exposing time. That is, the PEC is completed first, followed by reducing the exposing time. After finding the optimal linewidth reduction and the optimal spatial dose distribution for the PEC, the dose difference among the regions of a feature is minimized iteratively to reduce the exposing time by enhancing the beam utilization while maintaining the PEC result acceptable. The maximum dose among the regions is decreased and the reduced dose is distributed evenly in the nearby regions in each iteration until the minimum value of cost function is reached. The increased doses in the nearby regions compensate for the reduction of maximum dose in the resist development.

C. New optimization method

In the new optimization method, all the metrics, i.e., the CD error, LER, and exposing time, are considered in each iteration to adjust the control parameters. The following notations are used in the description of the method.

$C_{\min}$: The minimum value of cost function
$C_{\min,\Delta W}$: The minimum value of cost function for a given ΔW
$\{d_i\}$: The spatial dose distribution within a feature where i is the region index ($i = 1, 2, 3, 4, 5$)
x: The fraction of d_i to be added to or subtracted from d_i when adjusting dose where $0 < x < 1$
y: A small fraction (e.g., 0.01)
s_1: A parameter multiplied to x when it is required to increase x ($s_1 > 1$)
s_2: A parameter multiplied to x when it is required to decrease x ($0 < s_2 < 1$)
Δd_i: The amount of adjustment for d_i where $\Delta d_i = d_i \times x$
$\{d_i\}_{opt}$: The optimal spatial dose distribution
ΔW: The linewidth reduction
ΔW_{opt}: The optimal linewidth reduction
$d_{\max}$: The maximum dose among the regions of a feature

The new iterative optimization method to minimize the cost function for a single feature is described below, where k is the iteration index. The C_0, $C_{\min}$, and $C_{\min,\Delta W}$ are initially set to a large value [greater or equal to $(a_1 + a_2 + a_3)$]. To obtain the dose distribution realizable on an MPES given ΔW and $\{d_i\}$ (step 3), a method developed in our previous study[14] is employed. In step 9, if the contribution of the CD error and LER to the cost function is larger than that of exposing time, the dose distribution is adjusted through the shape+dose correction to get a resist profile closer to the target. If the contribution of the CD error and LER to the cost function is smaller than that of exposing time, the dose distribution is adjusted to enhance the beam utilization.[15] The iteration starts with a moderate value of x (e.g., 10%), and x can be adjusted in steps 7 and 8 if required. In step 7, to ensure a reasonable progress toward the minimal value of cost function in each iteration, x is multiplied by s_1 ($s_1 > 1$) if the improvement in C is too small, i.e., less than $C_{k-1} \times y$ where y is a small fraction (e.g., 0.01). A large y may cause x to be multiplied by s_1 too often, even when C is improving at a reasonable rate. This can cause the optimization method to overshoot the minimum and diverge. Similarly, a large s_1 may also cause the algorithm to overshoot the optimal point. If the value of cost function in the current iteration becomes larger than that in the previous iteration and the difference between them is at least $C_{k-1} \times y$, the $\Delta d_i = d_i \times x$ was larger than required. Hence, in such cases (see step 8), x is multiplied by s_2 where $0 < s_2 < 1$. However, a very small s_2 may cause the optimization method to converge slowly.

1. $\Delta W \leftarrow 0$, $C_0 \leftarrow$ a large value, $C_{\min} \leftarrow C_0$, and $C_{\min,\Delta W} \leftarrow C_0$.
2. Initialize $\{d_i\}$ according to the type of dose distribution.
3. Obtain the dose distribution realizable on an MPES given ΔW and $\{d_i | i = 1, 2, 3, 4, 5\}$.
4. Evaluate the cost function, C_k.
5. If $C_k < C_{\min}$,
$$C_{\min} \leftarrow C_k,\ \Delta W_{opt} \leftarrow \Delta W \text{ and } \{d_i\}_{opt} \leftarrow \{d_i\}.$$
6. If $| C_k - C_{k-1} | <$ *tolerance*, go to step 10.
7. If $C_{k-1} - C_{k-1} \times y < C_k < C_{k-1}$ where $0 < y < 1$ ($C_{k-1} \times y > tolerance$), i.e., the improvement in C is too small,
$$x \leftarrow s_1 x\ (s_1 > 1).$$
8. If $C_k > C_{k-1} + C_{k-1} \times y$, i.e., if the current value of cost function is larger than the previous one and the difference between them is at least $C_{k-1} \times y$,
$$x \leftarrow s_2 x\ (0 < s_2 < 1).$$
9. If $a_1 CD_error_{norm} + a_2 LER_{norm} > a_3 t_{exp_{norm}}$,
$d_i \leftarrow d_i \pm \Delta d_i$ (where $\Delta d_i = d_i \times x$) such that $d_{\min_allowed} < d_i \pm \Delta d_i < d_{\max_allowed}$ and the shape of dose distribution type is maintained. Go to step 3.
else
$d_{\max} \leftarrow d_{\max} - \Delta d_i$ (where $\Delta d_i = d_i \times x$) such that $d_{\max}$ is still the maximum of the doses among the five regions and Δd_i is evenly distributed to the nearby regions. Go to step 3.
10. If $C_{\min} < C_{\min,\Delta W}$,
$$C_{\min,\Delta W} \leftarrow C_{\min},\ \Delta W \leftarrow \Delta W + 2\text{ nm},\ \text{and go to step 2.}$$
11. Output ΔW_{opt} and $\{d_i\}_{opt}$.

26 June 2023 20:00:27

Published under an exclusive license by the AVS

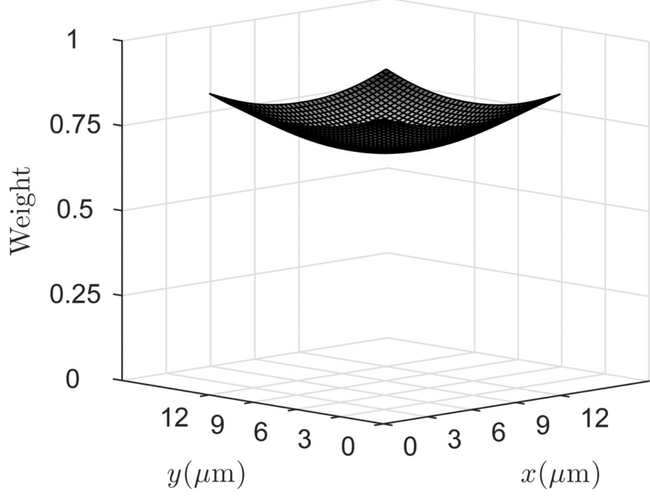


FIG. 5. Deconvolution surface for the pattern size of $12 \times 12 \mu m^2$ where each line is 50 nm wide, the space between lines is 50 nm, $\sigma_1 = 4$ nm, $\sigma_2 = 9.5 \mu m$, and $\eta = 0.74$.

V. OPTIMIZATION METHOD FOR A LARGE-SCALE PATTERN

The optimization for a large-scale pattern consists of three steps: (i) minimizing the cost function for a single feature in isolation, (ii) global adjustment of the dose distribution throughout the pattern, and (iii) minimizing the cost function iteratively considering all the features in the pattern.

In the first step, the cost function in Eq. (3) is minimized in correcting a single instance of the repeated feature in an L/S pattern in isolation (refer to Sec. IV C). The optimized dose distribution within the feature is denoted by $\{d_i\}_{single}$. The spatial average of the exposure resulted from $\{d_i\}_{single}$, denoted by $\overline{E}$, is computed for each layer of resist, i.e., $\overline{E} = \frac{1}{WL}\sum_{y=0}^{L}\sum_{x=0}^{W} e(x, y)$, where W and L are the width and length of the feature, respectively.

In the second step, the global distribution of dose throughout the pattern is derived by the following deconvolution:

$$A = {E_t}^{*-1} TF, \tag{5}$$

where E_t is a 2D matrix where the element (m, n) is the target (average) exposure, i.e., $\overline{E}$, for the location (m, n) of the feature in the pattern and TF is the transfer function (refer to Sec. II) sampled at the feature interval in both X and Y dimensions. The output of the deconvolution, i.e., matrix A, which specifies the global distribution of (average) dose required to achieve $\overline{E}$ at all locations of the feature, is referred to as a *deconvolution surface*[19] (see Fig. 5). This deconvolution is not computationally intensive as the spatial resolution involved is coarse (feature interval). To obtain the dose distribution within the feature at the location (m, n) in the pattern, $\{d_i\}_{single}$ is weighted (scalar multiplication) by the deconvolution surface A to result in

$$\{d_i\}_{(m,n)} = \{d_i\}_{single} \cdot A_{(m,n)}, \tag{6}$$

where $\{d_i\}_{(m,n)}$ is the dose distribution at the (m, n)th location of the feature.

Now, the average exposures are well balanced to be $\overline{E}$ at all locations of the features in a pattern. However, the exposure in the space between features may vary significantly with the location due to the spatial variation of global exposure. Then, the resist profiles of the feature at different locations might be different. In order to achieve a uniform resist profile throughout the pattern, the dose distribution in each feature is further adjusted in the next step so that the adjustment of exposure within a feature can compensate for the variation of exposure in the space. Note that the global adjustment in the optimization is a part of the PEC for a large-scale pattern. The proposed optimization method goes beyond the PEC for a single feature in isolation to address the spatial variation of global exposure among features in a large-scale pattern and achieve the uniform resist profile of feature throughout the pattern.

In the third step, the optimal linewidth reduction, ΔW_{opt}, and dose distributions, $\{d_i\}_{opt}$, for all the features in the pattern are obtained by minimizing the cost function in Eq. (3). The optimization method considers different amounts of ΔW in an outer loop, while the spatial dose distributions for all the features in the pattern are optimized for a given ΔW in the inner loop through iterations. The initial dose distribution is set to $\{d_i\}_{(m,n)}$ [refer to Eq. (6)] obtained in the second step because the average exposures are already well balanced at all locations of the features after the second step. The inner loop includes a nested loop, which is repeatedly executed to find the optimal dose distribution for each of the critical locations, i.e., corner, edge, and center [refer to Fig. 1(b)] individually, while the dose distributions of the other locations are fixed in each iteration. The optimization method used is the same as described in Sec. IV C, except that the contributions of global exposure from the other features in the pattern are also taken into account. In a large-scale uniform pattern, the same feature is replicated uniformly throughout the pattern. Then, the global spatial distribution of exposure is smooth (gradually decreasing from the center), and, in turn, the global distribution of dose required to achieve the same (average) exposure at all the features must be smooth (gradually increasing from the center), which is indicated by the deconvolution surface (Fig. 5). Therefore, it is sufficient to determine the optimal dose distributions for the features at the critical locations first. Then, the dose distributions of the features at other locations in the pattern can be computed using the deconvolution surface. After obtaining the optimal dose distribution at each critical location individually with the dose distributions at the other locations fixed in each iteration, the dose adjustment for a critical location, $\{\Delta d_i\}_{opt}$, is calculated as the difference between the old dose distribution (i.e., before entering the nested loop for the critical location) and the optimal dose distribution in the current iteration. However, the overadjustment of the dose distribution at a critical location may cause the dose distributions at other locations to move away from the optimal point. To reduce this recursive effect among the locations, an adjustment factor $(a.f.)$ is multiplied to the

26 June 2023 20:00:27

Published under an exclusive license by the AVS

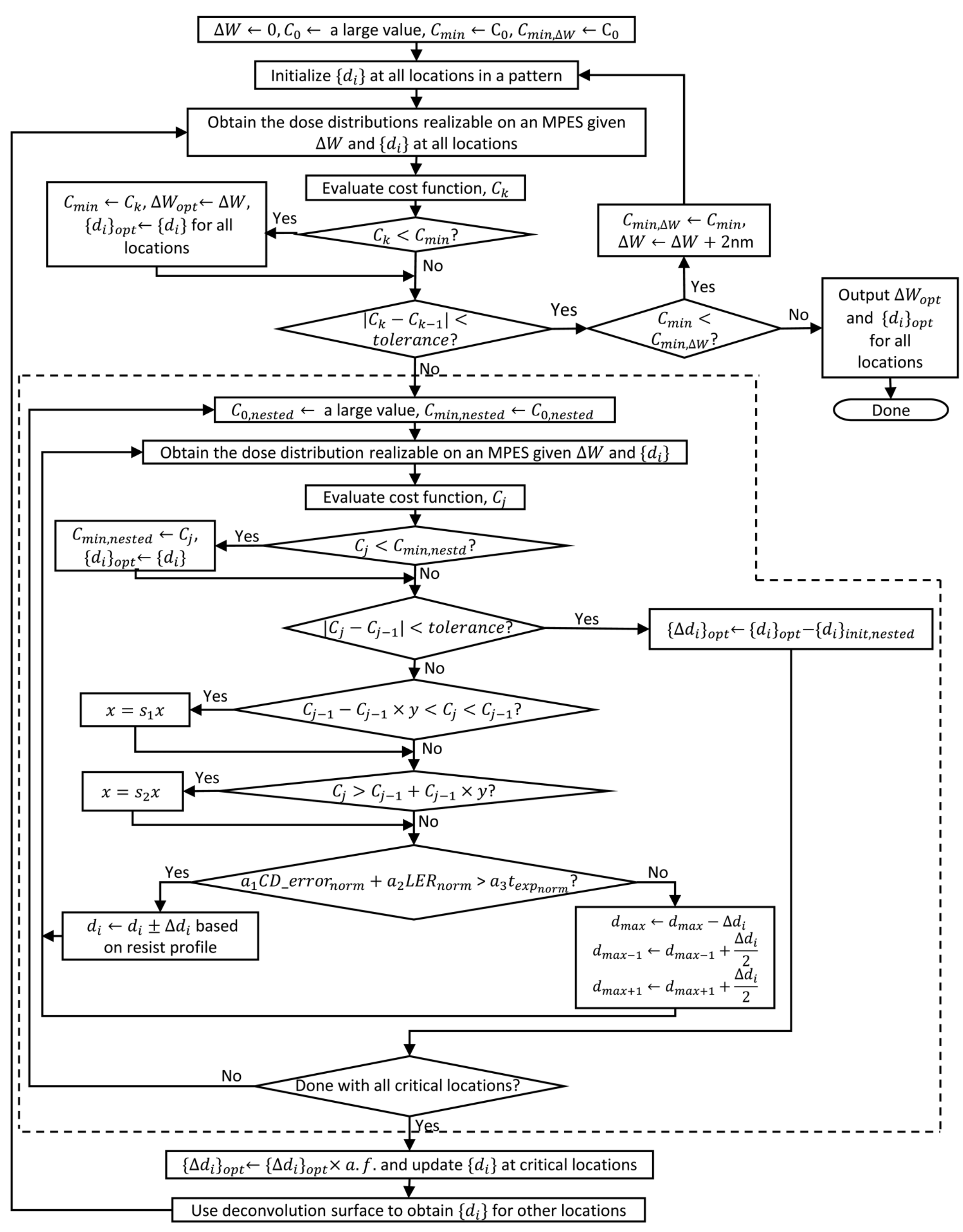


FIG. 6. Flow chart of the proposed optimization method for a large-scale pattern where k is the iteration index, $a.f.$ is the adjustment factor, $0 < x < 1$, y is a small fraction (e.g., 0.01), $s_1 > 1$, and $0 < s_2 < 1$. The part in the dashed box is executed to obtain the optimal dose distribution for each of the critical locations, i.e., corner, edge, and center, separately in each iteration where j is the index for nested iterations, $C_{0,nested}$ is the initial value of cost function, $C_{min,nested}$ is the minimum value of cost function, and $\{d_i\}_{init,nested}$ is the initial dose distribution at the critical location.

$\{\Delta d_i\}_{opt}$ obtained for each critical location where $0 < a.f. < 1$. After updating the dose distributions at the critical locations, the dose distribution at each location (m, n) is computed according to the deconvolution surface A; that is, $\{d_i\}_{(m,n)} = \{d_i\}_{(m_c,n_c)} \times \frac{A(m,n)}{A(m_c,n_c)}$, where (m_c, n_c) is a critical location. This computation is carried out with each of the critical locations (m_c, n_c) separately, and then the three results are averaged to obtain the final dose distribution at the location (m, n) in the current iteration. By adjusting the dose distributions within the feature at all the locations, the optimization method lets the adjustment of exposure within a feature compensate for the variation of exposure in the space, resulting in a uniform resist profile throughout the pattern. The flow chart of the third step

26 June 2023 20:00:27

Published under an exclusive license by the AVS

in the optimization method developed for large-scale patterns is provided in Fig. 6.

The computation time can be a limiting factor if a point-by-point or feature-by-feature optimization is performed on a large-scale pattern. However, in this study, the dose distributions only for the critical locations, i.e., corner, edge, and center, are directly optimized, and then those at other locations are derived from the optimal dose distributions at the critical locations using the deconvolution surface. This approach greatly reduces the computation time, making it possible to apply the optimization method to large-scale patterns.

VI. SIMULATION

The proposed optimization methods for a single feature and a large-scale pattern have been implemented, and their performance

TABLE I. Optimization results with the cost function $C = CD_error_{norm} + LER_{norm} + t_{exp\,norm}$. W is the feature width, σ_1 the forward scattering range of TF ($\sigma_2 = 9.5\,\mu$m), and t_{exp} the exposing time. The exposing time is calculated for a uniform pattern of 12 μm width, which consists of lines with the linewidths of 50 and 150 nm. The space between two lines is 40 nm, the beam interval is 200 nm, and the number of beams in a row is 21.

Method	W (nm)	σ_1 (nm)	Dose dist. type	ΔW (nm)	Dose ratio	CD error (nm)	LER (nm)	t_{exp} (cycle)	Cost function value
W/O optimization	50	1	A	0	1.0:2.0:7.0:2.0:1.0	1.28	0.42	4940	1.408
Old	50	1	A	4	1.8:3.7:7.3:3.7:1.8	0.32	0.15	5239	0.990
New	50	1	A	4	1.6:4.1:6.7:4.1:1.6	0.34	0.14	4843	0.926
W/O optimization	50	1	M	0	1.0:5.0:1.0:5.0:1.0	1.48	0.45	3550	1.272
Old	50	1	M	4	1.9:6.3:2.8:6.3:1.9	0.38	0.17	4541	0.919
New	50	1	M	4	1.8:5.8:3.0:5.8:1.8	0.40	0.17	4198	0.872
W/O optimization	50	1	V	0	4.0:2.0:1.0:2.0:4.0	4.05	0.59	2960	1.839
Old	50	1	V	6	5.6:3.9:3.5:3.9:5.6	0.29	0.14	4166	0.815
New	50	1	V	6	5.1:4.1:3.4:4.1:5.1	0.32	0.14	3805	0.768
W/O optimization	50	4	A	0	1.0:2.0:7.0:2.0:1.0	1.55	0.51	4940	1.552
Old	50	4	A	8	2.0:3.8:7.9:3.8:2.0	0.45	0.21	5661	1.139
New	50	4	A	6	1.0:3.4:6.9:3.4:1.0	0.48	0.21	4872	1.028
W/O optimization	50	4	M	0	1.0:5.0:1.0:5.0:1.0	1.69	0.55	3550	1.414
Old	50	4	M	8	2.7:7.3:4.2:7.3:2.7	0.46	0.21	5305	1.088
New	50	4	M	6	2.8:6.3:4.8:6.3:2.8	0.49	0.20	4630	0.984
W/O optimization	50	4	V	0	4.0:2.0:1.0:2.0:4.0	4.28	0.68	2960	1.975
Old	50	4	V	10	5.9:4.5:3.5:4.5:5.9	0.45	0.19	4397	0.931
New	50	4	V	10	5.8:4.7:4.0:4.7:5.8	0.42	0.18	4328	0.905
W/O optimization	150	1	A	0	1.0:2.0:7.0:2.0:1.0	5.09	0.76	4610	1.402
Old	150	1	A	4	2.4:4.7:8.0:4.7:2.4	0.45	0.21	5512	0.952
New	150	1	A	4	2.3:5.0:7.3:5.0:2.3	0.47	0.21	5091	0.891
W/O optimization	150	1	M	0	1.0:5.0:1.0:5.0:1.0	4.93	0.70	3590	1.211
Old	150	1	M	4	3.2:7.4:4.7:7.4:3.2	0.41	0.20	5382	0.925
New	150	1	M	4	2.0:7.0:5.2:7.0:2.0	0.42	0.20	5050	0.876
W/O optimization	150	1	V	0	4.0:2.0:1.0:2.0:4.0	4.82	0.67	2920	1.089
Old	150	1	V	4	7.0:4.9:4.0:4.9:7.0	0.36	0.18	5166	0.879
New	150	1	V	4	6.5:5.1:4.2:5.1:6.5	0.38	0.18	4819	0.829
W/O optimization	150	4	A	0	1.0:2.0:7.0:2.0:1.0	5.52	0.85	4610	1.476
Old	150	4	A	10	2.1:5.1:8.4:5.1:2.1	0.50	0.23	5784	1.005
New	150	4	A	8	2.9:4.7:7.3:4.7:2.9	0.49	0.24	5091	0.907
W/O optimization	150	4	M	0	1.0:5.0:1.0:5.0:1.0	5.28	0.79	3590	1.279
Old	150	4	M	10	2.6:8.0:5.4:8.0:2.6	0.44	0.21	5788	0.992
New	150	4	M	8	1.5:7.1:5.2:7.1:1.5	0.46	0.22	5101	0.896
W/O optimization	150	4	V	0	4.0:2.0:1.0:2.0:4.0	5.08	0.74	2920	1.141
Old	150	4	V	12	7.4:5.3:4.1:5.3:7.4	0.40	0.19	5466	0.931
New	150	4	V	10	6.5:4.5:3.8:4.5:6.5	0.43	0.20	4795	0.839

26 June 2023 20:00:27

Published under an exclusive license by the AVS

has been analyzed through simulation. In the case of a single feature, two different linewidths, i.e., 50 and 150 nm, are considered, and the length is fixed at 300 nm. In the case of a large-scale pattern, two L/S patterns of size $12 \times 12\,\mu m^2$ are considered. In one pattern, each line is 50 nm wide, and the space between lines is 50 nm. In the other pattern, each line is 100 nm wide, and the space between lines is 100 nm.

The TF is modeled based on the 3D point spread function (PSF) generated using a Monte Carlo simulation program CASINO,[20] where the resist (PMMA) thickness is 100 nm on the Si substrate, the beam energy is 50 keV, and the beam diameter is 6 nm. The resist is modeled by five layers to take into account the layer dependency of exposure. From the PSF, the ratios of total exposure and the forward scattering range among the five resist layers are determined, which are used in generating the TF of each layer. The beam size is set to $10 \times 10\,nm^2$. In the case of a single feature, both sharp and broad TFs are considered, i.e., $\sigma_1 = 1$ and 4 nm ($\sigma_2 = 9.5\,\mu m$, and $\eta = 0.74$).[21] In the case of a large-scale pattern, it is assumed that $\sigma_1 = 4$ nm, $\sigma_2 = 9.5\,\mu m$, and $\eta = 0.74$.

The 3D exposure distribution in the resist is computed through the convolution of TFs and the dose distribution in the pattern [see Eq. (1)]. The simulation interval (spatial resolution) is set to 0.5 nm. The developing-rate distribution is derived from the exposure distribution, and then the remaining resist profile is obtained through a fast path-based resist-development simulation.[22] The development simulation continues until the feature is fully developed to the bottom layer of resist. From the resist profile, the CD and LER are calculated. The middle 80% segment of the developed feature along the length dimension is used in the computation of CD and LER to exclude the edge effect (rounding at corners).

In this study, four different spatial dose distributions are employed, i.e., uniform, A-type, M-type, and V-type.[5] For realizing any spatial dose distribution with an arbitrary linewidth reduction, ΔW, on an MPES, a method developed in our previous study[14] is employed. The method utilizes multipass writing, i.e., exposes each writing path multiple times, to increase the dose range and/or resolution. The point of exposure on the substrate for each beam is shifted by the amount of $\frac{\Delta W}{2}$ in the direction of the substrate in each pass. Another benefit of using multipass is to spread the negative effects of abnormal beams spatially.[23]

The number of cycles required to realize a dose distribution is calculated by tracing the writing method on the entire pattern through simulation. Given a beam interval I_{bx} and $k(\frac{I_{bx}}{B}+1)$ beams in a row where k is a positive integer, the cycles to expose the first $\frac{I_{bx}}{B} \times \frac{I_{bx}}{B}$ pixels by the first $(\frac{I_{bx}}{B}+1)$ beams are repeated by the next $(\frac{I_{bx}}{B}+1)$ beams to expose the next $\frac{I_{bx}}{B} \times \frac{I_{bx}}{B}$ pixels, and so on.[15] Therefore, for simplicity, given a beam interval I_{bx}, $(\frac{I_{bx}}{B}+1)$ beams in a row are considered.

The weights (a_1, a_2, a_3) in the cost function are selected according to the relative importance of each metric for a given application. In the optimization procedure, $x = 0.5$ initially, $y = 0.01$, $s_1 = 2$, $s_2 = 0.5$, and $tolerance = 0.005$. The minimum dose is selected such that when all the regions of a feature receive that dose, the CD error is negative in all of the top, middle, and bottom layers, and the maximum CD error is 10% of the target CD at the bottom layer. There are 100 dose levels allowed between 0 and 10 times the minimum dose to achieve a fine dose resolution (around the operating point). Hence, the dose given to a pixel in each cycle is one tenth of the minimum dose. The selection of $a.f.$ depends on the $\{\Delta d_i\}_{opt}$ for the critical locations. The larger the $\{\Delta d_i\}_{opt}$, the higher possibility of a recursive effect between the critical locations, and consequently, the $a.f.$ is to be smaller. It is assumed that $a.f. = 0.7$ initially. If the average $\{\Delta d_i\}_{opt}$ is reduced between iterations, $a.f.$ is increased by 10% until $a.f. = 1$. Similarly, if the average $\{\Delta d_i\}_{opt}$ is increased between iterations, $a.f.$ is reduced by 10% until $a.f. = 0.5$.

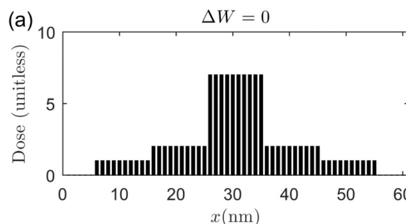
VII. RESULTS AND DISCUSSION

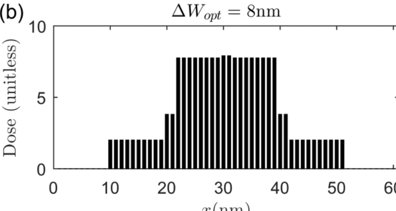
In this section, the simulation results of the old and new optimization methods for both a single feature and a large-scale pattern are compared through an extensive simulation. Multiple cost functions are considered in the simulation to demonstrate the adaptability of the new optimization method to different weights of the performance metrics. The effectiveness of using the $a.f.$ in the optimization method for a large-scale pattern is analyzed in handling the recursive effect among critical locations.

26 June 2023 20:00:27

A. Single feature

In Table I, the simulation results without any optimization, with the old optimization method, and with the new optimization method are compared. For the case of no optimization, the spatial dose distribution for each type is obtained from a previous study.[24] A cost function with equal weights to all the performance metrics is considered. To obtain a comprehensive set of results, the feature





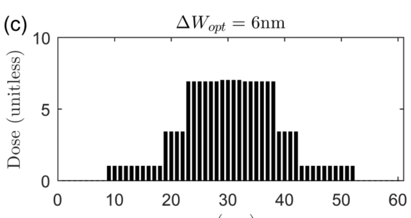


FIG. 7. A-type dose distribution obtained (a) without an optimization, (b) with the old method, and (c) with the new method for the case when the feature width $W = 50$ nm, $\sigma_1 = 4$ nm, and $\sigma_2 = 9.5\,\mu m$.

Published under an exclusive license by the AVS

TABLE II. Optimization results with four different cost functions. The feature width is 50 nm, σ_1 = 4 nm, σ_2 = 9.5 μm, l_{bx} = 100 nm, and the dose distribution is A-type. The exposing time is calculated for a uniform pattern of 12 μm width, which consists of lines with a linewidth of 50 nm. The space between two lines is 40 nm, the beam interval is 200 nm, and the number of beams in a row is 21.

Cost function	Method	ΔW (nm)	Dose ratio	CD error (nm)	LER (nm)	t_{exp} (cycle)	Cost function value
C1	Old	8	2.0:3.8:7.9:3.8:2.0	0.45	0.21	5661	1.139
C1	New	6	1.0:3.4:6.9:3.4:1.0	0.48	0.21	4872	1.028
C2	Old	8	2.1:4.1:7.5:4.1:2.1	0.58	0.24	5401	1.467
C2	New	6	1.3:3.8:6.4:3.8:1.3	0.66	0.26	4550	1.305
C3	Old	8	2.2:4.7:7.2:4.7:2.2	0.82	0.28	5225	1.770
C3	New	4	1.0:3.2:6.1:3.2:1.0	0.91	0.31	4379	1.544
C4	Old	8	2.3:5.0:7.0:5.0:2.3	1.19	0.35	5103	2.037
C4	New	2	1.0:2.4:6.0:2.4:1.0	1.30	0.37	4270	1.739

width, the forward scattering range σ_1, and the dose distribution type are varied.

It is clear from the results in Table I that the new method achieves a better result in minimizing the cost function compared to the old method, especially in the cases where the optimal ΔW is different between the two methods. The old method considers a fixed order of performance metrics to minimize the cost function, i.e., first, the CD error and LER for the PEC and then the exposing

TABLE III. Optimization results for a large-scale pattern with the cost function $C = CD_error_{norm} + LER_{norm} + t_{exp_{norm}}$. The beam interval is 200 nm, and the number of beams in each row is 21. The CD error and LER are the averages of CD errors and LERs at the critical locations, respectively.

Case	Method	Pattern	Dose dist. type	CD error (nm)	LER (nm)	t_{exp} (cycle)	Cost function value
1(a)	Single feature (new)	1	Uniform	2.11	0.27	1466	1.099
1(b)	Large-scale (old)	1	Uniform	1.23	0.26	1432	0.904
1(c)	Large-scale (new)	1	Uniform	0.93	0.25	1209	0.773
2(a)	Single feature (new)	2	Uniform	2.33	0.27	1611	0.815
2(b)	Large-scale (old)	2	Uniform	1.33	0.27	1554	0.700
2(c)	Large-scale (new)	2	Uniform	1.05	0.26	1400	0.623
3(a)	Single feature (new)	1	A	1.19	0.21	1547	0.878
3(b)	Large-scale (old)	1	A	0.70	0.20	1505	0.758
3(c)	Large-scale (new)	1	A	0.65	0.20	1309	0.695
4(a)	Single feature (new)	2	A	1.37	0.22	1675	0.712
4(b)	Large-scale (old)	2	A	0.81	0.21	1612	0.633
4(c)	Large-scale (new)	2	A	0.75	0.21	1462	0.586
5(a)	Single feature (new)	1	M	1.24	0.21	1557	0.890
5(b)	Large-scale (old)	1	M	0.72	0.21	1527	0.779
5(c)	Large-scale (new)	1	M	0.68	0.20	1379	0.719
6(a)	Single feature (new)	2	M	1.35	0.22	1644	0.702
6(b)	Large-scale (old)	2	M	0.78	0.21	1584	0.623
6(c)	Large-scale (new)	2	M	0.72	0.21	1449	0.579
7(a)	Single feature (new)	1	V	1.77	0.26	1687	1.082
7(b)	Large-scale (old)	1	V	1.06	0.24	1631	0.904
7(c)	Large-scale (new)	1	V	0.97	0.24	1445	0.836
8(a)	Single feature (new)	2	V	1.90	0.26	1664	0.782
8(b)	Large-scale (old)	2	V	1.11	0.25	1610	0.683
8(c)	Large-scale (new)	2	V	1.03	0.24	1481	0.634

26 June 2023 20:00:27

Published under an exclusive license by the AVS

time. As a result, the old method may miss the optimal ΔW in some cases and search for the optimal dose distribution for a ΔW that is optimal for the PEC but not optimal for both the PEC and reduction of exposing time. In most cases, the new method achieves a much shorter exposing time than the old method at the expense of increasing the CD error slightly.

In a previous paper,[5] it was shown that the A-type dose distribution is effective for the PEC when the aspect ratio (resist thickness to feature width) is relatively large, while the V-type distribution tends to work better for a relatively small aspect ratio. However, when considering both the PEC and exposing time, this may not be the case always. In Table I, the CD error for the V-type dose distribution is smaller than that for the A-type or M-type dose distribution when the aspect ratio is large (feature width: 50 nm and resist thickness: 100 nm). For the PEC only, the dose difference among the regions in A-type or M-type dose distributions tends to be larger than that in the V-type dose distribution.[14] However, as the exposing time is included in the cost function, the dose difference among the regions can be reduced considerably from that in the optimal dose distribution for the PEC only when the weight for the exposing time is comparable to that for the CD error. This may cause the CD error for the A-type or M-type dose distribution to be larger than that for the V-type dose distribution. Hence, the V-type dose distribution should be taken into account even when the aspect ratio is relatively large.

It is seen in Table I that the LER is low for both the old and new methods because when the edge location of a developed feature is outside the exposed area (when $\Delta W \neq 0$), the LER tends to be smaller compared to the case that the edge location is inside the exposed area.[18]

In Fig. 7, the dose distributions realized on an MPES with the optimal ΔW are illustrated for the case where the dose distribution is A-type, feature width $W = 50$ nm, and $\sigma_1 = 4$ nm. The dose distribution without an optimization is realized with one pass, while the optimal dose distributions obtained with the old and new optimization methods are realized with two passes.[14] To realize a dose distribution when ΔW is not an integer multiple of beam size B ($\Delta W \neq nB$), more than one pass is required and the point of exposure on the substrate for each beam is shifted by the amount of $\frac{\Delta W}{2}$ in the direction of substrate in each pass. Due to the overlap of

TABLE IV. CD error and LER at the top, middle, and bottom layers of the resist in the critical locations, i.e., corner, edge, and center, for the respective cases of Table III. For this table, the CD error is calculated as the difference between the actual and target CD instead of the absolute difference to provide an idea of the shape of the resist profile.

	CD error at the corner			CD error at the edge			CD error at the center					
Case	Top (nm)	Middle (nm)	Bottom (nm)	Top (nm)	Middle (nm)	Bottom (nm)	Top (nm)	Middle (nm)	Bottom (nm)	LER at corner (nm)	LER at edge (nm)	LER at center (nm)
1(a)	+1.02	−0.18	−2.70	+1.35	+0.20	−1.10	+8.25	+3.77	−0.39	0.27	0.26	0.28
1(b)	+1.42	+0.21	−1.01	+1.39	−0.18	−1.90	+1.10	−0.45	−3.41	0.26	0.26	0.27
1(c)	+1.33	+0.18	−1.41	+1.42	+0.27	−1.04	+1.47	+0.30	−0.98	0.26	0.25	0.25
2(a)	+1.13	−0.20	−2.99	+1.52	+0.25	−1.22	+9.08	+4.15	−0.45	0.27	0.26	0.28
2(b)	+1.57	−0.17	−2.04	+1.62	+0.26	−1.09	+1.24	−0.46	−3.56	0.27	0.26	0.28
2(c)	+1.58	+0.31	−1.18	+1.50	+0.27	−1.51	+1.65	+0.36	−1.05	0.26	0.27	0.26
3(a)	+0.79	−0.16	−1.08	+1.59	+0.71	−0.39	+4.01	+1.66	+0.32	0.20	0.21	0.23
3(b)	+0.68	−0.16	−1.44	+0.74	+0.07	−1.29	+0.81	+0.16	−0.96	0.22	0.20	0.19
3(c)	+0.83	−0.05	−1.07	+0.85	+0.10	−1.01	+0.73	−0.14	−1.11	0.20	0.20	0.20
4(a)	+0.93	−0.24	−1.26	+1.81	+0.83	−0.50	+4.47	+1.89	+0.41	0.21	0.22	0.24
4(b)	+0.85	+0.05	−1.45	+0.93	+0.15	−1.09	+0.79	−0.17	−1.77	0.21	0.20	0.23
4(c)	+0.88	−0.10	−1.27	+1.02	+0.15	−1.03	+0.93	−0.09	−1.25	0.21	0.20	0.21
5(a)	+0.81	−0.16	−1.13	+1.65	+0.74	−0.41	+4.19	+1.73	+0.32	0.20	0.21	0.23
5(b)	+0.87	+0.18	−0.96	+0.78	+0.07	−1.33	+0.73	−0.11	−1.47	0.20	0.21	0.22
5(c)	+0.86	−0.08	−1.08	+0.77	−0.15	−1.20	+0.91	+0.08	−0.99	0.20	0.21	0.20
6(a)	+0.89	−0.18	−1.24	+1.81	+0.79	−0.45	+4.57	+1.89	+0.35	0.20	0.22	0.24
6(b)	+0.80	−0.13	−1.74	+0.85	+0.07	−1.39	+0.89	+0.15	−1.03	0.23	0.21	0.20
6(c)	+0.86	−0.13	−1.24	+0.95	+0.13	−1.02	+0.93	−0.04	−1.16	0.21	0.20	0.21
7(a)	+1.37	−0.24	−1.16	+2.34	+0.98	−0.49	+6.29	+2.65	+0.38	0.25	0.26	0.28
7(b)	+1.38	−0.25	−1.98	+1.58	+0.22	−0.95	+1.51	+0.08	−1.55	0.25	0.23	0.24
7(c)	+1.50	+0.32	−1.20	+1.41	+0.07	−1.38	+1.32	−0.08	−1.49	0.24	0.24	0.24
8(a)	+1.55	−0.31	−1.21	+2.61	+1.04	−0.58	+6.77	+2.57	+0.48	0.26	0.26	0.28
8(b)	+1.48	−0.34	−1.99	+1.66	+0.23	−0.98	+1.61	+0.17	−1.53	0.26	0.24	0.25
8(c)	+1.45	−0.17	−1.47	+1.53	+0.18	−1.32	+1.63	+0.36	−1.15	0.24	0.24	0.23

26 June 2023 20:00:27

Published under an exclusive license by the AVS

TABLE V. CD error and LER at the top, middle, and bottom layers of the resist in the test locations for the respective cases of Table III. For this table, the CD error is calculated as the difference between the actual and target CD instead of the absolute difference to provide an idea of the shape of the resist profile.

	CD error at feature A			CD error at feature B			CD error at feature C			LER at feature A	LER at feature B	LER at feature C
Case	Top (nm)	Middle (nm)	Bottom (nm)	Top (nm)	Middle (nm)	Bottom (nm)	Top (nm)	Middle (nm)	Bottom (nm)	(nm)	(nm)	(nm)
1(a)	+1.20	+0.05	−1.90	+4.65	+1.80	−1.55	+4.80	+2.00	−0.75	0.27	0.28	0.27
1(b)	+1.41	+0.04	−1.43	+1.29	−0.12	−2.19	+1.22	−0.31	−2.63	0.26	0.27	0.27
1(c)	+1.39	+0.24	−1.20	+1.37	+0.24	−1.20	+1.46	+0.29	−1.05	0.26	0.26	0.25
2(a)	+1.35	+0.05	−2.10	+5.15	+1.95	−1.70	+5.35	+2.20	−0.86	0.27	0.28	0.27
2(b)	+1.61	+0.04	−1.53	+1.40	−0.34	−2.82	+1.44	−0.10	−2.34	0.27	0.28	0.27
2(c)	+1.56	+0.29	−1.35	+1.64	+0.35	−1.10	+1.60	+0.31	−1.27	0.27	0.26	0.27
3(a)	+1.23	+0.30	−0.75	+2.40	+0.75	−0.40	+2.85	+1.20	−0.07	0.21	0.22	0.22
3(b)	+0.69	−0.07	−1.38	+0.73	−0.03	−1.16	+0.76	+0.08	−1.09	0.21	0.21	0.20
3(c)	+0.84	+0.05	−1.05	+0.76	−0.10	−1.08	+0.80	−0.06	−1.08	0.20	0.20	0.20
4(a)	+1.40	+0.30	−0.91	+2.73	+0.87	−0.47	+3.23	+1.38	−0.11	0.22	0.23	0.23
4(b)	+0.87	+0.06	−1.25	+0.82	−0.06	−1.58	+0.87	−0.03	−1.40	0.21	0.22	0.22
4(c)	+0.97	+0.04	−1.18	+0.90	−0.11	−1.27	+0.98	−0.04	−1.16	0.21	0.21	0.21
5(a)	+1.25	+0.30	−0.75	+2.55	+0.83	−0.43	+2.95	+1.28	−0.10	0.21	0.22	0.22
5(b)	+0.79	+0.14	−1.16	+0.79	+0.04	−1.18	+0.75	+0.02	−1.42	0.21	0.21	0.22
5(c)	+0.83	−0.10	−1.13	+0.90	+0.04	−1.06	+0.82	−0.06	−1.15	0.21	0.20	0.21
6(a)	+1.33	+0.30	−0.90	+2.78	+0.90	−0.48	+3.21	+1.38	−0.10	0.21	0.22	0.23
6(b)	+0.82	−0.04	−1.60	+0.83	−0.03	−1.38	+0.88	+0.10	−1.18	0.22	0.22	0.21
6(c)	+0.92	+0.04	−1.13	+0.88	−0.12	−1.19	+0.95	+0.09	−1.06	0.21	0.21	0.20
7(a)	+1.90	+0.40	−0.80	+3.80	+1.19	−0.41	+4.30	+1.79	−0.11	0.26	0.27	0.27
7(b)	+1.45	−0.03	−1.48	+1.46	−0.11	−1.78	+1.53	+0.15	−1.24	0.24	0.25	0.24
7(c)	+1.46	+0.18	−1.28	+1.42	+0.14	−1.33	+1.40	+0.03	−1.42	0.24	0.24	0.24
8(a)	+2.10	+0.40	−0.88	+4.20	+1.18	−0.40	+4.70	+1.85	−0.09	0.26	0.27	0.27
8(b)	+1.55	−0.10	−1.55	+1.53	−0.09	−1.78	+1.65	+0.22	−1.20	0.25	0.26	0.24
8(c)	+1.51	+0.09	−1.42	+1.60	+0.16	−1.30	+1.57	+0.26	−1.26	0.24	0.24	0.24

26 June 2023 20:00:27

exposed regions, the dose change step is not equal to B (10 nm) as in Figs. 7(b) and 7(c) where $\Delta W \neq nB$.

In Table II, the optimization results obtained with four different cost functions, i.e., $C1 = CD_error_{norm} + LER_{norm} + t_{exp_{norm}}$, $C2 = 0.75CD_error_{norm} + 0.75LER_{norm} + 1.5t_{exp_{norm}}$, $C3 = 0.5CD_error_{norm} + 0.5LER_{norm} + 2t_{exp_{norm}}$, and $C4 = 0.25CD_error_{norm} + 0.25LER_{norm} + 2.5t_{exp_{norm}}$, are compared between the old and new optimization methods where the weight for the exposing time is increased gradually from $C1$ to $C4$. The larger the weight for the exposing time, the more significant the differences in the optimal ΔW and the minimal value of cost function between the two methods. The old optimization method considers performance metrics in the fixed order prioritizing the PEC, and therefore, the procedure may end up with a local minimum of the cost function instead of the global minimum when the weight for the exposing time is equal to or larger than those for the CD error and LER. Since the new optimization procedure considers all the metrics in each iteration, it is adaptable to any cost function with different combinations of weights.

B. Large-scale pattern

In order to analyze the performance of the new optimization method for a large-scale pattern, two L/S patterns of size $12 \times 12\,\mu m^2$ were employed. In one pattern, referred to as Pattern 1, each line is $12\,\mu m$ long and 50 nm wide, and the space between lines is 50 nm. In the other pattern, referred to as Pattern 2, each line is $12\,\mu m$ long and 100 nm wide, and the space between lines is 100 nm. Each line is partitioned into segments of 100 and 200 nm in Patterns 1 and 2, respectively, and each segment is identified with a location (m, n).

In Tables III–VI, the simulation results are compared among three methods: (a) the new optimization method for a single feature (as in Sec. IV C) directly applied to each of the features individually in a large-scale pattern, (b) the old optimization method for a large-scale pattern, and (c) the new optimization method for a large-scale pattern. The single feature considered in method (a) is a segment of line in a large-scale pattern. The old and new optimization methods for a large-scale pattern determine the optimal dose distribution for a critical location in each iteration of the third step (refer to Sec. V) according to the method described in Secs. IV B and IV C,

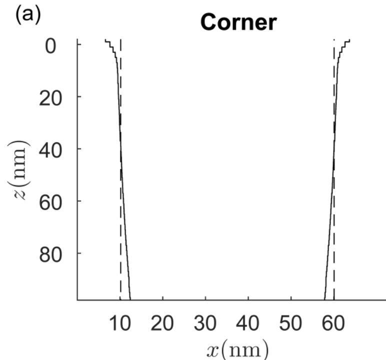
Published under an exclusive license by the AVS

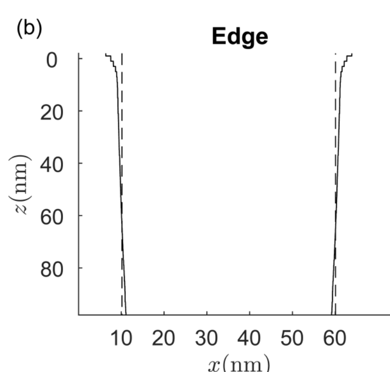
TABLE VI. Maximum absolute difference in CD among the features in the top, middle, and bottom layers of resist, i.e., $\Delta CD_{\max,t}$, $\Delta CD_{\max,m}$, and $\Delta CD_{\max,b}$, for the respective cases of Table III.

Case	$\Delta CD_{\max,t}$ (nm)	$\Delta CD_{\max,m}$ (nm)	$\Delta CD_{\max,b}$ (nm)
1(a)	7.23	3.95	2.31
1(b)	0.32	0.66	2.40
1(c)	0.14	0.12	0.43
2(a)	7.95	4.35	2.54
2(b)	0.38	0.72	2.47
2(c)	0.15	0.09	0.46
3(a)	3.22	1.82	1.40
3(b)	0.13	0.32	0.48
3(c)	0.12	0.24	0.10
4(a)	3.54	2.13	1.67
4(b)	0.14	0.32	0.68
4(c)	0.14	0.25	0.24
5(a)	3.38	1.89	1.45
5(b)	0.14	0.29	0.51
5(c)	0.14	0.23	0.19
6(a)	3.68	2.07	1.59
6(b)	0.09	0.28	0.71
6(c)	0.09	0.26	0.22
7(a)	4.92	2.89	1.54
7(b)	0.20	0.47	1.03
7(c)	0.18	0.40	0.29
8(a)	5.22	2.88	1.69
8(b)	0.18	0.57	1.01
8(c)	0.18	0.53	0.32

respectively. In the four tables, the rows with the same case number represent the results for the same optimization method, pattern, and dose distribution type.

It is observed in Table III that the new optimization method for a large-scale pattern outperforms the other two methods in minimizing the cost function. It achieves a smaller average CD error and a shorter exposing time than the other two methods. Note that the LER is low in all cases because the edge locations of the developed features are outside the exposed area ($\Delta W \neq 0$). In the old optimization method for a large-scale pattern, the PEC is completed first, followed by reducing the exposing time for each critical location. As discussed in Sec. VII A, the optimal ΔW for the PEC can be different from that for both the PEC and the reduction of exposing time. As a result, the old optimization method may end up with a local minimum of the cost function and a longer exposing time than the new optimization method. Tables IV and V present the simulation results obtained for the critical and test locations, respectively. The results show that the CD errors and LERs at the test locations are within the ranges of CD errors and LERs at the critical locations, respectively, and, hence, well justify the idea of adjusting the dose distributions at the critical locations only in the optimization method first and then computing the dose distributions at other locations according to the deconvolution surface. It is also seen in Tables IV and V that the minimum CD error, averaged over the top, middle, and bottom layers of the resist, among the critical and test locations considered, is from the old optimization methods in some cases. However, the difference in CD among the locations is larger for the old optimization methods than for the new optimization method, as shown in Table VI. The old optimization methods do not employ the $a.f.$ in determining the dose adjustment for each critical location. Without using the $a.f.$, the overadjustment of the dose distribution for one critical location can cause the dose distribution obtained for other critical locations to move further away from the optimal point. It





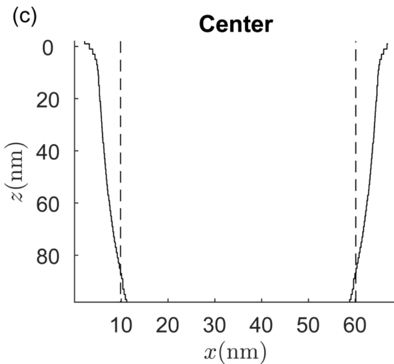


FIG. 8. Cross sections of resist profiles at the critical locations, i.e., (a) corner, (b) edge, and (c) center, obtained with the new optimization method for a single feature directly applied to each of the features individually in a large-scale pattern [case 1(a) in Table III]. In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively.

Published under an exclusive license by the AVS

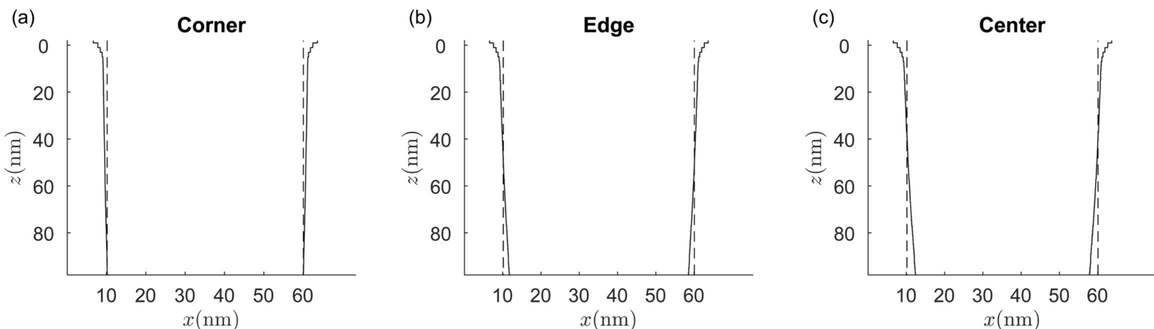


FIG. 9. Cross sections of resist profiles at the critical locations, i.e., (a) corner, (b) edge, and (c) center, obtained with the old optimization method for a large-scale pattern [case 1(b) in Table III]. In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively.

can also adversely affect the CD errors at other locations since the dose distributions at other locations are computed using the dose distributions at critical locations. This is evident from the simulation results at the test locations, as shown in Table V. The new optimization method prevents the overadjustment of dose using the $a.f.$ and leads to a better-balanced result throughout a large pattern. This is also demonstrated in Figs. 8–10, which clearly show that the new optimization method achieves a smaller average CD error and better linewidth uniformity among the features than the old optimization methods.

It is also observed from Tables III–VI that the optimization method for a large-scale pattern, either old or new, performs significantly better than the optimization method for a single feature directly applied to each of the features individually in a large-scale pattern in terms of the cost function and linewidth uniformity among the features (see Figs. 8–10). This is because the optimization method for a single feature directly applied to each feature individually in a large-scale pattern does not consider the contributions of global exposure from the other features and the variation of global exposure with the location in the pattern. As a result, this method suffers from large CD errors and a significant variation of CD among the features in a pattern. The CD error is larger at a location closer to the pattern center where the total exposure contribution from other features is the highest.

In Table VII, the simulation results with and without using the $a.f.$ in the new optimization method for a large-scale pattern are compared to evaluate the effectiveness of using the $a.f.$ It demonstrates that using an $a.f.$ achieves a smaller average CD error and a shorter exposing time than not using an $a.f.$ The main issue with not using an $a.f.$ is the recursive effect that the overadjustment of the dose distribution for one critical location can have on other critical locations in the pattern. This results in the dose distribution obtained for other critical locations moving further away from the optimal distribution, leading to a lack of linewidth uniformity among the locations. On the other hand, using an $a.f.$ avoids this overadjustment and leads to a better-balanced result throughout the entire pattern. Table VII shows that the difference in CD among the features without using the $a.f.$ is larger than that using the $a.f.$ Therefore, using an $a.f.$ in the optimization method makes it more effective, leading to more balanced results among critical locations in a pattern, a smaller minimum value of the cost function, and better overall results.

26 June 2023 20:00:27

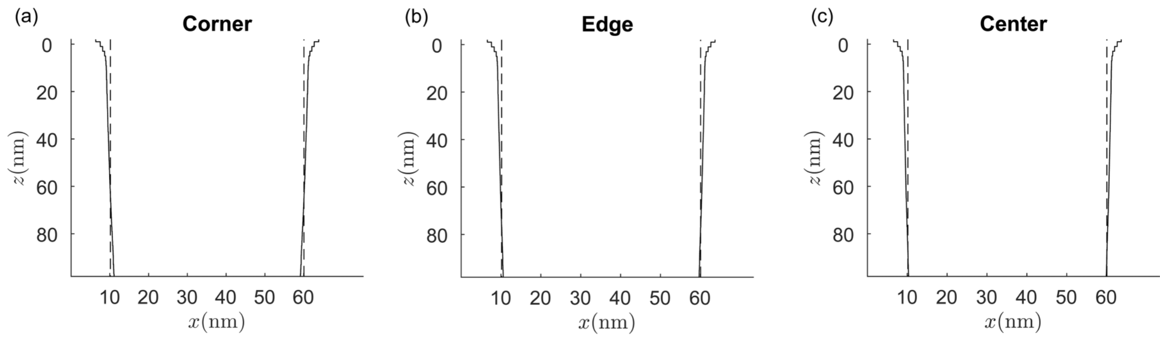


FIG. 10. Cross sections of resist profiles at the critical locations, i.e., (a) corner, (b) edge, and (c) center, obtained with the new optimization method for a large-scale pattern [case 1(c) in Table III]. In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively.

Published under an exclusive license by the AVS

TABLE VII. Optimization results with and without using the *a.f.* in the new optimization method for a large-scale pattern where the cost function $C = CD_error_{norm} + LER_{norm} + t_{exp\,norm}$. The beam interval is 200 nm, and the number of beams in each row is 21. The CD error and LER are the averages of CD errors and LERs at the critical locations, respectively. $\Delta CD_{\max,t}$, $\Delta CD_{\max,m}$, and $\Delta CD_{\max,b}$ are the maximum absolute difference in CD among the features in the top, middle, and bottom layers of resist, respectively.

a.f.	Pattern	Dose dist. type	CD error (nm)	LER (nm)	t_{exp} (cycle)	Cost function value	$\Delta CD_{\max,t}$ (nm)	$\Delta CD_{\max,m}$ (nm)	$\Delta CD_{\max,b}$ (nm)
No	1	Uniform	1.13	0.26	1239	0.830	0.21	0.42	1.52
Yes	1	Uniform	0.93	0.25	1209	0.773	0.14	0.12	0.43
No	2	Uniform	1.27	0.27	1435	0.660	0.33	0.49	1.48
Yes	2	Uniform	1.05	0.26	1400	0.623	0.15	0.09	0.46
No	1	A	0.73	0.20	1343	0.718	0.12	0.25	0.34
Yes	1	A	0.65	0.20	1309	0.695	0.12	0.24	0.10
No	2	A	0.84	0.21	1509	0.608	0.14	0.26	0.52
Yes	2	A	0.75	0.21	1462	0.586	0.14	0.25	0.24
No	1	M	0.76	0.21	1415	0.755	0.15	0.25	0.33
Yes	1	M	0.68	0.20	1379	0.719	0.14	0.23	0.19
No	2	M	0.83	0.21	1486	0.601	0.12	0.30	0.52
Yes	2	M	0.72	0.21	1449	0.579	0.09	0.26	0.22
No	1	V	1.12	0.24	1484	0.876	0.19	0.47	0.73
Yes	1	V	0.97	0.24	1445	0.836	0.18	0.40	0.29
No	2	V	1.18	0.24	1518	0.660	0.19	0.58	0.69
Yes	2	V	1.03	0.24	1481	0.634	0.18	0.53	0.32

VIII. SUMMARY

The major limiting factors in e-beam lithography are the proximity effect, LER, and writing throughput. The MPES has been considered to be a solution for improving the writing throughput significantly. In this system, there are a large number of programmable beams of which the optimal use is critical to maximizing the system's efficiency. However, the constraints of MPES compared to the single-beam systems, e.g., a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams, pose challenges in minimizing the critical dimension (CD) error and line edge roughness (LER) while maximizing the throughput. This paper proposes an adaptive optimization method designed specifically for the MPES, which considers all the performance metrics in a cost function, i.e., CD error, LER, and exposing time, in each iteration, unlike the old optimization method. The results demonstrate that the new optimization method outperforms the old optimization method in minimizing the cost function with different weights of performance metrics, making it more flexible. Also, the new optimization method is extended to large-scale patterns with uniform features, such as L/S patterns, by using a critical-location-based optimization method and an adjustment factor to reduce the recursive effect among critical locations. The simulation results show that the new optimization method achieves a smaller average CD error, a shorter exposing time, and a smaller minimum value of the cost function than the old optimization method. Notably, the new optimization method significantly improves the linewidth uniformity among features in large-scale patterns compared to the old optimization method.

The current implementation of the optimization method has been tested so far for patterns with a relatively low feature density (50%). For patterns with a higher feature density where the interdependency of exposure among features is greater, fine-tuning of the optimization procedure may be needed for achieving optimal results. Also, the application of the optimization method to large-scale patterns with nonuniform features is another potential area of future research.

26 June 2023 20:00:27

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Soo-Young Lee: Conceptualization (equal); Investigation (lead); Methodology (equal); Supervision (lead); Writing – review & editing (equal). **Md Nabid Hasan:** Investigation (equal); Software (lead); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1] S.-Y. Lee and B. D. Cook, IEEE Trans. Semicond. Manuf. **11**, 117 (1998).
[2] S. J. Wind, P. D. Greber, and H. Rothuizen, J. Vac. Sci. Technol. B **16**, 3262 (1998).

Published under an exclusive license by the AVS

[3]M. Osawa, K. Takahashi, M. Sato, H. Arimoto, K. Ogino, H. Hoshino, and Y. Machida, J. Vac. Sci. Technol. B **19**, 2483 (2001).
[4]S.-Y. Lee, S. C. Jeon, J. S. Kim, K. N. Kim, M. S. Hyun, J. J. Yoo, and J. W. Kim, J. Vac. Sci. Technol. B **27**, 2580 (2009).
[5]Q. Dai, S.-Y. Lee, S.-H. Lee, B.-G. Kim, and H.-K. Cho, J. Vac. Sci. Technol. B **30**, 06F307 (2012).
[6]C. Klein, H. Loeschner, and E. Platzgummer, J. Micro/Nanolith. MEMS MOEMS **11**, 031402 (2012).
[7]H. Matsumoto, H. Inoue, H. Yamashita, T. Tamura, and K. Ohtoshi, J. Micro/Nanolith. MEMS MOEMS **17**, 031205 (2018).
[8]A. Fay *et al.*, Proc. SPIE **9777**, 977714 (2016).
[9]C. Klein and E. Platzgummer, Proc. SPIE **9985**, 998505 (2016).
[10]S.-Y. Lee and K. Anbumony, J. Vac. Sci. Technol. B **25**, 2008 (2007).
[11]K. Ogino, H. Hoshino, and Y. Machida, J. Vac. Sci. Technol. B **26**, 2032 (2008).
[12]N. Unal, D. Mahalu, O. Raslin, D. Ritter, C. Sambale, and U. Hofmann, Microelectron. Eng. **87**, 940 (2010).
[13]M. Nagase, H. Namatsu, K. Kurihara, K. Iwadate, K. Murase, and T. Makino, Jpn. J. Appl. Phys. **35**, 4166 (1996).
[14]M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, and J.-S. Park, J. Vac. Sci. Technol. B **38**, 062603 (2020).
[15]M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, and J.-S. Park, J. Vac. Sci. Technol. B **40**, 032602 (2022).
[16]E. Seo, B. K. Choi, and O. Kim, Microelectron. Eng. **53**, 305 (2000).
[17]M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, S.-B. Kim, and C.-U. Jeon, J. Vac. Sci. Technol. B **37**, 061609 (2019).
[18]X. Zhao, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, C.-U. Jeon, B.-G. Kim, and H.-K. Cho, Microelectron. Eng. **133**, 78 (2015).
[19]Q. Dai, S.-Y. Lee, S.-H. Lee, B.-G. Kim, and H.-K. Cho, J. Vac. Sci. Technol. B **29**, 06F314 (2011).
[20]D. Drouin, A. R. Couture, D. Joly, X. Tastet, V. Aimez, and R. Gauvin, Scanning **29**, 92 (2007).
[21]L. D. Jackel, R. E. Howard, P. M. Mankiewich, H. G. Craighead, and R. W. Epworth, Appl. Phys. Lett. **45**, 698 (1984).
[22]Q. Dai, R. Guo, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, and C.-U. Jeon, Microelectron. Eng. **127**, 86 (2014).
[23]M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, and J.-S. Park, J. Vac. Sci. Technol. B **38**, 062601 (2020).
[24]X. Zhao, Q. Dai, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, and C.-U. Jeon, J. Vac. Sci. Technol. B **32**, 06F508 (2014).

26 June 2023 20:00:27

Published under an exclusive license by the AVS