Application of neural network to controlling three-dimensional electron-beam exposure distribution in resist

C. Guo and S.-Y. Lee[a)]
Department of Electrical and Computer Engineering, Auburn University, Auburn, Alabama 36849

S. H. Lee,[b)] B.-G. Kim, and H.-K. Cho
Samsung Electronics Co., Ltd., Photomask Division, Hwasung, Kyunggi-Do 446–711, Korea

(Received 21 July 2009; accepted 15 October 2009; published 1 December 2009)

Electron-beam (e-beam) lithography is often employed for the fabrication of binary patterns with nanoscale features and grayscale structures. In such applications, exposure distribution along the depth dimension of the resist layer can significantly affect the geometry of the written patterns or structures. Most of the previous e-beam dose control schemes adopted a two-dimensional exposure model, ignoring the depth-dependent exposure variation. In this paper, a method which utilizes a neural network for explicitly controlling three-dimensional (3D) (vertical as well as lateral) distribution of exposure is described. A neural network allows one to achieve more balanced exposure distributions compared to a conventional region-by-region recursive approach. Through an extensive computer simulation, performance of the proposed approach to controlling (3D) exposure distribution has been analyzed in detail. © *2009 American Vacuum Society.*
[DOI: 10.1116/1.3259959]

I. INTRODUCTION

One of the major factors that limit the feature resolution in a pattern is the proximity effect caused by electron scattering, leading to the "nonideal" exposure (energy deposited at a point) distribution in the resist. In order to be able to achieve a high resolution of nanoscale, it is essential to find proper methods to compensate for it. Almost all proximity effect correction schemes developed so far are based on two-dimensional (2D) models where the exposure variation along the resist depth dimension is ignored.[1–3] However, when the feature size is reduced down to nanoscale or a certain type of sidewall of the (remaining) resist profile is desired, three-dimensional (3D) exposure distribution needs be considered in order to have more explicit control of the resist profile. Note that the resist profile estimated using a 2D model can be significantly different from the actual 3D resist profile.[4] The goal of the 3D proximity effect correction is to determine the 2D dose (energy given to a point at the resist surface) distribution which achieves a certain target 3D exposure distribution in the resist. Its inherent difficulty is that one has only 2D controllability (dose) for 3D constraints (exposure).

In the past, there were efforts to utilize neural networks (NNs) for proximity effect correction.[5] A more recent study focused on a rigorous mathematical approach to enhancing edge (exposure) contrast.[6] However, in these studies, a 3D exposure model, where depth-dependent exposure variation is taken into account by using a 3D point spread function (PSF), was not adopted. It has been shown that this exposure variation along the depth dimension can have a significant effect on the type of feature sidewall in the resist profile and the achievable feature density.[4,7]

The main focus of this study is on controlling the 3D exposure distribution, i.e., given a target exposure distribution, find a dose distribution to achieve it. Hence, throughout this article, the term "correction" refers to the process of computing a dose distribution to achieve a given target exposure distribution. An earlier study on region-by-region correction has shown that 3D exposure distribution may be controlled to some extent.[7] However, such an approach has a limitation due to the recursive effect of sequential processing of regions and the characteristic of 3D correction problem, i.e., overdetermined. Hence, it is difficult to find a global or even areawise optimal solution (doses computed). Also, the solution tends to be spatially unbalanced. A possible approach to overcoming the limitation is to employ a neural network of manageable size to determine doses of regions in a window at a time, instead of one region at a time. The size of the window is to be carefully chosen to avoid an excessive amount of computation and achieve an acceptable level of accuracy.

In this article, a neural network approach to optimizing the dose distribution to achieve a target 3D exposure distribution is described, and it is compared to the region-by-region dose control[7] and also a 2D correction method.[8] Through computer simulation, it has been shown that the neural network approach allows one to have a better control of 3D exposure distribution in the resist.

The remainder of this article is organized as follows. In Sec. II, the task of the 3D correction is introduced. In Sec. III, a region-by-region dose control scheme implemented earlier is reviewed to be compared with the proposed method. In Sec. IV, the proposed neural network approach to

[a)]Electronic mail: leesooy@eng.auburn.edu
[b)]Electronic mail: sanghee27.lee@samsung.com

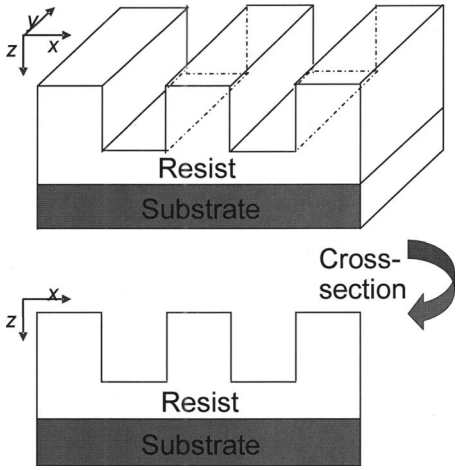


FIG. 1. 3D PSF for the substrate system of 500 nm PMMA on Si with the beam energy of 50 keV: (a) all layers, and (b) the top, middle, and bottom layers.

3D exposure control is described. In Sec. V, simulation results and discussion are provided, followed by a summary in Sec. VI.

II. 3D CORRECTION

A. Exposure model

A 3D substrate system is illustrated in Fig. 1. Circuit features considered in this study are lines of which lengths are long enough in the Y-dimension that exposure may be assumed not to vary along the Y-dimension. Then, it is sufficient to consider only the cross section (the X-Z plane) of the substrate system for correction, as shown in Fig. 1.

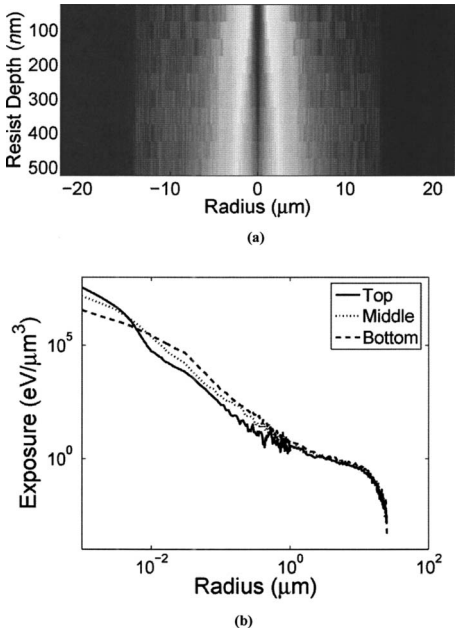


FIG. 2. 3D substrate system and its cross section: features (such as lines) are assumed to be long in the Y-dimension so that only the cross section perpendicular to the Y-axis can be considered.

In a 2D exposure model, exposure variation along the depth dimension in the resist (Z-axis in Fig. 1) is ignored. The 3D PSF $h(x,y,z)$ in Fig. 2 is averaged (integrated) over the Z-dimension to get the 2D PSF $h(x,y)$. A PSF describes the exposure distribution in the resist when a point is exposed. The 2D exposure distribution $e(x,y)$ can be computed by convolving a 2D dose distribution with the 2D PSF. Hence, any proximity effect correction scheme, which is based on a 2D model, does not have any explicit control of the exposure distribution along the resist depth dimension. However, exposure does vary significantly with depth in the resist, as illustrated in Fig. 2. In a 3D exposure model, a 3D PSF $h(x,y,z)$, which depicts the exposure distribution in the depth dimension as well as the lateral (X,Y) dimensions, is employed. By convolving the dose distribution at the top surface of the resist, $f(x,y,0)$, with the 3D PSF, a 3D exposure distribution $e(x,y,z)$ is obtained as in Eq. (1). For true 3D proximity effect correction, $e(x,y,z)$ instead of $e(x,y)$ must be referred to for explicit control of the 3D resist profiles of the written patterns,

$$
\begin{aligned}
e(x,y,z) &= \int_{x'}\int_{y'}\int_{z'} \mathrm{PSF}(x-x',y-y',z-z') \\
&\quad \times f(x',y',0)\,dx'\,dy'\,dz' \\
&= \int_{x'}\int_{y'}\int_{z'} \mathrm{PSF}(x-x',y-y',z-z') \\
&\quad \times f(x',y')\,\delta(z')\,dx'\,dy'\,dz' \\
&= \int_{x'}\int_{y'} \mathrm{PSF}(x-x',y-y',z) f(x',y')\,dx'\,dy'. \quad (1)
\end{aligned}
$$

Since features are assumed to be long in the Y-dimension, $e(x,z)$ instead of $e(x,y,z)$ will be considered for 3D correction in the remainder of this article.

B. Dose control

In a conventional 2D proximity effect correction, given a target exposure distribution $e(x,y)$ and a PSF $h(x,y)$, the dose distribution $f(x,y)$ is computed, which minimizes the difference between the target and the actual exposure distributions. Dose is controlled in the 2D space, i.e., the top surface of resist, and target exposures (constraints) are given in the 2D space, which makes the optimization problem of 2D proximity effect correction more solvable (than in 3D correction). However, in true 3D proximity effect correction, constraints are specified in the 3D space, while dose controllability still remains in the 2D space. That is, tighter constraints are required for the 3D correction, i.e., an overdetermined system. The resist on the substrate may be viewed as a stack of layers. A certain target exposure distribution in each layer needs to be satisfied by a dose distribution, which is common to all layers. Obviously, the dose distribution, which is optimal for a layer, may not be so for another layer, and therefore some kind of compromise must be made among the layers in order to minimize the difference between the target and the actual exposure distributions.

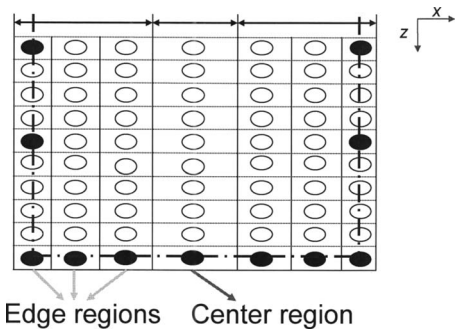


FIG. 3. Each feature is partitioned into regions and all resist layers (in the cross section) are accordingly partitioned. Each partition is assigned a critical point (oval) at which a target exposure can be specified for 3D exposure control. The filled ovals are the critical points where target exposures are specified.

III. REGION-BY-REGION CORRECTION (DOSE CONTROL)

In the previous study,[7] a 3D proximity effect correction scheme, which corrects one region at a time, was designed and implemented. Each feature in a circuit pattern is partitioned into a set of regions (refer to Fig. 3) for each of which a dose is to be determined such that the deviation from a given target exposure distribution is minimized. A target exposure distribution is specified at a set of "critical points" throughout the resist. Initially, all regions in the pattern are assigned with a single dose and the dose of each region is updated through iterations. In each iteration, all regions are sequentially corrected one region at a time. A new dose for each region is computed such that exposure errors at the critical points of multiple layers right below the region are minimized [refer to Fig. 4(a)]. The sum of exposure errors at the multiple critical points (layers) for a region, instead of individual errors, may be minimized as a way to handle the conflict among the layers. However, the optimal dose for a region may not be optimal for the neighboring regions. Hence, this region-by-region correction is repeated until region doses converge and the average exposure error is reduced to an acceptable level. The fundamental drawback of this correction procedure is that only one region is considered at a time, and therefore the relationship among the neighboring regions cannot be taken into account during correction (refer to Fig. 5). As a result, the solution (doses) is only locally (in fact regionwise) optimized and may not be spatially balanced. Therefore, it is expected that a more global procedure such as neural networks may be able to improve the solution substantially.

IV. NN APPROACH

A. NN for solving linear equations

A neural network is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionistic approach to computation. There are several different types of neural networks employed in various applications. For proximity

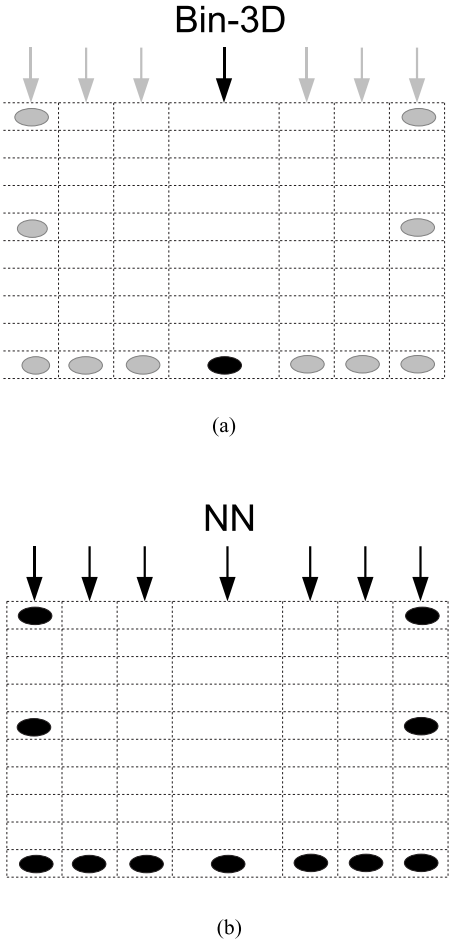


FIG. 4. (a) Bin-3D: one region and target exposures associated with it are considered together. (b) NN: all regions and all target exposures are considered together. The doses (arrows) and target exposures (ovals) that are active in a correction step are highlighted in black (the inactive ones in gray). In this illustration, only the target exposure is active in determining the dose for the center step in Bin-3D.

effect correction, the neural network designed for solving a system of linear equations, $Ax=b$, using the least-squares method is employed, noting that the relationship between dose and exposure is linear. The layout of the neural network is shown in Fig. 6.[9] The solution vector x, starting from a certain initial solution, is updated through iterations according to the learning rule $dx/dt=-\mu\varepsilon$, where the error vector $\varepsilon=A^TAx-A^Tb$, which is applicable to an overdetermined system.

B. Application to 3D correction

It should be clear that the neural network for solving a system of linear equations, described above, is readily applicable to proximity effect correction, especially 3D correction where the system is overdetermined, i.e., more constraints (target exposures) than unknowns (doses). In $Ax=b$, element x_i of the solution vector x (to be distinguished from the variable x for X-axis) corresponds to the dose of the ith region and element b_j of the vector b corresponds the jth constraint (target exposure). Each element a_{ij} of A, called weight ma-

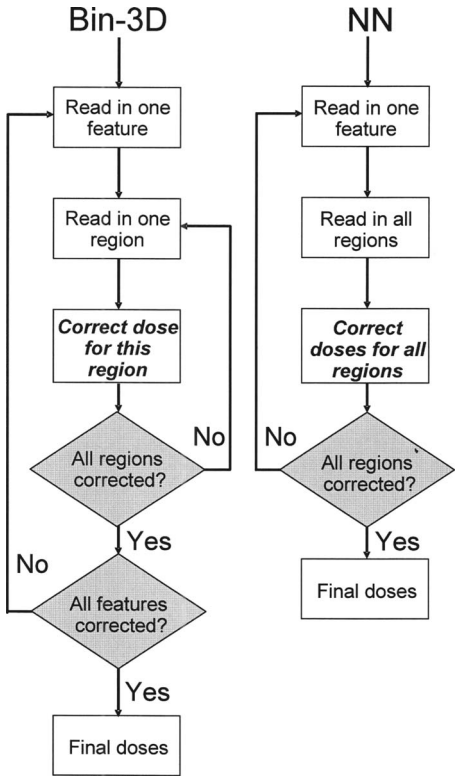


FIG. 5. Flowcharts of the two correction procedures; Bin-3D and NN.

trix, quantifies how much b_j is changed due to a unit change in x_i, i.e., $\Delta e_j/\Delta x_i$. The matrix A or a_{ij} depends on the PSF and geometric relationship among features in a circuit pattern. Now, finding the dose distribution that can achieve the given target exposure distribution minimizing exposure error is equivalent to solving $Ax=b$, minimizing the cost function,

$$C(x) = \|A^T Ax - A^T b\|^2 = \|\varepsilon\|^2 = \varepsilon^T \varepsilon. \quad (2)$$

Note that ε is a vector consisting of exposure errors at the critical points. The exposure error at the critical point i is defined as $\varepsilon_i = e_i - b_i$, where e_i is the exposure at the critical point i.

As is well known, it is possible to obtain a solution including negative doses when solving $Ax=b$. However, one may avoid it by properly selecting a realistic target exposure

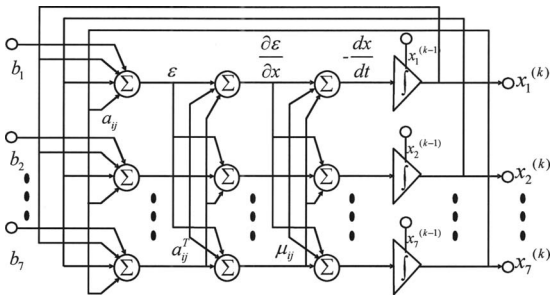


FIG. 6. Neural network for solving linear equations $Ax=b$ illustrated for seven unknowns where $a_{ij}=\Delta e_j/\Delta x_i$, a_{ij}^T:a_{ij} in A^T, $x_i^{(k)}$: dose for region i computed in the kth iteration, ε: exposure error vector $\{e_i-b_i\}$, μ_{ij}: learning rate for region i and critical point j.

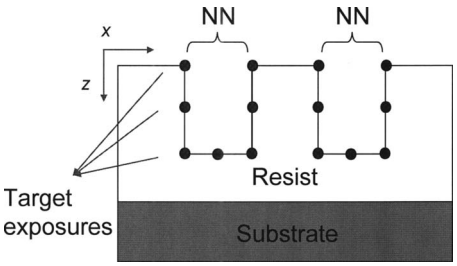


FIG. 7. NN is applied to each feature at a time with a set of target exposures specified over multiple layers of resist.

distribution or solving the equations with a constraint of nonnegative x, i.e., constrained optimization. The former approach was taken in this study.

C. Implementation

One potential drawback of using a neural network is a long computation time, especially when the network size that depends on the number of unknowns (doses or regions) to be considered together is large. On the other hand, a higher quality of solution calls for a larger number of regions to be corrected together. Therefore, a compromise between the solution quality and computational requirement must be made by choosing a proper size of the window, all regions within which are corrected together [refer to Figs. 4(b) and 5]. In the current implementation, the size of the window is set to be that of each feature, i.e., all regions generated from a feature are corrected together. Given a window, element a_{ij} in the weight matrix A is determined by computing the exposure change at the center of the jth region due to the unit change in the dose for the ith region. In order to ensure a proper convergence of the solution, elements of the learning matrix u are set to sufficiently small values. The neural network described earlier is applied to each window location, correcting all regions in the respective feature simultaneously. Nevertheless, there still exists sequentiality among features (refer to Fig. 7), which cannot be avoided unless all features in a pattern are corrected at the same time (which is practically impossible unless a pattern is relatively small in size). In an effort to take interfeature dependency into account and obtain a balanced solution, this correction procedure is repeated multiple times for the entire pattern.

V. RESULTS AND DISCUSSION

Performance of the proposed approach, "NN," has been analyzed through an extensive simulation, by comparing it to other methods, i.e., a 2D method, "GP" (GPYRAMID), and the region-by-region 3D correction method, "Bin-3D" (refer to Sec. III). GPYRAMID is a 2D correction method that does not take the variation in exposure along the depth dimension into account.[8] The three methods are compared in terms of the 3D exposure profiles achieved, how close to the target exposure profiles they are. Three different substrate

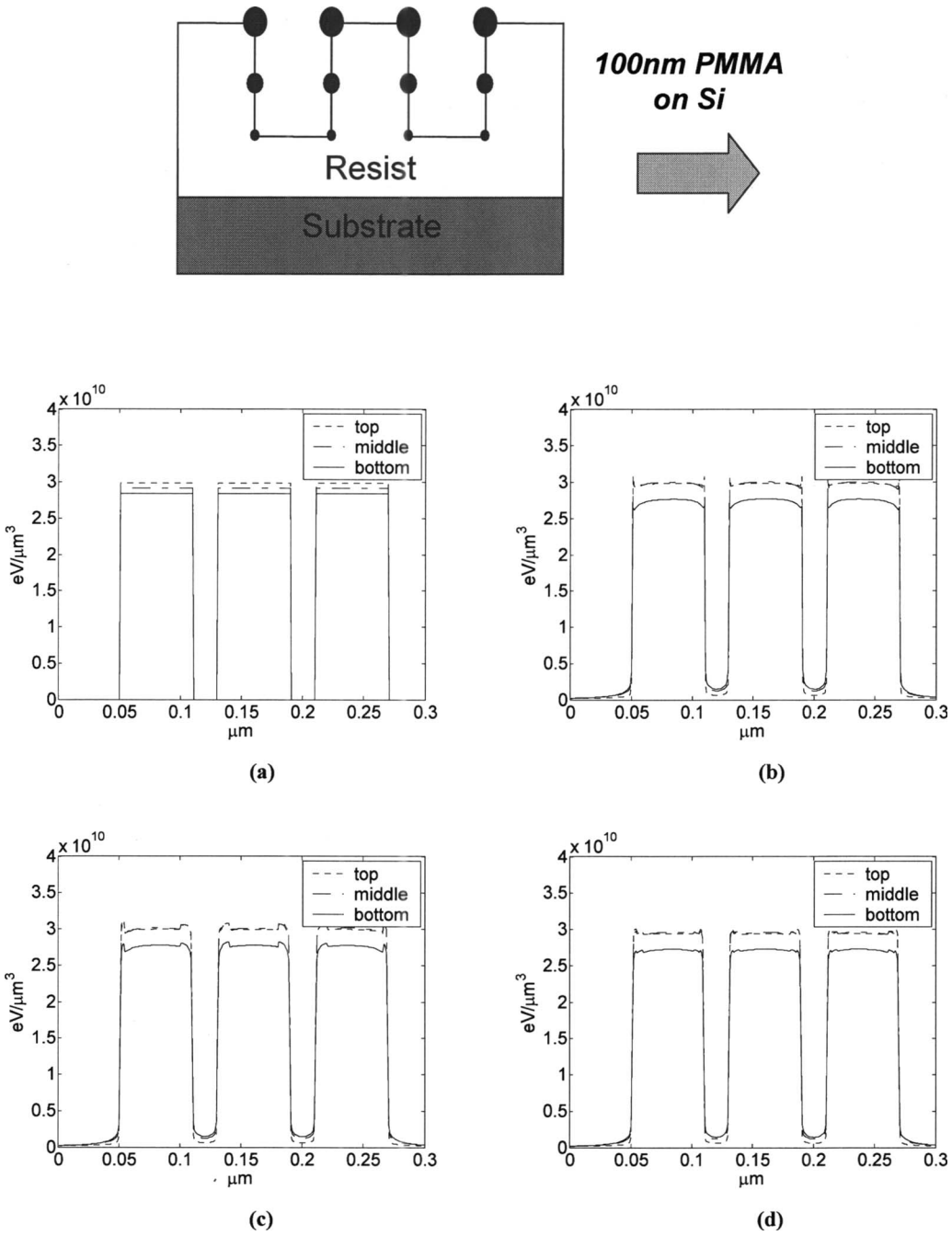


FIG. 8. Exposure distributions at the top, middle, and bottom layers on 100 nm PMMA on Si: (a) taget exposures, (b) GP, (c) Bin-3D, and (d) NN.

systems have been considered, i.e., 100, 500, and 1000 nm PMMA on Si with the beam energy of 50 keV. The test circuit patterns considered in the simulation study include a single line, multiple lines (lines and spaces), etc. The results for three-line patterns are provided in Figs. 8–10 where the target exposures are specified along the boundaries of each line in the 3D resist space as shown in each figure, specifically the top, middle, and bottom layers of the resist. For the 2D method, GP, the target exposures are averaged along the depth dimension and the average target exposures are referred to for correction.

In all three cases, the main difference among the exposure profiles achieved by the three methods lies in the edge regions. The 2D method (GP) does not consider the depth-dependent exposure variation and correction is done such that the average exposure errors (averaged over the three layers) are minimized. Therefore, it leads to a high overshoot of exposure at the top layer, which is canceled by an undershoot at the bottom layer in computing the average exposure errors. The region-by-region method (Bin-3D) better balances exposure profiles in the edge regions within a line than GP, but not among the lines. Note that the left and right lines

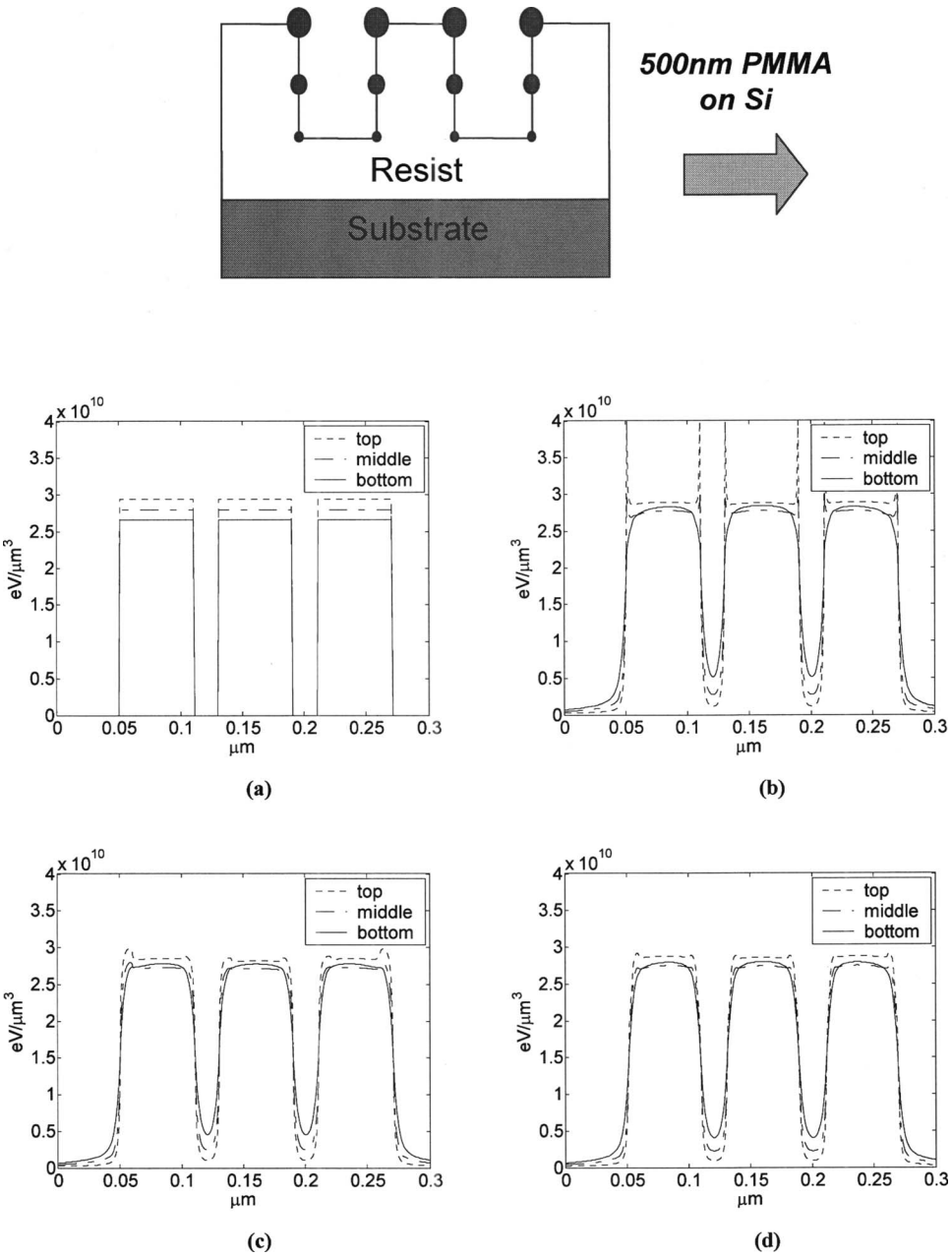
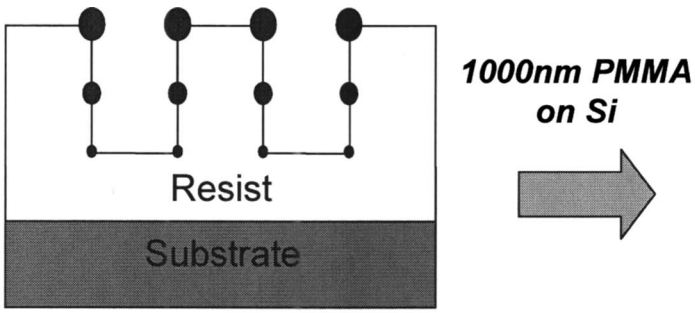


FIG. 9. Exposure distributions at the top, middle, and bottom layers on 500 nm PMMA on Si: (a) taget exposures, (b) GP, (c) Bin-3D, and (d) NN.

have higher overshoots of exposure than the center line. The NN method achieves the exposure profiles which are better balanced both within each line and among lines. That is, the exposure distribution within each line is symmetric with respect to its center, and all three lines show almost the identical exposure distribution. Also, the exposure distributions achieved by the NN method are closer to the target exposures (refer to Table I). The NN method corrects all regions within a line together, but one line at a time, in order to keep the computational complexity at a practical level for a large pattern. It is seen from these results that the recursive effect due to the sequential processing of lines is effectively offset through iterations over the entire pattern.

It needs to be pointed out that the exposure profiles at the center regions of lines (the average exposure level except for edge regions in each layer) do not show a substantial difference among the three methods. This is due to the fact that the center region of each line is larger than the edge regions and a single dose is assigned to the center region. Since there is no spatial control of dose within the center region, the exposure distribution along the depth dimension is largely determined by the 3D PSF (its layer-by-layer energy distribution). A finer partitioning of the center region would enhance the controllability of center exposure distribution over layers.

In Table I, the average and maximum absolute errors at the critical points are provided for more quantitative com-



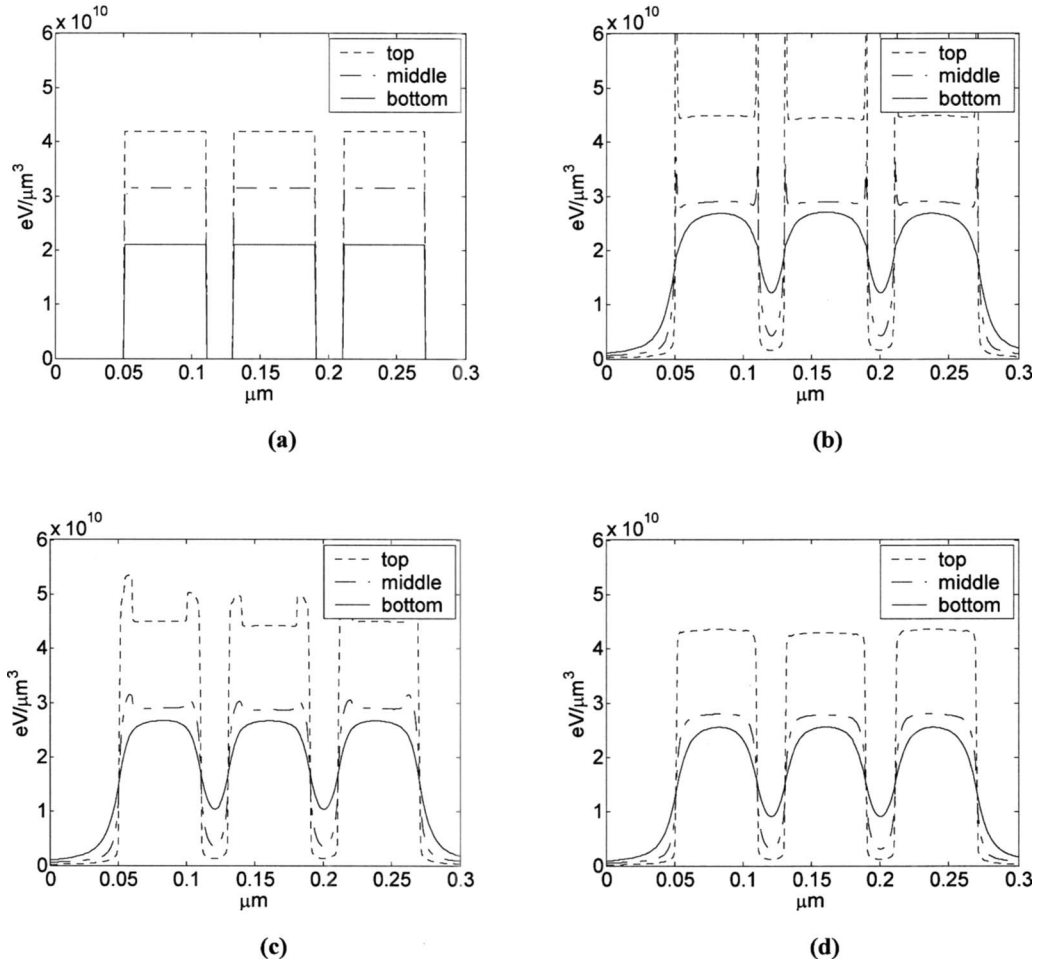


FIG. 10. Exposure distributions at the top, middle, and bottom layers on 1000 nm PMMA on Si: (a) taget exposures, (b) GP, (c) Bin-3D, and (d) NN.

TABLE I. Average and maximum absolute errors with respect to target exposures.

	100Si			500Si			1000Si		
	GP	Bin-3D	NN	GP	Bin-3D	NN	GP	Bin-3D	NN
Abs_Error (%)	6.00	5.00	3.00	16.00	9.50	7.00	27.00	14.00	10.00
Max_Error (%)	7.50	6.20	4.50	50.70	13.70	11.70	80.00	37.00	34.00

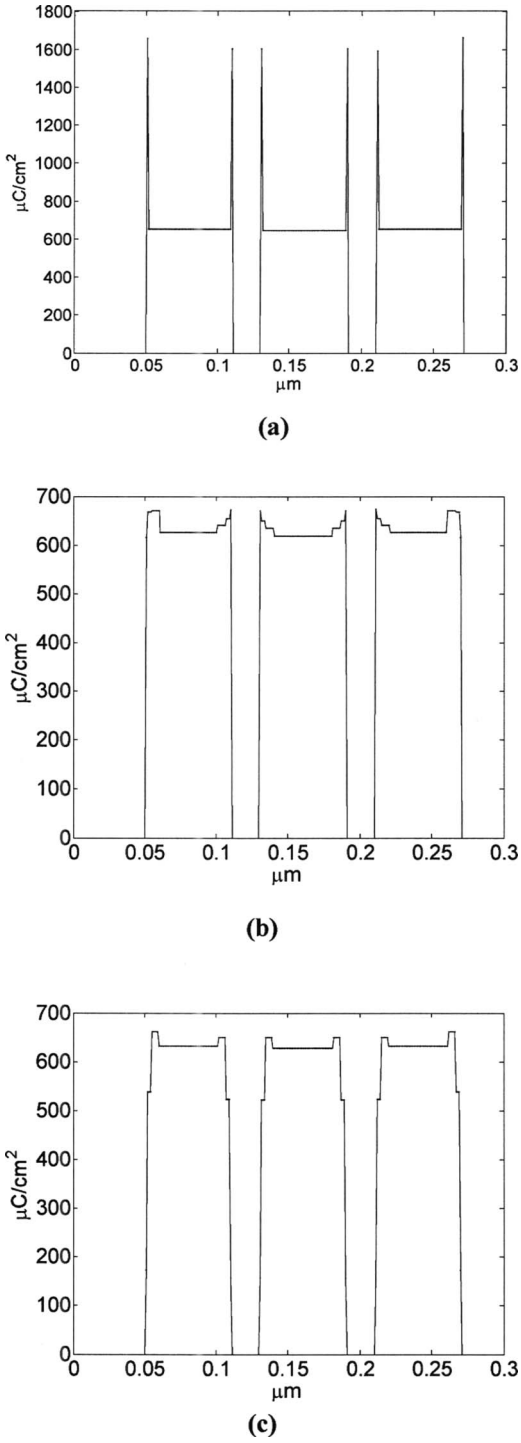


FIG. 11. Dose distributions corresponding to the exposure distributions in Fig. 9: (a) GP, (b) Bin-3D, and (c) NN.

parison of the three methods. The errors are computed as percent errors with respect to the target exposures. It is clear that both average and maximum errors are the largest for GP and smallest for NN. In particular, the 2D method GP has a much larger maximum error than the other two 3D methods in each case since it only considers the exposure distribution averaged over all layers without paying attention to the individual layers. This observation is particularly so for a thicker resist for which the exposure variation along the depth dimension is larger, and therefore a 2D method would be less accurate.

In Fig. 11, the spatial dose distributions computed by the three methods, which resulted in the exposure distributions shown in Fig. 9, are provided. GP considers only the average (among layers) exposure and ends up with very high doses in the edge regions. The doses computed by NN are different from those by Bin-3D (and GP) mostly in the edge regions, which led to the better balanced exposure distributions as discussed earlier.

VI. SUMMARY

In this study, controllability of the 3D exposure distribution in the resist by spatial control of dose has been addressed, which is an important issue in increasing feature resolution or circuit density by minimizing 3D proximity effect. Specifically, a 2D method and two 3D methods including the proposed neural network approach have been compared in terms of 3D exposure distributions achieved. It is seen that the 3D methods, especially NN, achieve an exposure distribution closer to the target one in the edge regions. However, the (exposure) differences at the center regions among the three methods are not significant (since dose is not controlled spatially within the large center region in the current implementation). NN is able to better balance the 3D exposure distribution within each feature and among features. These results must be helpful to future efforts in the area of 3D exposure control and proximity effect correction in e-beam lithography.

ACKNOWLEDGMENT

This work was supported by a research grant from Samsung Electronics Co., Ltd.

[1]S. J. Wind, P. D. Gerber, and H. Rothuizen, J. Vac. Sci. Technol. B **16**, 3262 (1998).
[2]D. P. Kern, in *Proceedings of the Ninth International Conference on Electron and Ion Beam Science and Technology*, edited by R. Bakish (Electrochemical Society, Princeton, NJ, 1980), p. 326.
[3]U. Hofmann, C. Kalus, A. Rosenbusch, R. Jonckheere, and A. Hourd, Proceedings of SPIE Conference on E-Beam, X-ray, EUV, and Ion-Beam Sub-micrometer Lithographies for Manufacturing VI, March 1996 (unpublished), Vol. 2723, p. 150.
[4]S.-Y. Lee and K. Anbumony, Microelectron. Eng. **83**, 336 (2006).
[5]K. D. Cummings, R. C. Frye, and E. A. Rietman, Appl. Phys. Lett. **57**, 1431 (1990).
[6]M. Peckerar, D. Sanders, A. Foli, A. Srivastava, and U. Vishkin, J. Vac. Sci. Technol. B **25**(6), 2288 (2007).
[7]K. Anbumony and S.-Y. Lee, J. Vac. Sci. Technol. B **24**, 3115 (2006).
[8]F. Hu and S.-Y. Lee, J. Vac. Sci. Technol. B **21**, 2672 (2003).
[9]Fredric M. Ham and Ivica Kostanic, *Principles of Neurocomputing for Science and Engineering* (McGraw-Hill, New York, 2001).