

Modeling and Optimization of Electron-Beam Lithographic Process

by

Dehua Li

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
August 7, 2021

Keywords: Electron-Beam Lithography, SEM Image Processing, Modeling and Optimization,
Neural Network, Estimation of Critical Dimension and LER

Copyright 2021 by Dehua Li

Approved by

Soo-Young Lee, Chair, Professor of Electrical and Computer Engineering
Stanley Reeves, Professor of Electrical and Computer Engineering
Guofu Niu, Professor of Electrical and Computer Engineering
Masoud Mahjouri-Samani, Assistant Professor of Electrical and Computer Engineering

Abstract

An important issue in electron-beam (e-beam) lithography is to minimize the critical dimension (CD) error and line edge roughness (LER). This dissertation presents a method to measure the CD and LER from scanning electron microscope (SEM) images, two methods to estimate the CD and LER, i.e., modeling the e-beam lithographic process and utilizing a neural network (NN), and the minimization results obtained based on the estimation. The 3-D modeling is also applied to SEM images from the massively-parallel e-beam system (MPES).

A common approach to the measurement of CD and LER from SEM images is to employ image processing techniques to detect feature boundaries from which the CD and LER are computed. SEM images usually contain a significant level of noise which affects the accuracy of measured CD and LER. This requires reducing the noise level by a certain type of low-pass filter before detecting feature boundaries. However, a low-pass filter also tends to destroy the boundary detail. Therefore, a careful selection of low-pass filter is necessary in order to achieve the high accuracy of CD and LER measurements. In this dissertation, a practical method to design a Gaussian filter for reducing the noise level in SEM images is developed. The method utilizes the information extracted from a given SEM image in adaptively determining the sharpness and size of a Gaussian filter.

Computational lithography is typically based on a model representing the lithographic process where a typical model consists of three components, i.e., line spread function (LSF), conversion formula (exposure-to-developing rate conversion), and noise process (exposure fluctuation). In a previous study, a practical approach to modeling the e-beam lithographic process by deriving the three components directly from SEM images was proposed. However, a 2-D model of a substrate system was employed, i.e., the exposure variation along the resist-depth dimension was not considered. In this dissertation, the possibility of improving the accuracy of modeling using a 3-D model is investigated. The 3-D model is iteratively determined by

modeling the estimated CD based on the model to those measured in SEM images. This dissertation describes the 3-D modeling approach and new optimization procedures, and discusses in detail the results from an extensive simulation for an accuracy analysis of the 3-D modeling approach.

The 3-D modeling involves several parameters to be determined and tends to require a long computation time. The possibility of improving the accuracy and efficiency of the CD and LER estimation using a NN is investigated. In the NN-based estimation, the explicit modeling of the e-beam lithographic process can be avoided. This dissertation describes the method of estimating the CD and LER using a NN, including the issues of training, tuning, and sample reduction, and presents results obtained through an extensive simulation.

The accuracy of the 3-D modeling is further verified through the proximity effect correction. A dose modification with the reduction of the feature width is performed utilizing the modeled LSF and noise. The CD error and LER are considered for the performance analysis.

The 3-D modeling is extended and applied to SEM images from the MPES deriving the PSF and the noise. The accuracy is verified by estimating the CD and LER for different doses.

Acknowledgments

First of all, the gratitude is extended to my major advisor Dr. Soo-Young Lee for guiding me into the research area of scanning electron microscope image processing, modeling of the electron-beam lithographic process, as well as the proximity effect correction in the e-beam lithography. My researches cannot be finished without the academic suggestions or the financial support from Dr. Lee in my doctoral program. His patience and precious time spent on my researches and dissertation are deeply appreciated. And also, his rigorous attitude to academic problems is an example which will be followed.

My appreciation also goes to Dr. Stanley Reeves, Dr. Guofu Niu, and Dr. Masoud Mahjouri-Samani for serving on my committee, and Dr. Ziqin Feng being my University Reader. Their precious comments on my researches are helpful and valued.

I would like to extend my gratitude to Dr. Thaddeus Roppel for his kind help at the initial period of my doctoral program.

The members in our research group, including Rui Guo, Hyesung Ji, Soomin Moon, and Nabid Hasan, are appreciated for the collaboration. I would like to thank my friends in Auburn, especially Yingsong Han, Xue Xia, Mi Yan, Yuan Bian, Huanyi Zhou, for the wonderful time. Thanks are also to Samsung Electronics Co., Ltd for the researching funding.

Finally, I am deeply indebted to my family for understanding and encouragement when I pursued a Ph.D degree in Auburn University.

Table of Contents

Abstract	ii
Acknowledgments	iv
List of Figures	viii
List of Tables	xi
List of Abbreviations	xiii
1 Introduction	1
1.1 Problem Definition	1
1.2 Review	2
1.3 Motivations and Objectives	4
1.4 Approach	5
1.5 Organization of Dissertation	6
2 Design of a Noise Filter for SEM Images	7
2.1 Noise Filtering	7
2.2 Filter Design	9
2.3 Detection of Feature Boundaries	14
2.4 Results and Discussion	16
2.5 Summary	21
3 3-D Modeling of Electron-Beam Lithographic Process from SEM Images	22

3.1	2-D Modeling Approach	22
3.2	3-D Modeling Approach	23
3.2.1	Overview	24
3.2.2	Exposure Distribution	25
3.2.3	Noise Model	26
3.2.4	LSF Model	27
3.2.5	Cost Function	28
3.2.6	Simulated Annealing	30
3.2.7	Initialization	30
3.2.8	Perturbation	31
3.3	Results and Discussion	32
3.3.1	Simulation	32
3.3.2	Modeling	32
3.4	Summary	33
4	Accurate Estimation of Critical Dimension and Line Edge Roughness Using a Neural Network	38
4.1	E-Beam Lithographic Process	38
4.2	Estimation of CD and LER	40
4.2.1	Structure of NN	40
4.2.2	Training with E-Beam Lithographic Samples	41
4.2.3	Estimation	45
4.3	Results and Discussion	45
4.3.1	Reference Remaining Resist Profiles	46
4.3.2	Training and Estimation Errors	46
4.3.3	Tuning of NN	47
4.3.4	Estimation of CD and LER	49

4.4	Reduction of Training Samples	50
4.5	Summary	51
5	Verification: Proximity Effect Correction	54
5.1	Width Reduction	54
5.2	Dose Correction	56
5.3	Results and Discussion	56
5.4	Summary	57
6	Application to Massively-Parallel E-Beam System	60
6.1	Massively-Parallel E-Beam System	60
6.2	Modifications of Modeling	61
6.2.1	PSF and TF	61
6.2.2	Noise Model	62
6.2.3	Cost Function and Constraints	63
6.3	Results and Discussion	64
6.3.1	Modeling	64
6.3.2	Estimation	66
6.4	Summary	68
7	Concluding Remarks	69
	References	71

List of Figures

2.1	An example of SEM image of a L/S pattern and (b) brightness distribution along the line A-A' in the SEM image in (a).	8
2.2	Edge detection results following (a) under-filtering and (b) over-filtering of noise.	8
2.3	A 2-D Gaussian filter of size $W_s \times W_s$ with the standard deviation of σ	9
2.4	(a) SEM image $i(x, y)$ of size 1024×1024 and (b) the frequency-domain representation $ I(u, v) $ plotted in the logarithmic scale.	10
2.5	(a) Generation of sampled noise $n(x, y)$ (before the DC is removed) and (b) the frequency-domain representation $ N(u, v) $ in the logarithmic scale.	11
2.6	The total signal and noise intensities as functions of the cut-off frequency σ_f of the Gaussian filter. In designing the Gaussian filter, σ_f at which the two intensities become equal (marked by “+”) is used.	13
2.7	Each boundary of a feature (line) is characterized by a bright band, and both edges of the band region are detected resulting in inner and outer edges.	14
2.8	Maximum and minimum brightness gradients correspond to (inner and outer) edge locations.	15
2.9	Edge regions are defined around peaks and the maximum and minimum gradients are found within edge regions.	16
2.10	An example of reference SEM image where the same noise level as in the corresponding original SEM image is added.	17
2.11	(a) An example of real SEM image and (b) Feature boundaries detected in SEM image in (a).	20
3.1	(a) An example of an SEM image in a L/S pattern and (b) detected boundaries overlapped on an SEM image in (a). The inner and outer edges are used to calculate the CD and LER.	23
3.2	The resist model includes 5 resist layers, where the middle layer of the resist is taken as the reference layer for the purpose of the 3-D modeling.	24

3.3	Overall flow of the modeling approach: the modeling is guided by simulated annealing.	25
3.4	An example of average exposure distribution along the y direction at different layers from the PMMA/Si system: linewidth of 120 nm, resist thickness of 300 nm and beam energy of 10 keV. The average exposure varies along resist-depth dimension.	26
3.5	An example of layer-dependency ratio: resist thickness of 200 nm, beam energy of 30 keV in PMMA/Si system.	27
3.6	Illustration of LSF adjustment: In an iteration, the length, location and scaling of the perturbed section in the main lobe are randomly determined.	29
3.7	The linewidth matching guided by SA is employed in the modeling process to minimize the cost function. In each iteration, the perturbed LSF and noise parameters are utilized going through the exposure and resist-development processes in order to compute the cost function.	31
3.8	The probability of acceptance decreases as the number of outer-loop iterations increases (corresponding to the decrease of annealing temperature) for $\Delta\epsilon > 0$	32
3.9	The deterministic LSFs from the 3-D modeling procedures for varying beam energy from 10 to 50 keV and resist thickness of a) 100 nm; b) 200 nm; c) 300 nm; d) 400 nm and e) 500 nm.	34
4.1	Contours of the remaining resist profiles at the top and bottom layers: substrate system of PMMA on Si, resist thickness of 100 nm, beam energy of 10 keV, beam diameter of 10 nm, and normalized dose of 0.335.	39
4.2	The dependency of (a) CD and (b) LER on the dose for varying linewidth when the development time is fixed: substrate system of PMMA on Si, resist thickness of 100 nm, beam energy of 10 keV, beam diameter of 10 nm, normalized dose of 0.335, exposure interval of 1 nm, and development time of 0.5 min.	40
4.3	The illustration of a neural network with an input layer of 5 inputs (resist thickness, beam energy, beam diameter, linewidth and normalized dose), N_h hidden layers with N_n neurons each, and an output layer of 2 outputs (CD and LER).	41
4.4	The NN training process employs the EBP algorithm using a mini-batch. Each mini-batch of samples is selected randomly in an iteration. The weights of nodes are updated after the training on a mini-batch of samples is finished.	42
4.5	Process for estimating CD and LER with one forward calculation using the trained NN.	45
4.6	LER estimation errors in the training process with varying (a) mini-batch size, (b) learning rate, (c) number of nodes in a hidden layer, and (d) number of nodes in a hidden layer.	48

4.7	CD and LER estimation errors with a reduced number of training samples for (a) linewidth, (b) normalized dose, (c) beam diameter, (d) resist thickness, and (e) beam energy.	52
5.1	The feature width is reduced by ΔW each side and the dose is determined for the reduced width.	54
5.2	The optimal dose is determined when the CD error (calculated with the modeled LSF, noise parameters, and the reduced feature linewidth to be exposed) is below a threshold, which is set to a small value.	55
6.1	Principle of a massively-parallel e-beam system with an APS. Each of 512×512 apertures controls the availability of electrons within a certain size, $10 \text{ nm} \times 10 \text{ nm}$ or $20 \text{ nm} \times 20 \text{ nm}$, onto the resist.	60
6.2	SEM images from the MPES with normalized dose of (a) 0.825, (b) 0.998, and (c) 1.247.	65
6.3	SEM images from the MPES with normalized dose of (a) 0.901, (b) 0.945, and (c) 1.113.	67

List of Tables

2.1	The known CD and LER in the four reference images generated from SEM images obtained with four different doses where the target CD is 120 nm. . . .	17
2.2	The CD and LER measured from the reference image 1, and the deviations from the known CD and LER, are provided with the SNR level varied from 13.2 dB to 27.9 dB. The filter parameters (W_s and σ) are also included in each case. . .	18
2.3	The CD and LER measured from the reference image 2, and the deviations from the known CD and LER, are provided with the SNR level varied from 13.2 dB to 28.0 dB. The filter parameters (W_s and σ) are also included in each case. . .	18
2.4	The CD and LER measured from the reference image 3, and the deviations from the known CD and LER, are provided with the SNR level varied from 13.8 dB to 28.5 dB. The filter parameters (W_s and σ) are also included in each case. . .	18
2.5	The CD and LER measured from the reference image 4, and the deviations from the known CD and LER, are provided with the SNR level varied from 14.3 dB to 28.8 dB. The filter parameters (W_s and σ) are also included in each case. . .	18
2.6	The filter parameters, and measured CD and LER for the SEM images obtained with different normalized doses, 0.705, 0.774, 0.845 and 1.082.	20
3.1	The results from 2-D and 3-D modeling with resist thickness varied from 100 to 300 nm, and beam energy from 10 to 50 keV.	35
3.2	The results from 2-D and 3-D modeling with resist thickness varied from 400 to 500 nm, and beam energy from 10 to 50 keV.	36
4.1	CD and LER estimation errors in nm and percentage for varying resist thickness and beam energy for the substrate system of PMMA on Si: Errors are averaged over the beam diameters of 2, 5, 10, 15 and 20 nm, linewidths of 60, 80, 100, 120 and 140 nm, and normalized doses.	50
5.1	Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 100 nm, and the target CDs of 40 nm and 70 nm.	57
5.2	Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 200 nm, and the target CDs of 40 nm and 70 nm.	58

5.3	Estimation of the CD and LER through the PEC in terms of the 2-D and 3-D modeling results for the resist thickness of 300 nm, and the target CDs of 40 nm and 70 nm. “NA” indicates that the resist is not fully developed.	58
5.4	Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 400 nm, and the target CDs of 40 nm and 70 nm. “NA” indicates that the resist is not fully developed.	59
5.5	Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 500 nm, and the target CDs of 40 nm and 70 nm. “NA” indicates that the resist is not fully developed.	59
6.1	CD and LER measured in SEM images from the MPES for modeling with varied normalized doses.	65
6.2	Modeled CD and LER according to the information in SEM images from a MPES.	66
6.3	CD and LER measured in SEM images from the MPES for estimation.	66
6.4	CD and LER estimated for different normalized doses.	67

List of Abbreviations

APS	Aperture Plate System
CD	Critical Dimension
E-Beam	Electron Beam
EBP	Error Back-Propagation
L/S	Line/Space
LER	Line Edge Roughness
LR	Layer-Dependency Ratio
LSF	Line Spread Function
MPES	Massively-Parallel E-Beam System
NN	Neural Network
PEC	Proximity Effect Correction
PMMA	Polymethyl Methacrylate
PSF	Point Spread Function
SA	Simulated Annealing
SEM	Scanning Electron Microscope
SNR	Signal-to-Noise Ratio
TF	Transfer Function

Chapter 1

Introduction

Electron beam (e-beam) lithography is a technique widely employed in pattern transfer onto resist [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. However, two of the main issues which limit the minimum feature size and maximum feature density achievable by e-beam lithography are the proximity effect and line edge roughness (LER) [11, 12, 13, 14, 15]. The minimization of proximity effect and LER often requires an accurate estimation of critical dimension (CD) and LER. The focus of this dissertation is on the estimation of CD and LER using a computational method of e-beam lithographic process, derived from experimental results. The CD and LER are to be measured from scanning electron microscope (SEM) images and are used to model the e-beam lithographic procedure.

1.1 Problem Definition

Design of a noise filter for SEM images

The fidelity of transferred patterns is often examined through SEM imaging [16, 17]. Feature boundaries in SEM images are determined, from which metrics such as CD and LER are computed [18, 19]. As the feature size continues to decrease, it has become even more important to be able to measure the CD and LER accurately from SEM images [20, 21]. A small error in determining feature boundaries can lead to a significant deviation of the CD and LER from the actual values.

A typical approach to determining feature boundaries in an SEM image is to employ image processing techniques such as edge detection or modeling. What makes the image processing non-trivial is the noise contained in SEM images [22]. A standard procedure is to apply a noise-reduction filter, e.g., low-pass filter, to SEM images before further processing. It is common

that the noise level is significant and varies with SEM image. Also, the type of noise may be different for a different image or a different location within an image. Therefore, it is not optimal to use a fixed filter in general. In this dissertation, the issue of designing a noise-reduction filter, taking the characteristics of noise into account, is addressed.

3-D modeling of electron-beam lithographic process from SEM images

Various methods have been developed in order to correct the proximity effect and reduce the LER, employing analytic, simulation or experimental approaches [23, 24, 25, 26]. In the computational lithography, such efforts are based on a model representing the lithographic process where a typical model consists of three components, i.e., the point spread function (PSF, equivalent to the line spread function, LSF), conversion formula (exposure-to-developing rate conversion), and noise process (exposure fluctuation). This model may be derived from SEM images.

Accurate estimation of critical dimension and line edge roughness using a neural network

An essential step in the computational lithography is to estimate the CD and LER. One may utilize an e-beam lithographic model to do the estimation through the exposure computation and resist-development simulation. To avoid the need for an explicit model of the e-beam lithographic process, one may employ a neural network (NN) for the task of estimating the CD and LER.

Application to massively-parallel e-beam system

The massively-parallel e-beam system (MPES), like the single beam system, suffers from the proximity effect because of electrons scattering in the resist. An effective method is desired to correct the proximity effect utilizing the information extracted from the SEM images obtained from the MPES.

1.2 Review

Design of a noise filter for SEM images

In the previous work [27], a fixed filter was employed to reduce the noise level in processing SEM images. That is, the image-dependent noise characteristics such as the noise level were not considered. Therefore, the subsequent detection of feature boundaries may not be always sufficiently accurate. Another method to detect feature boundaries with or without noise filtering is to rely on edge modeling, i.e., fitting the brightness distribution over a feature boundary to a model function. The effectiveness of this approach was well demonstrated using the Gaussian function as a model function through an elaborated investigation [28]. Nevertheless, it is worthwhile to point out the sensitivity of the method to the shape of brightness distribution. When the bright distribution has a significantly different shape from the model function, this can lead to a substantial error in the detected feature boundaries. Also, in their study, it appears that only the “simple noise” was considered in the accuracy analysis.

3-D modeling of electron-beam lithographic process from SEM images

In a study of Lee et al. [29], a practical approach to modeling the e-beam lithographic process by deriving the three components directly from SEM images was proposed. Based on the critical information extracted from SEM images, the modeling process derives the three components through iterations. However, a 2-D model was employed, i.e., the exposure variation along the resist-depth dimension was not considered. Since the dependency of the exposure energy deposited on the resist layer is ignored, such modeling approach is likely to lead to inaccurate results, especially for a thick resist, or a low beam energy.

Accurate estimation of critical dimension and line edge roughness using a neural network

In a previous study [29], a 2-D modeling approach was proposed to model the e-beam lithographic process with three components, i.e., LSF, conversion formula and noise process.

The CD and LER measured from SEM images are utilized in this approach, minimizing the modeling error through iterations. Then, the process model is employed in estimating the CD and LER for a given pattern. However, this modeling approach involves several parameters to derive (model) and tends to require a long computation time. Also, the estimation errors

were substantial in some cases, e.g., thick resist and low beam energy.

Application to massively-parallel e-beam system

Massively-parallel e-beam systems were developed recently to improve the writing throughput by using a large number of beams [30, 31, 32, 33]. But, no modeling of the e-beam lithographic process using SEM images obtained from the MPES was attempted.

1.3 Motivations and Objectives

Design of a noise filter for SEM images

For the accurate measurement of CD and LER, a noise filter which takes into account the SEM-image-dependent characteristics of noise is desired. A practical method to design such a filter is to be developed.

3-D modeling of electron-beam lithographic process from SEM images

The previous 2-D modeling approach does not consider the exposure variation along the depth dimension of resist and therefore may lead to an inaccurate estimation of CD and LER. A modeling method considering the layer-dependent exposure needs to be developed.

Accurate estimation of critical dimension and line edge roughness using a neural network

The modeling of the e-beam lithographic process and subsequent estimation of CD and LER require an explicit model and are computationally expensive. A neural network can be employed for the estimation, which does not involve the explicit model and is faster once trained.

Application to massively-parallel e-beam system

In order to further verify the accuracy of the 3-D modeling of the e-beam lithographic process, the SEM images from a MPES may be employed for the modeling and estimation. The CD error in the estimation can be used to evaluate the modeling accuracy.

1.4 Approach

Design of a noise filter for SEM images

A practical method for designing a Gaussian filter adaptive to the noise characteristics of each SEM image is described. In the method, the noise characteristics are extracted from the frequency-domain representation of SEM image, and the sharpness (standard deviation) and size of Gaussian filter are determined based on a certain measure derived from the characteristics.

3-D modeling of electron-beam lithographic process from SEM images

This study attempts to improve the modeling accuracy by taking into account the exposure dependency on the resist layer, i.e., a 3-D modeling approach. A 3-D model enables a more realistic representation of exposure distribution in the resist and in turn allows one to obtain a more accurate remaining resist profile through the resist-development simulation [34]. Therefore, it is more likely to derive a more realistic model of e-beam lithographic process by the 3-D approach. Also, the modeling procedures are simplified and optimized in order to reduce the computational complexity without sacrificing the accuracy of modeling.

Accurate estimation of critical dimension and line edge roughness using a neural network

The possibility of improving the estimation accuracy of CD and LER in the e-beam lithographic process using a neural network is investigated. The use of a NN for the estimation of CD and LER makes unnecessary an explicit modeling of the e-beam lithographic process. A NN is trained by a number of samples obtained from the e-beam lithographic process and the trained NN is used in estimating the CD and LER, given a set of the e-beam lithographic parameters, i.e., the resist thickness, beam energy, beam diameter, linewidth and normalized dose.

Application to massively-parallel e-beam system

The 3-D modeling is tested with SEM images obtained from a MPES. The modeling procedures are adjusted to improve the convergence and accuracy.

1.5 Organization of Dissertation

The rest of this dissertation is organized as follows:

- Chapter 2 describes a practical method for designing a Gaussian filter adaptive to the noise characteristics of each SEM image, to obtain the accurate measurement of CD and LER from SEM images.
- Chapter 3 describes the 3-D modeling procedures guided by the simulated annealing (SA) to derive the components of the e-beam lithographic process and to obtain a more accurate remaining resist profile.
- Chapter 4 describes a neural network approach to estimating the CD and LER based on the results from the e-beam lithographic process.
- Chapter 5 describes the verification through the proximity effect correction (PEC) using the modeling results to obtain the minimized CD error and LER.
- Chapter 6 describes the application of the 3-D modeling approach to the MPES, including the modeling and estimation using the SEM images.
- Chapter 7 provides concluding remarks of this dissertation.

Chapter 2

Design of a Noise Filter for SEM Images

In this chapter, a practical method for designing a Gaussian filter adaptive to the noise characteristics of each SEM image is described. In the method, the noise characteristics are extracted from the frequency-domain representation of an SEM image, and the sharpness (standard deviation) and size of Gaussian filter are determined based on a certain measure derived from the characteristics.

2.1 Noise Filtering

SEM images inherently contain a significant level of noise as shown in Fig. 2.1. Therefore, it is unavoidable to reduce the noise level before any image processing technique for detecting feature boundaries can be applied. The noise filtering greatly affects the accuracy of detected feature boundaries. That is, an inappropriate level of noise filtering can cause a significant error in the measured CD and LER as illustrated in Fig. 2.2 where a simulated SEM image with the known CD and LER is employed. An under-filtering leads to a rougher detected boundary than the actual one such that the measured LER, for example, is 31% larger than the actual LER (see Fig. 2.2(a)). On the other hand, an over-filtering makes the detected boundary smoother than the actual, leading to a negative error (-24%) in the measured LER (see Fig. 2.2(b)) for a line/space (L/S) pattern.

A 2-D Gaussian filter is selected for noise filtering in this study, for its applicability and controllability. The level of noise filtering can be adjusted by controlling the sharpness and size of a Gaussian filter. In the 2-D Gaussian function given in Eq. 2.1 and illustrated in Fig. 2.3, the sharpness is determined by the standard deviation σ . The larger the σ is, the higher level of noise filtering is achieved, i.e., the noise is filtered more, but more blurring occurs. On the

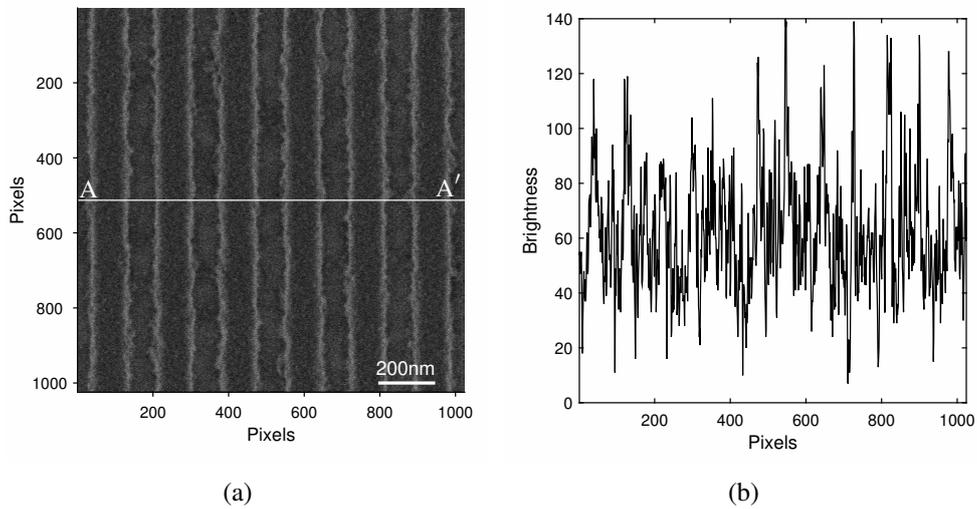


Figure 2.1: An example of SEM image of a L/S pattern and (b) brightness distribution along the line A-A' in the SEM image in (a).

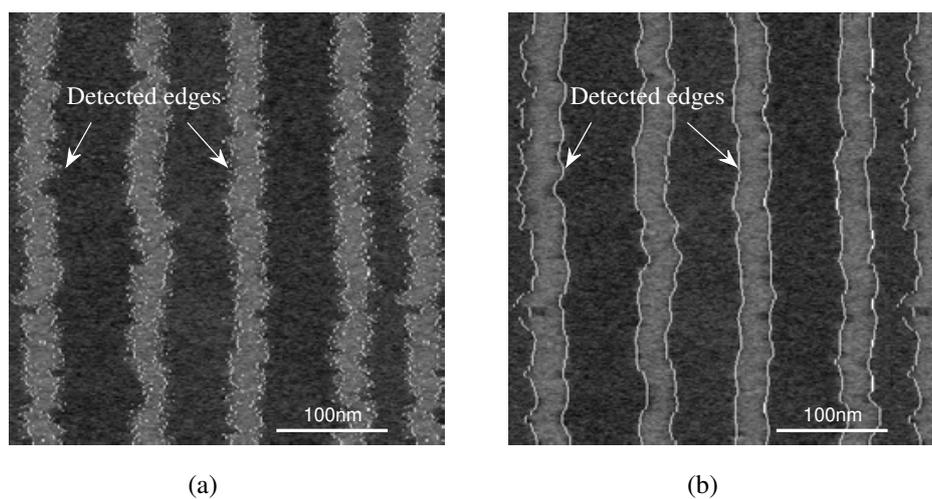


Figure 2.2: Edge detection results following (a) under-filtering and (b) over-filtering of noise.

other hand, for a smaller σ , the noise is filtered less. Given a σ , the filter size W_s (see Fig. 2.3) determines the domain over which the filtering is carried out for each pixel. A larger W_s tends to reduce the noise level more, but also blur the image more.

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2.1)$$

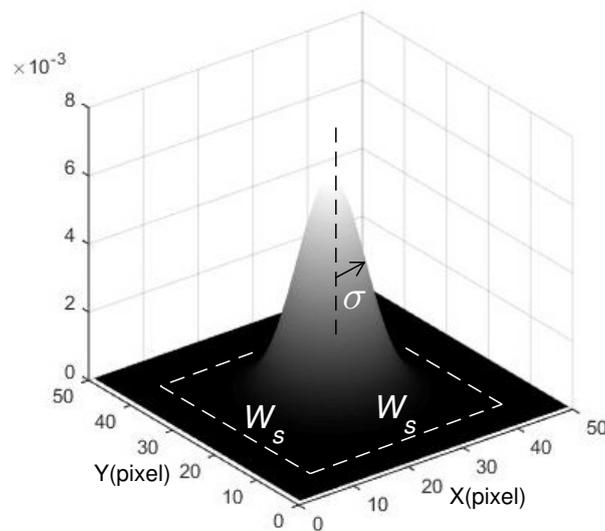


Figure 2.3: A 2-D Gaussian filter of size $W_s \times W_s$ with the standard deviation of σ .

The objective of the proposed method for designing the Gaussian filter is to determine the filter parameters, σ and W_s , which minimize the differences between the measured and actual CDs and LERs. The determination of the two parameters is guided by the characteristics of a given SEM image, in particular, the signal and noise characteristics.

2.2 Filter Design

The objective of noise filtering is to reduce the noise as much as possible while keeping the signal (feature boundaries) as much as possible. However, in general, reducing the noise also reduces the signal, and therefore the level of noise filtering must be properly selected such that the accuracy of boundary detection is sufficiently high. In the case of a Gaussian filter, the level of noise filtering can be controlled by the two parameters of filter, σ and W_s . In particular,

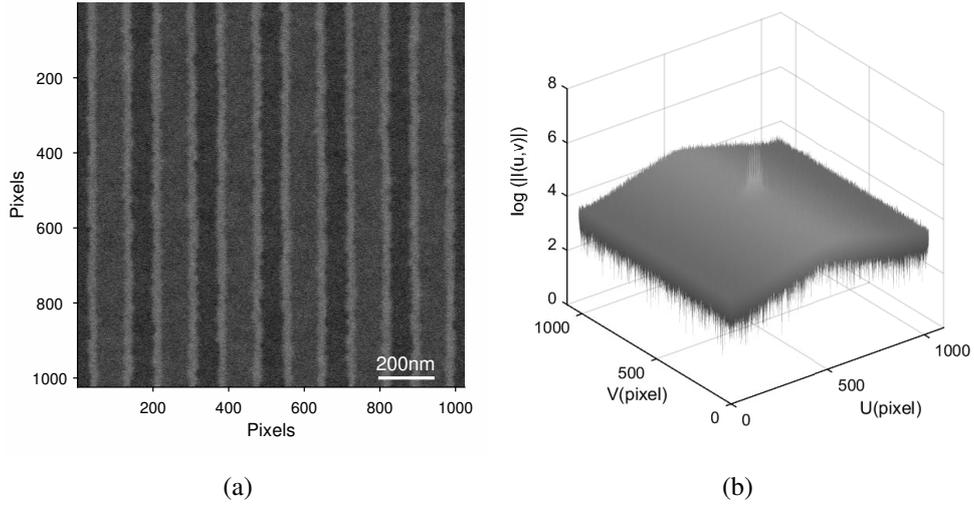


Figure 2.4: (a) SEM image $i(x, y)$ of size 1024×1024 and (b) the frequency-domain representation $|I(u, v)|$ plotted in the logarithmic scale.

the proposed design method determines σ and W_s using a metric which quantifies the relative signal level after filtering. For evaluating the metric, the frequency-domain representations of signal and noise are employed. However, the signal and noise are not readily separable in an image. Therefore, given an SEM image, they are estimated in the frequency domain.

Let $i(x, y)$ represent the brightness distribution in an SEM image of size $M \times M$. The frequency-domain representation, $I(u, v)$, is obtained through the 2-D discrete Fourier transform (just “Fourier transform” hereafter) as in Eq. 2.2.

$$I(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} i(x, y) e^{-j \frac{2\pi(u x + v y)}{M}} \quad (2.2)$$

An SEM image and the corresponding Fourier transform are shown in Fig. 2.4. The $|I(u, v)|$ shows different behaviors along the u and v dimensions. This is due to the fact that lines are vertically oriented and the noise is more spatially-correlated in the x direction than in the y direction. It should be noted that $I(u, v)$ includes both the signal and noise. In order to estimate the noise, “flat regions” (i.e., excluding boundary regions) are extracted from the SEM image, to generate an image of “sampled” noise as shown in Fig. 2.5(a). The sampled noise contains the noise and DC biases where the DC biases come from the different brightness levels of the inside and outside of a feature. The DC bias is removed from each flat region to

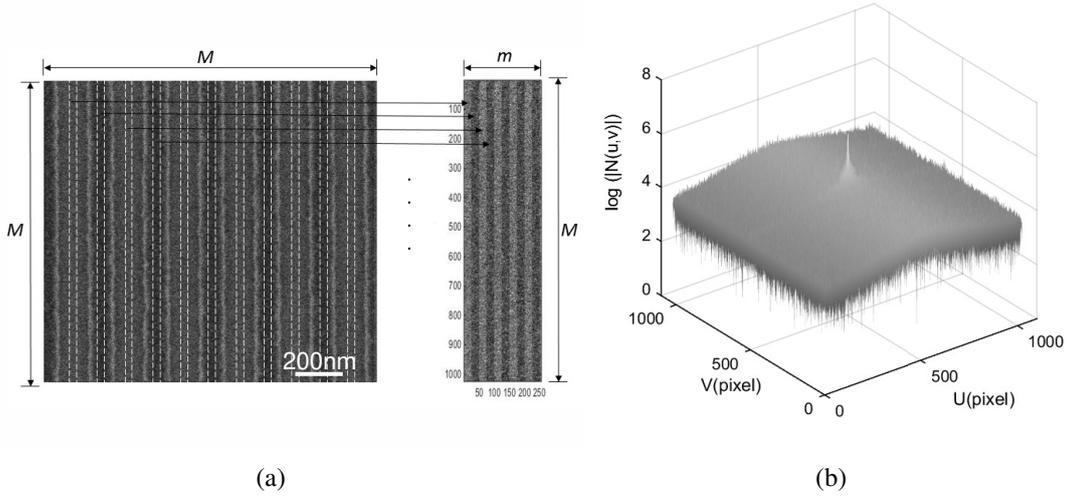


Figure 2.5: (a) Generation of sampled noise $n(x, y)$ (before the DC is removed) and (b) the frequency-domain representation $|N(u, v)|$ in the logarithmic scale.

leave only the noise. The sampled noise consisting of the extracted flat regions is smaller than the corresponding SEM image at least in one dimension, $m \times M$ in the case of a L/S pattern where $m < M$ (see Fig. 2.5(a)). The zero-padding is done to the sampled noise to make its size $M \times M$. The sampled noise after the DC removal and zero-padding is denoted by $n(x, y)$. The frequency-domain representation of $n(x, y)$ can be obtained through the Fourier transform. However, due to the fact that the size of sampled noise before the zero-padding is smaller than that of the SEM image, a scaling on the Fourier transform result makes it comparable to $I(u, v)$. Let $N(u, v)$ denote the scaled Fourier transform of $n(x, y)$, which can be expressed as in Eq. 2.3.

$$N(u, v) = \sqrt{\frac{M}{m}} \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} n(x, y) e^{-j \frac{2\pi(u x + v y)}{M}} \quad (2.3)$$

In Fig. 2.5(b), the plot of $|N(u, v)|$ is provided. It is seen that $|N(u, v)|$ does not vary with v while it does significantly with u . As pointed out earlier, this is due to the larger spatial correlation of noise in the x direction, which is related to the scanning direction of electron beam, contributing to the higher amplitudes of low frequency components. This may suggest the use of an anisotropic filter. However, in this study, the isotropic (radially symmetric) Gaussian filter is adopted (i) for the simplicity, (ii) since a pattern may include both vertical and horizontal boundaries of features, and (iii) the scanning direction can be in either direction.

The Fourier transform of a Gaussian function is also of Gaussian in the frequency domain. Let's denote the Fourier transform by $G(u, v)$ which may be expressed as in Eq. 2.4.

$$G(u, v) = \frac{1}{2\pi\sigma_f^2} e^{-\frac{u^2+v^2}{2\sigma_f^2}} \quad (2.4)$$

where σ_f is the standard deviation of $G(u, v)$ in the frequency domain, to be referred to as *cut-off frequency*.

The proposed method for designing a Gaussian filter determines the cut-off frequency σ_f from which the spatial-domain Gaussian filter can be derived. Let's define *total image intensity* to be $\sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |I(u, v)|G(u, v)$, which is the sum of all frequency components in an SEM image filtered by the Gaussian filter. Similarly, *total noise intensity* is defined as $\sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |N(u, v)|G(u, v)$, which is the sum of all frequency components in the sampled noise filtered by the Gaussian filter. Then, *total signal intensity* may be approximated to be $\sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |I(u, v)|G(u, v) - \sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |N(u, v)|G(u, v)$.

In Fig. 2.6, the total signal intensity and total noise intensity are plotted as functions of σ_f for a typical SEM image. When the cut-off frequency σ_f is very small (approaching to a zero), the total signal intensity is greater than the total noise intensity unless the noise is dominant. As σ_f increases, both the total signal intensity and total noise intensity monotonically decrease since $G(u, v)$ is a normalized Gaussian. But, the total signal intensity decreases faster since the signal is normally band-limited. Therefore, the total signal and noise intensities intersect at a certain σ_f . In the proposed method, σ_f at which *intensity ratio* defined in Eq. 2.5 becomes 1.0 is used in designing the Gaussian filter. The rationale behind this selection of the cut-off frequency is to reduce the noise level as much as possible under the condition that the signal level is not less than the noise level.

$$\text{Intensity ratio} = \frac{\sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |I(u, v)|G(u, v) - \sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |N(u, v)|G(u, v)}{\sum_{u=0}^{M-1} \sum_{v=0}^{M-1} |N(u, v)|G(u, v)} \quad (2.5)$$

In the continuous case, the Fourier transform of $e^{-\frac{x^2}{2\sigma^2}}$ is given as $\sqrt{2\sigma^2\pi}e^{-2\sigma^2(\pi f)^2}$. That is, the standard deviation in the spatial domain, σ , is related to the standard deviation in the

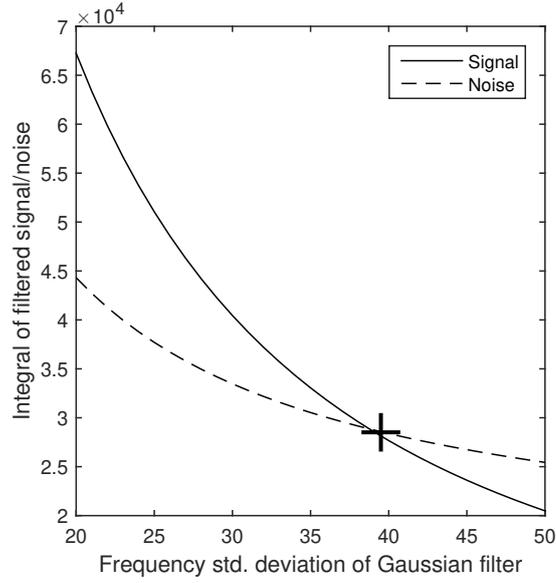


Figure 2.6: The total signal and noise intensities as functions of the cut-off frequency σ_f of the Gaussian filter. In designing the Gaussian filter, σ_f at which the two intensities become equal (marked by “+”) is used.

frequency domain, σ_f , as $2\pi\sigma\sigma_f = 1$. But, in the discrete case, the relationship becomes $2\pi\sigma\sigma_f = M$ where M is the size of an image. Therefore, the standard deviation of Gaussian filter in the spatial domain can be derived from the cut-off frequency as in Eq. 2.6.

$$\sigma = \frac{M}{2\pi\sigma_f} \quad (2.6)$$

The size of Gaussian filter, W_s , is determined such that most of the significant portion of Gaussian function is included. In this study, W_s is set to 6σ (note that the distance from the filter center of to the filter boundary is 3σ) as in Eq. 2.7. At 3σ a Gaussian function decreases to 1.11% of the peak value.

$$W_s = 6\sigma \quad (2.7)$$

A digital filter is normally symmetric and therefore 6σ is rounded up to the closest odd integer. It should be clear that the Gaussian filter specified by σ and W_s adapts itself to each individual SEM image according to the signal and noise levels in the image.

2.3 Detection of Feature Boundaries

In the remaining resist profile, the boundary of a feature is defined by a surrounding side-wall which is brighter than other regions in the (secondary-electron) SEM image (see Fig. 2.7). A band of white region appears along the feature boundary, and both sides of the band, i.e., *inner* and *outer* boundaries, are detected in this study. These boundaries are where the spatial change of image brightness is largest [35]. That is, the gradient of image brightness is either locally maximum or minimum at the boundaries (see Fig. 2.8). In the case of a L/S pattern, starting from the outside of a line, the first 4 maximum and minimum gradients correspond to outer, inner, inner and outer edges (boundaries), respectively (refer to Fig. 2.7).

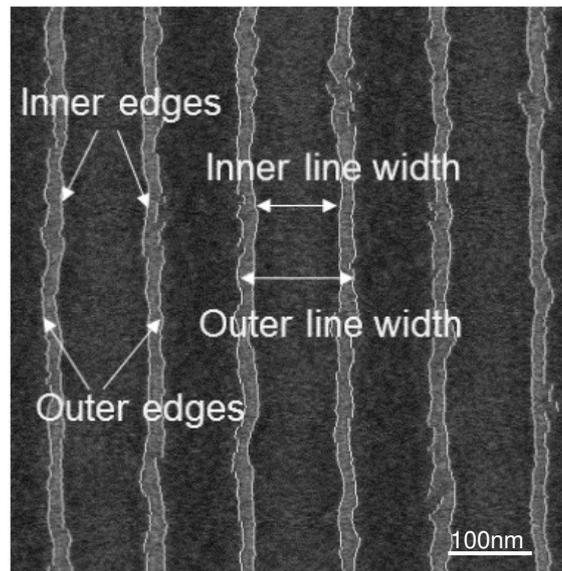


Figure 2.7: Each boundary of a feature (line) is characterized by a bright band, and both edges of the band region are detected resulting in inner and outer edges.

The designed Gaussian filter is applied to SEM images to reduce the noise level. Then, feature boundaries are detected by finding the local maximum and minimum of brightness gradient in the filtered image. A standard edge detector such as Sobel operator can be employed to compute the gradient of the brightness distribution. The Sobel operator for vertical boundaries (as in the SEM images in this study) is shown below.

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

The search of the maximum and minimum gradients is carried out row-by-row (for vertical lines). In order to guide the searching, an edge region is defined around each peak in the filtered SEM image as shown in Fig. 2.9. Then, the maximum and minimum gradients are searched within edge regions.

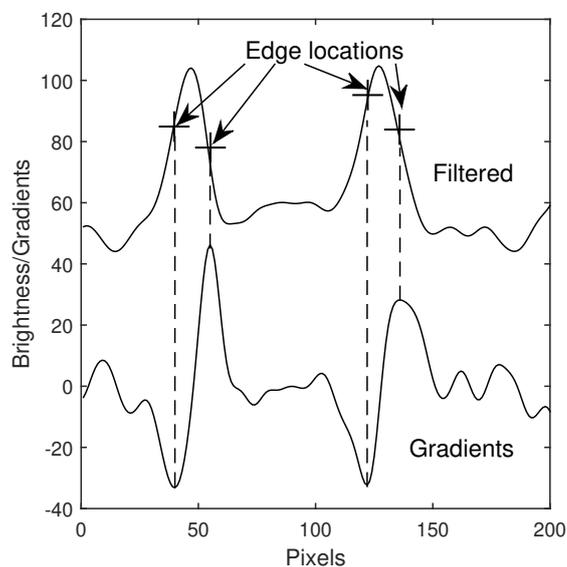


Figure 2.8: Maximum and minimum brightness gradients correspond to (inner and outer) edge locations.

Though detecting the inner and outer boundaries (edges) is considered in this study, one may want to find the boundary where the brightness is highest, i.e. peaks (see Fig. 2.9). The brightness peaks, which are normally between the inner and outer boundaries, may be searched within edge regions following the noise filtering (without computing the brightness gradient). Or, the peak detection can be done readily by locating the zero-crossings in the brightness gradient (refer to Fig. 2.8). Note that the brightness peak is where the brightness gradient changes its sign, from negative to positive or positive to negative.

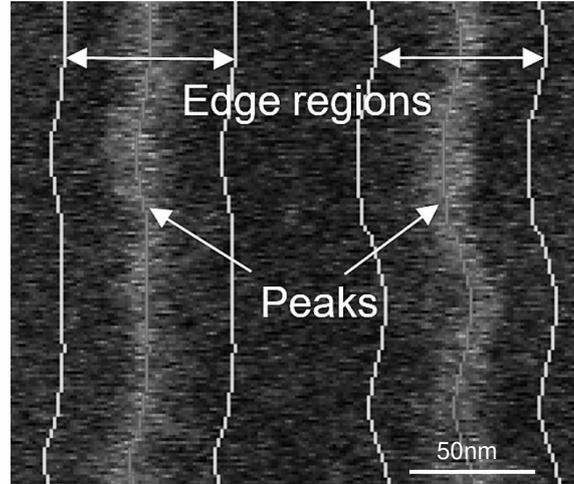


Figure 2.9: Edge regions are defined around peaks and the maximum and minimum gradients are found within edge regions.

2.4 Results and Discussion

For analyzing the effectiveness of a Gaussian filter designed by the proposed method, “reference (SEM) images” are generated from real SEM images. Feature boundaries in each reference image and therefore the CD and LER are known. To reference images, a spatially correlated noise rather than a simple noise such as shot (salt-and-pepper) noise is added to test the filter for realistic cases. The noise level in each reference image is varied for a thorough analysis. One of the four reference images used in this study is shown in Fig. 2.10, and the known CDs and LERs for the reference images are provided in Table 2.1.

After the noise filtering and boundary detection, the inner and outer linewidths are measured from the detected inner and outer edges, respectively (refer to Fig. 2.7). Then, the CD is computed as the average of the inner and outer linewidths, and the LER is quantified as the standard deviation of edge location, i.e., $1-\sigma$ LER, considering both inner and outer edges. The accuracy of measurement, i.e., detecting feature boundaries using the Gaussian filter designed by the proposed method is analyzed using the CD and LER deviations defined in Eq. 2.8 and Eq. 2.9:

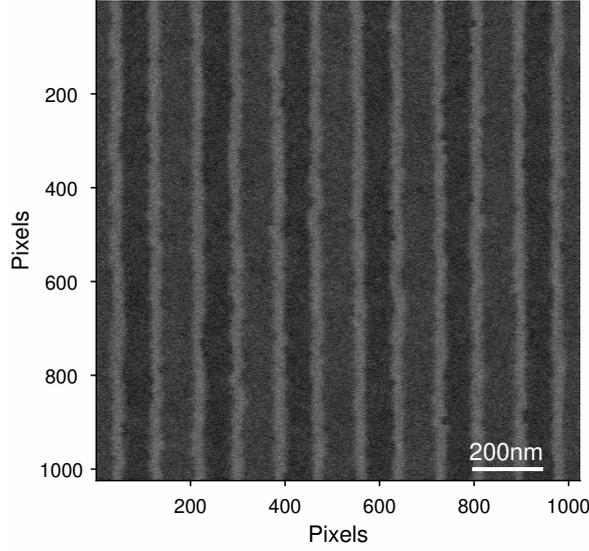


Figure 2.10: An example of reference SEM image where the same noise level as in the corresponding original SEM image is added.

Reference Image	LER (nm)	CD (nm)
Ref. Image 1	5.51	116.38
Ref. Image 2	3.42	128.15
Ref. Image 3	2.93	133.74
Ref. Image 4	2.71	144.50

Table 2.1: The known CD and LER in the four reference images generated from SEM images obtained with four different doses where the target CD is 120 nm.

$$LER_{Dev} = \frac{LER_m - LER_k}{LER_k} \times 100\% \quad (2.8)$$

$$CD_{Dev} = \frac{CD_m - CD_k}{CD_k} \times 100\% \quad (2.9)$$

where LER_m and LER_k are the measured and known LERs, respectively, and CD_m and CD_k the measured and known CDs, respectively.

In each reference image, four different noise levels are considered in order to test the adaptability of the Gaussian filter designed by the proposed method. The σ and W_s of the designed Gaussian filter, CD and LER measured, and percent CD and LER deviations for each noise level are provided for the four reference images in Tables 2.2-2.5.

SNR (dB)	Ref. Image 1					
	W_s (nm)	σ (nm)	LER (nm)	CD (nm)	LER Dev (%)	CD Dev (%)
13.2	35.25	5.60	5.59	116.26	1.51	-0.11
16.0	26.79	4.03	5.76	116.33	4.61	-0.05
20.0	18.33	2.64	5.81	116.33	5.57	-0.04
27.9	9.87	1.21	5.60	116.41	1.73	0.02

Table 2.2: The CD and LER measured from the reference image 1, and the deviations from the known CD and LER, are provided with the SNR level varied from 13.2 dB to 27.9 dB. The filter parameters (W_s and σ) are also included in each case.

SNR (dB)	Ref. Image 2					
	W_s (nm)	σ (nm)	LER (nm)	CD (nm)	LER Dev (%)	CD Dev (%)
13.2	46.53	7.41	3.38	128.11	-1.29	-0.03
16.0	32.43	5.34	3.55	128.17	3.71	0.02
20.1	21.15	3.48	3.70	128.11	8.27	-0.03
28.0	9.87	1.64	3.59	128.16	4.95	0.01

Table 2.3: The CD and LER measured from the reference image 2, and the deviations from the known CD and LER, are provided with the SNR level varied from 13.2 dB to 28.0 dB. The filter parameters (W_s and σ) are also included in each case.

SNR (dB)	Ref. Image 3					
	W_s (nm)	σ (nm)	LER (nm)	CD (nm)	LER Dev (%)	CD Dev (%)
13.8	46.53	7.41	2.89	133.66	-1.32	-0.06
16.5	32.43	5.34	3.07	133.74	4.95	0.00
20.6	21.15	3.48	3.19	133.77	8.86	0.02
28.5	9.87	1.64	3.09	133.76	5.67	0.01

Table 2.4: The CD and LER measured from the reference image 3, and the deviations from the known CD and LER, are provided with the SNR level varied from 13.8 dB to 28.5 dB. The filter parameters (W_s and σ) are also included in each case.

SNR (dB)	Ref. Image 4					
	W_s (nm)	σ (nm)	LER (nm)	CD (nm)	LER Dev (%)	CD Dev (%)
14.3	43.71	7.18	2.70	144.32	-0.38	-0.12
17.0	32.43	5.22	2.83	144.52	4.45	0.02
21.1	21.15	3.43	2.93	144.52	7.95	0.02
28.8	12.69	1.68	2.86	144.45	5.34	-0.03

Table 2.5: The CD and LER measured from the reference image 4, and the deviations from the known CD and LER, are provided with the SNR level varied from 14.3 dB to 28.8 dB. The filter parameters (W_s and σ) are also included in each case.

First, it can be seen from the tables that the Gaussian filter designed by the proposed method is effective in reducing the noise level and achieving the high accuracy in the measured CD and LER in all cases. Specifically, the average LER deviation is 4.4% and the maximum LER deviation is 8.86%. The CD deviation is less than 1% in all cases. The reason for the extremely small CD deviation is that the error in the detected edge location is not directly reflected in the CD deviation (note that the CD depends on both left and right edges) and the CD is averaged along the length dimension of feature (line). Second, one important observation is that the high accuracy is achieved independent of the noise level. This well demonstrates the good adaptability of the design method. As the SNR increases (i.e., the noise level decreases), the σ decreases since the noise needs to be filtered less. Third, the LER deviation tends to be smaller for a higher level of noise in some of the reference images and is negative in some of the highest levels of noise. This might be due to the slight over-filtering by the Gaussian filter, making the detected boundaries smoother.

It is worthwhile to point out that the LER deviation is caused by both the noise and filtering. The noise tends to make the measured LER increase. On the other hand, the filtering does the opposite by reducing the high frequency components of LER. Depending on which effect is more dominant, one may get an overestimation or underestimation of LER.

The proposed method is also applied to four real SEM images from which the reference images are generated. These SEM images were obtained by using four different dose levels for a L/S pattern (L=S=120 nm). A real SEM image is shown along with the detected boundaries in Fig. 2.11. The Gaussian filter designed, and CD and LER measured are provided in Table 2.6. It is not possible to quantify the accuracy of the detected boundaries in these images since the actual CD and LER are not known. However, it appears by visual inspection that the detected boundaries are accurate. As the dose level increases, the measured CD becomes larger as expected. Also, the measured LER is smaller for a higher dose level. This is consistent with the well-known behavior of LER. A higher dose leads to a smaller relative stochastic fluctuation of exposure (energy deposited in the resist), which makes the LER smaller.

The overall procedure including the filter design, boundary detection and computation of CD and LER is computationally efficient. It takes only a few seconds for an SEM image of

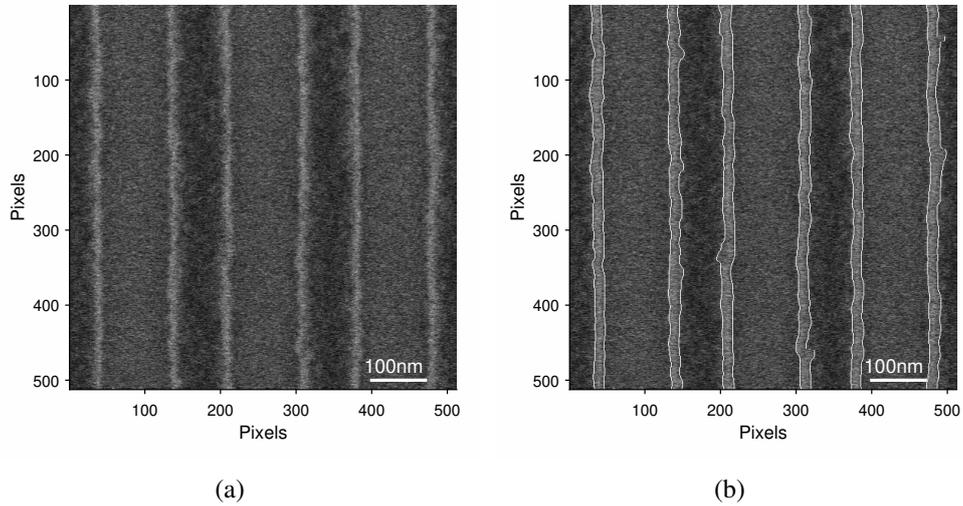


Figure 2.11: (a) An example of real SEM image and (b) Feature boundaries detected in SEM image in (a).

Dose *	Original images			
	W_s (nm)	σ (nm)	LER (nm)	CD (nm)
0.705	43.71	6.97	4.66	115.16
0.774	40.89	6.57	3.12	125.42
0.845	40.89	6.75	2.70	132.09
1.082	43.71	7.42	2.24	144.14

* Normalized Dose

Table 2.6: The filter parameters, and measured CD and LER for the SEM images obtained with different normalized doses, 0.705, 0.774, 0.845 and 1.082.

size 1024×1024 with 12 lines (24 edges per row) on a PC (Intel Xeon, 3.4 GHz, 8-G RAM). It is worthwhile to point out that the computational requirement of an edge modeling approach is much higher since each edge needs to be modeled individually through a time-consuming error-minimization procedure.

2.5 Summary

A practical method to design a Gaussian filter which plays a critical role in achieving a high accuracy of detecting feature boundaries in SEM images is described. Given an SEM image, the proposed method utilizes the frequency-domain representations of the signal and noise estimated from the image in order to determine the shape (standard deviation) and size of Gaussian filter. This allows to adaptively determine the level of filtering depending on the relative noise level. Using a set of reference images for which the CD and LER are known, the effectiveness of the Gaussian filter designed by the proposed method has been verified. In all cases considered, a high accuracy is achieved with the average LER deviation less than 5% and the average CD deviation less than 1%. It is worthwhile to point out that the high accuracy is achieved for various noise levels, which well demonstrates the good adaptability.

The proposed method is able to design an effective filter without requiring a complicated design procedure and is computationally efficient. Therefore, it is believed that the proposed design method has a good potential to be a useful tool.

Chapter 3

3-D Modeling of Electron-Beam Lithographic Process from SEM Images

A 3-D modeling approach enables a more realistic representation of exposure distribution in the resist and in turn allows one to obtain a more accurate remaining resist profile through the resist-development simulation [34]. In this chapter, the modeling procedures are described and the results from an extensive simulation are presented.

3.1 2-D Modeling Approach

In our previous study [29], a 2-D resist model was employed in modeling the e-beam lithographic process, i.e., the exposure variation along the resist-depth dimension was not taken into account. The three components of e-beam lithographic process are modeled through iterations by minimizing the modeling error, which consists of the modeling errors of the CD and LER. The CD modeling error is defined as the difference between the modeled CD and the CD measured in SEM images, and the LER modeling error as the difference between the modeled and the measured LERs. An example of SEM image with the detected edges is shown in Fig. 3.1. The CD and LER are measured from the detected inner and outer edges of a feature. In order to make the convergence easier to achieve, the 2-D modeling is carried out in two steps, i.e., the linewidth (or CD) matching followed by the LER matching.

The deterministic line spread function, LSF instead of the PSF, is modeled since L/S patterns are considered. In each iteration, the exposure distribution is computed from the stochastic LSF and converted into the developing rate. Then, the remaining resist profile is obtained from the exposure distribution through the development simulation using the path-based method [34]. From the remaining resist profile, the CD and LER are computed from the feature boundaries. Then, the model is adjusted such that the modeling error is reduced.

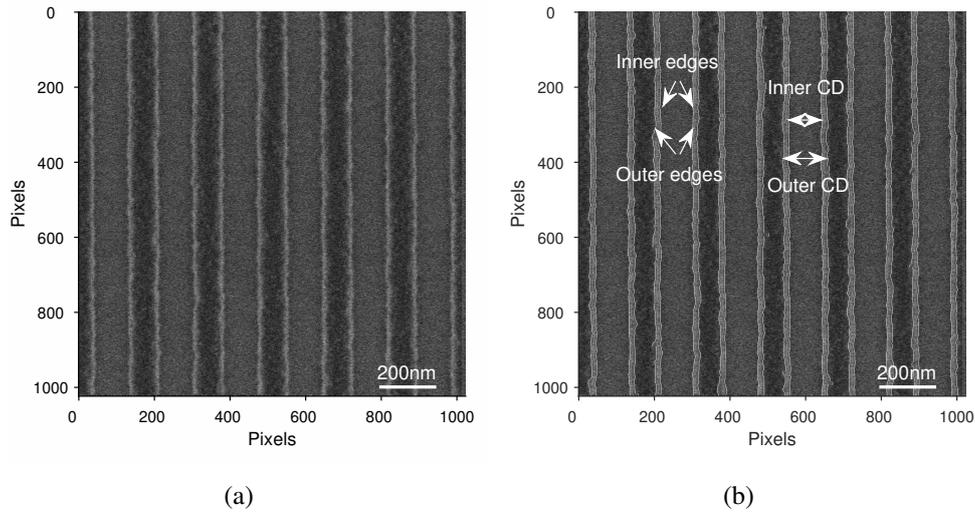


Figure 3.1: (a) An example of an SEM image in a L/S pattern and (b) detected boundaries overlapped on an SEM image in (a). The inner and outer edges are used to calculate the CD and LER.

The effectiveness of the 2-D modeling approach was well demonstrated for the case of relatively thin resist and high beam energy in the previous study. However, this modeling relies on a 2-D model, i.e., the exposure variation along the resist-depth dimension is not considered. Therefore, when the exposure variation among layers in a resist is significant, such as for a thicker resist, the 2-D modeling approach can lead to significant errors in modeling the e-beam lithographic process. In order to reduce the modeling error, a more realistic model needs to be employed. Hence, in the new approach, a 3-D model of resist layer is adopted and the modeling procedures are also improved.

The improved procedures are applicable to the 2-D approach, too. In comparison with the 3-D approach in Sec. 3.3, the improved 2-D approach will be referred to.

3.2 3-D Modeling Approach

In the proposed modeling approach, a 3-D resist model of 5 layers (see Fig. 3.2) is employed so that the exposure variation along the depth dimension of resist is taken into account in order to enhance the accuracy and applicability of modeling. An overview of the approach is presented first, and then the models and procedures involved in the modeling process are described in detail.

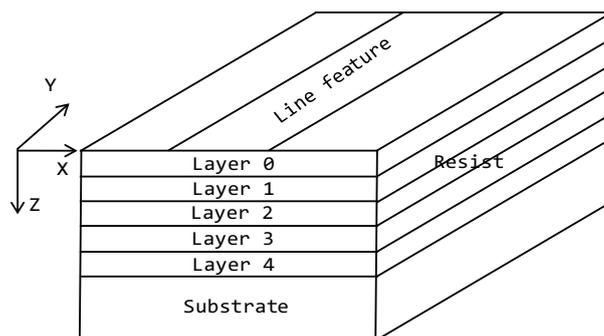


Figure 3.2: The resist model includes 5 resist layers, where the middle layer of the resist is taken as the reference layer for the purpose of the 3-D modeling.

3.2.1 Overview

The goal of the modeling is to derive the LSF and noise models from experimental results such that the models can be used in the estimation and optimization for the same e-beam lithographic process. The noise represents the stochastic fluctuation of exposure. Through an iterative procedure, the LSF and noise are modeled, being guided by CD measurements obtained from SEM images. In order to keep the complexity of modeling low, the conversion formula is fixed. The overall flow of the modeling process is depicted in Fig. 3.3, where the simulated annealing guides the iterations of modeling.

A set of parameters is used in representing each of LSF and noise models, and the iterative modeling process is structured following the simulated annealing procedure. In each iteration, (i) the model parameters are randomly perturbed (adjusted), (ii) the exposure distribution (of each of the circuit patterns from which the CD measurements are taken) is computed using the adjusted LSF and noise models, (iii) from the exposure distribution, the remaining resist profile is obtained through a resist-development simulation, (iv) the CD is measured from the remaining resist profile and the CD modeling error is computed. The steps from (i) to (iv) are repeated for each of the CD measurements from SEM images. The change in the total CD modeling error, i.e., cost function, determines the probability of accepting the adjusted models of LSF and noise. The iteration continues until an acceptable modeling error is achieved.

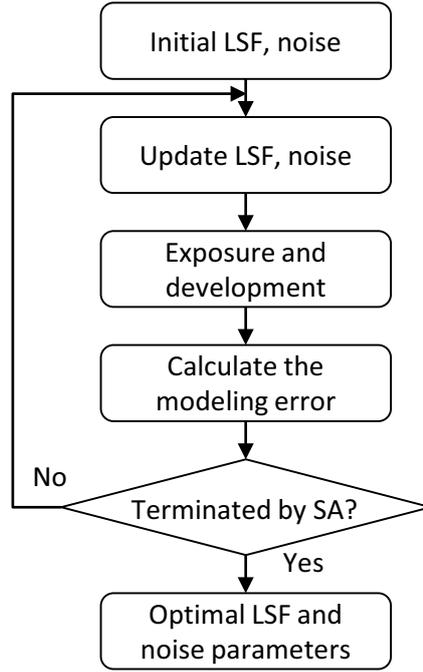


Figure 3.3: Overall flow of the modeling approach: the modeling is guided by simulated annealing.

3.2.2 Exposure Distribution

In Fig. 3.4, the deterministic exposure distribution for a single line is plotted for 5 layers, where the variation of the exposure distribution along the resist-depth dimension can be clearly seen.

Since L/S patterns are considered in this study, the LSF, instead of PSF, is employed to facilitate the modeling process. The stochastic LSF can be represented by the corresponding deterministic LSF, denoted by $lsf_d(x, z)$, to which the noise is added. Therefore, the deterministic LSF and noise are modeled. Modeling the deterministic LSFs of 5 layers independently would require a high complexity of computation and may not guarantee the convergence of modeling. To reduce the complexity, the LSF at a reference layer is modeled and the LSFs at other layers are derived from the LSF at the reference layer. This derivation is guided by the layer-dependency ratio, $LR(x, z)$, defined in Eq. 3.1.

$$LR(x, z) = \frac{lsf_d(x, z)}{lsf_d(x, z_0)} \quad (3.1)$$

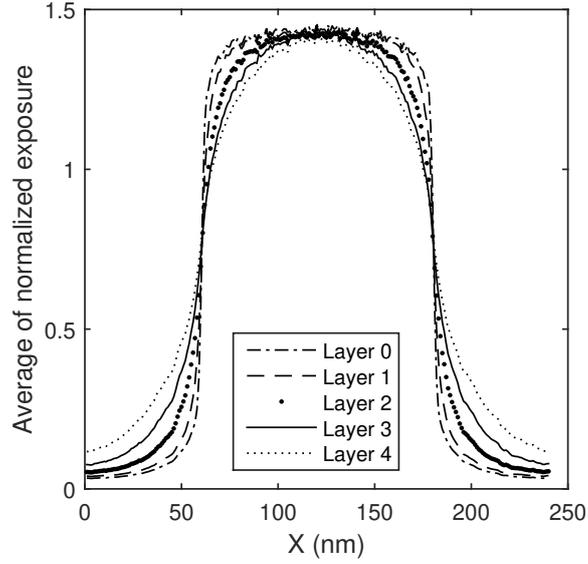


Figure 3.4: An example of average exposure distribution along the y direction at different layers from the PMMA/Si system: linewidth of 120 nm, resist thickness of 300 nm and beam energy of 10 keV. The average exposure varies along resist-depth dimension.

where $lsf_d(x, z)$ and $lsf_d(x, z_0)$ are the deterministic LSFs at a layer and the reference layer, respectively.

A number of instances of stochastic LSF are generated from a Monte Carlo simulation (CASINO) and then the deterministic LSF for each layer is obtained by averaging the instances for the layer [36]. Then, the LR can be computed from the deterministic LSFs. In Fig. 3.5, an example of the LR obtained using the CASINO software is provided. In the modeling, the stochastic LSF at a layer other than the reference layer is computed as in Eq. 3.2.

$$lsf(x, z) = lsf(x, z_0) \cdot LR(x, z) \quad (3.2)$$

3.2.3 Noise Model

The noise, i.e., exposure fluctuation, tends to have the following characteristics: when the exposure level, i.e., deterministic exposure, is higher, the absolute fluctuation of exposure is larger while the relative fluctuation, defined as the absolute fluctuation normalized to the average (deterministic) exposure, is smaller [29]. Noting that the deterministic exposure is

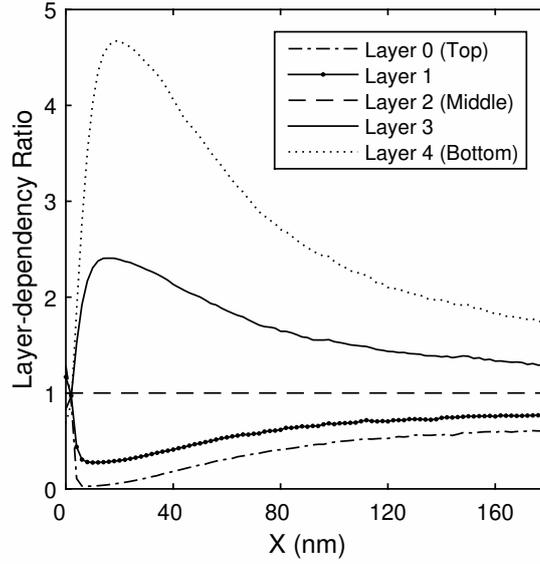


Figure 3.5: An example of layer-dependency ratio: resist thickness of 200 nm, beam energy of 30 keV in PMMA/Si system.

highest at the center ($x = 0$) of LSF. The noise, $e_n(x)$, at the reference layer ($z = z_0$) is modeled as in Eq. 3.3.

$$e_n(x) = rand \cdot e_0 \left(\frac{lsf_d(x, z_0)}{lsf_d(0, z_0)} \right)^\alpha \quad (3.3)$$

where $rand$ is a random number of the normal distribution $N(0, 1)$, e_0 is a noise level at the center of LSF and α is a parameter ($0 < \alpha < 1$).

Through iterations, e_0 and α are to be estimated. Then, the stochastic exposure at the reference layer is calculated as Eq. 3.4.

$$lsf(x, z_0) = lsf_d(x, z_0) + e_n(x) \quad (3.4)$$

To facilitate the modeling process, certain constraints to be imposed on the noise model are derived from the data obtained by a Monte Carlo simulation. The e_0 must be within 2 – 18% of the exposure level at the center (peak) of LSF, and α in the range of 0.3 - 0.7.

3.2.4 LSF Model

The computation of 3-D distribution of stochastic exposure starts from the deterministic LSF at the reference layer as described above. In this study, the middle layer of resist is used

as the reference layer. Therefore, the deterministic LSF at the middle layer is modeled. Since an LSF is symmetric with respect to its center ($x = 0$), only one side ($x \geq 0$) needs to be considered in the modeling.

In the previous study [29], the LSF was modeled point-by-point, which resulted in a high complexity of modeling and made the convergence hard to achieve. In this study, the characteristics of LSF are utilized to lower the complexity and also enable an easier convergence. A typical LSF monotonically decreases as the distance, x , from the center increases. It varies (decreases) fast where x is relatively small and very slow beyond a point, to be referred to as turning point. Therefore, the deterministic LSF can be represented by two parts, i.e., the fast-varying (“main lobe”) and slow-varying (“tail”) parts where the two parts are connected through the turning point as illustrated in Fig. 3.6. The main lobe is modeled point-by-point for sufficient flexibility. On the other hand, the tail is modeled by an exponential function as in Eq. 3.5, which involves only three parameters.

$$lsf_d(x, z_0) = V_t \cdot b^{(x-x_t)} \quad (3.5)$$

where V_t is the deterministic exposure at the turning point, x_t is the turning point and b is the base of exponential function ($b < 1$).

Since the domain of main lobe is much smaller than that of tail, the reduction of complexity is significant.

In order to facilitate the modeling process and achieve a convergence to an acceptable LSF, a few constraints, obtained from a Monte Carlo simulation, are imposed on the LSF, including the shape of main lobe (80 – 180% of the simulated LSF), the location of turning point (3 – 100 nm), the base b (0.96 – 0.99), and the ratio of the peak to the next point (< 1.25).

3.2.5 Cost Function

The cost function quantifies a modeling error and the iterative modeling process continues until the (value of) cost function is reduced below an acceptable threshold. In the previous study [29], both CD and LER modeling errors were included in the cost function. However, the LER

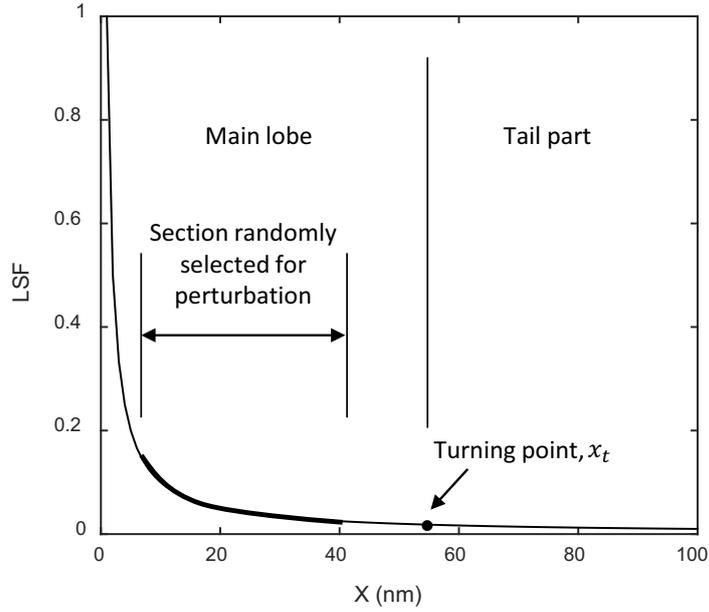


Figure 3.6: Illustration of LSF adjustment: In an iteration, the length, location and scaling of the perturbed section in the main lobe are randomly determined.

modeling error is excluded from the cost function in this study, to enable a lower complexity of modeling leading to a faster and more stable convergence. The exposure fluctuation (noise), which causes the LER, affects the CD, i.e., a higher level of fluctuation makes the CD larger. Therefore, this new cost function takes into account the effect of noise and, therefore, can be considered to reflect the LER modeling error indirectly.

The CD modeling error is defined as the difference between the modeled and reference CDs at a layer where the reference CD is defined as the CD measured in the reference remaining resist profile. In the previous study, the CD modeling error was computed using the average of the top and bottom CDs, i.e., the average of CDs at the top and bottom layers of resist. In this study, in order to obtain a more realistic modeling result, the CD modeling error is evaluated at each of the top and bottom layers. The cost function ϵ is defined as the average modeling error in Eq. 3.6.

$$\epsilon = \frac{1}{N} \sum_{i=1}^N |CD_{m,i} - CD_{ref,i}| \quad (3.6)$$

where $CD_{m,i}$ and $CD_{ref,i}$ are the i th modeled and reference CDs, respectively. N is the total number of the top and bottom CDs from the SEM images.

3.2.6 Simulated Annealing

The iterative modeling process is guided by the simulated annealing and the flow of modeling is shown in Fig. 3.7. The modeling process consists of outer and inner loops. The outer loop corresponds to the annealing temperature. The inner loop is finished after 3 successive perturbations are rejected, or iterated up to 10 times to allow a sufficient annealing. In each iteration, the deterministic LSF and noise are randomly perturbed. The stochastic exposure distribution is computed using the stochastic LSF derived from the perturbed deterministic LSF and noise. Through the resist-development simulation, the remaining resist profile is obtained, from which the modeled CD and subsequently CD modeling error are computed. This process is repeated for each measurement of CD. Then, the cost function is evaluated and the probability P of accepting the perturbed deterministic LSF and noise, is computed as in Eq. 3.7.

$$P = \min\{1, e^{-\frac{\epsilon_c - \epsilon_a}{T}}\} \quad (3.7)$$

where ϵ_c is the modeling error (cost function) in the current iteration, ϵ_a is the last accepted modeling error, and T is the annealing temperature.

The behavior of P (for $\Delta\epsilon > 0$) is plotted in Fig. 3.8, where $\Delta\epsilon = \epsilon_c - \epsilon_a$. The perturbed deterministic LSF and noise are accepted with the probability of P . This annealing process is continued until the cost function goes below a threshold (0.01 nm) or the maximum number of iterations allowed (or $T < 0.01$) is reached.

3.2.7 Initialization

The deterministic LSF and noise are initialized subject to the respective constraints (refer to Sections 3.2.3 and 3.2.4): The initial deterministic LSF for the iterative modeling process is obtained by randomly changing the LSF, up to 20%, used in generating the corresponding reference resist profiles. The location of turning point x_t is initialized to the location where the value of LSF is 1% of the peak, the base b to 0.97, e_0 to 8% of the peak of LSF, and α to 0.5.

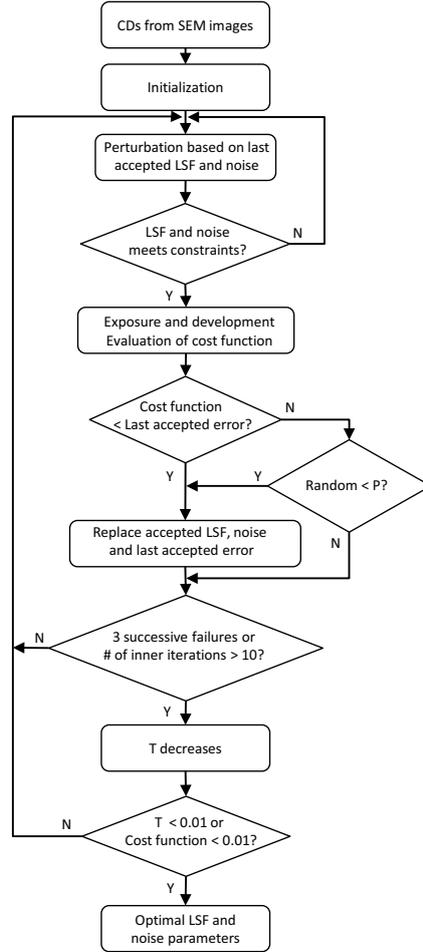


Figure 3.7: The linewidth matching guided by SA is employed in the modeling process to minimize the cost function. In each iteration, the perturbed LSF and noise parameters are utilized going through the exposure and resist-development processes in order to compute the cost function.

3.2.8 Perturbation

In each iteration, the deterministic LSF and noise are randomly perturbed with the following guidelines.

A selected section of the main lobe is scaled up or down where the location $S(0 \leq S \leq x_t)$ and length $L(1 \leq L \leq 8)$ of the section are randomly decided. The perturbation ranges of the scaling factor for the selected section of LSF, the base b of the exponential function for the tail part, the noise level e_0 and α for the noise are reduced from 150% (in the early stage) to 2% (in the final stage) of the allowed ranges mentioned in Fig. 3.2.4 through iterations. The location x_t of the turning point is changed up to ± 1 nm in each iteration.

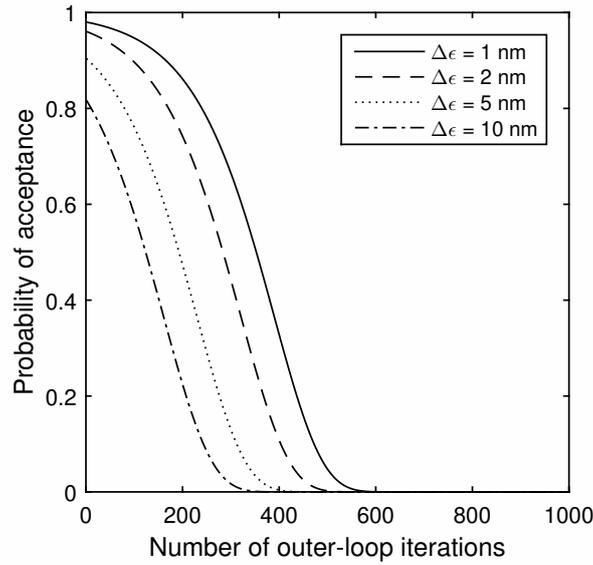


Figure 3.8: The probability of acceptance decreases as the number of outer-loop iterations increases (corresponding to the decrease of annealing temperature) for $\Delta\epsilon > 0$.

3.3 Results and Discussion

3.3.1 Simulation

Reference resist profiles, for which the stochastic LSFs are known, are used in the modeling. Note that, for arbitrary SEM images, the (true) LSF is not known and therefore the accuracy of modeling can not be analyzed. Each reference resist profile is obtained from the respective stochastic LSF obtained from a Monte Carlo simulation for the substrate system consisting of PMMA on Si. To analyze the versatility of the 3-D modeling, a number of reference resist profiles are generated by varying lithographic and feature parameters. The beam energy is varied from 10 to 50 keV, the resist thickness from 100 to 500 nm, the feature (line) width from 20 to 120 nm, and 4 different dose levels are employed for each feature width.

The features with widths of 20, 60 and 120 nm are employed in the modeling. The pixel sizes for exposure calculation and development simulation are 1 nm and 0.1 nm, respectively.

3.3.2 Modeling

The modeled deterministic LSFs are provided for varying resist thickness and beam energy in Fig. 3.9 where each LSF is normalized by its total energy (exposure). It can be seen that the

shape of LSF follows the expected general trend. The LSF is sharper for a higher beam energy or a thinner resist. When the beam energy is higher, electrons are scattered less in the lateral direction and deposit their energy mostly in a small area, leading to a sharper LSF. In the case of a thinner resist, electrons still maintain a relatively higher energy when they arrive at the substrate interface, i.e., there is a smaller (vertical) space through which they can be scattered and therefore they have been scattered less reaching the interface. This results in a sharper LSF. It is also observed that there is a smaller difference in the shape of LSF among different levels of beam energy when the resist is thinner, but a larger difference for a thicker resist. These can be understood through the same reasoning.

The accuracy of modeling has been analyzed in detail and the analysis results, i.e., average modeling error (cost function), are provided in Tables 3.1-3.2. From the tables, it is seen that the modeling error is much smaller for the 3-D modeling than for the 2-D modeling. Also, the modeling error tends to become significantly larger as the resist thickness increases in the case of 2-D modeling while it remains to be small. This well justifies the need to use the proposed 3-D modeling, compared to the 2-D modeling.

3.4 Summary

In this chapter, a practical approach to the 3-D modeling of the e-beam lithographic process is described. Modeling the e-beam lithographic process based on the information extracted from SEM images is a practical approach that has a potential to achieve high accuracy in the CD estimation and correction. The 2-D modeling method previously developed does not consider the exposure variation along the depth dimension of resist. This can result in lower accuracy when the resist is thicker or the beam energy is lower. Hence, in this study, a 3-D approach to the modeling has been investigated in order to develop a more general modeling method. The layer dependency of exposure is expressed by the layer-dependency ratio with the LSF of middle layer directly modeled. The iterative modeling process adopts the simulated annealing to enhance the possibility of finding a good quality of solution (model). Through an extensive simulation, the 3-D modeling has been tested for various combinations of resist thickness and beam energy. The modeling errors (CD and LER) are substantially smaller than those by the

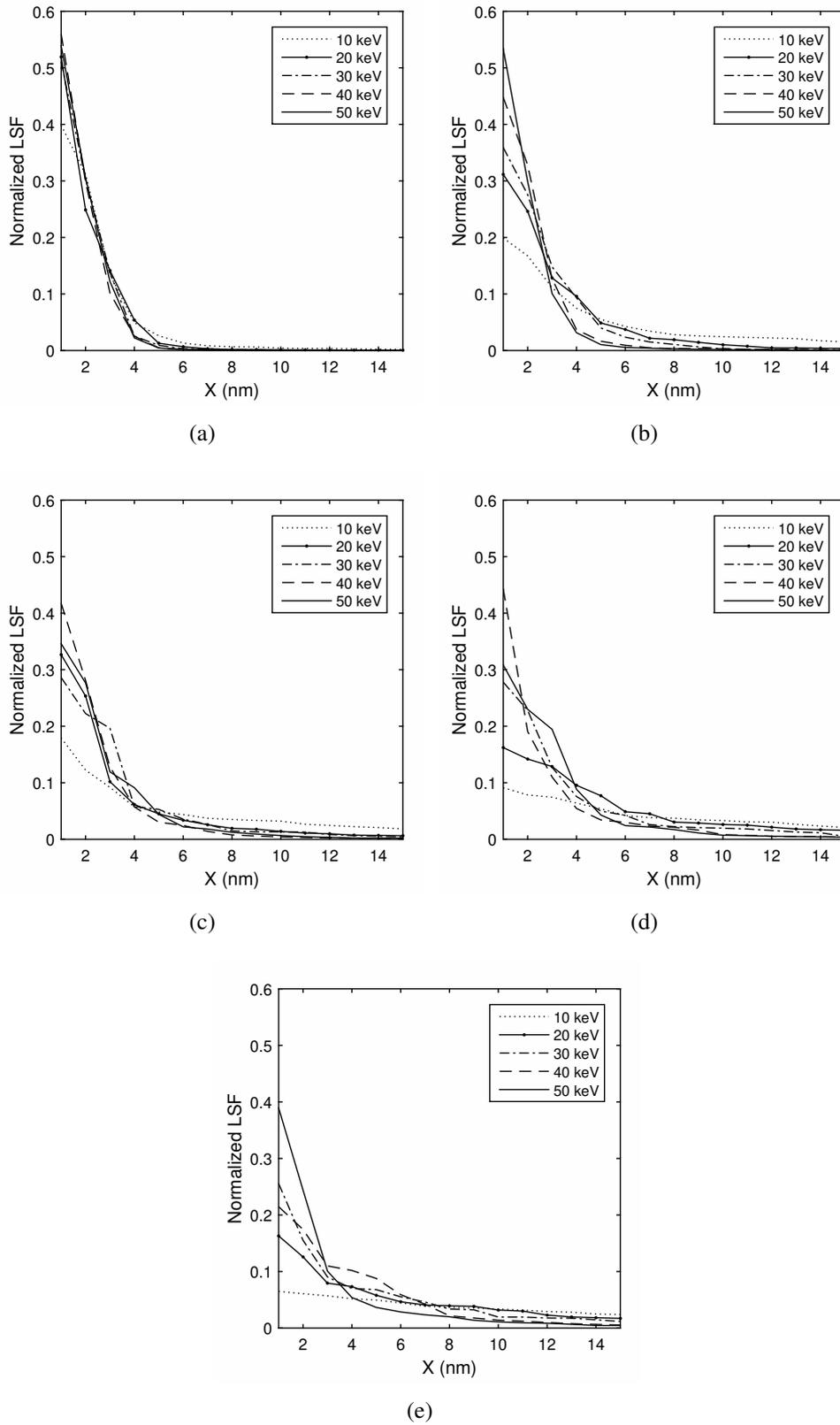


Figure 3.9: The deterministic LSFs from the 3-D modeling procedures for varying beam energy from 10 to 50 keV and resist thickness of a) 100 nm; b) 200 nm; c) 300 nm; d) 400 nm and e) 500 nm.

Thickness (nm)	Beam energy (keV)	3D/2D	Modeling error (nm/layer)	e_0	α	x_t (nm)	b	
100	10	3D	0.33	0.0037	0.517	35	0.965	
		2D	2.57	0.0070	0.519	20	0.977	
	20	3D	0.37	0.0038	0.532	46	0.971	
		2D	1.07	0.0025	0.543	28	0.970	
	30	3D	0.11	0.0023	0.463	91	0.976	
		2D	0.59	0.0018	0.484	18	0.978	
	40	3D	0.06	0.0046	0.505	14	0.961	
		2D	0.41	0.0013	0.518	37	0.965	
	50	3D	0.09	0.0083	0.522	21	0.970	
		2D	0.30	0.0013	0.488	90	0.971	
	200	10	3D	2.74	0.0288	0.519	18	0.961
			2D	8.66	0.0203	0.533	56	0.970
20		3D	0.97	0.0195	0.520	29	0.969	
		2D	5.41	0.0064	0.535	82	0.985	
30		3D	0.53	0.0053	0.544	26	0.966	
		2D	3.02	0.0057	0.521	76	0.967	
40		3D	0.17	0.0054	0.526	22	0.970	
		2D	2.37	0.0052	0.497	33	0.973	
50		3D	0.12	0.0020	0.507	86	0.988	
		2D	1.57	0.0023	0.491	24	0.964	
300		10	3D	1.14	0.0133	0.534	64	0.962
			2D	16.99	0.0138	0.465	79	0.966
	20	3D	0.34	0.0132	0.560	58	0.964	
		2D	9.21	0.0106	0.554	83	0.987	
	30	3D	0.94	0.0091	0.505	34	0.974	
		2D	7.43	0.0108	0.462	40	0.965	
	40	3D	0.14	0.0052	0.515	48	0.962	
		2D	5.24	0.0064	0.484	69	0.964	
	50	3D	1.83	0.0098	0.464	16	0.964	
		2D	4.71	0.0053	0.476	33	0.967	

Table 3.1: The results from 2-D and 3-D modeling with resist thickness varied from 100 to 300 nm, and beam energy from 10 to 50 keV.

Thickness (nm)	Beam energy (keV)	3D/2D	Modeling error (nm/layer)	e_0	α	x_t (nm)	b	
400	10	3D	0.84	0.0140	0.506	77	0.987	
		2D	28.90	0.0153	0.471	31	0.960	
	20	3D	0.71	0.0115	0.524	41	0.975	
		2D	15.98	0.0105	0.553	31	0.960	
	30	3D	0.31	0.0093	0.482	56	0.960	
		2D	10.66	0.0255	0.523	55	0.961	
	40	3D	1.19	0.0098	0.437	33	0.965	
		2D	8.33	0.0086	0.502	45	0.966	
	50	3D	0.53	0.0046	0.573	51	0.987	
		2D	6.40	0.0059	0.532	60	0.966	
	500	10	3D	6.25	0.0167	0.494	76	0.966
			2D	41.03	0.0161	0.514	77	0.967
20		3D	1.27	0.0128	0.455	32	0.960	
		2D	24.29	0.0163	0.464	29	0.968	
30		3D	0.64	0.0121	0.552	61	0.964	
		2D	15.66	0.0085	0.560	40	0.967	
40		3D	4.26	0.0102	0.474	18	0.962	
		2D	11.60	0.0188	0.471	96	0.979	
50		3D	2.03	0.0089	0.539	95	0.969	
		2D	10.54	0.0082	0.515	66	0.984	

Table 3.2: The results from 2-D and 3-D modeling with resist thickness varied from 400 to 500 nm, and beam energy from 10 to 50 keV.

2-D modeling in most cases and the shape of modeled LSF shows an expected dependency on the resist thickness and beam energy.

Chapter 4

Accurate Estimation of Critical Dimension and Line Edge Roughness Using a Neural Network

The use of a NN for the estimation of CD and LER makes unnecessary an explicit modeling of the e-beam lithographic process. A NN is trained by data of a number of samples obtained from the e-beam lithographic process and the trained NN is used in estimating the CD and LER, given a set of the e-beam lithographic parameters, i.e., the resist thickness, beam energy, beam diameter, linewidth and normalized dose. In this chapter, the procedures of training and tuning the NN are described and the estimation results obtained from the trained NN through an extensive simulation are discussed. Also, the issue of reducing the training samples is addressed to lower the computational complexity.

4.1 E-Beam Lithographic Process

A typical substrate system consists of a certain type of resist on top of the substrate. The resist layer is exposed according to the dose distribution required for a circuit pattern to be transferred. The exposure distribution in the resist, $e(x, y, z)$, is related to the dose distribution, $d(x, y, 0)$, and the point spread function, $psf(x, y, z)$, through the convolution as in Eq. 4.1.

$$e(x, y, z) = \int_{y'} \int_{x'} d(x - x', y - y', 0) psf(x', y', z) dx' dy' \quad (4.1)$$

Then, the substrate system goes through a resist-development process and the transferred pattern is obtained as the remaining resist profile. The resist-development is governed by the developing rate at each point of resist, which is determined by the exposure at the point. The relationship between the exposure and the developing rate is often specified by a conversion formula [34]. The contours (feature boundaries) of a remaining resist profile at the top and the

bottom layers are provided in Fig. 4.1 for a substrate system. From reference resist profiles (equivalent to SEM images), the feature boundaries at the top and bottom are detected, and the top and bottom CDs are determined, respectively. The CD is computed as the average of top and bottom CDs. And the LER is computed as the standard deviation of the inner and outer edge locations along the length dimension of a feature, i.e., $1-\sigma$ LER [29].

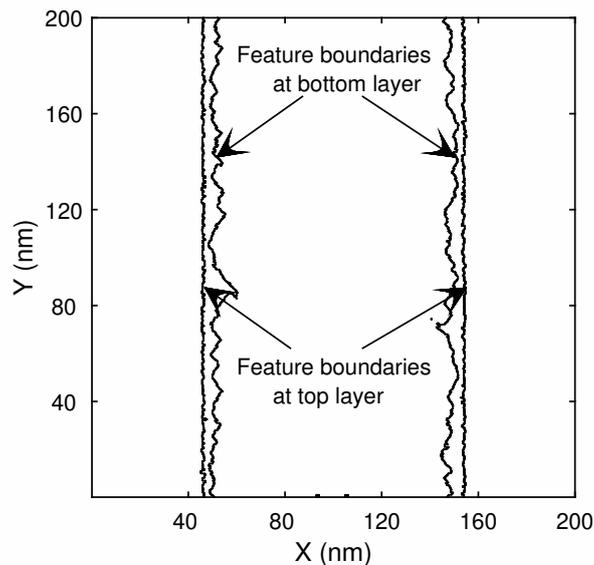


Figure 4.1: Contours of the remaining resist profiles at the top and bottom layers: substrate system of PMMA on Si, resist thickness of 100 nm, beam energy of 10 keV, beam diameter of 10 nm, and normalized dose of 0.335.

In a previous study [29], this e-beam lithographic process was modeled explicitly by the LSF, conversion formula, and exposure fluctuation, to estimate the CD and LER in a given pattern. An alternative approach, i.e., employing a NN, which avoids the explicit modeling and therefore reduces the computational requirement is taken in this study. The weights of all edges (links between neurons or nodes) in the NN may be considered to collectively model the e-beam lithographic process and are determined to realize the behaviors of CD and LER.

Both CD and LER vary with e-beam lithographic parameters in general. For example, the CD monotonically increases with the increment of dose as shown in Fig. 4.2(a). When the dose is relatively low, the CD increases fast with the dose. However, once the CD reaches the width of the exposed area, the rate of CD increase significantly drops as the dose continues to increase. This is due to the fact that the exposure level decreases rapidly over the edge of the

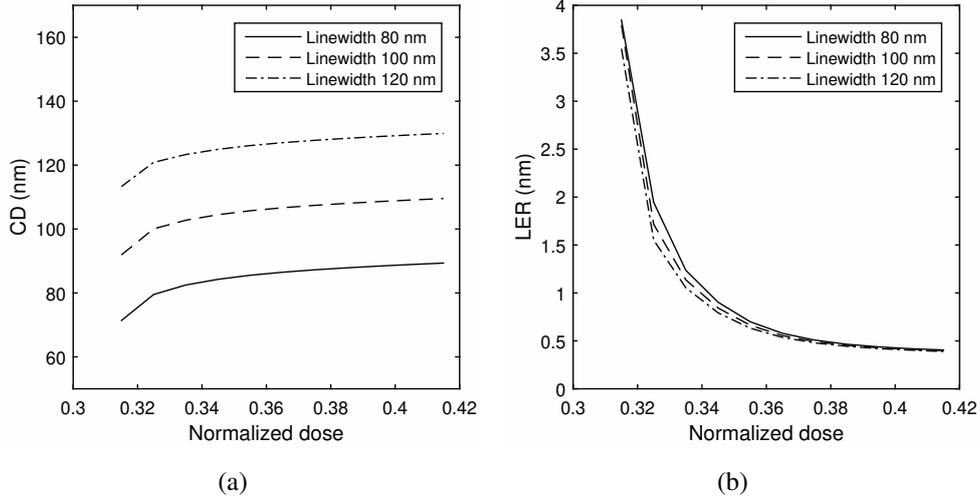


Figure 4.2: The dependency of (a) CD and (b) LER on the dose for varying linewidth when the development time is fixed: substrate system of PMMA on Si, resist thickness of 100 nm, beam energy of 10 keV, beam diameter of 10 nm, normalized dose of 0.335, exposure interval of 1 nm, and development time of 0.5 min.

exposed area. A similar dependency of LER on the dose can be seen in Fig. 4.2(b) except that the LER decreases with the dose. As the dose increases, the relative fluctuation of exposure decreases and the edge location moves from the inside of a feature to the outside, both making the LER smaller. Note that the absolute fluctuation of exposure is much larger in the inside than in the outside.

4.2 Estimation of CD and LER

The neural network employed for the estimation of the CD and LER in this study is a multi-layer NN which is trained by the error back-propagation (EBP) algorithm with a set of training samples. Each training sample is a known CD or LER for a combination of e-beam lithographic parameters.

4.2.1 Structure of NN

The NN consists of an input layer, an output layer, and hidden layers between the input and output layers as illustrated in Fig. 4.3. There are multiple nodes (neurons) in each layer and every node in a layer is connected to all the nodes in the following layer. The output from

node j in a layer is expressed as $O_j = f(\sum_{i=1}^N w_{ij}I_i)$ where $f()$ is the activation function, N is the number of inputs from the previous layer to node j , I_i is the i th input, i.e., the output from node i in the previous layer, and w_{ij} is the weight of the link (edge) between node i and node j . In this study, the Sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ is employed as the activation function.

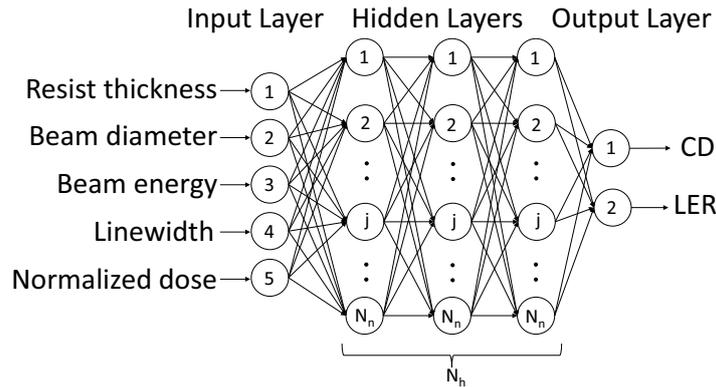


Figure 4.3: The illustration of a neural network with an input layer of 5 inputs (resist thickness, beam energy, beam diameter, linewidth and normalized dose), N_h hidden layers with N_n neurons each, and an output layer of 2 outputs (CD and LER).

Each training sample, which is an input vector of the NN, consists of e-beam lithographic parameters, i.e., resist thickness, beam energy, beam diameter, linewidth and normalized dose, and the output vector of the NN contains the CD and LER. For each training sample, the CD and LER are known. From this point on, each element (i.e., e-beam lithographic parameter) of the input vector is referred to as an input and, similarly, each element of the output vector as an output.

4.2.2 Training with E-Beam Lithographic Samples

Through the input layer, each training sample is fed into the NN. After the forward calculation, the corresponding CD and LER become available at the output layer. Then, the training errors (CD and LER) are computed, which are propagated backward through the layers to determine the amount of weight (w_{ij}) adjustment for every edge. This process of training is repeated until the NN converges, e.g., the training errors no longer decrease. The edge weights may be

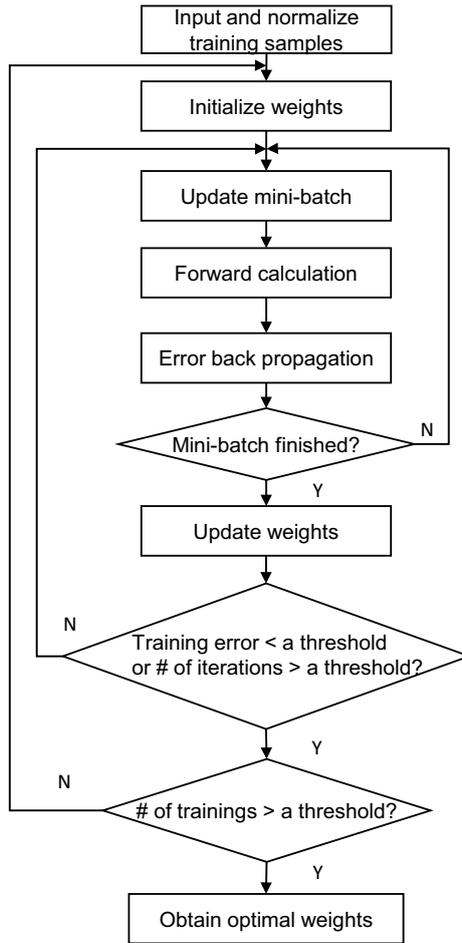


Figure 4.4: The NN training process employs the EBP algorithm using a mini-batch. Each mini-batch of samples is selected randomly in an iteration. The weights of nodes are updated after the training on a mini-batch of samples is finished.

updated after each training sample or a batch of training samples. The overall flow of training is shown in Fig. 4.4. The specific issues involved in the training are discussed below.

4.2.2.1 Normalization of Inputs

The range of an input may vary significantly with input, e.g., resist thickness (nm) from 100 to 500, beam energy (keV) from 10 to 50, and normalized dose from 0.3 to 3.5. Therefore, in order to have the same or similar range for all inputs in the NN, each input is normalized as in Eq. 4.2.

$$x'_k = \frac{x_k - E[x_k]}{\sigma} \quad (4.2)$$

where x_k is an input, x'_k is the normalized element, and $E[x_k]$ and σ are the mean and standard deviation of x_k , respectively.

4.2.2.2 Initialization of Weights

The initial weights of edges may have a substantial effect on the convergence and error of training. In order to avoid the saturation region (i.e., too small a derivative) of the activation function, the combined input ($\sum_{i=1}^N w_{ij}I_j$) to a node should not be too large. A reasonable choice is to maintain the range of the normalized input. Note that the input normalized according to Eq. 4.2 has the mean of 0 and the standard deviation of 1. The standard deviation of the sum of N uncorrelated samples of a random variable is \sqrt{N} times larger than the standard deviation of the random variable. Hence, each weight may be initialized to a value selected from the range of $[-1/\sqrt{N} \quad 1/\sqrt{N}]$ where N is the number of inputs to a node [37].

4.2.2.3 Gradient Descent Training

In the error back-propagation, the gradient descent training algorithm is employed [38]. The loss function consisting of the CD and LER training errors, i.e., half of the mean square error, is employed and the weights are adjusted such that the cost function decreases most.

The amount of weight adjustment for the link between node i in a layer and node j in the next layer is computed as in Eq. 4.3.

$$\Delta w_{ij} = -\eta \cdot \delta_j \cdot I_i \quad (4.3)$$

where η is the learning rate, δ_j is the error at node j , and I_i is the input to node j from node i .

For an output node j (a node in the output layer), δ_j is computed from the difference between the output O_j and the known (target) output T_j (CD or LER), as shown in Eq. 4.4.

$$\delta_j = (O_j - T_j) \cdot O_j \cdot (1 - O_j) \quad (4.4)$$

For a hidden layer node j (a node in a hidden layer) connected to N nodes in the next layer, δ_j is computed as in Eq. 4.5.

$$\delta_j = O_j \cdot (1 - O_j) \cdot \sum_{k=1}^N w_{jk} \delta_k \quad (4.5)$$

That is, the error at node j is derived from the errors at the nodes in the next layer.

4.2.2.4 Learning Rate

The learning rate sets the relative amount of weight adjustment that is allowed in each weight update. A too large learning rate makes a weight change faster, but possibly oscillate without converging. On the other hand, with a smaller learning rate, a weight tends to converge to a stable value, but the training takes longer and may be possible to be trapped locally. In the NN which generates normalized outputs, i.e., their absolute values <1 , as in this study, the weight updates through the back-propagation can become so small that an effective training is not possible, especially when the number of layers is large [39, 40]. Hence, a proper selection of learning rate is necessary in such cases. In this implementation, the learning rate is determined empirically.

4.2.2.5 Mini-batch

Updating weights for every training sample may cause them to oscillate, hampering their convergence and leading to a relatively large training error. On the other hand, accumulating weight changes for all training samples and reflecting the accumulated changes of weights in one step tend to make the behaviors of weights more stable, but the convergence slower. A reasonable approach to achieving an acceptable training error and speed would be to adopt the concept of mini-batch, i.e., weights are updated after weight changes are accumulated for a subset of training samples (a mini-batch).

The main purpose of employing a mini-batch in this application is to reduce the training time without sacrificing the accuracy substantially [41]. As the mini-batch size increases, the accuracy tends to improve, but the training time increases. Also, once the mini-batch size

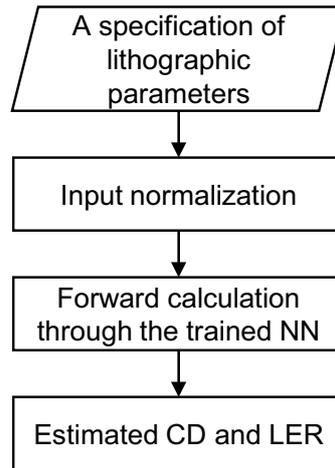


Figure 4.5: Process for estimating CD and LER with one forward calculation using the trained NN.

reaches a certain level, the accuracy does not improve significantly. That is, there exists a trade-off between the training accuracy and time.

Training samples may be stored in a specific order, e.g., one of the lithographic parameters is varied with the others fixed. Taking mini-batches of training samples in such an order is likely to follow a biased training path by the gradient-descent method, deviated from the path leading to the minimal training error. To enhance the possibility of achieving the minimal error, randomly-selected training samples are included in each mini-batch.

4.2.3 Estimation

Once the NN is trained, it is straightforward to use it for the estimation of CD and LER. Given a specification of the lithographic parameters, the forward computation through the NN is carried out and the estimated CD and LER become available at the output of NN without any iteration as shown in the flowchart in Fig. 4.5.

4.3 Results and Discussion

For training and testing the NN, remaining resist profiles for which the CD and LER are known are employed. A number of such remaining resist profiles are generated through

simulation for various combinations of the lithographic parameters. The CD and LER measured from the reference remaining resist profiles are used as the training and test samples.

4.3.1 Reference Remaining Resist Profiles

In creating reference remaining resist profiles, each of the lithographic parameters is varied: the resist thickness from 100 nm to 500 nm with an interval of 100 nm, the beam energy from 10 keV to 50 keV with an interval of 10 keV, the beam diameter of 2 nm, 5 nm, 10 nm, 15 nm and 20 nm. The exposure distribution in the resist is computed using the point spread function generated by the software CASINO [36]. The substrate system is composed of PMMA on Si. For each combination of the lithographic parameters, 50 instances of the 5-layer stochastic PSF are generated. Then, a fast path-based method [34] for simulation of resist-development is used to obtain the remaining resist profile from which the CD and LER are computed.

A line feature is considered where the line-length is 2,000 nm and the linewidth is varied from 50 nm to 150 nm with an interval of 10 nm. The dose is varied such that both under-developed and over-developed remaining resist profiles are obtained in each case. In each reference remaining resist profile, the CD and LER are measured according to the edge location defined in Sec. 4.1 from a single line.

4.3.2 Training and Estimation Errors

The CD training error is defined as the average difference between the trained CD (i.e., CD obtained after the training) and the known CD, and the CD estimation error as the average difference between the estimated CD and known CD as in Eq. 4.6.

$$\epsilon_{CD} = \frac{1}{N} \sum_{i=1}^N |CD_i - CD'_i| \quad (4.6)$$

where N is the number of samples used in the training or estimation, CD_i is the trained or estimated CD, and CD'_i is the known CD.

The LER training and estimation errors are defined similarly as in Eq. 4.7.

$$\epsilon_{LER} = \frac{1}{N} \sum_{i=1}^N |LER_i - LER'_i| \quad (4.7)$$

where LER_i is the trained or estimated LER, and LER'_i is the known LER.

The percent training and estimation errors of the CD and LER are computed as in Eq. 4.8 and Eq. 4.9.

$$\epsilon_{CD,p} = \frac{\frac{1}{N} \sum_{i=1}^N |CD_i - CD'_i|}{\frac{1}{N} \sum_{i=1}^N CD'_i} \times 100\% \quad (4.8)$$

$$\epsilon_{LER,p} = \frac{\frac{1}{N} \sum_{i=1}^N |LER_i - LER'_i|}{\frac{1}{N} \sum_{i=1}^N LER'_i} \times 100\% \quad (4.9)$$

This is to obtain quantifications of percent errors, not too sensitive to certain samples with small CD or LER which may result in large errors.

4.3.3 Tuning of NN

To achieve the best performance in both accuracy and computation time, the structure and parameters of NN need to be optimized. In the training of NN, the loss function is defined as in Eq. 4.10 [42].

$$loss = \frac{1}{2} \left(\sum_{i=1}^N (CD_i - CD'_i)^2 + \sum_{i=1}^N (LER_i - LER'_i)^2 \right) \quad (4.10)$$

The size of mini-batch affects the performance of NN substantially. As shown in Fig. 4.6(a), the LER estimation error is around 10% for the mini-batch size of 5. For the mini-batch size of 20 or larger, the LER estimation error is around 5% or less. On the other hand, the computation time is longer for a larger size of mini-batch. For this application, it turns out that the mini-batch size of 20 is a reasonable choice. The training takes about 50 hours on the PC with Intel Core i5-4210U and 12 Gbytes RAM running at 2.4 GHz when the entire 4,096 samples are used as a batch, but about 500 seconds with the mini-batch of size 20. As

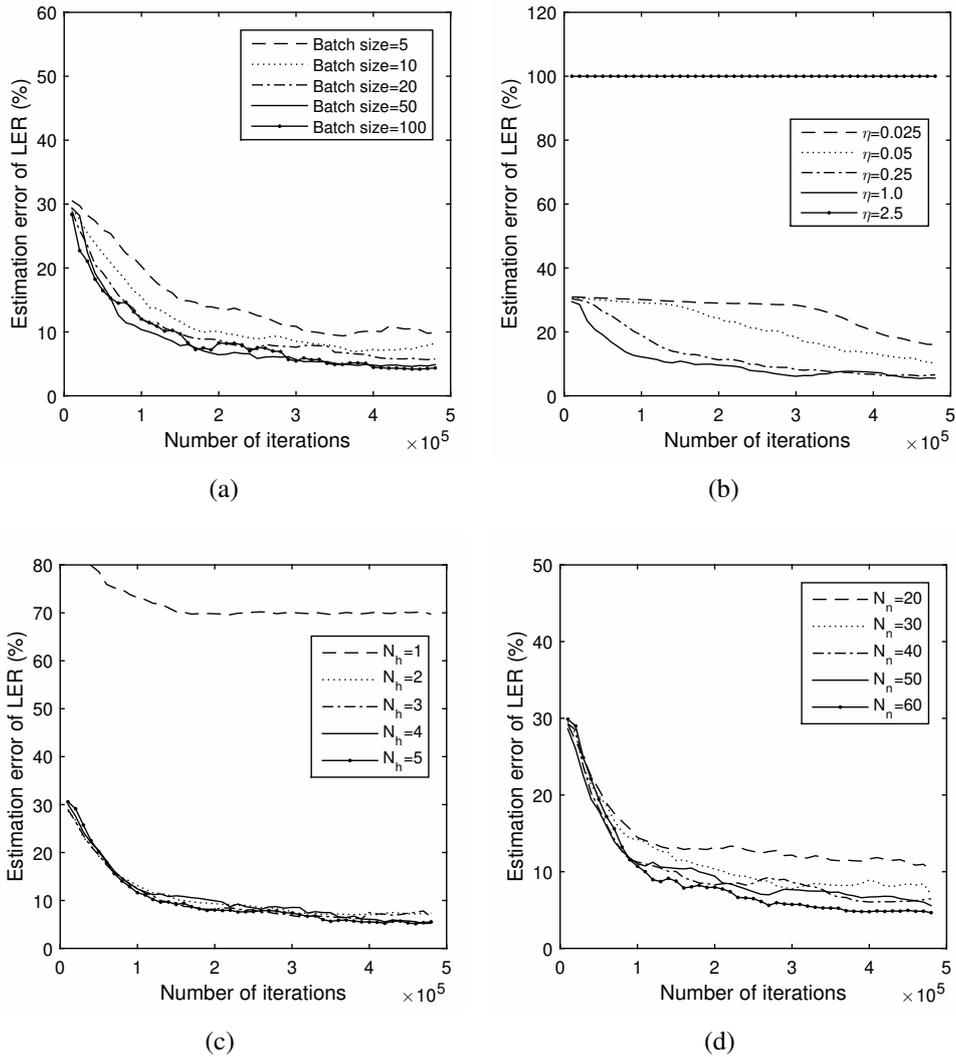


Figure 4.6: LER estimation errors in the training process with varying (a) mini-batch size, (b) learning rate, (c) number of nodes in a hidden layer, and (d) number of nodes in a hidden layer.

explained above, using a mini-batch reduces the training time greatly without sacrificing the accuracy substantially.

The effect of the learning rate on the LER estimation error is shown in Fig. 4.6(b). It is seen that the LER estimation error is larger than 15% for the small learning rate of 0.025 and decreases close to 5% as the learning rate increases to 1.0. A larger learning rate tends to make the convergence faster, however, too large a learning rate such as 2.5 in Fig. 4.6(b) leads to the divergence of the NN. Hence, in this study, the learning rate of 1.0 is used. Though there are fluctuations of the training error in the training process, this learning rate makes a fast training and achieves smaller training/estimation errors for both CD and LER.

The structure of NN may be characterized by the number of hidden layers (N_h) and the number of nodes (neurons) per hidden layer (N_n). In Fig. 4.6(c), the dependency of the LER estimation error on N_h is shown. The LER estimation error is as high as 70% when only a single hidden layer is employed and decreases down to around 5% as N_h increases to 4. Adding more hidden layers beyond 4 does not decrease the LER estimation error further. More nodes in each hidden layer result in a smaller LER estimation error as can be seen in Fig. 4.6(d), but require a longer computation time. The LER estimation error is around 13% for $N_n = 20$, and decreases down to 5% as N_n increases to 50.

4.3.4 Estimation of CD and LER

The optimized NN consists of 5 layers in total, i.e., 3 hidden layers, an input layer and an output layer. There are 5 nodes in the input layer for the 5 lithographic parameters: resist thickness, beam diameter, beam energy, linewidth, and dose. Each hidden layer is composed of 50 nodes. The output layer has 2 nodes for the CD and LER. In the optimized NN, the mini-batch size is set to 20 and the learning rate is set to 1.0. Each lithographic parameter is sampled for the training and estimation as described in Section 4.3.1. In the optimized NN, the linewidths and doses for the estimation are selected to be different from those used in the training. The linewidth is varied from 50 nm to 150 nm with an interval of 20 nm for the training, and from 60 nm to 140 nm with an interval of 20 nm for the estimation.

The optimized NN is trained with a subset of samples such that the training errors are minimized, and the CD and LER are estimated for various combinations of the lithographic parameters using the rest of the samples. The estimation errors are provided for various combinations of resist thickness and beam energy in Table 4.1 where each error is the error averaged over the other three lithographic parameters, i.e., beam diameter, linewidth and dose. The average CD and LER estimation errors for all the cases are 0.68% and 4.12%, respectively. It may be said that the accuracy of estimation is high enough for the practical use. One observation is that the largest estimation errors are obtained for the resist thickness of 500 nm and beam energy of 10 keV. The thicker the resist is and the lower the beam energy is, the broader (less sharp) the point spread function is. Hence, the exposure contrast over the feature boundary is

smaller, which makes the location of the feature edge less stable. That is, the samples for the estimation are likely less consistent with those for the training. This results in larger estimation errors.

Resist thickness (nm)	Beam energy (keV)	ϵ_{CD} (nm)	$\epsilon_{CD,p}$ (%)	ϵ_{LER} (nm)	$\epsilon_{LER,p}$ (%)
100	10	0.41	0.40	0.05	4.39
	20	0.58	0.57	0.07	5.87
	30	0.45	0.44	0.05	4.16
	40	0.39	0.39	0.05	4.09
	50	0.58	0.58	0.07	4.73
200	10	0.65	0.65	0.07	4.18
	20	0.50	0.50	0.04	2.78
	30	0.45	0.45	0.07	4.16
	40	0.69	0.67	0.08	5.11
	50	0.62	0.61	0.05	3.27
300	10	0.62	0.64	0.06	3.35
	20	0.43	0.42	0.07	3.65
	30	0.69	0.68	0.10	4.86
	40	0.40	0.39	0.10	5.08
	50	0.61	0.60	0.09	5.02
400	10	1.08	1.15	0.07	4.07
	20	0.87	0.92	0.10	4.30
	30	0.64	0.65	0.08	3.37
	40	0.55	0.55	0.07	3.19
	50	0.54	0.54	0.09	4.50
500	10	1.79	2.68	0.10	5.87
	20	1.09	1.58	0.08	3.24
	30	0.78	1.07	0.07	2.74
	40	1.17	1.55	0.10	4.30
	50	0.89	1.16	0.10	3.98

Table 4.1: CD and LER estimation errors in nm and percentage for varying resist thickness and beam energy for the substrate system of PMMA on Si: Errors are averaged over the beam diameters of 2, 5, 10, 15 and 20 nm, linewidths of 60, 80, 100, 120 and 140 nm, and normalized doses.

4.4 Reduction of Training Samples

In general, using a larger number of training samples results in a smaller estimation error, however, requires the preparation of more samples and a longer training time. Also, adding more training samples beyond some point may not improve the estimation accuracy. Therefore,

it is desirable to minimize the number of training samples without increasing the estimation error significantly. For each lithographic parameter, the effect of reducing training samples (from the training samples used to obtain the results in Sec. 4.3.4) on the estimation errors is analyzed. The results are provided in Figs. 4.7(a)-4.7(e).

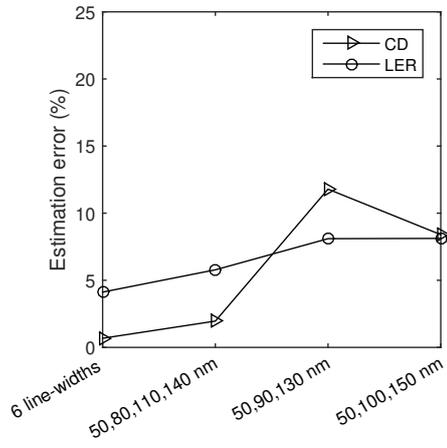
In Fig. 4.7(a), it is observed that when the number of training samples for the linewidth is reduced to 3 from 6, the estimation errors, especially the CD estimation error, is increased significantly. But, when 4 different linewidths are kept including the two linewidths at or close to the two ends of the original range, the increase of the estimation errors is not significant. Hence, the number of linewidths may be reduced to 4 at the expense of a minor increase in the estimation errors. Both CD and LER vary with the dose rapidly when the dose level is relatively low, however, changes vary slowly once the dose exceeds a certain level. Therefore, as long as the dose is sampled sufficiently in the fast-varying region, a significant increase of the estimation errors can be avoided as indicated in Fig. 4.7(b). Depending on the acceptable accuracy of estimation, one may readily reduce the number of training samples down to 4.

The effect of the change in the beam diameter on the sharpness of PSF and in turn on the CD and LER is smaller when the beam diameter is relatively large. Hence, removing one of the large beam diameters does not increase the estimation errors much as shown in Fig. 4.7(c).

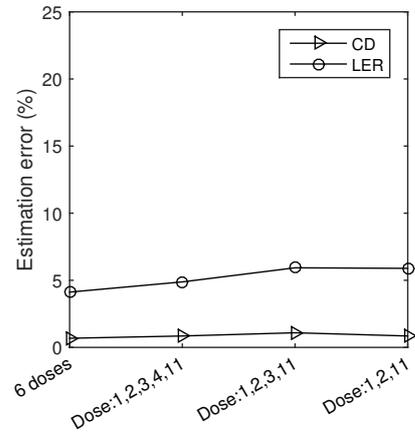
It is seen in Fig. 4.7(d) and Fig. 4.7(e) that the removal of any training sample for the resist thickness and beam energy increases the estimation errors, the LER estimation error in particular, significantly. Therefore, it is advisable to maintain the sampling intervals of 100 nm and 10 keV for the resist thickness and beam energy, respectively, which are already sufficiently coarse.

4.5 Summary

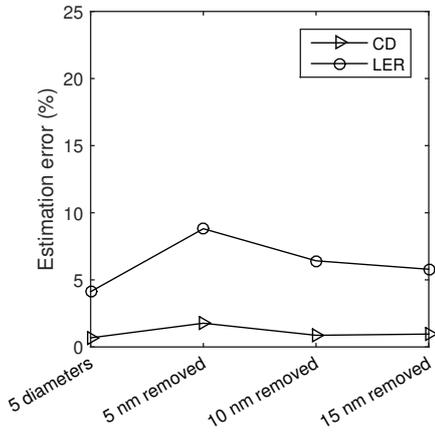
The line edge roughness becomes a more critical issue as the CD decreases down to nanometers. An accurate estimation of CD and LER is an important task by itself and also a necessary step toward the minimization of the proximity effect and LER in the e-beam lithography. The estimation of the CD and LER for the e-beam lithographic process using a NN is investigated.



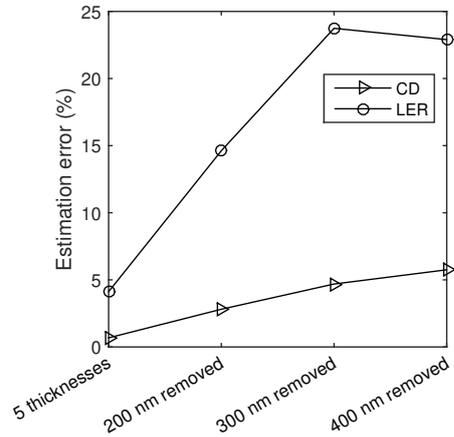
(a)



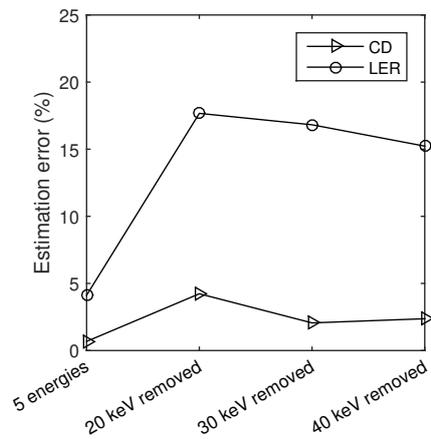
(b)



(c)



(d)



(e)

Figure 4.7: CD and LER estimation errors with a reduced number of training samples for (a) linewidth, (b) normalized dose, (c) beam diameter, (d) resist thickness, and (e) beam energy.

A number of remaining resist profiles are generated using the exposure and development simulation programs and from each profile, the CD and LER are measured. The NN is trained and tuned by the measured CD and LER. Given a line feature with the specification of e-beam lithographic parameters, the CD and LER are estimated by the NN. Through an extensive performance analysis, it has been shown that, with proper network structure and training, the NN can accurately estimate the CD and LER for wide ranges of the lithographic parameters. Using a NN does not require a direct modeling of the e-beam lithographic process and therefore avoids the involvement of the process components. Also, an addition or deletion of certain lithographic parameter can be readily done in the training and estimation.

A large number of training samples is required for an acceptable estimation accuracy. However, it has been shown that, for this application to the e-beam lithography, it is possible to reduce the sample size substantially without degrading the accuracy much. As training samples from experiments are accumulated, the accuracy of CD and LER estimation by a NN would be improved.

Chapter 5

Verification: Proximity Effect Correction

The accuracy of the modeling results in Chapter 3 is further verified through the proximity effect correction. A dose modification with the reduction of the feature width is performed based on the modeled LSF and noise. Then the CD error and LER are analyzed, and compared with those by the 2-D approach.

5.1 Width Reduction

In this study, for the correction of the proximity effect, the width of a feature to be exposed is reduced from the target (designed) width, W , by $2\Delta W$ where ΔW is the width reduction on each side of the feature as illustrated in Fig. 5.1, and the dose to be given to the reduced feature width is computed such that the CD error is minimized through an iterative procedure shown in Fig. 5.2).

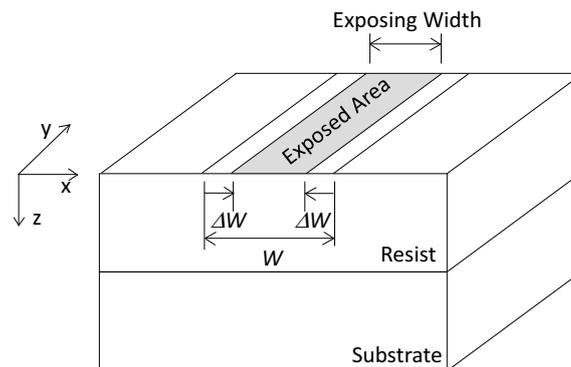


Figure 5.1: The feature width is reduced by ΔW each side and the dose is determined for the reduced width.

When the CD error is minimized with $2\Delta W > 0$, the boundary of a written feature would be outside the exposed area. The LER tends to be much smaller outside than inside an exposed

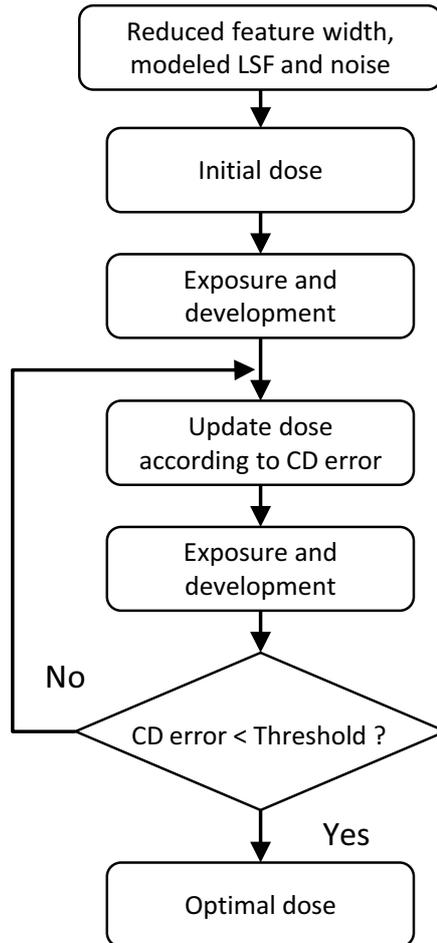


Figure 5.2: The optimal dose is determined when the CD error (calculated with the modeled LSF, noise parameters, and the reduced feature linewidth to be exposed) is below a threshold, which is set to a small value.

area. This is due to the fact that the exposure level is lower and the absolute fluctuation of exposure is smaller in the unexposed area. Therefore, the reduction of feature width helps achieving a smaller LER.

A proper ΔW can be larger for a broader LSF, i.e., proportional to the standard deviation of fitted LSF from modeling. However, the width reduction also decreases the exposure contrast over the feature boundary, which is not desirable from the viewpoint of process stability. Also, when ΔW is too large, an impractically high dose would be required or the feature may not be developed to its full width. Hence, an upper limit of ΔW is necessary, and the ΔW is

determined empirically as in Eq. 5.1.

$$\Delta W = \min\{\lfloor 2\sigma + \frac{1}{2} \rfloor, 7\} \quad (5.1)$$

where σ is the standard deviation of the fitted main lobe of the modeled deterministic LSF.

5.2 Dose Correction

With the reduced width of a feature, the optimal dose which minimizes the CD error is derived iteratively using a bisection procedure. In this derivation, the modeled deterministic LSF and noise are employed. The CD error is defined as the difference between the target CD and the average CD on the top and bottom layers.

In this study, the 3-D modeling is carried out for the reference cases for which the stochastic LSFs are known (generated through a Monte Carlo simulation). In each case, the optimal dose is used with the respective stochastic LSF to compute the exposure distribution and obtain the CD and LER following the development simulation [34].

5.3 Results and Discussion

The results from the dose correction, i.e., the CD error and LER, are provided in Tables 5.1-5.5 for the linewidths of 40 nm and 70 nm. As in the modeling, various thicknesses of resist and levels of beam energy are considered. Overall, it is seen that the CD error and LER achieved by the 3-D model are significantly smaller than those by the 2-D model. The improvement by the 3-D model is larger when the resist is thicker or the beam energy is lower. These are the cases where the exposure distribution varies more with the resist layer and therefore the 3-D model can describe the exposure distribution in the resist more accurately. Note that the 2-D model does not take into account the layer-dependent exposure variation.

The optimal dose computed using the 2-D model is lower than that using the 3-D model as can be seen in the tables. This makes the resist under-developed, especially at the bottom layer, leading a negative CD error. In the cases where the resist is thick and the beam energy is low, it can happen that the bottom layer is not developed at all, especially for a narrow feature. For

a narrower feature, the exposure distribution is more layer-dependent. In such a case, the LER cannot be evaluated properly from the remaining resist profile (see “NA” in Tables 5.3-5.5). This occurs in 11 cases of the 2-D results, compared to only one case of the 3-D results.

These verification results further confirm the advantages of and the need for the 3-D modeling.

Target width (nm)	Beam energy (keV)	ΔW (nm)		Optimal dose		CD error (nm)		LER (nm)	
		3D	2D	3D	2D	3D	2D	3D	2D
40	10	3	3	0.438	0.414	-0.17	-1.96	0.31	0.49
	20	3	2	0.825	0.731	-0.40	-1.00	0.20	0.49
	30	2	2	1.063	1.047	-0.03	-0.40	0.23	0.30
	40	2	2	1.190	1.167	0.05	-0.36	0.20	0.27
	50	2	2	1.375	1.383	-0.42	-0.27	0.27	0.24
70	10	3	3	0.432	0.403	-0.01	-2.48	0.35	0.63
	20	3	2	0.820	0.726	-0.36	-1.22	0.19	0.55
	30	2	2	1.061	1.047	-0.05	-0.38	0.24	0.30
	40	2	2	1.195	1.168	0.16	-0.29	0.20	0.27
	50	2	2	1.392	1.379	-0.11	-0.31	0.21	0.25

Table 5.1: Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 100 nm, and the target CDs of 40 nm and 70 nm.

5.4 Summary

The accuracy verification is realized through the proximity effect correction based on the modeling results on substrate systems with varying resist thickness and beam energy. According to the comparison between the results of the 2-D and 3-D approaches, the 3-D approach demonstrates a clear advantage over the 2-D approach in terms of the CD error and LER. And this tendency becomes more obvious for a substrate system with a broader LSF.

Target width (nm)	Beam energy (keV)	ΔW (nm)		Optimal dose		CD error (nm)		LER (nm)	
		3D	2D	3D	2D	3D	2D	3D	2D
40	10	6	6	0.839	0.743	0.81	-7.94	0.34	0.80
	20	4	4	1.330	1.209	0.36	-4.79	0.21	0.68
	30	4	3	1.854	1.690	-0.75	-2.51	0.15	0.50
	40	3	3	1.979	1.871	0.09	-1.82	0.14	0.31
	50	2	3	2.203	2.246	0.01	-1.15	0.30	0.17
70	10	6	6	0.727	0.702	-2.75	-5.69	0.49	0.65
	20	4	4	1.275	1.195	-0.57	-4.54	0.26	0.68
	30	4	3	1.819	1.675	-1.07	-3.12	0.16	0.62
	40	3	3	1.950	1.861	-0.23	-2.05	0.15	0.35
	50	2	3	2.199	2.238	0.03	-1.18	0.31	0.19

Table 5.2: Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 200 nm, and the target CDs of 40 nm and 70 nm.

Target width (nm)	Beam energy (keV)	ΔW (nm)		Optimal dose		CD error (nm)		LER (nm)	
		3D	2D	3D	2D	3D	2D	3D	2D
40	10	7	7	1.328	1.078	2.44	-20.99	0.65	NA
	20	3	4	1.690	1.592	1.77	-10.34	0.55	NA
	30	4	3	2.328	2.213	-0.19	-2.40	0.31	0.76
	40	3	3	2.500	2.416	0.25	-3.22	0.28	0.88
	50	4	4	2.953	2.839	-1.27	-3.53	0.20	0.43
70	10	7	7	1.006	0.929	-0.96	-16.43	0.83	1.74
	20	3	4	1.583	1.553	-0.54	-5.83	0.76	1.21
	30	4	3	2.247	2.167	-1.24	-3.04	0.34	0.88
	40	3	3	2.462	2.387	-0.18	-4.05	0.29	0.99
	50	4	4	2.870	2.815	-2.30	-3.72	0.24	0.45

Table 5.3: Estimation of the CD and LER through the PEC in terms of the 2-D and 3-D modeling results for the resist thickness of 300 nm, and the target CDs of 40 nm and 70 nm. “NA” indicates that the resist is not fully developed.

Target width (nm)	Beam energy (keV)	ΔW (nm)		Optimal dose		CD error (nm)		LER (nm)	
		3D	2D	3D	2D	3D	2D	3D	2D
40	10	7	7	2.021	1.535	-0.01	-18.23	1.05	NA
	20	7	5	2.469	2.072	0.15	-20.07	0.49	NA
	30	4	6	2.846	2.795	-0.91	-17.45	0.47	NA
	40	3	4	3.065	2.952	1.73	-7.53	0.32	1.39
	50	4	3	3.555	3.378	-0.55	-6.24	0.26	1.51
70	10	7	7	1.357	1.196	0.32	-33.65	0.96	NA
	20	7	5	2.086	1.919	-0.17	-11.81	0.45	1.35
	30	4	6	2.715	2.665	-0.36	-7.79	0.42	0.72
	40	3	4	2.937	2.895	-0.16	-4.75	0.48	0.83
	50	4	3	3.480	3.346	-0.76	-4.70	0.28	1.14

Table 5.4: Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 400 nm, and the target CDs of 40 nm and 70 nm. “NA” indicates that the resist is not fully developed.

Target width (nm)	Beam energy (keV)	ΔW (nm)		Optimal dose		CD error (nm)		LER (nm)	
		3D	2D	3D	2D	3D	2D	3D	2D
40	10	7	7	2.859	1.534	-12.96	-16.61	NA	NA
	20	7	7	3.282	2.959	3.14	-20.29	0.54	NA
	30	6	5	3.651	3.258	-1.98	-20.20	0.57	NA
	40	6	6	3.906	3.578	1.69	-5.57	0.29	NA
	50	3	5	4.097	4.031	1.13	-9.66	0.37	1.40
70	10	7	7	1.792	1.509	-5.32	-30.81	1.20	NA
	20	7	7	2.513	2.352	2.01	-15.30	0.54	1.46
	30	6	5	3.227	3.045	-0.28	-13.18	0.49	1.70
	40	6	6	3.475	3.353	-1.83	-7.75	0.40	0.94
	50	3	5	3.916	3.873	-1.04	-9.50	0.60	1.19

Table 5.5: Estimation of the CD and LER through the PEC based on the 2-D and 3-D modeling results for the resist thickness of 500 nm, and the target CDs of 40 nm and 70 nm. “NA” indicates that the resist is not fully developed.

Chapter 6

Application to Massively-Parallel E-Beam System

New e-beam lithography systems with a large number of beams, i.e., MPES, were recently developed to improve writing throughput significantly. The 3-D modeling method is applied to the MPES for verification with SEM images.

6.1 Massively-Parallel E-Beam System

The low throughput is a main drawback of the single-beam lithography system. Massively-parallel e-beam system is one of the new technologies, which use a large array of beams simultaneously to increase the writing speed significantly [30, 31, 32, 33]. Fig. 6.1 shows the principle of a mask exposure tool with a massively-parallel e-beam system using an aperture plate system (APS) [31].

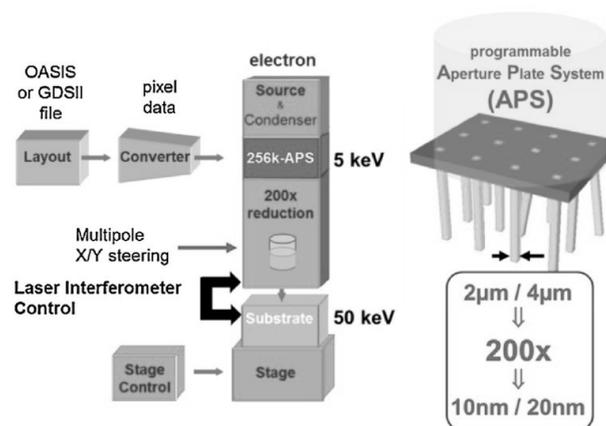


Figure 6.1: Principle of a massively-parallel e-beam system with an APS. Each of 512×512 apertures controls the availability of electrons within a certain size, $10 \text{ nm} \times 10 \text{ nm}$ or $20 \text{ nm} \times 20 \text{ nm}$, onto the resist.

In the current massively-parallel e-beam system, the size of each beam is relatively large, compared to that in a single-beam system. Therefore, the energy deposition profile due to a beam is described by a transfer function (TF) rather than a PSF.

6.2 Modifications of Modeling

To enable more general application of modeling and consider the characteristic of a MPES, the 3-D modeling procedures described in Chapter 3 have been modified.

6.2.1 PSF and TF

The 1-D deterministic PSF, instead of LSF, is modeled at the middle layer of the resist. Then, the TF is computed from the PSF. The main lobe of the 1-D PSF is guided by a Gaussian curve since it typically follows the Gaussian shape. PSFs at other layers are derived from the PSF at the middle layer.

The main lobe of PSF at each layer is fitted to a Gaussian curve of which the standard deviation is used to characterize the PSF. The ratio of standard deviation of PSF at a layer (z) to that of the reference layer (middle layer) is denoted by $ratio_\sigma(z)$ as in Eq. 6.1.

$$ratio_\sigma(z) = \frac{\sigma(z)}{\sigma(z_0)} \quad (6.1)$$

where $\sigma(z)$ and $\sigma(z_0)$ are the standard deviations of PSFs at a layer (z) and reference layer, respectively.

Similarly, the ratio of the amplitude of PSF at a layer (z) to that of the reference layer is denoted by $ratio_A(z)$ as in Eq. 6.2.

$$ratio_A(z) = \frac{A(z)}{A(z_0)} \quad (6.2)$$

where $A(z)$ and $A(z_0)$ are the amplitudes of PSFs at a layer (z) and reference layer, respectively.

The amplitude and standard deviation of PSF at a layer (z) can be derived as $A(z) = A(z_0) \cdot ratio_A(z)$ and $\sigma(z) = \sigma(z_0) \cdot ratio_\sigma(z)$, respectively. Then, the main lobe of the 1-D deterministic PSF at a layer (z), $psf(x, z)$, can be expressed as in Eq. 6.3.

$$psf(x, z) = A(z) \cdot \exp\left(-\frac{x^2}{2(\sigma(z))^2}\right) + R(x, z) \quad (6.3)$$

where $R(x, z)$ is a random value allowing the deviation from a Gaussian function for a more realistic PDF curve.

The 2-D PSF at a layer (z), $psf_{2D}(x, y, z)$, is derived by rotating the 1-D PSF around its center. The deterministic TF at a layer (z), $TF(x, y, z)$, can be obtained from the convolution between the 2-D PSF and the ideal TF, $TF_{ideal}(x, y)$, as in Eq. 6.4 and Eq. 6.5.

$$TF(x, y, z) = \int_{y'} \int_{x'} TF_{ideal}(x - x', y - y') psf_{2D}(x', y', z) dx' dy' \quad (6.4)$$

$$TF_{ideal}(x, y) = \begin{cases} 1 & \text{when } -\frac{a}{2} \leq x \leq \frac{a}{2} \text{ and } -\frac{a}{2} \leq y \leq \frac{a}{2} \\ 0 & \text{otherwise} \end{cases} \quad (6.5)$$

where the beam size at the top surface of resist is $a \times a$.

6.2.2 Noise Model

To obtain the stochastic exposure, the noise (exposure fluctuation) is added to the TF, instead of LSF, for a MPES.

The stochastic TF at a layer (z), $TF_s(x, y, z)$, may be expressed as in Eq. 6.6.

$$TF_s(x, y, z) = TF(x, y, z) + e_n(x, y, z) \quad (6.6)$$

where $e_n(x, y, z)$ is the noise at a layer (z).

The noise added on the TF is modeled as in Eq. 6.7.

$$e_n(x, y, z) = rand \cdot e_0(z) \left(\frac{TF(x, y, z)}{TF(0, 0, z)} \right)^\alpha \quad (6.7)$$

where $e_0(z)$ is the level of the exposure fluctuation, $rand$ is a random number of the normal distribution $N(0, 1)$ and $TF(0, 0, z)$ is the maximum value of the deterministic TF at a layer (z).

The stochastic TF is employed to calculate the exposure distribution for a feature as in Eq. 6.8.

$$e(x, y, z) = \int_{y'} \int_{x'} d(x - x', y - y') \cdot TF_s(x', y', z) dx' dy' \quad (6.8)$$

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

6.2.3 Cost Function and Constraints

From SEM images, the outer and inner boundaries of a feature may be detected as shown in Chapter 2. However, without knowing the shape of sidewall, it is often uncertain that the outer and inner boundaries actually correspond to the feature boundaries at the top and bottom layers of resist, respectively. Alternatively, the feature boundaries may be defined by the peak brightness in the edge region (See Sec. 2.3). In this study, the CD measured from the peak-based feature boundaries, which can be considered to be the middle CD (the CD at the middle layer), is adopted in the formulation of cost function, as shown in Eq. 6.9.

$$\epsilon = \frac{1}{N} \sum_{i=1}^N |CD_i - CD'_i| \quad (6.9)$$

where N is the number of middle CDs from SEM images, and CD_i and CD'_i are the modeled and measured middle CDs, respectively.

While the inner CD may not be the bottom CD, it is clear that the latter can not be smaller than the former. Hence, a constraint that the (modeled) bottom CD is not less than the inner CD is applied to the modeling.

Also, a constraint on the total energy is added so that the exposure distribution inside the feature is close to the middle section of the conversion formula. That is, in each iteration, the 2-D PSF is scaled such that the total energy of 2-D PSF does not change. Using a fixed energy of PSF facilitates the convergence of the modeling for consistent results.

6.3 Results and Discussion

A L/S pattern was exposed by a MPES at Samsung, with the normalized dose ranged from 0.825 to 1.247 and developed to obtain a number of SEM images. Most of the lithographic pattern parameters are confidential. Hence, based on the provided hints, it is assumed that the resist thickness is 80 nm, both linewidth and space are 100 nm and the developing time is 30 seconds. The beam energy is known to be 50 keV.

SEM images are of 1024×1024 pixels and the size of a pixel is calculated as 0.5634 nm. In each SEM image, 4 sidewalls of 2 complete line features are included (See Figs. 6.2 and 6.3).

6.3.1 Modeling

The SEM images with 3 different normalized doses employed for the modeling are shown in Fig. 6.2, and the CD and LER measured from them in Table 6.1. It can be seen that the middle and inner CDs increase with the normalized dose, while the LERs fluctuate within a small range (from 0.38 nm to 0.44 nm).

The modeling results are provided in Table 6.2. It can be seen that the modeled middle CDs are closely matched with the ones measured from the SEM images and modeled middle LERs are close to the measured ones. On the other hand, the modeled bottom CDs are larger than the measured inner CDs. This is most probably due to the fact that the measured inner CD may not be the actual bottom CD and the constraint on the bottom CD is applied in the modeling, in addition to the inherent modeling error. The modeled bottom LER shows a larger variation among the doses, compared to the measured inner LER. This may be explained similarly as for the bottom/inner CDs.

Normalized dose	Middle CD (nm)	Inner CD (nm)	Middle LER (nm)	Inner LER (nm)
0.825	103.35	96.95	0.44	0.44
0.998	105.81	99.29	0.39	0.38
1.247	108.31	102.05	0.43	0.44

Table 6.1: CD and LER measured in SEM images from the MPES for modeling with varied normalized doses.

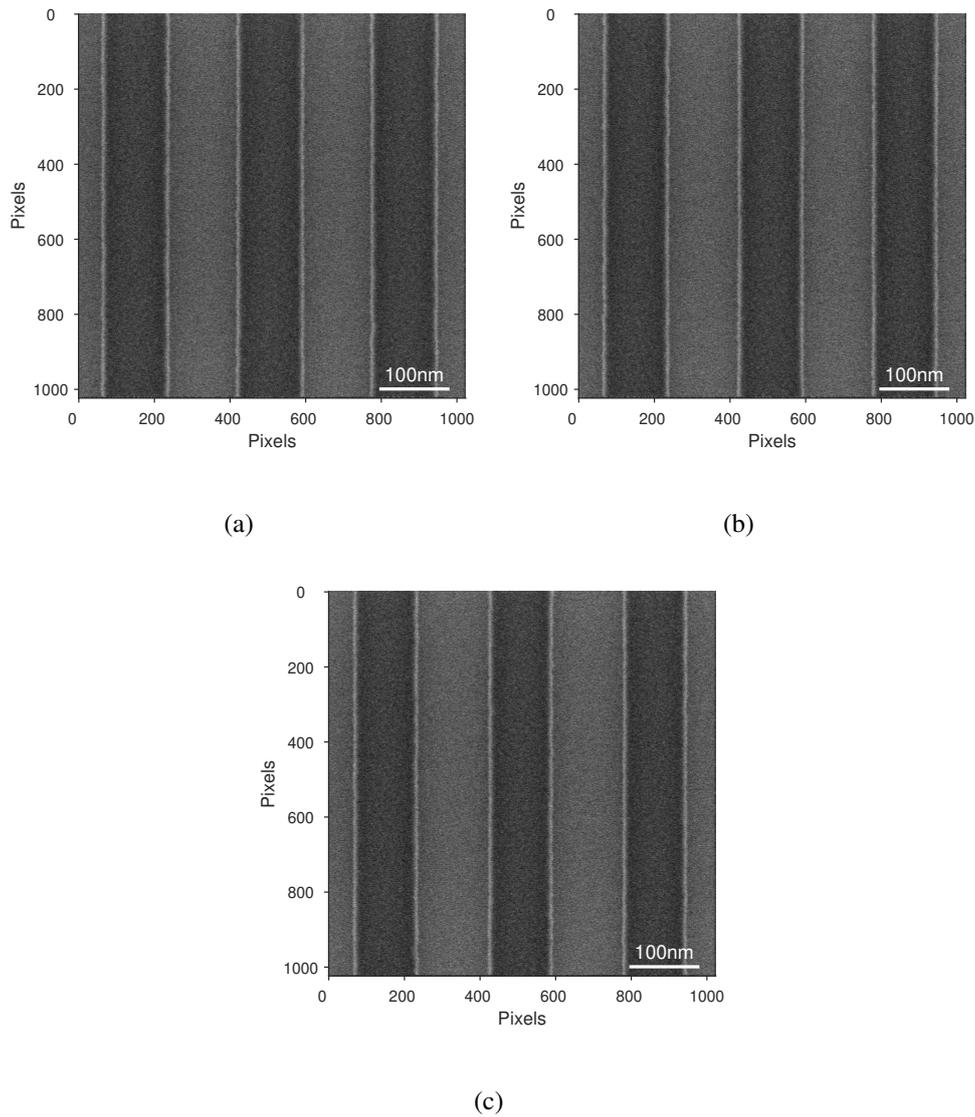


Figure 6.2: SEM images from the MPES with normalized dose of (a) 0.825, (b) 0.998, and (c) 1.247.

Normalized dose	Middle CD (nm)	Bottom CD (nm)	Middle LER (nm)	Bottom LER (nm)
0.825	103.34	99.26	0.36	0.71
0.998	105.83	104.80	0.37	0.40
1.247	108.31	107.69	0.38	0.28

Table 6.2: Modeled CD and LER according to the information in SEM images from a MPES.

Normalized dose	Middle CD (nm)	Inner CD (nm)	Middle LER (nm)	Inner LER (nm)
0.901	104.38	98.01	0.46	0.45
0.945	105.39	98.01	0.40	0.45
1.113	107.00	100.69	0.45	0.44

Table 6.3: CD and LER measured in SEM images from the MPES for estimation.

6.3.2 Estimation

The SEM images with varied normalized doses different from those for the modeling are used for the estimation and are shown in Fig. 6.3. The CD and LER measured from these SEM images are provided in Table 6.3.

Based on the modeled e-beam lithographic process, the resist profile is obtained for each dose through the simulation. The CD and LER are estimated (calculated) from the resist profiles, which are provided in Table 6.4. A good match between the measurement and estimation can be seen for the middle CD and LERs. As in the modeling, the estimated bottom CD is larger than the measurement inner CD.

Normalized dose	Middle CD (nm)	Bottom CD (nm)	Middle LER (nm)	Bottom LER (nm)
0.901	104.66	102.70	0.36	0.41
0.945	105.50	103.71	0.32	0.40
1.113	106.93	106.41	0.28	0.41

Table 6.4: CD and LER estimated for different normalized doses.

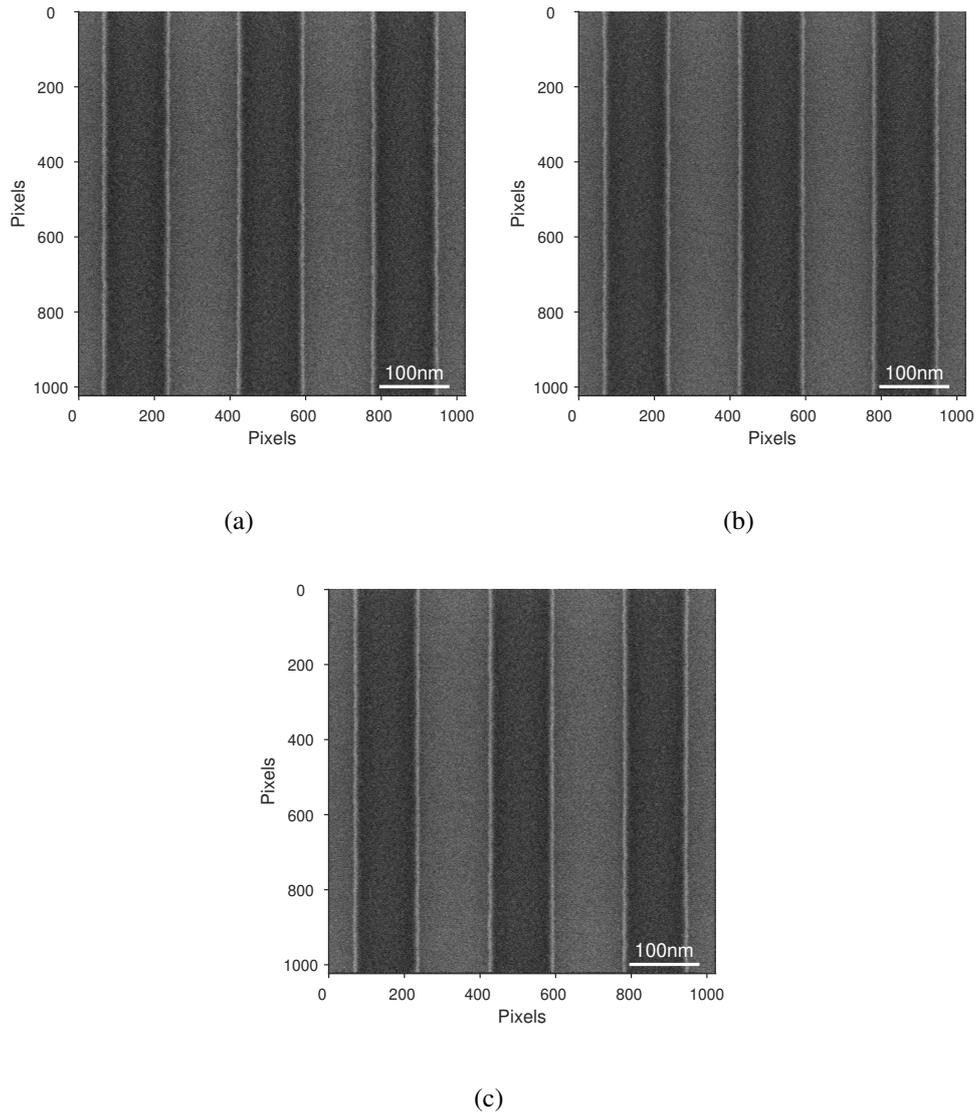


Figure 6.3: SEM images from the MPES with normalized dose of (a) 0.901, (b) 0.945, and (c) 1.113.

6.4 Summary

The 3-D modeling approach has been tested with SEM images obtained using a MPES, with some of the modeling details modified. The modeling is carried out first, of which results are used in the estimation of CD and LER for different doses. It may be said that the overall results indicate a successful accuracy verification of the 3-D modeling. Nevertheless, it is to be mentioned that the bottom CD shows a non-negligible deviation from the measured inner CD in both modeling and estimation.

It should be pointed out that, in the application to a MPES, the PSF is modeled, instead of the LSF, which makes the applicability of the modified 3-D modeling wider.

Chapter 7

Concluding Remarks

This dissertation describes a method to measure the CD and LER from SEM images, two methods to estimate the CD and LER, i.e., modeling the e-beam lithographic process and utilizing a NN, and the dose correction through the PEC for the e-beam lithographic process, for the single beam system and the massively-parallel e-beam system.

A practical method to design a noise filter for accurate measurement of CD and LER on SEM images is developed. The noise characteristics are extracted from an SEM image to determine the sharpness of Gaussian filter for each image. The cut-off frequency is employed to compute the sharpness and the size of the 2-D spatial low-pass filter, which is used to filter the noise in an SEM image to calculate the CD and LER. The results demonstrate that the measured CD and LER with the noise filtering are closely matched with the known ones in SEM images.

The 3-D approach, improved over the previous 2-D approach, is described to model the e-beam lithographic process based on the extracted information from SEM images. The 3-D modeling approach provides a more accurate model to the realistic process with the introduction of LR. In the modeling, the new method of perturbation on the LSF facilitates the convergence. The results demonstrate that the derived LSFs follow the expected tendency for the different sharpness and the 3-D approach has an advantage over the 2-D method. Also, a new method to determine the optimal ΔW , adaptive to different substrate systems, is introduced based on the sharpness of the modeled LSF. The verification through the PEC using the modeled LSF and noise characteristics with the 3-D approach demonstrates a high performance over the 2-D approach, with a smaller CD error and LER.

The explicit modeling of the e-beam lithographic process is avoided by using the NN approach for the CD and LER estimation. The tuned multi-layer NN is trained with EBP and the optimal weights for each node are obtained with the cost function minimized. The results demonstrate that the NN can provide the accurate estimation of CD and LER after training. Also, the investigation of possibility to reduce the number of training samples with the acceptable estimation accuracy is carried out.

The modified 3-D modeling approach is further tested with SEM images from a MPES. The results of modeling and estimation show a close match between the modeled or estimated middle CD and the measured one from an SEM image. This confirms the accuracy of the 3-D modeling approach.

References

- [1] A. A. Tseng, K. Chen, C. D. Chen, and K. J. Ma, "Electron beam lithography in nanoscale fabrication: recent development," *IEEE Transactions on electronics packaging manufacturing*, vol. 26, no. 2, pp. 141–149, 2003.
- [2] R. Murali, D. K. Brown, K. P. Martin, and J. D. Meindl, "Process optimization and proximity effect correction for gray scale e-beam lithography," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 24, no. 6, pp. 2936–2939, 2006.
- [3] M. J. Burek and J. R. Greer, "Fabrication and microstructure control of nanoscale mechanical testing specimens via electron beam lithography and electroplating," *Nano letters*, vol. 10, no. 1, pp. 69–76, 2010.
- [4] Q. Dai, S.-Y. Lee, S. Lee, B.-G. Kim, and H.-K. Cho, "Estimation of resist profile for line/space patterns using layer-based exposure modeling in electron-beam lithography," *Microelectronic engineering*, vol. 88, no. 6, pp. 902–908, 2011.
- [5] W. Chen and H. Ahmed, "Fabrication of 5-7 nm wide etched lines in silicon using 100 keV electron-beam lithography and polymethylmethacrylate resist," *Applied physics letters*, vol. 62, no. 13, pp. 1499–1501, 1993.
- [6] F. Hu and S.-Y. Lee, "Dose control for fabrication of grayscale structures using a single step electron-beam lithographic process," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 21, no. 6, pp. 2672–2679, 2003.

- [7] Q. Dai, S.-Y. Lee, S.-H. Lee, B.-G. Kim, and H.-K. Cho, “Three-dimensional proximity effect correction for large-scale uniform patterns,” *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 29, no. 6, p. 06F314, 2011.
- [8] G. P. Watson, L. A. Fetter, and J. A. Liddle, “Dose modification proximity effect correction scheme with inherent forward scattering corrections,” *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 15, no. 6, pp. 2309–2312, 1997.
- [9] R. Rau, J. H. McClellan, and T. J. Drabik, “Proximity effect correction for nanolithography,” *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 14, no. 4, pp. 2445–2455, 1996.
- [10] X. Zhao, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, and C.-U. Jeon, “Minimization of line edge roughness and critical dimension error in electron-beam lithography,” *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 32, no. 6, p. 06F505, 2014.
- [11] S.-Y. Lee and B. D. Cook, “Pyramid-a hierarchical, rule-based approach toward proximity effect correction. i. exposure estimation,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 1, pp. 108–116, 1998.
- [12] C. Shin, “Line edge roughness (ler),” in *Variation-Aware Advanced CMOS Devices and SRAM*, pp. 19–35, Springer, 2016.
- [13] T. Linton, M. Chandhok, B. J. Rice, and G. Schrom, “Determination of the line edge roughness specification for 34 nm devices,” in *Digest. International Electron Devices Meeting*, pp. 303–306, 2002.
- [14] T. E. A. Khan, S. Sheena, and T. A. Shahul Hameed, “An investigation on the effect of ler on threshold voltage and on current of soi, bulk and gaa finfets,” in *2018 International*

Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), pp. 1–4, 2018.

- [15] C. A. Mack, “Line-edge roughness and the impact of stochastic processes on lithography scaling for moore’s law,” in *Photonic Innovations and Solutions for Complex Environments and Systems (PISCES) II*, vol. 9189, p. 91890D, International Society for Optics and Photonics, 2014.
- [16] M. Okada and S. Matsui, “High-resolution line and space pattern fabrication by electron beam lithography using neb-22 resist,” *Japanese Journal of Applied Physics*, vol. 54, no. 11, p. 118004, 2015.
- [17] S. Babin, K. Bay, and J. J. Hwu, “Application of analytic scanning electron microscopy to critical dimensions metrology at nanometer scale,” *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 28, no. 6, pp. C6H1–C6H5, 2010.
- [18] A. Yamaguchi, R. Steffen, H. Kawada, and T. Iizumi, “Bias-free measurement of LER/LWR with low damage by CD-SEM,” in *Metrology, Inspection, and Process Control for Microlithography XX* (C. N. Archie, ed.), vol. 6152, pp. 829 – 836, International Society for Optics and Photonics, SPIE, 2006.
- [19] B. D. Bunday, M. Bishop, D. W. M. Jr., J. S. Villarrubia, A. E. Vladar, R. Dixon, T. V. Vorburger, N. G. Orji, and J. A. Allgair, “Determination of optimal parameters for CD-SEM measurement of line-edge roughness,” in *Metrology, Inspection, and Process Control for Microlithography XVIII* (R. M. Silver, ed.), vol. 5375, pp. 515 – 533, International Society for Optics and Photonics, SPIE, 2004.
- [20] A. Yamaguchi, R. Tsuchiya, H. Fukuda, O. Komuro, H. Kawada, and T. Iizumi, “Characterization of line-edge roughness in resist patterns and estimations of its effect on device performance,” in *Metrology, Inspection, and Process Control for Microlithography XVII*, vol. 5038, pp. 689–698, International Society for Optics and Photonics, 2003.

- [21] M. Yoshizawa and S. Moriya, "Study of the acid-diffusion effect on line edge roughness using the edge roughness evaluation method," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 20, no. 4, pp. 1342–1347, 2002.
- [22] C. A. Mack and B. D. Bunday, "Using the analytical linescan model for SEM metrology," in *Metrology, Inspection, and Process Control for Microlithography XXXI* (M. I. Sanchez, ed.), vol. 10145, pp. 564 – 571, International Society for Optics and Photonics, SPIE, 2017.
- [23] G. Patsis, N. Tsirikas, D. Drygiannakis, and I. Raptis, "Simulation of shot noise effect on cd and ler of electron-beam lithography in 32 nm designs," *Microelectronic engineering*, vol. 87, no. 5-8, pp. 1575–1578, 2010.
- [24] Z. Yu, L. Chen, W. Wu, H. Ge, and S. Y. Chou, "Fabrication of nanoscale gratings with reduced line edge roughness using nanoimprint lithography," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 21, no. 5, pp. 2089–2092, 2003.
- [25] Y. Ma, G. Tsvid, and F. Cerrina, "Line edge roughness of sub-100 nm dense and isolated features: Experimental study," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 21, no. 6, pp. 3124–3130, 2003.
- [26] R. Guo, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, C.-U. Jeon, B.-G. Kim, and H. Cho, "Derivation of line edge roughness based on analytic model of stochastic exposure distribution," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 31, no. 6, p. 06F408, 2013.
- [27] G. Patsis, V. Constantoudis, A. Tserepi, E. Gogolides, and G. Grozev, "Quantification of line-edge roughness of photoresists. i. a comparison between off-line and on-line analysis of top-down scanning electron microscopy images," *Journal of Vacuum Science &*

- Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 21, no. 3, pp. 1008–1018, 2003.
- [28] T. Verduin, P. Kruit, and C. W. Hagen, “Determination of line edge roughness in low-dose top-down scanning electron microscopy images,” *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 13, no. 3, p. 033009, 2014.
- [29] R. Guo, S.-Y. Lee, J. Choi, S.-H. Park, I.-K. Shin, and C.-U. Jeon, “Practical approach to modeling e-beam lithographic process from sem images for minimization of line edge roughness and critical dimension error,” *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 34, no. 1, p. 011601, 2016.
- [30] H. Matsumoto, H. Yamashita, T. Tamura, and K. Ohtoshi, “Multi-beam mask writer mbm-1000,” in *Photomask Japan 2017: XXIV Symposium on Photomask and Next-Generation Lithography Mask Technology*, vol. 10454, p. 104540E, International Society for Optics and Photonics, 2017.
- [31] C. Klein, H. Loeschner, and E. Platzgummer, “50-keV electron multibeam mask writer for the 11-nm hp node: first results of the proof-of-concept electron multibeam mask exposure tool,” *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 11, no. 3, p. 031402, 2012.
- [32] B. J. Lin, “Future of multiple-e-beam direct-write systems,” *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 11, no. 3, p. 033011, 2012.
- [33] E. Platzgummer, C. Klein, and H. Loeschner, “Electron multibeam technology for mask and wafer writing at 0.1 nm address grid,” *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 12, no. 3, p. 031108, 2013.
- [34] Q. Dai, R. Guo, S.-Y. Lee, J. Choi, S.-H. Lee, I.-K. Shin, C.-U. Jeon, B.-G. Kim, and H.-K. Cho, “A fast path-based method for 3-d resist development simulation,” *Microelectronic engineering*, vol. 127, pp. 86–96, 2014.

- [35] M. A. Sutton, N. Li, D. Joy, A. P. Reynolds, and X. Li, “Scanning electron microscopy for quantitative small and large deformation measurements part i: Sem imaging at magnifications from 200 to 10,000,” *Experimental mechanics*, vol. 47, no. 6, pp. 775–787, 2007.
- [36] D. Drouin, A. R. Couture, D. Joly, X. Tastet, V. Aimez, and R. Gauvin, “Casino v2. 42—a fast and easy-to-use modeling tool for scanning electron microscopy and microanalysis users,” *Scanning: The Journal of Scanning Microscopies*, vol. 29, no. 3, pp. 92–101, 2007.
- [37] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, 2010.
- [38] T. P. Vogl, J. Mangis, A. Rigler, W. Zink, and D. Alkon, “Accelerating the convergence of the back-propagation method,” *Biological cybernetics*, vol. 59, no. 4-5, pp. 257–263, 1988.
- [39] B. M. Wilamowski and J. Korniak, “Learning architectures with enhanced capabilities and easier training,” in *2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)*, pp. 21–29, IEEE, 2015.
- [40] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.
- [41] A. Cotter, O. Shamir, N. Srebro, and K. Sridharan, “Better mini-batch algorithms via accelerated gradient methods,” in *Advances in neural information processing systems*, pp. 1647–1655, 2011.
- [42] W. Kristjanpoller and E. Hernández, “Volatility of main metals forecasted by a hybrid ann-garch model with regressors,” *Expert Systems with Applications*, vol. 84, pp. 290–300, 2017.