Analysis and Optimization of Lithographic Performance on Massively-Parallel Electron-beam Systems

by

Md Nabid Hasan

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 6, 2023

Keywords: Electron-beam lithography, Massively-parallel electron-beam system, Critical dimension error, Exposing time, Cost function, Optimization

Copyright 2023 by Md Nabid Hasan

Approved by

Soo-Young Lee, Professor Emeritus of Electrical and Computer Engineering
Guofu Niu, Professor of Electrical and Computer Engineering
Masoud Mahjouri-Samani, Assistant Professor of Electrical and Computer Engineering
Mehdi Sadi, Assistant Professor of Electrical and Computer Engineering

Abstract

While electron-beam (e-beam) lithography is widely used in transferring fine-feature patterns onto a substrate, its major drawback is the low throughput, especially for large-scale patterns. The massively-parallel electron-beam systems (MPES) were developed to increase the writing throughput and demonstrated to be able to write large-scale patterns significantly faster compared to conventional single-beam systems. However, the MPES comes with several constraints, such as abnormal beams (e.g., permanently on or off, spatial and temporal beam current fluctuation, and beam-positioning error), relatively large beam size, the realization of non-uniform dose distribution with a sub-beam-size resolution, and optimizing one performance metric affecting other metrics.

To address these issues and maximize the efficiency of MPES, this study has several objectives. Firstly, the effects of abnormal beams on performance metrics are analyzed through simulation, comparing different writing methods, *Single-row writing I, Single-row writing II, Multi-row writing*, to suggest ways of reducing their negative effects. Secondly, the *multi-row writing* and *multi-pass writing* methods are compared in reducing the adverse effects of abnormal beams. Thirdly, a shape and dose control procedure with a sub-beam-size resolution for the proximity effect correction (PEC) is developed. Fourthly, two different methods for increasing beam utilization and reducing exposing time are investigated, i.e., lowering the dose difference among regions of a feature during the PEC and a new writing method for non-uniform distribution by adjusting the deflection angle of beams. Finally, an adaptive optimization method is designed that can handle any combination of performance metrics in a cost function for both a single feature and a large-scale pattern.

Overall, this study aims at maximizing the efficiency of MPES by addressing various constraints, improving performance metrics individually, and optimizing multiple performance metrics simultaneously.

Acknowledgments

I am deeply grateful to Dr. Soo-Young Lee, my advisor, for leading me into the exciting research area of Massively-parallel electron-beam systems and providing invaluable support throughout my doctoral program. Without his detailed guidance and helpful suggestions, this dissertation would not have been possible. Dr. Lee's patience and encouragement carried me through difficult times, and his vast knowledge and skills significantly contributed to my research and this dissertation.

I would also like to express my sincere appreciation to Dr. Guofu Niu, Dr. Masoud Mahjouri-Samani, and Dr. Mehdi Sadi, who served on my committee, and to Dr. Hans Werner Van Wyk, who acted as my university reader. Their valuable comments and assistance greatly benefited my research.

I am also thankful to former and current members of our research group for their contributions and support. This research was made possible by the generous funding of Samsung Electronics Co., Ltd., and I am grateful for their support.

Last but not least, I want to extend my heartfelt gratitude to my family, including my parents, Md Reazul Hasan and Shahida Parvin, and my wife, Homayra Asima. Their unwavering love, patience, and understanding have been my pillars of strength and motivation throughout my academic journey.

Table of Contents

Abstract ii

Acknowledgments iii

List of Figures viii

List of Tables xvii

1 Introduction 1

1.1 Problem definition 1

1.2 Background review 2

1.3 Motivation, objectives and approaches 4

1.4 Organization of Dissertation 6

2 Models and Performance Metrics in Massively-parallel E-beam Systems 7

2.1 Models 7

2.1.1 Massively-parallel electron-beam system 7

2.1.2 Substrate system 8

2.1.3 Transfer function 9

2.1.4 E-beam lithography process 11

2.2 Performance metrics 12

2.2.1 Exposure variation 12

2.2.2 Critical dimension error 12

2.2.3 Line edge roughness 13

2.2.4 Maximum indent and maximum outdent 13
2.2.5 Exposing time 14

3 Effects of Abnormal Beams on Performance Metrics 16
3.1 Writing methods 16
3.2 Simulation 18
3.3 Results and discussion 22
3.3.1 Always-off faulty beams 22
3.3.2 Multiple sets of beams 24
3.3.3 Always-on faulty beams 24
3.3.4 Spatial fluctuation of beam current 25
3.3.5 Temporal fluctuation of beam current 26
3.3.6 Beam-positioning error 27
3.3.7 Dependency on σ_t 28
3.4 Summary 28

4 Effectiveness of Multi-pass and Multi-row Writing Methods 29
4.1 Optimization of writing methods 29
4.1.1 Multi-pass (MP) writing 30
4.1.2 Multi-row writing (MRW) 31
4.2 Simulation 32
4.3 Results and discussion 35
4.3.1 Multi-pass (MP) writing 36
4.3.2 Multi-row writing (MRW) 38
4.3.3 Comparison of the MP writing and MRW 40
4.4 Summary 43

5 Shape and Dose Control for Proximity Effect Correction 44
5.1 Realization of spatial dose distribution 44
5.1.1 Uniform dose distribution 47
5.1.2 V-type dose distribution 50
5.1.3 A-type dose distribution 52
5.1.4 M-type dose distribution 54
5.2 Minimization of CD error and LER 56
5.2.1 Cost function 56
5.2.2 Shape+dose correction 56
5.3 Results and discussion 61
5.4 Summary 65
6 Reduction of Exposing Time 66
6.1 Methods for reducing exposing time 66
6.1.1 Writing method without empty cycle 66
6.1.2 Reduction of dose difference among the regions 68
6.2 Simulation 72
6.3 Results and discussion 75
6.3.1 Reduction of exposing time by removing empty cycles 75
6.3.2 Reduction of exposing time by decreasing dose difference among the regions 86
6.4 Summary 90
7 A Generalized Performance Optimization 92
7.1 Transfer function 92
7.2 Optimization method for a single feature 93
7.2.1 Cost function 94

7.2.2 Old optimization method . . . 95

7.2.3 New optimization method . . . 96

7.3 Optimization method for large-scale pattern . . . 98

7.4 Simulation . . . 103

7.5 Results and discussion . . . 104

7.5.1 Single feature . . . 105

7.5.2 Large-scale pattern . . . 109

7.6 Summary . . . 116

8 Concluding Remarks . . . 118

References . . . 121

List of Figures

2.1 A substrate moves in the x-direction exposed by parallel beams: the beam size is $B \times B$, and the beam interval is I_{bx}. 8

2.2 A substrate system when the feature is exposed with a dose D. 9

2.3 (a) The ideal transfer function for the beam aperture of size 10 nm $\times$ 10nm, (b) the Gaussian function with $\sigma_t = 1.5$nm, (c) the transfer function obtained from the Gaussian function in (b), (d) the PSF generated from the Monte-Carlo simulation (CASINO) for the substrate system of 100nm PMMA on Si and the beam energy of 30keV: the estimated σ_t is 1.22nm, and (e) the transfer function obtained from the PSF in (d). 10

2.4 (a) The deterministic transfer function for the beam aperture of size 10nm $\times$ 10nm, and (b) the same transfer function with 7% noise added. 11

2.5 Peak-peak exposure variation. The feature size is 40nm $\times$ 120nm, and σ_t = 2nm. 12

2.6 The 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. 13

2.7 (a) maximum indent and (b) maximum outdent. The feature size is 40nm $\times$ 120nm and σ_t = 2nm. 14

3.1 A row of pixels (larger squares) being exposed by 4 beams (smaller squares) in the single-row writing method I where $I_{bx} = 3B$ and $n_s = 4$, and the substrate moves to the right: this is a snapshot at the beginning of 1st step of a cycle. The pixels with the same shade of gray are exposed by the same beam with the corresponding shade. 17

3.2 A row of pixels (larger squares) being exposed by 4 beams (smaller squares) in the single-row writing method II where $n_s = 4$, $I_{bx} = 2B$, $n_g = 2$ and $I_{gx} = B$, and the substrate moves to the right. A group of two beams exposes each pixel and this is a snapshot right before the 2nd beam in each group starts to expose the pixel partially exposed by the first beam. The pixels with the same shade of gray are exposed by the group of beams with the corresponding shade. 17

3.3 4 rows of pixels (larger squares) being exposed by 4 beams (smaller squares) in the multi-row writing method where $I_{bx} = 3B$ and $n_s = 4$, and the substrate moves to the right. This is a snapshot at the beginning of the 3rd step of a cycle. The pixels with the same shade of gray are exposed by the beam with the same shade and each pixel is exposed by a beam, once in a cycle. 18

3.4 An example of spatial fluctuation of beam current following a Gaussian distribution with the average of 1 and standard deviation of 5%. 20

3.5 An example of temporal fluctuation of beam current following a Gaussian distribution with the average of 1 and standard deviation of 5%. 21

3.6 Beam-positioning error. The dotted line represents the target exposure of a beam while the solid line represents the actual exposure of the beam deviating from the target position by $(\Delta x, \Delta y)$. 21

3.7 (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm. 23

3.8 Contour of a rectangular feature (40nm × 120nm) when a faulty beam affects edge pixels and $\sigma_t = 2$nm: (a) single-row writing method I, (b) single-row writing method II, and (c) multi-row writing method 23

3.9 (a) Exposure variation, (b) LER, and (c) maximum indent with the number of cycles (n_c) varied. $\sigma_t = 2$nm, $n_f = 1$. 24

3.10 (a) Exposure variation, (b) LER, and (c) maximum outdent with the number of always-on faulty beams (n_f) varied. $\sigma_t = 2$nm. 25

3.11 (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 26

3.12 Contour of a rectangular feature (40nm × 120nm) when the spatial fluctuation of beam current is 10% with no faulty beam and $\sigma_t = 2$nm: (a) single-row writing method I, (b) single-row writing method II, and (c) multi-row writing method. 26

3.13 (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 27

3.14 (a) Exposure variation, (b)LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$. 27

3.15 (a) Exposure variation, (b) LER, and (c) maximum indent with the blurring factor σ_t varied. $n_f = 1$. 28

4.1 An illustration of the pixels exposed by a beam in the multi-pass (MP) writing where (a) $\Delta x = B$, and (b) $\Delta x = 2B$. The gray pixels represent the pixels exposed once by the beam with a small gray square on its top. The black pixels represent the pixels exposed by the same gray beam more than once. In both cases, $n_s = 4$ and $I_{bx} = 3B$. 31

4.2 Different realizable patterns for the MRW with $n_s = 4$ and $I_{bx} = 3B$, where (a) $\Delta H = 1$ and $w_p = 4$, (b) $\Delta H = 3$ and $w_p = 4$, (c) $\Delta H = 2$ and $w_p = 3$, and (d) $\Delta H = 2$ and $w_p = 7$. In all cases, $\Delta V = 1$. 32

4.3 Beam abnormalities: (a) an example of the spatial fluctuation of beam currents, where the doses given by the beams are slowly decreasing from the center of the aperture array, and (b) a vector field of the beam-positioning error guided by the concentric ellipses, $ax^2 + by^2 = 1$ where a and b are varied. 35

4.4 (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm. 36

4.5 (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 37

4.6 (a) Exposure variation, (b) LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$. 37

4.7 (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 38

4.8 (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm. 39

4.9 (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 39

4.10 (a) Exposure variation, (b) LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$. 40

4.11 (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 40

4.12 (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm. 41

4.13 (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 42

4.14 (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$. 42

4.15 (a) Exposure variation, (b) LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$. 43

5.1 A feature is partitioned into 5 regions such that the dose can be spatially controlled region-wise. The spatial dose distribution $\{d_i\}$ is symmetric with respect to the center region, i.e., $d_1 = d_5$ and $d_2 = d_4$. 45

5.2 The total dose is given in one pass when (a) $\Delta W = 0$, and (b) $\Delta W = B$. More than one pass is used to realize a dose distribution when (c) $0 < \Delta W < B$. . . . 46

5.3 The realization of uniform dose distribution in multiple passes for (a) $\Delta W =$ 0nm and $n_p = 1$, (b) $\Delta W = 2$ and 4nm, and $n_p = 4$, (c) $\Delta W =$ 6nm and $n_p = 4$, and (d) $\Delta W =$ 8nm and $n_p = 4$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes. . . 48

5.4 The ideal (top) and realized (bottom) uniform dose distributions for each of $\Delta W = 2, 4, 6$, and 8nm. 49

5.5 The realization of V-type dose distribution in multiple passes with a spatial-dose-distribution ratio of 5:3:2:3:5 for (a) $\Delta W =$ 0nm and $n_p = 1$, and (b) $\Delta W = 2, 4, 6$ and 8nm, and $n_p = 2$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes. 50

5.6 The ideal (top) and realized (bottom) V-type dose distributions for each of $\Delta W = 2, 4, 6$, and 8nm. 51

5.7 The realization of A-type dose distribution in multiple passes with a spatial-dose-distribution ratio of 1:2:5:2:1 for (a)$\Delta W =$ 0nm and $n_p = 1$, and (b)$\Delta W =$ $2, 4, 6$ and 8nm, and $n_p = 2$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes. 52

5.8 The ideal (top) and realized (bottom) A-type dose distributions for each of $\Delta W = 2, 4, 6$, and 8nm. 53

5.9 The realization of M-type dose distribution in multiple passes with a spatial-dose-distribution ratio of 1:5:2:5:1 for (a) $\Delta W =$ 0nm and $n_p = 1$, and (b) $\Delta W = 2, 4, 6$ and 8nm, and $n_p = 3$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes. 54

5.10 The ideal (top) and realized (bottom) M-type dose distributions for each of $\Delta W = 2, 4, 6$, and 8nm. 55

5.11 The flowchart of the shape + dose correction method where k is the iteration index, and x and y are fractional numbers. 58

5.12 The dose adjustment in the case of V-type dose distribution guided by the cross-section of the resist profile. 59

5.13 The dose adjustment in the case of A-type dose distribution guided by the cross-section of the resist profile. 60

5.14 The dose adjustment in the case of M-type dose distribution guided by the cross-section of the resist profile. 61

5.15 (a) The average CD error over the top, middle and bottom layers of resist, (b) total dose required, and (c) LER at the middle layer are plotted with respect to the line-width reduction ΔW for $\sigma_t = 1$nm where feature size: 50nm $\times$ 300nm, beam size: 10nm $\times$ 10nm, beam interval: 40nm, resist thickness: 100nm. For each ΔW, the optimal spatial-dose-distribution ratio is determined using the iterative procedure described in Section 5.2. 63

5.16 (a) The average CD error over the top, middle and bottom layers of resist, (b) total dose required, and (c) LER at the middle layer are plotted with respect to the line-width reduction ΔW for $\sigma_t = 4$nm where feature size: 50nm $\times$ 300nm, beam size: 10nm $\times$ 10nm, beam interval: 40nm, resist thickness: 100nm. For each ΔW, the optimal spatial-dose-distribution ratio is determined using the iterative procedure described in Section 5.2. 63

6.1 Comparison of the writing methods: (a) the conventional (single-row) writing method, and (b) the proposed writing method. 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 a line/space pattern and s_i is the space between two adjacent lines l_i and l_{i+1}. 67

6.2 Reduction of dose difference among the five regions to improve the beam utilization and reduce the exposing time. 70

6.3 The flowchart of reducing the dose difference among the regions of a feature to improve the beam utilization and shorten the exposing time where k is the iteration index, $0 < x < 1$, and $0 < y < 1$. 72

6.4 (a) The cross-section view and (b) the top-down view of the remaining resist profile. The feature size is 50nm$\times$300nm and the resist thickness is 100nm. . . 73

6.5 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm. 76

6.6 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c) $l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm. 77

6.7 The realization of non-uniform target dose distribution in multiple passes. The feature is divided into five regions and each region consists of a pixel with size 10nm×10nm. The same figure can be referred to when each region consists of more than a pixel. 78

6.8 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 50$nm and the dose distribution ratio is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1. 79

6.9 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l+s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1, 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1, 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1, 1:4:4:4:4:1 in three different widths of features. 79

6.10 Tracing pixels exposed by the beams in multiple cycles in the steady state for the conventional writing method when $I_{bx} > l + s$ and $l_i = l_j$. In this figure, $I_{bx} = 70$nm, $l = 30$nm, and $s = 30$nm. The gray and white squares represent the exposed and unexposed pixels, respectively. 80

6.11 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $s = 40$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm. 80

6.12 Tracing pixels exposed by the beams in the system in the second cycle of the steady state for the conventional writing method when $I_{bx} > l + s$ and $l_i \neq l_j$. The gray and white squares represent the exposed and unexposed pixels, respectively. 81

6.13 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c)$l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm. 81

6.14 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 40$nm and the dose distribution ratio is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1. 82

6.15 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is non-uniform, $l_1 = 40$nm, $l_2 = 50$nm, and $l_3 = 60$nm and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1 and 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1 and 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1 and 1:4:4:4:4:1, respectively, for three different widths of features. 82

6.16 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 83

6.17 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c) $l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 83

6.18 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 50$nm and the dose distribution ratio in each feature is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 84

6.19 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l+s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1, 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1, 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1, 1:4:4:4:4:1 in three different widths of features. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 84

6.20 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 85

6.21 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c) $l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 85

6.22 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 40$nm and the dose distribution ratio in each feature is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 85

6.23 Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, $l_1 = 40$nm, $l_2 = 50$nm, and $l_3 = 60$nm, and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1 and 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1 and 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1 and 1:4:4:4:4:1, respectively, in three different widths of features. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation. 86

6.24 The A-type dose distribution obtained (a) without the PEC optimization, (b) with the PEC optimization which does not apply the method in Section 6.1.2, and (c) with the PEC optimization which applies the method in Section 6.1.2 for the case when $l = 50$nm and $\sigma_t = 1$nm. 89

6.25 The M-type dose distribution obtained (a) without the PEC optimization, (b) with the PEC optimization which does not apply the method in Section 6.1.2, and (c) with the PEC optimization which applies the method in Section 6.1.2 for the case when $l = 50$nm and $\sigma_t = 1$nm. 89

6.26 The V-type dose distribution obtained (a) without the PEC optimization, (b) with the PEC optimization which does not apply the method in Section 6.1.2, and (c) with the PEC optimization which applies the method in Section 6.1.2 for the case when $l = 50$nm and $\sigma_t = 1$nm. 90

7.1 The deterministic transfer function for the beam aperture of size 10nm$\times$10nm where (a) $\sigma_1 = 2$nm and $\sigma_2 = 0$, and (b) $\sigma_1 = 4$nm and $\sigma_2 = 9.5\mu$m. 93

7.2 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. . . . 98

7.3 Deconvolution surface. 100

7.4 The flowchart 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. 102

7.5 The 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. 108

7.6 The 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 7.3). In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively. 114

7.7 The 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 7.3). In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively. 115

7.8 The 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 7.3). In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively. 115

List of Tables

3.1 A summary of simulation set-up 22

5.1 The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing only the average CD error for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 4$nm 64

5.2 The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing the cost function for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 4$nm 64

5.3 The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing only the average CD error for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 1$nm 64

5.4 The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing the cost function for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 1$nm 65

6.1 The CD error, total dose, beam utilization, exposing time with the conventional writing method, $t_{exp,old}$, and exposing time with the proposed writing method, $t_{exp,new}$, each obtained for three cases, i.e., without the PEC optimization (no use of the cost function C), the PEC optimization without applying the method described in Section 6.1.2 (the rows with C) and the PEC optimization applying the method described in Section 6.1.2 (the rows with C^*). It is considered that $C = CD_error_{norm} + 0.2 total_dose_{norm} + 0.5\frac{1}{U_{norm}}$, $I_{bx} \neq l + s$, $s = 40$nm, $I_{bx} = 100$nm for $l = 50$nm and $I_{bx} = 200$nm for $l = 150$nm. 88

7.1 The 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 12μm width which consists of lines with the linewidths of 50nm and 150nm. The space between two lines is 40nm, the beam interval is 200nm, and the number of beams in a row is 21. 107

7.2 The optimization results with four different cost functions. The feature width is 50nm, $\sigma_1 = 4$nm, $\sigma_2 = 9.5\mu$m, $I_{bx} = 100$nm, and the dose distribution is A-type. The exposing time is calculated for a uniform pattern of 12μm width which consists of lines with a linewidth of 50nm. The space between two lines is 40nm, the beam interval is 200nm, and the number of beams in a row is 21. . . 109

7.3 The optimization results for a large-scale pattern with the cost function $C = CD_error_{norm} + LER_{norm} + t_{exp_{norm}}$. The beam interval is 200nm 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. 111

7.4 The 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 7.3. 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. 112

7.5 The CD error and LER at the top, middle and bottom layers of the resist in the test locations for the respective cases of Table 7.3. 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. 113

7.6 The 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 7.3. 114

7.7 The 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 200nm 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. 116

Chapter 1

Introduction

The efficient use of a large number of programmable beams in a massively-parallel electron-beam (e-beam) system (MPES) is crucial for maximizing the system's efficiency. This chapter provides the problem definition and background review along with the motivation, objectives, and approach of this study.

1.1 Problem definition

E-beam lithography has been known for its capability to transfer complex and fine-feature patterns directly onto a substrate.[1–5] However, the main drawback of this maskless lithography technique is its low throughput, which limits its use for the fabrication of photomasks, in particular of large size. One practical solution to the throughput limitation is to utilize multiple beams in a system. Recent developments in multi-beam systems have offered improved productivity over the conventional technology (single variable-shaped-beam system), with a mask write time of less than 10 hours.[6–10] In such a system (e.g., Multi-Beam Mask Writer-101),[6] there are a large number of beams, of which optimal use is critical to maximizing the efficiency of the system.

It is highly likely that some of the beams in the MPES are abnormal, e.g., always off or always on faulty beams, spatial and temporal fluctuations in beam current, and beam-positioning error. Hence, it is crucial to understand the behavior of beam abnormalities and develop an optimal writing method that can minimize the adverse effects of abnormal beams on the performance metrics. Another major limiting factor in e-beam lithography is the geometric distortion of written features due to electron scattering, i.e., the proximity effect, which causes critical dimension (CD) error and line edge roughness (LER). A common approach to the proximity

effect correction (PEC) is the shape and dose control method. However, the optimal adjustment of feature size for the shape correction and the optimal spatial modulation of dose for the dose correction may not be possible due to the constraints in the MPES, e.g., a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams. On the other hand, the reduction of CD error and LER during PEC may increase the total dose, which may worsen the exposing time and charging effect. Also, the optimal spatial dose distribution for the PEC may decrease the beam utilization since some beams are selectively turned off to realize non-uniform distribution. Evidently, all the MPES performance metrics are correlated. Therefore, a combined approach is essential to minimize the cost function consisting of the CD error, LER, and exposing time.

1.2 Background review

In the MPES, e.g., the eMET,[7] MBM-1000[9] and MAPPER Lithography FLX-1200 tool,[11] there can be various physical defects in the beams. Physical defects include the dose deviation due to faulty beams, spatial and temporal fluctuations in beam current, beam deviating from its target position, etc. The impact of defective beams on performance metrics was not considered in the past. One recent paper[12] discussed briefly the effect of one faulty beam in the massively parallel e-beam system. However, the main focus of this paper was on the effects of lithographic parameters, i.e., exposing interval, beam shape, etc. on different performance metrics such as exposure fluctuation, LER, maximum indent, etc. Also, this paper did not consider the non-ideal conditions, for example, fluctuating beam current density, beam-positioning error, etc. Several writing methods have been introduced for the MPES, i.e., *single-row writing I* (SRW-I), *single-row writing II* (SRW-II) and *multi-row writing* (MRW). In the single-row writing,[13] each pixel may be exposed by one beam (SRW-I) or a group of beams (SRW-II). In both methods, a beam is confined to exposing a single row, which makes the effects of abnormal beams localized in the respective rows. The multi-pass (MP) writing[14] can be employed in both SRW methods, i.e., SRW-I-MP and SRW-II-MP. However, it requires each writing path to be exposed multiple times and a positional shift of the substrate in the scanning direction in each pass. In the multi-row writing (MRW),[15] a beam exposes pixels from multiple rows in each writing path to

spread affected pixels and decrease the number of times a pixel is exposed by the same beam. These writing methods need to be compared in terms of the performance metrics, e.g., exposure variation, LER, and maximum indent, to find the optimal writing method for minimizing the effect of beam abnormalities.

The importance of PEC has been well recognized and extensively investigated, and various effective schemes have been developed to improve the accuracy of CD.[3,4,16–19] A typical approach to the PEC employs one of the dose, shape, and shape+dose corrections. In a previous study,[19] four different spatial dose distributions were considered for the PEC, i.e., uniform, V-type, A-type, and M-type. It was shown that the A-type distribution is effective when the aspect ratio (resist thickness to feature width) is relatively large, while the uniform or V-type distribution tends to work better for a relatively small aspect ratio. Also, in another study,[20] it was observed that when the edge location of a developed feature is outside the exposed area, the LER tends to be small compared to that when the edge location is inside the exposed area. In a single-beam e-beam system, whether a Gaussian or variable-shape beam, the beam size on the resist surface is or can be made small so that a sufficient spatial resolution can be achieved in both shape and dose corrections. On the other hand, in the MPES, the beam size is relatively large ($\sim$10nm), and the exposing interval is fixed, usually the same as the beam size. Therefore, the adjustment of feature size for the shape correction and the spatial modulation of dose for the dose correction can be limited leading to a non-optimal correction result, unless the optimal reduction (ΔW) of line-width is an integer multiple of the beam size. Therefore, an effective method enabling an adjustment of feature size and a spatial dose modulation at the resolution of sub-beam-size is required.

Each writing method was proposed to achieve a uniform dose, i.e., give the same dose to all pixels within a feature or features. While realizing a non-uniform dose distribution, certain beams were selectively turned off during the exposing process for a uniform dose distribution. The beams to be turned off in each step are determined based on the shape of the non-uniform dose distribution. However, this method does not guarantee the shortest exposing time possible by a MPES. This is because certain beams may be turned off in multiple cycles when they fall over the spaces between features in a pattern or when they fall over the pixels requiring a

lower dose than others. Therefore, a new writing method is required to remove the cycles with turned-off beams. One of the effective PEC methods is to adjust the spatial dose distribution in order to achieve a resist profile close to the target profile. If the utilization of beams is not considered, the maximum dose difference between any two regions can be significant, causing a lot of beams to be turned off during the exposing process. The decrease in beam utilization may increase the exposing time. However, no study has been done to incorporate the exposing time during the PEC. It is necessary to minimize the CD error and LER to achieve the highest feature resolution possible by the e-beam lithography while ensuring the minimal exposing time for the maximum throughput in the MPES.

1.3 Motivation, objectives and approaches

To achieve the optimal performance on MPES, it is crucial to minimize the negative impact of beam abnormalities and proximity effect while maximizing the utilization of beams and writing speed. However, a comprehensive study to observe the impact of defective beams on performance metrics was not conducted in the past. Also, it remains to be determined what conditions are optimal for achieving the best MP writing and MRW with a given set of lithographic parameters. Additionally, no study has been done on the comparison of performance metrics of the MP writing and MRW in the case of different beam abnormalities. Although the importance of PEC has been well recognized, no effective method of PEC has been developed considering the constraints of the MPES, such as a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams. Moreover, no study has focused on reducing exposing time while using conventional writing methods and implementing PEC. It is essential to develop an optimization method to minimize a cost function consisting of all major performance metrics since optimizing one performance metric independently can result in another performance metric being sub-optimal.

Below are the main objectives of this study and the respective approaches that are taken to achieve the objectives.

- Analyzing the effects of beam abnormalities on performance metrics for three writing methods, SRW-I, SRW-II, and MRW:

  An extensive simulation is conducted with beam abnormalities, including always-off and always-on faulty beams, spatial and temporal fluctuations in beam current, and beam-positioning error, and their effects on performance metrics, such as exposure variation, LER, and maximum indent/outdent, are observed.

- Deriving general procedures to find the optimal MP writing and MRW that minimize the effect of abnormal beams in an MPES and comparing the performance metrics between them for various beam abnormalities:

  In the case of MP writing, the optimal number of passes and amount of shift in each writing path are determined. In the case of MRW, the optimal spatial distribution of pixels to be exposed by each beam is determined. The performance metrics of the optimized MP writing and MRW methods are then compared under different types of beam abnormalities.

- Developing a systematic method for realizing various types of spatial dose distributions with an arbitrary reduction of feature size on an MPES and designing a shape+dose correction method with the dose distributions realizable on an MPES:

  The MP writing is utilized to realize a dose distribution at the resolution of sub-beam-size while considering the constraints of the MPES, such as a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams. The effectiveness of the shape+dose correction method under the constraints of the MPES is analyzed in terms of CD error, LER, and total dose.

- Developing methods to increase the beam utilization and reduce the exposing time:

  Two different approaches to increasing the beam utilization and thereby reducing the exposing time are investigated, i.e., lowering the dose difference among the regions of a feature while implementing the PEC and utilizing the cycles with all the beams turned off.

- Designing an adaptive optimization method to minimize a cost function consisting of all major performance metrics, i.e., CD error, LER, and exposing time:

  Instead of optimizing performance metrics one at a time, all the performance metrics are considered in each iteration to adjust the control parameters for obtaining a balanced result among all metrics. The optimization method is made applicable to both a single feature and a large-scale pattern.

1.4 Organization of Dissertation

The rest of this dissertation is organized as follows:

- Chapter 2 introduces the models and performance metrics for the MPES.
- Chapter 3 describes the writing methods, single-row writing I, single-row writing II, and multi-row writing, and analyzes the effects of abnormal beams for each of the writing methods.[21]
- Chapter 4 discusses the general procedures to find the optimal multi-pass writing and the optimal multi-row writing method, and compares several combinations of writing methods.[22]
- Chapter 5 introduces the realization of uniform, V-type, A-type, and M-type dose distributions for different ΔW in an MPES.[23] The shape+dose correction algorithm for minimizing the CD error and LER is also described.[23]
- Chapter 6 investigates two methods for reducing exposing time, e.g., a new writing method by controlling the deflection angle of beams and lowering the dose difference among regions of a feature during PEC.[24]
- Chapter 7 provides a detailed description of the new optimization method for both a single feature and a large-scale pattern where all the performance metrics, i.e., CD error, LER, and exposing time, are considered in each iteration.[25]
- Chapter 8 presents the concluding remarks of this Dissertation.

Chapter 2

Models and Performance Metrics in Massively-parallel E-beam Systems

In this chapter, the exposure process in the MPES is described along with the coordinate and substrate systems, and the modeling of the transfer function is explained. Also, the e-beam lithography process is described in detail. Furthermore, the performance metrics, i.e., exposure variation within a feature, critical dimension (CD) error, line edge roughness (LER), maximum indent, maximum outdent, and exposing time are explained.

2.1 Models

2.1.1 Massively-parallel electron-beam system

In the MPES (Fig. 2.1) like the eMET[7] and MBM-1000,[26] a large number of beams are generated through a 2-D array of apertures, each of which defines a beam. The beams can be individually turned on and off, and are deflected in a synchronized manner, i.e., the same deflection angle for all beams. The substrate to be exposed can be modeled as an array of square-shaped pixels with size $B \times B$, the same as the beam size on the substrate (Fig. 2.1). The exposing interval, I_{ex}, can be larger, smaller or equal to B. For simplicity, it is assumed that $I_{ex} = B$ throughout this study.

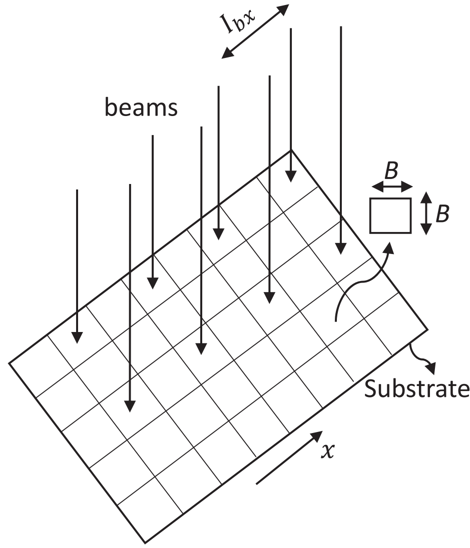

Figure 2.1. A substrate moves in the x-direction exposed by parallel beams: the beam size is $B \times B$, and the beam interval is I_{bx}.

2.1.2 Substrate system

In a typical substrate system[27] (Fig. 2.2), a resist layer is on top of the substrate, where the X–Y plane corresponds to the top surface of resist and the resist depth is along the Z-dimension. The length and width of a feature are denoted by L and W, respectively. The 3-D transfer function is denoted by $TF(x, y, z)$, describing the exposure distribution in the resist when a point (pixel) is exposed by a beam. Then, the 3-D spatial distribution of exposure, $E(x, y, z)$, can be expressed by the following convolution:

$$E(x, y, z) = \int_{y'} \int_{x'} d(x - x', y - y') TF(x', y', z) dx'dy' \tag{2.1}$$

where $d(x, y)$ represents the dose distribution given by all beams.

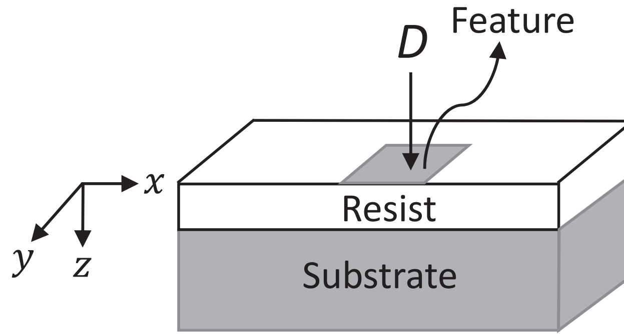

Figure 2.2. A substrate system when the feature is exposed with a dose D.

2.1.3 Transfer function

The cross-section of a beam is assumed to be of square in this study. As a beam travels from the electron source to the resist surface, the beam gets blurred.[28] Also, electrons experience scattering in the resist layer. Therefore, the transfer function is modeled by characterizing the beam blur and electron scattering, given a set of parameters (resist thickness, beam energy, etc.). The transfer function (Fig. 2.3-(c)) is generated by convolving the ideal transfer function (i.e., constant within a square of $B \times B$, see Fig. 2.3-(a)) with a Gaussian function (Fig. 2.3-(b)) of which the standard deviation σ_t, to be referred to as *blurring factor*, quantifies the level of beam blur and electron scattering. A smaller σ_t results in a transfer function closer to the ideal one and leads to a lower level of proximity effect. The σ_t may be varied explicitly to study its effects on the performance metrics.

Note that this Gaussian function is equivalent to the point spread function (PSF). The CASINO software[29] can generate point spread functions (PSFs) for different beam energies and resist thicknesses by simulating the scattering of electrons in the resist. The software allows users to define the material properties of resist, such as density and chemical composition, as well as the characteristics of the electron beam, such as beam energy and beam size. The user can also set up the simulation parameters, such as the number of electrons and the area of resist. The Monte Carlo simulation performed by CASINO involves simulating the trajectories of electrons as they interact with the resist. Through this simulation, CASINO calculates how the electrons deposit their energy in the resist, which determines the PSF.

The transfer function mainly models the forward scattering of electrons though the back-scattering is partially represented by the long tail of a Gaussian function. However, the back-scattering component of exposure has a very low amplitude and spatially varies slow. Therefore, the effect of the back-scattering component on the performance metrics (considered in this study, see Section 2.2) would be practically negligible. For comparison, the PSF generated from the Monte Carlo simulation (CASINO) and the corresponding transfer function obtained through the same process of convolution are also provided in Fig. 2.3-(d) and Fig. 2.3-(e), respectively. It is seen that the TFs generated by the Gaussian function and the PSF look very similar.

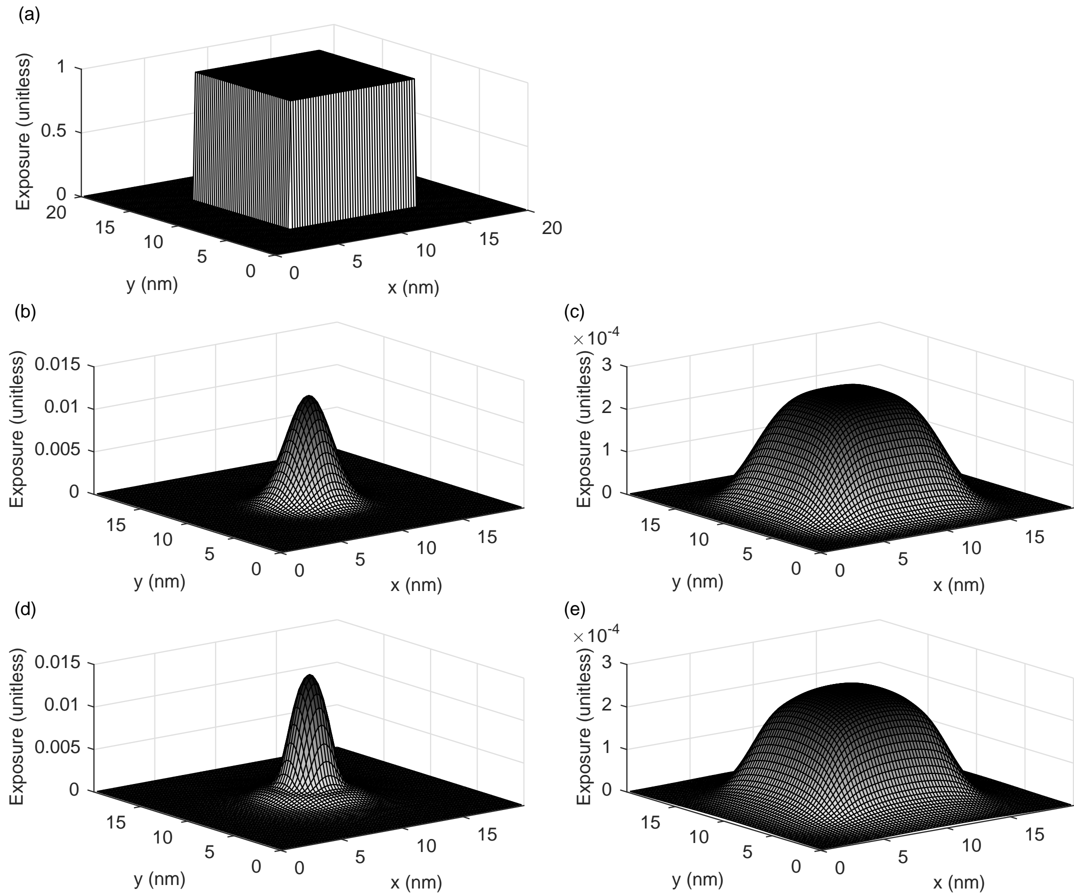

Figure 2.3. (a) The ideal transfer function for the beam aperture of size 10 nm $\times$ 10nm, (b) the Gaussian function with $\sigma_t = 1.5$nm, (c) the transfer function obtained from the Gaussian function in (b), (d) the PSF generated from the Monte-Carlo simulation (CASINO) for the substrate system of 100nm PMMA on Si and the beam energy of 30keV: the estimated σ_t is 1.22nm, and (e) the transfer function obtained from the PSF in (d).

The CD of a feature, estimated based on the stochastic exposure, can be larger than the CD based on the deterministic exposure.[30] Since the exposure in the resist is inherently stochastic, a stochastic model of transfer function is necessary to obtain realistic results. A certain percentage of Gaussian noise is added to the (deterministic) transfer function, $TF_d(x, y, z)$, to model the stochastic exposure as follows:

$$TF_s(x, y, z) = TF_d(x, y, z) + TF_n(x, y, z) \tag{2.2}$$

$$TF_n(x, y, z) = c\epsilon TF_d(x, y, z) \left(\frac{TF_d(x, y, z)}{\max_{x,y}(TF_d(x, y, z))} \right)^{\alpha} \tag{2.3}$$

where $TF_s(x, y, z)$ represents the stochastic transfer function, $TF_n(x, y, z)$ is the absolute fluctuation (noise) added to the transfer function, c is a random number from normal distribution $N(0, 1)$, ϵ is the percantage of noise to be added (e.g. 0.05 or 5%), and $0 < \alpha < 1$.

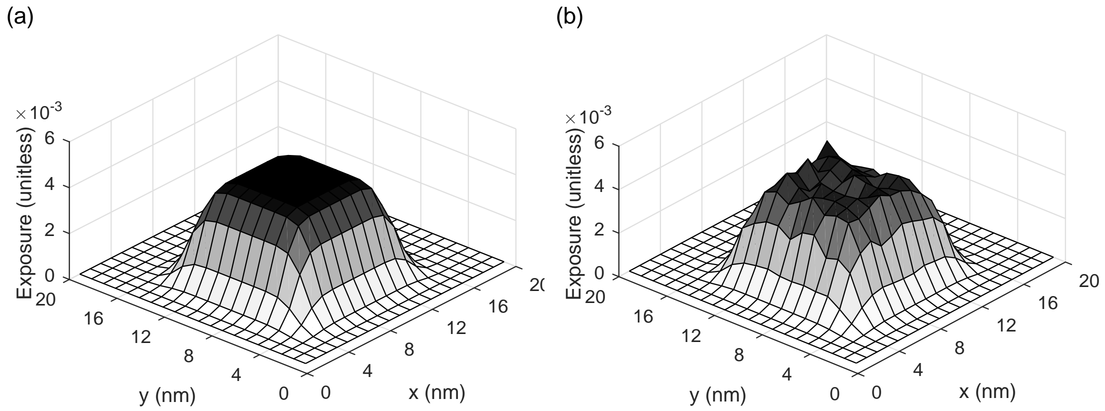

Figure 2.4. (a) The deterministic transfer function for the beam aperture of size 10nm × 10nm, and (b) the same transfer function with 7% noise added.

2.1.4 E-beam lithography process

The e-beam lithography[1–5] process starts with coating a substrate, e.g., a silicon wafer, with a layer of resist material. PMMA and HSQ are commonly used as e-beam lithography resists. The e-beam exposes the resist layer according to a desired pattern, which makes the resist undergo a chemical reaction and allows for direct writing of the pattern into the resist layer. Then, the resist is developed through chemical treatment of the substrate system using

solvents such as acetone or alcohols to complete the pattern transfer. The e-beam lithography is capable of writing patterns in the resist with extremely high resolution and used in fabrication of photomasks, low-volume production of semiconductor devices, experimental circuit patterns, etc.

2.2 Performance metrics

The performance metrics considered in this study, such as the spatial exposure variation, critical dimension (CD) error, line edge roughness (LER), maximum indent, maximum outdent, and exposing time, are described in this section.

2.2.1 Exposure variation

The exposure variation is defined as the standard deviation of exposure within a feature at a resist layer, excluding the exposure drop over the feature boundary (Fig. 2.5), which is normalized to the average exposure. It can be expressed as follows:

$$E_{SD} = \frac{1}{\overline{E}}\sqrt{\frac{1}{WL}\int_{y=0}^{L}\int_{x=0}^{W}(E(x,y,z)-\overline{E})^2 dxdy} \tag{2.4}$$

where $\overline{E}$ is the average exposure.

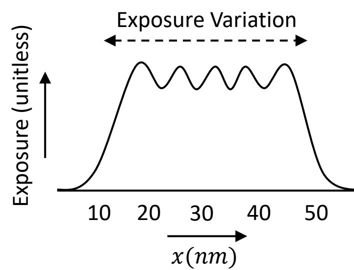

Figure 2.5. Peak-peak exposure variation. The feature size is 40nm × 120nm, and σ_t = 2nm.

2.2.2 Critical dimension error

The critical dimension (CD) error, CD_error, is defined as the absolute difference between the actual and target CDs at a resist layer, which can be expressed as follows:

$$CD_error = |CD_{actual} - CD_{target}| \tag{2.5}$$

where CD_{actual} and CD_{target} are the actual and target CDs at the resist layer, respectively, as illustrated in Fig. 2.6.

2.2.3 Line edge roughness

The line edge roughness (LER) is defined as the standard deviation of edge location along the length dimension of a feature at a resist layer (see Fig. 2.6), which can be expressed as follows:

$$LER = \sqrt{\frac{1}{L}\int_0^L (x_e(y) - \overline{x}_e)^2 dy} \tag{2.6}$$

where $x_e(y)$ is the edge location and $\overline{x}_e$ is the average edge location.

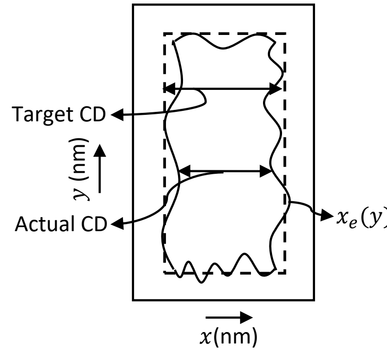

Figure 2.6. The top-down view of the remaining resist profile where the solid and dashed lines represent the actual and target resist profiles, respectively. The CD error is calculated as the absolute difference between the actual and target CD. The LER is computed as the standard deviation of edge location $x_e(y)$ from the average edge location.

2.2.4 Maximum indent and maximum outdent

Always-off faulty beams may introduce an exposure drop near the feature boundary, which can cause a large indent in the feature boundary. Such an indent is to be considered in addition to the LER since it can increase the resistance of a signal path directly. Hence, the maximum indent is defined as the largest edge deviation inward from the average edge location (Fig. 2.7(a)). The maximum indent Δx_{in} can be represented as:

$$\Delta x_{in} = \max_{y}(\mid x_{e,in}(y) - \overline{x}_e \mid) \tag{2.7}$$

where $x_{e,in}(y)$ is the edge location inside the average feature boundary.

Always-on faulty beams, exposing outside the feature boundary, can cause a significant resist development beyond the feature boundary, possibly causing a merge with a neighboring feature. Therefore, the maximum outdent is considered, which is defined as the largest edge deviation outward from the average edge location (Fig. 2.7(b)). The maximum outdent Δx_{out} can be expressed as:

$$\Delta x_{out} = \max_{y}(\mid x_{e,out}(y) - \overline{x}_e \mid) \tag{2.8}$$

where $x_{e,out}(y)$ is the edge location outside the average feature boundary.

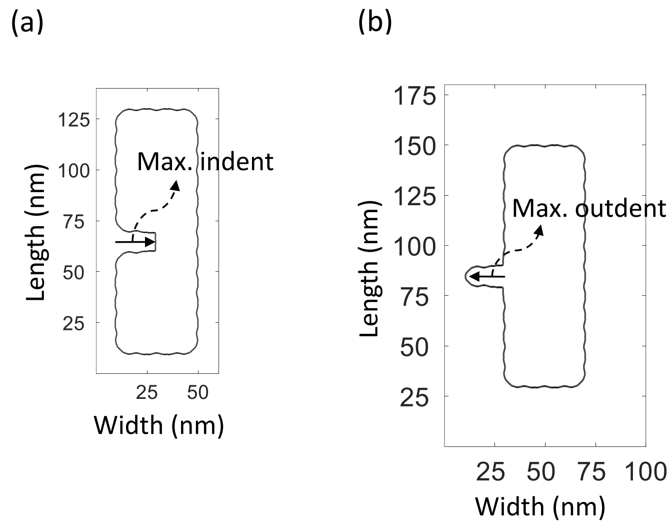

Figure 2.7. (a) maximum indent and (b) maximum outdent. The feature size is 40nm $\times$ 120nm and σ_t = 2nm.

2.2.5 Exposing time

The exposing time of a circuit pattern is the amount of time required for the beams on the MPES to deliver the target doses to all the pixels in the pattern. It is calculated in *cycles*, with each cycle representing the duration for which a beam follows a pixel to deliver the dose of

$n_s d$, where n_s is the number of steps and d is the dose given to a pixel by the beam in a step (refer to Section 3.1 for more detail).

Chapter 3

Effects of Abnormal Beams on Performance Metrics

In this chapter, the effects of beam abnormalities on performance metrics are analyzed for three different writing methods: single-row writing I, single-row writing II, and multi-row writing.[21] The beam abnormalities include always-off and always-on faulty beams, spatial and temporal fluctuations in beam current, and beam-positioning error. The performance metrics considered are exposure variation, LER, and maximum indent/outdent.

3.1 Writing methods

The three writing methods, single-row writing I, single-row writing II and multi-row writing, to be compared in terms of the effects of abnormal beams, are briefly described in this section. The dose a beam gives to a pixel in a *step* is denoted by d and the beam interval, I_{bx}, is considered an integer multiple of B. All beams are controlled in a synchronized manner.

In the single-row writing method I[13] (Fig. 3.1), the writing path of each beam is on a single row of pixels. As the substrate moves continuously underneath the array of beams, each beam follows a pixel, being deflected in each step, to give a total (or target) dose of $D = n_s d$ to the pixel through n_s steps. The duration of n_s steps to give the target dose of D to each pixel is referred to as a *cycle*. Then, the beam is reset back to its vertical orientation and exposes another pixel in the next cycle. It is assumed that the time required to reset beams is much shorter than a step. The condition of $n_s = \frac{I_{bx}}{B} + 1$ is required to be able to achieve a uniform dose distribution without any hole in an exposed region. When a pixel is exposed by a (always-off) faulty beam, the pixel receives no dose, i.e., the dose reduction for the pixel would be

$\Delta D = n_s d$. Each pixel may be exposed through more than one cycle using n_c sets of beams where n_c is the number of cycles.

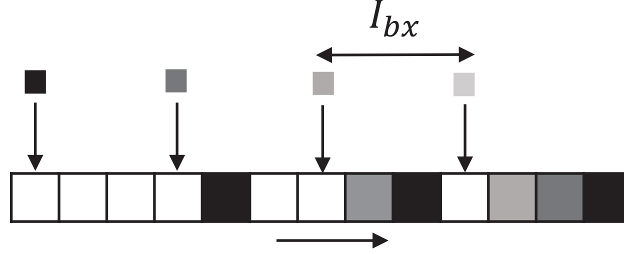

Figure 3.1. A row of pixels (larger squares) being exposed by 4 beams (smaller squares) in the single-row writing method I where $I_{bx} = 3B$ and $n_s = 4$, and the substrate moves to the right: this is a snapshot at the beginning of 1st step of a cycle. The pixels with the same shade of gray are exposed by the same beam with the corresponding shade.

In the single-row writing method II[13] (Fig. 3.2), the writing path of each beam is still on a single row of pixels. But, the main difference from the single-row writing method I is that each pixel is exposed jointly by a group of n_g consecutive beams, instead of a beam. This provides a better fault tolerance, decreasing the dose reduction by a faulty beam to $\Delta D = \frac{n_s}{n_g} d$. I_{bx} refers to the beam interval within each group of beams and I_{gx} the group interval which is the distance between groups. Note that beams need to be reset back to their vertical orientation after every $\frac{n_s}{n_g}$ steps, i.e., n_g times per cycle.

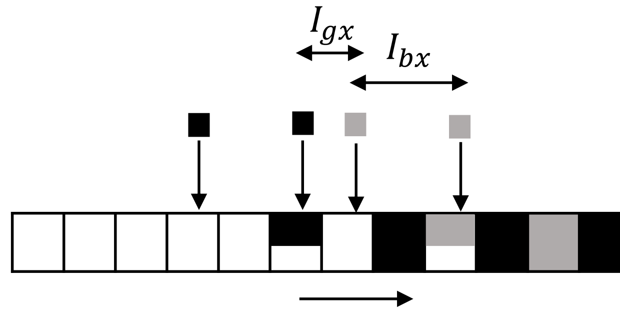

Figure 3.2. A row of pixels (larger squares) being exposed by 4 beams (smaller squares) in the single-row writing method II where $n_s = 4$, $I_{bx} = 2B$, $n_g = 2$ and $I_{gx} = B$, and the substrate moves to the right. A group of two beams exposes each pixel and this is a snapshot right before the 2nd beam in each group starts to expose the pixel partially exposed by the first beam. The pixels with the same shade of gray are exposed by the group of beams with the corresponding shade.

In the multi-row writing method[15] (Fig. 3.3), the writing path of each beam is over multiple rows of pixels and each pixel is exposed by a particular beam only once. That is, each beam exposes a set of n_s pixels, each pixel once, distributed over multiple rows. If the target dose for a pixel is $D = n_s d$, it is exposed by n_s different beams in a row. Therefore, the dose reduction for a pixel due to a (always-off) faulty beam is minimized, i.e., $\Delta D = d$ independent of n_s. This method also enables the de-localization of the affected pixels. There can be more than one combination of selecting n_s pixels over multiple rows. In this study, the one shown in Fig. 3.3 is used, where the writing path consists of $\frac{I_{bx}}{B} + 1$ rows and each beam exposes one pixel from each row, being deflected n_s times in the diagonal direction.

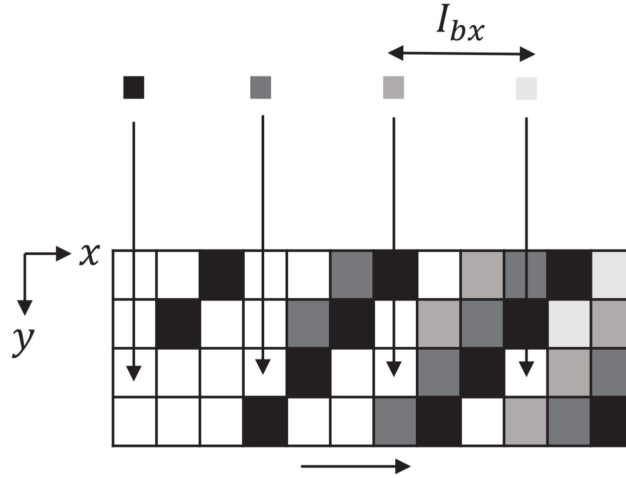

Figure 3.3. 4 rows of pixels (larger squares) being exposed by 4 beams (smaller squares) in the multi-row writing method where $I_{bx} = 3B$ and $n_s = 4$, and the substrate moves to the right. This is a snapshot at the beginning of the 3rd step of a cycle. The pixels with the same shade of gray are exposed by the beam with the same shade and each pixel is exposed by a beam, once in a cycle.

3.2 Simulation

A feature of width 80nm and length 320nm is exposed by the MPES along its width dimension on a typical substrate system. The cross-section of beam at the resist surface is 10nm × 10nm (B = 10nm), and the exposing interval I_{ex} is 10nm ($I_{ex} = B$). The beam interval is 30nm ($I_{bx} = 3B$) and, therefore, the number of beams in one set is 4 ($n_s = \frac{I_{bx}}{B} + 1$). For the single-row writing method II, n_g is set to 2.

For modeling the transfer function, a 3-D point spread function is generated using a Monte Carlo simulation program CASINO[29] for the substrate system of 100nm PMMA on Si and the beam energy of 50keV. From the point spread function, the total exposure and the forward scattering range are extracted in each of 5 resist layers which set the total exposure and blurring factor of Gaussian function used to generate the transfer function of the respective layer through the convolution as described earlier.

The 3-D exposure distribution in the resist is computed at the resolution, I_{sm}, referred to as *simulation interval*, which is set to $\frac{1}{4}$nm so that a detailed analysis is possible. From the exposure distribution, the exposure variation is computed. The developing-rate distribution is derived from the exposure distribution and then the remaining resist profile is obtained through a resist-development simulation.[27] The development simulation continues until the feature is fully developed to the bottom layer of resist. The dose is adjusted such that the width of developed feature at the middle layer is as close to the target width as possible when there is no abnormal beam in the system. From the resist profile, the CD, LER, maximum indent, and maximum outdent are measured. The middle 80% segment of the feature is used in the computation of the metrics to exclude the edge effect (rounding at corners). Note that all the performance metrics are averaged over 10 typical cases.

The MPES has a large array of beams, e.g., 512 $\times$ 512 beams. Though the probability p_b of a beam being faulty (always-on or always-off) is very small, the number of faulty beams in a system can be substantial. However, out of 128 beams required to expose the feature considered in this study, the average number of faulty beams is well under 1 for the realistic value of p_b. Therefore, in this analysis, the number of faulty beams affecting the feature, denoted by n_f, is considered instead of p_b.

The parameters which are varied when considering different beam abnormalities are described below.

Always-Off and Always-On Faulty Beams

The number of always-off or always-on beams n_f is varied from 0 to 5. Pixels in each writing path are exposed by one set of beams. Both the worst case(when n_f faulty beams are in

the same row) and the best case (when none of the faulty beams exposes any boundary pixel) are considered. Also, to observe the effect of multiple sets of beams exposing each pixel, the number of sets of beams is varied from 1 to 5, while the number of faulty beams is kept at 1.

Spatial Fluctuation of Beam Current

The beam current may not be the same for all beams (Fig. 3.4). This spatial fluctuation is assumed to follow a truncated Gaussian distribution with a standard deviation σ, where the maximum fluctuation is limited to 20% of the average current (or dose per step), and σ is varied from 0.5% to 10%.

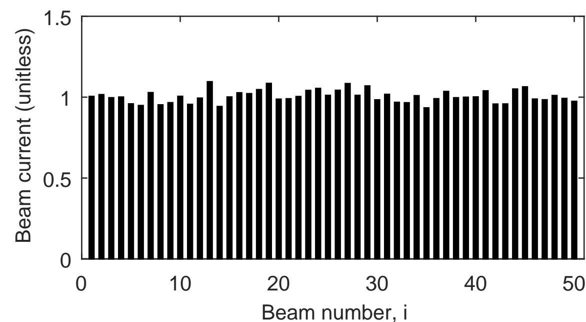

Figure 3.4. An example of spatial fluctuation of beam current following a Gaussian distribution with the average of 1 and standard deviation of 5%.

Temporal Fluctuation of Beam Current

The current of a beam may fluctuate with time (Fig. 3.5). This temporal fluctuation of beam current is also assumed to follow a truncated Gaussian distribution with a standard deviation σ, where the maximum fluctuation is limited to 20% of the average current (or dose per step), and σ is varied from 0.5% to 10%. The average currents of all beams are assumed to be the same to exclude the effect of spatial fluctuation.

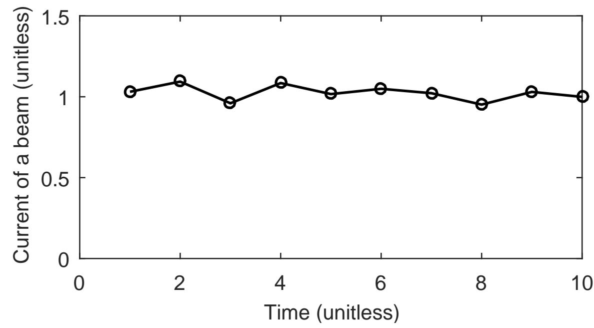

Figure 3.5. An example of temporal fluctuation of beam current following a Gaussian distribution with the average of 1 and standard deviation of 5%.

Beam-positioning Error

The actual point of exposure of a beam may deviate from its target position, in both x and y directions (Fig. 3.6). The beam-positioning error in both directions is assumed to follow a truncated Gaussian distribution with a standard deviation σ, and a maximum deviation of 5nm from its center. σ is varied from 0 to 1nm.

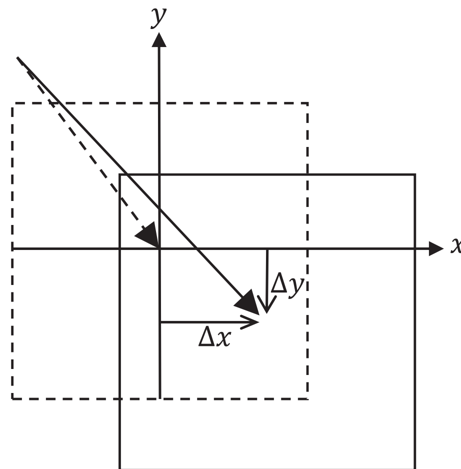

Figure 3.6. Beam-positioning error. The dotted line represents the target exposure of a beam while the solid line represents the actual exposure of the beam deviating from the target position by $(\Delta x, \Delta y)$.

The values of all the parameters considered in this study are provided in Table 3.1.

Parameter	Value
Resist (PMMA) layer	Middle
Resist (PMMA) thickness	100nm
Beam energy	50KeV
Feature size	$80nm \times 320nm$
Beam size, B	10nm
Beam interval, I_{bx}	30nm
Total dose, D	$4d$
Blurring factor, σ_t	1 to 4nm
# of faulty beams, n_f	Varied from 0 to 5
Spatial beam current fluctuations	Gaussian. σ is varied from 0.5% to 10%
Temporal beam current fluctuations	Gaussian. σ is varied from 0.5% to 10%
Beam-positioning error	Gaussian. σ is varied from 0 to 1nm

TABLE 3.1. A summary of simulation set-up

3.3 Results and discussion

The simulation results considering only the always-off faulty beams are first presented. Then, the improvement by using multiple sets of beams is demonstrated. After that, the simulation results considering each of always-on faulty beams, spatial fluctuation of beam current, temporal fluctuation of beam current, and beam-positioning error are discussed separately. All the performance metrics are evaluated at the middle layer of resist for $\sigma_t = 2$nm except when σ_t is varied. The metrics show similar tendencies at other layers.

3.3.1 Always-off faulty beams

The simulation results of the three performance metrics obtained with n_f varied are provided in Fig. 3.7. As n_f increases, the exposure variation becomes larger due to the increasing number of affected pixels. A larger variation of exposure tends to result in a rougher feature boundary and more affected pixels close to the feature edges make an indent larger. Therefore, for a larger n_f, the LER and maximum indent become larger. It is also observed that all of the performance metrics are significantly improved by the multi-row writing method, compared to both single-row writing methods. The main reason is that the dose reduction at a pixel due to

faulty beams is lower in the multi-row writing. Between the two single-row writing methods, the single-row writing method II performs better since the dose reduction is n_g times lower. Since n_g is set to 2 in the simulation set-up, the affected pixels receive only 50% of the target dose and, therefore, remain mostly underdeveloped. As a result, the LER and maximum indent are not improved much by the single-row writing method II. The feature contours in the remaining resist profiles obtained by the three writing methods when there is a faulty beam affecting edge pixels are compared in Fig. 3.8. It is clear that the maximum indent is significantly smaller for the multi-row writing method.

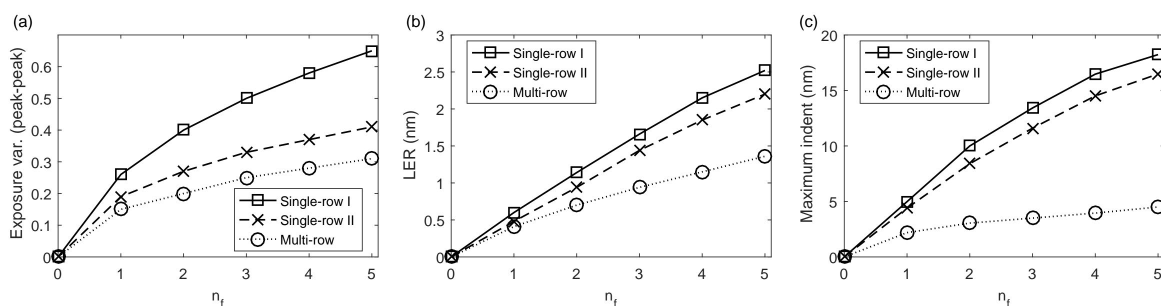

Figure 3.7. (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm.

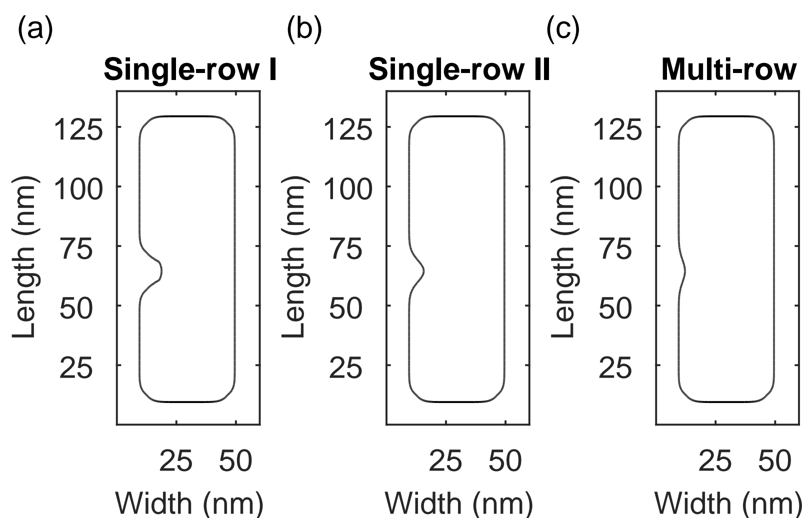

Figure 3.8. Contour of a rectangular feature (40nm × 120nm) when a faulty beam affects edge pixels and $\sigma_t = 2$nm: (a) single-row writing method I, (b) single-row writing method II, and (c) multi-row writing method

3.3.2 Multiple sets of beams

From Fig. 3.9, it is seen that, as n_c increases, the exposure variation becomes smaller since the dose reduction at a pixel due to a (always-off) faulty beam is decreased by the factor of n_c. Also, the decreasing rate of exposure variation becomes lower with increasing n_c, since the percentage increase of the total dose at a pixel gets smaller as n_c increases. The LER and maximum indent also become smaller for a larger n_c. In Fig. 3.9, a large drop in the LER and maximum indent is observed for the multi-row writing method when n_c is increased from 1 to 2, and for the single-row writing method II when n_c is increased from 3 to 4. In both cases, the total dose received by the affected pixels is increased close to 90% of the target dose. Therefore, most parts of the affected pixels are developed, which improves the LER and maximum indent substantially. However, for the single-row writing method I, even when $n_c = 5$, the affected pixels receive well below 90% of the target dose. Hence, the affected pixels remain mostly underdeveloped, and the improvement in the LER and maximum indent is marginal, although the affected pixels get higher dose with increasing n_c.

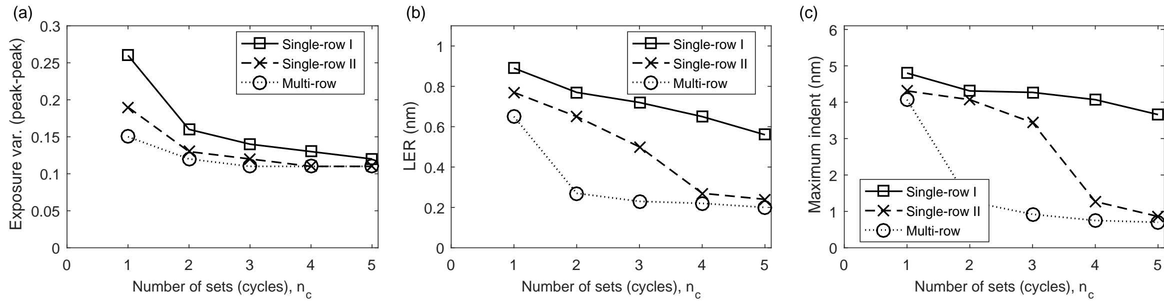

Figure 3.9. (a) Exposure variation, (b) LER, and (c) maximum indent with the number of cycles (n_c) varied. $\sigma_t = 2$nm, $n_f = 1$.

3.3.3 Always-on faulty beams

The simulation results of the three performance metrics with the number of always-on faulty beams (n_f) varied are plotted in Fig. 3.10. The increase in the exposure variation with n_f is very small because always-on faulty beams do not affect the exposure distribution inside the feature boundary except edge pixels. However, the pixels affected by always-on beams just outside the feature boundary can contribute to the LER significantly, and consecutive affected

pixels can make the outdent extremely large. In the single-row writing methods, each pixel is exposed by a beam multiple times. When the pixels just outside the feature boundary are exposed by an always-on beam, it causes the resist to develop well beyond the target boundary, and results in a larger increase of the LER and maximum outdent. On the other hand, in the multi-row writing method, each beam exposes a pixel only once and, therefore, the LER and maximum outdent due to always-on beams are smaller compared to both single-row writing methods.

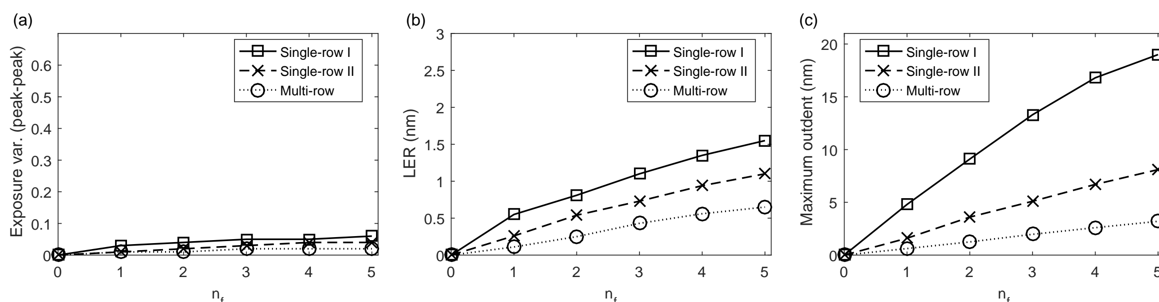

Figure 3.10. (a) Exposure variation, (b) LER, and (c) maximum outdent with the number of always-on faulty beams (n_f) varied. $\sigma_t = 2$nm.

3.3.4 Spatial fluctuation of beam current

Fig. 3.11 shows the effects of spatial fluctuation of beam current on the performance metrics. As the spatial fluctuation increases, the spatial variation of exposure becomes more significant, resulting in larger LER and maximum indent. Also, the improvement in the performance metrics by the multi-row writing method compared to both single-row writing methods is notable. This is because, in the multi-row writing method, each pixel is exposed by a larger number of beams than the single-row writing methods and therefore different doses from multiple beams have a better chance to be averaged out. The number of beams exposing each pixel in the single-row writing method II is n_g times larger than that in the single-row writing method I and therefore the performance metrics are significantly better for the former. The feature contours in the remaining resist profiles obtained by the three writing methods with a spatial fluctuation of beam current are compared in Fig. 3.12. Evidently, the edge of the feature is smoother for the multi-row writing method.

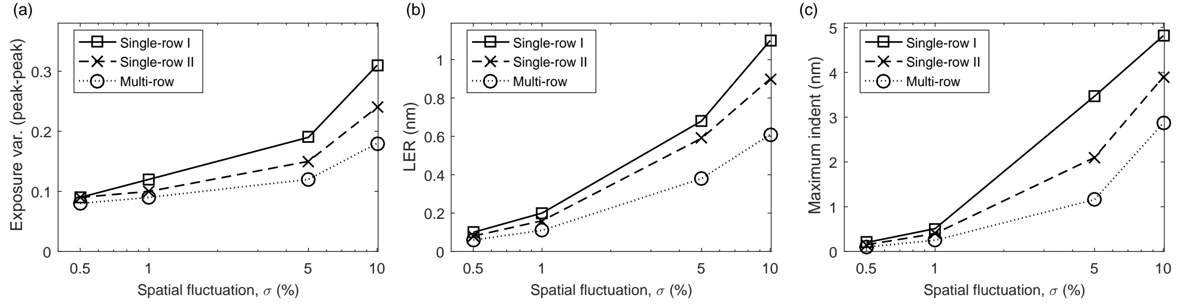

Figure 3.11. (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

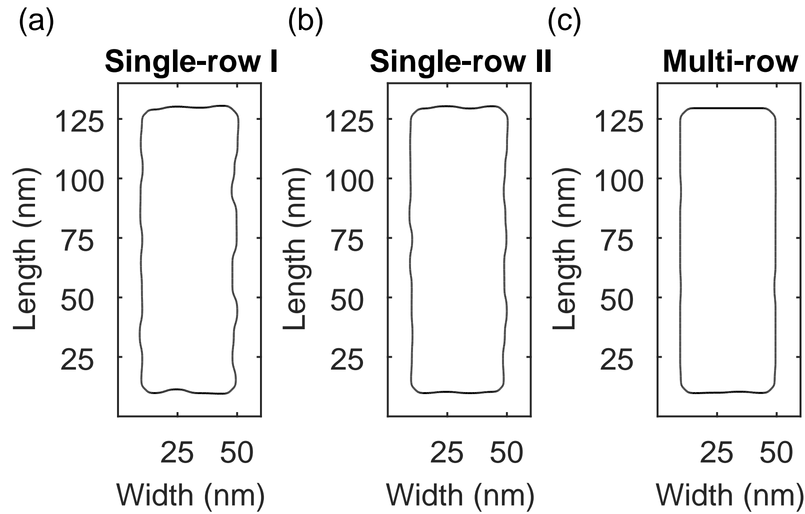

Figure 3.12. Contour of a rectangular feature (40nm $\times$ 120nm) when the spatial fluctuation of beam current is 10% with no faulty beam and $\sigma_t = 2$nm: (a) single-row writing method I, (b) single-row writing method II, and (c) multi-row writing method.

3.3.5 Temporal fluctuation of beam current

In Fig. 3.13, the dependency of the performance metrics on the temporal fluctuation of beam current is shown. It is observed that all three metrics are the same for all three writing methods. In all three writing methods, each pixel is exposed the same number of times, either by one or multiple beams, leading to the same level of time-averaging of the temporal fluctuation. Another observation is that the metrics with the temporal fluctuation of beam current are substantially smaller than those with the spatial fluctuation for the single-row writing methods,

but not for the multi-row writing method. This is because there is always a time-averaging effect when a pixel is exposed multiple times. But, for a pixel not exposed by multiple beams, there is no averaging effect to compensate the spatial fluctuation.

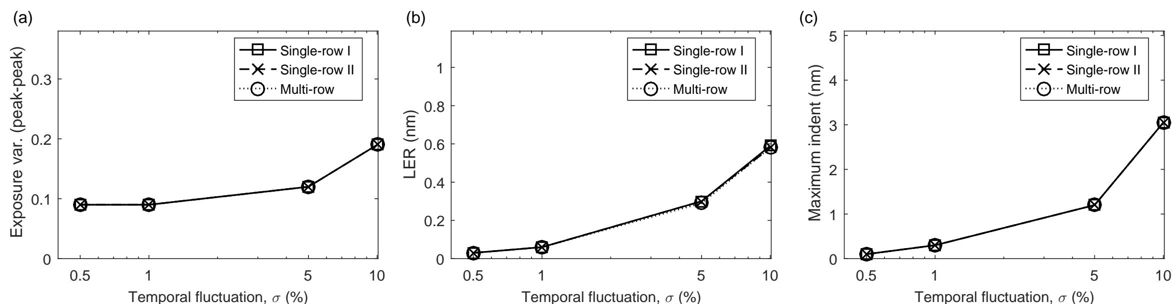

Figure 3.13. (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

3.3.6 Beam-positioning error

The simulation results of the three performance metrics obtained with the beam-positioning error (σ) are given in Fig. 3.14. The more a beam deviates from its target position, the greater the exposure variation, LER, and maximum indent. Also, all the performance metrics are improved greatly by the multi-row writing method compared to both single-row writing methods. In the single-row writing method I, when a beam deviates from its target position, all the shots from that beam received by the affected pixels are out-of-position. In the single-row writing method II, the effect of beam-positioning error is averaged among n_g different beams. In the multi-row writing method, the level of averaging effect of the beam-positioning error is the highest among the three methods, resulting in the lowest exposure variation, LER, and maximum indent.

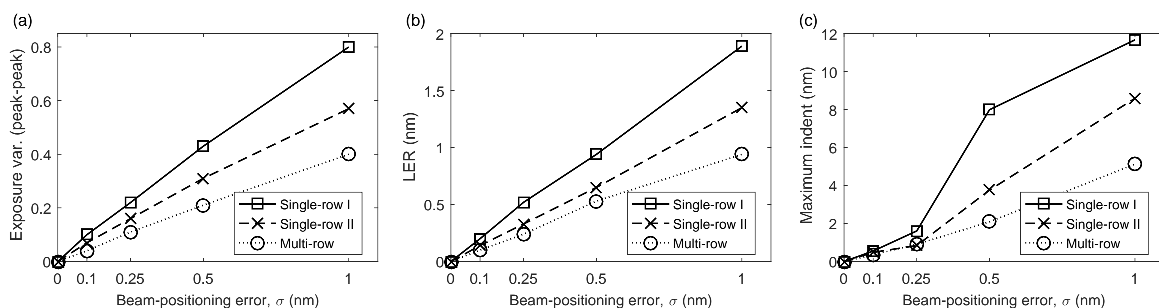

Figure 3.14. (a) Exposure variation, (b)LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$.

3.3.7 Dependency on σ_t

All the results so far are for $\sigma_t = 2$nm. In Fig. 3.15, the results with σ_t varied are provided for the case when there is one always-off beam with no other abnormalities. The transfer function with a larger σ_t spreads more and drops slower from its maximum. Therefore, when σ_t is larger, the pixels exposed by the always-off beam would receive more exposure contributions from nearby normal shots. Therefore, the performance metrics get improved (reduced) as σ_t increases. The same trend is expected for other abnormalities.

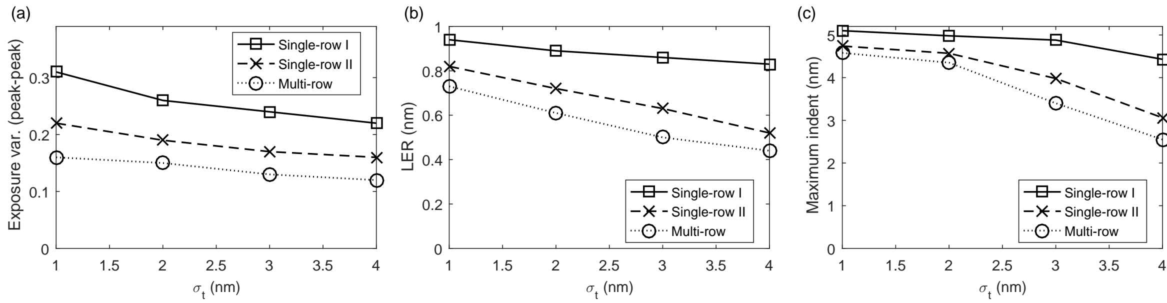

Figure 3.15. (a) Exposure variation, (b) LER, and (c) maximum indent with the blurring factor σ_t varied. $n_f = 1$.

3.4 Summary

In this chapter, the effects of abnormal beams on the performance metrics are analyzed in terms of the performance metrics and the writing methods, single-row writing I, single-row writing II, and multi-row writing, are compared. The study shows that the multi-row writing performs better than the single-row writing methods I and II because of its capability to spread the effects of abnormal beams spatially. The next chapter includes the comparison of the multi-row writing with the multi-pass writing method since the multi-pass writing is also able to mitigate the effects of faulty beams by exposing each row of pixels multiple times.

Chapter 4

Effectiveness of Multi-pass and Multi-row Writing Methods

In this chapter, general procedures are derived to find the optimal multi-pass writing and multi-row writing which minimize the effect of abnormal beams in the MPES, and various combinations of single-row and multi-row writing methods along with the multi-pass writing are compared in terms of the performance metrics.[22] The beam abnormalities considered in this chapter are always-off faulty beams, spatial and temporal fluctuations in beam current, and beam-positioning error.

4.1 Optimization of writing methods

In the single-row writing methods (SRW-I or SRW-II),[13] the writing path of each beam is confined to a single row of pixels. In the multi-row writing (MRW),[15] the writing path of each beam is over multiple rows of pixels and each pixel is exposed by a beam only once. That is, each beam exposes a set of n_s pixels, each pixel once, distributed over multiple rows. If the target dose for a pixel is $D = n_s d$, each pixel is exposed by n_s different beams in a row. There are more than one way of selecting a set of n_s pixels over multiple rows, where the set is referred to as *pattern*. A basic requirement for a pattern to achieve a uniform dose is *stackabilty*,[15] that is, being able to replicate and stack the pattern in the X dimension without a hole or overlap.

In order to optimize the performance metrics, the effect of abnormal beams needs to be spread maximally. The procedure to minimize the localization of pixels affected by abnormal beams is described for each of the MP writing and MRW in the following. In both cases, it is assumed that $n_s = \frac{I_{bx}}{B} + 1 = 4$ and the exposing interval, $I_{ex} = B$. The positional shift, Δx,

in the MP writing is defined as the absolute amount of positional shift in the direction of the substrate after each pass. The number of passes is denoted by n_p.

4.1.1 Multi-pass (MP) writing

In the multi-pass (MP) writing,[14] the effect of abnormal beams is delocalized by increasing the number of passes n_p with a positional shift Δx of the substrate. The relation between n_p and Δx required for the optimal MP writing needs to be determined. In Fig. 4.1, two different realizations of the MP writing with the SRW-I method are illustrated. When $\Delta x = B$ as in Fig. 4.1(a), all pixels in the writing path are exposed by each beam through n_s passes. Increasing n_p beyond n_s results in certain pixels exposed by the same beam more than once. If that beam is defective, the negative effect on those pixels will be larger. Hence, the optimal number of passes in this case is n_s. When $\Delta x = 2B$ as in Fig. 4.1(b), every other pixel is exposed by the same beam once after $\frac{n_s}{2}$ passes. Increasing n_p beyond $\frac{n_s}{2}$ results in specific pixels exposed by the same beam twice. Therefore, in general, when $\Delta x = nB$, the optimal number of passes is n_s/n where n is an integer. But, when $\Delta x = B$, the negative effect of an abnormal beam is spread most (to the entire row of pixels) with n_s passes. Therefore, when $\Delta x = B$ and $n_p = n_s$, the MP writing is optimized.

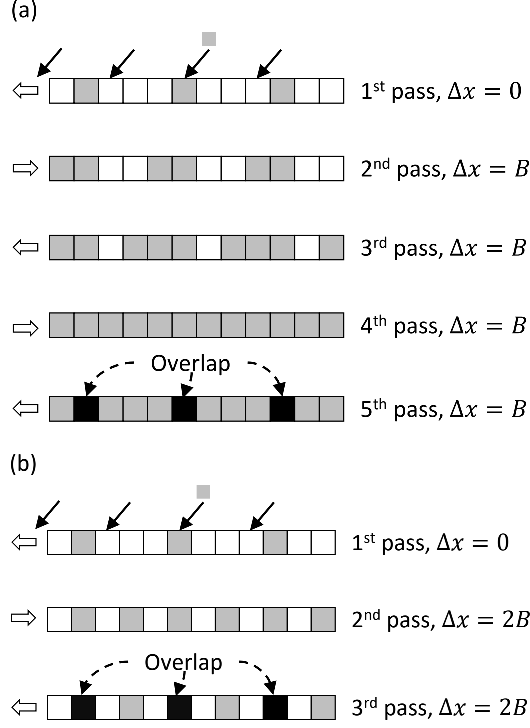

Figure 4.1. An illustration of the pixels exposed by a beam in the multi-pass (MP) writing where (a) $\Delta x = B$, and (b) $\Delta x = 2B$. The gray pixels represent the pixels exposed once by the beam with a small gray square on its top. The black pixels represent the pixels exposed by the same gray beam more than once. In both cases, $n_s = 4$ and $I_{bx} = 3B$.

4.1.2 Multi-row writing (MRW)

An optimal MRW pattern needs to be selected such that the distance among the pixels exposed by the same beam is maximized. In this way, the localization of pixels affected by an abnormal beam is minimized. Let $(\Delta H, \Delta V)$ denote the displacement vector where ΔH and ΔV represented in pixel are the horizontal and vertical displacements between the two pixels exposed consecutively by a beam (assuming that pixels in a pattern are exposed from the top to bottom row of the pattern). Also, $(\Delta X, \Delta Y)$ denotes the displacement vector between the two closest pixels exposed by the same beam at the completion of exposing process. The larger the $\sqrt{\Delta X^2 + \Delta Y^2}$ is, the less the localization of affected pixels is. In Fig. 4.2, four different MRW patterns for $n_s = 4$ and $I_{bx} = 3B$ are shown. In all patterns, $\Delta V = 1$ to achieve the stackability in one pass. When ΔH is less than $\frac{n_s}{2}$ as in Fig. 4.2(a) or greater than $\frac{n_s}{2}$ as in Fig. 4.2(b), ΔX is less than $\frac{n_s}{2}$. On the other hand, when $\Delta H = \frac{n_s}{2}$ as in the patterns of Figs. 4.2(c) and 4.2(d), $\Delta X = \frac{n_s}{2}$. Hence, when $\Delta H = \frac{n_s}{2}$, the distance $\sqrt{\Delta X^2 + \Delta Y^2}$ is maximized.

Therefore, in general, the MRW pattern with $\Delta H = \frac{n_s}{2}$ performs better than any other MRW patterns.

Let w_p denote the width of a pattern represented in pixel. The two patterns in Figs. 4.2(c) and 4.2(d) have the same $\Delta H = \frac{n_s}{2}$, but different w_p, i.e., $w_p = \frac{n_s}{2} + 1 = 3$ in the former pattern and $w_p = 3\frac{n_s}{2} + 1 = 7$ in the latter pattern. Therefore, the deflection angles of beams to realize the pattern in Fig. 4.2(c) are smaller. Also, the exposing time is shorter for the pattern of Fig. 4.2(c) compared to the pattern in Fig. 4.2(d) where the initial and final latencies of exposing a feature are longer. Hence, the optimal MRW pattern is obtained when $\Delta H = \frac{n_s}{2}$ and $w_p = \frac{n_s}{2} + 1$.

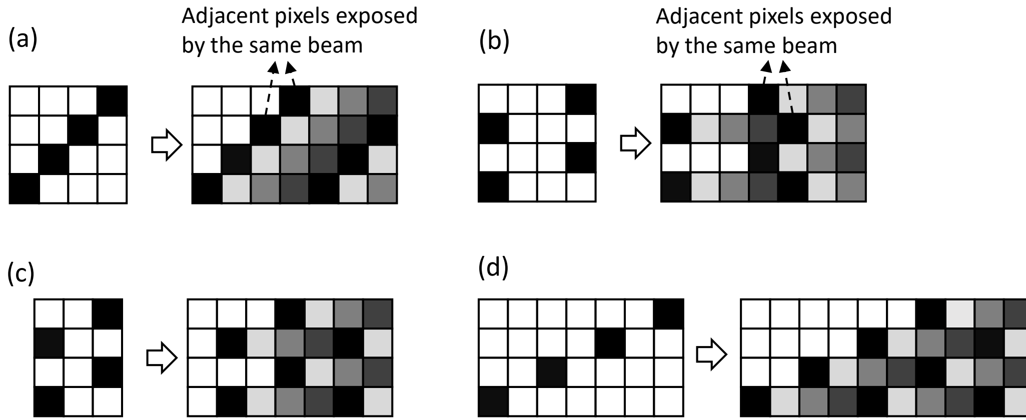

Figure 4.2. Different realizable patterns for the MRW with $n_s = 4$ and $I_{bx} = 3B$, where (a) $\Delta H = 1$ and $w_p = 4$, (b) $\Delta H = 3$ and $w_p = 4$, (c) $\Delta H = 2$ and $w_p = 3$, and (d) $\Delta H = 2$ and $w_p = 7$. In all cases, $\Delta V = 1$.

4.2 Simulation

A single line feature of width 100nm and length 500nm is exposed by the MPES along its width dimension on a typical substrate system. The resist is modeled by five layers to consider the layer dependency of exposure. The beam cross-section at the surface of resist is 10nm $\times$ 10nm (B = 10nm). Therefore, the exposing interval, $I_{ex} = B = 10$ nm. All the pixels are given the dose, $D = 10d$ ($n_s = 10$). The number of beams in a set is 10 and the beam interval, $I_{bx} = 9B = 90$ nm ($n_s = \frac{I_{bx}}{B} + 1$). For the SRW-II, n_g is set to 2.

The TF is modeled based on the 3-D point spread function (PSF) generated using a Monte Carlo simulation program CASINO[29] for the substrate system of 100 nm PMMA on Si, the beam energy of 50 keV, and the beam diameter of 6nm. The total exposure and forward scattering range (the standard deviation of Gaussian) are extracted from the PSF in each of the five resist layers. The ratios of the total energy and forward scattering range among the five resist layers are referred to in setting the total exposure and σ_t of the Gaussian function used to generate the TF of each layer.

The 3-D exposure distribution in the resist is computed at the resolution I_{sm}, referred to as *simulation interval*, which is set to $\frac{1}{2}$nm. The exposure variation is calculated from the exposure distribution. The developing-rate distribution is obtained from the exposure distribution and then the remaining resist profile is obtained through a fast path-based simulation of resist development.[27] The development simulation continues until the resist is fully developed to the bottom layer. The total dose is set such that the CD of developed feature at the middle layer without considering the beam abnormalities is as close to the target CD as possible. To measure the LER and maximum indent, the middle 80% segment of the developed feature along the length dimension is considered to exclude the edge effect (corner rounding). All the performance metrics are averaged over simulation runs of 10 typical cases.

The parameters which are varied when considering different beam abnormalities are described below.

Always-off faulty beams

The probability of a beam being faulty is very low, e.g., 0.05%, in an MPES. However, due to the large number of beams, there would be always a few defective beams (statistically distributed), of which the beam currents are lower than normal. Such beams are considered to be always-off in the simulation. The number of faulty beams, n_f, is varied from 0 to 5.

Spatial fluctuation of beam current

As electrons travel from the electron source to the 2D array of beam apertures, they experience blurring, which results in a high beam current density in the center of the aperture

array and a gradually decreasing beam current density away from the center. This distribution of beam currents may be modeled by clipping a square array from the center of a 2D Gaussian distribution as shown in Fig. 4.3(a). The value associated with each element (i, j) of the array represents the current of beam (i, j). The standard deviation σ of all the beam currents represents the level of spatial fluctuation. σ is varied from 0.5% to 20% where the maximum fluctuation is limited to 40% of the average current (or dose per step).

Temporal fluctuation of beam current

The current of a beam may vary (fluctuate) with time. To exclude the spatial fluctuation, it is assumed that the temporal fluctuation is the same for all beams. In this study, the case where the beam current slowly drifts is considered, i.e., the current (dose) of a beam increases linearly in each step from a minimum to a maximum value. The standard deviation σ of the dose given by a beam in each step over the entire exposing process is used to represent the level of temporal fluctuation. σ is varied from 0.5% to 20% where the maximum fluctuation is limited to 40% of the average current (or dose per step).

Beam-positioning error

The actual point of exposure of a beam may deviate from its target position, in both x and y directions, due to the imperfect deflecting electric and magnetic fields. The deviation of each beam, i.e., beam-positioning error, may be expressed by a vector $(\delta x, \delta y)$. While various spatial distributions of beam-positioning error may be considered, this study deals with only a spatially correlated beam-positioning error, which may be modeled to follow a parameterized function. The vector field of beam-positioning error shown in Fig. 4.3(b), considered in this study, is generated based on the equation of ellipse $ax^2 + by^2 = 1$ where a and b are varied. The standard deviation σ of $\delta r = \sqrt{\delta x^2 + \delta y^2}$ among all beams represents the overall level of beam-positioning error. σ is changed from 0 to 1nm with a maximum deviation of 5nm from the target position.

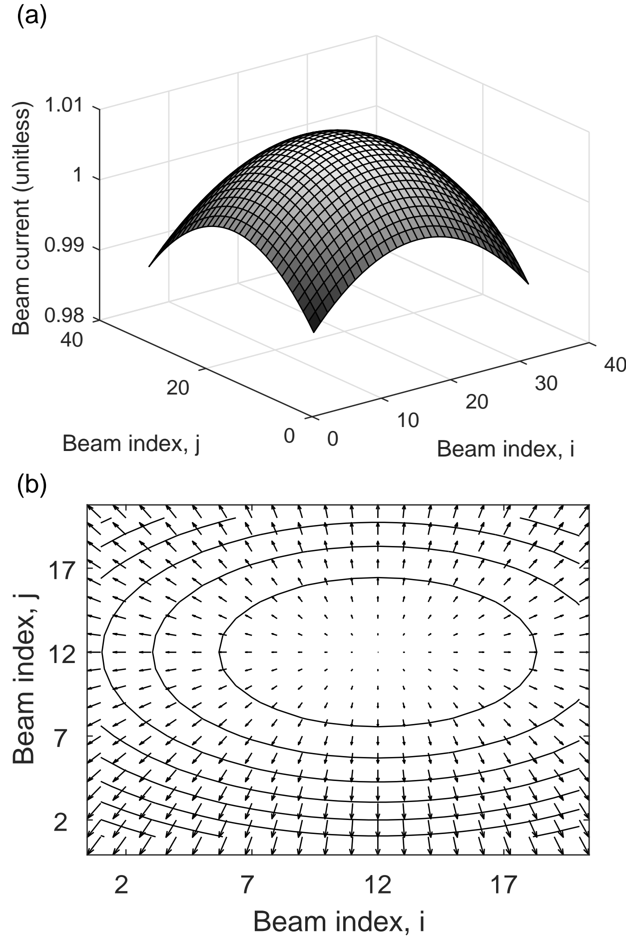

Figure 4.3. Beam abnormalities: (a) an example of the spatial fluctuation of beam currents, where the doses given by the beams are slowly decreasing from the center of the aperture array, and (b) a vector field of the beam-positioning error guided by the concentric ellipses, $ax^2 + by^2 = 1$ where a and b are varied.

4.3 Results and discussion

In this section, the results are presented in three parts. First, three different implementations of the MP writing in the SRW-I, i.e., SRW-I-MP, are compared. Then, simulation results for two different patterns of the MRW are discussed. Finally, the results from an extensive simulation comparing the SRW-I, SRW-II, and MRW with and without the MP writing are presented.

4.3.1 Multi-pass (MP) writing

The three performance metrics, e.g., exposure variation, LER, and maximum indent, obtained with n_f varied are provided in Fig. 4.4 for three different implementations of the SRW-I-MP, i.e., ($n_p = 10$, $\Delta x = B$), ($n_p = 5$, $\Delta x = 2B$), and ($n_p = 2$, $\Delta x = 5B$). In the SRW-I-MP with $n_p = 10$ and $\Delta x = B$, each beam exposes all the pixels in a row. Hence, each pixel is exposed by all of the 10 beams in a set. Consequently, the dose reduction due to a faulty beam is d. For $n_p = 5$ and $\Delta x = 2B$, each pixel is exposed by 5 different beams in a set and the dose reduction due to a faulty beam is $2d$. Similarly, since each pixel is exposed by 2 different beams when $n_p = 2$ and $\Delta x = 5B$, the dose reduction due to a faulty beam is $5d$. Therefore, the smallest dose reduction occurs when $n_p = 10$ and $\Delta x = B$. Hence, the SRW-I-MP with $n_p = n_s$ and $\Delta x = B$ gives the best result in terms of the performance metrics.

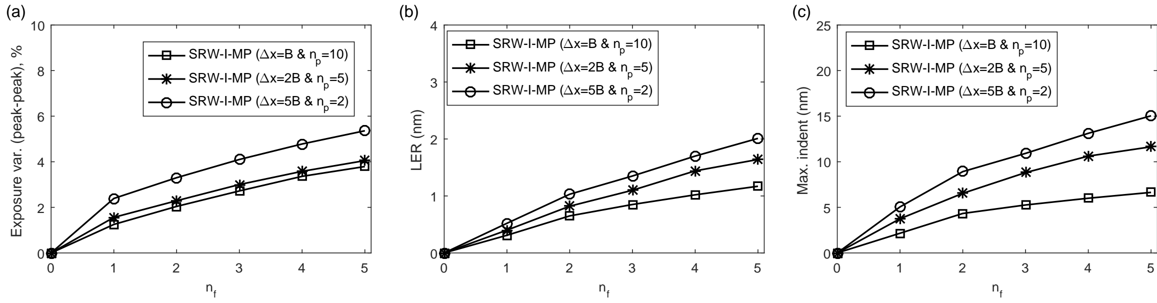

Figure 4.4. (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm.

In Fig. 4.5, the level of spatial fluctuations of beam current is varied for three different implementations of the SRW-I-MP. The performance metrics are improved most when $n_p = 10$ and $\Delta x = B$. This is because, in this case, each pixel is exposed by a larger number of beams than the other implementations and, therefore, different doses from multiple beams have a better chance to be averaged out. For the same reason, the SRW-I-MP with $n_p = 10$ and $\Delta x = B$ gives the best results in the case of the beam-positioning error, as shown in Fig. 4.6.

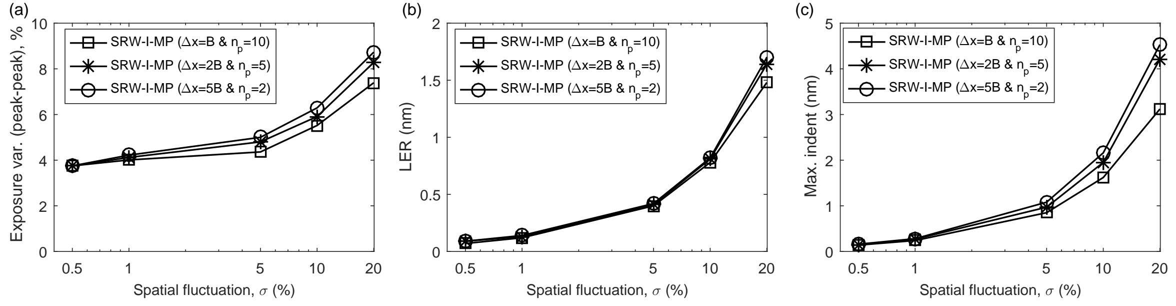

Figure 4.5. (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2\text{nm}$, $n_f = 0$.

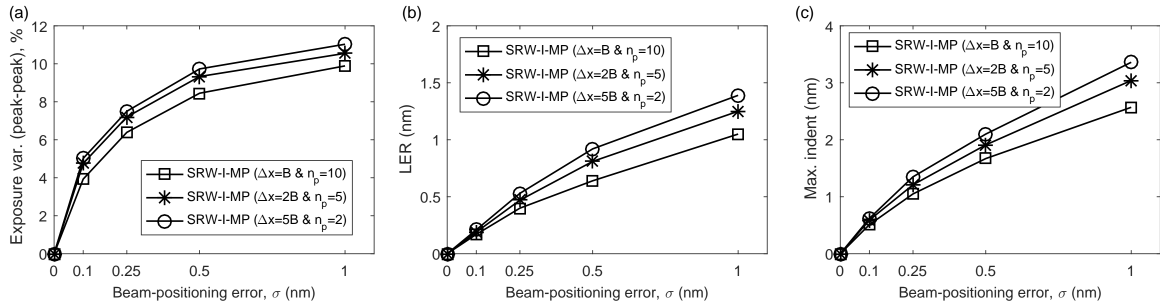

Figure 4.6. (a) Exposure variation, (b) LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2\text{nm}$, $n_f = 0$.

The three performance metrics for the varying level of temporal beam-current fluctuation are plotted in Fig. 4.7. For a fixed σ of the temporal fluctuation, the difference between the initial and final beam currents (of a beam), ΔI, is the same for all beams. In the MP writing, the higher n_p, the larger the total number of steps. Since ΔI is the same independent of n_p, the current increment per step is inversely proportional to n_p. As the variation of beam current with time decreases with increasing n_p, the difference between the minimum and maximum doses in pixels along the boundary of a feature decreases with increasing n_p. Therefore, the SRW-I-MP with the highest n_p ($n_p = 10$) has the smallest exposure variation, LER and maximum indent.

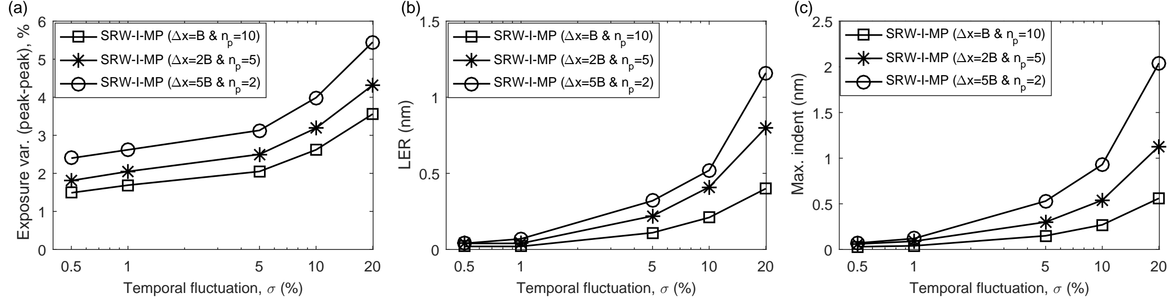

Figure 4.7. (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

4.3.2 Multi-row writing (MRW)

The optimal pattern of the MRW ($\Delta H = \frac{n_s}{2} = 5$ and $w_p = \frac{n_s}{2} + 1 = 6$) is compared with another pattern ($\Delta H = 1$ and $w_p = n_s = 10$) in the cases of always-off faulty beams, spatial and temporal fluctuations of beam current, and beam-positioning error. From the results shown in Figs. 4.8-4.11, it is observed that the optimal pattern improves the performance metrics significantly. The results for the MRW with the pattern $\Delta H = \frac{n_s}{2}$ and $w_p = \frac{n_s}{2} + 1$ combined with the optimal MP writing ($\Delta x = B$ and $n_p = n_s$) are also presented. In all cases, the MRW-MP improves the performance metrics most.

The dose reduction due to an always-off faulty beam is d in any realizable MRW pattern. When $(\Delta H, \Delta V) = (1, 1)$, adjacent pixels are affected. But, when $(\Delta H, \Delta V) = (5, 1)$, an affected pixel is surrounded by non-affected pixels which can compensate for the dose reduction in the affected pixel. Therefore, the exposure variation for the pattern $(\Delta H, \Delta V) = (5, 1)$ is smaller than that for the pattern $(\Delta H, \Delta V) = (1, 1)$, resulting in smaller LER and maximum indent as can be seen in Fig. 4.8. In the optimal MRW-MP, each beam exposes all the pixels in its writing path of 10 rows in 10 passes. Hence, each pixel is exposed by 10 times as many beams as in the MRW. Consequently, the amount of dose reduction due to a faulty beam is reduced 10 times. Therefore, the MRW-MP improves the exposure variation, LER, and maximum indent significantly compared to the MRW.

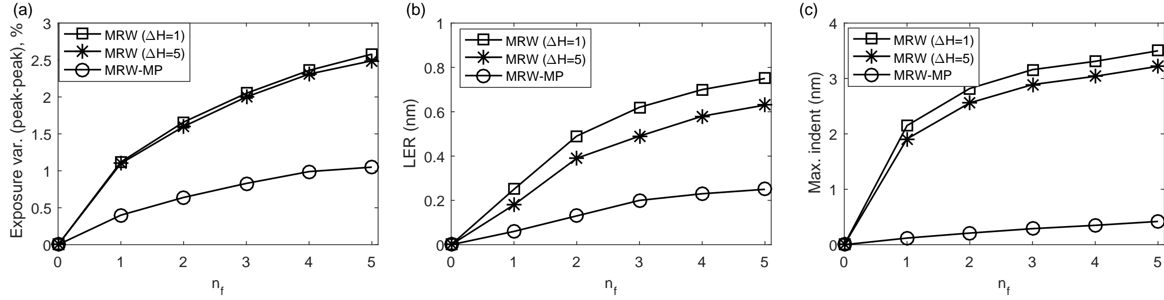

Figure 4.8. (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. $\sigma_t = 2$nm.

Fig. 4.9 shows that the optimal MRW pattern performs better than the pattern with $\Delta H = 1$ in the case of spatial fluctuation of beam current. In the optimal MRW pattern with $(\Delta H, \Delta V) = (5, 1)$, the distance between two pixels exposed by the same beam is maximized. The pixels between the two affected pixels are exposed by other beams in the same writing path. On the other hand, when $(\Delta H, \Delta V) = (1, 1)$ in a MRW pattern, adjacent pixels are affected by the same beam, resulting in a lower number of surrounding pixels exposed by other beams. Hence, the averaging effect of spatial fluctuation of beam current is larger in the optimal MRW pattern. In the optimal MRW-MP, each pixel is exposed by all the beams in each writing path. Therefore, the averaging effect is maximized, resulting in the smallest exposure variation, LER, and maximum indent. For the same reasons, similar trends are observed in the case of beam-positioning error as shown in Fig. 4.10.

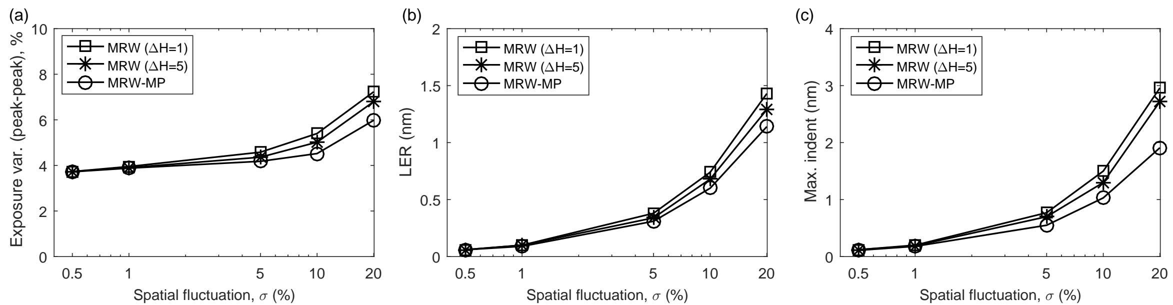

Figure 4.9. (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

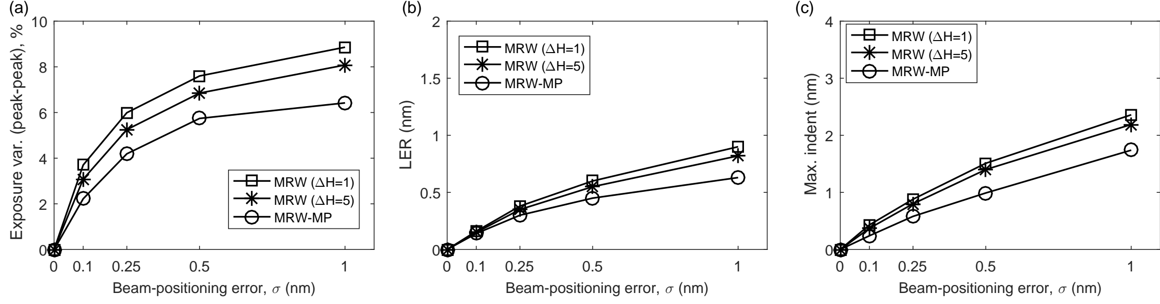

Figure 4.10. (a) Exposure variation, (b) LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$.

The three performance metrics obtained with the temporal fluctuation of beam current varied are provided in Fig. 4.11. In the MRW, the number of steps required to expose all the pixels in a writing path as well as the current increment per step are the same for both $\Delta H = 1$ and $\Delta H = 5$. As a result, the difference between the minimum and maximum doses in pixels along the boundary of a feature is almost the same in both cases. Therefore, the improvement of the performance metrics by the optimal pattern ($\Delta H = 5$) over the non-optimal pattern ($\Delta H = 1$) is not substantial in the case of temporal fluctuation.

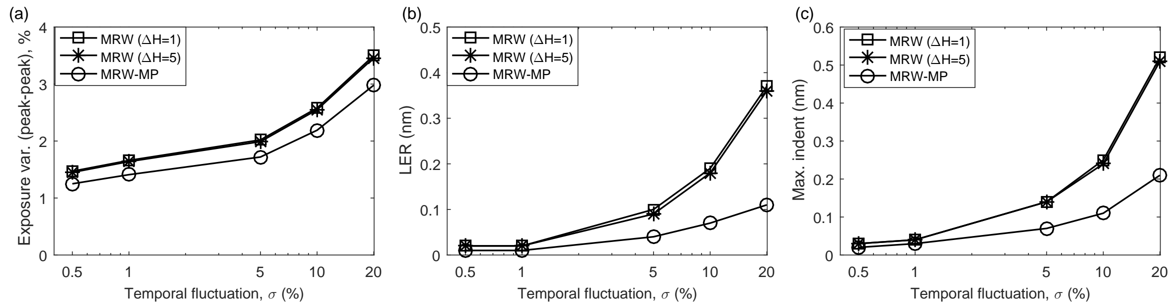

Figure 4.11. (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

4.3.3 Comparison of the MP writing and MRW

In this comparison study, the optimal MRW pattern with $\Delta H = \frac{n_s}{2}$ and $w_p = \frac{n_s}{2} + 1$ is used. Also, in the SRW-I-MP, SRW-II-MP and MRW-MP, the optimal MP writing with $n_p = n_s$ and $\Delta x = B$ is applied.

The performance metrics of the SRW-I, SRW-I-MP, SRW-II, SRW-II-MP, MRW, and MRW-MP are compared in the cases of always-off faulty beams (Fig. 4.12), spatial beam

current fluctuations (Fig. 4.13), temporal beam current fluctuations (Fig. 4.14), and beam-positioning error (Fig. 4.15). In all cases, the MRW performs better than the SRW-I-MP and SRW-II-MP. The MRW-MP performs even better than the MRW since each pixel in n_p rows is exposed by all the beams in the writing path. Also, the performance of the SRW-I-MP is identical with that of the SRW-II-MP. This is because, in both cases, each of the pixels in a writing path is exposed by all the beams in that writing path.

The dose reduction due to a faulty beam in the SRW-I-MP, SRW-II-MP, and MRW is d. However, the affected pixels are confined to only one row in the SRW-I-MP and SRW-II-MP while they are spread over multiple rows in the MRW. Therefore, even though the dose reduction in an affected pixel is the same, the MRW performs better than the SRW-I-MP and SRW-II-MP in terms of the performance metrics.

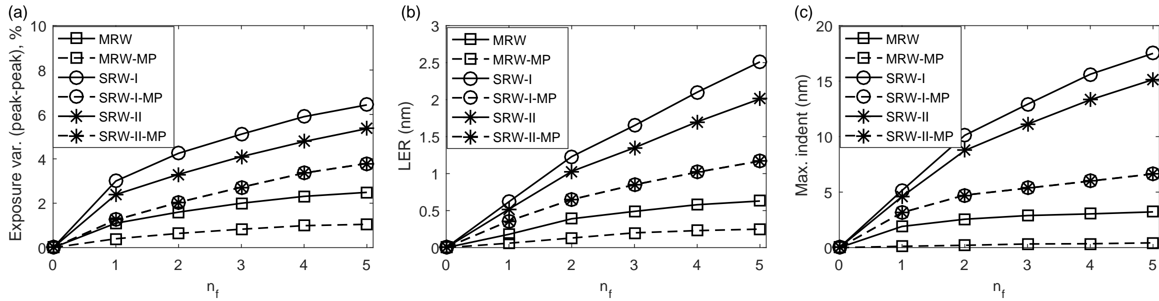

Figure 4.12. (a) Exposure variation, (b) LER, and (c) maximum indent with the number of always-off faulty beams (n_f) varied. σ_t = 2nm.

In the case of spatial beam-current fluctuation, the higher the number of beams exposing a pixel, the higher the averaging effect of spatial fluctuation of beam current. The number of beams exposing a pixel is the same, i.e., n_s, for all of the SRW-I-MP, SRW-II-MP and MRW, but there is a slight improvement of performance metrics by the MRW compared to the other two (Fig. 4.13). In the SRW-I-MP or SRW-II-MP, each beam exposes pixels in a single row, while in the MRW, each beam exposes pixels over multiple rows spreading its effect. Therefore, the overall exposure distribution is smoother in the MRW resulting in smaller LER and maximum indent. In the MRW-MP, as each pixel is exposed by each of the n_s^2 beams in a writing path, the averaging effect of spatial fluctuation of beam currents is the maximum among all the writing methods. As a result, the performance metrics are improved most by the MRW-MP.

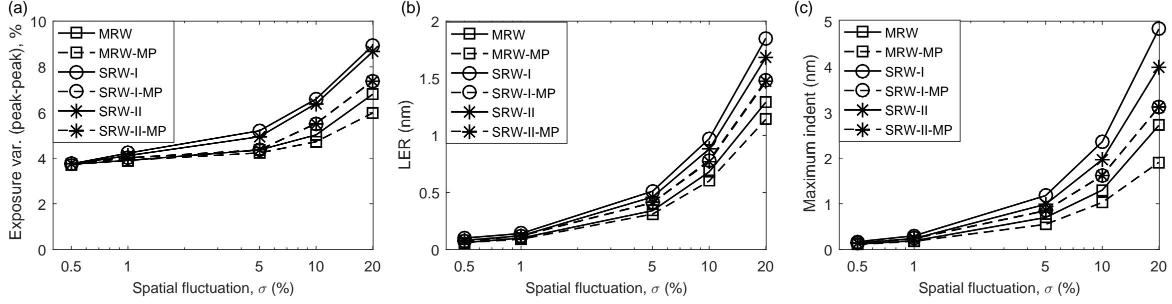

Figure 4.13. (a) Exposure variation, (b) LER, and (c) maximum indent with the spatial fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

From Fig. 4.14, it is seen that the performance metrics are similar for the SRW-I and SRW-II when temporal fluctuation of beam current is varied. This is because the number of steps required to expose all the pixels in a writing path of a single row is the same for the SRW-I and SRW-II. In the SRW-I-MP, SRW-II-MP, and MRW, the required number of steps is n_s times ($n_s = n_p$) that of the SRW-I or SRW-II. The current increment per step decreases with the increase in the number of steps, which results in a smaller variation of beam current and improves the performance metrics. On the other hand, in the MRW-MP, the number of steps required to expose all the pixels in a writing path of n_s rows is n_p times ($n_p = n_s$) that of the MRW. Therefore, the improvement in the performance metrics by the MRW-MP is largest.

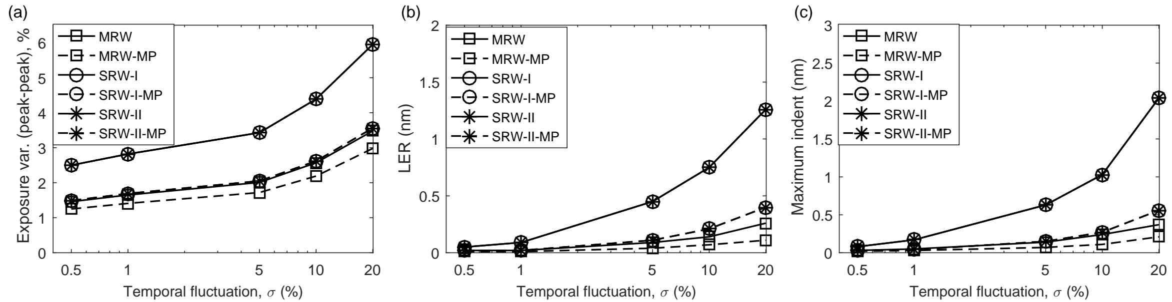

Figure 4.14. (a) Exposure variation, (b) LER, and (c) maximum indent with the temporal fluctuation (σ) of beam current varied. $\sigma_t = 2$nm, $n_f = 0$.

In the case of the beam-positioning error, the higher the number of beams exposing a pixel, the higher the probability of averaging out the effect of the beam-positioning error. The numbers of beams exposing a pixel in the cases of SRW-I, SRW-II, SRW-I-MP, SRW-II-MP, MRW, and MRW-MP are 1, n_g, n_s, n_s, n_s, and n_s^2, respectively. As a result, the MRW-MP achieves the best result in terms of the performance metrics (Fig. 4.15). Although the number of

beams exposing a pixel is the same for the SRW-I-MP, SRW-II-MP, and MRW, the performance of the MRW is slightly better (Fig. 4.15). Each pixel is exposed by n_s consecutive beams in the SRW-I-MP and SRW-II-MP while by one in every n_s beams in the MRW. Since the beam-positioning error is spatially correlated, the beams located close to each other have similar error vectors. Therefore, the averaging effect of beam-positioning error is smaller in the SRW-I-MP and SRW-II-MP than in the MRW.

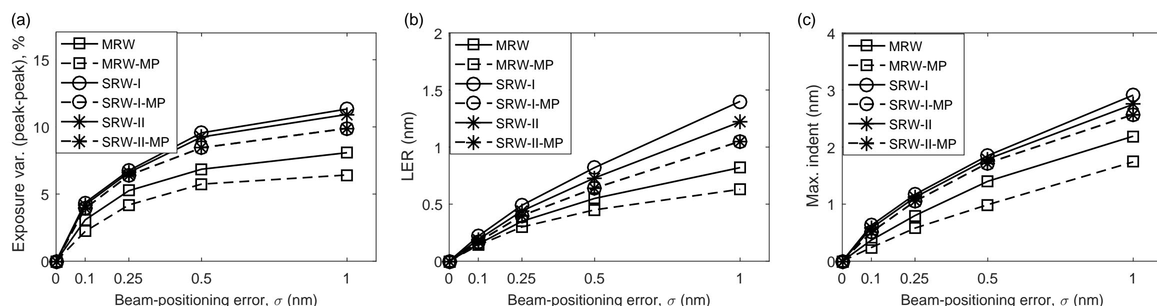

Figure 4.15. (a) Exposure variation, (b) LER, and (c) maximum indent with the beam-positioning error (σ) varied. $\sigma_t = 2$nm, $n_f = 0$.

4.4 Summary

This chapter presents the derivation of the optimal conditions for achieving the best MP writing and MRW in a MPES with a given set of lithographic parameters, which includes determining the optimal number of passes and positional shift for MP writing, and the optimal writing pattern for MRW. The study also analyzes various combinations of SRW and MRW, and their performance in reducing the negative effects of beam abnormalities. The results demonstrate that the MRW method (without MP) performs better than SRW-MP methods, and MRW-MP is the most effective in reducing the impact of beam abnormalities. The results from this chapter may be referred to in designing an effective writing method for current and future MPES's.

Chapter 5

Shape and Dose Control for Proximity Effect Correction

In this chapter, a systematic method for realizing various types of spatial dose distributions, e.g., uniform, V-type, A-type, and M-type, with any linewidth reduction of feature is presented.[23] Also, the shape and dose control procedure for proximity effect correction (PEC) with spatial dose distributions realizable on an MPES is described.[23]

5.1 Realization of spatial dose distribution

To have a sufficient spatial control of the dose distribution, the line feature of width W is partitioned into five regions along its length dimension and a dose d_i is determined for each region i where $i = 1, 2, 3, 4, 5$ (Fig. 5.1). In a previous study,[19] four different types of spatial dose distributions are introduced, i.e., uniform, V-type, A-type, and M-type. It is shown in another study[20] that reducing the feature width to be exposed results in a smaller CD error and LER. But, due to a relatively large beam size B and a fixed exposing interval I_{ex} in the current MPES, the optimal line-width reduction ΔW and spatial distribution of dose required for the PEC may not be achievable. Therefore, a practical method for realizing a given type of dose distribution with any ΔW is needed.

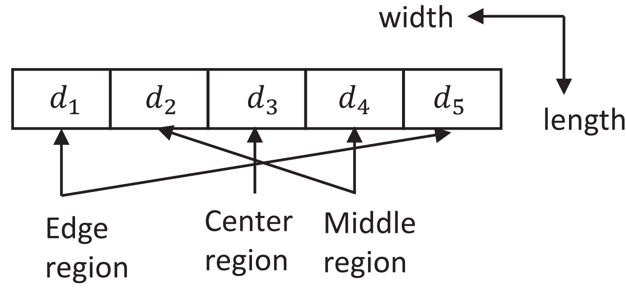

Figure 5.1. A feature is partitioned into 5 regions such that the dose can be spatially controlled region-wise. The spatial dose distribution $\{d_i\}$ is symmetric with respect to the center region, i.e., $d_1 = d_5$ and $d_2 = d_4$.

From Fig. 5.2, it can be seen that, when $\Delta W = nB$ where n is a non-negative integer, any type of dose distribution can be realized in one pass as the width of feature to be exposed is an integer multiple of B. But, when $\Delta W \neq nB$, the number of passes, n_p, is greater than one to realize the dose distribution as close to the target dose distribution as possible. Also, the point of exposure on the substrate for each beam needs to be shifted by the amount of $\Delta x = \frac{\Delta W}{2}$ in the direction of substrate as illustrated in Fig. 5.2(c). In such cases, there can be some regions where the given dose is higher than the target dose due to the overlap of exposed regions (Fig. 5.2(c)). Hence, the dose to be given to each region in each pass needs to be determined such that the overlap and dose deviation are minimized.

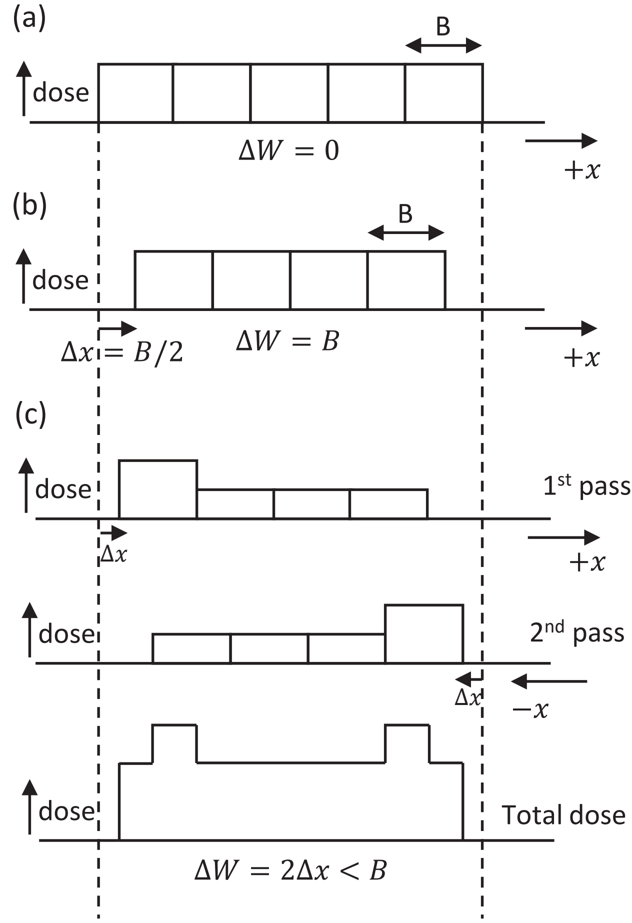

Figure 5.2. The total dose is given in one pass when (a) $\Delta W = 0$, and (b) $\Delta W = B$. More than one pass is used to realize a dose distribution when (c) $0 < \Delta W < B$.

Another important issue to consider is the utilization of beams in an MPES. Utilization U is defined as the average fraction of time the beams are on with respect to the total exposing time. Suppose that five beams are exposing a feature of five pixels. Each beam is deflected five times exposing a pixel. If all the beams are on in all 25 shots (5 shots per beam), $U = 100\%$. The utilization may also be computed as the ratio of the average dose per pixel to the maximum dose given to a pixel.

The realization of uniform, V-type, A-type, and M-type dose distributions for $0 < \Delta W < B$ is described in the following. A dose distribution with the line-width reduction of $\Delta W + nB$ can be realized in the same way as for the line-width reduction of ΔW. It is assumed that the feature is divided into five regions where each region consists of a pixel of size B and five beams are used to expose the feature. The beam size of $B = 10$nm and the beam interval of $I_{bx} = 4B$ are considered. In a *step*, a beam gives a unit dose of d to a pixel. During the

exposing process, each beam follows a pixel being deflected five times, to give a total dose of $D = 5d$ to the pixel through a cycle of 5 steps.

5.1.1 Uniform dose distribution

In Fig. 5.3, the realization of uniform dose distribution is illustrated with different line-width reductions, ΔW. When $\Delta W = nB$, a uniform dose distribution can be achieved in one pass (Fig. 5.3(a)). On the other hand, when $\Delta W \neq nB$, a (completely) uniform dose distribution cannot be achieved though the deviation from the uniform distribution may be minimized through multiple passes. In Figs. 5.3(b)-(d), the dose to be given to each region in each pass is shown to realize a uniform dose distribution for $0 < \Delta W < B$. The doses in the gray areas in Fig. 5.3 are higher than the target (desired) dose due to the overlap of exposed regions over multiple passes. To avoid too high a dose in the gray areas, the final dose in the edge regions may be made smaller than that in the center and middle regions. In Fig. 5.4, the ideal and realized dose distributions are illustrated. The utilization U with $\Delta W \neq nB$ is smaller than that with $\Delta W = nB$ as shown in Fig. 5.3. This is because some beams remain off when a region gets no dose or a smaller dose than other regions in different passes.

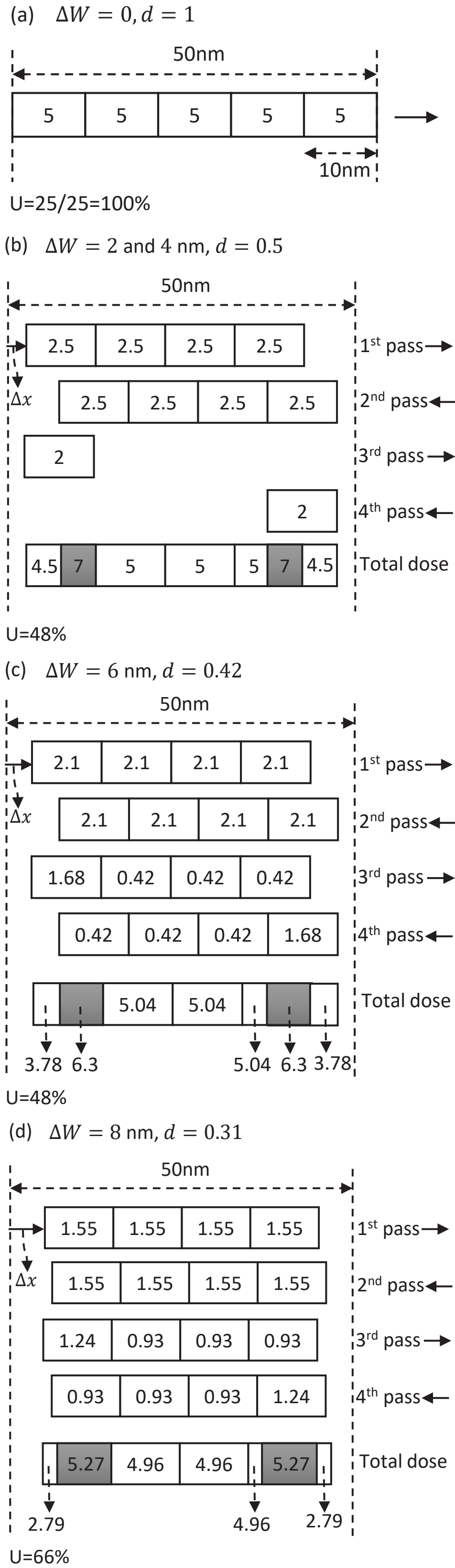

Figure 5.3. The realization of uniform dose distribution in multiple passes for (a) $\Delta W = 0$nm and $n_p = 1$, (b) $\Delta W = 2$ and 4nm, and $n_p = 4$, (c) $\Delta W = 6$nm and $n_p = 4$, and (d) $\Delta W = 8$nm and $n_p = 4$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes.

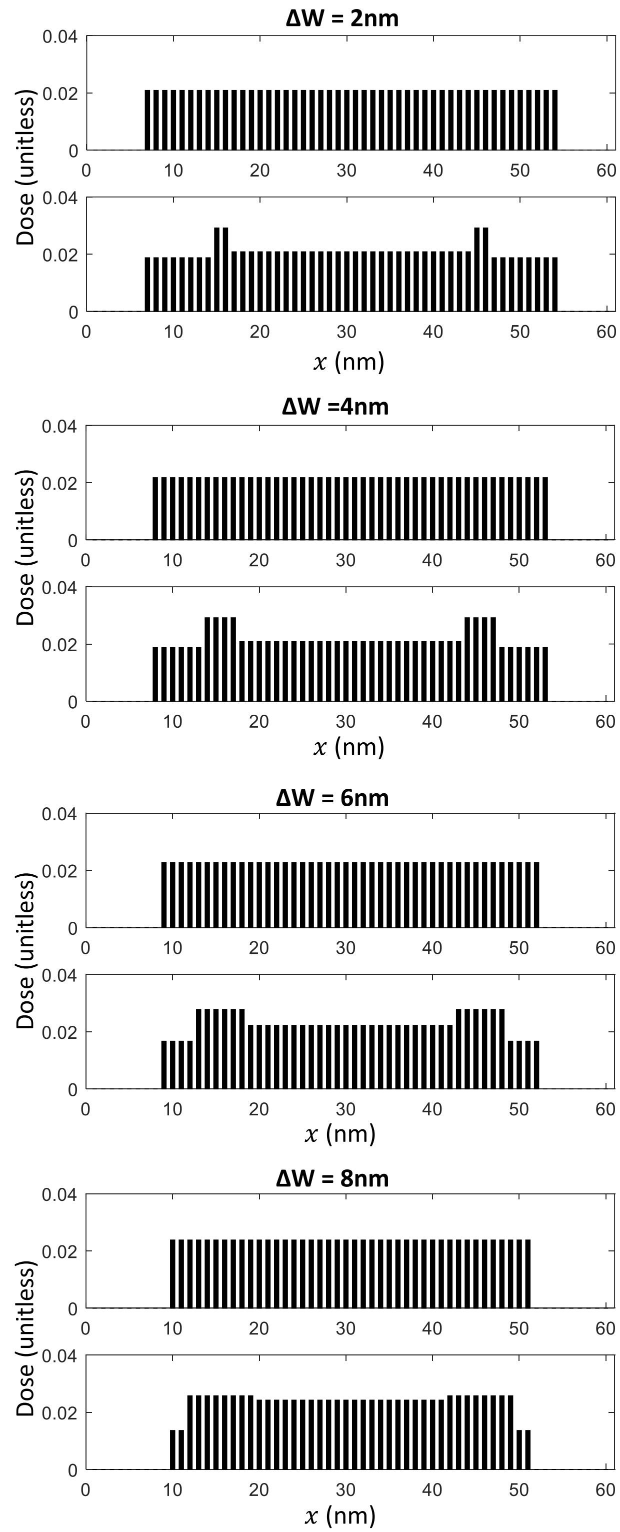

Figure 5.4. The ideal (top) and realized (bottom) uniform dose distributions for each of $\Delta W = 2, 4, 6$, and 8nm.

5.1.2 V-type dose distribution

The realization of V-type dose distribution with varying ΔW is illustrated in Fig. 5.5. The number of passes may vary depending on the spatial-dose-distribution ratio and d. In general, more passes are needed with $\Delta W \neq nB$ than with $\Delta W = nB$. The overlapped (gray) areas between the middle and center regions when $\Delta W \neq nB$ cause the realized V-type dose distribution to deviate from the ideal V-shape as shown in Fig. 5.6. The utilization U with $\Delta W \neq nB$ is smaller than that with $\Delta W = nB$, as more beams are turned off in more passes when $\Delta W \neq nB$. Also, U for the V-type dose distribution tends to be lower than that for the uniform dose distribution (see Fig. 5.5).

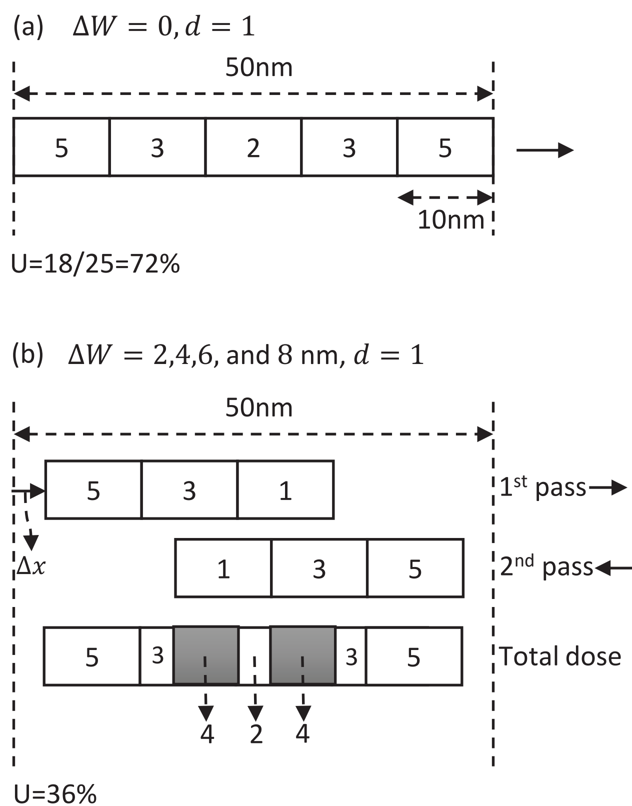

Figure 5.5. The realization of V-type dose distribution in multiple passes with a spatial-dose-distribution ratio of 5:3:2:3:5 for (a) $\Delta W = 0$nm and $n_p = 1$, and (b) $\Delta W = 2, 4, 6$ and 8nm, and $n_p = 2$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes.

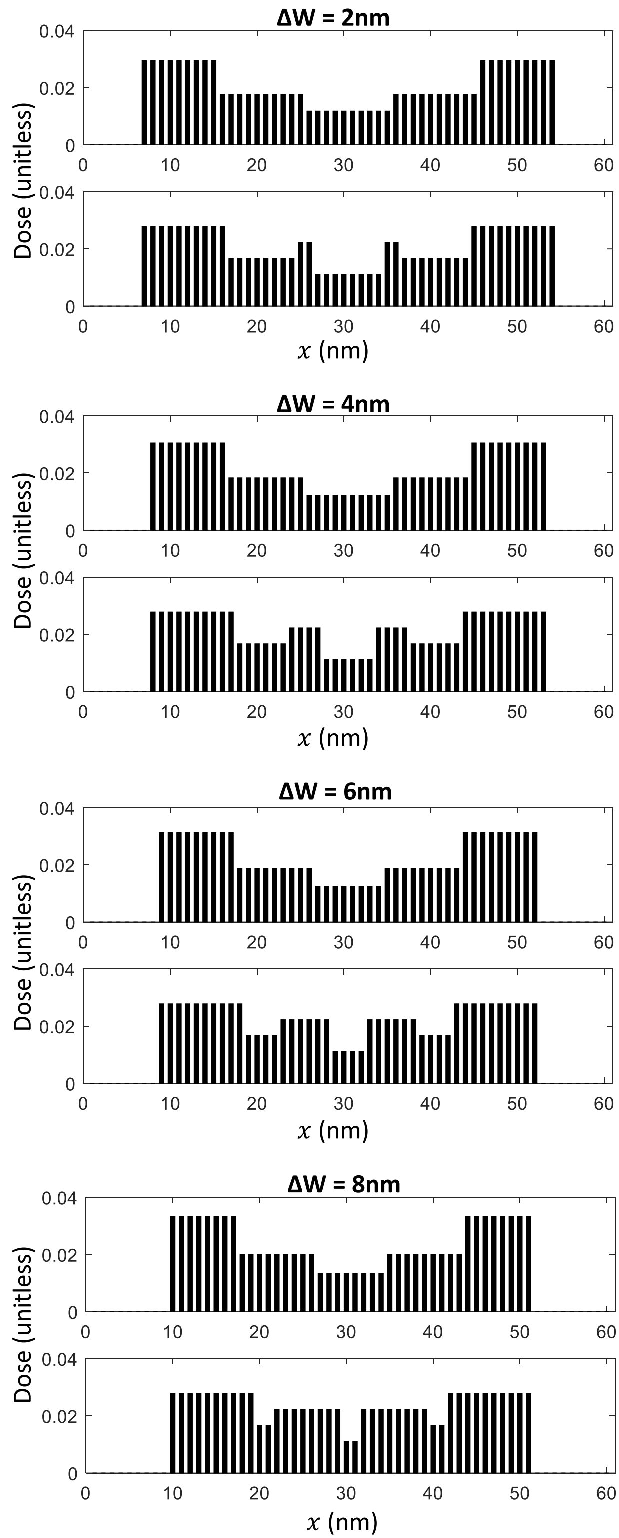

Figure 5.6. The ideal (top) and realized (bottom) V-type dose distributions for each of $\Delta W = 2, 4, 6$, and 8nm.

5.1.3 A-type dose distribution

The realization of A-type dose distribution with varying ΔW is illustrated in Fig. 5.7. Because of the overlapped (gray) areas between the middle and center regions, the realized A-type dose distribution is smoother than the ideal A-shape as shown in Fig. 5.8. Also, the utilization U for the A-type may change depending on the spatial-dose-distribution ratio which also determines the number of passes required. U tends to be lower when the dose in the center region is higher. This is because the number of steps increases to provide the high dose in the center region, but the beams remain off most of the time while exposing other regions.

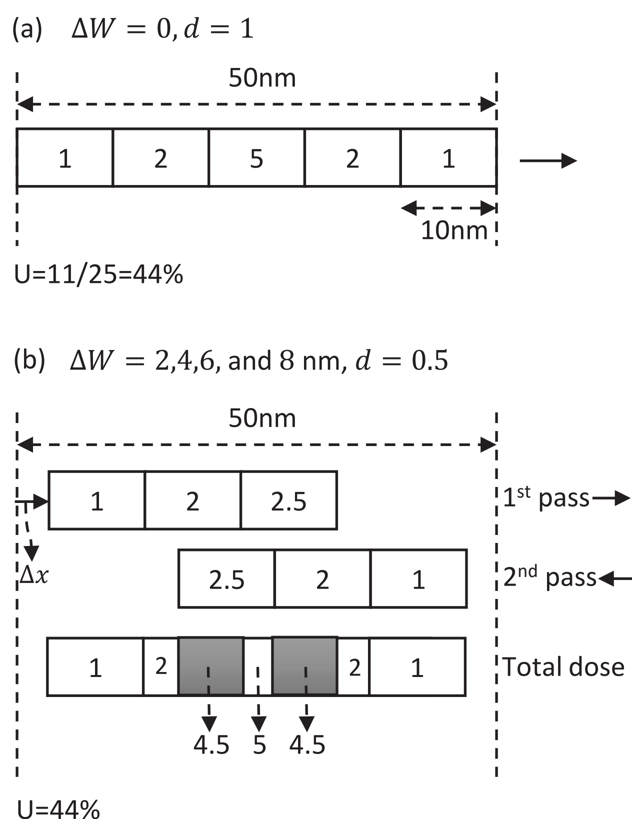

Figure 5.7. The realization of A-type dose distribution in multiple passes with a spatial-dose-distribution ratio of 1:2:5:2:1 for (a)$\Delta W = 0$nm and $n_p = 1$, and (b)$\Delta W = 2, 4, 6$ and 8nm, and $n_p = 2$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes.

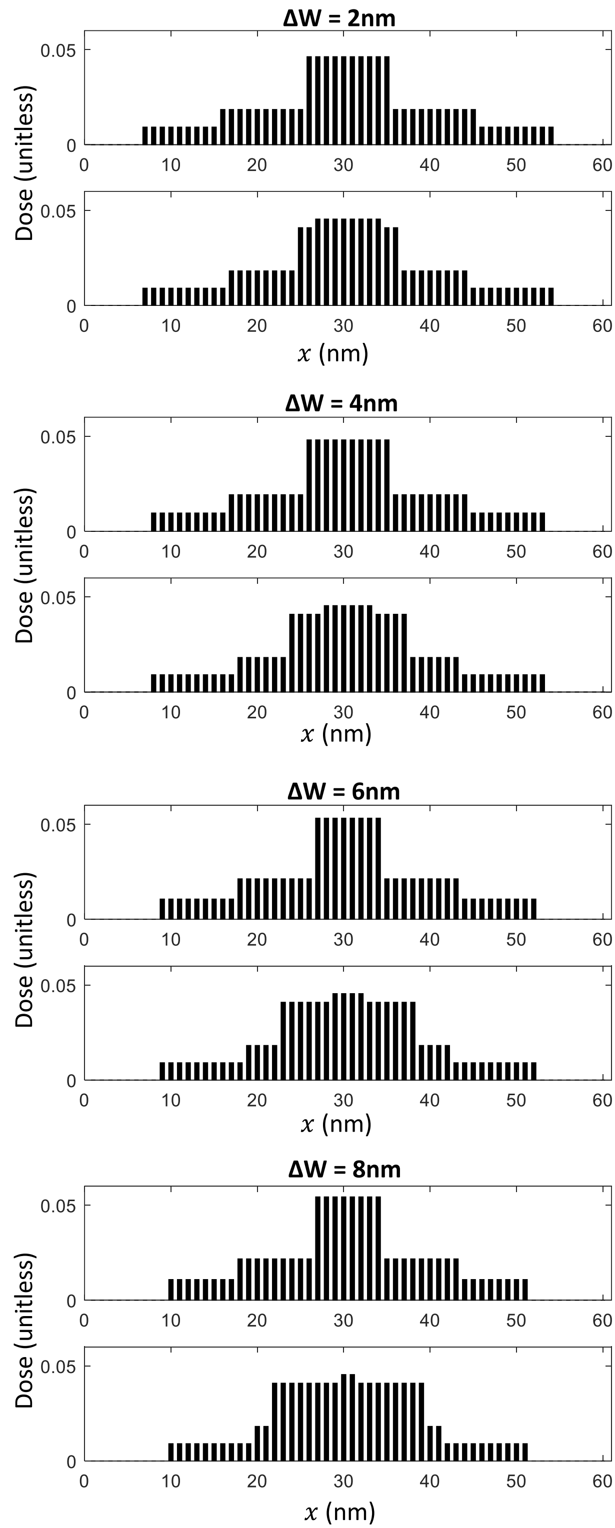

Figure 5.8. The ideal (top) and realized (bottom) A-type dose distributions for each of $\Delta W = 2, 4, 6,$ and 8nm.

5.1.4 M-type dose distribution

The realization of M-type dose distribution with varying ΔW is demonstrated in Fig. 5.9. There are overlapped (gray) areas between the edge, middle and center regions (Fig. 5.9), which results in a blurred transition between the regions as can be seen in Fig. 5.10, where the ideal and realized dose distributions are compared with varying ΔW. In general, the utilization U with $\Delta W \neq nB$ is smaller than that with $\Delta W = nB$, since more passes are required when $\Delta W \neq nB$ where more beams stay off.

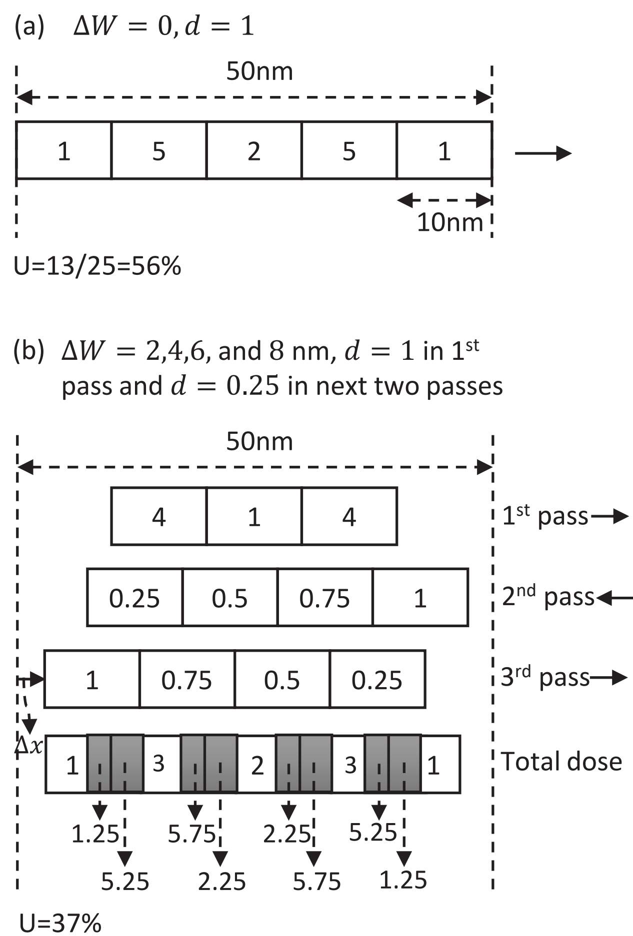

Figure 5.9. The realization of M-type dose distribution in multiple passes with a spatial-dose-distribution ratio of 1:5:2:5:1 for (a) $\Delta W = 0$nm and $n_p = 1$, and (b) $\Delta W = 2, 4, 6$ and 8nm, and $n_p = 3$. The gray areas represent higher dose than the target dose due to the overlap of exposed regions in multiple passes.

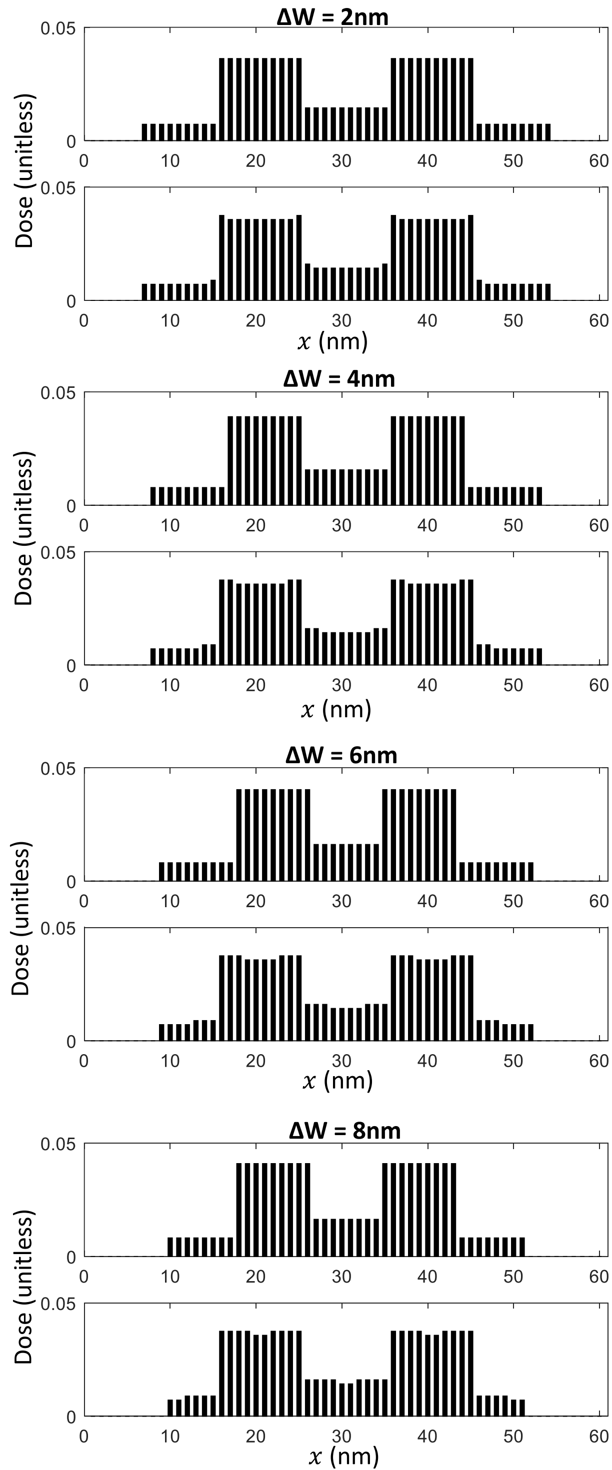

Figure 5.10. The ideal (top) and realized (bottom) M-type dose distributions for each of $\Delta W = 2, 4, 6,$ and 8nm.

5.2 Minimization of CD error and LER

In this section, the realization of the shape+dose correction method is described under the constraints of the MPES to minimize the CD error and LER.

5.2.1 Cost function

In a previous work,[20] it is observed that the CD error and LER stay low when the edge of the developed feature is outside the exposed area. This is because the exposure level quickly drops from the exposed area (inside the feature) to the unexposed area (outside the feature) and the absolute variation of exposure is smaller in the unexposed area. It is possible that the optimal ΔW for the minimal CD error is not the same as that for the minimal LER. Also, the required total dose may increase significantly while reducing the line-width to be exposed. Therefore, the cost function, C, consisting of the CD error, LER, and total dose is considered

$$C = a_1 CD_error + a_2 LER + a_3 total_dose \tag{5.1}$$

where a_1, a_2, and a_3 are the weights given to CD_error, LER, and $total_dose$, 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, measured at each layer of resist. As a way to reflect the CD error and LER at all of the top, middle and bottom layers of resist, they are averaged over the three layers. The $total_dose$ is the integration of area dose over the feature width.

5.2.2 Shape+dose correction

The optimal line-width reduction and the dose to be given to each region for different types of dose distributions are determined through an iterative procedure such that the cost function is minimized. The correction procedure does not allow the dose of a region to exceed a certain value to avoid too low a utilization. Also, it exploits the fact that the cost function shows a

bitonic behavior with respect to ΔW. The following notations are used in the description of the procedure.

C_{min}:	The minimum value of cost function
$C_{min,\Delta W}$:	The minimum value of cost function for a ΔW
C_0:	The initial value of cost function
$d_{max_allowed}$:	The maximum dose allowed in a region
Δd_i:	The amount of adjustment for d_i. $\Delta d_i = d_i \times x$, where x is a fractional number.
ΔW_{opt}:	The optimal line-width reduction
$\{d_i\}_{opt}$:	The optimal spatial dose distribution
$\{d_i\}$:	Initialized to {3,2,1,2,3} for V-type, {1,2,3,2,1} for A-type, and {1,2,1,2,1} for M-type.

The iterative correction procedure for a given type of dose distribution is described below where k is the iteration index, and its flowchart is also provided in Fig. 5.11. It consists of two nested loops where different amounts of width reduction are considered through the outer loop while the spatial dose distribution is optimized for a given feature width in the inner loop. It is assumed that the minimum possible reduction of width is 2nm, i.e., 1nm on each side of feature.

1. $\Delta W \leftarrow 0$, $C_{min} \leftarrow$ a large value, $C_{min,\Delta W} \leftarrow$ a large value, and $C_0 \leftarrow$ a large value.
2. Initialize $\{d_i\}$ according to the type of dose distribution and x.
3. Obtain the realizable dose distribution given ΔW and $\{d_i | i = 1, 2, 3, 4, 5\}$. Refer to Section 5.1.
4. Evaluate the cost function, C_k.
5. If $C_k < C_{min}$,

   $C_{min} \leftarrow C_k$, $\Delta W_{opt} \leftarrow \Delta W$ 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 y is a fractional number, i.e., the improvement in C is too small,

   $x \leftarrow s_1 x$ ($s_1 > 1$).
8. If $C_k > C_{k-1} + C_{k-1} \times y$,

   $x \leftarrow s_2 x$ ($0 < s_2 < 1$).

9. $d_i \leftarrow d_i \pm \Delta d_i$ such that $0 < d_i \pm \Delta d_i < d_{max_allowed}$ and the shape of the dose distribution type is maintained, and go to step 3.

10. If $C_{min} < C_{min,\Delta W}$,

$$C_{min,\Delta W} \leftarrow C_{min},\ \Delta W \leftarrow \Delta W + 2\text{nm, and go to step 2.}$$

11. Output ΔW_{opt} and $\{d_i\}_{opt}$.

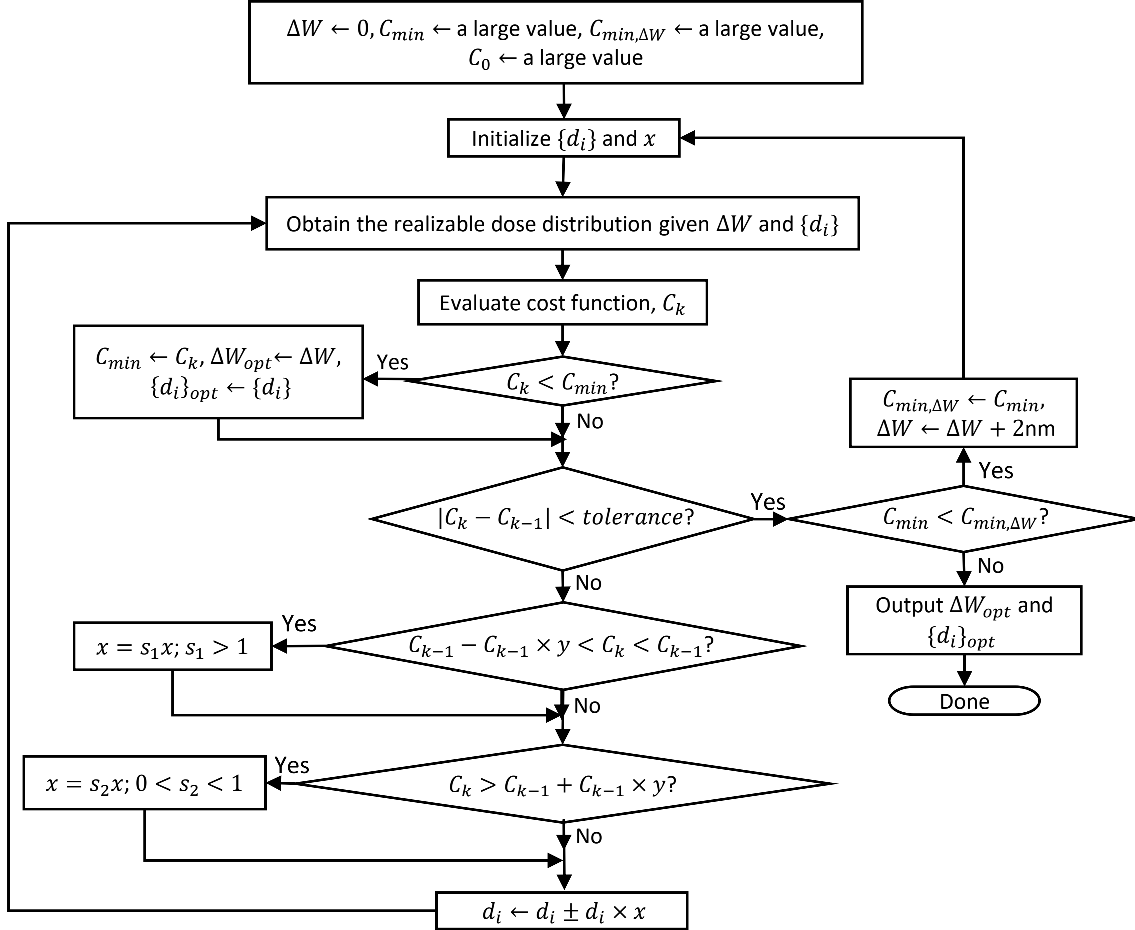

Figure 5.11. The flowchart of the shape + dose correction method where k is the iteration index, and x and y are fractional numbers.

In step 9, the dose d_i is adjusted maintaining the relationship among region doses for the given type of dose distribution, i.e., $d_1 > d_2 > d_3$ for V-type, $d_1 < d_2 < d_3$ for A-type, and $d_1 \leq d_3 < d_2$ for M-type. With any type of dose distribution, if $d_i \pm \Delta d_i$ is greater than $d_{max_allowed}$, or does not maintain the shape of the dose distribution type, the adjustment is attempted in the adjacent region or Δd_i is adjusted by reducing the fraction, x, of d_i to maintain the relationships. Figs. 5.12-5.14 illustrate for different types of dose distribution

how a region is selected and the dose adjustment in the region is determined. With the V-type dose distribution, there can be an inner CD error shown in Fig.5.12(a) due to a low dose in the center regions. In this case, the dose in the center region is increased to remove the inner CD error. Also, with the V-type, the top layer of resist develops earlier than the bottom layer within the edge regions resulting in an overcut sidewall as shown in Fig.5.12(b). This problem can be alleviated by lowering the dose in the edge regions which reduces the top-layer CD error more than it reduces the bottom-layer CD error. In this way, the resist profiles are more balanced in the CD error among layers.

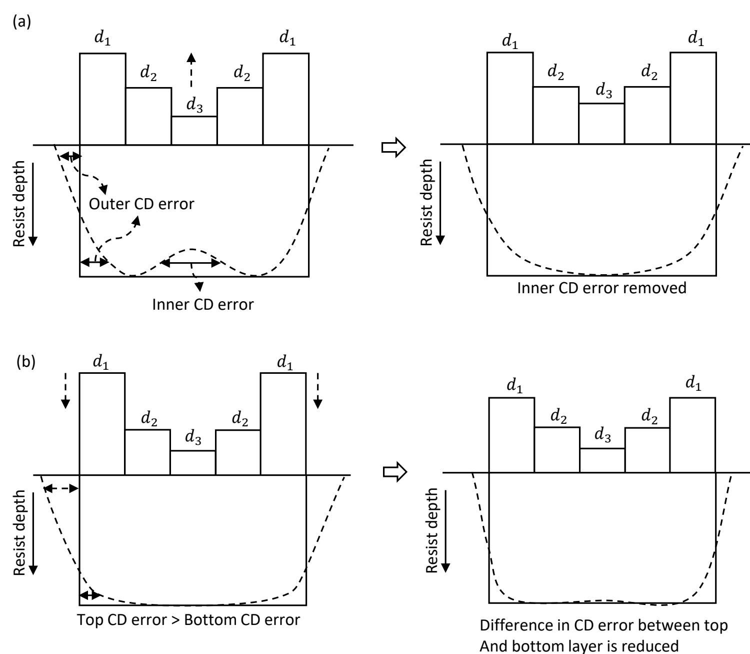

Figure 5.12. The dose adjustment in the case of V-type dose distribution guided by the cross-section of the resist profile.

The CD in the top layer can be smaller than the target CD in the case of A-type dose distribution (Fig. 5.13(a)) due to a low dose in the edge regions. In such a case, the dose in the edge regions is increased to match the top-layer CD to the target CD. Also, if the CD error in the bottom layer is larger than that in the top layer with the A-type (Fig. 5.13(b)), the dose in the center region is increased so that the resist development in the bottom layer can catch up with that in the top layer.

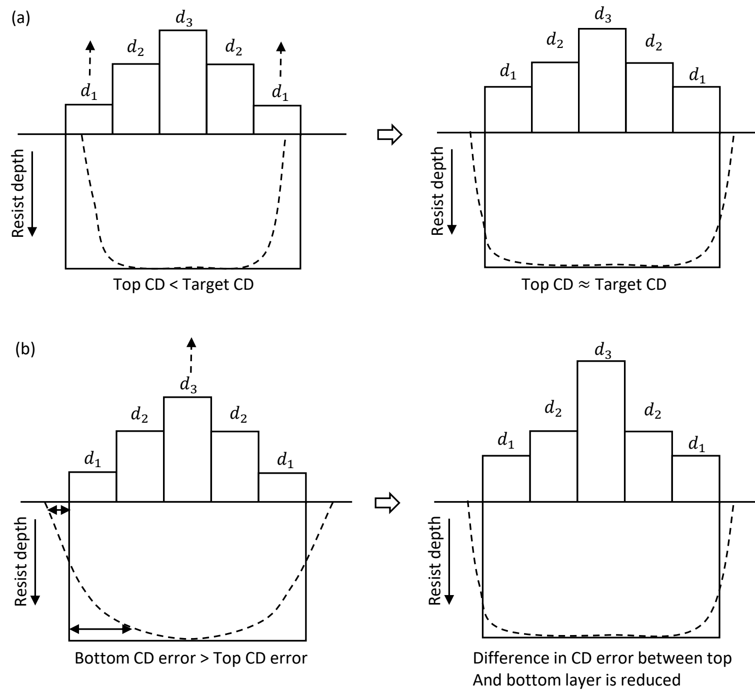

Figure 5.13. The dose adjustment in the case of A-type dose distribution guided by the cross-section of the resist profile.

An inner CD error may occur also in the case of M-type dose distribution as illustrated in 5.14(a), which is eliminated by increasing the dose in the center region. Also, if the CD error in the bottom layer is larger than that in the top layer with the M-type (Fig. 5.14(b)), the dose in the center region is increased to match the resist development in the bottom layer to that in the top layer.

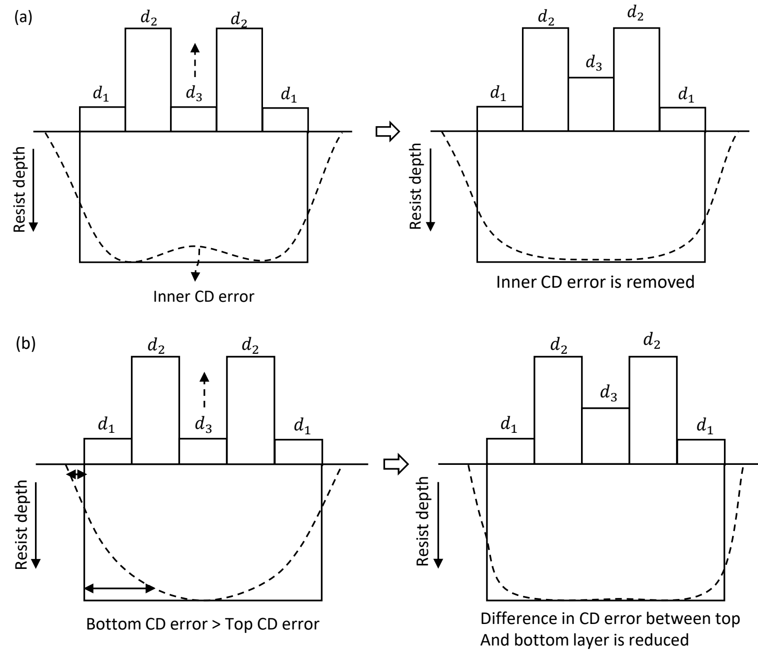

Figure 5.14. The dose adjustment in the case of M-type dose distribution guided by the cross-section of the resist profile.

5.3 Results and discussion

The effectiveness of the shape+dose correction algorithm under the constraints of the MPES has been analyzed for different types of dose distributions through simulation. The TF is modeled based on the 3-D point spread function (PSF) generated using a Monte Carlo simulation program CASINO[29] for the substrate system of 100nm PMMA on Si, the beam energy of 50 keV, and the beam diameter of 6nm. The total exposure and forward scattering range (the standard deviation of Gaussian) are extracted from the PSF for each of the five resist layers. The ratios of the total energy and forward scattering range among the five resist layers are referred to in setting the total exposure and σ_t of the Gaussian function used to generate the TF of each layer.

The 3-D exposure distribution in the resist is computed at the resolution I_{sm}, referred to as simulation interval, which is set to $\frac{1}{2}$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.[27] 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

measured. The middle 80% segment of the developed feature along the length dimension is used in the computation of the CD and LER to exclude the edge effect (rounding at corners). The CD and LER in each case are averaged over 5 simulations.

Two different widths of a single-line are considered for performance analysis, i.e., 50 and 150nm, and the length of the line is fixed at 300nm. Each row of pixels is exposed by five beams with the beam size of 10nm. Both sharp and broad TFs are considered, i.e., $\sigma_t = 1$ and 4nm. The line features are corrected with four types of dose distributions, i.e., uniform, V-type, M-type, and A-type minimizing the CD error and LER at the top, middle, and bottom layers. The values of a_1, a_2, and a_3 in the cost function are selected to be 1, 0, and 0.2, respectively. a_2 is set to 0 since the effect of the LER is not significant in any type of dose distribution as long as $\Delta W > 0$. As the main objective is to minimize the CD error, the weight of the CD error is 5 times larger than that of the total dose. In the optimization procedure, $x = 0.5$ initially, $s_1 = 2$, $s_2 = 0.5$, $y = 0.01$ and $tolerance = 0.005$.

The average CD error, required total dose and LER obtained with ΔW varied are provided in Figs. 5.15 and 5.16 for the feature width of 50nm. From the figures, it is seen that as ΔW increases, the CD error decreases up to a certain ΔW, and then starts to increase. For a larger ΔW, the resist development in the bottom layer is more likely to catch up the resist development in the top layer and the resist profile achieves a more vertical sidewall. However, when ΔW is too large, the resist development in any layer cannot reach the target edge location unless a significantly high dose is given to the feature. For this reason, the CD error starts to increase after a certain ΔW. The required total dose increases exponentially with ΔW in order to develop the feature outside the exposed area. The LER decreases with the increase in ΔW, since the absolute variation of exposure is smaller in the unexposed area. The similar trends are observed in the results for the feature width of 150nm.

These results show that it is possible to achieve the minimal CD error and very small LER with an acceptable dose level by the shape+dose correction which utilizes the spatial dose distribution realizable on an MPES.

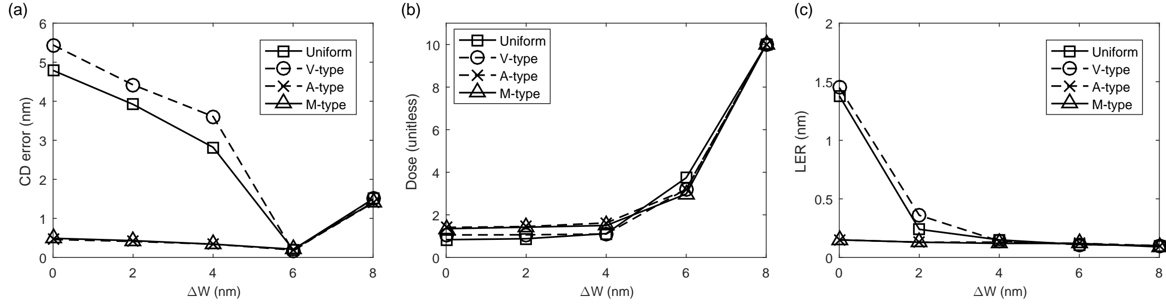

Figure 5.15. (a) The average CD error over the top, middle and bottom layers of resist, (b) total dose required, and (c) LER at the middle layer are plotted with respect to the line-width reduction ΔW for $\sigma_t = 1$nm where feature size: 50nm $\times$ 300nm, beam size: 10nm $\times$ 10nm, beam interval: 40nm, resist thickness: 100nm. For each ΔW, the optimal spatial-dose-distribution ratio is determined using the iterative procedure described in Section 5.2.

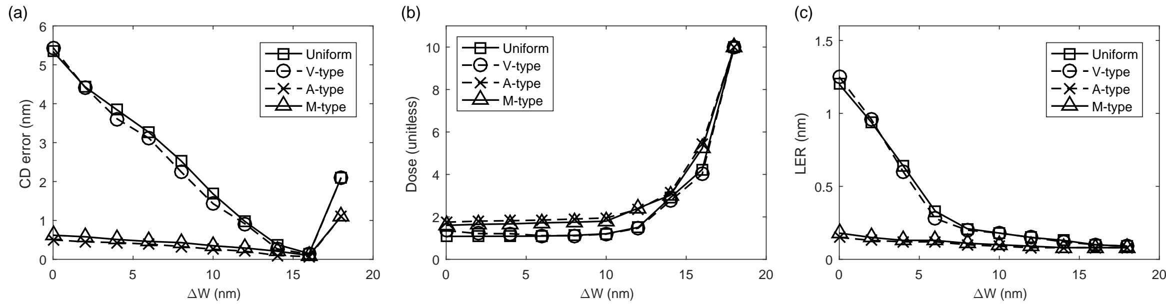

Figure 5.16. (a) The average CD error over the top, middle and bottom layers of resist, (b) total dose required, and (c) LER at the middle layer are plotted with respect to the line-width reduction ΔW for $\sigma_t = 4$nm where feature size: 50nm $\times$ 300nm, beam size: 10nm $\times$ 10nm, beam interval: 40nm, resist thickness: 100nm. For each ΔW, the optimal spatial-dose-distribution ratio is determined using the iterative procedure described in Section 5.2.

In Tables 5.1 and 5.3, the optimal ΔW and spatial-dose-distribution ratio obtained by minimizing only the average CD error ($a_1 = 1, a_2 = 0$, and $a_3 = 0$) with different types of dose distributions are provided for the feature width of 50nm, while in Tables 5.2 and 5.4, those obtained by minimizing the cost function ($a_1 = 1, a_2 = 0$, and $a_3 = 0.2$). First, in none of the cases considered, the optimal ΔW is an integer multiple of beam size, B. This well demonstrates that a spatial dose control for the cases when $\Delta W \neq nB$ considering the constraints of the MPES is necessary to obtain the optimal correction results. Second, it is seen that the CD error achieved by the shape+dose correction method is very small in all cases. That is, the spatial dose distributions realized on an MPES are effective though they deviate from the ideal distributions. Third, the non-uniform dose distributions give better results in most

of the cases than the uniform dose distribution, which illustrates the importance of realizing non-uniform dose distributions with varying ΔW. Fourth, the optimal ΔW and spatial-dose-distribution ratios are different between without (i.e., CD error only) and with considering the cost function in most cases for both σ_t of 1nm and 4nm. That is, minimizing the CD error only is not necessarily optimal when the total dose also needs to be taken into account. The required total dose can be made smaller (as seen in Tables 5.2 and 5.4) by increasing its weight in the cost function such that the average CD error still remains relatively small.

Type	Spatial-dose-distribution ratio	ΔW_{opt} (nm)	CD error (nm)	Dose (unitless)
Uniform	1:1:1:1:1	16	0.15	4.23
V	1.17:1.06:1:1.06:1.17	16	0.13	4.01
A	1:2:3.15:2:1	16	0.06	5.48
M	1:3.9:2:3.9:1	16	0.11	5.15

TABLE 5.1. The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing only the average CD error for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 4$nm

Type	Spatial-dose-distribution ratio	ΔW_{opt} (nm)	CD error (nm)	Dose (unitless)
Uniform	1:1:1:1:1	14	0.30	3.11
V	4.5:2:1:2:4.5	14	0.24	2.99
A	1:2:5.4:2:1	10	0.22	2.15
M	1:7.5:2:7.5:1	10	0.20	2.26

TABLE 5.2. The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing the cost function for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 4$nm

Type	Spatial-dose-distribution ratio	ΔW_{opt} (nm)	CD error (nm)	Dose (unitless)
Uniform	1:1:1:1:1	6	0.20	3.74
V	1.21:1.16:1:1.16:1.21	6	0.15	3.19
A	1:2:4.49:2:1	6	0.18	3.21
M	1:6.01:2:6.01:1	6	0.20	2.98

TABLE 5.3. The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing only the average CD error for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 1$nm

Type	Spatial-dose-distribution ratio	ΔW_{opt} (nm)	CD error (nm)	Dose (unitless)
Uniform	1:1:1:1:1	6	0.18	3.84
V	1.21:1.16:1:1.16:1.21	6	0.15	3.24
A	1:2:5.06:2:1	4	0.30	1.74
M	1:6.75:2:6.75:1	4	0.32	1.69

TABLE 5.4. The optimal ΔW and spatial-dose-distribution ratio obtained by minimizing the cost function for uniform, V-type, A-type, and M-type dose distribution. $W = 50$nm, $\sigma_t = 1$nm

5.4 Summary

In this chapter, a practical method of realizing four types of dose distributions, i.e., uniform, V-type, A-type, and M-type, with varying line-width reduction (ΔW) under the constraints of the MPES is presented. Also, a shape+dose control procedure which uses the spatial dose distribution (with a width reduction) realizable on an MPES is described. The results show that the shape+dose control procedure can optimally decrease the CD error and LER with a relatively low total dose. In most cases, the optimal ΔW is not an integer multiple of the beam size and the non-uniform dose distributions perform better than the uniform dose distribution. These results clearly show that it is essential to have practical methods to realize various spatial dose distributions with any ΔW on an MPES.

Chapter 6

Reduction of Exposing Time

With the existing writing methods, the beams are turned off when they fall on the spaces between two features in a pattern. Also, the beams are turned off selectively for realizing non-uniform distributions. As a result, the utilization of beams is reduced. In this chapter, two methods are described which are developed to increase the beam utilization and, therefore, reduce the exposing time.[24]

6.1 Methods for reducing exposing time

In this section, two methods to increase the utilization of beams and reduce the exposing time, utilizing the cycles with all the beams turned off and lowering the dose difference among the regions of a feature while implementing the PEC, are described.

6.1.1 Writing method without empty cycle

In the conventional (single-row and multi-row) writing methods, the substrate moves continuously underneath the array of beams and each beam follows a pixel (single-row) or a set of pixels (multi-row), by deflecting in each step during the $(\frac{I_{bx}}{B} + 1)$ steps of a cycle. Then, the beam is reset back to its vertical orientation and exposes another pixel or a set of pixels 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." The number of empty cycles varies with the width of the features (l), space regions (s) and pattern, target dose distribution, I_{bx} and the number of beams (n) in the system.

To remove the empty cycles in the conventional writing method while realizing a dose distribution, the deflection angle of beams is adjusted in the proposed writing method such that the beams skip to the next cycle in which at least one beam falls on a pixel requiring a dose. Fig. 6.1 illustrates the idea of the proposed writing method. All the beams are turned off during the 4th cycle in the conventional (single-row) writing method (Fig. 6.1(a)), while the deflection angle of beams is adjusted during the 4th cycle in the proposed writing method to expose the next available pixels in the pattern (Fig. 6.1(b)). Clearly, the higher the number of empty cycles in the conventional writing method, the larger the reduction in exposing time by the proposed writing method.

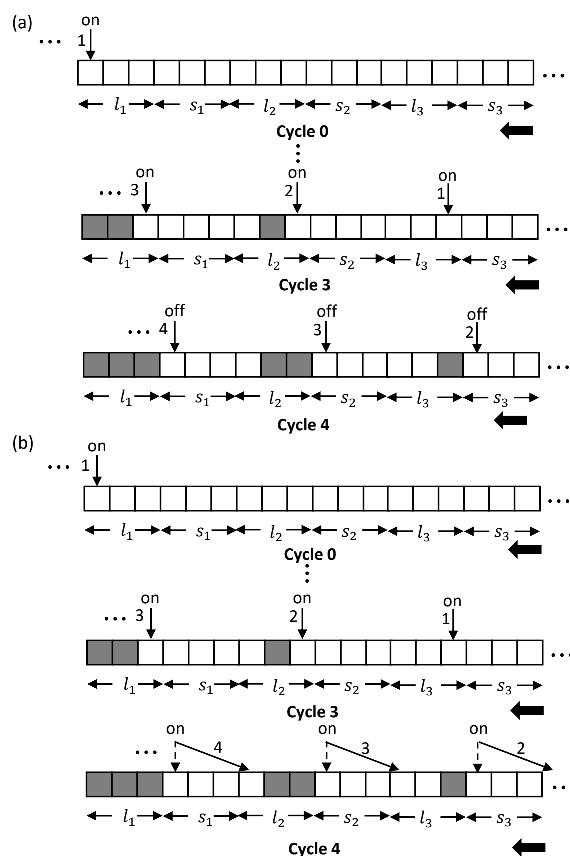

Figure 6.1. Comparison of the writing methods: (a) the conventional (single-row) writing method, and (b) the proposed writing method. The solid arrows indicate the orientation of beams, the numbers denote the beam IDs, and the squares represent a row of pixels. The substrate is moving to the left as 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 a line/space pattern and s_i is the space between two adjacent lines l_i and l_{i+1}.

Suppose that a pattern of width W_p is exposed by n beams in the system, where $W_p = (l_1 + s_1) + (l_2 + s_2) + ... + (l_j + s_j)$. Two states can be considered while exposing a pattern, *transient* and *steady* states. The transient state occurs at the beginning and end of the exposure when all the beams are not over the pattern. On the other hand, the steady state occurs when all the beams are over the pattern. Suppose that m cycles are required to give the target dose distribution to the entire line/space pattern in the conventional writing method. Due to the spaces between features and some pixels not requiring any dose based on the shape of the dose distribution, all the beams in the system are turned off in some cycles. By tracing the writing method through simulation, the number of empty cycles, m_ϕ, in the conventional writing method can be calculated. In the proposed writing method, each empty cycle is skipped to the next available non-empty cycle by changing the initial vertical orientation ($\theta = 0$) to an angle θ. Therefore, the number of cycles required to give the target dose distribution to the entire pattern in the proposed writing method is $(m - m_\phi)$.

The percentage reduction of exposing time by the proposed writing method with respect to the conventional writing method is denoted by t_{red}. t_{red} is measured as the ratio of the difference between the required cycles of the two writing methods to the required cycles in the conventional writing method and multiplied by 100.

$$t_{red} = \frac{m - (m - m_\phi)}{m} \times 100\% = \frac{m_\phi}{m} \times 100\% \tag{6.1}$$

6.1.2 Reduction of dose difference among the regions

In Chapter 5, the shape+dose correction was implemented through an iterative procedure to find the optimal linewidth reduction, ΔW, and the optimal spatial-dose-distribution ratio. To have a sufficient spatial control of the dose distribution, the line feature is partitioned into five regions along its length dimension. However, the higher the dose difference among the regions, the lower the beam utilization and, therefore, the longer the exposing time. After finding the optimal dose-distribution ratio, the difference of doses among the regions is carefully reduced by lowering the maximum dose among the five regions with a certain amount and distributing that amount to the nearby regions evenly so that the beam utilization is increased and, hence,

the exposing time is reduced (Fig. 6.2). The cost function to be optimized, C, consists of four metrics, i.e., the CD error, LER, total dose and beam utilization U,

$$C = a_1 CD_error_{norm} + a_2 LER_{norm} + a_3 total_dose_{norm} + a_4 \frac{1}{U_{norm}} \tag{6.2}$$

where a_1, a_2, a_3, and a_4 are the weights given to the normalized CD_error, LER, $total_dose$, and U, respectively.

The CD_error is measured as the absolute difference between the target and actual edge locations, averaged over the top, middle, and bottom layers. The LER is quantified as the standard deviation of edge location. The $total_dose$ is the integration of area dose over the feature width. The beam utilization, U, is defined as the fraction of time the beams are on with respect to the total exposing time. 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:

$$\tilde{X} = \frac{X - X_{min}}{X_{range}} \tag{6.3}$$

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 procedure, 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 in a region of the feature, 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 (e.g., $P = 10$). 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.[20] 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 uniform or non-uniform 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 minimum beam utilization is obtained when one of the regions in a feature receives the dose $d_{max_allowed}$ and the rest receive the dose $d_{min_allowed}$. The maximum beam utilization is 1.

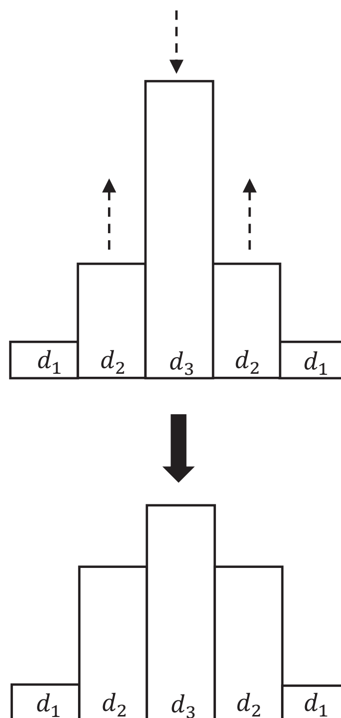

Figure 6.2. Reduction of dose difference among the five regions to improve the beam utilization and reduce the exposing time.

The optimal reduction of dose difference among the five regions of a feature is determined through an iterative procedure such that the cost function in Eq. 6.2 is minimized. The following notations are used in the description of the procedure.

C_{min}: The minimum value of cost function
$\{d_i\}$: The spatial dose distribution
Δd_i: The amount of adjustment for d_i. $\Delta d_i = d_i \times x$ where $0 < x < 1$.
$\{d_i\}_{opt}$: The optimal spatial dose distribution
d_{max}: The maximum dose among the regions of a feature

The iterative procedure to reduce the dose difference among the regions of a feature is described below where k is the iteration index, and its flowchart is also provided in Fig. 6.3.

1. $C_{min} \leftarrow$ a large value.
2. Initialize $\{d_i\}$ with the optimal dose distribution obtained by the shape+dose correction procedure (see Chapter 5) and determine the exposing time required for $\{d_i\}$.
3. Obtain the realizable dose distribution given the optimal ΔW and $\{d_i | i = 1, 2, 3, 4, 5\}$.
4. Evaluate the cost function, C_k.
5. If $C_k < C_{min}$,

   $C_{min} \leftarrow C_k$, 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 $0 < y < 1$, i.e., the improvement in C is too small,

   $x \leftarrow s_1 x$ ($s_1 > 1$).
8. If $C_k > C_{k-1} + C_{k-1} \times y$,

   $x \leftarrow s_2 x$ ($0 < s_2 < 1$).
9. $d_{max} \leftarrow d_{max} - \Delta d_i$ 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. Output $\{d_i\}_{opt}$ and the exposing time required for $\{d_i\}_{opt}$.

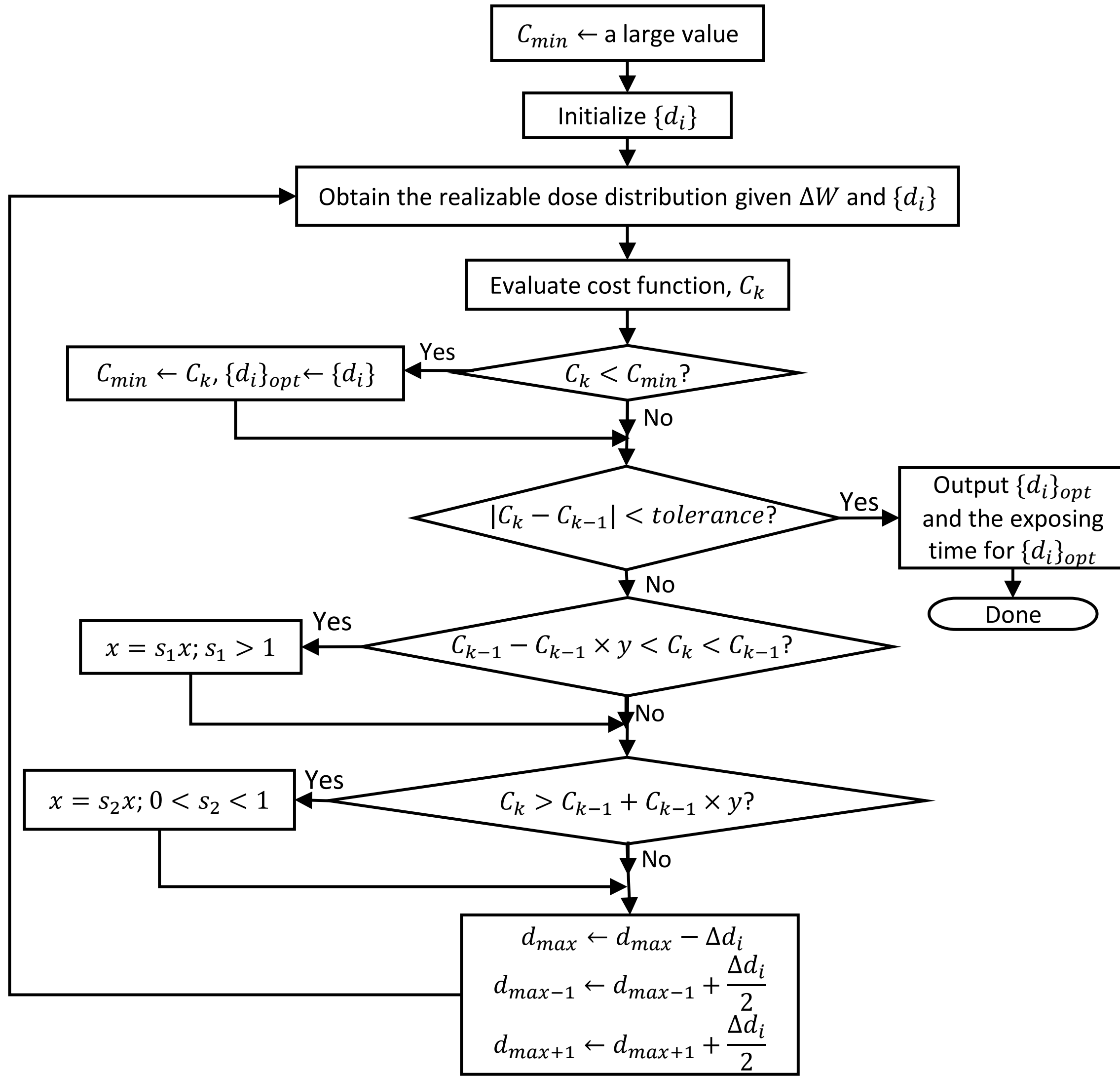

Figure 6.3. The flowchart of reducing the dose difference among the regions of a feature to improve the beam utilization and shorten the exposing time where k is the iteration index, $0 < x < 1$, and $0 < y < 1$.

6.2 Simulation

The effectiveness of the proposed methods to reduce the exposing time has been analyzed through simulation. The TF is modeled based on the 3-D point spread function (PSF) generated using a Monte Carlo simulation program CASINO[29] for the substrate system of 100nm PMMA on Si, the beam energy of 50 keV, and the beam diameter of 6nm. The total exposure and forward scattering range (the standard deviation of Gaussian) are extracted from the PSF for each of the five resist layers. The ratios of the total energy and forward scattering range among

the five resist layers are referred to in setting the total exposure and σ_t of the Gaussian function used to generate the TF of each layer.

The 3-D exposure distribution in the resist is computed at the resolution I_{sm}, referred to as simulation interval, which is set to $\frac{1}{2}$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.[27] 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 measured. The cross-section and top-down views of the remaining resist profile are provided in Fig. 6.4 where the top, middle, and bottom layers and their respective edges are illustrated. The CD error is quantified as the absolute difference between the target and actual edge locations and the LER is quantified as the standard deviation of edge location along the length dimension of a feature, both averaged over the top, middle, and bottom layers. 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).

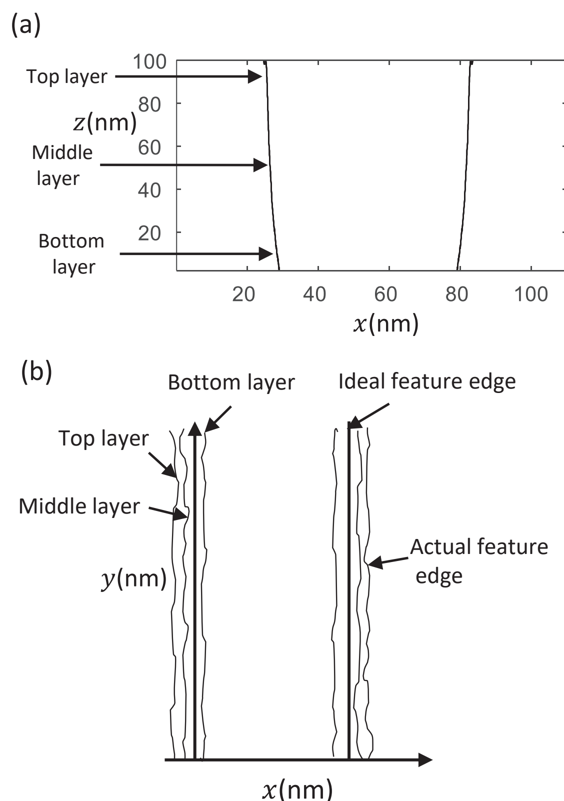

Figure 6.4. (a) The cross-section view and (b) the top-down view of the remaining resist profile. The feature size is 50nm×300nm and the resist thickness is 100nm.

To analyze the percent reduction of exposing time, t_{red}, by the proposed writing method, the widths of the features (l), spaces (s) and pattern (W_p) are varied while the beam interval I_{bx} is kept at 100nm. The numbers of cycles and empty cycles in the conventional writing method are calculated by tracing the writing procedure on the entire pattern through simulation. Then, t_{red} is calculated using Eq. 6.1. 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 in the steady state 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. Therefore, for simplicity, given a beam interval I_{bx}, $(\frac{I_{bx}}{B}+1)$ beams in a row are considered.

To observe the performance of lowering the dose difference among the regions in reducing the exposing time, two different widths of a single-line, i.e., 50 and 150nm, are considered and the length of the line is fixed at 300nm. The beam size is 10nm $\times$ 10nm, and both sharp and broad TFs are considered, i.e., σ_t = 1nm and 4nm. The line features are corrected with four types of dose distributions, i.e., uniform, V-type, M-type, and A-type.[5] It was shown previously that the A-type dose distribution is effective when the aspect ratio (resist thickness to feature width) is relatively large while the uniform or V-type distribution tends to work better for a relatively small aspect ratio. In Chapter 5, a systematic method for realizing various types of spatial dose distributions with an arbitrary reduction of feature size on an MPES was developed. The method utilizes the multi-pass writing, i.e., exposes each writing path multiple times, to increase the dose range and/or dose 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 substrate in each pass. Another benefit of using multi-pass is to spread the negative effects of abnormal beams spatially (see Chapter 4). The weights (a_1, a_2, a_3, a_4) in the cost function are selected according to the relative importance of each metric. In this study, a_1, a_2, a_3, and a_4 are selected to be 1, 0, 0.2, and 0.5, respectively. Also, a_2 is set to 0 since the effect of the LER is not significant in any type of dose distribution as long as $\Delta W > 0$. As the main objective of the PEC is to minimize the CD error, the weight of the CD error (a_1) is 5 times larger than that of the total dose and 2 times larger than that of U. In the optimization procedure, $x = 0.5$ initially, $s_1 = 2$, $s_2 = 0.5$, $y = 0.01$ 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) that can lead to a very good PEC accuracy. To calculate the exposing time, the beam interval $I_{bx} = 100$nm, number of beams in a row $n = 11$ ($n = \frac{I_{bx}}{B} + 1$), space width $s = 40$nm ($I_{bx} \neq l + s$), pattern width $W_p = 4.5\mu$m and the dose given in each step $d = 0.1$ are considered.

6.3 Results and discussion

In this section, the results are presented in two parts. The percent reduction in exposing time by the proposed writing method compared to the conventional writing methods is analyzed for various cases varying the widths of the features (l), space regions (s) and pattern (W_p), and the target dose distribution in a line/space pattern. Note that the exposing time is the same for the single-row and multi-row writing methods. Hence, the single-row writing method is used as the conventional writing method to compare to in this study. Then, the results without the PEC optimization, and with the PEC optimization both applying and without applying the method of reducing the dose difference among the regions are observed for two different linewidths and blurring factors of TF to compare the exposing time.

6.3.1 Reduction of exposing time by removing empty cycles

To analyze the percent reduction of exposing time, t_{red}, by the proposed writing method, various cases are considered which are presented in two parts, non-varying and varying total dose from feature to feature in a line/space pattern. Each part considers the relationship among I_{bx}, l and s, i.e., $I_{bx} = l + s$ and $I_{bx} \neq l + s$, and different dose distribution types, i.e., uniform and non-uniform dose distributions. In a uniform dose distribution, the same dose is given to all the regions of a feature, and in a non-uniform dose distribution, different doses are given to different regions of a feature.

In each case, t_{red} is plotted against the pattern width (size). In the transient state, the number of empty cycles, and hence t_{red}, can fluctuate significantly. A wide range of pattern width is considered so that the behavior of t_{red} in both transient and steady states can be observed.

A. Non-varying total dose from feature to feature

A.1 $I_{bx} = l + s$

A.1.1 Uniform dose distribution

The percent reduction of exposing time, t_{red}, by the proposed writing method in the case of $I_{bx} = l + s$, $l_i = l_j$ and uniform dose distribution is provided in Fig. 6.5. It is observed that t_{red} is proportional to s, the space between two adjacent features. In this case, all the beams fall on the pixels which have the same positions in their respective features. Hence, in the conventional writing method, when the beams fall on the spaces, they are turned off together (see Fig. 6.1(a)). As a result, all the beams are turned off $\frac{s}{I_{bx}}$ of the total exposing time. In the proposed writing method, when the beams fall on the spaces, the deflection angle of the beams is adjusted to expose the pixels requiring doses (see Fig. 6.1(b)). As a result, the beams are always on. Hence, in the steady state, t_{red} is proportional to s and can be expressed as in Eq. 6.4.

$$t_{red} = \frac{I_{bx} - l}{I_{bx}} \times 100\% = \frac{s}{I_{bx}} \times 100\% \tag{6.4}$$

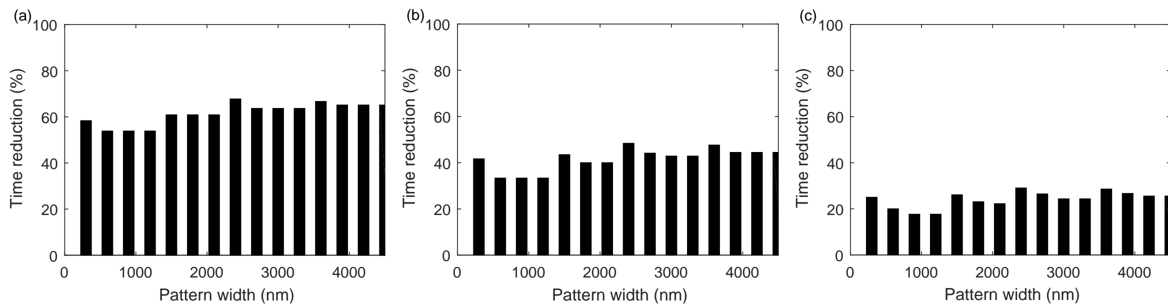

Figure 6.5. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm.

The percent reduction of exposing time, t_{red}, by the proposed writing method in the case of $I_{bx} = l + s$, $l_i \neq l_j$ and uniform dose distribution is provided in Fig. 6.6. It is seen that t_{red} is proportional to the smallest s, i.e., s_{min}. In this case, at least one beam is turned on $\frac{l_{max}}{I_{bx}}$ of the total exposing time in the conventional writing method since each beam falls on the same position in its corresponding $(l + s)$ domain. Hence, all the beams are turned off together $\frac{s_{min}}{I_{bx}}$ of the total exposing time in the conventional writing method. In the proposed writing method, the empty cycles are removed. Therefore, in the steady state, t_{red} is proportional to s_{min} and can be calculated as in Eq. 6.5.

$$t_{red} = \frac{I_{bx} - l_{max}}{I_{bx}} \times 100\% = \frac{s_{min}}{I_{bx}} \times 100\% \quad (6.5)$$

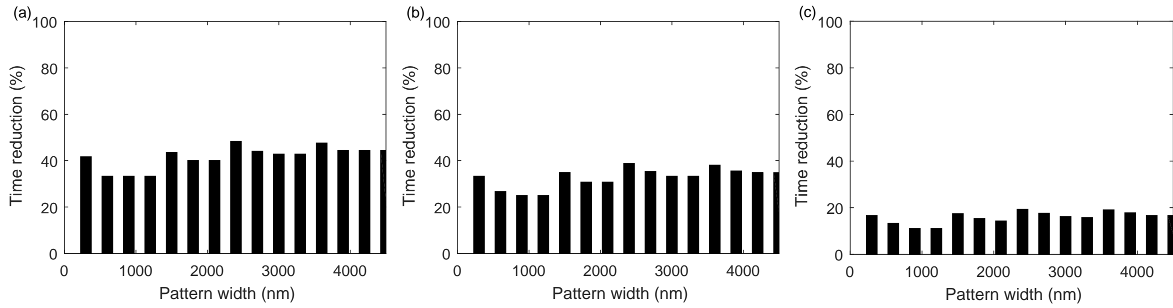

Figure 6.6. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c) $l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm.

A.1.2 Non-uniform dose distribution

In the case of non-uniform dose distribution, the number of passes to achieve the target dose is determined by the maximum dose, d_{max}, required by a pixel in the feature. However, all the pixels in a feature do not require doses in each pass. The number of pixels in the feature requiring non-zero doses in each pass is referred to as *effective feature width*, denoted by l_{eff}. This is illustrated in Fig. 6.7 where the five pixels in the feature require the doses $D, 2D, 3D, 2D$ and D, and three passes are used to achieve the target dose distribution. In the first pass, $l_{eff} = 5$, in the second pass, $l_{eff} = 3$, and in the third pass, $l_{eff} = 1$. As stated in Section 6.3.1.A.1.1, when $I_{bx} = l + s$ and $l_i = l_j$, all the beams give doses to the pixels at the

same positions in their respective features in the conventional writing method. Therefore, at least one beam is always on in l_{eff}/I_{bx} of the total cycles in each pass, and the beams are off together in the rest of the cycles. That is, a smaller l_{eff} in a pass leads to more empty cycles. Since empty cycles are removed in the proposed writing method, t_{red} in the steady state can be calculated as in the following equation

$$t_{red} = \frac{N \times I_{bx} - \sum_{k=1}^{N} l_{eff_k}}{N \times I_{bx}} \times 100\% = \frac{\frac{d_{max}}{D} \times I_{bx} - \sum_{k=1}^{\frac{d_{max}}{D}} l_{eff_k}}{\frac{d_{max}}{D} \times I_{bx}} \times 100\% \qquad (6.6)$$

where l_{eff_k} is the effective width of the feature in the k-th pass and N is the number of passes ($N = \frac{d_{max}}{D}$).

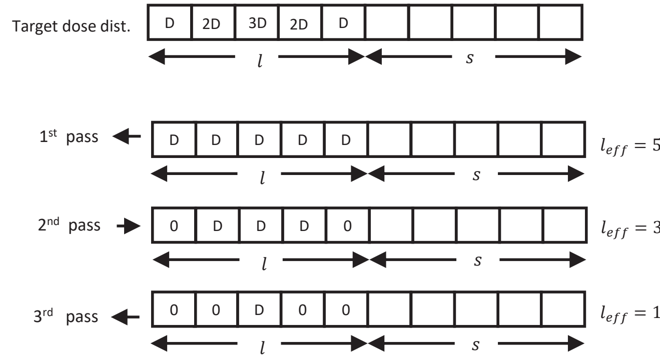

Figure 6.7. The realization of non-uniform target dose distribution in multiple passes. The feature is divided into five regions and each region consists of a pixel with size 10nm×10nm. The same figure can be referred to when each region consists of more than a pixel.

In Fig. 6.8, t_{red} is provided for $I_{bx} = l + s$, $l_i = l_j$ and non-uniform dose distribution. It is seen that the higher the dose difference among the pixels in the feature, the larger t_{red}. This is because l_{eff_k} becomes smaller in the later passes compared to the first pass when the dose difference among the pixels is larger.

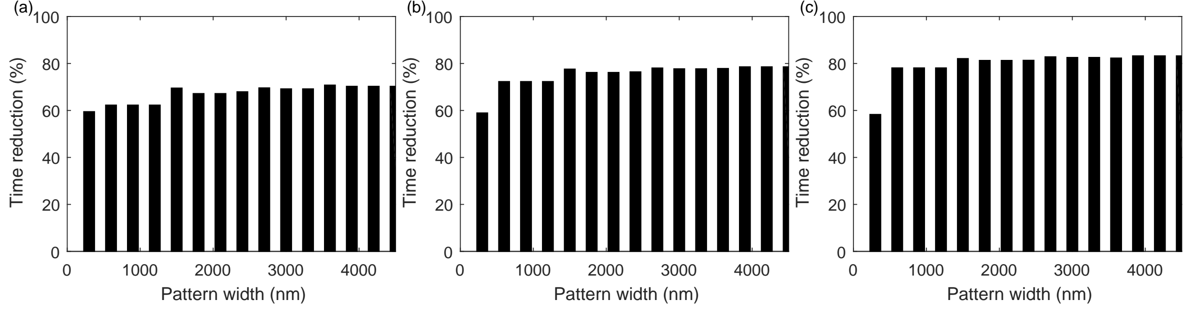

Figure 6.8. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 50$nm and the dose distribution ratio is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1.

When $I_{bx} = l + s$ and $l_i \neq l_j$, t_{red} in the steady state depends on the largest l_{eff} among features in each pass, i.e., $l_{eff_{max}}$. This is because at least one beam is turned on $\frac{l_{eff_{max}}}{I_{bx}}$ of the total exposing time in each pass in the conventional writing method similar to the respective case ($I_{bx} = l + s$ and $l_i \neq l_j$) in Section 6.3.1.A.1.1. Therefore, t_{red} in the steady state can be calculated as in Eq. 6.7

$$t_{red} = \frac{\frac{d_{max}}{D} \times I_{bx} - \sum_{k=1}^{\frac{d_{max}}{D}} l_{eff_{max_k}}}{\frac{d_{max}}{D} \times I_{bx}} \times 100\% \tag{6.7}$$

where $l_{eff_{max_k}}$ is the largest effective width among features in the k-th pass.

The simulation results of t_{red} for $I_{bx} = l + s$, $l_i \neq l_j$ and non-uniform dose distribution are provided in Fig. 6.9. Similar to the case in Fig. 6.8, the higher the dose difference among the pixels in the feature, the larger t_{red}.

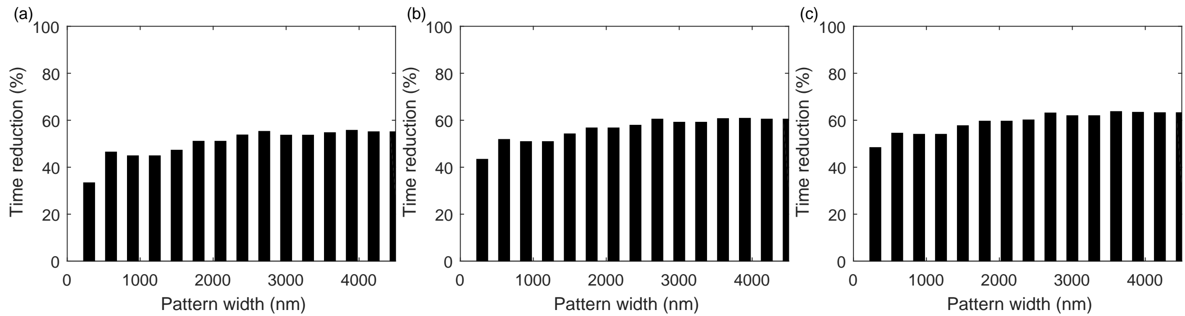

Figure 6.9. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1, 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1, 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1, 1:4:4:4:4:1 in three different widths of features.

A.2 $I_{bx} \neq l + s$

A.2.1 Uniform dose distribution

When $I_{bx} \neq l + s$, $l_i = l_j$ and the dose distribution is uniform, empty cycles are not found in the steady state in the conventional writing method as illustrated in Fig. 6.10 where $I_{bx} > l + s$. In Fig. 6.10, the pixels exposed by the beams in four consecutive cycles are shown for the conventional writing method. The four cycles are repeated in the steady state as the substrate moves underneath the beams. One or more beams are always on in the steady state. Hence, there are no empty cycles to be removed by the proposed writing method. Few empty cycles may be found in the transient state which is a small fraction of the total exposing time. Hence, t_{red} is very small in this case (Fig. 6.11).

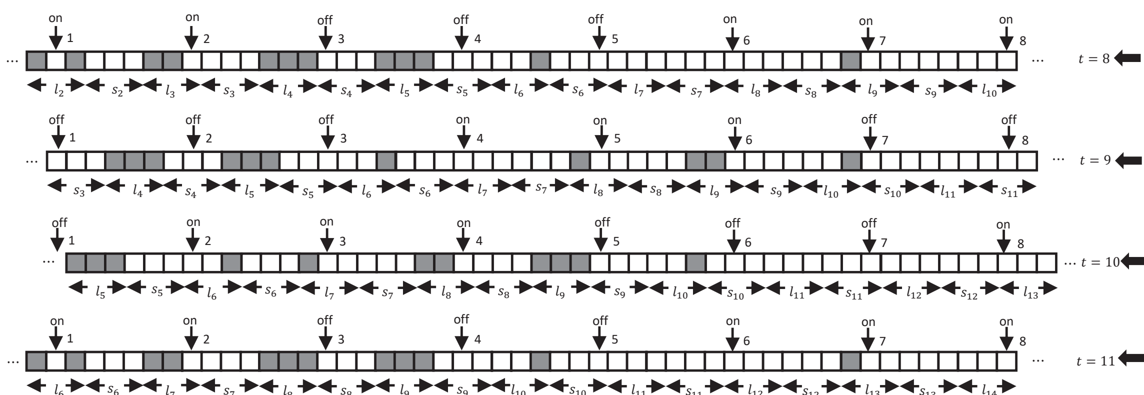

Figure 6.10. Tracing pixels exposed by the beams in multiple cycles in the steady state for the conventional writing method when $I_{bx} > l + s$ and $l_i = l_j$. In this figure, $I_{bx} = 70$nm, $l = 30$nm, and $s = 30$nm. The gray and white squares represent the exposed and unexposed pixels, respectively.

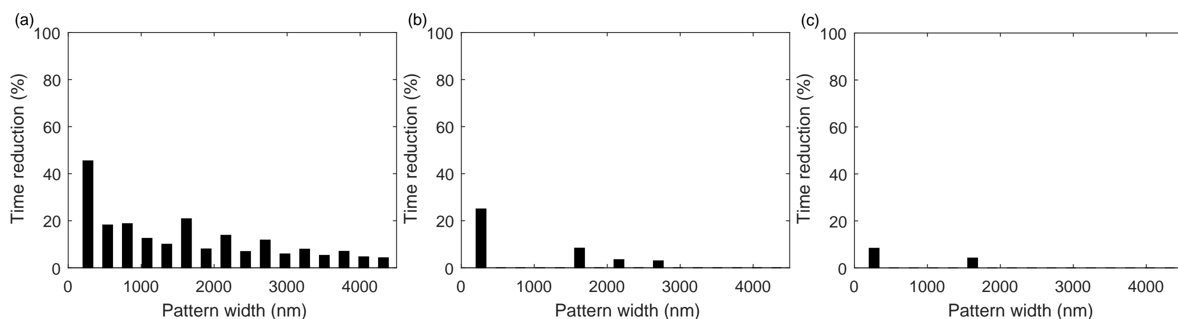

Figure 6.11. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $s = 40$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm.

In the case of $I_{bx} \neq l + s$, $l_i \neq l_j$ and uniform dose distribution, empty cycles may be observed (see Fig. 6.12). A random selection of the feature widths and spaces may cause the beams to be turned off together in some cycles. The simulation results for such a case are provided in Fig. 6.13.

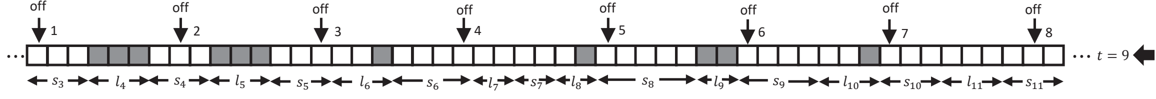

Figure 6.12. Tracing pixels exposed by the beams in the system in the second cycle of the steady state for the conventional writing method when $I_{bx} > l + s$ and $l_i \neq l_j$. The gray and white squares represent the exposed and unexposed pixels, respectively.

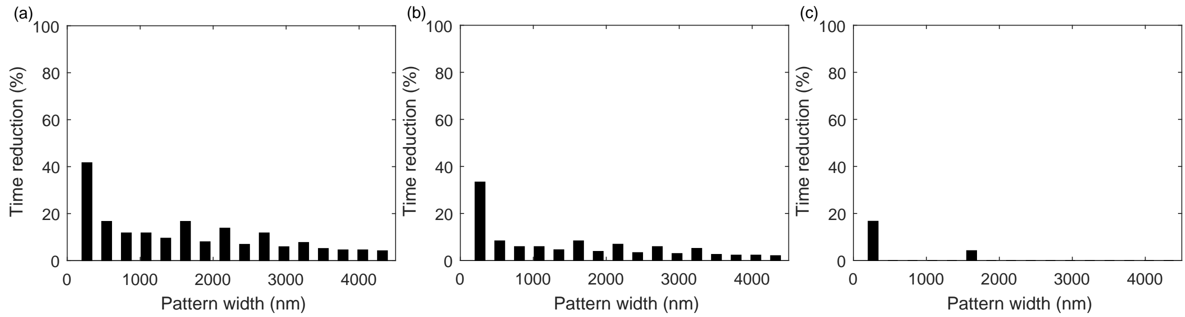

Figure 6.13. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c)$l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm.

A.2.2 Non-uniform dose distribution

When $I_{bx} \neq l + s$ (either $l_i = l_j$ or $l_i \neq l_j$) and the dose distribution is non-uniform, empty cycles are not found in the conventional writing method in the first pass (similar to the illustration in Fig. 6.10) in the steady state. In the later passes, the number of pixels in the feature requiring non-zero doses, l_{eff}, gets smaller depending on the shape of dose distribution which may lead to a few empty cycles in the conventional writing method in the steady state (similar to the illustration in Fig. 6.12). The smaller l_{eff} becomes in a pass, the higher the chance of finding empty cycles in that pass. Therefore, t_{red} is relatively small in this case (Figs. 6.14 and 6.15), but larger than the case when $I_{bx} \neq l + s$ and the dose distribution is uniform.

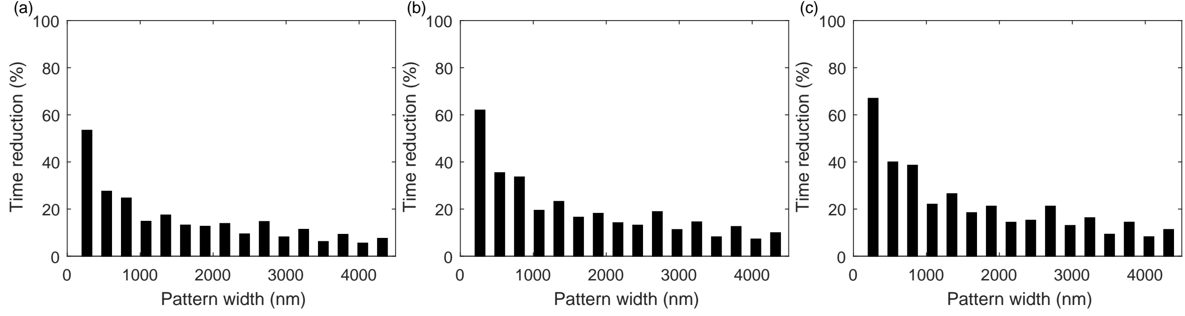

Figure 6.14. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 40$nm and the dose distribution ratio is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1.

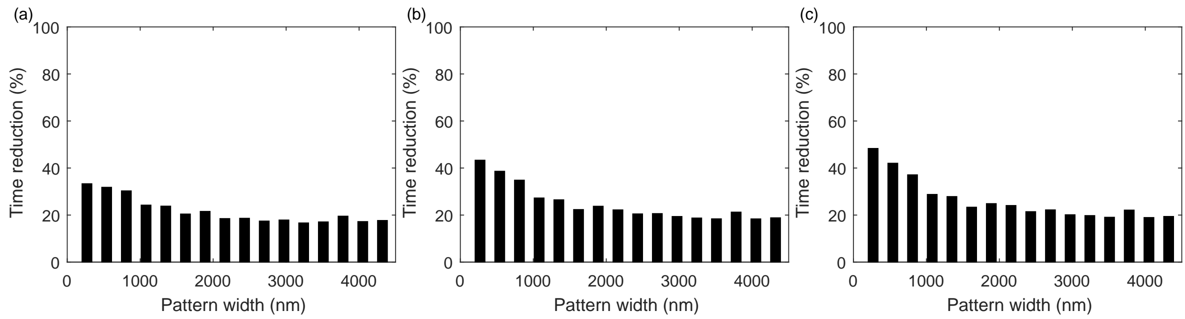

Figure 6.15. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is non-uniform, $l_1 = 40$nm, $l_2 = 50$nm, and $l_3 = 60$nm and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1 and 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1 and 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1 and 1:4:4:4:4:1, respectively, for three different widths of features.

B. Varying total dose from feature to feature

For correcting the proximity effect, the features on the edges of a pattern tend to get higher doses than the features in the center. This type of variation of total dose from feature to feature increases the probability of getting empty cycles in the conventional writing method compared to the non-varying total dose from feature to feature. This is because after giving the required doses to the features in the center in the first few passes, the later passes are used only to give the remaining doses to the features on the edges and the beams are turned off together when they fall on the features in the center. Therefore, t_{red} is larger when the total dose varies from feature to feature compared to when the total dose does not vary. Also, the higher the dose difference among the features of a pattern, the larger t_{red} since the number of passes to give doses to the features on the edges, which cause empty cycles, becomes larger.

In Figs. 6.16 and 6.17, t_{red} in the case of $I_{bx} = l + s$ and uniform dose distribution is provided for $l_i = l_j$ and $l_i \neq l_j$, respectively. The only difference between this case and the case considered in Section 6.3.1.A.1.1 is the variation of total dose from feature to feature. It is observed that the trend of t_{red} with respect to the pattern width is similar between the cases, but t_{red} is larger in Figs. 6.16 and 6.17 compared to that in Figs. 6.5 and 6.6, respectively. Similarly, when $I_{bx} = l + s$, $l_i = l_j$ or $l_i \neq l_j$ and the dose distribution is non-uniform, t_{red} is larger in Figs. 6.18 and 6.19 for varying total dose from feature to feature compared to that in Figs. 6.8 and 6.9, respectively, for non-varying total dose from feature to feature.

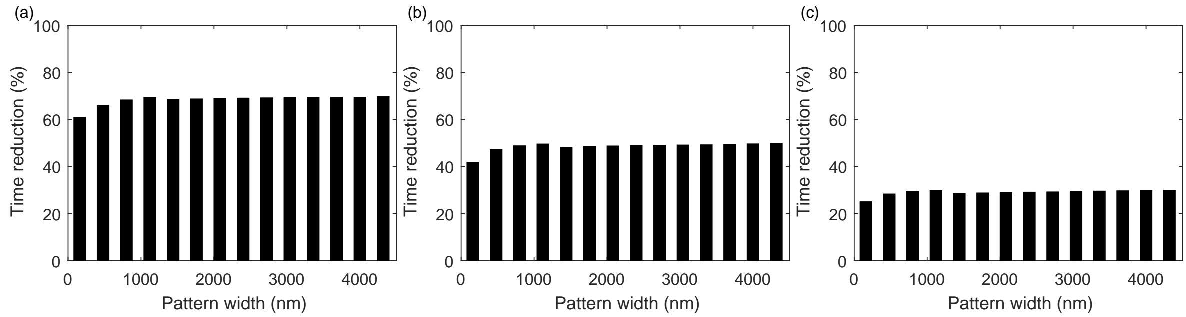

Figure 6.16. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

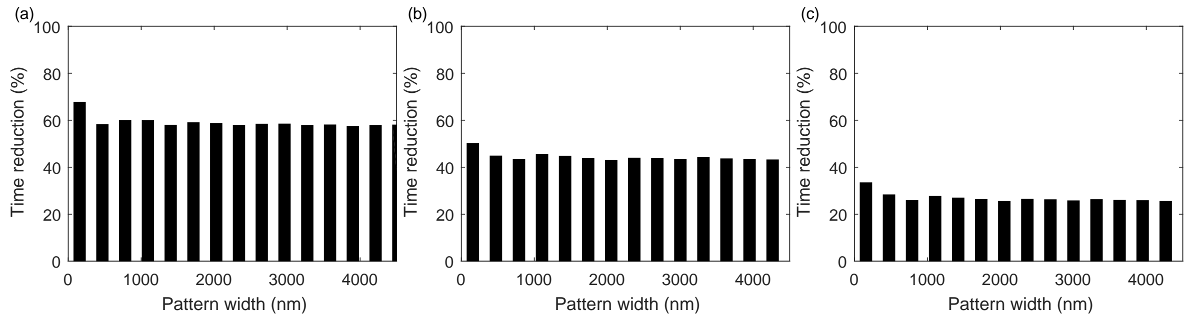

Figure 6.17. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c) $l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

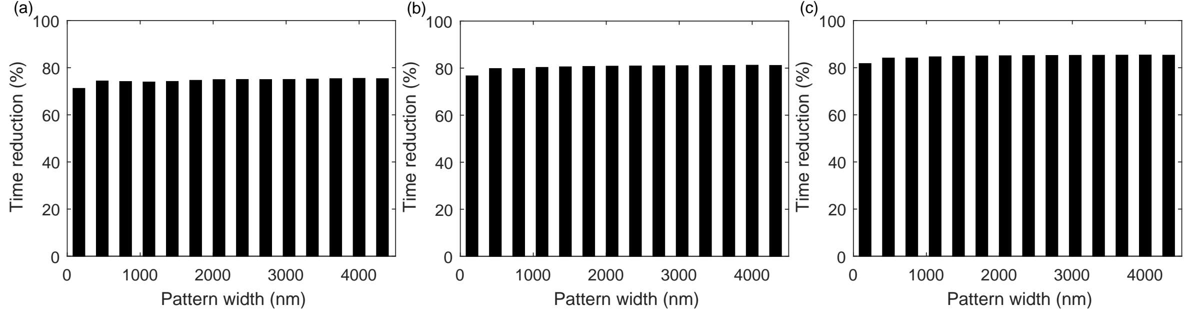

Figure 6.18. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 50$nm and the dose distribution ratio in each feature is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

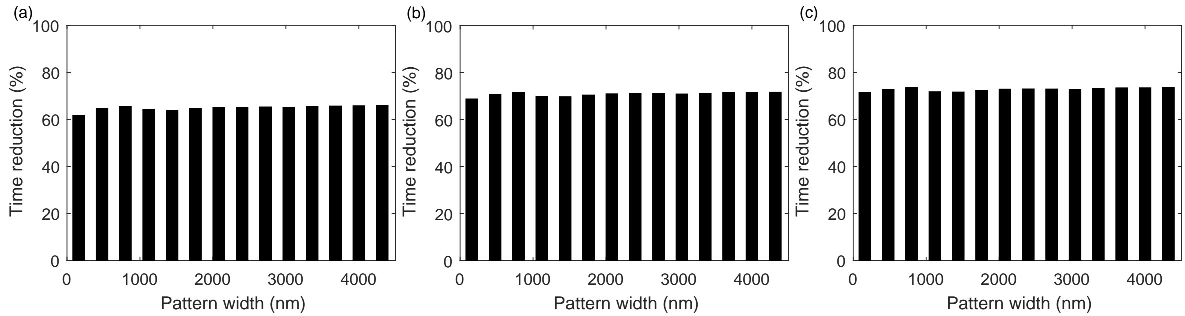

Figure 6.19. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} = l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1, 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1, 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1, 1:4:4:4:4:1 in three different widths of features. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

In Figs. 6.20 and 6.21, t_{red} in the case of $I_{bx} \neq l + s$ and uniform dose distribution is provided for $l_i = l_j$ and $l_i \neq l_j$, respectively. It is observed that t_{red} in Figs. 6.20 and 6.21 is significantly larger than that in Figs. 6.11 and 6.13, respectively, where the total dose does not vary from feature to feature. This is because the variation of total dose from feature to feature results in empty cycles in the steady state in the conventional writing method even when $I_{bx} \neq l + s$, but empty cycles are not found in the steady state in the case of non-varying total dose from feature to feature. Similarly, when $I_{bx} \neq l + s$, $l_i = l_j$ or $l_i \neq l_j$ and the dose distribution is non-uniform, t_{red} is substantially larger in Figs. 6.22 and 6.23 than that in Figs. 6.14 and 6.15, respectively, because of the variation of total dose and a small l_{eff} in the passes for giving doses on the edges of the pattern.

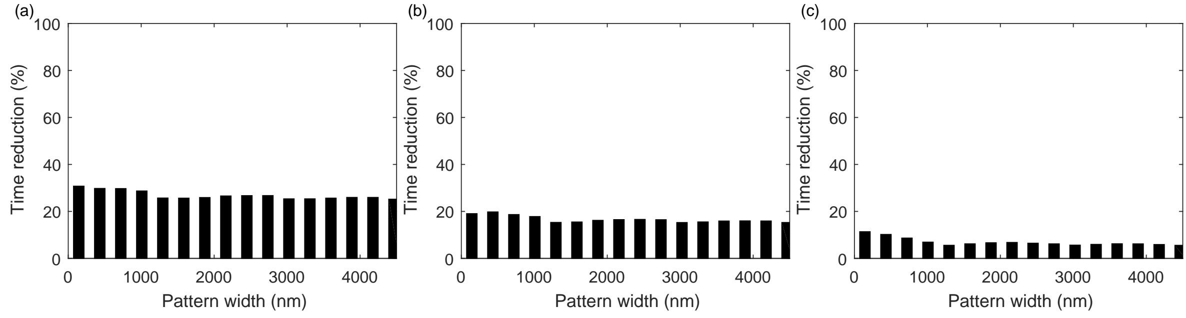

Figure 6.20. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is uniform and (a) $l = 30$nm, (b) $l = 50$nm, and (c) $l = 70$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

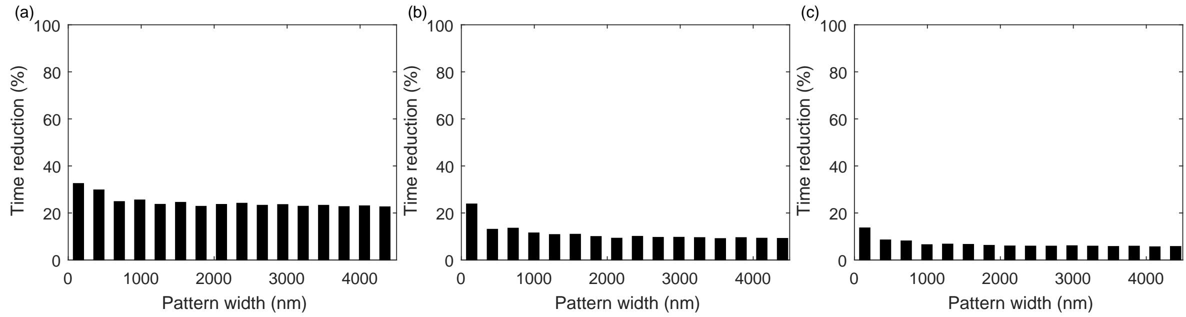

Figure 6.21. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l + s = 90$nm, the dose distribution is uniform and (a) $l_1 = 10$nm, $l_2 = 30$nm, $l_3 = 50$nm, (b) $l_1 = 40$nm, $l_2 = 50$nm, $l_3 = 60$nm, and (c) $l_1 = 60$nm, $l_2 = 70$nm, $l_3 = 80$nm. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

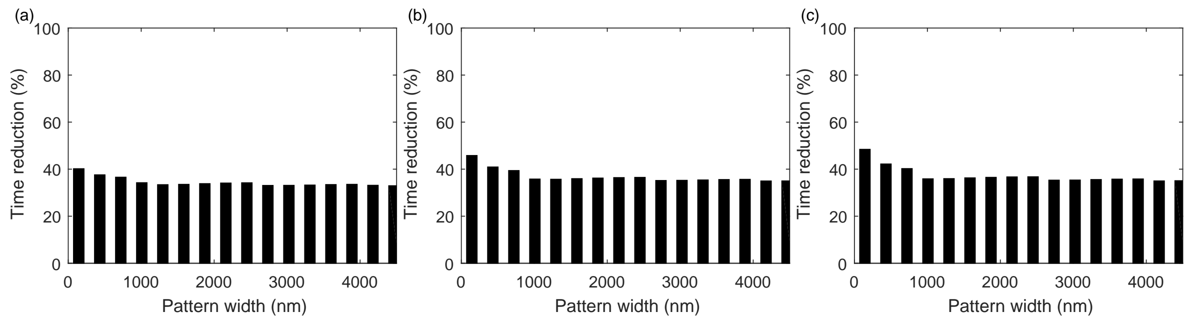

Figure 6.22. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l + s$, $l_i = l_j$, $I_{bx} = 100$nm, $l = 50$nm, $s = 40$nm and the dose distribution ratio in each feature is (a) 1:3:5:3:1, (b) 1:2:7:2:1, and (c) 1:1:9:1:1. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

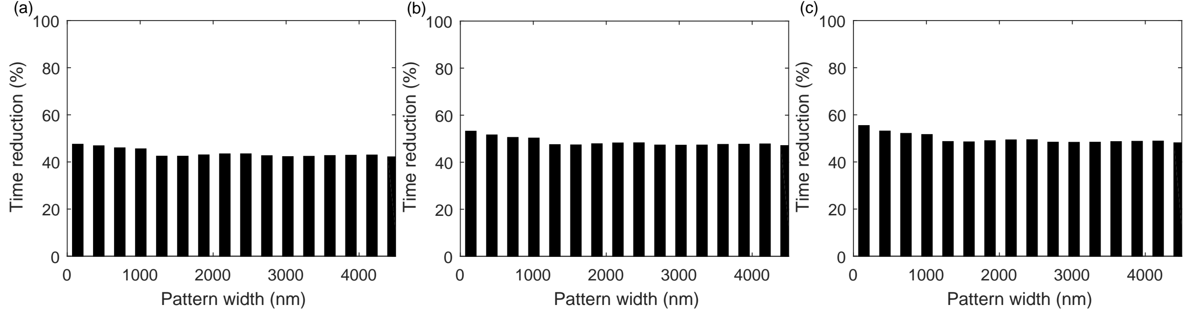

Figure 6.23. Reduction of exposing time compared to the conventional writing method, t_{red}, when $I_{bx} \neq l+s$, $l_i \neq l_j$, $I_{bx} = 100$nm, $l+s = 90$nm, $l_1 = 40$nm, $l_2 = 50$nm, and $l_3 = 60$nm, and the dose distribution ratios are (a) 1:2:2:1, 1:1:2:1:1 and 1:1:1:1:1:1, (b) 1:4:4:1, 1:2:4:2:1 and 1:2:2:2:2:1, and (c) 1:8:8:1, 1:4:8:4:1 and 1:4:4:4:4:1, respectively, in three different widths of features. The dose ratio between the edge and center feature is 1.5:1 and the doses at other features are computed by the linear interpolation.

6.3.2 Reduction of exposing time by decreasing dose difference among the regions

In Table 6.1, the results are compared in terms of the CD error, total dose, beam utilization and exposing time among three cases, i.e., without employing the PEC optimization, the PEC optimization without applying the method described in Section 6.1.2 and the PEC optimization applying the method described in Section 6.1.2. To obtain a comprehensive set of results, the feature width, blurring factor of TF and dose distribution type are varied. While calculating the exposing time, it is considered that $I_{bx} \neq l + s$. The A-type, M-type, and V-type dose distributions obtained for each of the above-mentioned cases when $l = 50$nm and $\sigma_t = 1$nm are presented in Figs. 6.24, 6.25, and 6.26, respectively. From Table 6.1, it can be seen that, compared to the case without the PEC optimization, the cases with the PEC optimization both with and without applying the method described in Section 6.1.2 achieve a significantly smaller CD error while maintaining the total dose within an acceptable limit. By comparing the optimization results when the method in Section 6.1.2 is applied with those when not applied, it is found that the increase in CD error is marginal but the increase in U and the reduction in exposing time are significant for both the conventional and the proposed writing methods. The beam utilization depends on the dose difference among the regions and the exposing time depends on the total dose and the maximum dose d_{max} among the regions. In the procedure to reduce the dose difference among the regions, the total dose is kept constant by lowering d_{max} with a certain amount and evenly distributing that amount to the nearby regions. The increased doses

in the nearby regions compensate for the reduced d_{max} in the resist development. Hence, the increase in CD error is unsubstantial while the increase in beam utilization and the reduction in exposing time are significant. Furthermore, Table 6.1 shows the exposing time required for both the conventional writing (single-row writing) method and the proposed writing method in Section 6.1.1. Hence, the reduction in exposing time by both the proposed writing method and the lowering of dose difference among the regions while implementing the PEC can also be observed by comparing the exposing time in the conventional writing method when the method in Section 6.1.2 is not applied and the exposing time in the proposed writing method when the method in Section 6.1.2 is applied. It is seen that even when $I_{bx} \neq l + s$, the reduction in exposing time is significant.

l (nm)	σ_t (nm)	Type	Cost function	ΔW (nm)	Dose ratio	CD error (nm)	Total dose (unitless)	U (%)	$t_{exp,old}$ (cycle)	$t_{exp,new}$ (cycle)
50	1	A	-	0	1:2:7:2:1	1.25	1.30	37.14	1881	1761
50	1	A	C	4	1:2.19:4.36:2.19:1	0.34	1.76	49.27	2084	1915
50	1	A	C^*	4	1:2.25:4.22:2.25:1	0.37	1.74	50.81	1940	1828
50	1	M	-	0	1:5:1:5:1	1.55	1.30	52.00	2542	2486
50	1	M	C	4	1:5.75:2.51:5.75:1	0.37	1.81	55.68	2763	2635
50	1	M	C^*	4	1:5.45:3.0:5.45:1	0.40	1.78	58.35	2641	2528
50	1	V	-	0	4:2:1:2:4	4.64	1.30	65.00	1663	1520
50	1	V	C	8	1.31:1.08:1:1.08:1.31	0.19	3.18	88.24	2668	2532
50	1	V	C^*	8	1.26:1.16:1:1.16:1.26	0.19	3.20	92.58	2618	2489
50	4	A	-	0	1:2:7:2:1	1.46	1.50	37.14	1780	1671
50	4	A	C	10	1:2.10:5.25:2.10:1	0.25	2.19	43.62	2520	2352
50	4	A	C^*	10	1:2.34:5.02:2.34:1	0.28	2.22	46.62	2369	2213
50	4	M	-	0	1:5:1:5:1	1.61	1.50	52.00	1995	1809
50	4	M	C	10	1:6.68:2.80:6.68:1	0.26	2.24	54.37	3249	3110
50	4	M	C^*	10	1:6.55:3.05:6.55:1	0.29	2.21	55.42	3115	2985
50	4	V	-	0	4:2:1:2:4	5.11	1.50	65.00	1493	1344
50	4	V	C	14	4.25:2.09:1:2.09:4.25	0.25	2.94	64.37	2154	2044
50	4	V	C^*	14	4.10:2.29:1:2.29:4.10	0.29	2.95	67.22	2050	1945
150	1	A	-	0	1:2:7:2:1	5.26	3.50	37.14	1916	1780
150	1	A	C	4	1:2.08:5.82:2.08:1	0.47	4.86	41.16	2718	2515
150	1	A	C^*	4	1:2.33:5.35:2.33:1	0.51	4.88	44.90	2508	2345
150	1	M	-	0	1:5:1:5:1	4.90	3.50	52.00	2248	2010
150	1	M	C	4	1:6.60:2.11:6.60:1	0.53	4.87	52.45	3225	3068
150	1	M	C^*	4	1:6.34:2.94:6.34:1	0.58	4.91	55.58	3059	2935
150	1	V	-	0	4:2:1:2:4	4.63	3.50	65.00	2110	1821
150	1	V	C	8	1.24:1.12:1:1.12:1.24	0.29	6.30	91.84	3761	3352
150	1	V	C^*	8	1.18:1.10:1:1.10:1.18	0.33	6.25	94.23	3627	3202
150	4	A	-	0	1:2:7:2:1	5.46	4.50	37.14	1917	1808
150	4	A	C	10	1:2.01:4.87:2.01:1	0.52	5.93	44.72	2390	2228
150	4	A	C^*	10	1:2.22:4.57:2.22:1	0.57	5.95	48.18	2156	2017
150	4	M	-	0	1:5:1:5:1	5.27	4.50	52.00	2642	2505
150	4	M	C	10	1:6.57:2.55:6.57:1	0.54	5.85	53.85	3310	3172
150	4	M	C^*	10	1:6.48:2.90:6.48:1	0.57	5.87	55.12	3225	3098
150	4	V	-	0	4:2:1:2:4	4.92	4.50	65.00	2705	2514
150	4	V	C	14	4.05:2.20:1:2.20:4.05	0.45	6.04	66.67	3115	2907
150	4	V	C^*	14	3.91:2.34:1:2.34:3.91	0.43	6.01	69.05	2954	2759

TABLE 6.1. The CD error, total dose, beam utilization, exposing time with the conventional writing method, $t_{exp,old}$, and exposing time with the proposed writing method, $t_{exp,new}$, each obtained for three cases, i.e., without the PEC optimization (no use of the cost function C), the PEC optimization without applying the method described in Section 6.1.2 (the rows with C) and the PEC optimization applying the method described in Section 6.1.2 (the rows with C^*). It is considered that $C = CD_error_{norm} + 0.2 total_dose_{norm} + 0.5\frac{1}{U_{norm}}$, $I_{bx} \neq l + s$, $s = 40$nm, $I_{bx} = 100$nm for $l = 50$nm and $I_{bx} = 200$nm for $l = 150$nm.

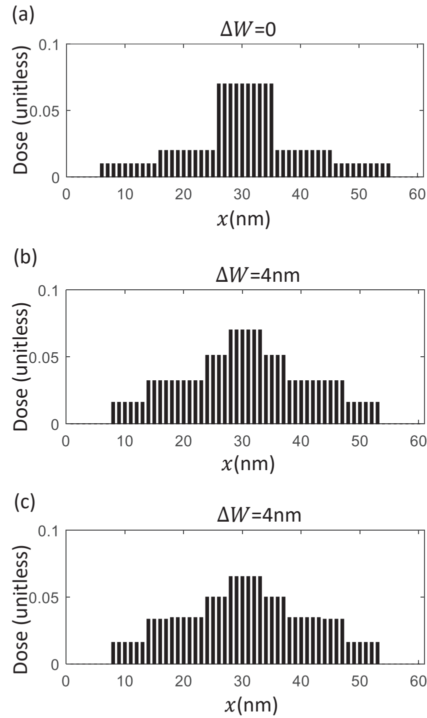

Figure 6.24. The A-type dose distribution obtained (a) without the PEC optimization, (b) with the PEC optimization which does not apply the method in Section 6.1.2, and (c) with the PEC optimization which applies the method in Section 6.1.2 for the case when $l = 50\text{nm}$ and $\sigma_t = 1\text{nm}$.

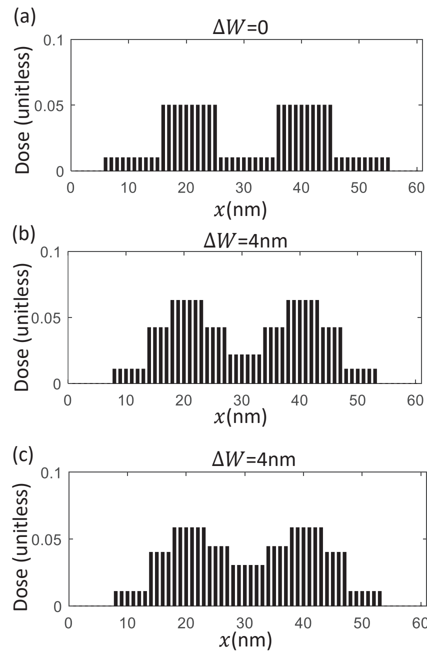

Figure 6.25. The M-type dose distribution obtained (a) without the PEC optimization, (b) with the PEC optimization which does not apply the method in Section 6.1.2, and (c) with the PEC optimization which applies the method in Section 6.1.2 for the case when $l = 50\text{nm}$ and $\sigma_t = 1\text{nm}$.

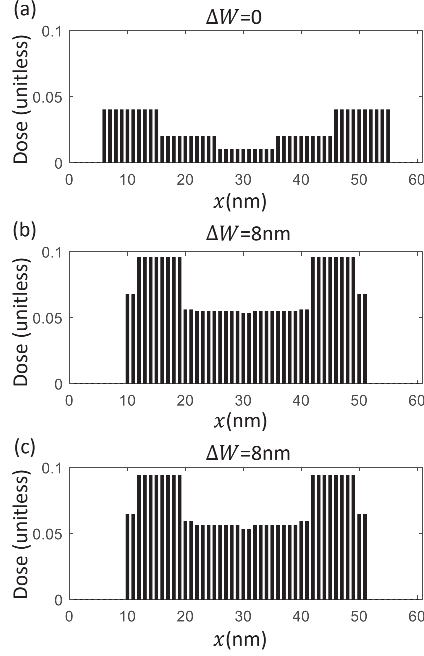

Figure 6.26. The V-type dose distribution obtained (a) without the PEC optimization, (b) with the PEC optimization which does not apply the method in Section 6.1.2, and (c) with the PEC optimization which applies the method in Section 6.1.2 for the case when $l = 50$nm and $\sigma_t = 1$nm.

6.4 Summary

Several writing methods have been introduced in previous studies for the MPES, e.g., single-row writing (SRW) and multi-row writing (MRW). A straight-forward application of these methods for realizing non-uniform dose distributions including line/space patterns would turn off certain beams selectively in several cycles. Depending on the target dose distribution and the relationship among I_{bx}, l_i and s_i (the beam interval, the width of the feature l_i in a pattern and the space between two adjacent features l_i and l_{i+1}), sometimes all the beams in the system may be turned off resulting in empty cycles. The proposed writing method removes the empty cycles in the conventional writing methods by adjusting the deflection angle of beams. From the results, it is observed that the proposed writing method can shorten the exposing time significantly, especially the cases with $I_{bx} = l + s$. In some cases with $I_{bx} \neq l + s$, the proposed writing method may not reduce the exposing time notably. In another method to reduce the exposing time, the dose difference among the regions of a feature in the optimal dose distribution for the PEC is reduced through an iterative procedure to improve the beam

utilization. It is demonstrated that the exposing time can be reduced substantially with little effect on the accuracy of CD. Moreover, the two proposed methods can be combined to shorten the exposing time significantly even when $I_{bx} \neq l + s$.

Chapter 7

A Generalized Performance Optimization

This chapter presents an adaptive optimization method that minimizes a cost function consisting of all major performance metrics, i.e., CD error, LER, and exposing time.[25] Instead of optimizing the performance metrics in a fixed order, this method considers all metrics in each iteration to adjust control parameters. The optimization method is also extended for large-scale patterns with uniform features such as L/S patterns. An effective way to handle the recursive effect among critical locations in a large-scale pattern is described.

7.1 Transfer function

The TF (refer to Fig. 7.1) 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:[31]

$$f(r) = \frac{1}{\pi(1+\eta)} \left\{ \frac{1}{\sigma_1^2} exp(-\frac{r^2}{\sigma_1^2}) + \frac{\eta}{\sigma_2^2} exp(-\frac{r^2}{\sigma_2^2}) \right\} \tag{7.1}$$

where η is the ratio of the back-scattered 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 back-scattering 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 chapter. Also, a certain level of Gaussian noise is added to the (deterministic) transfer function since the exposure in the resist is stochastic in nature (refer to Section 2.1.3).

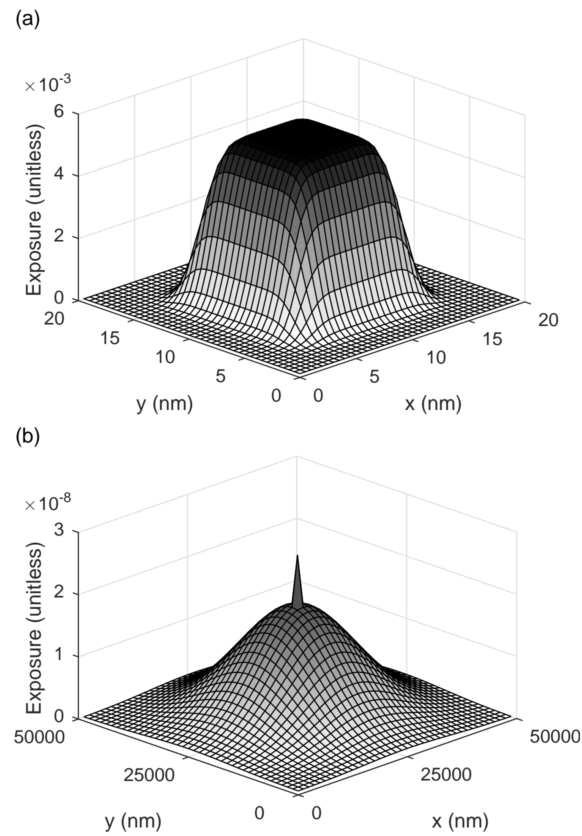

Figure 7.1. The deterministic transfer function for the beam aperture of size 10nm×10nm where (a) $\sigma_1 = 2$nm and $\sigma_2 = 0$, and (b) $\sigma_1 = 4$nm and $\sigma_2 = 9.5\mu$m.

7.2 Optimization method for a single feature

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

7.2.1 Cost function

The cost function to be minimized, C, consists of three metrics, i.e., the CD error, LER and exposing time t_{exp},

$$C = a_1 CD_error_{norm} + a_2 LER_{norm} + a_3 t_{exp_{norm}} \tag{7.2}$$

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

The cost function in Eq. 6.2 includes the total dose and beam utilization as metrics. However, reducing the total dose and increasing the beam utilization both contribute to shortening the overall exposing time. In order to directly focus on the goal of optimizing the exposing time, the cost function in Eq. 7.2 replaces the two metrics by the exposing time.

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:

$$\tilde{X} = \frac{X - X_{min}}{X_{range}} \tag{7.3}$$

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, 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 (e.g., $P = 10$). 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.[20] 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 uniform or non-uniform 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 $d_{min_allowed}$.

7.2.2 Old optimization method

In Chapter 5, 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. That is, the CD error is computed considering the top, middle, and bottom layers of resist and the maximum of the three errors is minimized such that a profile as close to the target resist profile as possible is obtained. In the optimal dose distribution for the PEC, the dose difference between any two regions can be significant causing a lot of 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 Chapter 6, 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.

7.2.3 New optimization method

In the new optimization method, all the metrics, i.e., 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 Chapter 5 is employed. In step 9, if the contribution of CD error and LER to the cost function is larger than that of exposing time, the dose distribution is adjusted to get a resist profile closer to the target. The layer with the maximum CD error is determined, and the dose of the selected region (the region of which the dose can affect the CD of that layer the most) is either increased/decreased to match the CD of that layer with the target CD (refer to Chapter 5). If the contribution of 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(refer to Chapter 6). 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 $C_k < C_{min}$,

   $C_{min} \leftarrow C_k$, $\Delta W_{opt} \leftarrow \Delta W$ 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 $C_{min} < C_{min,\Delta W}$,

    $C_{min,\Delta W} \leftarrow C_{min}$, $\Delta W \leftarrow \Delta W + 2$nm, and go to step 2.
11. Output ΔW_{opt} and $\{d_i\}_{opt}$.

7.3 Optimization method for large-scale pattern

A large-scale uniform pattern considered in this study is illustrated in Fig. 7.2, where the critical and test locations are shown.

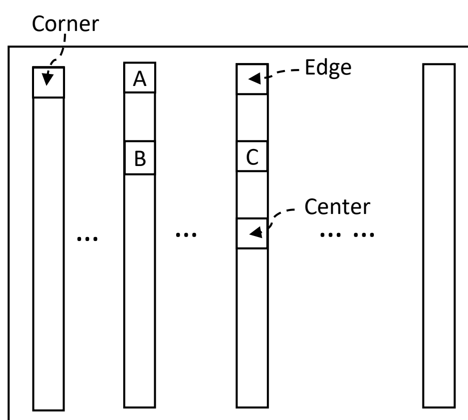

Figure 7.2. A large-scale uniform pattern with line features where the three critical locations (corner, edge and center) and test locations (A, B and C) are shown.

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 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. 7.2 is minimized in correcting a single instance of the repeated feature in a L/S pattern in isolation (refer to Section 7.2.3). 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:

$$A = E_t *^{-1} TF \tag{7.4}$$

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 Section 7.1) 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 *deconvolution surface*[32] (see Fig. 7.3). 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:

$$\{d_i\}_{(m,n)} = \{d_i\}_{single} \cdot A_{(m,n)} \tag{7.5}$$

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.

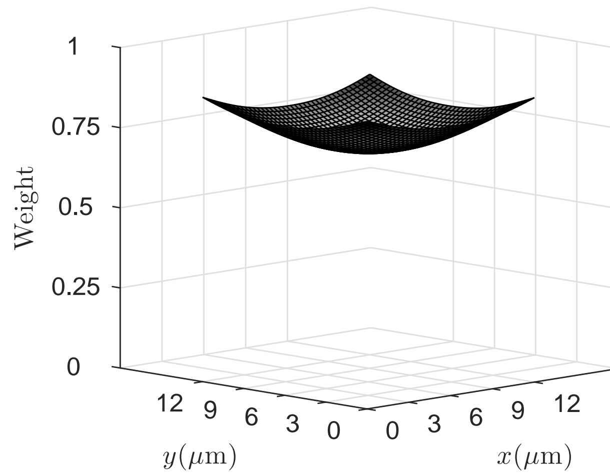

Figure 7.3. Deconvolution surface.

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. 7.2. 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. 7.5) 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. 7.2(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 Section 7.2.3, 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. 7.3). 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 over-adjustment of 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 $\{\Delta d_i\}_{opt}$ 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 flowchart of the third step in the optimization method developed for large-scale patterns is provided in Fig. 7.4.

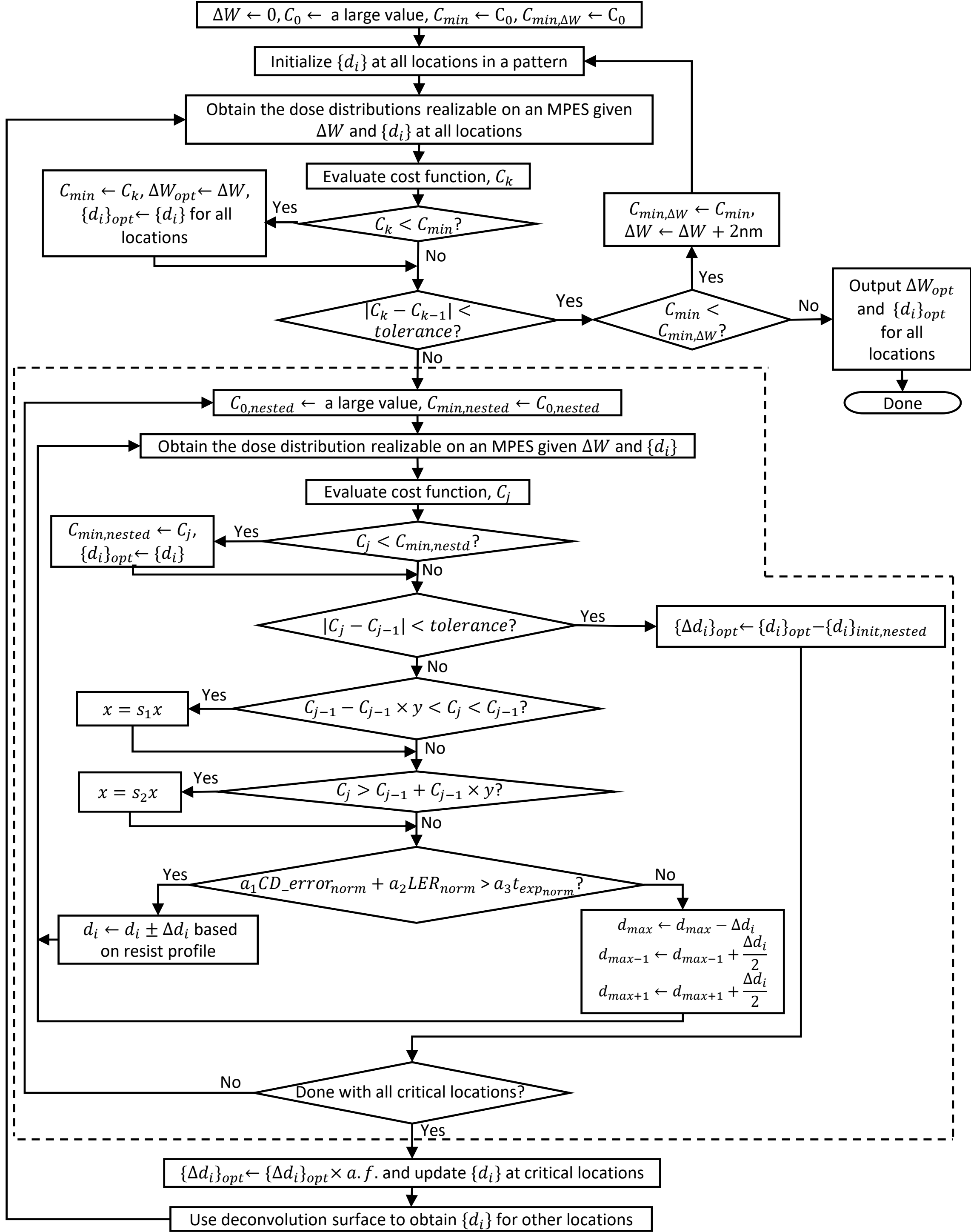

Figure 7.4. The flowchart of the proposed optimization method for a large-scale pattern where k is the iteration index, $a.f.$ is the adjustment factor, $0 < x < 1$, y is a small fraction (e.g., 0.01), $s_1 > 1$ and $0 < s_2 < 1$. The part in the dashed box is executed to obtain the optimal dose distribution 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.

7.4 Simulation

The proposed optimization methods for a single feature and a large-scale pattern have been implemented, and their performance has been analyzed through simulation. In the case of a single feature, two different linewidths, i.e., 50nm and 150nm, are considered, and the length is fixed at 300nm. In the case of a large-scale pattern, two L/S patterns of size 12μm $\times$ 12μm are considered. In one pattern, each line is 50nm wide, and the space between lines is 50nm. In the other pattern, each line is 100 nm wide, and the space between lines is 100nm.

The TF is modeled based on the 3-D point spread function (PSF) generated using a Monte Carlo simulation program CASINO,[29] where the resist (PMMA) thickness is 100nm on Si Substrate, the beam energy is 50keV, and the beam diameter is 6nm. 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 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 10nm $\times$ 10nm. In the case of a single feature, both sharp and broad TFs are considered, i.e., $\sigma_1 = 1$nm and 4nm ($\sigma_2 = 9.5\mu$m, and $\eta = 0.74$).[33] 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. 2.1). The simulation interval (spatial resolution) is set to 0.5nm. 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.[27] 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 Chapter Chapter 5 is employed. The method utilizes multi-pass 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 multi-pass is to spread the negative effects of abnormal beams spatially (refer to Chapter 4).

The number of cycles required to realize a dose distribution is calculated by tracing the writing method on the entire pattern through simulation. The writing method employed in the simulation is described in Section 6.1.1. 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 (refer to Chapter 6). 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 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 $a.f. = 0.5$.

7.5 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.

7.5.1 Single feature

In Table 7.1, 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.[34] A cost function with equal weights to all the performance metrics is considered. To obtain a comprehensive set of results, the feature width, forward scattering range σ_1 and dose distribution type are varied.

It is clear from the results in Table 7.1 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 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 7.1, 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: 50nm and resist thickness: 100nm). 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 V-type dose distribution (refer to Chapter 5). But, 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 7.1 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.[20]

In Figure 7.5, 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.

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

TABLE 7.1. The 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 12μm width which consists of lines with the linewidths of 50nm and 150nm. The space between two lines is 40nm, the beam interval is 200nm, and the number of beams in a row is 21.

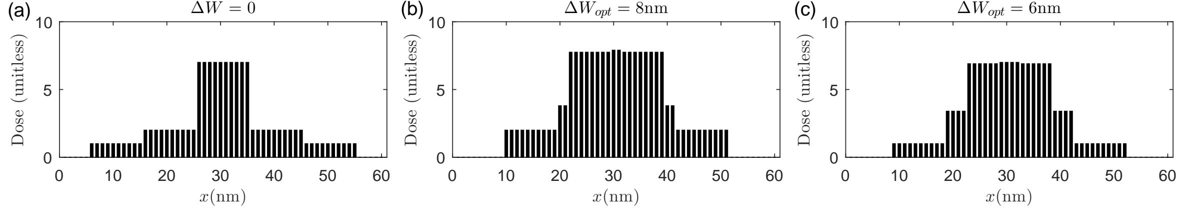

Figure 7.5. The A-type dose distribution obtained (a) without an optimization, (b) with the old method, and (c) with the new method for the case when the feature width $W = 50$nm, $\sigma_1 = 4$nm and $\sigma_2 = 9.5\mu$m.

In Table 7.2, the optimization results obtained with four different cost functions are compared between the old and new optimization methods where the weight for the exposing time is increased gradually from $C1$ to $C4$:

$$C1 = CD_error_{norm} + LER_{norm} + t_{exp_{norm}}$$

$$C2 = 0.75CD_error_{norm} + 0.75LER_{norm} + 1.5t_{exp_{norm}}$$

$$C3 = 0.5CD_error_{norm} + 0.5LER_{norm} + 2t_{exp_{norm}}$$

$$C4 = 0.25CD_error_{norm} + 0.25LER_{norm} + 2.5t_{exp_{norm}}$$

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.

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

TABLE 7.2. The optimization results with four different cost functions. The feature width is 50nm, $\sigma_1 = 4$nm, $\sigma_2 = 9.5\mu$m, $I_{bx} = 100$nm, and the dose distribution is A-type. The exposing time is calculated for a uniform pattern of 12μm width which consists of lines with a linewidth of 50nm. The space between two lines is 40nm, the beam interval is 200nm, and the number of beams in a row is 21.

7.5.2 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 12μm $\times$ 12μm were employed. In one pattern, referred to as Pattern 1, each line is 12μm long and 50nm wide, and the space between lines is 50nm. In the other pattern, referred to as Pattern 2, each line is 12μm long and 100 nm wide, and the space between lines is 100nm. Each line is partitioned into segments of 100nm and 200nm in Patterns 1 and 2, respectively, and each segment is identified with a location (m, n).

In Tables 7.3-7.6, the simulation results are compared among three methods: (a) the new optimization method for a single feature (as in Section 7.2.3) 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 Section 7.3) according to the method described in Section 7.2.2 and Section 7.2.3, 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 7.3 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 Section 7.5.1, 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 7.4 and 7.5 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 7.4 and 7.5 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 7.6. 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 over-adjustment of 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 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 7.5. The new optimization method prevents the over-adjustment of dose using the $a.f.$ and leads to a better-balanced result throughout a large pattern. This is also demonstrated in Figs. 7.6-7.8, 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 7.3-7.6 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. 7.6-7.8). 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 highest.

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

TABLE 7.3. The optimization results for a large-scale pattern with the cost function $C = CD_error_{norm} + LER_{norm} + t_{exp_{norm}}$. The beam interval is 200nm 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	CD error at corner			CD error at edge			CD error at center			LER at corner (nm)	LER at edge (nm)	LER at center (nm)
	Top (nm)	Middle (nm)	Bottom (nm)	Top (nm)	Middle (nm)	Bottom (nm)	Top (nm)	Middle (nm)	Bottom (nm)			
1(a)	+1.02	-0.18	-2.70	+1.37	+0.22	-1.12	+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

TABLE 7.4. The 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 7.3. 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.

Case	CD error at feature A			CD error at feature B			CD error at feature C			LER at	LER at	LER at
	Top	Middle	Bottom	Top	Middle	Bottom	Top	Middle	Bottom	feature A	feature B	feature C
	(nm)	(nm)	(nm)	(nm)	(nm)	(nm)	(nm)	(nm)	(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

TABLE 7.5. The CD error and LER at the top, middle and bottom layers of the resist in the test locations for the respective cases of Table 7.3. 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.

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

TABLE 7.6. The 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 7.3.

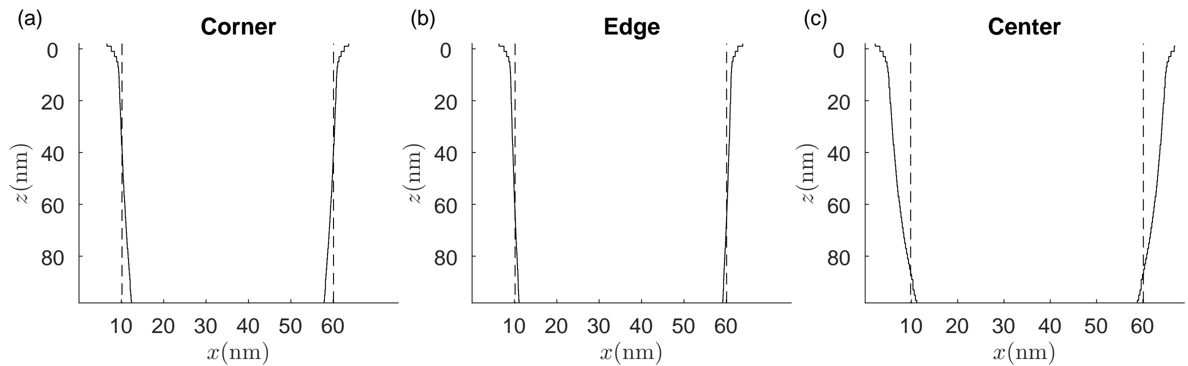

Figure 7.6. The cross-sections of resist profiles at the critical locations, i.e., (a) corner, (b) edge and (c) center, obtained with the new optimization method for a single feature directly applied to each of the features individually in a large-scale pattern (case 1(a) in Table 7.3). In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively.

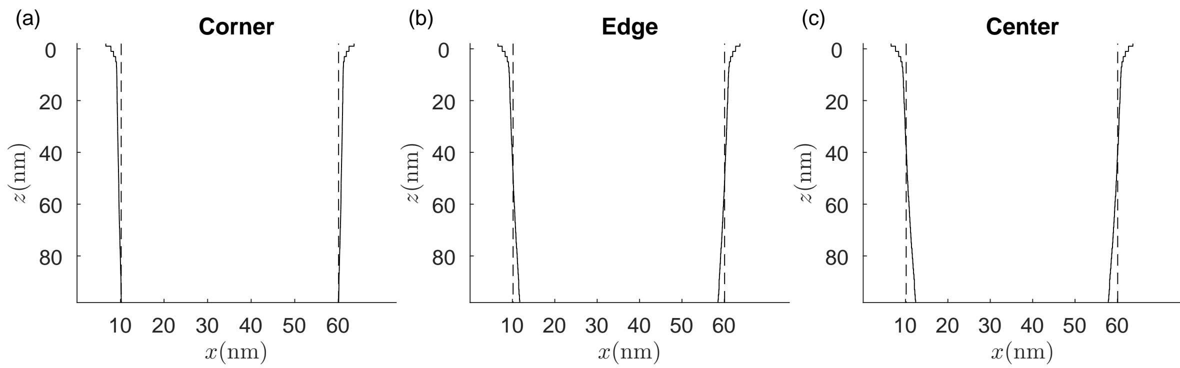

Figure 7.7. The cross-sections of resist profiles at the critical locations, i.e., (a) corner, (b) edge and (c) center, obtained with the old optimization method for a large-scale pattern (case 1(b) in Table 7.3). In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively.

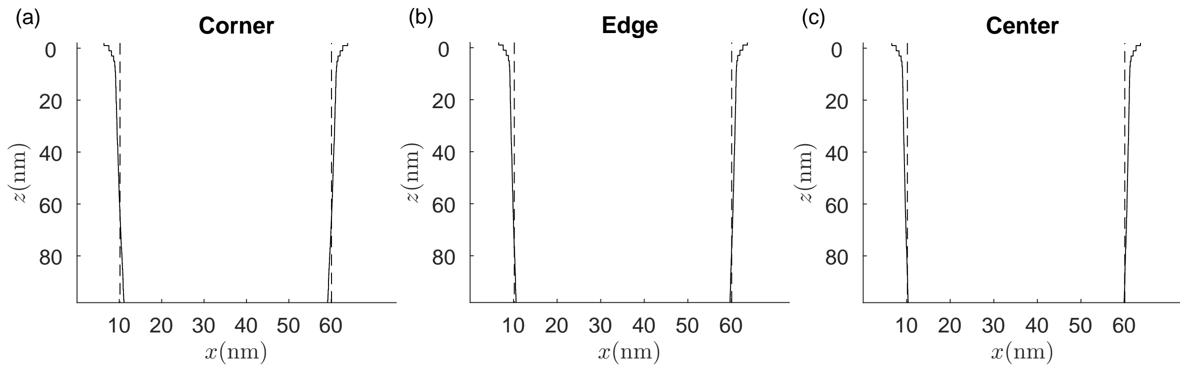

Figure 7.8. The cross-sections of resist profiles at the critical locations, i.e., (a) corner, (b) edge and (c) center, obtained with the new optimization method for a large-scale pattern (case 1(c) in Table 7.3). In each figure, the dashed and solid lines represent the target and actual resist profiles, respectively.

In Table 7.7, 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 over-adjustment of 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 over-adjustment and leads to a better-balanced result throughout the entire pattern. Table 7.7 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.

$a.f.$	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

TABLE 7.7. The 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 200nm 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.

7.6 Summary

This chapter presents an adaptive optimization method, which considers all the major 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.

Chapter 8

Concluding Remarks

The major drawback of an e-beam lithographic system has been its low throughput. Using the recently developed massively-parallel e-beam systems (MPES) could considerably improve the throughput. However, MPES systems have certain limitations. Firstly, the beams with abnormal behaviors, e.g., permanently off or on, spatial and temporal fluctuations of beam current, and beam-positioning error, may worsen the performance metrics. Secondly, a relatively large beam size, a fixed exposing interval, and the same deflection angle for all beams may cause a non-optimal shape and dose correction for proximity effect correction (PEC). Thirdly, while realizing the optimal non-uniform dose distribution for PEC, a significant number of beams (sometimes all) may be turned off in several cycles, leading to decreased beam utilization and increased exposing time. Finally, optimizing one performance metric can worsen another metric since the performance metrics are correlated, e.g., reducing the CD error and LER with PEC may worsen the exposing time. Chapters 3-7 in this dissertation address these limitations, improve performance metrics individually and optimize multiple performance metrics simultaneously for the efficient use of MPES.

The outcomes of this research for the dissertation are summarized below.

Chapter 3: The negative effects of abnormal beams on performance metrics, such as the exposure variation, LER, maximum indent, and maximum outdent, have been analyzed for three writing methods, single-row writing I, single-row writing II, and multi-row writing. Results from an extensive simulation show that the multi-row writing method always performs better than the single-row writing methods I and II.

Chapter 4: Various combinations of single-row and multi-row writing methods along with multi-pass writing have been compared in terms of mitigating the negative effect of abnormal beams. The notable results include: (a) the optimal number of passes and positional shift for the MP writing, (b) the optimal writing pattern for the MRW, (c) the optimal SRW and MRW methods achieve significant performance improvements, (d) the MRW (without MP) performs better than the SRW-MP's, and (e) the MRW-MP is superior to all other methods considered in this study.

Chapter 5: A practical method of realizing four types of dose distributions, i.e., uniform, V-type, A-type, and M-type, with varying linewidth reduction (ΔW) under the constraints of the MPES has been developed and is applicable to any spatial-dose-distribution ratio. Also, a shape+dose correction procedure that uses the spatial dose distribution (with a width reduction) realizable on an MPES has been designed and implemented. Through an extensive simulation, it has been shown that the shape+dose correction procedure can optimally reduce the CD error and LER while maintaining a relatively low total dose.

Chapter 6: Two methods have been investigated for increasing beam utilization and reducing exposing time: a writing method utilizing empty cycles (the cycles with all beams turned off) and an optimization method lowering the dose difference among the regions of a feature while implementing PEC. The writing method adjusts the deflection angle of beams to remove empty cycles and shorten exposing time, particularly in cases where the beam interval is equal to the feature width and space. The other method reduces the dose difference in the optimal dose distribution for PEC through an iterative process, enhancing beam utilization with minimal impact on CD accuracy. These two methods can also be combined for further reducing the exposing time.

Chapter 7: An adaptive optimization method has been developed, which considers all the performance metrics in a cost function, i.e., CD error, LER, and exposing time, in each iteration. The method is shown to outperform the optimization method that considers the performance metrics in a fixed order in minimizing the cost function for both a single feature and a large-scale pattern with uniform features and obtaining overall balanced results among all the

performance metrics. Also, it is adaptable to different weights of performance metrics in the cost function.

As summarized above, this dissertation makes several key contributions to the field of MPES. Firstly, it offers a set of general procedures to find the optimal MP writing and MRW, which can help reduce the effects of abnormal beams and design more effective writing methods. Also, it describes a novel method of spatial-dose realization and a procedure of shape+dose correction, which can be useful for correcting the proximity effect for MPES. Furthermore, it presents new methods for reducing exposing time and an adaptive optimization method that simultaneously optimizes multiple key performance metrics. These results may be referred to for optimizing the performance of MPES.

Future research could build upon these contributions by extending the application of the developed optimization method to large-scale patterns with non-uniform features. Another interesting future work would be to consider the beam abnormalities in the simulation to analyze the performance of the optimization method with different writing methods. Also, processing the large amount of data involved with exposing very large patterns using nearly all the beams in the MPES active is an important area for further investigation.

References

[1] S.-Y. Lee and B. D. Cook, IEEE Trans. Semicond. Manuf. **11**, 117 (1998).

[2] S.J. Wind, P.D. Greber, H. Rothuizen, J. Vac. Sci. Technol. B **16**, 3262 (1998).

[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 and E. Platzgummer, Proceedings Volume 9985, Photomask Technology 2016; 998505 (2016).

[7] C. Klein, H. Loeschner, and E. Platzgummer, J. Micro/Nanolith. MEMS MOEMS. **11**(3), 031402 (2012).

[8] B. J. Lin, J. Micro Nanolithogr. MEMS MOEMS **11**, 033011 (2012).

[9] H. Matsumoto, H. Inoue, H. Yamashita, H. Morita, S. Hirose, M. Ogasawara, H. Yamada, and K. Hattori, BACUS News **32**, 1 (2016).

[10] P. Petric, C. Bevis, A. Carroll, H. Percy, M. Zywno, K. Standiford, A. Brodie, N. Bareket, and L. Grella, J. Vac. Sci. Technol. B **27**, 161 (2009).

[11] A. Fay, C. Browning, P. Brandt, J. Chartoire, S. B.-Bergery, J. Hazart, A. Chagoya, S. Postnikov, M. Saib, L. Lattard, P. Schavione, in Proc. SPIE, 977714 (2016).

[12] Soomin Moon, S.-Y. Lee, J. Choi, S.-B. Kim, and C.-U. Jeon, J. Vac. Sci. Technol. B **36**, 06JA03 (2018).

[13] H. Fragner and E. Platzgummer, U.S. Patent No. 7777201 B2, 17 Aug (2010).

[14] E. Platzgummer, C. Klein, and H. Loeschner, J. Micro/Nanolith. MEMS MOEMS. **12**(3), 031108 (2013).

[15] S.-Y. Lee, B.-S. Ahn, J. Choi, S.-B. Kim and C.-U. Jeon, J. Vac. Sci. Technol. B **37**, 061602 (2019).

[16] R. Murali, D. K. Brown, K. P. Martin, and J. D. Meindl, J. Vac. Sci. Technol. B **24**, 2936 (2006).

[17] K. Ogino, H. Hoshino, and Y. Machida, J. Vac. Sci. Technol. B **26**, 2032 (2008).

[18] N. Unal, D. Mahalu, O. Raslin, D. Ritter, C. Sambale, and U. Hofmann, Microelectron. Eng. **87**, 940 (2010).

[19] Q. Dai, S.-Y. Lee, S.-H. Lee, B.-G. Kim, and H.-K. Cho, J. Vac. Sci. Technol. B **30**, 06F307 (2012).

[20] 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).

[21] 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).

[22] M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, and J.-S. Park, J. Vac. Sci. Technol. B **38**, 062601 (2020).

[23] M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, J.-S. Park, J. Vac. Sci. Technol. B **38**, 062603 (2020).

[24] M. N. Hasan, S.-Y. Lee, B.-S. Ahn, J. Choi, and J.-S. Park, J. Vac. Sci. Technol. B **40**, 032602 (2022).

[25] M. N. Hasan and S.-Y. Lee, J. Vac. Sci. Technol. B (2023) [Manuscript submitted for publication].

[26] H. Matsumoto, H. Yamashita, T. Tamura, and K. Ohtoshi, Proc. SPIE **10454**, 10454E (2017).

[27] Q. Dai, R. Guo, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, and C.-U. Jeon, Microelectronic Engineering, **127**, 86 (2014).

[28] T. Kamikubo, K. Ohtoshi, S. Golladay. V. Katsap, R. Kendall, H. Sunaoshi, and S. Tamamushi, Physics Procedia, 119 (2008).

[29] D. Drouin, A. R. Couture, D. Joly, X. Tastet, V. Aimez, and R. Gauvin, Scanning **29**, 92 (2007).

[30] H. Ji, S.-Y. Lee, J. Choi, S.-B. Kim, I.-K. Shin, and C.-U. Jeon, J. Vac. Sci. Technol. B **35**, 06G503 (2017).

[31] E. Seo, B. K. Choi, and O. Kim, Microelectron. Eng. **53**, 305-308 (2000).

[32] Q. Dai, S.-Y. Lee, S.-H. Lee, B.-G. Kim, and H.-K. Cho, J. Vac. Sci. Technol. B **29**, 06F314 (2011).

[33] L. D. Jackel, R. E. Howard, P. M. Mankiewich, H. G. Craighead, and R. W. Epworth, Appl. Phys. Lett. **45**, 698 (1984).

[34] 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).