Copyright © 2002, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

CHARACTERIZATION OF SYSTEMATIC SPATIAL VARIATION IN PHOTOLITHOGRAPHY

by

Jason P. Cain

Memorandum No. UCB/ERL M02/43

19 December 2002

cory

CHARACTERIZATION OF SYSTEMATIC SPATIAL VARIATION IN PHOTOLITHOGRAPHY

by

Jason P. Cain

Memorandum No. UCB/ERL M02/43

19 December 2002

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Characterization of Systematic Spatial Variation in Photolithography

M.S. Thesis

Jason P. Cain
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

Abstract

As the dimensions of minimum-sized features in ULSI processing continue to decrease, control of these features becomes increasingly important. In this report, sources of variability in the photolithography pattern transfer process are discussed, followed by the presentation of a framework for characterization of spatial variation by dense linewidth measurements.
The variability is found to be highly systematic in nature, and methods are developed to extract the parameters of this systematic variation. Knowledge of the structure of the systematic variance allows the selection of an optimum sampling plan to best capture the variance in future measurements of wafers from the same process. This allows the systematic variation in the process to be monitored over time at a much lower cost than the initial characterization.

Contents

1 Introduction 1

1.1 Motivation . . . 1

1.2 Thesis Organization . . . 2

2 Overview of the Lithographic Pattern Transfer Process 4

2.1 Photoresist . . . 4

2.1.1 Adhesive Materials . . . 6

2.1.2 Anti-Reflective Coatings (ARCs) . . . 7

2.2 Exposure . . . 8

2.3 Post-Exposure Bake and Develop . . . 12

2.4 Etch . . . 12

2.5 State-of-the-Art in Optical Lithography . . . 13

3 Sources of Critical Dimension Variation 16

3.1 Spin Coat Effects . . . 16

3.2 Exposure Effects . . . 17

3.2.1 Mask Errors . . . 18

3.2.2 Lens Aberrations . . . 20

3.2.3 Flare . . . 27

3.2.4 Lens Heating . . . 28

3.2.5 Exposure System Vibration . . . 29

3.2.6 Focus Errors . . . 29

3.2.7 Dose Errors . . . 32

3.2.8 Illumination Variations . . . 32

3.3 Post Exposure Bake Effects . . . 35

3.4 Develop Effects . . . 35

3.5 Etch Effects . . . 36

4 **Methods for Critical Dimension Metrology** **38**
4.1 Scanning Electron Microscopy 38
4.2 Atomic Force Microscopy 40
4.3 Scatterometry 42
4.4 Electrical Linewidth Metrology 43
4.4.1 Measuring Electrical CD 44
4.4.2 Measuring Sheet Resistance 46
4.4.3 Bias Between Physical and Electrical CD 47
4.4.4 Reproducibility of ELM Measurements 48
4.4.5 Comparison ELM, CD-SEM, and Scatterometry 49

5 **Systematic Variation in the Lithographic Patterning Process** **52**
5.1 Nested Variance Model of CD Variation 52
5.2 Experimental Design 54
5.2.1 Design of ELM2 Mask 55
5.2.2 Fabrication Process Flow 56
5.3 Extraction of Systematic Variation 57

6 **Selection of Optimum Sampling Plan** **60**
6.1 Motivation for Optimum Sampling 60
6.2 Selection of Optimum Sampling Plan 61
6.2.1 Optimization Framework 62

7 **Experimental Results** **65**
7.1 CD-SEM Experiment 65
7.1.1 Exhaustive Sampling Plan 65
7.1.2 Extraction of Across-Field Systematic Variation Parameters 67
7.1.3 Optimum Sampling Plan Selection 81
7.2 ELM Experiment 87
7.2.1 Experiment Setup 87
7.2.2 Extraction of Systematic Variation Using ELM 88
7.2.3 Exhaustive Sampling Plan 89
7.2.4 Baseline Etch Lot Results 89
7.2.5 Feed-Forward Control Results 104
7.2.6 Optimum Sampling Plan Selection 117
7.2.7 Effectiveness of Optimum Sampling Plan 118
7.2.8 Tracking of Systematic Variation Model Parameters . . 121

8 Conclusions **127**

9 Acknowledgements **128**

List of Figures

2.1 Illustration of ASML ProCell™photoresist processing track . . 7
2.2 Projection system for ASML TWINSCAN™AT:1100 9
2.3 Schematic of step-and-repeat exposure system. 10
2.4 Schematic of step-and-scan exposure system. 11
2.5 Illustration of ASML TWINSCAN™AT:1100 15

3.1 Effect of decreasing linewidth on MEF 20
3.2 Primary defocus aberration 23
3.3 Primary spherical aberration 23
3.4 Primary coma aberration 24
3.5 Primary astigmatism aberration 24
3.6 Diagram of imaging system. 26
3.7 Examples of aberrated point spread functions 27
3.8 Effect of defocus on image intensity 31
3.9 Effect of dose errors on linewidth 33
3.10 Definition of partial coherence 34

4.1 Sample top-down CD-SEM image. 39
4.2 Profile dependence of top-down CD-SEM measurements 40
4.3 Sample cross-sectional SEM image 41
4.4 Schematic diagram of spectroscopic ellipsometry 43
4.5 Kelvin structure for electrical CD 45
4.6 Van der Pauw test structure 46
4.7 ELM bias 47
4.8 ELM reproducibility 49
4.9 Correlation between CD-SEM, SCD, and ELM 50

5.1 ELM2 module layout 56
5.2 ELM2 film stack before patterning 57

5.3 ELM2 film stack after patterning 57

7.1 CD-SEM Across-Field Sampling Plan 66
7.2 CD-SEM Across-Wafer Sampling Plan 66
7.3 Average field for no-OPC vertical isolated lines 67
7.4 ELM2b mask errors for no-OPC vertical isolated lines 68
7.5 Average field with mask errors removed for no-OPC vertical isolated lines 69
7.6 Average field with mask errors and polynomial model removed for no-OPC vertical isolated lines 71
7.7 Average field with mask errors and polynomial and bivariate Gaussian models removed for no-OPC vertical isolated lines . 73
7.8 Components of across-field systematic variation 74
7.9 Average wafer for no-OPC vertical isolated lines 75
7.10 Average wafer with mask errors and across-field systematic variation removed for no-OPC vertical isolated lines 76
7.11 Average wafer with mask errors, across-field systematic variation, and across-wafer bivariate Gaussian models removed for no-OPC vertical isolated lines 78
7.12 Average wafer with mask errors, across-field systematic variation, and across-wafer bivariate Gaussian and polynomial models removed for no-OPC vertical isolated lines 79
7.13 Components of across-wafer systematic variation 80
7.14 Distribution of CD measurements for no-OPC vertical isolated lines 82
7.15 CD-SEM sampling plan 1 83
7.16 CD-SEM sampling plan 2 83
7.17 CD-SEM Sampling Plan 3 83
7.18 CD-SEM Sampling Plan 4 84
7.19 CD-SEM Sampling Plan 5 84
7.20 Comparison of widths of 95% confidence intervals for across-field systematic variation model parameters 85
7.21 Comparison of widths of 95% confidence intervals for across-wafer systematic variation model parameters 86
7.22 ELM exhaustive sampling plan 90
7.23 Baseline trim etch wafer 1 90
7.24 Baseline trim etch wafer 2 91
7.25 Baseline trim etch wafer 3 91

7.26 Baseline trim etch wafer 4 92
7.27 Baseline trim etch wafer 5 92
7.28 Average wafer from baseline trim etch lot 93
7.29 Average field from baseline trim etch lot 94
7.30 Mask error scaling factor for baseline lot 95
7.31 Scaled DUT3 mask errors for baseline lot 96
7.32 Average field with scaled mask errors removed 96
7.33 Across-field systematic variation model for baseline lot 98
7.34 Residuals after removal of across-field polynomial model . . . 98
7.35 Average wafer after removal of across-field systematic variation 99
7.36 Gaussian model for wafer spot effect in baseline lot 100
7.37 Gaussian model for wafer spot effect in baseline lot 101
7.38 Polynomial model for across-wafer variation in baseline lot . . 102
7.39 Residuals after removal of all systematic variation models . . . 103
7.40 Feed-forward control wafer 1 104
7.41 Feed-forward control wafer 2 105
7.42 Feed-forward control wafer 3 105
7.43 Feed-forward control wafer 4 106
7.44 Feed-forward control wafer 5 106
7.45 Average wafer from feed-forward control lot 107
7.46 Average field from feed-forward control lot 108
7.47 Mask error scaling factor for feed-forward control lot 109
7.48 Scaled DUT3 mask errors for feed-forward control lot 109
7.49 Average field with scaled mask errors removed 110
7.50 Across-field systematic variation model for FF control lot . . . 112
7.51 Residuals after removal of across-field polynomial model . . . 112
7.52 Average wafer after removal of across-field systematic variation 113
7.53 Gaussian model for wafer spot effect in feed-forward control lot 114
7.54 Gaussian model for wafer spot effect in baseline lot 115
7.55 Polynomial model for across-wafer variation in FF control lot . 116
7.56 Residuals after removal of all systematic variation models . . . 117
7.57 ELM candidate sampling plan 1 119
7.58 ELM candidate sampling plan 2 119
7.59 ELM candidate sampling plan 3 119
7.60 ELM candidate sampling plan 4 120
7.61 Sampling plan comparison 120
7.62 ELM "intermediate" sampling plan 121
7.63 Variation of Across-Field X^2 Term 122

7.64 Variation of Across-Field Y^2 Term 123
7.65 Variation of Across-Field X Term 123
7.66 Variation of Across-Field Y Term 124
7.67 Variation of Across-Wafer X^2 Term 124
7.68 Variation of Across-Wafer Y^2 Term 125
7.69 Variation of Across-Wafer X Term 125
7.70 Variation of Across-Wafer Y Term 126

List of Tables

3.1 The first eleven Zernike polynomials 22

4.1 Comparison of metrology methods 51

7.1 Summary of across-field systematic variation model fit for polynomial model shown in Equation 7.2. 70
7.2 ANOVA table for CD-SEM across-field polynomial model. . . 70
7.3 Summary of across-field systematic variation model fit for bivariate Gaussian model shown in Equations 7.3–7.5. 72
7.4 ANOVA table for CD-SEM field spot-effect Gaussian model. . 72
7.5 Comparison of estimated variance for each successive step in the analysis of across-field systematic variation. 73
7.6 Summary of across-wafer systematic variation model fit for bivariate Gaussian model. 77
7.7 ANOVA table for CD-SEM field spot-effect Gaussian model. . 77
7.8 Summary of across-wafer systematic variation model fit for polynomial model shown in Equation 7.8. 78
7.9 ANOVA table for CD-SEM across-wafer polynomial model. . . 79
7.10 Comparison of variance for each successive step in the analysis of across-wafer systematic variation. 80
7.11 Candiate sampling plans for optimum sampling example . . . 82
7.12 Summary of across-field systematic variation model fit for polynomial model shown in Equation 7.10. 97
7.13 ANOVA table for XWFR_B across-field polynomial model. . . 97
7.14 Summary of across-wafer systematic variation model fit for bivariate Gaussian model shown in Equations 7.3–7.5. 100
7.15 ANOVA table for XWFR_B center spot Gaussian model. . . . 100
7.16 Summary of across-wafer systematic variation model fit for polynomial model shown in Equation 7.11. 102

7.17 ANOVA table for XWFR_B across-wafer polynomial model. . 102
7.18 Summary ANOVA table for XWFR_B systematic variation. . 103
7.19 Summary of across-field systematic variation model fit for polynomial model shown in Equation 7.12. 111
7.20 ANOVA table for MIX1 across-field polynomial model. 111
7.21 Summary of across-wafer systematic variation model fit for bivariate Gaussian model shown in Equations 7.3–7.5. 113
7.22 ANOVA table for MIX1 center spot Gaussian model. 114
7.23 Summary of across-wafer systematic variation model fit for polynomial model shown in Equation 7.11. 116
7.24 ANOVA table for MIX1 across-wafer polynomial model. 116
7.25 Summary ANOVA table for MIX1 systematic variation. 117
7.26 Summary of reduced sampling plan candidates. 118

List of Symbols

(ρ, ϕ) Polar coordinates within the pupil of the optical system, see equation (3.5)

(x, y) Rectangular coordinates within the pupil of the optical system, see equation (3.7) [mm]

(x_c, y_c) Center of bivariate Gaussian function, see equation (7.3)

(x_f, y_f) Coordinates within the lithographic field, see equation (7.2)

(x_w, y_w) Coordinates on the wafer, see equation (7.8)

ΔCD_{mask} Deviation from nominal critical dimension on the mask, see equation (3.2) [nm]

ΔCD_{wafer} Deviation from nominal critical dimension on the wafer, see equation (3.2) [nm]

$\Delta E(x)$ Exposure dose error, see equation (3.13) [μJ/cm^2]

ΔL Change in linewidth, see equation (3.1) [nm]

$\Delta \zeta$ Change in amount of defocus, see equation (3.12) [μm]

γ Photoresist contrast, see equation (3.1)

λ Wavelength of light source, see equation (2.1) [nm]

$\mathcal{P}(x, y)$ Generalized pupil function, see equation (3.7)

ρ	Resistivity, see equation (4.1)	[$\Omega\cdot$ cm]
σ	Partial coherence factor, see equation (3.14)	
σ_ε^2	Variance of random component of systematic variation, see equation (5.3)	[nm^2]
σ_x	"Spread" of bivariate Gaussian in the x direction, see equation (7.5)	
σ_y	"Spread" of bivariate Gaussian in the y direction, see equation (7.5)	
σ_{xy}	Interaction term between x and y "spread" in bivariate Gaussian, see equation (7.4)	
ε	Random component of critical dimension variation , see equation (5.1)	[nm]
ζ	Amount of defocus, see equation (3.11)	[μm]
A	Point spread function amplitude, see equation (3.9)	[μJ/s]
a_j	Coefficient of the jth Zernike term in waves, see equation (3.5)	[λ]
CD_{ijkl}	Critical dimension at index $ijkl$ on wafer , see equation (5.1)	[nm]
CD_{mask}	Critical dimension on the mask, see equation (3.4)	[nm]
CD_{nom}	Nominal critical dimension, see equation (5.1)	[nm]
d_o	Objective aperture diameter, see equation (3.14)	[mm]
d_s	Source diameter, see equation (3.14)	[mm]
d_s'	Effective source diameter, see equation (3.14)	[mm]
DOF	Depth of focus, see equation (3.10)	[μm]
$E(x)$	Exposure energy dose, see equation (3.1)	[μJ/cm^2]
E_0	Dose-to-clear in an open field of photoresist, see equation (3.1)	[μJ/cm^2]
$f(p)$	Optimization objective function, see equation (6.1)	
f_i	Across-field systematic variation at point i , see equation (5.1)	[nm]
h	Height of Kelvin test structure, see equation (4.1)	[nm]

$h(u,v)$	Point spread function, see equation (3.9)	[μJ/(cm^2·s)]
I	Current applied through electrical test structure, see equation (4.1)	[A]
$I(x)$	Image intensity, see equation (3.1)	[W/cm^2]
k	Scalar wavenumber, $k = 2\pi/\lambda$, see equation (3.7)	[nm^{-1}]
k_1	Lithography "difficulty" parameter, see equation (2.2)	
k_2	Depth of focus constant, see equation (3.10)	
L	Length of Kelvin test structure, see equation (4.1)	[μm]
l_k	Wafer-to-wafer systematic variation at wafer k , see equation (5.1)	[nm]
L_l	Lot-to-lot systematic variation at lot l , see equation (5.1)	[nm]
M	Magnification factor of projection system, see equation (3.2)	
MEF	Mask error factor, see equation (3.2)	
N_f	Number of points in field, see equation (5.2)	
N_L	Number of lots in the manufacturing run, see equation (5.2)	
N_l	Number of wafers in the lot, see equation (5.2)	
N_w	Number of fields on the wafer, see equation (5.2)	
NA	Numerical aperture of projection system, see equation (2.1)	
OPD	Optical path difference, see equation (3.5)	[λ]
$P(x,y)$	Real-valued pupil function, see equation (3.7)	
PSF	Point spread function, see equation (3.6)	
R_s	Sheet resistance, see equation (4.2)	[$\Omega/\Box$]
T	Nominal photoresist thickness, see equation (3.1)	[Å]
t	Time, see equation (3.1)	[s]
T_2	Equal to NA/λ for projection system, see equation (3.4)	[nm^{-1}]

t_r Resist threshold, see equation (3.4)

V Voltage measured across electrical test structure, see equation (4.1)[V]

w Electrically measured linewidth, see equation (4.1) [nm]

$W(x, y)$ Effective path length error or optical path difference, see equation (3.7) [nm]

w_i Weight for ith model parameter, see equation (6.1)

w_j Across-wafer systematic variation at point j , see equation (5.1) [nm]

w_{min} Minimum linewidth, see equation (2.1) [nm]

z_i Distance from exit pupil to image plane, see equation (3.9) [cm]

Z_j jth Zernike term, see equation (3.5)

NA_c Condenser lens numerical aperture, see equation (3.14)

NA_o Objective lens numerical aperture, see equation (3.14)

Chapter 1

Introduction

1.1 Motivation

In recent years, as feature sizes in integrated circuits have continued to shrink according to Moore's Law [1], increasing importance has been given not just to making the devices smaller, but in controlling the variation of minimum-sized features.

One of the main areas of concern in modern semiconductor manufacturing is the control of transistor gate length (also known as gate critical dimension, or simply CD, and typically expressed in either μm or nm) in MOS integrated circuits. The performance of an integrated circuit is limited by the speed of the slowest device, since the synchronous design of most integrated circuits requires that all logic paths operate at the speed of the slowest path to ensure consistent operation. Device speed is inversely proportional to

gate length, and therefore the CD directly influences the performance of the circuit. Clearly, CD variation must be tightly controlled in order to ensure maximum circuit performance.

Much of the variation observed in the gate CD in manufacturing processes is systematic in nature. In many cases, the systematic components of overall CD variation can be attributed to a specific part of the manufacturing process (for example, the lithographic exposure or post-exposure bake). Obviously, it is highly desireable to find the sources of systematic CD variation so that these sources may be reduced or compensated for.

In order to characterize the variation in the lithography pattern transfer process, it is necessary to have a reliable and accurate metrology system for measuring critical dimension. Also, it is important to make intelligent decisions about how many measurements to make and how to distribute the measurements in order to best capture the variances present.

1.2 Thesis Organization

The remainder of this thesis will be organized as follows. An overview of the lithographic pattern transfer process will be presented in Chapter 2. Sources of systematic variation will be discussed will be discussed in Chapter 3. Critical dimension metrology systems will be discussed in Chapter 4. Next, a method for the extraction of systematic variance parameters will be presented in 5. Chapter 6 will cover the selection of an optimum sampling

plan to best capture the systematic variance. Experimental results will be presented in Chapter 7. Finally, Chapter 8 will present a summary of the work and give conclusions, along with possible directions for future work.

Chapter 2

Overview of the Lithographic Pattern Transfer Process

The process most often used to transfer a circuit pattern into a thin film on a silicon wafer is photolithography. This technique is conceptually very simple, but has evolved into an extremely complex and sophisticated process as the demand for smaller and smaller circuit features has continued.

2.1 Photoresist

To begin the process, the thin film to be patterned is deposited onto the wafer in a uniform coat using one of a variety of techniques. The wafer is then coated with a photo-sensitive material called photoresist (or often simply resist).

Photoresists can be divided into two categories: positive and negative resists. Positive resists contain photoactive compounds (PACs)which consist of large molecules that separate into smaller molecules when exposed to radiation.

This makes the parts of the resist that have been exposed more soluble than those parts of the resist that have not. Therefore, a suitable "developer" chemical will wash away the exposed portions of the photoresist and leave behind the unexposed resist.

Negative resists work in the opposite manner. These resists contain a PAC that consists of small molecules which bond together when exposed to radiation, making these areas less soluble than the unexposed areas. Therefore, when a developer is used, the exposed photoresist remains on the wafer while the unexposed resist is washed away.

Traditional photoresists typically consist of three components: a matrix, a solvent, and a sensitizer or PAC [2]. The matrix component consists of a low-molecular weight material such as a novolac resin which determines the film properties of the photoresist. The sensitizers in the traditional positive resists used for patterning at mid-UV wavelengths are diazonaphthaquinones [3].

As the semiconductor industry made the transition to photolithography at deep-UV wavelengths around the 0.25 μm techology node in the mid 1990's, a new resist chemistry with very high efficiency was needed due to the limited output of available sources at a wavelength of 248 nm. To address these needs, chemically-amplified resists were developed [4, 5]. These resists are radically different from the previous DNQ/novolac-based resists. Chemically-amplified resists involve the preparation of an acid-reactive polymer (typically hydroxystyrene-based) formulated with an "onium salt" pho-

toacid generator. The critical feature is that the acid-labile group attached to the polymer reacts with the photogenerated acid in such a way that a new molecule of acid is generated, thus beginning a catalytic cycle. Therefore, a molecule of photogenerated acid might produce 500 to 1000 chemical reactions [6], thereby increasing the resist efficiency.

The most common method for coating the wafer with photoresist is to dispense a small amount of resist in liquid form onto the center of the wafer [7]. The wafer is then spun at a high rate of speed, causing the resist to spread out into a uniform layer across the wafer. After the resist coating, a "soft-bake" process is used in which the wafer is heated at a relatively low temperature for a short time. This is necessary because resist films are coated from a polymer solution, and the amount of solvent present in the photoresist must be reduced to change the resist into a more solid state [7]. An example of a modern photoresist processing track, the ASML ProCell™, is shown in Figure 2.1.

2.1.1 Adhesive Materials

In many cases the material properties of the photoresist and the underlying thin film layers prevent good adhesion of the resist. In these cases, a material such as hexamethyldisilazane (HMDS) can be used to promote adhesion between the two. The wafer is coated with the adhesive layer (typically by exposing the wafer to a vapor of the adhesion-promoting chemical [8]) before the resist is applied to the wafer.

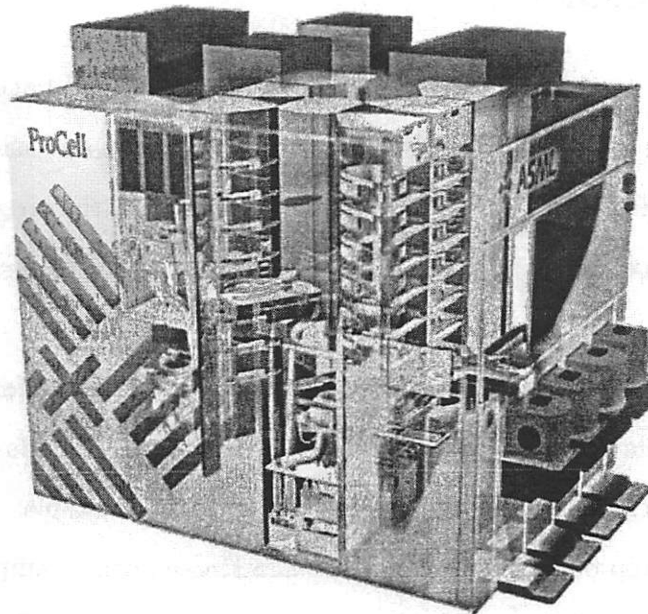

Figure 2.1: An illustration of the ASML ProCell™photoresist processing track. Source: ASML promotional material.

2.1.2 Anti-Reflective Coatings (ARCs)

Additionally, in more advanced lithography processes it is desireable to minimize reflections from the films on the wafer to prevent effects such as standing waves. Anti-reflective coatings (ARCs) may be used to reduce these effects [9]. Anti-reflective coatings may be organic or inorganic, and may be deposited before the resist (bottom anti-reflective coating, BARC) or after the resist (top anti-reflective coating, TARC). While these coatings offer the benefit of reduced reflections, they increase the complexity of the process.

2.2 Exposure

Once the wafer has been coated with photoresist the desired pattern can be transferred into the photoresist. The pattern is created on a reticle, or mask. This is typically a thin piece of high-quality quartz with the circuit pattern in chrome on one side. The mask pattern is transferred to the resist with an exposure tool.

Exposure tools have evolved considerably as the requirements for manufacturing integrated circuits became more demanding [10]. In the early days of IC manufacturing, exposure systems were relatively simple. The mask was placed on top of the wafer and then illuminated with a lamp to expose the photoresist. This process is known as contact printing. However, this leads to contamination on the mask from the resist, so the industry moved to proximity printing, in which the mask is held slightly above the wafer during the exposure. As circuit pattern features began to get smaller, fabrication of the mask features became too difficult, and projection printing came into use. This allows the mask pattern to be demagnified onto the wafer, thus reducing the requirements on the mask feature sizes. Projection printing has dominated the industry for the last few decades, and virtually all modern lithography exposure tools use this method. A diagram of the illumination and projection system for a modern projection lithography tool is shown in Figure 2.2.

Because projection systems print a demagnified image of the mask it is no

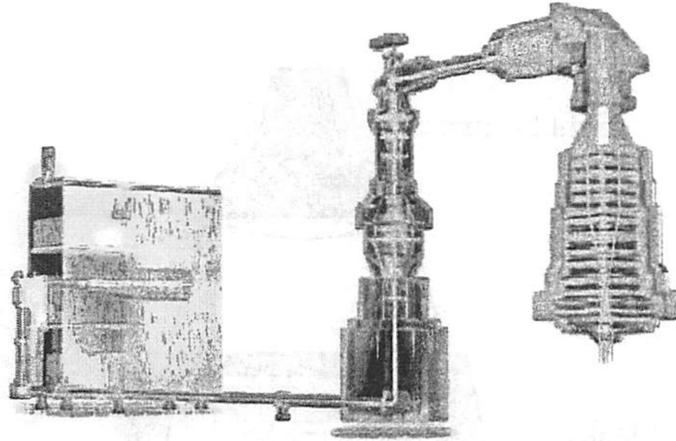

Figure 2.2: The illumination and lens system for the ASML TWINSCAN™AT:1100, a state-of-the-art projection lithography system. Source: ASML promotional material.

longer practical to print the whole wafer at once as was the case with contact and proximity lithography. Therefore, the mask pattern is printed once, covering a relatively small portion of the wafer (called a field), and then the wafer is moved so that the mask pattern may be exposed again in a different location on the wafer. These systems are called step-and-repeat systems. A schematic of this type of exposure system is shown in Figure 2.3.

Step-and-repeat exposure systems have been extremely successful, but problems begin to occur when trying to expose a large field. This is due to the fact that it becomes difficult to control lens aberrations in the projection optics across the entire field. For this reason, step-and-scan systems were developed as an improvement to step-and-repeat systems. Step-and-scan exposure systems operate in a very similar fashion to their predecessors. The difference

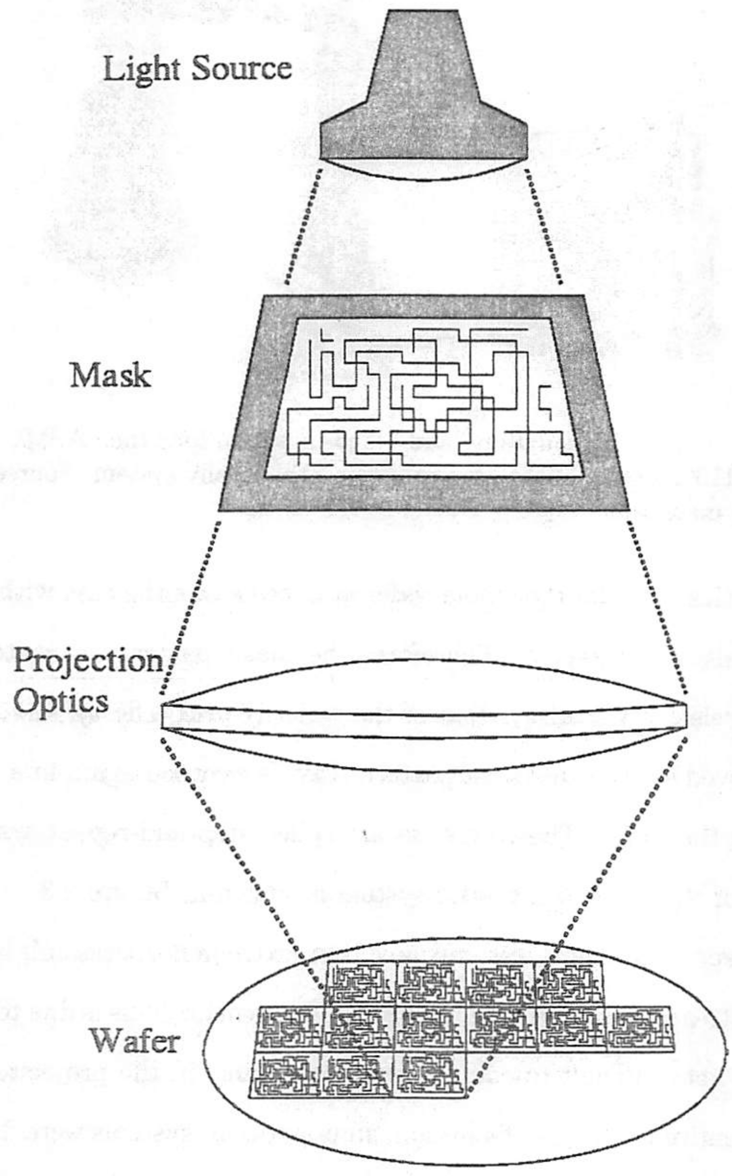

Figure 2.3: Schematic diagram of step-and-repeat exposure system.

is that instead of projecting the entire field at once, a narrow slit aperture is used to confine the image to a narrow portion in the center of the optics, where it is easier to control aberrations. The mask and the wafer are then scanned back and forth to print the image. A schematic of a step-and-scan exposure system is shown in Figure 2.4.

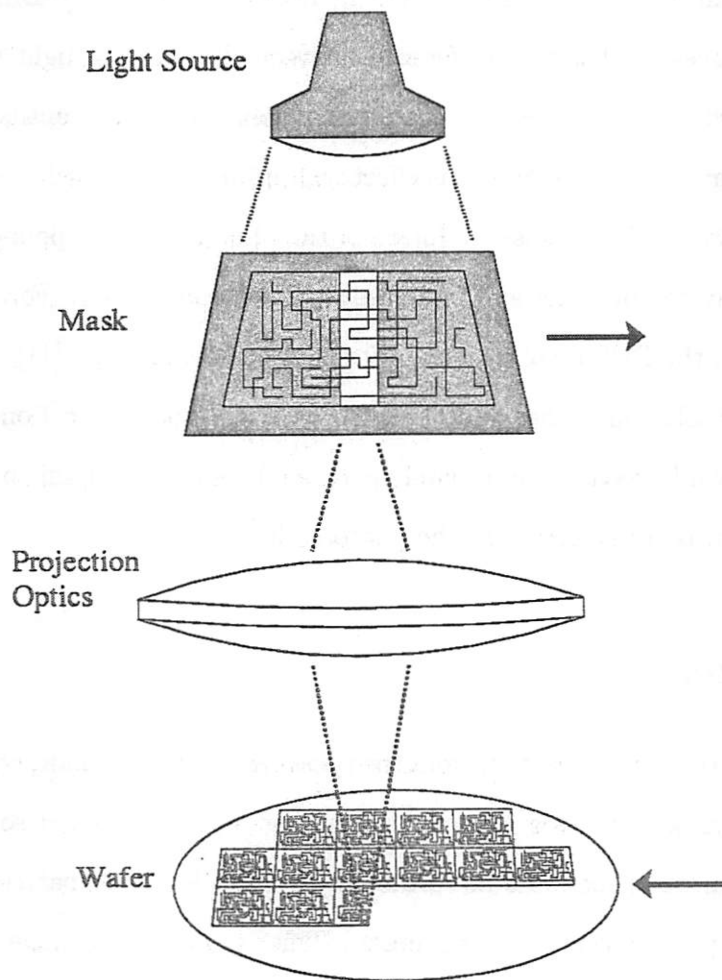

Figure 2.4: Schematic diagram of step-and-scan exposure system.

2.3 Post-Exposure Bake and Develop

After the photoresist has been exposed, a post-exposure bake (PEB) is performed at slightly higher temperatures than the soft-bake. This produces a slight reduction in solvent content, but the main advantage of PEB is to reduce the effects of standing waves on the resist line profile [7]. During the exposure process, coherent interference between the incident light and the light reflected from the substrate produces a distribution of intensity within the resist film. Left unchecked, this effect will produce "scalloped" resist line profiles. However, PEB causes diffusion of the photoactive compound which leads to an averaging effect and produces smoother line profiles. For modern DUV resists, the PEB step is highly critical in controlling CD. [11].

A developer solution is then used to wash away the soluble portions of the photoresist while leaving the insoluble portions behind [7]. At this point the mask pattern is transferred into the photoresist.

2.4 Etch

In order to transfer the pattern from the photoresist into the underlying thin film, an etching process is used. The etchant must be selected so that it attacks the areas of the wafer not protected by the photoresist pattern, while leaving the photoresist relatively intact. This is most often done using a plasma etch process, in which the wafer is exposed to a plasma containing chemical species which attack the thin film to be patterned.

Once the etching process is complete, any remaining photoresist can be removed by a variety of methods. One common technique is to "burn" off the remaining resist by using an oxygen plasma [12]. This method, called ashing, has a significant advantage in that it can often be done on the same tool as the plasma etch step.

2.5 State-of-the-Art in Optical Lithography

Modern optical lithography systems have become incredibly complicated and expensive (in excess of US$15 million for a state-of-the-art system) in the search for higher resolution, and have been extended to capabilities previously thought impossible. These systems offer extremely high resolution while still maintaining a throughput of around 100 300mm wafers per hour.

The minimum feature size resolvable by an optical system is often described by the well-known Rayleigh criterion [13],

$$w_{min} = 0.61 \frac{\lambda}{NA} \quad , \qquad (2.1)$$

where w_{min} is the width of the smallest printable isolated line in nm, λ is the wavelength of light used for the exposure in nm, and NA is the numerical aperture of the projection optics. This relation corresponds to the minimum spacing between two points which will give a 20% drop in intensity in between the two points.

In photolithography, however, the resolution can be extended beyond the

Rayleigh criterion by using a variety of resolution enhancement techniques including optical proximity corrections, off-axis illumination, and phase-shifting masks [14]. The minimum printable feature size for a lithography system can then be described by the relation [15]

$$w_{min} = k_1 \frac{\lambda}{NA} \quad , \tag{2.2}$$

where k_1 is a parameter which may be thought of as describing the "difficulty" of the lithography process. Typical values of k_1 in a state-of-the-art lithography process are 0.3-0.4.

Systems for production of state-of-the-art ICs typically use a deep ultraviolet excimer laser source [16, 17, 18] with a wavelength of 248 nm (KrF plasma source) or 193 nm (ArF plasma source). Development of systems which operate at 157 nm (F_2 plasma source) is under way. The numerical aperture of the projection optics used can be as high as 0.85, with the potential for even higher NA values in the future. The k_1 parameter accounts for factors such as advanced resists and resolution enhancement techniques (including off-axis illumination, phase-shifting masks, and optical proximity corrections of the mask pattern), and can be as low as 0.3 in modern lithography processes. These values combine to give a minimum resolveable feature size of approximately 70 nm with a 193 nm source, and if past performance is any indication, this may be pushed even lower. A diagram of a state-of-the-art step-and-scan lithography system is shown in Figure 2.5.

Figure 2.5: An illustration of the ASML TWINSCAN™AT:1100, a state-of-the-art step-and-scan lithography system. Source: ASML promotional material.

Given the complexity of the process used to print such small features, the need for careful control of each step is clear.

Chapter 3

Sources of Critical Dimension Variation

The lithographic pattern transfer process is extremely complex, consisting of many steps. Critical dimension variation may come from multiple sources at each step. In this chapter, some of the primary sources of critical dimension variation will be discussed. These sources are categorized by the step within the pattern transfer process at which each CD variation source is introduced.

3.1 Spin Coat Effects

The spin-coating process used to coat the wafer with photoresist has long been a staple of the semiconductor manufacturing process. Unfortunately, the resist may not always be as uniform as desired, which may lead to undesired CD variations across the wafer [19, 7]. A deviation from nominal resist thickness may result in a change in resist linewidth. This may be quantified using a thin-resist model [8], simplified to one dimension for clarity. This model assumes that the thickness $T(x)$ of photoresist remaining after devel-

opment is determined by the exposure energy dose $E(x) = I(x)t$, where $I(x)$ is image intensity and t is exposure time.

Under this model, the change in the photoresist linewidth ΔL that occurs as a result of a change in photoresist thickness ΔT is given by [8]

$$\Delta L = 2\left(\frac{\Delta T}{\gamma T}\right)\left(\frac{1}{E_0}\frac{dE}{dx}\right)^{-1} \quad , \tag{3.1}$$

where γ is the resist contrast, T is the nominal resist thickness, E_0 is the dose-to-clear, and E is the exposure dose.

3.2 Exposure Effects

One of the primary sources of critical dimension variation is the exposure step in the lithographic patterning process. The exposure is the process by which a circuit pattern on a reticle (or mask) is projected onto a wafer coated with a photosensitive material called photoresist. Pattern fidelity between the pattern on the reticle and the pattern projected onto the wafer is essential, and many factors can lead to a reduction in image quality. These factors will be discussed in the following sections.

It is important to note that in all modern ULSI applications, the exposure system is of the step-and-repeat (Figure 2.3) or step-and-scan (Figure 2.4) type. These systems expose one field at a time, then move (or step) to the next field. Therefore, the systematic variation introduced by the exposure process occurs at the across-field level.

3.2.1 Mask Errors

One obvious reason for failing to achieve the desired pattern on the wafer is errors on the mask. Mask errors can be thought of as any deviation of the pattern on the mask from the desired pattern described by a circuit layout engineer. For the purposes of CD variation, a mask error is a linewidth that is either larger or smaller than the designed width.

Most modern masks are printed using electron-beam lithography. While this method offers extremely high resolution, it is not perfect.

The Mask Error Factor

For systems that operate near the resolution limits of optical lithography, a linewidth error on the mask will lead to an error in the printed linewidth that is larger than the mask error by a certain factor. This factor is known as the mask error factor (MEF) or, alternatively, the mask error enhancement factor (MEEF), and is defined as

$$MEF = \frac{\Delta CD_{wafer}}{\Delta CD_{mask}/M} \quad . \tag{3.2}$$

Here ΔCD_{wafer} is the deviation from the target CD on the wafer, ΔCD_{mask} is the deviation from the target CD on the mask, and M is the magnification of the optical projection system (typically either 4 or 5).

The primary cause of greater than unity MEF is degradation of aerial image (the image of the mask projected onto the photoresist) quality for k_1 values

below 0.8 [20]. The MEF can also have a strong dependence on pitch for line-space patterns [21], and can be much larger for contacts than for lines of equivalent CD [22]. In addition, the mask type (binary chrome, alternating phase-shifting mask, etc) can affect the MEF [23]. In addition, the nonlinearity in the resist response can affect MEF dramatically [24, 25].

The MEF may be controlled by numerical aperture (NA), illumination settings, process conditions, and resist type [26, 27]. However, the optimal settings for controlling MEF may be different from the optimal settings for depth of focus (DOF). Therefore, the need to minimize the MEF must be balanced against the desire for greater depth of field in order to obtain minimum CD variation.

Recently, a compact theoretical model of the MEF has been developed [28]. This analysis assumes that the illumination is coherent and is derived based on physical diffraction principles applied to isolated features. The expression for MEF is

$$MEF = \frac{1}{2}\frac{\mathrm{sinc}(u) + \mathrm{sinc}(v)}{\mathrm{sinc}(u) - \mathrm{sinc}(v)} \tag{3.3}$$

$$\text{where} \quad u = T_2\left(t_r + \frac{CD_{mask}}{2}\right) \quad \text{and} \quad v = T_2\left(t_r - \frac{CD_{mask}}{2}\right) \quad . \tag{3.4}$$

In Equations 3.3 and 3.4, T_2 is NA/λ for the lithography system, t_r is the resist threshold, and CD_{mask} is the mask CD. A plot of MEF versus CD is

shown in Figure 3.1. The MEF is calculated with the expression shown in Equation 3.3 and the condition that MEF ≥ 1, and is intended to show the trend in MEF under feature scaling. As the figure shows, the MEF becomes increasingly important as feature sizes decrease.

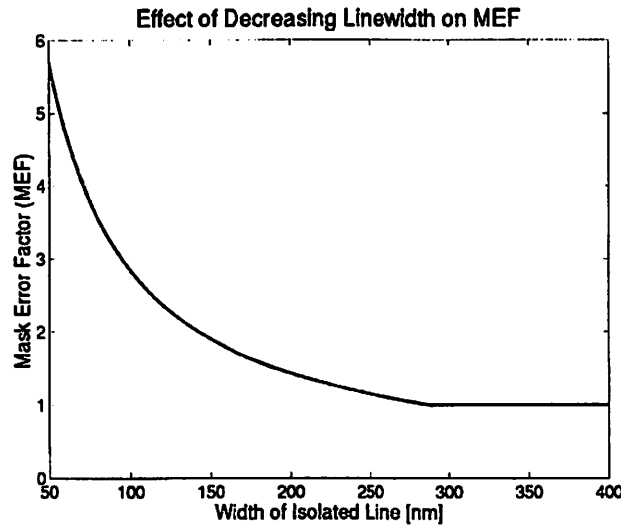

Figure 3.1: Effect of decreasing linewidth on the mask error factor (MEF) calculated with the theoretical model of Equation 3.3 and the condition that MEF ≥ 1.

3.2.2 Lens Aberrations

Despite the fact that modern optical projection lithography systems have some of the highest quality optics in the world, these optics still have some slight imperfections which affect the image quality. These effects become particularly severe when large image field size is combined with high numerical aperture [29].

Lens aberrations are the result of low spatial frequency imperfections of the lens surface, and lead to wavefront errors in the light passing through the optical system [30]. These wavefront errors can affect the resolution of the exposure system.

One way to model the aberrated wavefront is with an optical path difference (OPD) surface. The optical path difference of a particular ray is defined as the difference between the optical path length of that ray and the optical path length of a reference ray which passes through the center of the exit pupil. The location of a ray within the aperture is defined in cylindrical coordinates (ρ, ϕ), where ρ is the normalized radius within the aperture (*i.e.*, $\rho = 0$ in the center of the pupil, $\rho = 1$ at the extreme edge of the pupil). The OPD (in waves, or units of wavelength λ) can then be represented as a sum of Zernike circle polynomials [31, 32, 13],

$$OPD\left(\rho, \phi\right) = \sum_{j} a_j \lambda Z_j\left(\rho, \phi\right) \quad , \tag{3.5}$$

where a_j determines the contribution of the jth Zernike term (Z_j) measured in waves.

Zernike polynomials are convenient because they are balanced (*i.e.*, an aberration of a certain order in the power series expansion of the aberration function is mixed with aberrations of a lower order such that the variance of the net aberration is minimized), orthonormal, and form a complete set (any arbitrary two-dimensional function may be represented as a combina-

Table 3.1: The first eleven Zernike polynomials. After Levinson [8].

Z_j	Interpretation	Equation	Imaging consequence
Z_1	Piston	1	None
Z_2	X Tilt	$\rho\cos(\phi)$	Transverse pattern shift
Z_3	Y Tilt	$\rho\sin(\phi)$	Transverse pattern shift
Z_4	Defocus	$2\rho^2 - 1$	Image degradation
Z_5	3rd-order Astigmatism 45°	$\rho^2\sin(2\phi)$	Orientation-dependent focus shift
Z_6	3rd-order X Astigmatism	$\rho^2\cos(2\phi)$	Orientation-dependent focus shift
Z_7	3rd-order X Coma	$(3\rho^3 - 2\rho)\cos(\phi)$	Image asymmetry, pattern-dependent shift
Z_8	3rd-order Y Coma	$(3\rho^3 - 2\rho)\sin(\phi)$	Image asymmetry, pattern-dependent shift
Z_9	3rd-order Spherical	$6\rho^4 - 6\rho^2 + 1$	Pattern dependent focus shift
Z_{10}	X Trifoil	$\rho^3\cos(3\phi)$	Pattern shift, image asymmetry
Z_{11}	Y Trifoil	$\rho^3\sin(3\phi)$	Pattern shift, image asymmetry

tion of Zernike polynomials) [30]. Additionally, several individual balanced aberrations may be identified with specific Zernike terms, such as piston, tilt, spherical aberration, coma, and astigmatism. The first eleven Zernike polynomial terms are shown in Table 3.1.

To illustrate the Zernike polynomials, a few important terms are shown in Figures 3.2–3.5.

One way to consider the effect of aberrations on the aerial image is to consider their effect on the point spread function (PSF). The point spread function is defined as the image in the observation plane of a point in the object plane. Under the assumption that the optical system is a space-invariant (isoplanic) linear system, the aerial image can then be found by performing the convolution of the point spread function with the mask pattern:

$$image = mask \otimes PSF \quad . \tag{3.6}$$

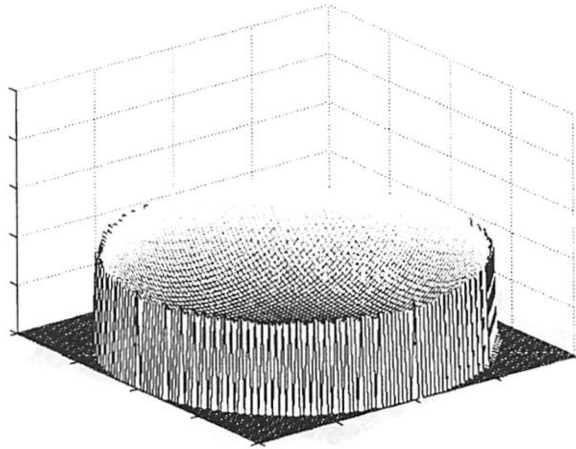

Figure 3.2: Primary defocus aberration.

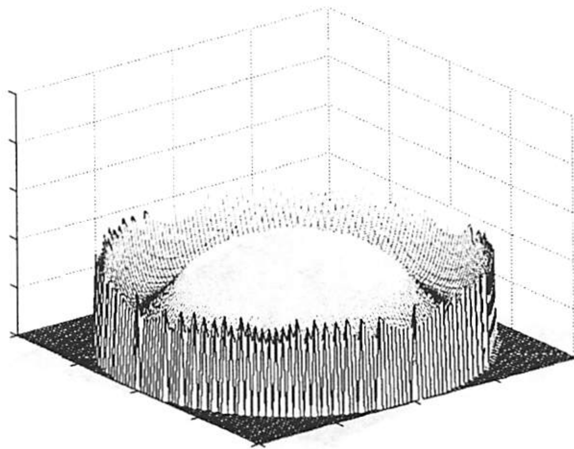

Figure 3.3: Primary spherical aberration.

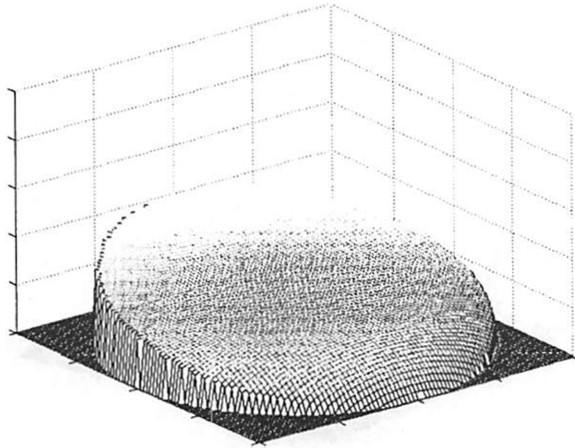

Figure 3.4: Primary coma aberration.

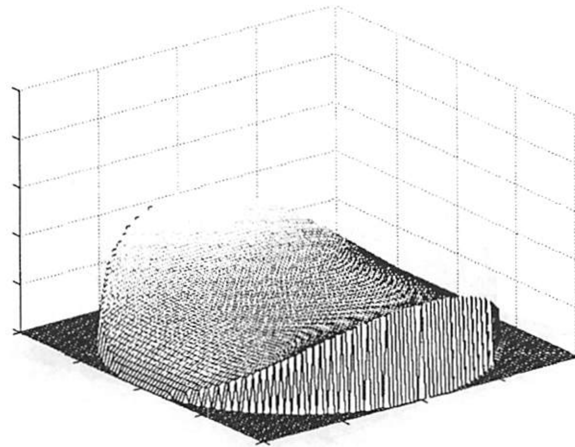

Figure 3.5: Primary astigatism aberration.

In practice, imaging systems are rarely space-invariant over the entire object field, but it is usually possible to divide the field into small regions (called isoplanic patches) within which the system is approximately space-invariant [33].

Lens aberrations are often included in a generalized pupil function $\mathcal{P}(x,y)$ so that

$$\mathcal{P}(x,y) = P(x,y)\exp[jkW(x,y)] \quad , \tag{3.7}$$

where $P(x,y)$ is the real-valued pupil function defined by

$$P(x,y) = \begin{cases} 1 & \text{inside the lens aperture} \\ 0 & \text{otherwise} \end{cases} \quad , \tag{3.8}$$

$k = 2\pi/\lambda$ is the scalar wavenumber, and $W(x,y)$ is the path length error or optical path difference (shown in cylindrical coordinates in Equation 3.5).

The point spread function $h(u,v)$ for an optical system with aberrated lenses can then be found with a scaled Fourier transform of the generalized pupil function [33]:

$$h(u,v) = \frac{A}{\lambda z_i} \iint\limits_{-\infty}^{\infty} \mathcal{P}(x,y) \exp\left\{-j\frac{2\pi}{\lambda z_i}(ux+vy)\right\} \mathrm{d}x\, \mathrm{d}y \quad , \tag{3.9}$$

where A is a constant amplitude, z_i is the distance from the exit pupil to the image plane, (x,y) are coordinates in the exit pupil, and (u,v) are coordinates in the image plane. A diagram showing the key elements of an imaging system

are shown in Figure 3.6.

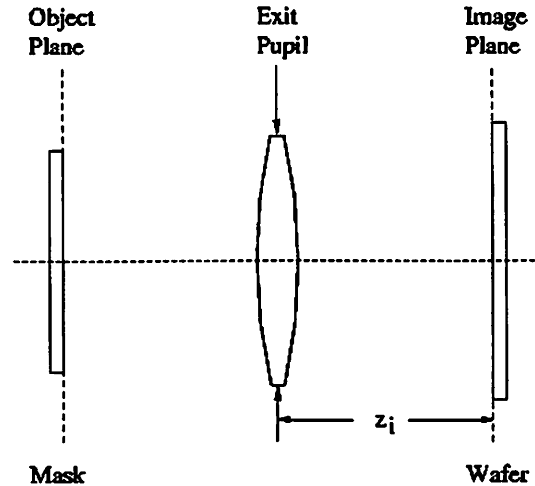

Figure 3.6: Diagram showing position of key elements of an imaging system.

Examples of some point spread functions for aberrated circular pupils are shown in Figure 3.7. These examples were calculated using a MATLAB program to solve Equation 3.9 for various levels of aberrations. Clearly, lens aberrations can lead to complicated deviations from the expected image.

Lens aberrations have very complex interactions with the geometry being imaged. For example, it has been shown that patterns composed of small dense lines are more robust to lens aberrations than patterns with small isolated lines [34]. Therefore, the particular imaging application (and therefore the dominant mask geometry) must be taken into consideration when determining the impact of lens aberrations.

One common method of measuring image quality as a result of lens aberrations is to consider the distribution of energy in an intensity point spread

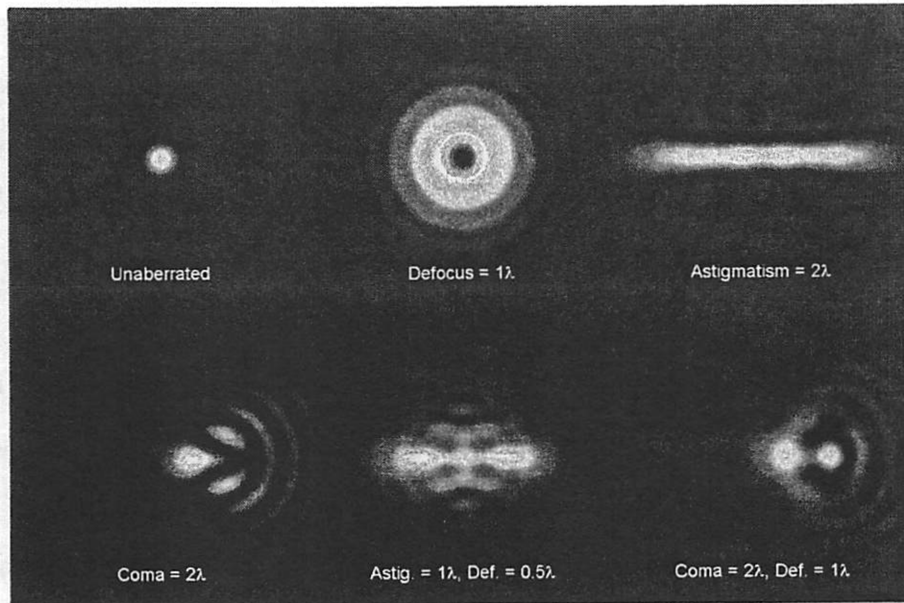

Figure 3.7: Sample point spread functions for an unaberrated circular pupil and for important aberration types. Note that the magnitudes of the aberrations shown are much larger than would be tolerable in a lithography application, and are shown for illustrative purposes only.

function. The ratio of energy at the center of an aberrated point image to the energy at the center of an unaberrated point image is known as the Strehl ratio [35]. Modern lithography applications typically require a Strehl ratio value of 0.975 or higher.

3.2.3 Flare

Undesired scattering of light in the exposure system is often referred to as flare or stray light. Many sources contribute to flare, including mid-spatial frequency imperfections of the lens surface (as opposed to low-spatial fre-

quency imperfections, which are referred to as aberrations, and high-spatial frequency imperfections, which scatter light out of the system and therefore cause a loss of optical power). Flare may also result from inefficient anti-reflective coatings on lens elements, contamination on the lens surface from dust or resist out-gassing, and reflections from internal lens hardware surfaces, wafer surface, and front lens element [36].
Although lens flare has often been treated as a simple "DC" background, it is becoming clear that flare can be a strong function of field position and can also vary significantly across the slit in a step-and-scan system [37]. In addition, since contamination is a significant source of flare [38], these effects may vary over time as contamination builds up on the lens and then is removed by periodic cleaning.

3.2.4 Lens Heating

Over the course of repeated exposures, the lens system may absorb some some of the energy and as a result the temperature of the lens may rise. This causes the lens to expand slightly, which if left uncorrected leads to focusing errors and an unwanted change in system magnification [39]. Fortunately, these effects can be modeled and controlled quite effectively [40, 41].

3.2.5 Exposure System Vibration

In modern step-and-scan systems, the wafer is moved at a high rate of speed (on the order of 0.5–1 m/s) as the image is scanned onto the wafer. However, this rapid movement can introduce vibrations into the system which can affect pattern fidelity. In addition, vibrations from external sources may be transmitted to the exposure system through the floor.

The effect of system vibrations on image quality may be modeled as the spatial convolution of an appropriate probability density function with the unperturbed aerial image [42]. The impact on critical dimension is then given by the time averaged intensity distribution. A common metric used to describe the magnitude of these vibrations is the moving standard deviation (MSD) [43].

In order to ensure optimal image quality, vibrations must be minimized. This topic has received increasing attention of late, and sophisticated new techniques such as balance masses [44] and active vibration control [45, 46] are being implemented to deal with the challenges presented by exposure system vibration.

3.2.6 Focus Errors

Another area of extreme concern in optical lithography is depth of focus. Focusing errors (defocus) occur when the actual distance between the wafer and the lens is different from the optimum distance. Increasing defocus leads

to degradation in image quality. However, there is a range of defocus over which the image is still of acceptable quality. This range is called the depth of focus (DOF). Using the criteria that the peak intensity of a point source image should not decrease by more than 20% with defocus results in the following expression for depth of focus (DOF) [13]:

$$DOF = k_2 \frac{\lambda}{(NA)^2} \quad , \tag{3.10}$$

where k_2 is a constant (of value on the order of 0.5) which varies with parameters of the particular exposure system and resist used. As wavelengths have decreased and numerical aperture values have increased in the search for higher resolution (Equation 2.2), depth of focus has decreased dramatically. The effect of defocus on can be seen in Figure 3.8. Here the image intensity pattern for a 0.25 μm space exposed at λ = 248 nm, NA = 0.5, and partial coherence $\sigma = 0.6$ at varying levels of defocus. The simulations were performed using SPLAT [47], an aerial image simulator based on physical principles of the imaging process. Simulations were performed at three different focus levels (perfect focus and 0.05 λ RMS and 0.1 λ RMS defocus) to show the effects of defocus on the image pattern. Note that there are two points at which the three curves intersect. These are known as the isofocal points. If the resist threshold is chosen so that it is near these points, there will be minimal change in linewidth through focus (although the line profile will change).

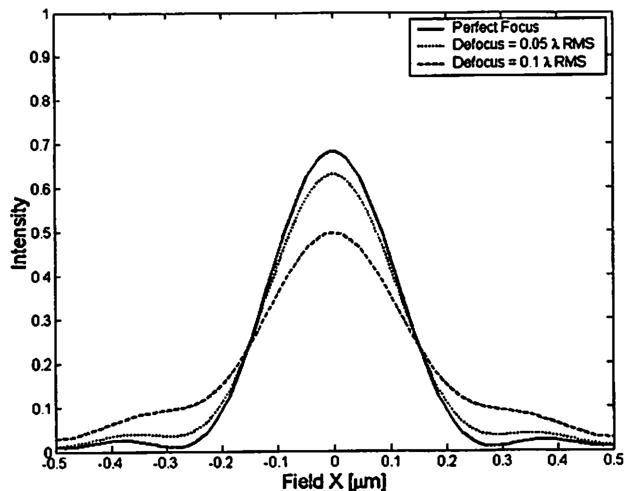

Figure 3.8: Effect of increasing defocus on image intensity for a 0.25 μm space exposed at λ = 248 nm, NA = 0.5, and σ = 0.6. Simulation performed using SPLAT [47].

The thin resist model of Section 3.1 can be expanded to include defocus effects, so that exposure dose (again simplified to one lateral dimension) now becomes a function of position in the image plane and of defocus:

$$E(x) \rightarrow E(x, \zeta) \tag{3.11}$$

where x represents position in the image plane and ζ represents the amount of defocus.

Under this model, the linewidth change ΔL due to a change in the amount of defocus $\Delta\zeta$ is given by [8]

$$\Delta L = -\frac{1}{2}(\Delta\zeta)^2 \left[\frac{1}{E_0}\frac{\partial^2 E}{\partial \zeta^2}\right] \left[\frac{1}{E_0}\frac{\partial E}{\partial x}\right]^{-1} \quad . \tag{3.12}$$

3.2.7 Dose Errors

Also of great interest in modern photolithography is the sensitivity of the linewidth to exposure dose, also known as exposure latitude. The change in linewidth ΔL resulting from a dose error $\Delta E(x)$ is given by [8]

$$\Delta L = 2\frac{\Delta E(x)}{E(x)} \left(\frac{1}{E(x)}\frac{d}{dx}E(x)\right)^{-1} \tag{3.13}$$

This effect can be visualized by considering Figure 3.9. Here the intensity profile resulting from an isolated space is shown for two different doses (the diffraction patterns were calculated using MATLAB [48]). The dashed line represents the intensity profile of a relatively low dose, while the solid line represents the intensity profile of a relatively high dose. The horizontal dotted line represents the resist threshold. The gap between the two intensity profiles at the resist threshold will be the difference in printed linewidth. As the figure shows, the higher dose leads to a wider line. Note that this effect is exaggerated in the figure for illustrative purposes.

3.2.8 Illumination Variations

The concept of spatial coherence is of extreme importance in photolithography. A source is said to be spatially coherent if the wavefronts diverging from

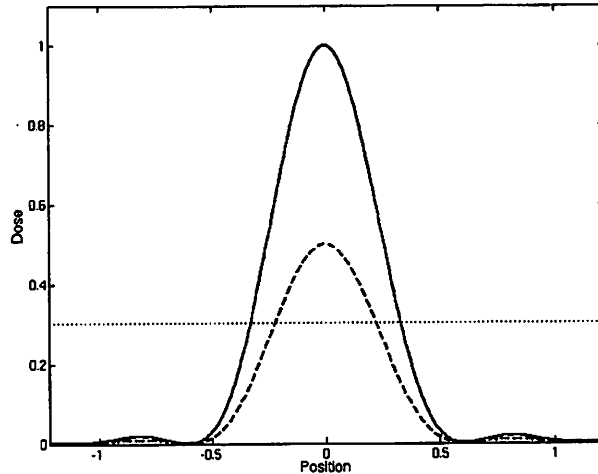

Figure 3.9: Effect of dose errors on linewidth. The dashed line represents a relatively low dose, while the solid line represents a relatively high dose. The horizontal dotted line represents the resist threshold. This effect is exaggerated here for illustrative purposes.

each point on the source are correlated, meaning that interference effects between the various wavefronts will be observed. Conversely, a source is said to be spatially incoherent if the wavefronts from each point on the source are uncorrelated, meaning no interference is possible. A source which lies in between these two extremes is said to be partially coherent [49]. Köhler illumination, in which an image of the source is formed at the entrance pupil of the objective lens, allows control of the partial coherence via a pupil at the condenser lens [15]. When the pupil is closed down, the effective source size is reduced, and the source becomes more coherent. Conversely, opening the pupil can increase the partial coherence of the source.

This concept is captured in the partial coherence factor σ, which is conventionally defined as the ratio of the effective source size to the full objective aperture size or the ratio of the condenser lens NA to the objective lens NA (Figure 3.10)[15]:

$$\sigma = \frac{d'_s}{d_o} = \frac{\text{NA}_c}{\text{NA}_o} \quad , \tag{3.14}$$

where d_s is the source diameter, d'_s is the effective source diameter, d_o is the objective aperture diameter, NA_c is the numerical aperture of the condenser optics, and NA_o is the numerical aperture of the objective optics. Note that $\sigma = 0$ describes a perfectly coherent source and $\sigma = \infty$ describes a perfectly incoherent source (although $\sigma \geq 1$ is generally considered to be incoherent).

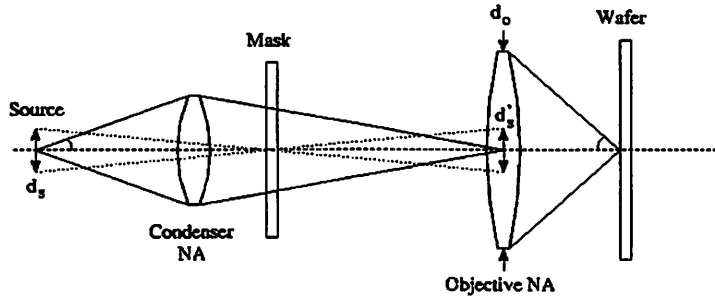

Figure 3.10: Diagram illustrating the definition of partial coherence under Köhler illumination. After [15].

Illumination sources used in photolithography are partially coherent, and in many cases are tunable over a certain range of σ. In fact, partially coherent illumination is essential in achieving small linewidths since it allows for resolution beyond the Rayleigh criterion (Equation 2.1) [33, 35]. However,

local variations in partial coherence can have an adverse effect on across-field CD variation. The primary cause of local partial coherence is aberrations in the illumination (condenser) optics [50]. Therefore, the optical quality of the illumination optics must be optimized in order to minimize these effects.

3.3 Post Exposure Bake Effects

A key step in the lithography pattern transfer process is the post exposure bake (PEB), in which thermal flow is used to reduce standing waves in the exposed resist [15]. For the chemically amplified resists used in DUV lithography, the PEB step is also essential in driving chemical reactions in the resist. Therefore, the choice of key PEB parameters such as duration and temperature can have a dramatic effect on the final resist linewidth.

It is becoming increasingly clear that complex and dynamic temperature variations may occur across the bake plate used in this process, and this may introduce substantial CD variation across the wafer [11].

3.4 Develop Effects

The development process used to dissolve the unwanted resist, leaving behind the desired pattern, can also have an effect on CD variation [15]. Perhaps the most common effect is related to the method in which the developer fluid is dispensed onto the wafer. Typically, the developer fluid is dispensed into a

"puddle" in the center of the wafer, after which the wafer is spun to smooth out the developer. However, there is usually a brief delay between the start of dispensation and the start of the spinning. During this delay, the developer fluid begins to dissolve the resist in the center of the wafer. The net effect is that a small area in the center of the wafer is developed for a longer period of time than the rest of the wafer, leading to overdevelopment (*i.e.*, narrower lines) in this center area.

3.5 Etch Effects

The plasma etching process typically used to transfer the resist pattern into the underlying thin film also introduces critical dimension variation into the pattern. The etch-related uniformity problems can be grouped into two categories: aspect ratio-dependent etching (ARDE) and pattern-dependent etching, known as microloading [51]. In both cases, the cause of the problem is ion and neutral (non-ionized species) transport.

Since plasma etching is used to etch trenches (long narrow openings) into films, the aspect ratio of the trenches desired is important. Aspect ratio is defined as the ratio of trench depth to trench width. Typically, trenches with a large aspect ratio etch more slowly than trenches with a small aspect ratio. This is often referred to as aspect ratio-dependent etching (ARDE).

Microloading causes areas with differences in local pattern density to etch at different rates. The reason is that areas with a locally higher-density

unmasked area will tend to deplete the etching species at a faster rate than areas with a locally lower-density unmasked area.

In addition, the degree of anisotropy affects the line profile and therefore the critical dimension [52]. An isotropic etching process etches at an equal rate in all directions. Conversely, an anisotropic etching process has an etch rate which is directionally dependent. In general, a high degree of anisotropy is desired, as this will provide the most faithful reproduction of the pattern in the masking film. However, in practice an etching process will never be completely anisotropic, and therefore the line sidewalls will not be perfectly vertical. This must be controlled in order to achieve good CD uniformity (although a tradeoff often exists between etch anisotropy and selectivity to other films [2]).

Chapter 4

Methods for Critical Dimension Metrology

4.1 Scanning Electron Microscopy

The primary metrology tool for measuring critical dimension has been the top-down critical dimension scanning electron microscope, or CD-SEM. These metrology tools operate by focusing a stream of electrons into a beam a few nanometers in diameter [53]. The beam is then directed onto the region of interest of the wafer, and scanned back and forth in a raster pattern. Typically, secondary electrons (electrons generated at the sample with an energy between 1 and 50 eV) are collected and the resulting signal is converted to an image. Alternatively, backscattered electrons may be used to form the image. CD-SEM tools are highly specialized to the task of measuring linewidths (including image pattern recognition software to detect line edges) and have a high degree of automation.

CD-SEM has minimal requirements for test-structure real estate, and has

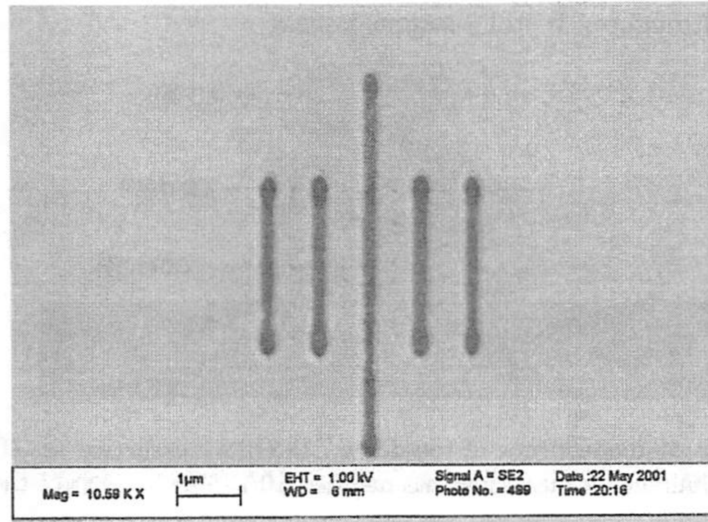

Figure 4.1: Sample top-down CD-SEM image of photoresist pattern. Courtesy of Mike Williamson, EECS Dept., UC Berkeley.

the added advantage of being able to measure nonconducting films such as photoresist. However, this tool has some limitations that become important in characterizing the across-chip linewidth variation. CD-SEM is relatively costly in terms of measurement time, which means that few points can be measured in a high-volume manufacturing environment where cost and throughput are of the utmost importance. In addition, top-down CD-SEM is profile dependent. If the profile of the feature under measurement is tapered rather than vertical, the measured CD of the feature will depend on the height along the feature at which the measurement is made (Figure 4.2). Although the recent development of CD-SEM tools with variable-angle electron columns [54] has allowed some extraction of the feature profile, the resolution of these tools is still limited by the small angle of rotation of the

source, and throughput is still a serious issue.

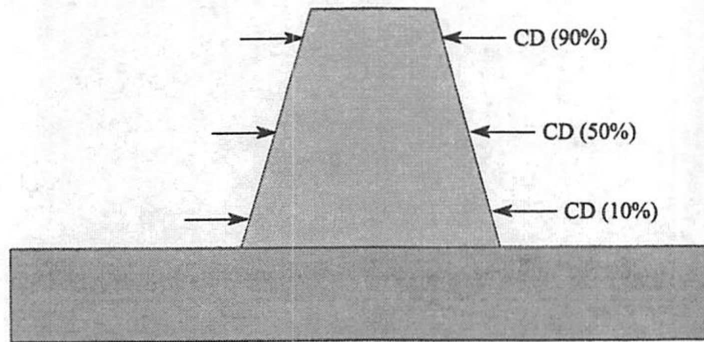

Figure 4.2: Profile dependence of top-down CD-SEM measurements. The measured CD is different for measurements made at 10%, 50%, or 90% of the feature height.

A second form of scanning electron microscope metrology is the cross-sectional SEM method. Here, the wafer is cleaved through the feature of interest, then examined with a SEM tool. This method allows the feature profile to be examined. However, this method is obviously destructive, and is also slow and expensive. Therefore, it is used only infrequently as a tool for diagnosis of serious process problems or for process development.

4.2 Atomic Force Microscopy

Atomic force microscopy (AFM) is occasionally used for CD metrology. AFM tools measure the topography of the wafer surface by bringing a sharp probe to within a few angstroms of the surface [55, 56]. At such small distances, a strong repulsive force occurs between the atoms in the probe tip and the atoms on the surface due to the Pauli exclusion principle from quantum

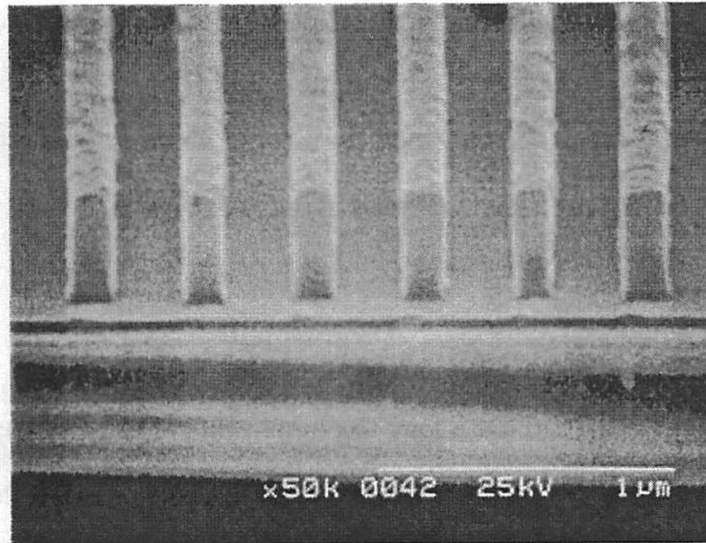

Figure 4.3: Sample cross-sectional SEM image of photoresist lines.

mechanics [57], which states that no two electrons in an atom may share the same state (*i.e.*, an identical set of quantum numbers). When two atoms are widely separated, the wave functions of their core electrons do not overlap significantly, and the electrons in the different atoms can have the same quantum numbers. However, as the atoms are brought closer together (when the probe tip is moved close to the sample surface), the overlap of the two wave functions becomes significant. Since no two electrons can share the same state, a new set of energy states is created for the composite two-nucleus system, with some electrons being pushed into higher energy states. This requires energy, which is observed as a repulsion which resists bringing the two atoms closer together.

Atomic force microscopy has excellent resolution and does not require that

the test-structure be conductive. This metrology tool is also capable of capturing the entire profile of the lines being measured. However, the probe-tip requires maintenance, and the size of the probe-tip limits the density of the pattern that can be measured. In addition, AFM is relatively slow.

4.3 Scatterometry

Scatterometry is an optical metrology technique which has been the subject of intense development over the past decade [58, 59]. The technique involves shining light of a known polarization state onto a periodic grating and collecting the reflected light. The polarization state of the reflected light can be used to infer the profile of the lines in the grating.

Scatterometry has several key advantages as a metrology tool. Perhaps the chief advantage is that scatterometry provides information about the entire profile of the structure being measured. It is also noncontact and nondestructive, and can be performed on standard ellipsometers already available in most production fabrication facilities. Scatterometry is also a relatively fast method for measurements, and is suitable for in line or *in situ* use for process control. However, scatterometry requires a test grating, which consumes precious chip real estate. In addition, the profile measured by this method can be thought of as the "average" profile of all lines in the grating, which must be considered when interpreting the results. Scatterometry has been shown through simulation to have the potential to make accurate CD

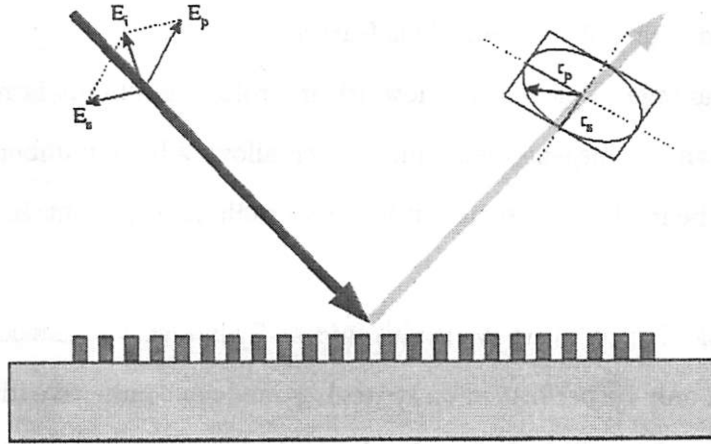

Figure 4.4: Schematic diagram of spectroscopic ellipsometry.

measurements down to at least the 50 nm technology node, and possibly further with improvements in ellipsometry equipment [60].

4.4 Electrical Linewidth Metrology

One alternative metrology tool to CD-SEM is electrical linewidth metrology (ELM). ELM relies on electrical measurements rather than optical measurements to extract the CD. Consider a conductive line pattern in a thin film. This feature may have any arbitrary profile and may include surface roughness. However, the feature may be represented electrically by a three-dimensional rectangle that has the same height and length as the actual feature, along with a width corresponding to the "average" width of the patterned feature, provided that both the physical and equivalent lines have the same electrical properties (*e.g.* resistivity profile). The electrical CD is

defined as the width of the equivalent feature.

One key advantage of electrical linewidth metrology is that it is relatively low-cost in terms of measurement time, which allows a large number of measurements to be made. In addition, it is fairly profile-independent, in contrast to CD-SEM.

However, ELM is not without disadvantages. This form of linewidth measurement can only be performed on structures made in conductive films, such as polysilicon or metal, whereas optical linewidth measurements can be made on any type of film, including insulators such as silicon nitride, silicon dioxide, and photoresist. The sensitivity of ELM is generally limited by the precision of the voltage measurements. A higher applied current will give larger voltages which are easier to measure with high precision, but higher currents will also lead to heating of the test structure which causes a nonlinear change in resistance and therefore inaccurate linewidth measurements.

4.4.1 Measuring Electrical CD

Electrical CD is often measured using a Kelvin test structure as shown in Figure 4.5 [61]. Applying a current I through a line of height h, width w, and length L results in a voltage drop V. The ratio of measured voltage to applied current (the electrical resistance of the line) is related to the dimensions of the line according to the expression

$$\frac{V}{I} = \frac{\rho L}{wh} \quad , \tag{4.1}$$

where ρ is the resistivity of the material used to fabricate the test structure. By replacing the term ρ/h with the sheet resistance R_s and rearranging terms, the following expression for electrical width w is obtained:

$$w = \frac{L}{V/I} R_s \quad . \tag{4.2}$$

Note that although R_s is nominally in units of ohms (Ω), it measures the resistance of a square (of arbitrary size) of the material being characterized. For this reason, R_s is typically expressed in units of $\Omega/\square$.

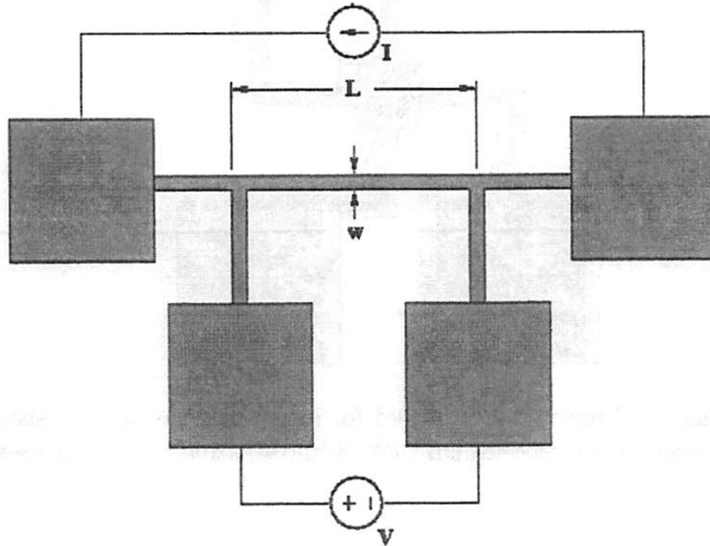

Figure 4.5: Kelvin test structure for measuring electrical CD. Current I is applied through the two outer pads while voltage V is measured between the two inner pads. The electrical CD is then calculated according to Equation 2.1.

4.4.2 Measuring Sheet Resistance

The sheet resistance R_s of the film must be measured accurately in order to obtain a good measurement of electrical CD. The most popular method for performing this measurement is with the use of some variation of the Van der Pauw structure (Figure 4.6)[62].

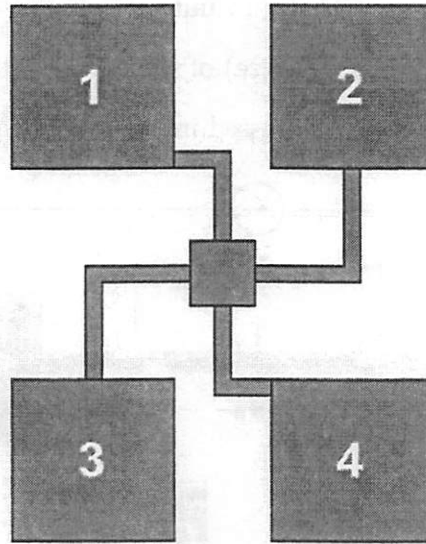

Figure 4.6: Van der Pauw structure used for measuring the sheet resistance of a thin film. Current is forced through pads 3 and 4 while voltage is measured at pads 1 and 2.

By applying a current I through pads 3 and 4 while measuring the voltage V between pads 1 and 2, the sheet resistance R_s can be calculated according to Equation 4.3 [61]:

$$R_S = \frac{\pi}{\ln 2}\left(\frac{V}{I}\right) \tag{4.3}$$

4.4.3 Bias Between Physical and Electrical CD

There may be a significant bias between the physical CD and the electrically-measured CD, especially in process flows in which polysilicon is doped after patterning. An example of this bias is shown in Figure 4.7. All statistical analysis in this chapter was performed with the JMP software package [63].

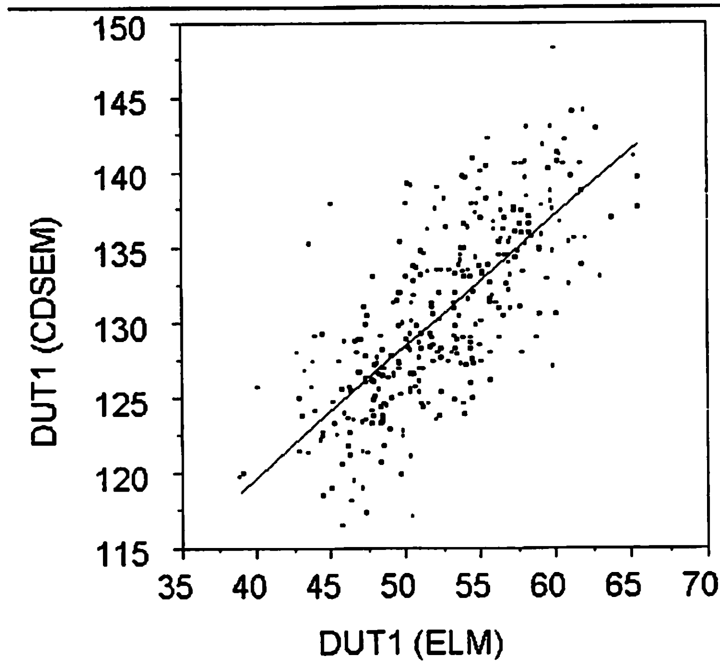

Figure 4.7: Example of bias between CD-SEM measurements (generally accepted as a reasonably good indicator of physical CD) and ELM measurements. The estimated slope of the regression line is 0.87 ± 0.09 (95% confidence interval) and the intercept (bias) is 84.97 ± 4.78 (95% confidence interval). The R^2 for this model is 0.5317.

One of the probable causes for the increased bias is dopant sidewall out-diffusion during the anneal step when polysilicon is doped after patterning

[64]. In addition, surface depletion in the conductive layer being measured may lead to non-uniformity of the electrical conductivity, which would also contribute to bias [65]. The bias between physical and electrical CD can be as much as 100 nm, which obviously is very significant as linewidths approach that same dimension. Therefore, it is extremely important to characterize this bias in order to make ELM a reliable metrology tool.

4.4.4 Reproducibility of ELM Measurements

In order to study the reproducibility of ELM measurements, an experiment was run in which a wafer with a large number of electrical linewidth test structures (see Chapters 5 and 7) was measured twice, with thre measurement runs one week apart. The results are shown in Figure 4.8. As the results show, the results are quite good. A regression line was fitted to the data, and the estimated slope was found to be 0.9965 with a symmetric 95% confidence interval of ± 0.0017. The intercept term of the model was not significant at a 95% level. The estimated mean of the difference between the two measurement sets was 0.209 nm, while the estimated standard deviation of the difference was 0.243 nm. This small difference may be explained by differences in environmental conditions (*e.g.* temperature).

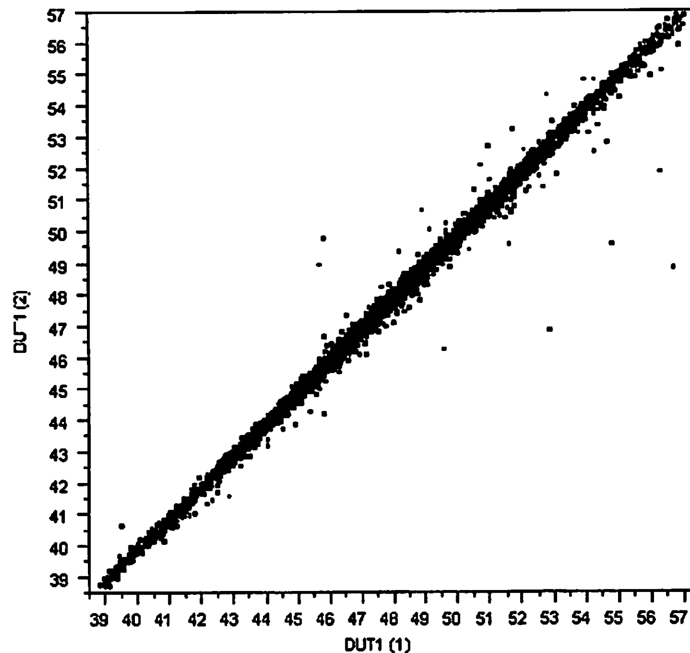

Figure 4.8: Reproducibility of electrical measurements. The wafer from slot 1 of the baseline trim etch lot was measured twice using the exhaustive sampling plan. The measurements were made one week apart. The slope of the regression line is 0.9965±0.0017 (95% confidence interval). The intercept term was not significant at a 95% level. The model R^2 statistic is 0.9949, and the number of points $N = 6403$.

4.4.5 Comparison ELM, CD-SEM, and Scatterometry

In order to compare the results of ELM, CD-SEM, and scatterometry measurements, a set of test structures was measured with each method. A matrix of scatterplots was then constructed (Figure 4.9) to study the correlations among the measurements from each metrology method. A 95% bivariate normal density ellipse is superimposed on each scatterplot.

Figure 4.9 suggests that the correlation between scatterometry and CD-SEM

is stronger than the correlation between ELM and scatterometry and between ELM and CD-SEM. This is most likely related to the nonuniform dopant distribution within the Kelvin test structure, which causes unpredictable differences between physical CD and electrically-measured CD.

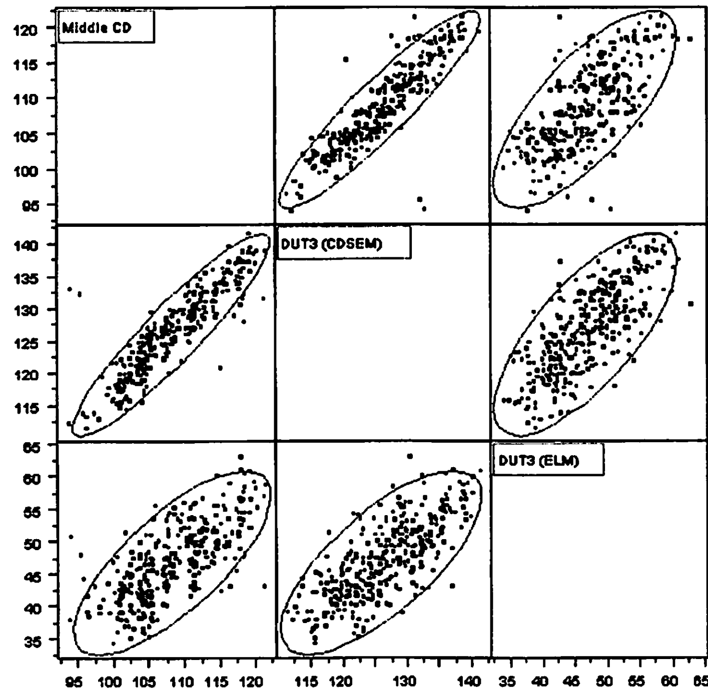

Figure 4.9: Correlation between CD-SEM, SCD, and ELM measurements. Note that "Middle CD" refers to the 50% CD as measured with scatterometry.

Table 4.1: Comparison of metrology methods

	CD-SEM	**Scatterometry**	**ELM**
Throughput	Slow	Fast	Fast
Test Structure	Line	Grating	Kelvin
Profile Measurement	No	Yes	No
Resolution Limiting Factors	Charging	Detector sensitivity	Bias

Chapter 5

Systematic Variation in the Lithographic Patterning Process

5.1 Nested Variance Model of CD Variation

A certain amount of variation impacts any manufacturing process, and the semiconductor manufacturing process is no exception. While some of the variation is assumed to be random, it has become apparent that much of the variation in the lithographic patterning process is systematic (*i.e.*, not due to chance) in nature. This gives hope that if the sources of the systematic variation can be found, methods for controlling systematic variation can be developed.

Systematic variation of critical dimension has been found to at several different scales in the lithographic patterning process [66]. One type of systematic variation that has received much attention is variation within the

lithographic field. This type of variation affects each field on the wafer in the same way, and may be related to the optics of the exposure system. In addition, there may be systematic variation across the wafer as a whole. This type of variation may be due to nonuniformities in temperature during a processing step, resist thickness nonuniformity, and other problems that affect the whole wafer. Systematic variation is also seen from wafer to wafer within a lot, possibly due to some tool problem which affects every wafer processed on that tool. Finally, systematic variation can be seen over longer periods of time from lot to lot. This may be due to long-term machine drifts, operator practices, and similar long-term effects.

In order to help understand the nested character of the CD variation, the following model is proposed, in which the variation from the nominal CD at a given point on the wafer is assumed to be the sum of the nominal CD plus each of the systematic variance components and a random component:

$$CD_{ijkl} = CD_{nom} + f_i + w_j + l_k + L_l + \varepsilon, \tag{5.1}$$

where

$$i = 1 \ldots N_f, \quad j = 1 \ldots N_w, \quad k = 1 \ldots N_l, \quad l = 1 \ldots N_L \tag{5.2}$$

and

$$\varepsilon \sim N(0, \sigma_\varepsilon^2). \tag{5.3}$$

Here CD_{ijkl} is the CD at a given point on the wafer, CD_{nom} is the true mean of all critical dimensions in the process, f_i is the across-field systematic variation at index i within the field, w_j is the across-wafer systematic variation at index k within the wafer, l_k is the systematic variation from wafer to wafer within a lot at index k within the lot, L_l is the systematic variation from lot to lot at index l within the manufacturing run, and ε is random variation (including variation in the measurement process). The systematic variation components f_i, w_j, l_k, and L_l are assumed to be fixed for a given manufacturing process. Also, N_f represents the number of points in the field, N_w represents the number of fields on the wafer, N_l represents the number of wafers in the lot, and N_L represents the number of lots in the manufacturing run. The random variation ε is assumed to be normally distributed with mean 0 and variance σ_ε^2.

5.2 Experimental Design

An large scale experiment (covering 200 wafers over a time period of several weeks) was designed to evaluate systematic variation in the lithographic pattern transfer process. A mask was designed to allow for dense CD measurements with a variety of metrology methods. This mask was then used to print a large number of wafers, which were then measured to extract systematic variation.

5.2.1 Design of ELM2 Mask

In order to characterize the systematic variation in the pattern transfer process, dense measurements are needed to obtain a detailed picture of CD variation (a "variance map") for each of the desired variance scales (across chip, across wafer, etc.). In addition, it is desirable to compare various metrology methods to evaluate their effectiveness in characterizing systematic variation. With this in mind, a test mask was designed to facilitate the investigation of systematic linewidth variation through a 2^3 factorial experiment [67] with pattern density (isolated and dense), orientation (horizontal and vertical), and the presence or absence of optical proximity corrections (OPC) as the three factors. This mask is composed of a group of test structures, or module, repeated many times. Each module consists of two similar submodules. Both submodules have eight test structures (two of each combination of pattern density and orientation) for measuring the linewidth either electrically or via CD-SEM. In addition, both submodules have a Van der Pauw structure for measuring the resistivity of the film in which the test structures are fabricated (polysilicon in the case of this experiment) to facilitate ELM. The submodules also contain a variety of periodic gratings for scatterometry measurements and several lines that run the full width of the submodule and are continuous across the mask, so that the wafer may be cross-sectioned at any point and still yield usable cross-sectional SEM features. The factorial design is completed by alternating submodules (which are otherwise identical) between the presence or absence of OPC. A block diagram of the figure

is shown in Figure 5.1.

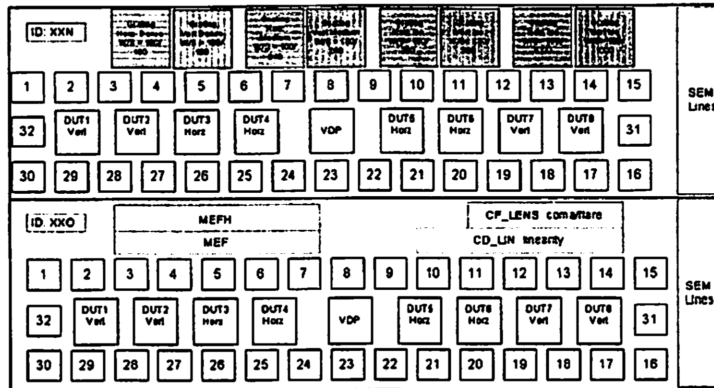

Figure 5.1: Block diagram of ELM2 module.

This module is then repeated in a 19x22 array across the mask. This array is designed to be large enough to cover the maximum field size for modern steppers and scanners. When the mask pattern is reproduced many times on the wafer using a stepper or scanner exposure system, the result is that many thousand linewidth test structures are created on the wafer. This allows a highly detailed linewidth variance map to be created.

5.2.2 Fabrication Process Flow

The fabrication process used for the wafers in the experiment is quite simple. A thin gate oxide is first deposited on eight-inch silicon wafers. A layer of polycrystalline silicon is then deposited, followed by a layer of silicon-rich nitride (SiN). The nitride serves as both an anti-reflective coating and a hard mask for etching the poly-silicon. A diagram of this thin film stack is shown

in Figure 5.2.

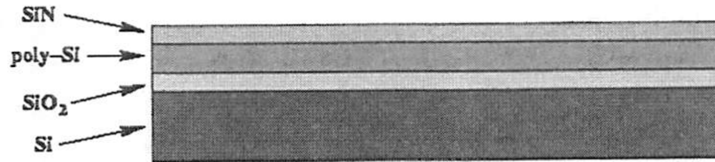

Figure 5.2: The film stack used in the fabrication of the ELM2 wafers.

This film stack is then patterned using the lithography pattern transfer process. A resist trim etch is then performed to shrink the resist lines from a nominal CD of 180 nm to a nominal CD of 130 nm. Finally, the nitride layer is removed to allow for electrical contact to the underlying polysilicon layer for electrical linewidth measurements. A diagram of the final film stack after patterning is shown in Figure 5.3.

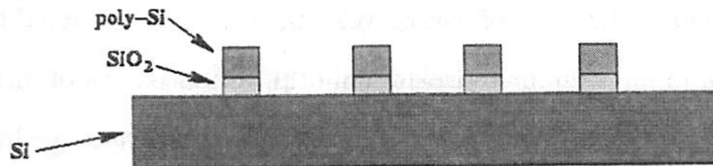

Figure 5.3: The ELM2 film stack after lithographic patterning, resist trim etch, polysilicon etch, and ARC strip.

5.3 Extraction of Systematic Variation

The process for extracting systematic variation is fairly simple. Essentially, the data is averaged at each level of the variance hierarchy discussed in

Section 5.1. For example, to calculate the "average" field, the following expression would be used:

$$CD_{avg,i} = \frac{1}{N_w N_l N_L} \sum_{j=1}^{N_w} \sum_{k=1}^{N_l} \sum_{l=1}^{N_L} CD_{ijkl}, \tag{5.4}$$

where $CD_{avg,i}$ is the CD value at index i within the field. Similarly, the "average" wafer can be calculated according to:

$$CD_{avg,ij} = \frac{1}{N_l N_L} \sum_{k=1}^{N_l} \sum_{l=1}^{N_L} CD_{ijkl}, \tag{5.5}$$

where $CD_{avg,ij}$ is the CD value at index i within the field and index j across the wafer.

It is assumed that this process "averages out" all higher levels of the variance hierarchy. Any the structure of lower levels may then be subtracted from the averaged data in an attempt to isolate only the variance level of interest.

For an example, suppose we start at the bottom of the variance hierarchy, across-field systematic variation. The average is computed at each point in the field across all fields and all wafers. This is then the "average" field. If mask measurements are available, the mask errors (magnified by a suitable scaling factor) may be subtracted from the average field. The mask errors will be constant for a given process. Statistical models (*e.g.*, a linear model [68] with coordinates within the field as independent variables) may then be fit to the data to characterize the across-field systematic variation that is not the result of mask errors.

We can then consider the next level in the variance hierarchy, across-wafer systematic variation. First the average is computed at each point on the wafer across all wafers to give the "average" wafer. The mask errors and statistical models of across-field systematic variation can then be subtracted to isolate only the across-wafer variation. New statistical models can then be fit to this data. This process can be repeated for higher and higher levels in the variance hierarchy. Examples and experimental results will be presented in Chapter 7.

Chapter 6

Selection of Optimum Sampling Plan

6.1 Motivation for Optimum Sampling

It is not generally practical to measure all of the test structures on the ELM2 test wafers. For example, consider the cost of exhaustively measuring an entire wafer patterned with the ELM2 mask. Each module contains 16 features, and there are 19x22 = 418 modules per field ($N_f = 418$ in Equation 5.2). If the each wafer has 21 complete fields ($N_w = 21$), this corresponds to 140,448 features per wafer. This would take several days to measure even with high-speed electrical measurements, and possibly an order of magnitude longer using CD-SEM. In a high-volume production environment processing tens of thousands of wafers per month, this would be out of the question. Even the "exhaustive" sampling plan used to characterize systematic variation in Chapter 5 required nearly 10,000 measurements (roughly 25 hours of measurement time). Clearly, this is not a practical solution.

Therefore, decisions must be made concerning how many structures to measure and how to distribute the measurements across the wafer. The need to accurately capture the systematic variance must be balanced against the cost of making the measurements. Knowledge about the structure of the systematic variance allows the selection of a so-called optimum sampling plan which will "best" capture the systematic variance at a fixed cost.

6.2 Selection of Optimum Sampling Plan

Once a detailed variance map has been created and the systematic variance has been extracted as described in Chapter 5, an optimum sampling plan can be selected. The first step is to set a fixed measurement cost, *i.e.* the maximum allowable measurement time for each wafer. This is important because it describes the delay in processing required to measure each wafer, and increased delay means that the wafer throughput (and, therefore, profit) is decreased. In setting the measurement cost, the need to sample each wafer many times (requiring a high measurement cost per wafer) to capture within-wafer systematic variation must be balanced against the need to sample many wafers (requiring a low measurement cost per wafer) to capture wafer-to-wafer systematic variation.

Once a fixed cost is determined, different sampling plans that have the same cost but varying distributions of the measurements can be evaluated. This is done by trying the sampling plans on the data from the exhaustively sampled

wafers and fitting the various systematic variation models to the data from the reduced sampling plans. The best plan is then defined as the plan which obtains the best fit for the different models. One measure of best fit is the length of the 95% confidence intervals for the different model parameters. Of course, different priority may be given to different model parameters, and this may be taken into account when selecting the best plan (*e.g.*, by using the weighting objectives method discussed in Section 6.2.1). Examples of this method will be presented in Chapter 7.

6.2.1 Optimization Framework

The task of measuring features on a wafer to characterize across-field and across-wafer systematic variation involves two competing objectives: the need to measure many points per field to capture across-field variation and the need to spread the points across the wafer to capture across-wafer variation. The cost (time) of a measurement plan is directly related to the number of measurements in the plan. In order to maintain a high wafer throughput, measurement cost must be minimized. Therefore, this is a multiobjective optimization problem.

Many approaches have been developed to deal with multiobjective optimization problems [69][70]. Here a relatively simple approach known as the weighting objectives method is presented as an illustrative example.

First the individual objectives must be clearly defined. If the goal is to minimize the lengths of the confidence intervals of the individual parameters, then

a set of objective functions $f_i(p)$ can be defined to represent the minimum width of the confidence interval of parameter i for sampling plan p. Note that the sampling plan contains information on the number and distribution of measurements to be made, and the set of sampling plans considered will all have the same total number of measurements (and therefore have the same measurement time cost).

The weighting objectives method involves adding the individual objective functions together using arbitrary weighting conditions for each [71]. This has the effect of changing the multiobjective optimization problem into a scalar optimization problem by creating one objective function of the form

$$f(p) = \sum_{i=1}^{k} w_i f_i(p) \quad , \tag{6.1}$$

where k is the number of model parameters and w_i is the weight for the i_{th} parameter. The weights are meant to represent the relative importance of the model parameters, and are defined such that $0 \leq w_i \leq 1$ and

$$\sum_{i=1}^{k} w_i = 1 \quad . \tag{6.2}$$

Once the objective function of Equation 6.1 has been defined, any of a variety of minimization methods (*e.g.* simplex optimization [70]) may be used to find a sampling plan that minimizes the objective function and therefore may be considered the "optimum" sampling plan given the fixed cost and model parameter weights set by the engineer. This optimum sampling plan may

then be used to track the systematic variation model parameters over time at a much lower cost than the exhaustive sampling plan.

Chapter 7

Experimental Results

7.1 CD-SEM Experiment

An experiment was performed in order to evaluate the effectiveness of the techniques discussed in Chapters 5 and 6. Wafers were patterned in photoresist at an industrial fab using the ELM2 mask, and the resist patterns were then measured using CD-SEM at another industrial partner.

7.1.1 Exhaustive Sampling Plan

Since this work focuses primarily on systematic spatial variation, the sampling plans used to collect the detailed variance maps needed for extraction of systematic variation were chosen with an emphasis on across-field and across-wafer systematic variation. Two sampling plans were used. The first plan is intended to capture across-field systematic variation. In this sampling plan, 20 modules per field are measured, with five fields per wafer for 50 wafers (Figure 7.1). The second plan is targeted toward across-wafer systematic

variation, and measured three modules per field with all 23 fields per wafer for 69 wafers (Figure 7.2). For each module measured, one vertical isolated line with no OPC, one horizontal isolated line with no OPC, one horizontal dense line with no OPC, and one horizontal line with OPC were measured.

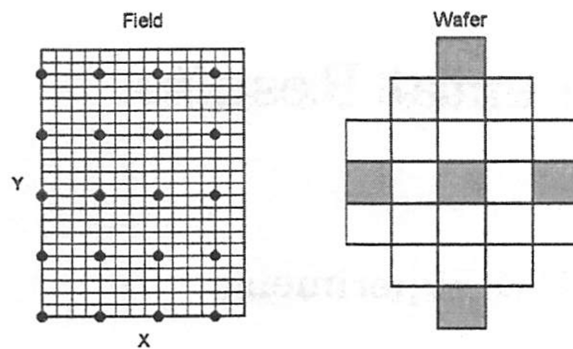

Figure 7.1: CD-SEM sampling plan intended to capture across-field systematic variation. 50 wafers were measured.

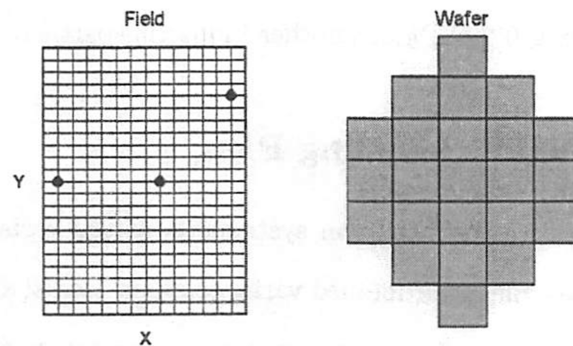

Figure 7.2: CD-SEM sampling plan intended to capture across-wafer systematic variation. 69 wafers were measured.

7.1.2 Extraction of Across-Field Systematic Variation Parameters

In order to extract the across-field systematic variation, the average field is computed by calculating the average across all fields and all wafers at each point in the field according to Equation 5.4. It is assumed that this will have the effect of averaging out all higher levels of the variance hierarchy (w_j, l_k, L_l, and ε in Equation 5.1), leaving only the across-field effects. The average field calculated for vertical isolated lines without OPC is shown in Figure 7.3.

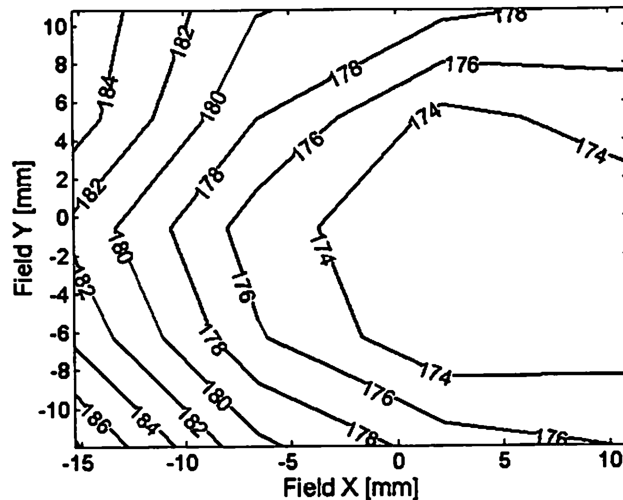

Figure 7.3: CD contour map for average field calculated for vertical isolated lines without OPC. CD values are in nm, field dimensions are in mm.

Mask CD measurements can be used in an attempt to remove the mask errors from the average field. The mask errors for the ELM2b mask, magnified by

the approximated MEF, are shown in Figure 7.4.

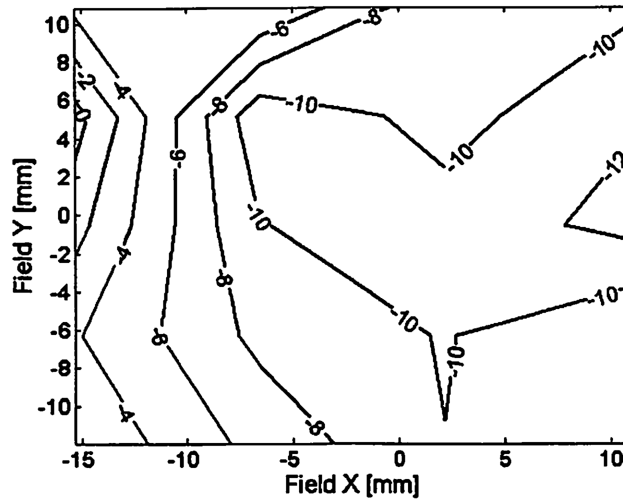

Figure 7.4: Contour map for ELM2b mask errors magnified by the approximated MEF for vertical isolated lines without OPC. Mask error values are in nm, field dimensions are in mm.

The mask errors can then be subtracted from the average field data to obtain the results shown in Figure 7.5.

An ANOVA model is then proposed to extract the across-field systematic variation. The ANOVA model is

$$CD(x_f, y_f) = \mu + f(x_f, y_f) + \varepsilon_{x_f, y_f}, \tag{7.1}$$

where $f(x_f, y_f)$ is a treatment effect describing across-field systematic variation.

For a scanner system of the type used for exposure in this experiment, it is expected that there will be a large relatively large variation in the vertical

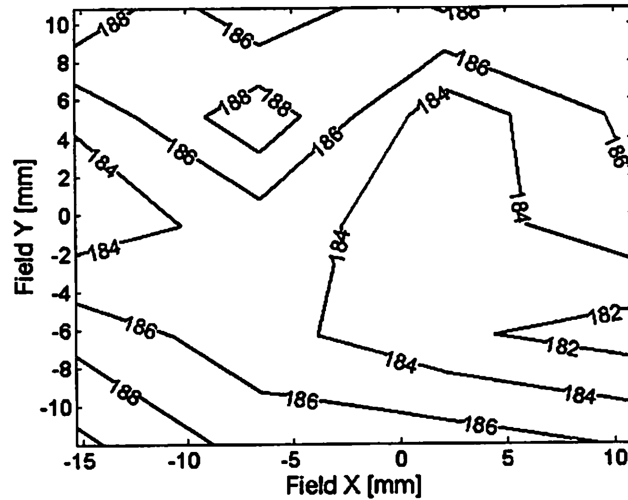

Figure 7.5: CD Contour map for average field with mask errors (magnified by the approximated MEF) removed for vertical isolated lines without OPC. CD values are in nm, field dimensions are in mm.

direction (across the slit) due to lens aberrations. Also, there may be a linear variation in the horizontal (scan) direction due to variations in the source intensity.

The treatment model chosen includes a second-order polynomial across the slit direction (due to the apparent curvature in the vertical direction in Figure 7.5) with a linear term in the scan direction and an interaction term. This model is shown in Equation 7.2.

$$f(x_f, y_f) = a \cdot y_f^2 + b \cdot x_f + c \cdot y_f + d \cdot x_f \cdot y_f + intercept + \varepsilon \tag{7.2}$$

Here x_f and y_f are the coordinates within the field, and a, b, c, d, and

Table 7.1: Summary of across-field systematic variation model fit for polynomial model shown in Equation 7.2.

Parameter	Estimate	Probability > $\|t\|$	Symmetric 95% Confidence Interval Width
Intercept	183.832 nm	0.0000	±0.179 nm
y_f^2	0.027 nm/mm^2	< 0.0001	±0.0021 nm/mm^2
x_f	-0.073 nm/mm	< 0.0001	±0.012 nm/mm
y_f	0.076 nm/mm	< 0.0001	±0.0147 nm/mm
$x_f * y_f$	0.008 nm/mm^2	< 0.0001	±0.0014 nm/mm^2

Table 7.2: ANOVA table for CD-SEM across-field polynomial model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	4	28207.99	7052.00	240.1620
Error	8676	254757.82	29.36	*p*>F
Total	8680	282965.82		< 0.0001

intercept are parameters to be estimated using least squares regression. The error in the model is represented by ε. The parameter estimates of the model fit are shown in Table 7.1, while an ANOVA table for the model is shown in 7.2. All statistical model fitting was performed using the JMP software package [63]. Contour plots were produced using MATLAB [48].

This fitted polynomial model can then be subtracted from the data as shown in Figure 7.6. Examination of the results shown in Figure 7.6 reveals that there is a highly localized "spot" effect in the upper-left portion of the field. This is most likely a mask-related issue (either a defect on the mask or an incorrect MEF used in subtracting the mask errors). A model may be used in an attempt to remove the effect from the data.

The model chosen to remove the "spot" effect is a bivariate Gaussian of the

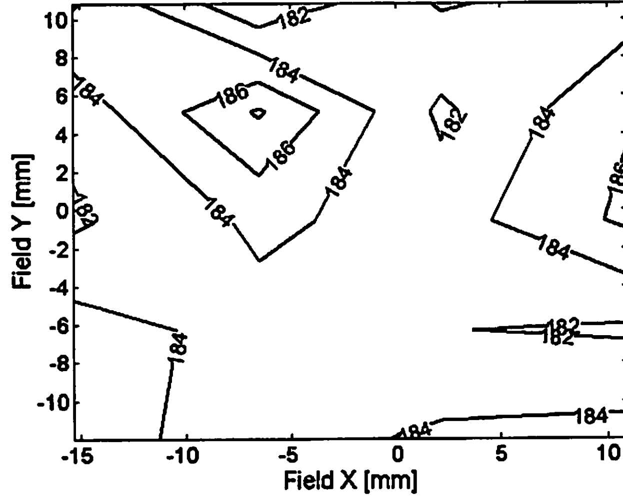

Figure 7.6: CD Contour map for average field with mask errors (magnified by the approximated MEF) and polynomial model of across-field systematic variation removed for vertical isolated lines without OPC. CD values are in nm, field dimensions are in mm.

form

$$z = \frac{(x_f - x_c)^2}{\sigma_x^2} - \frac{2\rho (x_f - x_c)(y_f - y_c)}{\sigma_x \sigma_y} + \frac{(y_f - y_c)^2}{\sigma_y^2} \quad , \tag{7.3}$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{7.4}$$

$$CD(x_f, y_f) = \frac{scale}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{z}{2(1-\rho)^2}\right] + constant \tag{7.5}$$

where *scale* is a scaling parameter, x_c and y_c are the coordinates of the center of the bivariate Gaussian function, σ_x and σ_y describe the "spread" in each

Table 7.3: Summary of across-field systematic variation model fit for bivariate Gaussian model shown in Equations 7.3–7.5.

Parameter	Estimate
scale	90.03
constant	183.59 nm
x_c	-6.48 mm
y_c	3.59 mm
σ_x	0.55 mm
σ_y	3.67 mm
σ_{xy}	1.06 mm^2

Table 7.4: ANOVA table for CD-SEM field spot-effect Gaussian model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	7	8.91×10^3	1272.86	44.88
Error	8673	2.46×10^5	28.36	p>F
Total	8680	2.55×10^5		< 0.0001

direction, σ_{xy} is an interaction term, and *constant* is used to set the value of the function far away from the center point. It is assumed that nonlinear regression may be used to find a maximum likelihood estimate for each of these parameters. The results of this model fit are shown in Table 7.3, and an ANOVA table is shown in Table 7.4.

The results of "removing" the bivariate Gaussian model from the data are shown in Figure 7.7. In order to compare the variation removed by each step in this process, Table 7.5 shows the estimated variance present in the data at each successive step. The table suggests that the main component of across-field systematic variation in this process is due to mask errors. After the last step, there is no clear geometric pattern remaining in the data. The

Table 7.5: Comparison of estimated variance for each successive step in the analysis of across-field systematic variation.

Average field	$\hat{\sigma}^2 = (4.79\text{nm})^2$
... after removal of mask errors (magnified by MEF)	$\hat{\sigma}^2 = (2.69\text{nm})^2$
... after additional removal of polynomial model of across-field systematic variation	$\hat{\sigma}^2 = (1.95\text{nm})^2$
... after additional removal of bivariate Gaussian model for field spot effect	$\hat{\sigma}^2 = (1.65\text{nm})^2$

remaining variance is assumed to be random. These components are shown visually in Figure 7.8.

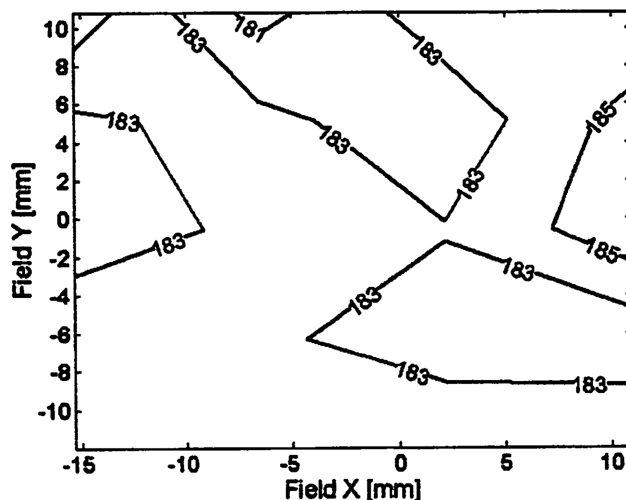

Figure 7.7: CD contour map for average field with mask errors (magnified by the approximated MEF) and polynomial and bivariate Gaussian models of across-field systematic variation removed for vertical isolated lines without OPC. CD values are in nm, field dimensions are in mm.

The next level in the systematic variance hierarchy is across-wafer systematic variation. This variance component can be extracted in much the same way as the across-field systematic variation. The average over all wafers

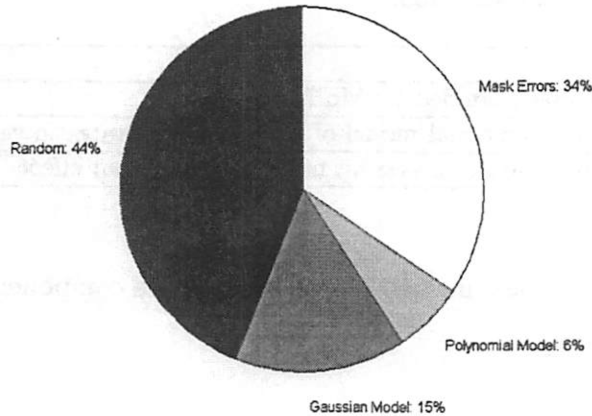

Figure 7.8: Components of across-field systematic variation. Mask errors are magnified by the mask error factor.

is computed for each point on the wafer, which results in the estimated "average" wafer, shown in Figure 7.9. Note the periodic signature of across-field systematic variation, visible as vertical bands in the contour map.

The CD variance due to across-field effects can be "removed" using the mask errors (magnified by MEF) and the systematic across-field variation models calculated previously. The results of removing these across-field systematic variation effects are shown in Figure 7.10. Note that the periodic signature seen in Figure 7.9 is now gone.

Examination of the CD contour map with across-field systematic variation removed as shown in Figure 7.10 suggests that there is a localized spot effect in the center of the wafer. This is most likely related to the way in which the

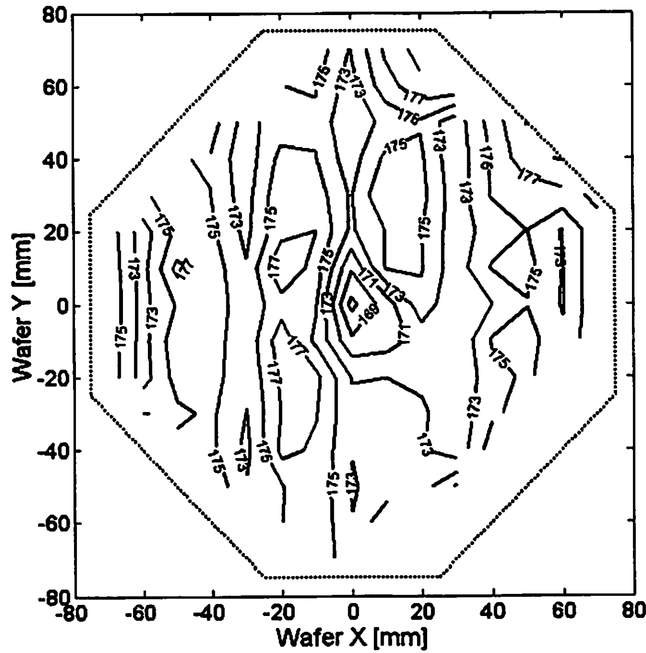

Figure 7.9: CD contour map for average wafer for vertical isolated lines without OPC. CD values are in nm, wafer dimensions are in mm.

develop fluid is dispensed onto the wafer. An ANOVA model is again used to model this effect. This model has the form

$$CD(x_w, y_w) = \mu + w_{spot}(x_w, y_w) + \varepsilon_{x_w, y_w}, \tag{7.6}$$

where $w_{spot}(x_w, y_w)$ is the treatment effect of the center spot.

A bivariate Gaussian function is used as the treatment model. This model is similar in form to the model in Equations 7.3– 7.5 with the exception that the coordinates now refer to position on the wafer rather than position within

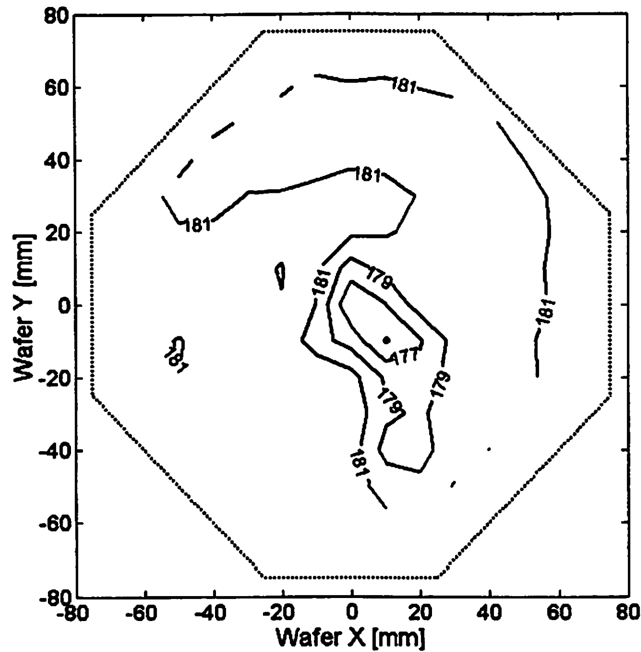

Figure 7.10: CD contour map for average wafer with mask errors (magnified by MEF) and across-field systematic variation models removed for vertical isolated lines without OPC. CD values are in nm, wafer dimensions are in mm.

the field. A summary of this model fit is shown in Table 7.6, and an ANOVA table is shown in Table 7.7.

The across-wafer bivariate Gaussian model is then removed from the data as shown in Figure 7.11.

At this point it is possible to fit an ANOVA model to the remaining across-wafer systematic variation. This model takes the form

$$CD(x_w, y_w) = \mu + w(x_w, y_w) + \varepsilon_{x_w, y_w}, \tag{7.7}$$

Table 7.6: Summary of across-wafer systematic variation model fit for bivariate Gaussian model.

Parameter	Estimate
scale	-3761.68
constant	181.35 nm
x_c	6.99 mm
y_c	-5.65 mm
σ_x	5.38 mm
σ_y	7.12 mm
σ_{xy}	-3.73 mm^2

Table 7.7: ANOVA table for CD-SEM field spot-effect Gaussian model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	7	1.12×10^4	1600	85.27
Error	11565	2.17×10^5	18.76	**p>F**
Total	11572	2.28×10^5		< 0.0001

where $w(x_w, y_w)$ is the treatment effect of across-wafer systematic variation.The function chosen to model the treatment effect is a radially-symmetric second-order polynomial with tilt terms in x and y of the form

$$CD(x_w, y_w) = a * r_w^2 + b * r_w + c * x_w + d * y_w + intercept \tag{7.8}$$

where x_w and y_w are coordinates across the wafer, r_w is the radius from the center of the wafer, and a, b, c, d, and *intercept* are paramters to be estimated using least squares. The parameter estimates from fitting this model to the data shown in Figure 7.11 are summarized in Table 7.8. An ANOVA table for the model is shown in Table 7.9.

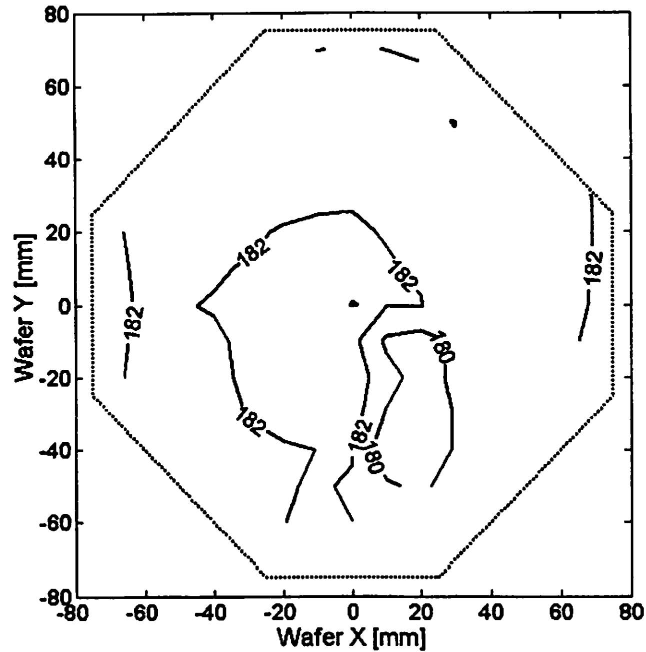

Figure 7.11: CD contour map for average wafer with mask errors (magnified by MEF), across-field systematic variation models, and across-wafer bivariate Gaussian model removed for vertical isolated lines without OPC. CD values are in nm, wafer dimensions are in mm.

Table 7.8: Summary of across-wafer systematic variation model fit for polynomial model shown in Equation 7.8.

Parameter	**Estimate**	**Probability > $\|t\|$**	**95% Confidence Interval**
Intercept	184.90 nm	0.0000	±0.445 nm
r_w^2	0.002 nm/mm^2	< 0.0001	±0.0002 nm/mm^2
r_w	-0.19 nm/mm	< 0.0001	±0.020 nm/mm
x_w	-0.08 nm/mm	< 0.0001	±0.005 nm/mm
y_f	-0.002 nm/mm^2	0.0088	±0.002 nm/mm^2

Table 7.9: ANOVA table for CD-SEM across-wafer polynomial model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	4	8179.58	2044.90	113.4906
Error	11568	208434.52	18.02	$p>F$
Total	11572	216614.11		< 0.0001

The results of removing the polynomial model of across-wafer systematic variation from the data is shown in Figure 7.12.

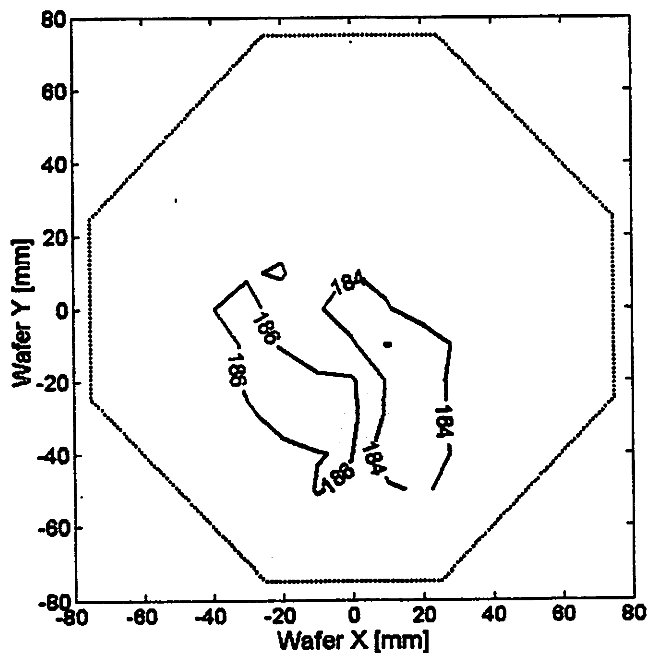

Figure 7.12: CD contour map for average wafer with mask errors (magnified by MEF), across-field systematic variation models, and across-wafer bivariate Gaussian and polynomial models removed for vertical isolated lines without OPC. CD values are in nm, wafer dimensions are in mm.

In order to compare the variation removed by each step in this process, Table

Table 7.10: Comparison of variance for each successive step in the analysis of across-wafer systematic variation.

Average wafer	$\hat{\sigma}^2 = (3.34\text{nm})^2$
... after removal of mask errors and polynomial model of across-field systematic variation	$\hat{\sigma}^2 = (1.72\text{nm})^2$
... after additional removal of bivariate Gaussian model for wafer spot effect	$\hat{\sigma}^2 = (1.34\text{nm})^2$
... after additional removal of polynomial model of across-wafer systematic variation	$\hat{\sigma}^2 = (1.05\text{nm})^2$

7.10 shows the estimated variance remaining in the data at each successive step. After the last step, there is no clear geometric pattern remaining in the data. The remaining variance is assumed to be the random. These components are shown by a pie chart in Figure 7.13.

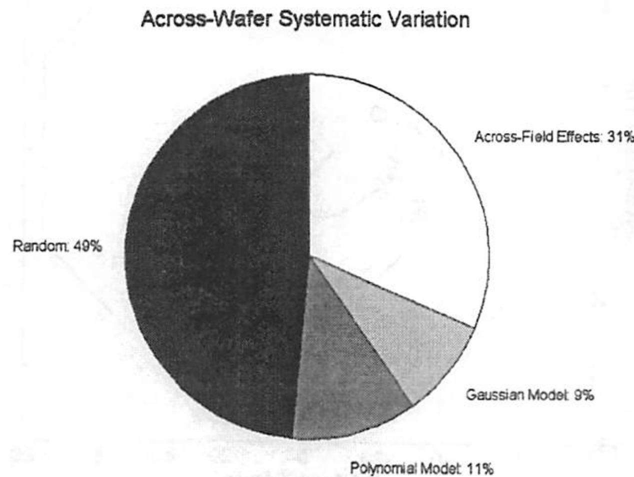

Figure 7.13: Components of across-wafer systematic variation. Across-field effects includes mask errors and polynomial and bivariate Gaussian models of across-field systematic variation.

7.1.3 Optimum Sampling Plan Selection

The first step in selecting an optimum sampling plan is to determine the maximum number of measurements per wafer permissible for the given application. Then simulations can be run to test the sampling plan. Random data sets are generated by combining the calculated systematic variation models with random noise. The amount of random noise should be chosen so that the distribution of the random data set reasonably matches the distribution of the actual measurements (shown in Figure 7.14) as closely as possible. The effectiveness of each plan will then be evaluated based on the width of the model parameter confidence intervals as estimated by the particular sampling plan.

This sampling plan analysis method was tested using the data from the example presented in Section 2. The models for across-field and across-wafer systematic variation were combined with random noise from a normally-distributed pseudo-random number generator to create random data sets for several different sampling plans. These plans were chosen to compare the effects of changing the sampling plan emphasis from capturing across-field systematic variation to across-wafer systematic variation. At one extreme, each point in the field is measured (294 points) but only two fields are measured. At the other extreme, only 25 points per field are measured, but all 23 fields on the wafer are measured. In all cases the plans are chosen so that the total number of measurements remains roughly constant at around 600 measurements, which might require a few hours of CD-SEM time, but could

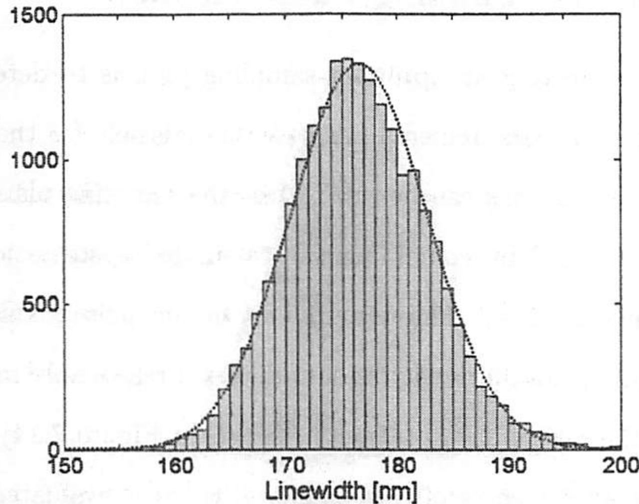

Figure 7.14: Histogram showing the distribution of line width measurements for vertical isolated lines without OPC. Note that the distribution is remarkably close to Normal (overlaid curve). The mean is 176.4 nm with a range of 34.7 nm (from 158.7 nm to 193.4 nm), and the 3σ variation is 18.2 nm.

Table 7.11: Candidate sampling plans chosen to evaluate the effects of changing the emphasis from capturing across-field to across-wafer systematic variation.

Plan Number	# points/field	# fields/wafer	Total # samples/wafer
1	294	2	588
2	147	4	588
3	77	7	539
4	47	13	611
5	25	23	575

be accelerated by an approximate factor of ten if one uses electrical linewidth metrology. The sampling plans are summarized in Table 7.11, while the details are shown in Figures 7.15–7.19.

Each plan is designed to have roughly the same number of measurements per

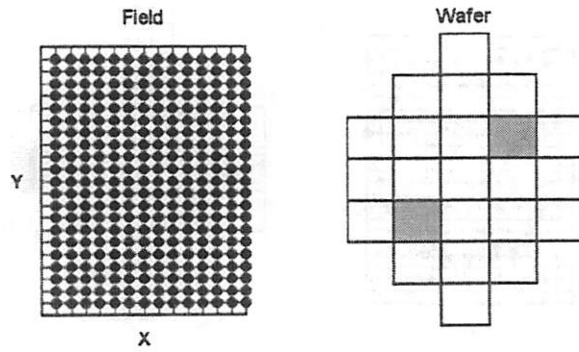

Figure 7.15: Candidate sampling plan 1.

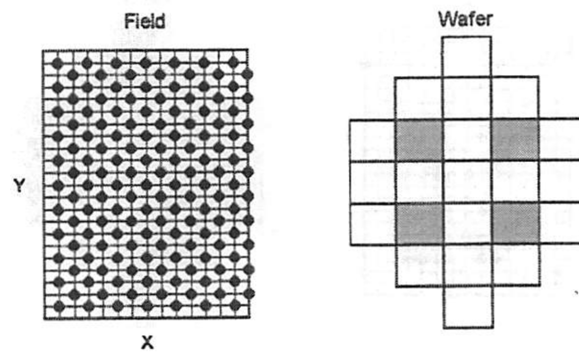

Figure 7.16: Candidate sampling plan 2.

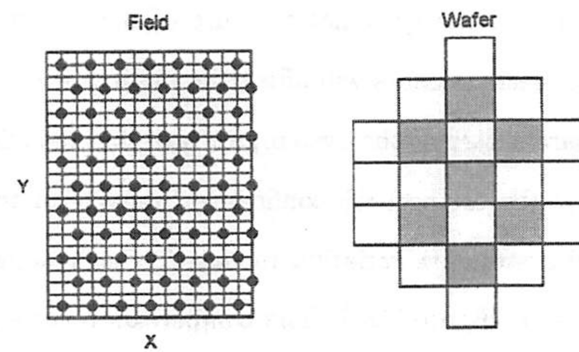

Figure 7.17: Candidate sampling plan 3.

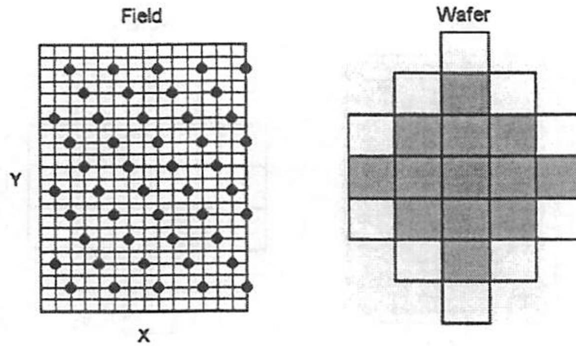

Figure 7.18: Candidate sampling plan 4.

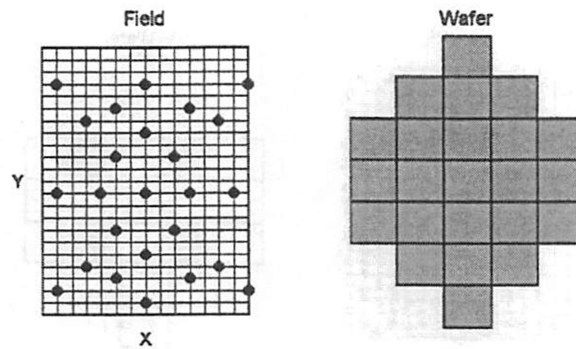

Figure 7.19: Candidate sampling plan 5.

wafer. However, due to the systematic nature of the variance, the spatial distribution of the measurements will affect the ability of the sampling plan to estimate the parameters of the systematic variation models. To demonstrate this, the widths of the 95% confidence intervals of the parameters of the across-field systematic variation models for each sampling plan are shown in a bar graph (Figure 7.20). The comparison reveals that the sampling plans vary in their ability to estimate the parameters of the across-field

systematic variation model. For example, Plan 4 is the "best" sampling plan for capturing the x_f term, but it is also the worst plan for capturing the $x_f * y_f$ interaction term. Therefore, plans must be carefully evaluated based on which model parameters are deemed the most important.

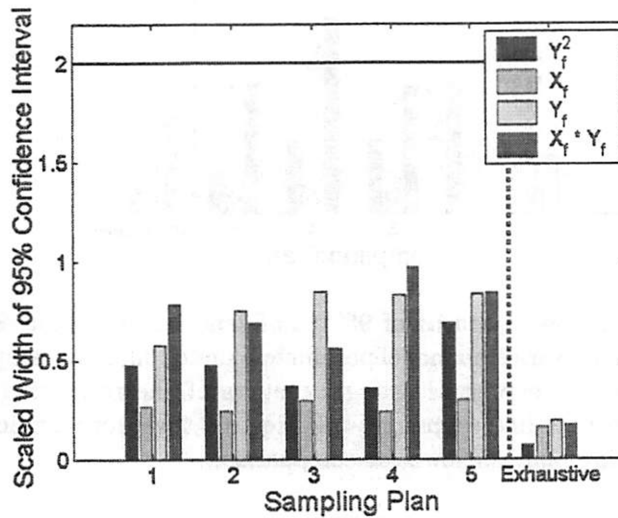

Figure 7.20: Comparison of widths of 95% confidence intervals for estimates of across-field systematic variation model parameters under different sampling plans. Confidence intervals are expressed as a percentage of the respective parameter estimate. The solid blue line represents the required level for significance. The "exhaustive" sampling plan is shown for comparison.

In addition, the widths of the 95% confidence intervals of the parameters of the across-wafer systematic variation models were plotted in a similar graph (Figure 7.21). None of the sampling plans were able to obtain a statistically significant estimate (at a 95% level) of the x_f term of the model. Plans 1 and 2 were both unable to estimate the y_f term of the model.

Figure 7.21 suggests that sampling plans with more fields per wafer (i.e. a

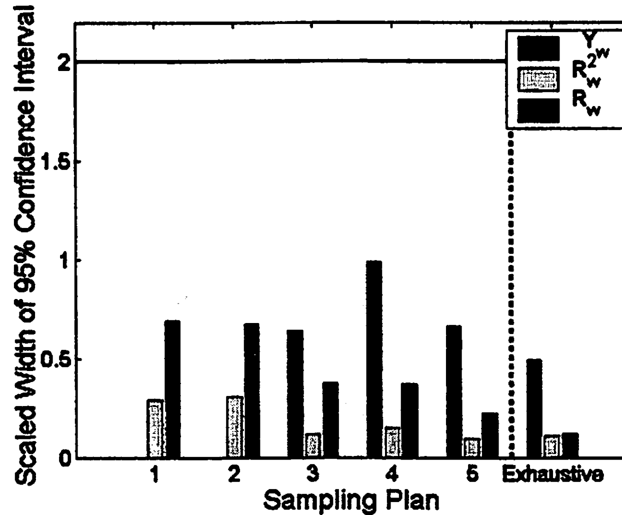

Figure 7.21: Comparison of widths of 95% confidence intervals for estimates of across-wafer systematic variation model parameters under different sampling plans. Confidence intervals are expressed as a percentage of the respective parameter estimate. The solid blue line represents the required level for significance. The "exhaustive" sampling plan is shown for comparison.

wider distribution of samples across the wafer) do a better job of estimating the across-wafer model parameters in general. However, this is not universally true, as Plans 3 and 4 perform comparably in estimating the r_w^2 and r_w terms, but Plan 4 is much worse than Plan 3 at estimating the y_w term, despite the fact that Plan 4 measures more fields than Plan 3.

This analysis shows that capturing across-field and across-wafer systematic variation involves two competing objectives: the need to measure many points per field to capture across-field variation and the need to spread the points across the wafer to capture across-wafer variation. These two objectives must be balanced against each other to find an optimum sampling plan.

In addition, individual plans must be evaluated to determine their ability to capture the most important model parameters. This plan can then be used to track systematic variation in the fabrication process over time. An additional possibility is to generate "mixed" plans, where a subset of the wafers is sampled with an emphasis in extracting the field variability, while another subset is sampled with an emphasis in wafer variability. Similar plans can be used for analyzing the within lot and across lot variability components.

7.2 ELM Experiment

7.2.1 Experiment Setup

The framework for characterizing systematic variation and developing an optimum sampling plan described in Chapters 5 and 6 was implemented for a series of wafers patterned with the ELM2 mask at a state-of-the-art industrial fab.

A number of eight-wafer sublots were exposed at different time intervals in order to study lot-to-lot variation at different time scales. After development, the wafers were randomized into new lots of 24 wafers to eliminate the effects of temporal variation in the lithography process from subsequent processing steps. These wafers were measured using CD-SEM and scatterometry after development, after resist trim etch, and after poly etch. In addition, the wafers were measured after poly etch using electrical linewidth metrology (ELM), which offers increased speed over CD-SEM. ELM has previously been

used to characterize systematic spatial variation [72, 73], however, this work offers more complete metrology at each processing step along with complete wafer-processing history. The use of multiple metrology methods allows for a comparison of the effectiveness of each method, as well as investigation of the correlation between methods. Also, care was taken to record the processing history of each wafer in detail in order to allow find similarities in processing which might account for a systematic variation signature. Two of these 24-wafer lots were included in this study: a baseline trim etch lot and a feed-forward control lot in which the trim etch time was varied to control the CD mean based on CD-SEM measurements after development.

7.2.2 Extraction of Systematic Variation Using ELM

After the completion of wafer processing and CD-SEM and scatterometry measurements, the wafers were measured using ELM in the UC Berkeley Device Characterization Lab. The goal was to apply the concepts in Chapters 5 and 6 to characterize the systematic variation in the two lots. In addition, the scatterometry and CD-SEM measurements, combined with the wafer processing history, can be used to separate different sources of systematic variation.

7.2.3 Exhaustive Sampling Plan

The exhaustive sampling plan is designed to allow a very high degree of spatial resolution to provide detailed variance maps. This plan measures DUT1 (horizontal isolated), DUT3 (vertical isolated), and DUT4 (vertical dense) features for the no-OPC submodules. These features were chosen to match the features measured on the mask. The OPC submodules were not sampled due to the fact that some OPC features (particularly dense features) did not print correctly.

Every module in the field (22 rows × 14 columns = 308 modules) was measured, and every field on the wafer (23 fields) was sampled. This means the total number of measurements in the exhaustive sampling plan is

$$3\frac{\text{features}}{\text{module}} \times 308\frac{\text{modules}}{\text{field}} \times 23\frac{\text{fields}}{\text{wafer}} = 21252\frac{\text{features}}{\text{wafer}} \quad . \tag{7.9}$$

The distribution of measurements in the exhaustive sampling plan is shown in Figure 7.22. The required measurement time for this sampling plan is about 24 hours and 45 minutes per wafer. The first five wafers from each lot were measured using this plan.

7.2.4 Baseline Etch Lot Results

CD contour maps for the five exhaustively sampled wafers from the baseline trim etch lot are shown in Figures 7.23 through 7.27. Note that the CD values are much smaller than the 130 nm nominal CD, indicating that there

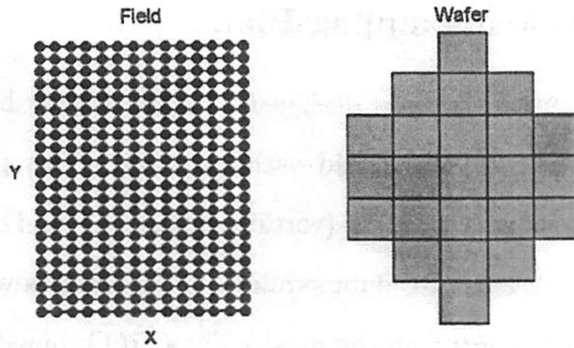

Figure 7.22: ELM exhaustive sampling plan.

is a significant bias between the physical CD and the electrical CD.

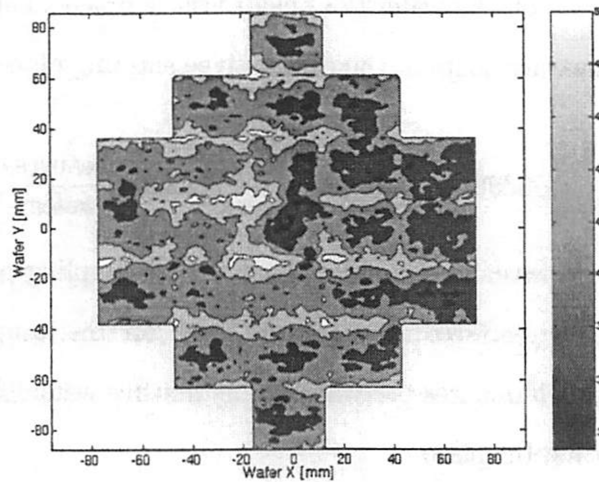

Figure 7.23: A CD contour map of the wafer in slot 1 of the baseline trim etch lot.

The first step in characterizing the systematic variation is to calculate the "average" wafer (Figure 7.28). This is calculated by averaging the ELM measurements for the five exhaustively sampled wafers at each point on the

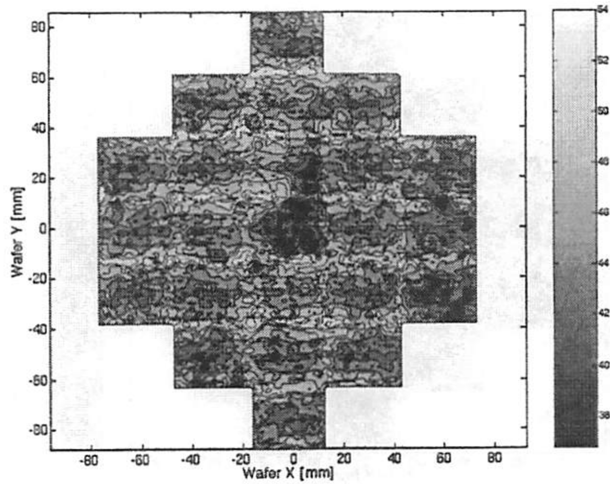

Figure 7.24: A CD contour map of the wafer in slot 2 of the baseline trim etch lot.

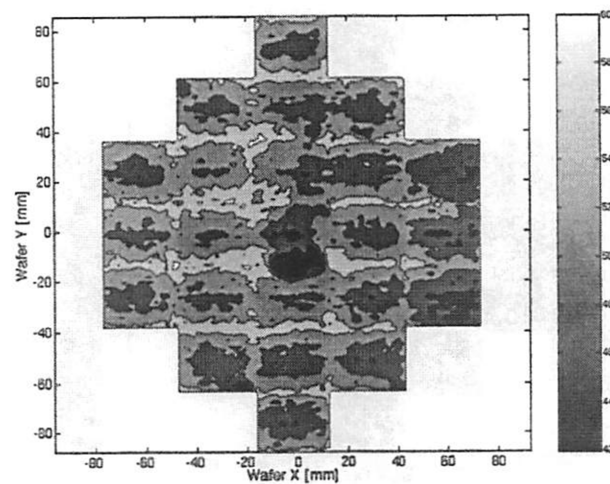

Figure 7.25: A CD contour map of the wafer in slot 3 of the baseline trim etch lot.

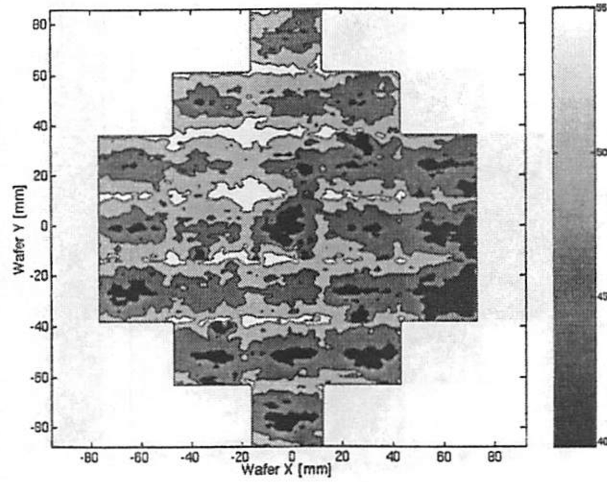

Figure 7.26: A CD contour map of the wafer in slot 4 of the baseline trim etch lot.

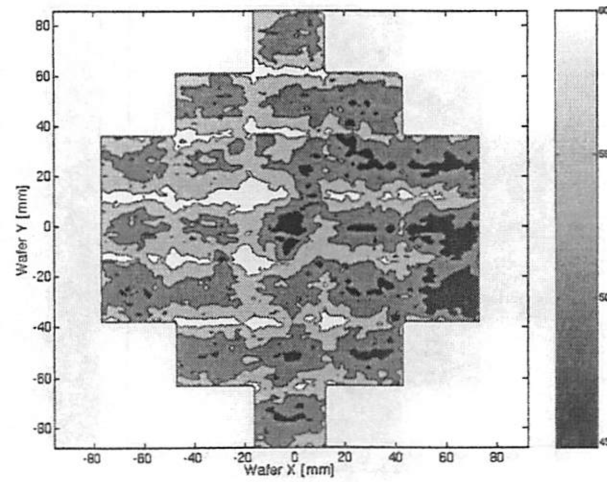

Figure 7.27: A CD contour map of the wafer in slot 5 of the baseline trim etch lot.

wafer. Inspection of the CD contour maps for the five wafers in Figures 7.23 through 7.27 shows that the wafers are similar enough to justify averaging them together.

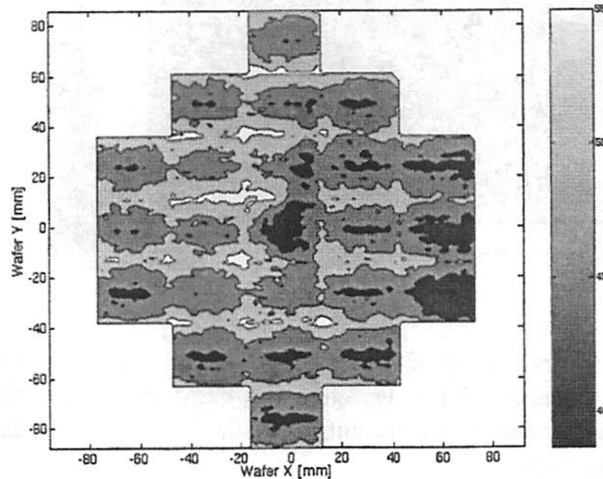

Figure 7.28: A CD contour map of the "average" wafer from the baseline trim etch lot. The average is computed by averaging the ELM measurements for the five exhaustively sampled wafers at each point on the wafer.

Inspection of the average wafer data reveals that there is a spot effect in the center of the wafer where the CD is smaller than the surrounding fields. Also, the right edge shows a smaller CD than the rest of the wafer.

The next step is to use this information to calculate the "average" field by averaging the CD measurements for the five wafers at each point in the field (Figure 7.29). Fields which seem atypical (those in the center and at the far right edge) are excluded in calculating the average field.

At this point measurements of the ELM2 mask can be used to isolate the

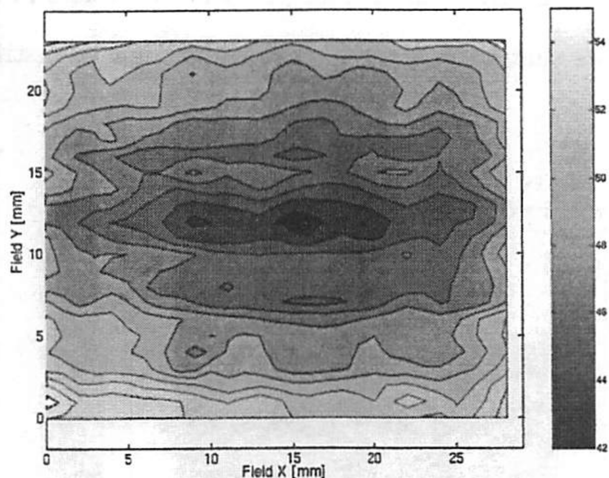

Figure 7.29: A CD contour map of the "average" field from the baseline trim etch lot. The average is computed by averaging the ELM measurements for the five exhaustively sampled wafers at each point in the field. Atypical fields are excluded in calculating the average.

effects of mask errors. As mentioned in Section 3.2.1, the mask errors are magnified by the mask error factor (MEF). However, the MEF applies to the aerial image, whereas electrical measurments are performed after several subsequent processing steps. Therefore, the aerial image MEF may not be the correct factor to use in scaling the mask errors.

The mask error scaling factor is therefore chosen to be the factor which minimizes the variance of the residuals when the scaled mask errors are subtracted from the average field data. Figure 7.30 shows the variance of the residuals for varying scaling factors. The value of the scaling factor which minimizes the variance of the residuals is found to be 0.68.

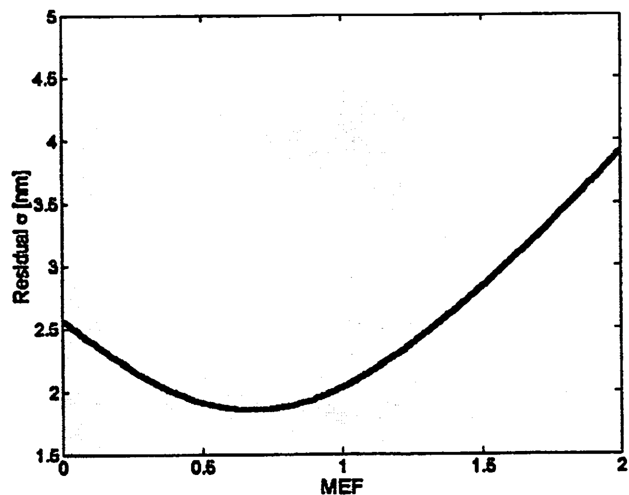

Figure 7.30: Plot used to find the mask error scaling factor which minimizes the variance of the residuals when the scaled mask errors are subtracted from the average field data.

The mask errors scaled by the appropriate factor are shown in Figure 7.31. The scaled mask errors are then subtracted from the average field data (Figure 7.32). The remaining variance is assumed to be the result of other imperfections in the imaging system (e.g. lens aberrations, illumination variations, etc.).

An ANOVA model of the form shown in Equation 7.1 is again used in an attempt to extract the across-field systematic variation. The model chosen for the treatment effect is of the form

$$f(x_f, y_f) = a \cdot x_f^2 + b \cdot y_f^2 + c \cdot x_f + d \cdot y_f \quad , \qquad (7.10)$$

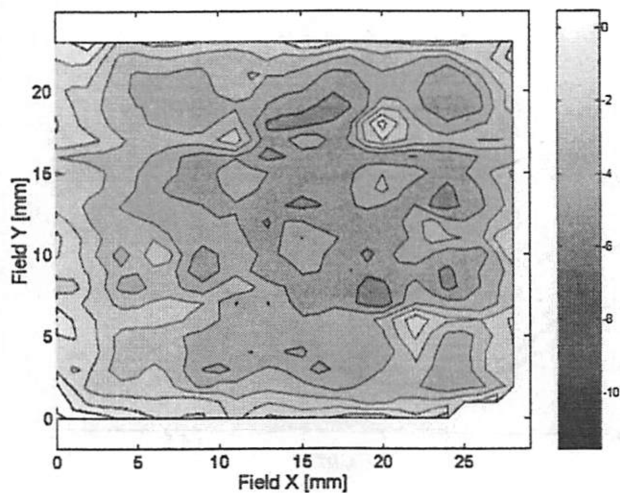

Figure 7.31: DUT3 (vertical isolated) feature mask errors scaled by the factor which minimizes variance in the residuals when the scaled mask errors are subtracted from the average field.

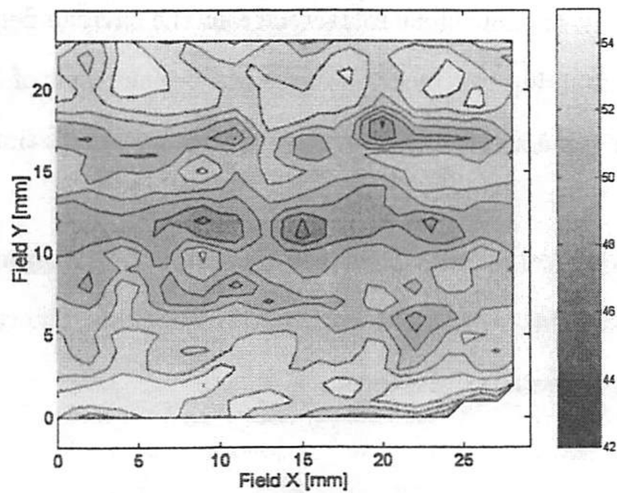

Figure 7.32: Average field with scaled mask errors removed.

Table 7.12: Summary of across-field systematic variation model fit for polynomial model shown in Equation 7.10.

Parameter	Estimate	Standard Error	p-value
Intercept	55.58 nm	0.0515	$p < 0.0001$
x_f^2	0.0019 nm/mm^2	0.0004	$p < 0.0001$
y_f^2	0.0264 nm/mm^2	0.0006	$p < 0.0001$
x_f	-0.0408 nm/mm	0.0122	$p = 0.0008$
y_f	-0.6225 nm/mm	0.0154	$p < 0.0001$

Table 7.13: ANOVA table for XWFR_B across-field polynomial model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	5	0.521×10^5	10420	369.15
Error	34414	9.714×10^5	28.227	p>F
Total	34419	1.0235×10^6		< 0.0001

where (x_f, y_f) are coordinates within the field in millimeters and a, b, c, and d are coefficients to be estimated using linear regression. The parameter estimates from the model fit are shown in Table 7.12, while an ANOVA table for the fit is shown in Table 7.13. In addition, a graphical representation of this model is shown in Figure 7.33.

This across-field systematic variation model is then subtracted from the average field data with scaled mask errors removed. The results are shown in Figure 7.34.

The scaled mask errors and the polynomial model of across-field systematic variation can then be subtracted from each field in the average wafer data shown in Figure 7.28. The results are shown in Figure 7.35.

The wafer map in Figure 7.35 reveals a spot effect in the center of the wafer

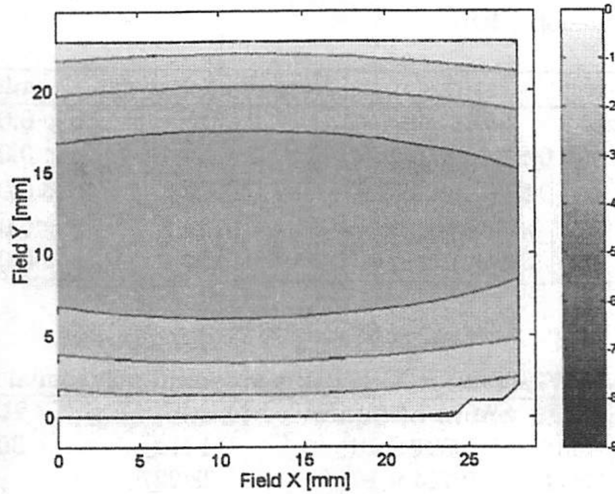

Figure 7.33: Graphical representation of second-order polynomial model (Eq. 7.10) of across-field systematic variation for DUT3 (vertical isolated) features.

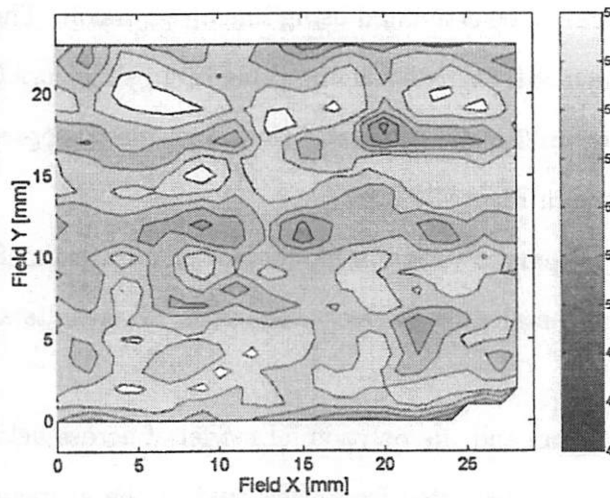

Figure 7.34: Residuals after the second-order polynomial model (Eq. 7.10) of across-field systematic variation is removed from average field data with scaled mask errors removed.

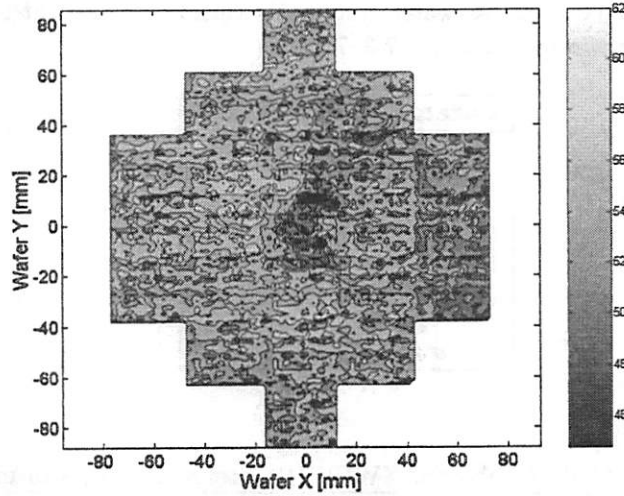

Figure 7.35: Average wafer after removal of scaled mask errors and polynomial model of across-field systematic variation from each field on the wafer.

where the CD is smaller than the surrounding area. An ANOVA model of the form shown in Equation 7.6 is again used to model this effect. This function is of the form shown in Equations 7.3 through 7.5 is used (with (x_f, y_f) replaced with (x_w, y_w) to represent coordinates across the wafer rather than across the field). The parameter estimates from the model fit are shown in Table 7.14, and an ANOVA table is shown in Table 7.15. In addition, a graphical representation of the model is shown in Figure 7.36.

The bivariate Gaussian model is then subtracted from the average wafer with across-field effects removed, and the results are shown in Figure 7.37.

Finally, another ANOVA model of the form shown in Equation 7.7 is used to extract the remaining across-wafer systematic variation. A second-order

Table 7.14: Summary of across-wafer systematic variation model fit for bivariate Gaussian model shown in Equations 7.3–7.5.

Parameter	Estimate
scale	-1391.04
constant	54.51 nm
x_c	1.71 mm
y_c	6.26 mm
σ_x	5.13 mm
σ_y	8.17 mm
σ_{xy}	27.12 mm^2

Table 7.15: ANOVA table for XWFR_B center spot Gaussian model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	7	9.2×10^3	1314.29	47.06
Error	34414	9.611×10^5	27.93	**p>F**
Total	34408	9.714×10^5		< 0.0001

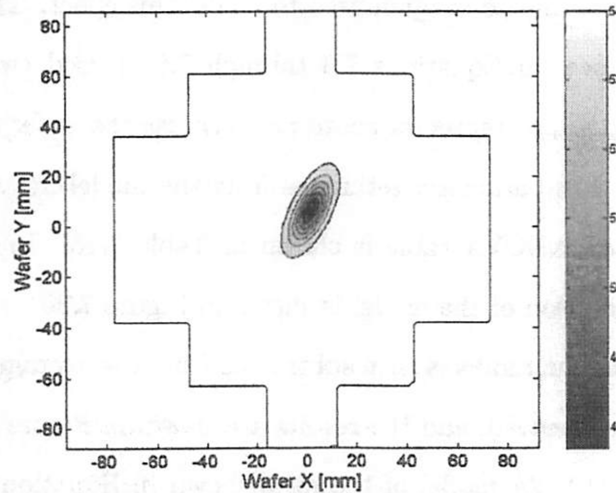

Figure 7.36: Average wafer after removal of scaled mask errors and polynomial model of across-field systematic variation from each field on the wafer.

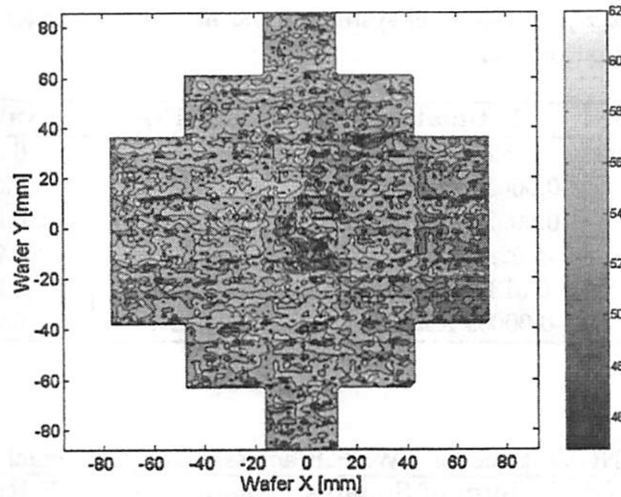

Figure 7.37: Average wafer after removal of across-field systematic variation and bivariate Gaussian spot effect model.

polynomial model was chosen model this treatment, and it is of the form

$$w(x_w, y_w) = a \cdot x_w^2 + b \cdot y_w^2 + c \cdot x_w + d \cdot y_w + e \cdot x_w \cdot y_w \quad , \tag{7.11}$$

where (x_w, y_w) are coordinates on the wafer in millimeters and a, b, c, d, and e are coefficients to be estimated using linear regression. The parameter estimates from the model fit are shown in Table7.16, and an ANOVA table for the fit is shown in Table 7.17. A graphical representation is shown in Figure 7.38.

This across-wafer polynomial model is subsequently removed from the data, and the results are shown in Figure 7.39. The remaining variance is assumed to be random. A summary ANOVA table of the systematic variation

Table 7.16: Summary of across-wafer systematic variation model fit for polynomial model shown in Equation 7.11.

Parameter	Estimate	Standard Error	p-value
Intercept	55.59 nm	0.0515	$p < 0.0001$
x_w^2	-0.00055 nm/mm^2	0.00002	$p < 0.0001$
y_w^2	-0.00021 nm/mm^2	0.00002	$p < 0.0001$
x_w	-0.0281 nm/mm	0.0007	$p < 0.0001$
y_w	0.0137 nm/mm	0.0007	$p < 0.0001$
$x_w \cdot y_w$	-0.00009 nm/mm	0.00003	$p = 0.0002$

Table 7.17: ANOVA table for XWFR_B across-wafer polynomial model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	6	6.72×10^4	11200	430.99
Error	34402	8.94×10^5	25.99	**p>F**
Total	34408	9.61×10^5		< 0.0001

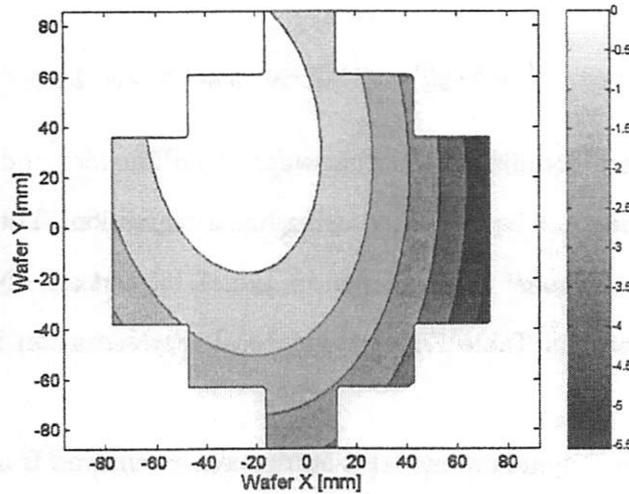

Figure 7.38: Graphical representation of second-order polynomial model (Eq. 7.11) of across-wafer systematic variation for DUT3 (vertical isolated) features.

Table 7.18: Summary ANOVA table for XWFR_B systematic variation.

Source	DOF	Sum of Squares	Mean Square	F Ratio	p-value
Mask Errors	305	2.36×10^5	773.77	29.77	$p < 0.0001$
Field Polynomial	5	5.21×10^4	10420	400.92	$p < 0.0001$
Wafer Gaussian	7	9.20×10^3	1314.29	50.57	$p < 0.0001$
Wafer Polynomial	6	6.72×10^4	11200	430.93	$p < 0.0001$
Error	34401	8.94×10^5	25.99		
Total	34724	1.13×10^6			

extracted from the XWFR_B lot is shown in Table 7.18.

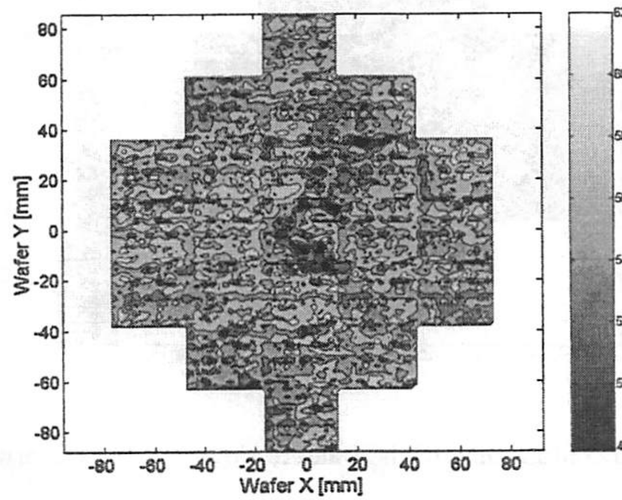

Figure 7.39: Residuals after removal of scaled mask errors, polynomial across-field systematic variation model, and Gaussian and polynomial across-wafer systematic variation models.

7.2.5 Feed-Forward Control Results

CD contour maps for the five exhaustively sampled wafers from the baseline trim etch lot are shown in Figures 7.40 through 7.44. Note that the CD values are much smaller than the 130 nm nominal CD, indicating that there is a significant bias between the physical CD and the electrical CD.

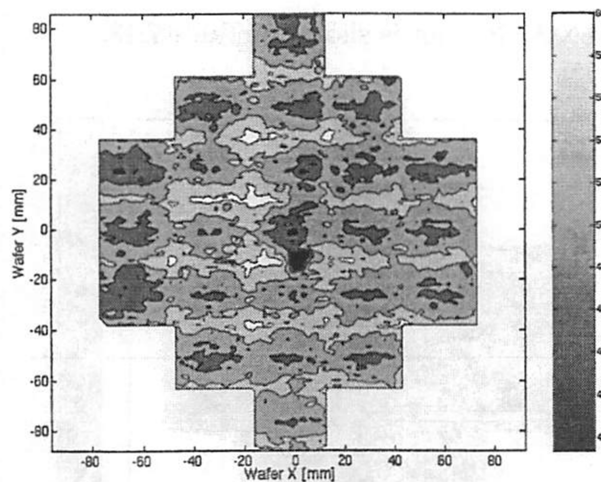

Figure 7.40: A CD contour map of the wafer in slot 1 of the feed-forward control lot.

The first step in characterizing the systematic variation is to calculate the "average" wafer (Figure 7.45). This is calculated by averaging the ELM measurements for the five exhaustively sampled wafers at each point on the wafer. Inspection of the CD contour maps for the five wafers in Figures 7.40 through 7.44 shows that the wafers are similar enough to justify averaging them together.

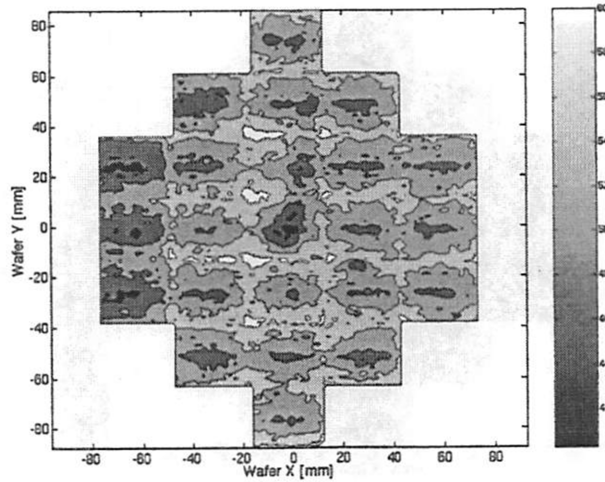

Figure 7.41: A CD contour map of the wafer in slot 2 of the feed-forward control lot.

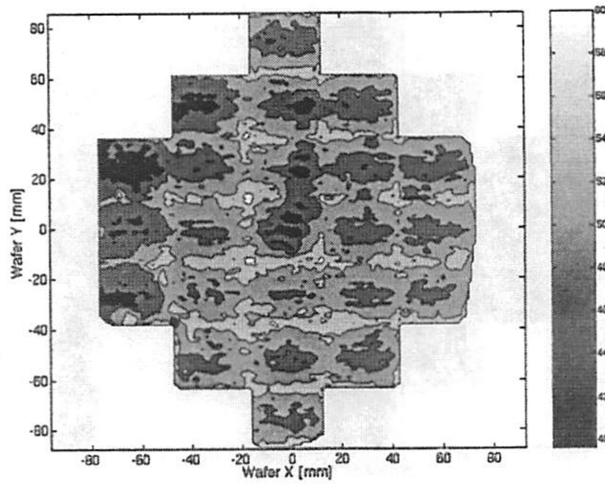

Figure 7.42: A CD contour map of the wafer in slot 3 of the feed-forward control lot.

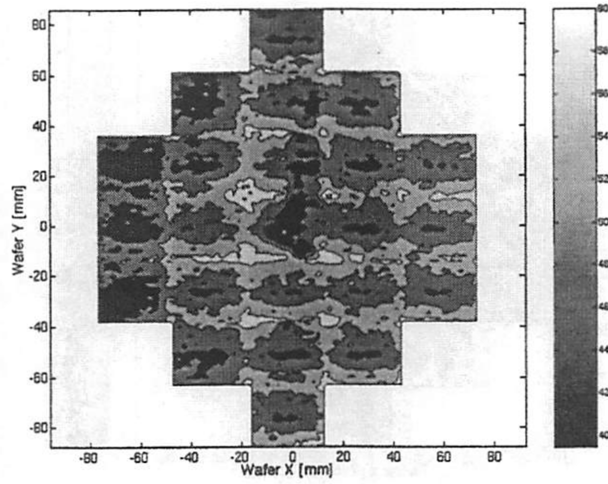

Figure 7.43: A CD contour map of the wafer in slot 4 of the feed-forward control lot.

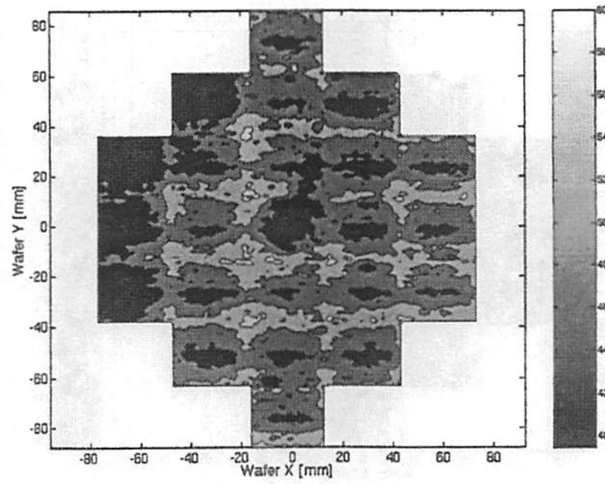

Figure 7.44: A CD contour map of the wafer in slot 5 of the feed-forward control lot.

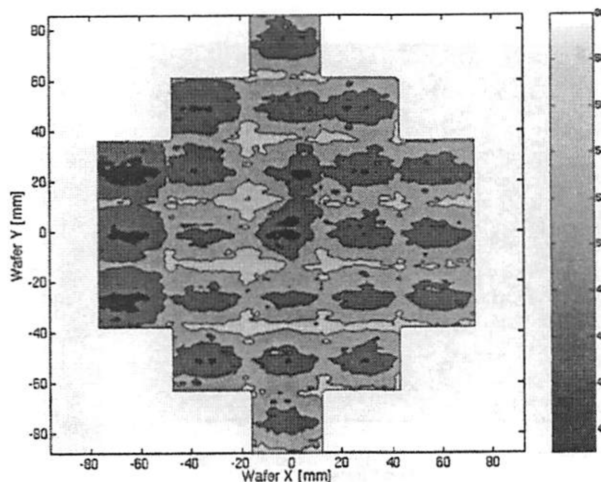

Figure 7.45: A CD contour map of the "average" wafer from the feed-forward control lot. The average is computed by averaging the ELM measurements for the five exhaustively sampled wafers at each point on the wafer.

Inspection of the average wafer data reveals that there is a spot effect in the center of the wafer where the CD is smaller than the surrounding fields. Also, the left edge shows a smaller CD than the rest of the wafer. Note that this is the opposite edge as the baseline trim etch lot.

The next step is to use this information to calculate the "average" field by averaging the CD measurements for the five wafers at each point in the field (Figure 7.46). Fields which seem atypical (those in the center and at the far left edge) are excluded in calculating the average field.

At this point measurements of the ELM2 mask can be used to isolate the effects of mask errors. As mentioned in Section 3.2.1, the mask errors are magnified by the mask error factor (MEF). However, the MEF applies to

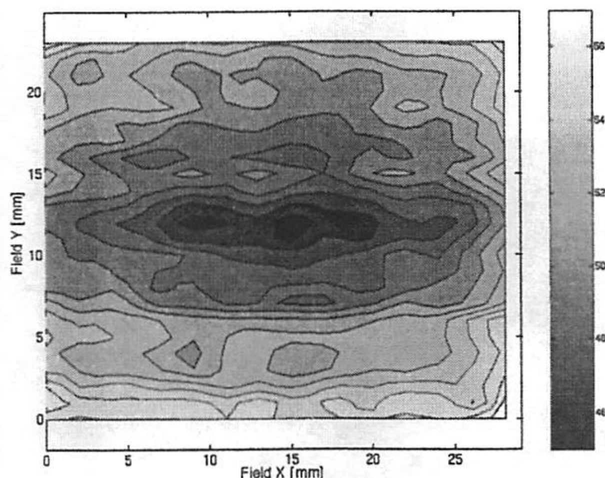

Figure 7.46: A CD contour map of the "average" field from the feed-forward control lot. The average is computed by averaging the ELM measurements for the five exhaustively sampled wafers at each point in the field. Atypical fields are excluded in calculating the average.

the aerial image, whereas electrical measurments are performed after several subsequent processing steps. Therefore, the aerial image MEF may not be the correct factor to use in scaling the mask errors.

The mask error scaling factor is therefore chosen to be the factor which minimizes the variance of the residuals when the scaled mask errors are subtracted from the average field data. Figure 7.47 shows the variance of the residuals for varying scaling factors. The value of the scaling factor which minimizes the variance of the residuals is found to be 0.64.

The mask errors scaled by the appropriate factor are shown in Figure 7.48. The scaled mask errors are then subtracted from the average field data (Fig-

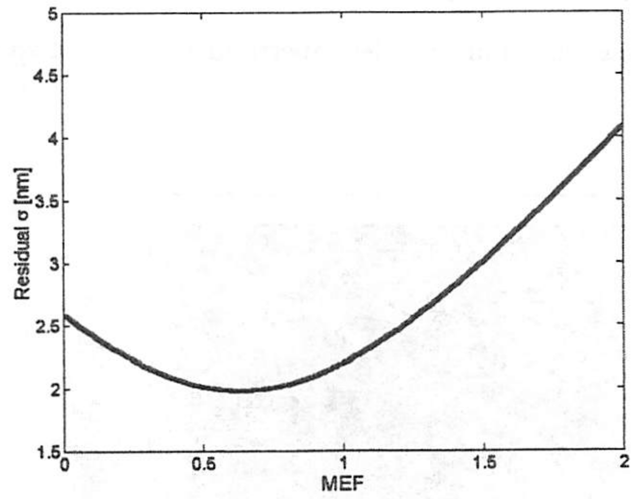

Figure 7.47: Plot used to find the mask error scaling factor which minimizes the variance of the residuals when the scaled mask errors are subtracted from the average field data.

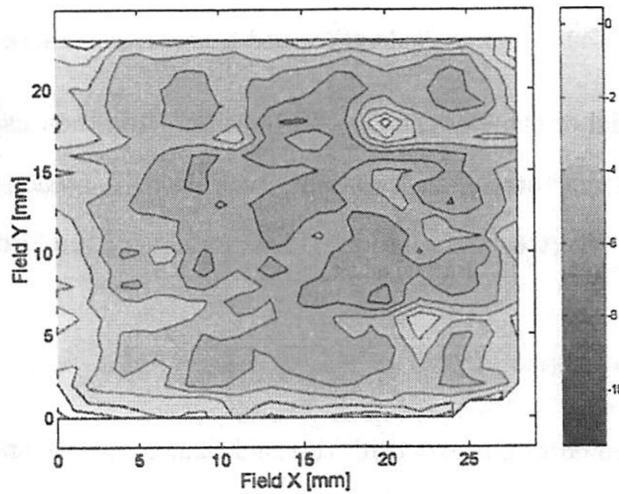

Figure 7.48: DUT3 (vertical isolated) feature mask errors scaled by the factor which minimizes variance in the residuals when the scaled mask errors are subtracted from the average field.

ure 7.49). The remaining variance is assumed to be the result of other imperfections in the imaging system (e.g. lens aberrations, illumination variations, etc.).

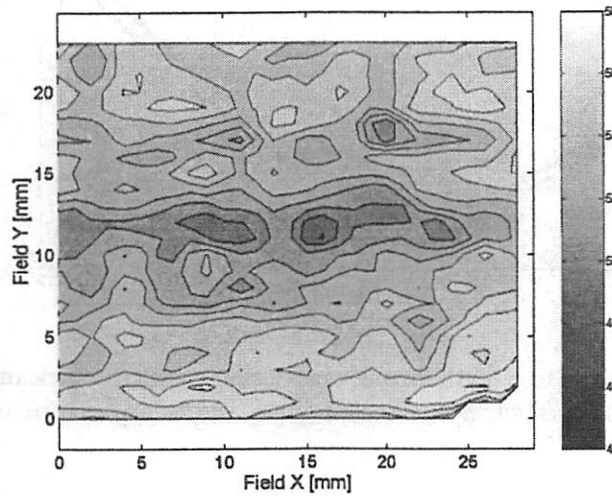

Figure 7.49: Average field with scaled mask errors removed.

An ANOVA model of the form shown in Equation 7.1 is then used in an attempt to extract the across-field systematic variation. A second-order polynomial is used as the treatment model. This model is of the form

$$CD = a \cdot x_f^2 + b \cdot y_f^2 + c \cdot x_f + d \cdot y_f + e \cdot x_f^2 \cdot y_f^2 + f \quad , \tag{7.12}$$

where (x_f, y_f) are coordinates within the field in millimeters and a, b, c, d, e, and f are coefficients to be estimated using linear regression. The results of the model fit are shown in Table 7.19, and an ANOVA table for the fit is shown in Table 7.20. In addition, a graphical representation of this model is

Table 7.19: Summary of across-field systematic variation model fit for polynomial model shown in Equation 7.12.

Parameter	Estimate	Standard Error	p-value
Intercept	56.00 nm	0.0927	$p < 0.0001$
x_f^2	0.0053 nm/mm^2	0.0003	$p < 0.0001$
y_f^2	0.0242 nm/mm^2	0.0004	$p < 0.0001$
x_f	-0.0587 nm/mm	0.0095	$p < 0.0001$
y_f	-0.6004 nm/mm	0.0118	$p < 0.0001$
$x_f \cdot y_f$	-0.0015 nm/mm^2	0.0003	$p < 0.0001$

Table 7.20: ANOVA table for MIX1 across-field polynomial model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	6	6.24×10^4	10400	727.68
Error	33935	4.85×10^5	14.29	p>F
Total	33941	5.47×10^5		< 0.0001

shown in Figure 7.50.

This systematic variation model is then subtracted from the average field data with scaled mask errors removed. The results are shown in Figure 7.51. The scaled mask errors and the polynomial model of across-field systematic variation can then be subtracted from each field in the average wafer data shown in Figure 7.45. The results are shown in Figure 7.52.

The wafer map in Figure 7.52 reveals a spot effect in the center of the wafer where the CD is smaller than the surrounding area. An ANOVA model of the form shown in Equation 7.6 is used to model this effect with a bivariate Gaussian of the form shown in Equations 7.3 through 7.5 (with (x_f, y_f) replaced with (x_w, y_w) to represent coordinates across the wafer rather than

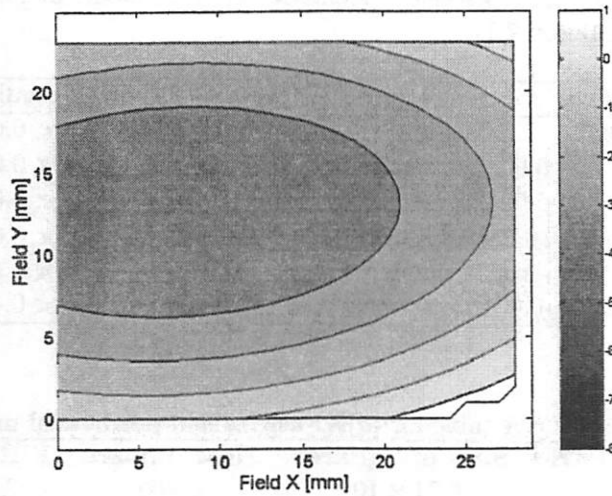

Figure 7.50: Graphical representation of second-order polynomial model (Eq. 7.10) of across-field systematic variation for DUT3 (vertical isolated) features.

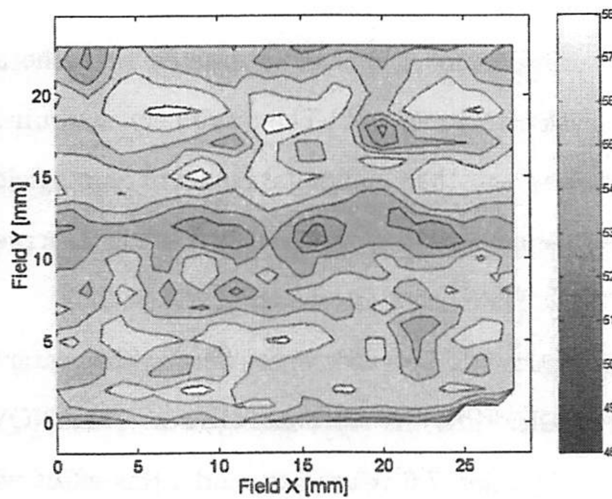

Figure 7.51: Residuals after the second-order polynomial model (Eq. 7.10) of across-field systematic variation is removed from average field data with scaled mask errors removed.

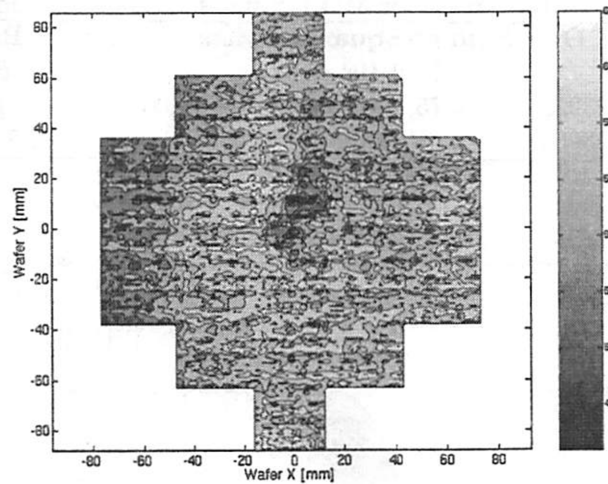

Figure 7.52: Average wafer after removal of scaled mask errors and polynomial model of across-field systematic variation from each field on the wafer.

Table 7.21: Summary of across-wafer systematic variation model fit for bivariate Gaussian model shown in Equations 7.3–7.5.

Parameter	Estimate
scale	-1204.92
constant	56.07 nm
x_c	4.42 mm
y_c	10.56 mm
σ_x	4.87 mm
σ_y	8.24 mm
σ_{xy}	24.01 mm^2

across the field) used to model the treatement effect. The results of the model fitting are shown in Table 7.21. In addition, a graphical representation of the model is shown in Figure 7.53.

The bivariate Gaussian model is subsequently subtracted from the average

Table 7.22: ANOVA table for MIX1 center spot Gaussian model.

Source	**DOF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	7	6.79×10^3	970	69.28
Error	33928	4.75×10^5	14.00	p>F
Total	33935	4.85×10^5		< 0.0001

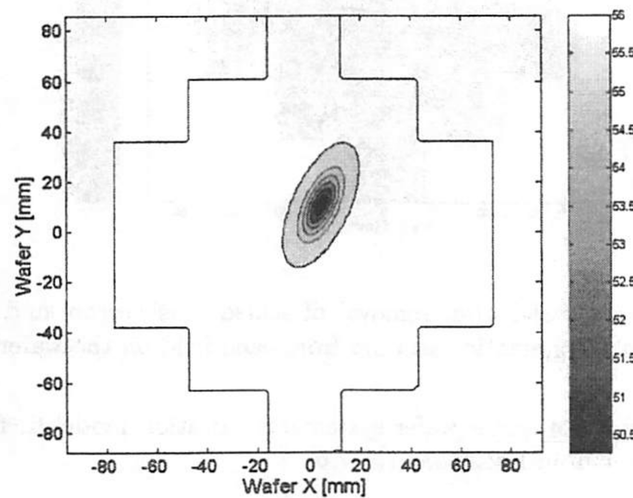

Figure 7.53: Average wafer after removal of scaled mask errors and polynomial model of across-field systematic variation from each field on the wafer.

wafer with across-field effects removed, and the results are shown in Figure 7.54.

Finally, another ANOVA model of the form shown in Equation 7.7 was used in an attempt to extract the across-wafer systematic variation. A second-order polynomial model was chosen to model the treatment effect. This polynomial model is of the form

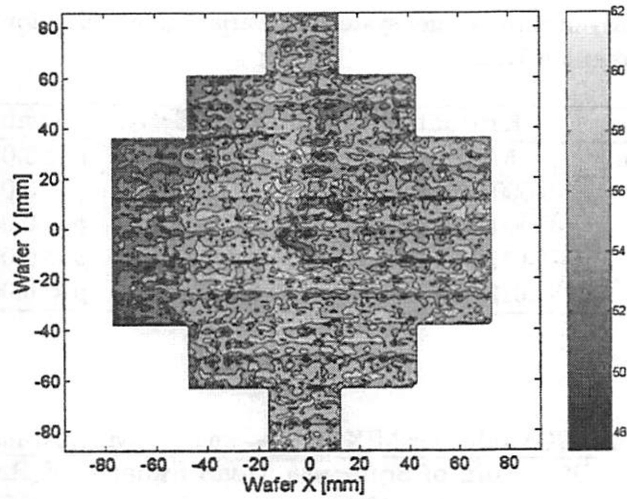

Figure 7.54: Average wafer after removal of across-field systematic variation and bivariate Gaussian spot effect model.

$$w(x_w, y_w) = a \cdot x_w^2 + b \cdot y_w^2 + c \cdot x_w + d \cdot y_w + e, \tag{7.13}$$

where (x_w, y_w) are coordinates on the wafer in millimeters and a, b, c, d, and e are coefficients to be estimated using linear regression. The results of the model fit are shown in Table 7.23, and an ANOVA table for the model fit is shown in Table 7.24. A graphical representation is shown in Figure 7.55.

This across-wafer polynomial model is subsequently removed from the data, and the results are shown in Figure 7.56. The remaining variance is assumed to be random.

Table 7.23: Summary of across-wafer systematic variation model fit for polynomial model shown in Equation 7.11.

Parameter	Estimate	Standard Error	p-value
Intercept	57.31 nm	0.0356	$p < 0.0001$
x_w^2	-0.0006 nm/mm^2	0.00001	$p < 0.0001$
y_w^2	-0.0002 nm/mm^2	0.00001	$p < 0.0001$
x_w	-0.0190 nm/mm	0.0005	$p < 0.0001$
y_w	-0.0105 nm/mm	0.0005	$p < 0.0001$

Table 7.24: ANOVA table for MIX1 across-wafer polynomial model.

Source	DOF	Sum of Squares	Mean Square	F Ratio
Model	5	5.87×10^4	11740	957.35
Error	33923	4.16×10^5	12.26	**p>F**
Total	33928	4.75×10^5		< 0.0001

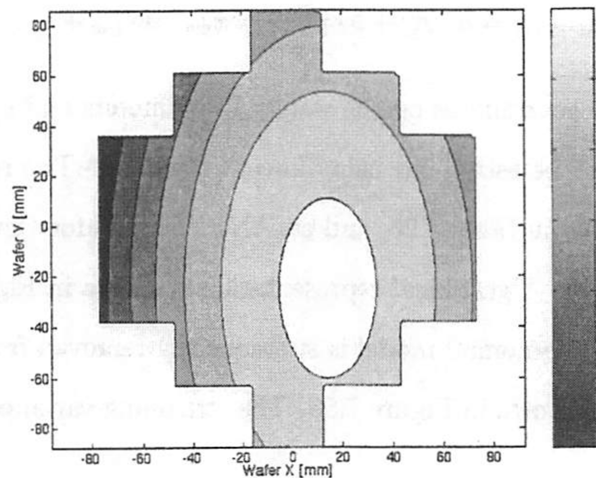

Figure 7.55: Graphical representation of second-order polynomial model (Eq. 7.13) of across-wafer systematic variation for DUT3 (vertical isolated) features.

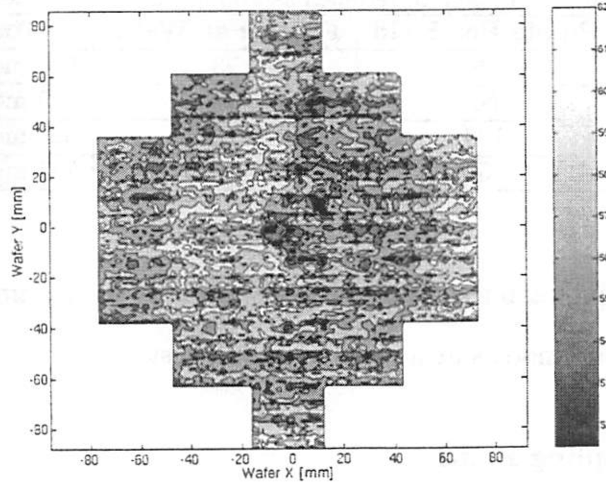

Figure 7.56: Residuals after removal of scaled mask errors, polynomial across-field systematic variation model, and Gaussian and polynomial across-wafer systematic variation models.

Table 7.25: Summary ANOVA table for MIX1 systematic variation.

Source	DOF	Sum of Squares	Mean Square	F Ratio	p-value
Mask Errors	305	2.34×10^5	767.21	62.56	$p < 0.0001$
Field Polynomial	6	6.24×10^4	10400	848.29	$p < 0.0001$
Wafer Gaussian	7	6.79×10^3	970	79.12	$p < 0.0001$
Wafer Polynomial	5	5.87×10^4	11740	957.58	$p < 0.0001$
Error	33923	4.16×10^5	12.26		
Total	34247	1.13×10^6			

7.2.6 Optimum Sampling Plan Selection

Since it is not practical to measure a large number of wafers using the exhaustive sampling plan (especially in a production environment), a reduced sampling plan must be used for the remaining wafers. The knowledge of systematic variation gained from the exhaustively sampled wafers may be used

Table 7.26: Summary of reduced sampling plan candidates.

Plan Number	Points Per Field	Fields Per Wafer	Total Cost
1	40	23	920 measurements
2	60	15	900 measurements
3	103	9	927 measurements
4	154	5	785 measurements

to choose a sampling plan that is best able to estimate the parameters of the systematic variation models at a given sampling cost.

Candidate Sampling Plans

Although a true optimization was not performed, a small number of sampling plans with similar costs but different measurement distributions were compared to find the plan which is best able to estimate the parameters of the systematic variation models. The plans considered are summarized in Table 7.26, while the details are shown in Figures 7.57–7.60. Each plan requires roughly one hour of measurement time (the exhaustive sampling plan requires nearly 25 hours of measurement time).

7.2.7 Effectiveness of Optimum Sampling Plan

The four candidate sampling plans were used to sample the data collected with the exhaustive sampling plan to evaluate the effectiveness of each plan. The plans were compared on the basis of the widths of the 95% confidence interval of the model parameter estimates. A comparison is shown in Figure

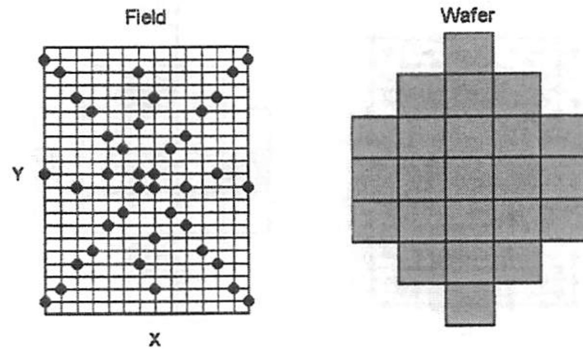

Figure 7.57: ELM candidate sampling plan 1.

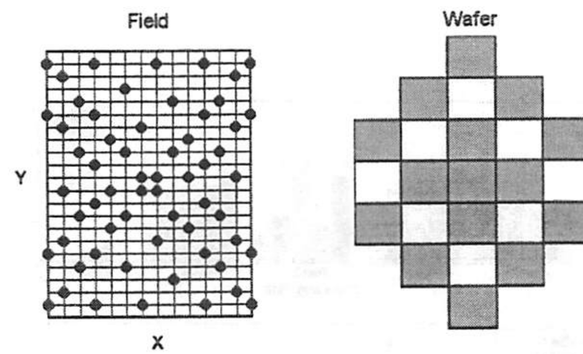

Figure 7.58: ELM candidate sampling plan 2.

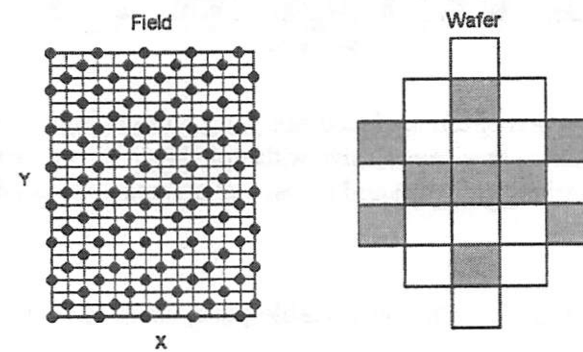

Figure 7.59: ELM candidate sampling plan 3.

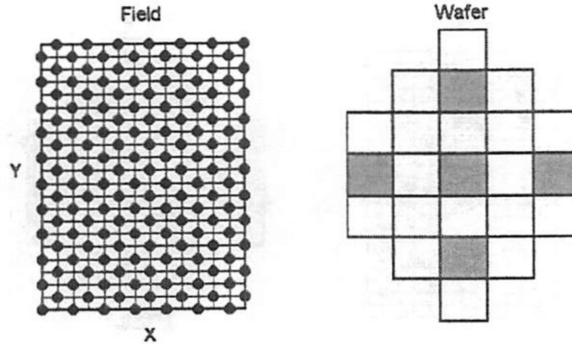

Figure 7.60: ELM candidate sampling plan 4.

7.61.

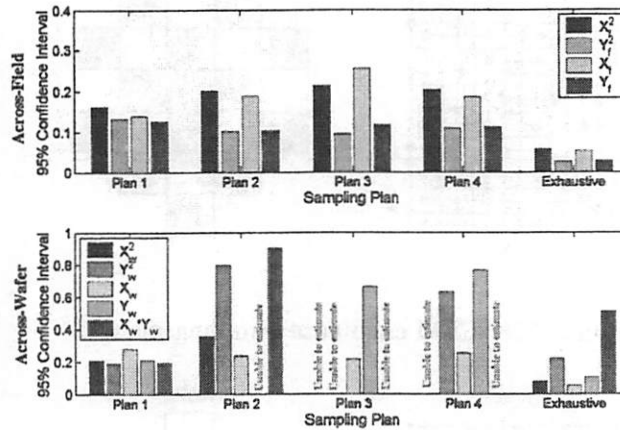

Figure 7.61: Comparison of four reduced sampling plans and exhaustive sampling plan. The vertical axes show the relative width of the 95% confidence interval for parameters of the across-field (top) and across-wafer (bottom) systematic variation models.

Clearly, sampling plan 1 is the only viable plan, since it is the only reduced sampling plan capable of estimating all of the across-wafer systematic variation model parameters and also offers a good estimate of all of the across-

field model parameters as compared with the other reduced sampling plans. Therefore, plan 1 is selected as the "optimum" sampling plan.

7.2.8 Tracking of Systematic Variation Model Parameters

The selected reduced sampling plan (plan 1) was then used to measure the remaining wafers. This plan was used for two features (vertical isolated and vertical dense lines), making the total measurement time required roughly two hours per wafer. In addition, an "intermediate" sampling plan was created to measure wafers overnight to ensure near-100% utilization of the ELM equipment. This plan measured three features (horizontal isolated, vertical isolated, and vertical dense lines) in 50% of the modules in the field in a checkerboard pattern, and measured all of the fields on the wafer (Figure 7.62). The "intermediate" sampling plan required roughly 12 hours to measure a single wafer.

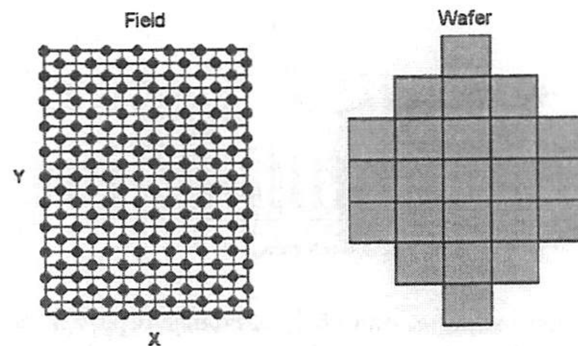

Figure 7.62: ELM "intermediate" sampling plan.

The wafer in slot 22 also exhibited an unexplained deviation from the nominal CD as measured by CD-SEM, so this wafer as well as the wafers in slot 21 and 23 were measured using the exhaustive sampling plan to see if this would provide any insight into this phenomenon. However, the measurements from slot 22 were excessively noisy, and most of the measurements had to be discarded (especially in the center of the field). It is likely that the wafer was misprocessed in the lithography step (this wafer was processed immediately following tool maintenance), resulting in patterning problems.

After the wafers were measured, the across-field and across-wafer systematic variation models were fit to the data for each wafer. This allows the variation in the model parameter estimates to be tracked from wafer to wafer. The results are shown in Figures 7.63 through 7.70.

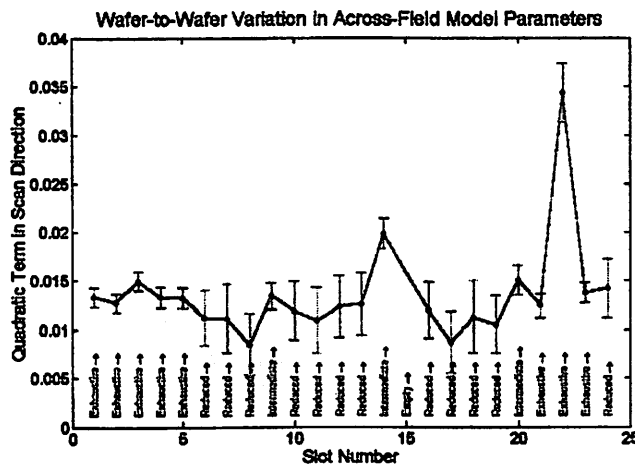

Figure 7.63: Parameter estimates and 95% confidence intervals for quadratic term in the direction of the slit of the across-field systematic variation model.

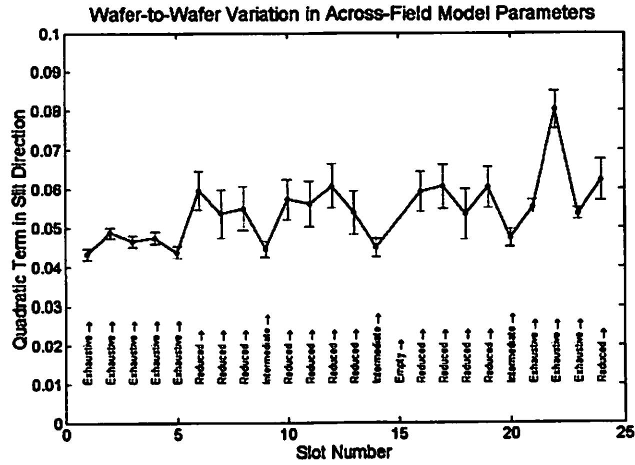

Figure 7.64: Parameter estimates and 95% confidence intervals for quadratic term in the direction of the scan of the across-field systematic variation model.

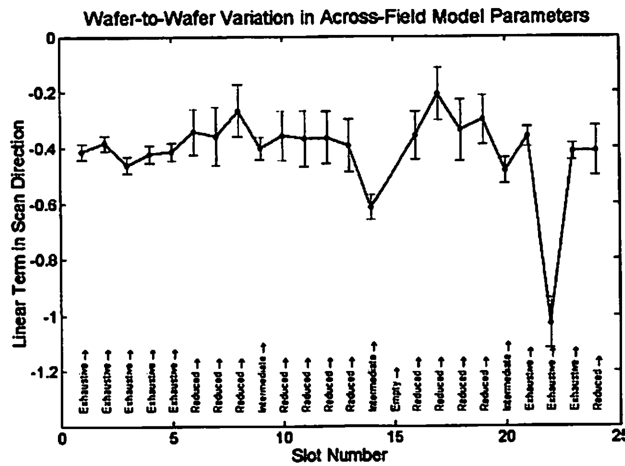

Figure 7.65: Parameter estimates and 95% confidence intervals for linear term in the direction of the slit of the across-field systematic variation model.

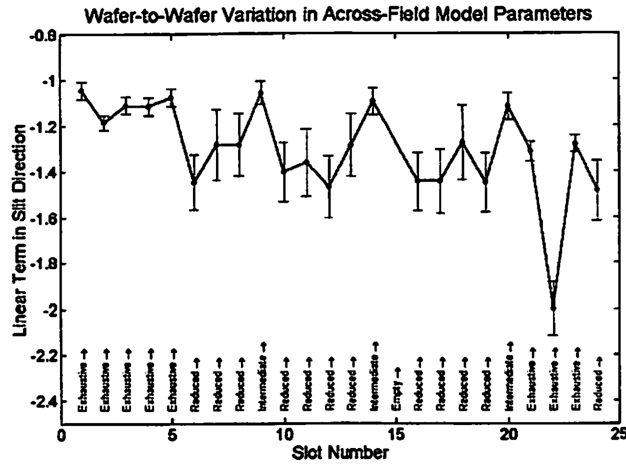

Figure 7.66: Parameter estimates and 95% confidence intervals for linear term in the direction of the scan of the across-field systematic variation model.

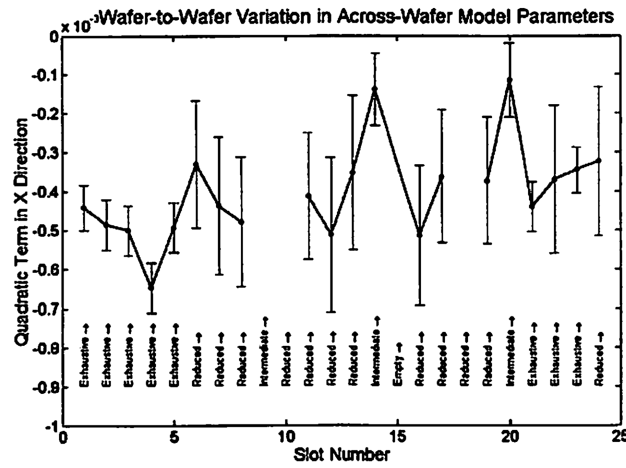

Figure 7.67: Parameter estimates and 95% confidence intervals for quadratic term in the X direction of the across-wafer systematic variation model. Gaps in the curve indicate that the parameter was not able to be estimated for that wafer.

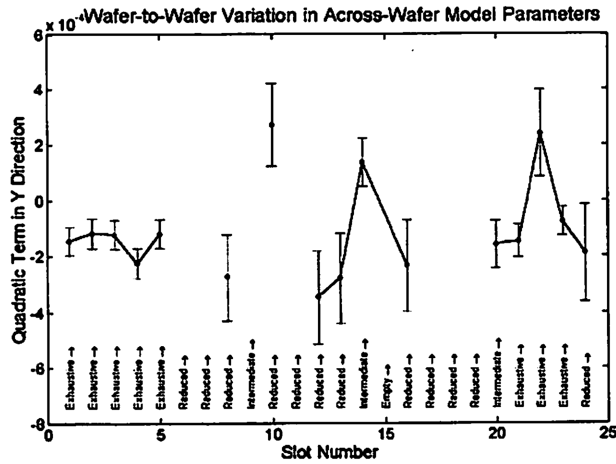

Figure 7.68: Parameter estimates and 95% confidence intervals for quadratic term in the Y direction of the across-wafer systematic variation model. Gaps in the curve indicate that the parameter was not able to be estimated for that wafer.

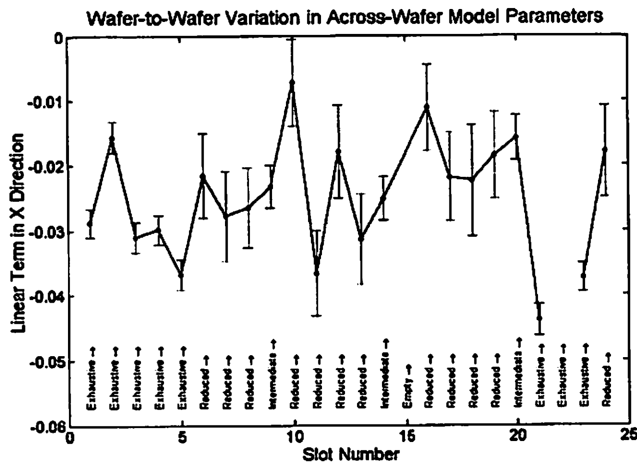

Figure 7.69: Parameter estimates and 95% confidence intervals for linear term in the X direction of the across-wafer systematic variation model. Gaps in the curve indicate that the parameter was not able to be estimated for that wafer.

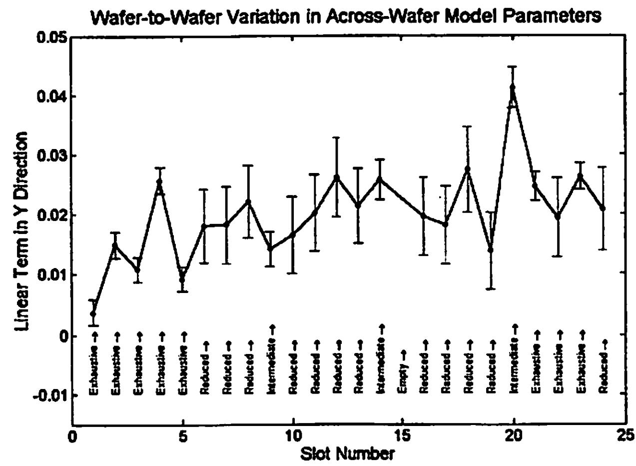

Figure 7.70: Parameter estimates and 95% confidence intervals for linear term in the Y direction of the across-wafer systematic variation model.

Chapter 8

Conclusions

A complete framework for characterization of systematic variation in the lithographic patterning process has been presented. In addition, a method for selecting an optimum sampling plan for monitoring the systematic variation present in the process has been developed. Therefore, a two-step process can be used to track systematic CD variation over time. In the first step, the various components of systematic variation are characterized using a few test wafers sampled extensively. The second step involves the use of an optimum sampling plan to monitor the systematic variance of the process over time. Future work on this topic includes consideration of higher-level systematic variation, including across-lot and lot-to-lot variation. In addition, the use of electrical CD measurements may be used to collect much more dense CD variation maps than are generally possible with CD-SEM.

Chapter 9

Acknowledgements

I must first offer my sincere thanks to my research advisor, Professor Costas Spanos, for his guidance and insight throughout this project. Thanks also go to Professor David Brillinger, who provided many valuable comments and whose statistical expertise helped to improve the clarity of this work. In addition, thanks to Professor Andrew Neureuther, who provided insight into the science of lithography.

Several members of the semiconductor industry helped me complete this work. Joe Bendik of Dynamic Intelligence and Matt Hankinson of KLA-Tencor both offered support that was essential to this project. Also, Tom Ni of Lam Research and Will Conley and Danny Miller, both of International SEMATECH, assisted with wafer processing.

All of the electrical measurements, in addition to some wafer processing, were performed in the UC Berkeley Microfabrication Laboratory. Tim Duncan

modified the probing software to work with the ELM2 structures. TK Chen and Todd Merport also provided computer support. Thanks to all of the Microlab staff for their invaluable assistance. In particular, thanks to Katalin Voros, Sia Parsa, Laszlo Voros, and Bob Hamilton.

I also wish to thank all the past and present members of BCAM, whose friendship and encouragement I value greatly: Junwei Bao, Runzi (Tiger) Chang, Weng Loong (Ralph) Foong, Mason Freed, Paul Friedberg, Anna Ison, Nickhil Jakatdar, Payman Jula, Michiel Kruger, Jae-Wook Lee, Jing Xue, Jiangxin Wang, Charlie Zhang, Haolin Zhang, and Dong Wu Zhao.

Last but certainly not least, I offer my deepest thanks to my wife Melissa, without whose patience, love, and support this work would not have been possible.

Bibliography

[1] G. E. Moore, "Lithography and the future of Moore's law," in *Optical/Laser Microlithography VIII*, T. A. Brunner, ed., *Proc. SPIE* **2440**, pp. 2–17, 1995.

[2] S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era*, vol. 1, Lattice Press, Sunset Beach, CA, 1986.

[3] R. Dammel, *Diazonaphthoquinone-based Resists*, SPIE Optical Engineering Press, Bellingham, WA, 1993.

[4] R. D. Allen, W. E. Conley, and R. R. Kunz, "Deep-UV resist technology," in *Handbook of Microlithography, Micromachining, and Microfabrication*, P. Rai-Choudhury, ed., pp. 321–375, SPIE Optical Engineering Press, Bellingham, WA, 1997.

[5] T. Ueno, "Chemistry of photoresist materials," in Sheats and Smith [15], pp. 429–514.

[6] D. R. McKean, U. Schaedeli, and S. A. MacDonald, "Acid photogeneration from sulfonium salts in solid polymer matrices," *J. Polymer Sci. A: Polymer Chem.* **27**, pp. 3927–3935, 1989.

[7] B. W. Smith, "Resist processing," in Sheats and Smith [15], pp. 515–565.

[8] H. J. Levinson, *Principles of Lithography*, SPIE Press, Bellingham, WA, 2001.

[9] B. W. Smith and M. Hanratty, "Multilayer resist technology," in Sheats and Smith [15], pp. 567–614.

[10] J. H. Bruning, "Optical lithography–thirty years and three orders of magnitude," in *Advances in Resist Technology and Processing XIV*, R. G. Tarascon-Auriol, ed., *Proc. SPIE* **3049**, pp. 14–27, 1997.

[11] D. Steele, A. Coniglio, C. Tang, B. Sing, S. Nip, and C. Spanos, "Characterizing post exposure bake processing for transient and steady state conditions, in the context of critical dimension control," in Herr [74], pp. 517–530.

[12] R. R. Kunz, "Dry etching of photoresists," in Sheats and Smith [15], pp. 615–644.

[13] M. Born and E. Wolf, *Principles of Optics*, Cambridge University Press, Cambridge, UK, seventh ed., 1999.

[14] A. K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, WA, 1990.

[15] J. R. Sheats and B. W. Smith, eds., *Microlithography: Science and Technology*, Marcel Dekker, Inc., New York, 1998.

[16] K. Jain, *Excimer Laser Lithography*, SPIE Optical Engineering Press, Bellingham, WA, 1990.

[17] L. D. Laude, ed., *Excimer Lasers*, Kluwer Academic Publishers, Boston, MA, 1994.

[18] P. Das and U. Sengupta, "Krypton fluoride excimer laser for advanced microlithography," in Sheats and Smith [15], pp. 271–315.

[19] G. A. Luurtsema, "Spin coating for rectangular substrates," Master's thesis, University of California, Berkeley, 1997.

[20] A. K. Wong, R. A. Ferguson, and S. M. Mansfield, "The mask error factor in optical lithography," *IEEE Trans. Semiconduct. Manufact.* **13**, pp. 235–242, May 2000.

[21] A. K. Wong, R. A. Ferguson, L. W. Liebmann, S. M. Mansfield, A. F. Molless, and M. O.Neisser, "Lithographic effects of mask critical dimension error," in den Hove [75], pp. 106–116.

[22] F. M. Schellenberg, V. Boksha, N. Cobb, J. C. Lai, C. H. Chen, and C. Mack, "Impact of mask errors on full chip error budgets," in den Hove [76], pp. 261–275.

[23] M. Plat, K. Nguyen, C. Spence, C. Lyons, and A. Wilkison, "The impact of optical enhancement techniques on the mask error enhancement function (meef)," in Progler [77], pp. 206–214.

[24] J. Randall, A. Tritchkov, R. Jonckheere, P. Jaenen, and K. Ronse, "Reduction of mask induced Cd errors by optical proximity correction," in den Hove [75], pp. 124–130.

[25] K. Adam, R. Socha, M. Dusa, and A. Neureuther, "Effect of reticle CD uniformity on wafer CD uniformity in the presence of scattering bar optical proximity correction," in *BACUS Symposium on Photomask Technology and Management*, B. J. Grenon and F. E. Abboud, eds., *Proc. SPIE* **3546**, pp. 642–650, 1998.

[26] J. van Schoot, J. Finders, K. van Ingen Schenau, M. Klaassen, and C. Buijk, "The mask error factor: Causes and implications for process latitude," in den Hove [76], pp. 250–260.

[27] W. Conley, C. Garza, M. Dusa, R. Socha, J. Bendik, and C. Mack, "The MEEF shall inherit the earth," in Progler [78], pp. 251–258.

[28] H. Zhang, J. P. Cain, and C. J. Spanos, "Compact formulation of mask error factor for critical dimension control in optical lithography," in Herr [74], pp. 462–465.

[29] T. A. Brunner, "Impact of lens aberrations on optical lithography," *IBM J. Res. Develop.* **41**, pp. 57–67, Jan./Mar. 1997.

[30] V. N. Mahajan, *Aberration Theory Made Simple*, SPIE Optical Engineering Press, Bellingham, WA, 1991.

[31] F. Zernike, "Diffraction theory of the knife-edge test and its improved form, the phase-contrast method," *J. Microlithography, Microfabrication, and Microsystems* **1**, pp. 87–94, July 2002.

[32] V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils," *Appl. Optics* **33**, pp. 8121–8124, Dec. 1994.

[33] J. W. Goodman, *Introduction to Fourier Optics*, McGraw Hill, Boston, MA, second ed., 1996.

[34] B. W. Smith and R. Schlief, "Understanding lens aberration and influences to lithographic imaging," in Progler [77], pp. 294–306.

[35] B. W. Smith, "Optics for photolithography," in Sheats and Smith [15], pp. 171–270.

[36] A. Bourov, L. C. Litt, and L. Zavyalova, "Impact of flare on CD variation for 248nm and 193nm lithography systems," in Progler [78], pp. 1388–1393.

[37] B. La Fontaine, M. Dusa, A. Acheta, C. Chen, A. Bourov, H. Levinson, L. C. Litt, M. Mulder, R. Seltman, and J. Van Praagh, "Analysis of flare and its impact on low-k1 KrF and ArF lithography," in Yen [79].

[38] D. Nam, E. Lee, Y. Kang, G. Yeo, J. Lee, H. Cho, W. Ham, and J. Moon, "Effectiveness and confirmation of local area flare measurement method in various pattern layouts," in Yen [79].

[39] J. Schefske, E. Kent, U. Okoroanyanwu, H. Levinson, C. Masud, B. Streefkerk, R. Hanzen, and J. Bruebach, "Issues and non-issues on a 193 nm step-and-scan system in production," in Progler [77], pp. 460–471.

[40] G. H. Ho, A. T. Cheng, C.-J. Chen, C.-K. Fang, M.-C. Li, I.-C. Chang, P.-T. Chu, Y. Chu, K.-Y. Shu, C.-Y. Huang, H.-L. Yeh, H. Shiao, and H. Lan, "Lens heating induced focus drift of i-line step & scan: Correction and control in a manufacturing environment," in Sullivan [80], pp. 289–296.

[41] B.-J. Cheng, H. Liu, Y. Cui, and J. Guo, "Improve the image control by correcting the lens heating focus drift," in Progler [77], pp. 818–826.

[42] M. Klaassen, M. Reuhman, A. Loock, M. Rademaker, and J. Gemen, "Dymanic performance of DUV step & scan systems and process latitude," in Progler [77], pp. 776–784.

[43] D. G. Flagello, J. Mulkens, and C. Wagner, "Optical lithography into the millennium: Sensitivity to aberrations, vibration and polarization," in Progler [77], pp. 172–183.

[44] B. Sluijk, T. Castenmiller, and R. du Croo de Jongh and Hans Jasper and Theo Modderman and Leon Levasier and Erik Loopstra and Guustaaf Savenije and Marc Boonman and Harry Cox, "Performance results of a new generation of 300 mm lithography systems," in Progler [78], pp. 544–557.

[45] R. Spangler, J. Vipperman, B. Pletner, and D. Brown, "Enabling advanced lithography through active vibration control," *MICRO* **19**, pp. 45–55, October 2001.

[46] R. Spangler, "Vibration control gets active," *Photonics Spectra* **35**, pp. 98–104, December 2001.

[47] K. K. H. Toh and A. R. Neureuther, "Identifying and monitoring effects of lens aberrations in projection printing," in *Optical Microlithography VI*, H. Stover, ed., *Proc. SPIE* **772**, pp. 202–209, 1987.

[48] The Mathworks, Inc. `http://www.mathworks.com`.

[49] J. W. Goodman, *Statistical Optics*, John Wiley & Sons, New York, 2000.

[50] T. Yao, T. Hiraike, K. Kobayashi, S. Asai, and I. Hanyu, "Improvement of CD uniformity in 180 nm LSI manufacturing by optimizing illumination system," in *IEEE International Symposium on Semiconductor Manufacturing*, B. Sohn, ed., pp. 125–128, 2001.

[51] Y. J. T. Lii, "Etching," in *ULSI Technology*, C. Y. Chang and S. M. Sze, eds., pp. 329–370, McGraw Hill, New York, 1996.

[52] J. A. Mucha, D. W. Hess, and E. S. Aydil, "Plasma etching," in *Introduction to Microlithography*, L. F. Thompson, C. G. Wilson, and M. J. Bowden, eds., pp. 377–507, American Chemical Society, Washington, DC, second ed., 1994.

[53] M. T. Postek and A. E. Vladár, "Critical-dimension metrology and the scanning electron microscope," in Diebold [81], pp. 295–333.

[54] A. E. Braun, "3-D capability turns CD-SEM into metrology system," *Semiconductor International* **23**, January 2000.

[55] G. Binning, H. Rohrer, C. Gerber, and E. Weibel, "Surface studies by scanning tunneling microscopy," *Phys. Rev. Let.* **49**, pp. 57–61, July 1982.

[56] H. M. Marchman and J. E. Griffin, "Scanned probe microscope dimensional metrology," in Diebold [81], pp. 335–375.

[57] R. Eisberg and R. Resnick, *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*, John Wiley & Sons, New York, 1985.

[58] X. Niu, *An Integrated System of Optical Metrology for Deep Sub-Micron Lithography.* PhD thesis, University of California, Berkeley, 1999.

[59] X. Niu, N. Jakatdar, J. Bao, and C. J. Spanos, "Specular spectroscopic scatterometry," *IEEE Trans. Semiconduct. Manufact.* **14**, pp. 97–111, May 2001.

[60] W. L. Foong, "Characterizing the sensitivity of scatterometry for sub-100nm technologies," Master's thesis, University of California, Berkeley, 2001.

[61] M. W. Cresswell and R. A. Allen, "Electrical CD metrology and related reference materials," in Diebold [81], pp. 377–410.

[62] L. J. van der Pauw, "A method of measuring specific resistivity and Hall effect of discs of arbitrary shape," *Phillips Research Reports* **13**, pp. 1–9, 1958.

[63] SAS Institute Inc. `http://www.jmp.com`.

[64] A. Grenville, B. Coombs, J. Hutchison, K. Kuhn, D. Miller, and P. Troccolo, "Electrical critical dimension metrology for 100-nm linewidths and below," in Progler [77], pp. 452–459.

[65] J. Kye and H. Levinson, "Electrical linewidth measurement for next generation lithography," in Sullivan [80], pp. 637–643.

[66] D. Drain, *Statistical Methods for Industrial Process Control*, Chapman & Hall, New York, 1997.

[67] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, John Wiley & Sons, New York, 1978.

[68] C. R. Rao and H. Toutenburg, *Linear Models: Least Squares and Alternatives*, Springer-Verlag, New York, 1995.

[69] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation, and Application*, Robert E. Krieger Publishing Company, Malabar, FL, 1989.

[70] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, MA, 1999.

[71] A. Osyczka, *Multicriterion Optimization in Engineering with FORTRAN Programs*, John Wiley & Sons, New York, 1984.

[72] C. Yu, T. Maung, C. Spanos, D. Boning, J. Chung, H.-Y. Liu, K.-J. Chang, and D. Bartelink, "Use of short-loop electrical measurements for yield improvement," *IEEE Trans. Semiconduct. Manufact.* **8**, pp. 150–159, May 1995.

[73] A. K. Wong, A. F. Molless, T. A. Brunner, E. Coker, R. H. Fair, G. L. Mack, and S. M. Mansfield, "Linewidth variation characterization by spatial decomposition," *J. Microlithography, Microfabrication, and Microsystems* **1**, pp. 106–116, July 2002.

[74] D. J. Herr, ed., *Metrology, Inspection, and Process Control for Microlithography XVI*, *Proc. SPIE* **4689**, 2002.

[75] L. V. den Hove, ed., *Optical Microlithography XI*, *Proc. SPIE* **3334**, 1998.

[76] L. V. den Hove, ed., *Optical Microlithography XII*, *Proc. SPIE* **3679**, 1999.

[77] C. J. Progler, ed., *Optical Microlithography XIII*, *Proc. SPIE* **4000**, 2000.

[78] C. J. Progler, ed., *Optical Microlithography XIV*, *Proc. SPIE* **4346**, 2001.

[79] A. Yen, ed., *Optical Microlithography XV*, *Proc. SPIE* **4691**, 2002.

[80] N. T. Sullivan, ed., *Metrology, Inspection, and Process Control for Microlithography XV*, *Proc. SPIE* **4344**, 2001.

[81] A. C. Diebold, ed., *Handbook of Silicon Semiconductor Metrology*, Marcel Dekker, Inc., New York, 2001.