Dose control for fabrication of grayscale structures using a single step electron-beam lithographic process

Fei Hu and Soo-Young Lee[a)]
Department of Electrical and Computer Engineering, Auburn University, Auburn, Alabama 36849

(Received 1 July 2003; accepted 29 September 2003; published 4 December 2003)

Many optoelectronic devices include multidimensional (grayscale) semiconductor structures such as diffractive optical elements (DOE's), blazed gratings, and photonic band gap (PBG) crystals. Their performance is known to be highly sensitive to dimensional accuracy of features in the multidimensional structure. Therefore, it is essential to have a tight control of the feature size in the fabrication process. *Grayscale* electron(e)-beam lithography enables a single step lithographic process for fabrication of such multidimensional structures and therefore has several advantages over using a binary lithographic process multiple times. As the feature size decreases, proximity effect in the e-beam lithographic process can make dimensions of the written features in a device substantially different from the ideal ones so that performance of the device is significantly degraded. In this paper, an efficient dose control scheme for grayscale e-beam lithography, which attempts to minimize the difference between the spatial exposure distribution and the "shape" of a multidimensional structure, is described along with results from an extensive computer simulation.
© *2003 American Vacuum Society.* [DOI: 10.1116/1.1627808]

I. INTRODUCTION

Many optoelectronic devices include multidimensional (grayscale) semiconductor structures such as diffractive optical elements (DOE's), blazed gratings, and photonic band gap (PBG) crystals.[1,2] Their performance is known to be highly sensitive to dimensional accuracy of features in the multidimensional structure. Therefore, it is essential to have capability of controlling the feature size accurately, in order to be able to fabricate such devices with the desired performance properties. One of the ways to fabricate such structures is to rely on electron-beam (e-beam) lithography. However, as the feature size decreases, proximity effect due to electron scattering can make dimensions of the written features in a device substantially different from the ideal ones so that its performance is significantly degraded.

For fabricating grayscale structures, *grayscale lithography* can be employed (refer to an illustration in Fig. 1). A substrate is partitioned into a number of "features," each of which is to be etched to a specified depth. A different depth of a feature is achieved by a different thickness of the remaining resist, which in turn depends on the exposure level within the feature. Therefore, it is critical to control feature-wise exposure distribution precisely. Another way to fabricate such grayscale structures is to use the binary lithographic process multiple times, i.e., once for each depth. However, this approach has a few drawbacks, especially when the number of different depths is much greater than two: a longer total process time, the alignment problem between processes, a higher probability for lower yield due to the complicated process, and a higher cost. Grayscale lithography can eliminate such drawbacks since it requires only one lithographic process. However, there are more stringent requirements in grayscale lithography than in binary lithography. Not only the blurring between features is to be corrected, but also exposure level within each feature should be as close to the desired level (to be referred to as *ideal exposure*) of the feature, i.e., deviation from the ideal exposure is to be minimized.

All of the existing proximity effect correction schemes are for *binary lithography* where one only concerns whether each feature in the resist is developed or not.[3–11] In this paper, an efficient dose control scheme for grayscale e-beam lithography is described and results from an extensive computer simulation are presented. It is the first proximity effect correction scheme ever developed for grayscale lithography. Minimization of exposure variation within each feature requires sufficient control of spatial dose distribution. However, too fine a spatial control of dose, such as a pixel-by-pixel control, would be too time-consuming to be practical for realistic-size circuit patterns. In the proposed dose control scheme, each feature is partitioned into regions for region-wise dose control, where the number and shapes of regions are judiciously selected in order to minimize correction time and not to sacrifice correction accuracy.

The remainder of this paper is organized as follows: Sec. II depicts how circuit features are partitioned for dose control. Section III describes an iterative scheme to derive dose distribution within each feature. Section IV provides simulation results, followed by a summary in Sec. V.

II. DECOMPOSITION OF CIRCUIT FEATURES

In binary lithography, the main concern is minimization of CD error and the spatial exposure distribution within each feature is not very critical as long as it is above (in the case of a positive resist) the developing threshold. However, in the case of grayscale lithography, exposure level within a feature is to be as close to the ideal exposure of the feature as

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

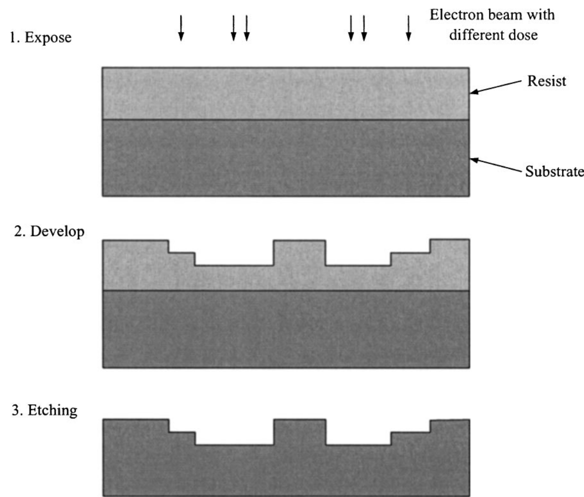


FIG. 1. Illustration of the grayscale lithographic process.

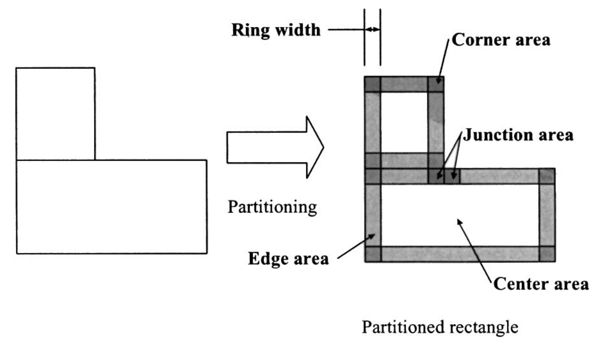


FIG. 2. First step of region partitioning where each feature (rectangle) is divided into center, corner, edge, and junction areas.

possible. This requirement in addition to the CD error minimization makes it necessary to have a fine control of dose in each feature. Hence, each feature (currently only *rectangular* features are supported) is partitioned into a set of regions where a dose is to be determined for each region, to be referred to as *region partitioning*. That is, a region is the smallest spatial unit for controlling dose, and a *critical point* is set up at the center of region (refer to Sec. III).

Depending on the size of a feature (a rectangle) and its relationship with neighboring or adjacent features, a feature may be divided into smaller features, to be referred to as *rectangle partitioning*, each of which is partitioned into regions.

A. Rectangle partitioning

For a rectangle that has neighboring rectangles, exposure contributions from neighbors may cause an unbalanced exposure distribution within the rectangle, when it is given a single dose. Therefore, such a rectangle needs to be partitioned to gain more control of dose, and hence balance the exposure distribution within the rectangle. In Ref. 11, an adaptive rectangle partitioning scheme, which considers only the geometric relationship among adjacent rectangles, was implemented for binary lithography. The scheme has been augmented for grayscale lithography, by taking into account the ideal exposures of adjacent rectangles as well as the geometric relationship among them. A rectangle is partitioned where ideal exposures of adjacent rectangles are significantly different from each other.

B. Region partitioning

From the viewpoint of accuracy (minimizing error between the ideal exposure and actual exposure achieved), a pixel-by-pixel dose control may be preferred. However, it would require too long a computation time to be practical even for a relatively small circuit pattern. Also, it is unnecessary to have such a fine spatial control of dose. Note that the ideal exposure is constant within each feature (rectangle) and therefore dose does not need to be varied much in the center area of a rectangle. Region partitioning is to be optimized such that the number of regions is minimized without sacrificing accuracy.

Region partitioning conceptually consists of two steps. In the first step, a rectangle is divided into a *ring* area and a *center* area. The width of a ring area, referred to as *ring width*, is adjust able depending on the sharpness of the PSF (point spread function describing the spatial energy deposition in the resist when a point is exposed). The ring area is further divided into *corner, edge*, and *junction areas* as illustrated in Fig. 2. Each of *corner area* and *junction area* is always a square. In the second step, an area is partitioned into one or more regions.

1. Center area

In the center area of each rectangle, it is not necessary to vary dose much spatially. Therefore, in most cases, the center area does not need to be further partitioned. However, when the ideal exposures of adjacent rectangles are much different from each other, the center area may be partitioned. The current implementation allows a center area to be divided horizontally and/or vertically (once in each direction), generating up to 4 regions of the same size. For example, if the left and right neighbors of a rectangle have significantly different ideal exposures, its center area is partitioned vertically into two halves (regions).

2. Edge area

An edge area is divided into one or more regions depending on the partitioning mode. Two partitioning modes may be considered, *fixed partitioning* and *adaptive partitioning*. In the fixed partitioning mode, edges are partitioned into a fixed number of regions of predetermined widths as illustrated in Fig. 3(a). Therefore, an edge area may be divided unnecessarily. For example, when two adjacent rectangles have the same or very similar ideal exposures, edge areas between them may not need to be partitioned. In the adaptive partitioning mode, two options are available. One is *simple adaptive partitioning*. In this option, as shown in Fig. 3(b), an edge area is not partitioned when the *exposure step* (difference of the ideal exposure between two adjacent rectangles)

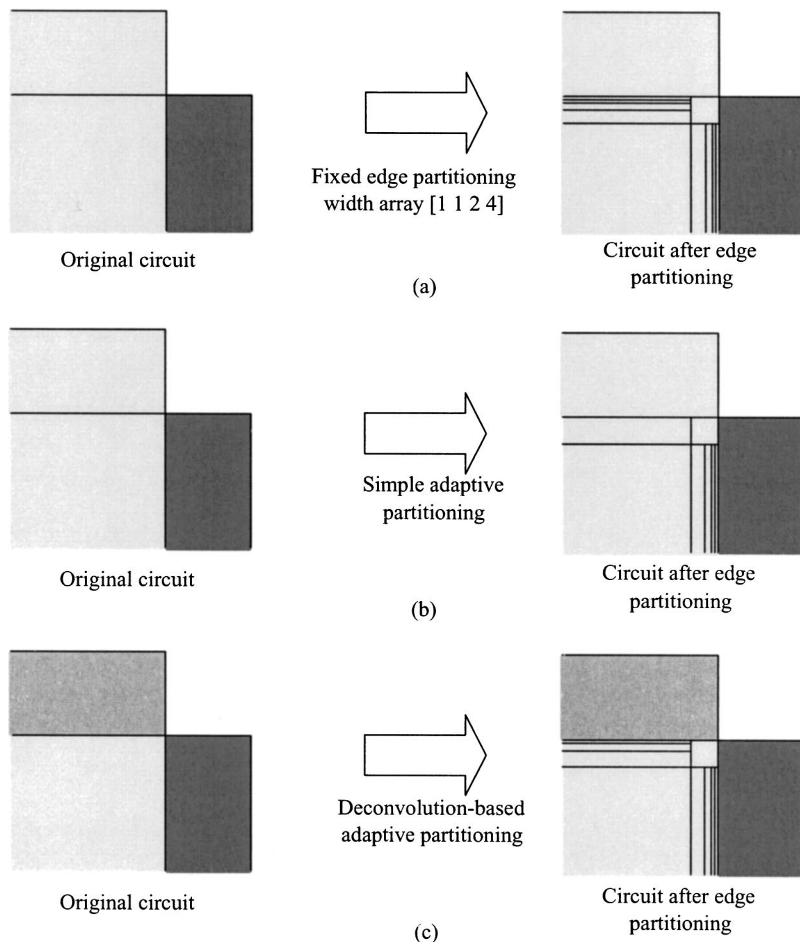


FIG. 3. Partitioning of edge areas: (a) fixed partitioning, (b) simple adaptive partitioning, (c) deconvolution based adaptive partitioning. The gray level of each feature represents its ideal exposure. Some regions are not shown.

at the edge boundary is zero. In other cases, the fixed partitioning is used. Note that the simple adaptive partitioning includes the fixed partitioning. The other option is *deconvolution-based adaptive partitioning* shown in Fig. 3(c). Deconvolution of the (exposure) step at the edge by the PSF is calculated first. Then, the edge area is partitioned into one or more regions according to variation of the deconvolution result (dose distribution). As shown in Fig. 3(c), a larger exposure step leads to a larger spatial variation of deconvolution result, which makes the corresponding edge area partitioned more finely.

3. Corner and junction areas

The spatial variation of the ideal exposure is large around a corner area. Hence, a corner area requires a finer partitioning (compared to other types of areas) in order to have sufficient spatial control of dose within the corner area. Corner partitioning adopts the same partitioning options as for edge partitioning, i.e., the fixed partitioning, the simple adaptive partitioning and the deconvolution-based adaptive partitioning. In the fixed partitioning, a corner area is partitioned, in two directions, by extending the fixed partitions of the two adjacent edge areas, as illustrated in Fig. 4(a). In the simple adaptive partitioning, a corner area is normally partitioned (in two directions) in the same way as in the fixed partitioning. However, when the ideal exposure varies only in one direction around a corner area, it is partitioned in one direction by extending partitions of one of the two adjacent edge areas as illustrated in Fig. 4(b). In the deconvolution-based adaptive partitioning, a corner area is partitioned by extending the two edge partitions obtained by the deconvolution-based edge partitioning as shown in Fig. 4(c).

A junction area is partitioned in the similar way with the corner partitioning illustrated in Fig. 4.

III. DOSE CALCULATION

After the rectangle and region partitioning, the initial dose factor for every region is derived. Then, an iterative procedure is carried out. In each iteration, the dose factor of each region is calculated (adjusted) based on the exposure estimate at the corresponding critical point and the corresponding ideal exposure. The iteration continues until a certain termination condition (e.g., the average exposure error is less than a certain value) is met.

The main objective of proximity effect correction in binary lithography is to minimize the CD error and therefore critical points are set up along the (external) boundaries of each feature. Another requirement (to the minimization of the CD error) for grayscale lithography is that the exposure

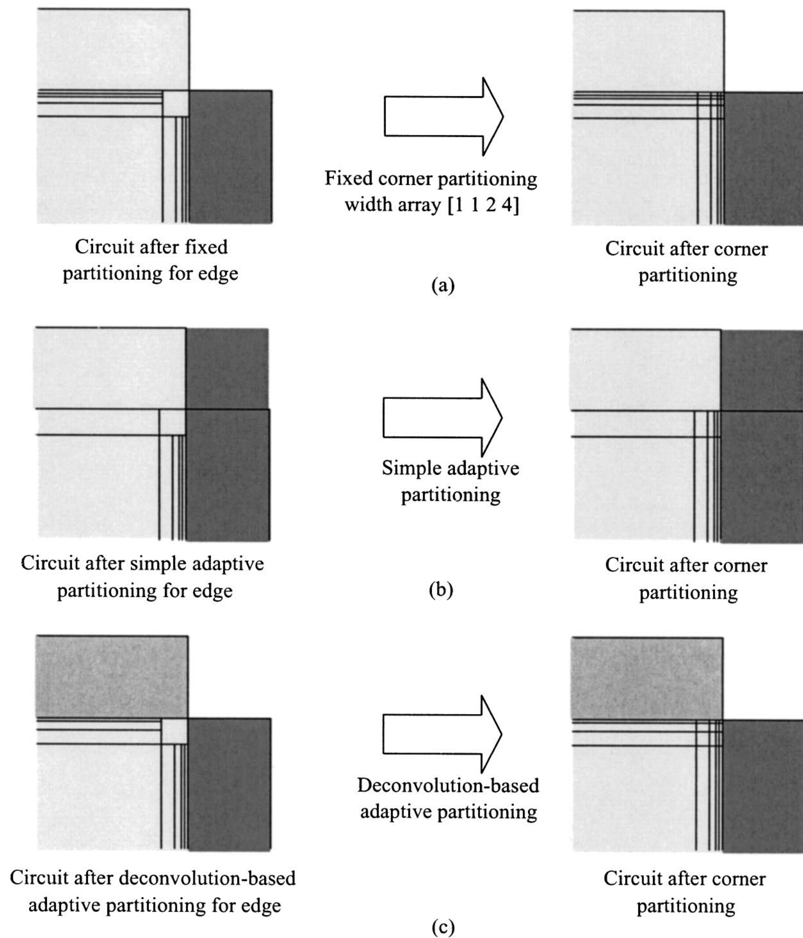


FIG. 4. Partitioning of corner areas: (a) fixed partitioning, (b) simple adaptive partitioning, (c) deconvolution based adaptive partitioning. The gray level of each rectangle represents its ideal exposure. Some regions are not shown.

distribution within each feature should be as uniform and close to its ideal exposure as possible. Accordingly, critical points are distributed throughout each feature, one at the center of each region, as shown in Fig. 5. When an edge of the feature is external (i.e., not adjacent to another feature), each region along the edge has an additional critical point right outside the middle point of its external edge (as in binary lithography).

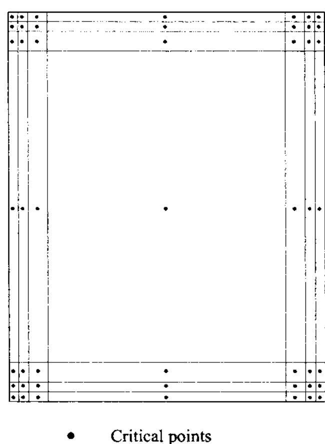
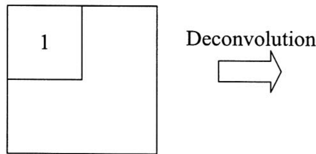
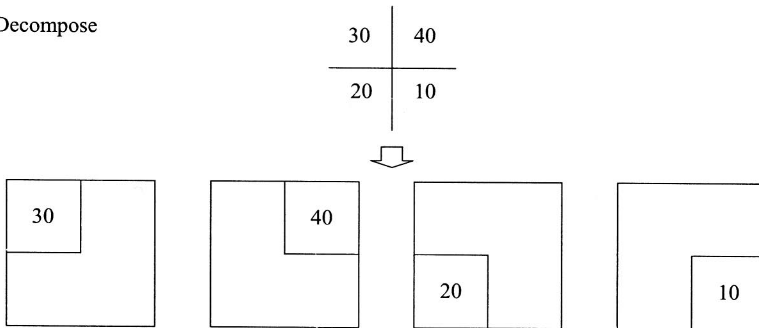


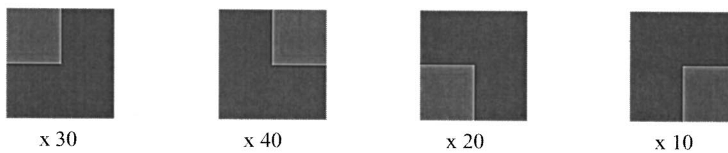
FIG. 5. Critical points in a rectangle that has been partitioned into multiple regions.

A. Derivation of initial dose

In an iterative dose correction scheme, a ''good'' initial solution (initial dose for each region) can lead to faster convergence of correction and a better correction result. In the previous work,[10] a deconvolution-based initial dose set-up method was employed. In binary lithography, a unit edge (corner, junction) is deconvolved by a PSF *a priori* and then the deconvolution result can be used for all edges (corners, junctions) as long as the same PSF is assumed. This is possible because every edge (corner, junction) has the same height (exposure step), i.e., *binary* structure. However, for a grayscale structure, exposure step varies with edge (corner, junction). In order to avoid carrying out deconvolution for every edge, corner, and junction, the *linear combination* method which exploits the linearity of deconvolution has been designed. For an edge of arbitrary exposure step, the deconvolution result of a unit edge is scaled by the exposure step to derive the initial dose distribution along the edge. For an *intersection* area where two or more features (rectangles) meet, the deconvolution result of a unit corner is scaled by the exposure level of each rectangle and then the scaled deconvolution results are combined (added) to obtain the deconvolution result of the intersection area, as illustrated in Fig. 6. Note that deconvolution of a unit edge and a unit corner needs to be done only once during correction of a







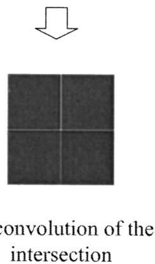


FIG. 6. Illustration of the linear combination method to derive deconvolution of an arbitrary intersection. The number in each rectangle is its ideal exposure and each grayscale image represents the dose distribution.

grayscale structure. The linear combination method saves computation time for the initial dose set-up greatly.

B. Calculation of dose factor

Dose factor is the dose normalized by the base dose. The first step in determining dose factor of a region is to estimate exposure at the critical point of the region. The efficient two-level approach to exposure estimation, which was developed earlier,[8] is employed, where exposure at a point is decomposed into two components, *local* exposure and *global* exposure. Local exposure is composed of exposure contributions from features within a small window and is computed exactly by a lookup table method. Global exposure consists of exposure contributions from features outside the window and is estimated by an approximation method. This approach saves computation time for exposure estimation by several orders of magnitude without degrading accuracy substantially.

Let us denote the total exposure at the critical point of a region by E_{tl} which is the sum of global exposure E_{gb} and local exposures E_{lc} i.e., $E_{tl}=E_{gb}+E_{lc}$.

Define E_{rem} as follows:

$$E_{rem}=E_{id}-(E_{tl}-E_{reg}), \tag{1}$$

where E_{id} denotes the ideal exposure of the region and E_{reg} refers to exposure contribution from the region itself.

That is, E_{rem} is the *remaining* exposure when the dose factor of the region is set to zero. The exposure contribution from the region itself is determined to compensate for E_{rem}. Let E_{reg1} be the exposure contribution from the region when the dose factor of the region is 1. Then, a temporary dose factor, D_{tmp}, of the region is derived as E_{rem}/E_{reg1}. D_{tmp} is computed *nonrecursively* within each iteration, i.e., assuming that dose factors of all other regions are maintained the same as in the previous iteration. In general, this would lead to overcorrection. Hence, an adjustment factor F_{adj} is used to compensate overcorrection and thereby avoid oscillation of dose factors. The dose factor of the region, D_{reg}, is adjusted as follows:

$$D_{reg}=\begin{cases} F_{adj}\cdot(D_{tmp}-D_{old})+D_{old} & \text{if } D_{reg}\geqslant 0 \\ 0 & \text{if } D_{reg}<0 \end{cases} . \tag{2}$$

In Eq. (2), D_{old} is the dose factor computed in the previous iteration, and $0<F_{adj}<1$. The new dose factors com-

TABLE I. Exposure errors and correction time. The substrate is 1000 nm PMMA on Si and the beam energy is 50 keV: (a) checkerboard pattern, (b) Etec pattern. No-cor, DC Cor, Simp. Adpt, and Deconv represent No-correction, DC correction, simple adaptive partitioning and deconvolution-based partitioning, respectively. A Sun workstation (Sparc V9 450 MHz) was used for correction.

		Edge error (μC/cm^2)			Corner error (μC/cm^2)			Center error (μC/cm^2)			Correction time
	Procedure	Avg	Avg Per	Max	Avg	Avg Per	Max	Avg	Avg Per	Max	(mm:ss)
(a)	No-cor.	12.615	192.50%	31.564	12.873	242.99%	33.481	12.367	44.07%	22.781	0:00
	DC Cor.	2.224	19.00%	18.564	3.429	40.18%	24.125	0.238	0.93%	1.031	0:00
	Simp. Adpt	0.542	4.43%	17.559	1.047	15.83%	18.997	0.037	0.16%	0.259	31:43
	Deconv	0.585	4.86%	17.562	1.108	16.51%	18.982	0.043	0.18%	0.262	28:53
(b)	No-cor.	17.498	114.54%	30.421	19.732	122.44%	35.106	13.648	44.61%	21.897	0:00
	DC Cor.	5.210	21.29%	19.149	7.829	40.48%	28.588	0.305	1.02%	1.148	0:00
	Simp. Adpt	0.355	1.67%	20.497	0.451	9.57%	21.425	0.095	0.33%	1.702	17:23
	Deconv	0.432	2.44%	20.446	0.584	10.00%	20.592	0.098	0.34%	1.712	9:18

puted in an iteration do not take effect until the beginning of next iteration. In general, F_{adj} depends on the degree of proximity effect. The higher the proximity effect is, the smaller F_{adj} will be.

IV. SIMULATION RESULTS AND DISCUSSION

The main emphasis of this first grayscale proximity correction scheme was on achieving the exposure distribution as close to the grayscale structure (ideal exposure distribution) as possible. Hence, *exposure error*, which is defined to be the difference between the actual and ideal exposures, is considered in analyzing performance of the dose control scheme described in this paper. Since critical points are not uniformly distributed (see Fig. 5), *average* exposure error, ϵ, is computed with a weight of region area for each region, i.e., exposure error at the critical point of a larger region is weighted more. Suppose that there are N regions in a circuit pattern. Let ϵ_i and A_i be the exposure error at the critical point of region i and the area of region i, respectively. The average exposure error ϵ is computed as follows:

$$\epsilon = \frac{\sum_{i=1}^{N} \epsilon_i \cdot A_i}{\sum_{i=1}^{N} A_i}. \tag{3}$$

Also, *percentage exposure error* of region i is defined as $\epsilon_i / E_{\text{ref}_i} \times 100$ where E_{ref_i} is the *reference exposure* level for region i, which varies with type of region. Then, *average* percentage exposure error is computed as a weighted sum similarly with the average exposure error [refer to Eq. (3)]. The reference exposure for a center region is the ideal exposure of the center region. The reference exposure for an edge region is the exposure step of the corresponding edge. And the reference exposure for a corner (junction) region is the average of the exposure steps of all edges associated with the corner (junction).

The dose control scheme described in this paper is compared to *No-correction* and *DC correction*. In No-correction, features (rectangles) are not partitioned and the dose factor corresponding to the ideal exposure of a rectangle is assigned to the rectangle. In DC correction, correction is carried out, but without partitioning of rectangles, i.e., a dose factor for each rectangle (feature) is determined. For the proposed dose control scheme, the two region partitioning options, simple adaptive partitioning and deconvolution-based adaptive partitioning, are used.

Two circuit patterns, a checkerboard pattern of size 800×800 nm^2 and Etec pattern of size 2500×2500 nm^2, both with the minimum feature size (MFS) of 100 nm, are corrected for the substrate system of 1000 nm PMMA on silicon and the e-beam energy of 50 keV. Ideal exposure of each feature is randomly set to avoid pattern-dependent results as much as possible. The range of ideal exposure is 17–44 μC/cm^2 and 16–43 μC/cm^2 for the checkerboard and Etec patterns, respectively. The minimum exposure step is 1 μC/cm^2 for both. Since there is no critical point outside features (in this implementation), the exposure error is computed only for the internal regions of features. Some of the simulation results, exposure error and correction time, are provided in Table I. Also, the exposure distributions after correction are shown in Fig. 7.

In Table I, it can be seen that DC correction achieves exposure errors much smaller than those by No-correction with both taking a negligible amount of computation time (less than 1 s). However, edge and corner errors are still large with the average percentage exposure errors ranging from 19% to 40.48%. The proposed dose control scheme greatly reduces edge error to below 5% and corner error to below 17%, by effectively controlling spatial dose distribution within each feature. The center exposure error is very small (less than 1%) for the proposed scheme.

Note that the maximum edge and corner exposure errors are large even for the proposed scheme. Over edges and corners, the ideal exposure varies abruptly. Such exposure steps are not realizable by any practical dose control method. That is, the exposure slope (contrast) cannot be infinite, which leads to a highly localized large difference between the ideal and actual exposures at the edge and corner. However, one should notice that edge and corner areas are much smaller than a center area. Therefore, the average (percentage) exposure error would be close to the center error which

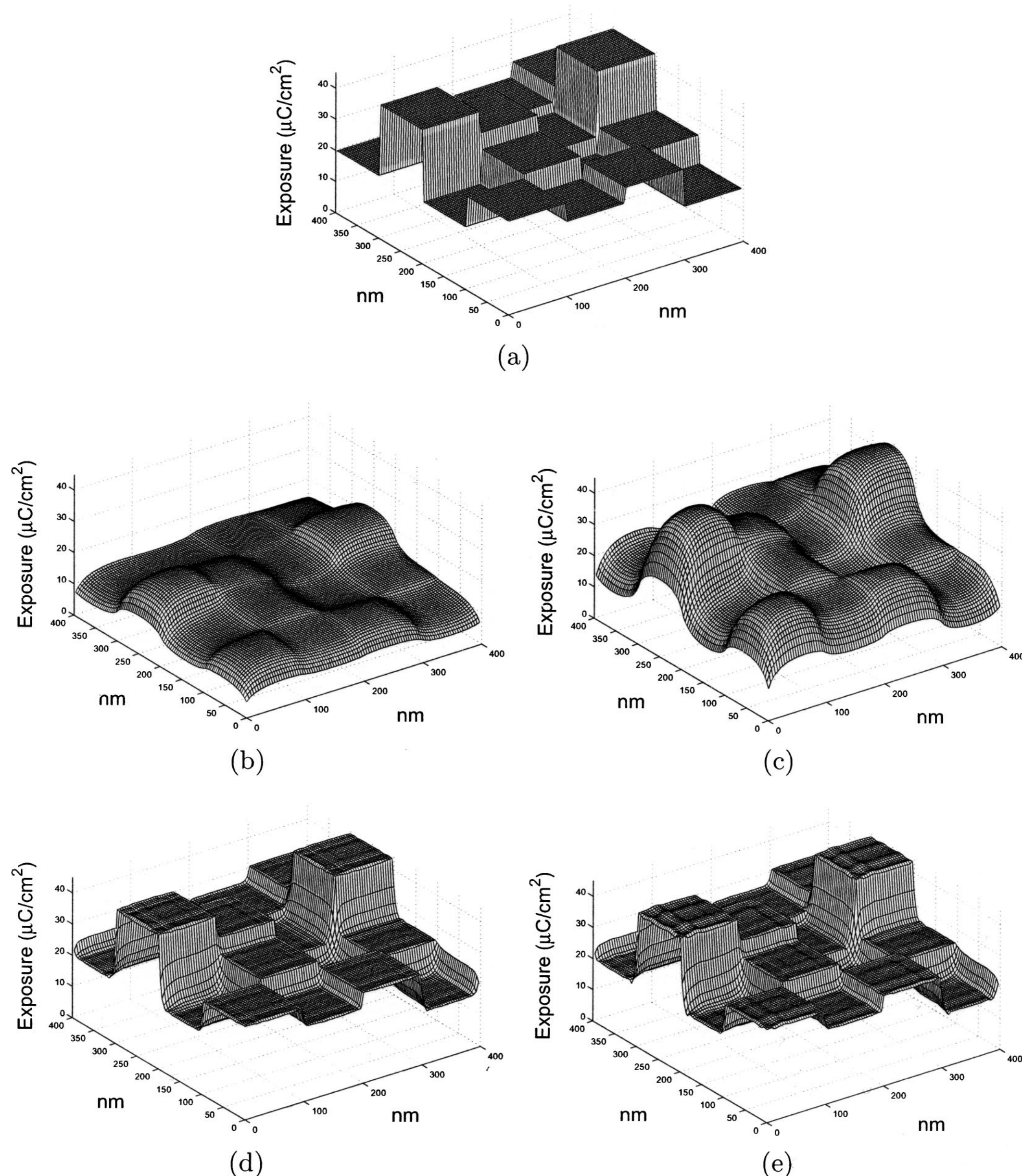


FIG. 7. Exposure distribution for the checkerboard pattern corrected for the substrate of 1000 nm PMMA on Si (50 keV): (a) Ideal exposure, (b) No-correction, (c) DC correction, (d) Simple adaptive partitioning, (e) Deconvolution-based adaptive partitioning.

is much smaller than the edge and corner errors.

Deconvolution-based adaptive partitioning has a better adaptability to individual edges, corners and junctions than simple adaptive partitioning. Hence, it tends to generate a smaller number of regions, which leads to a substantially shorter correction time at the expense of slightly larger exposure errors.

The feature size and density spatially vary more in the Etec pattern than in the checkerboard pattern. This makes a simple correction procedure like dc correction less effective for the Etec pattern. Note that the relative improvement by the proposed dose control scheme over dc correction is larger for the Etec pattern.

It is seen in Fig. 7 that the overall exposure distributions achieved by the proposed dose control scheme are much closer to the ideal exposure distribution than those by No-correction and DC correction.

For the results in Table I, each feature is partitioned into at least 169 regions (usually more than 200 regions in Etec pattern) where the minimum region size is one pixel (2 nm). This large number of regions per feature lead to a relatively long correction time for such small patterns. However, the number of regions per feature would be significantly less in practice.

Improvements: Significant improvement over this first version of dose control scheme is possible. For example, this version mainly attempts to minimize the exposure error, but an explicit constraint for minimizing CD error may be incorporated into the correction procedure. Also, the computation

requirement may be reduced substantially by optimizing the exposure estimation and dose calculation procedures.

V. SUMMARY

In this paper, the first grayscale proximity effect correction scheme which takes a dose control approach is described and some of its performance analysis results obtained via computer simulation are presented. In grayscale lithography, one is to deal with multiple exposure levels to be realized within a single circuit pattern. Hence, not only CD error between features, but also exposure error within each feature must be minimized. This requires much finer spatial control of dose, particularly along the boundaries of features, than that required in proximity effect correction for binary lithography.

For sufficient spatial control of dose, a (large) feature is broken up into smaller ones if necessary. Each feature is partitioned into a set of regions where a dose factor is computed for each region. Therefore, how each feature is partitioned directly affects correction time and accuracy. Finer partitioning is required in the edge, corner, and junction areas where the ideal exposure varies abruptly. Both fixed and adaptive partitioning methods have been considered.

Simulation results indicate that the proposed dose control scheme can achieve the exposure distribution closely following the multidimensional structure of a grayscale circuit pattern and therefore has a good potential to be applicable to the single step e-beam lithographic process for fabricating multilevel structures. One of the current efforts is to shorten correction time by minimizing the number of iterations required, reducing the number of regions per feature, and optimizing the correction algorithm.

[1]O. Painter, R. Lee, A. Scherer, A. Yariv, J. O'Brien, P. Dapkus, and I. Kim, Science **284**, 1819 (1999).
[2]S.-Y. Lin and J. Fleming, IEEE J. Lightwave Technol. **17**, 1944 (1999).
[3]D. P. Kern, in *Proceedings of the 9th International Conference on Electron and Ion Beam Science and Technology*, edited by R. Bakish (Electrochemical Society, Princeton, 1980), pp. 326–339.
[4]J. Greeneich, J. Vac. Sci. Technol. A **19**, 1269 (1981).
[5]G. Owen, J. Vac. Sci. Technol. B **8**, 1889 (1990).
[6]U. Hofmann, C. Kalus, A. Rosenbusch, R. Jonckheere, and A. Hourd, in *Proceedings of the SPIE Conference on E-Beam, X-Ray, EUV, and Ion-Beam Submicrometer Lithographies for Manufacturing VI*, March 1996, Vol. 2723, pp. 150–158 (unpublished).
[7]M. Peckerar, R. Bass, and K. Rhee, J. Vac. Sci. Technol. B **18**, 3143 (2000).
[8]S.-Y. Lee and B. D. Cook, IEEE Trans. Semiconductor Manufacturing **11**, 108 (1998).
[9]B. D. Cook and S.-Y. Lee, IEEE Trans. Semiconductor Manufacturing **11**, 117 (1998).
[10]S.-Y. Lee and J. Laddha, Microelectron. Eng. **57–58**, 303 (2001).
[11]S.-Y. Lee and D. He, "Simultaneous dose modification for balance proximity effect correction in E-beam lithography," Microelectron Eng. (to be published).