DEEP SUB-MICRON PHOTOLITHOGRAPHY CONTROL THROUGH IN-LINE METROLOGY

by

Nickhil Harsh Jakatdar

Memorandum No. UCB/ERL M00/11

15 January 2000

DEEP SUB-MICRON PHOTOLITHOGRAPHY CONTROL THROUGH IN-LINE METROLOGY

by

Nickhil Harsh Jakatdar

Memorandum No. UCB/ERL M00/11

15 January 2000

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Deep Sub-Micron Photolithography Control through In-Line Metrology

by
Nickhil Harsh Jakatdar

B.S. (College of Engineering, Pune, India) 1995
M.S. (University of California, Berkeley, CA) 1998

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Engineering - Electrical Engineering and Computer Sciences
in the
GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor Costas J. Spanos, Chair
Professor Andrew R. Neureuther
Professor Kameshwar Poolla

Spring, 2000

The dissertation of Nickhil Jakatdar is approved:

Chair Date

Date

Date

University of California, Berkeley

Spring, 2000

Abstract

Deep Sub-Micron Photolithography Control through In-Line Metrology

by

Nickhil Harsh Jakatdar

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Costas J. Spanos, Chair

The exponential increase of integrated circuit density and semiconductor manufacturing cost is well described by Moore's Law. In order to provide affordable lithography at and below 100nm, *in-situ* and in-line metrology is becoming increasingly critical for advanced process control and rapid yield learning. The successful implementation of a real time or run-to-run controller requires metrology for the intermediate and final quantities of interest as well as robust process models.

In this thesis, a metrology framework is developed for each of the process steps by identifying the observable that is related to the final quantity of interest, identifying the associated sensors and developing efficient algorithms to analyze the sensor data. This framework is demonstrated for optical constant measurements of thin films before and after the spin-coat and soft-bake steps, deprotection and associated thickness loss measurements after the exposure and post-exposure bake steps, and the cross-section resist and polysilicon profile measurements after the lithography and etch steps. Simple optical sensors, such as spectroscopic reflectometry and spectroscopic ellipsometry, are used for the different metrology steps in order to facilitate integrated metrology with the lithography process equipment i.e. wafer track and stepper.

Modeling the chemistry and physics of the deep ultraviolet lithography is critical in the deep sub-micron pattern transfer process for effective CD prediction. In this thesis, novel process models are developed for each step of the lithography process to provide an effective simulation environment. To make the simulation predictive, a hierarchical architecture is developed that calibrates the simulator model coefficients based on experimental cross-section profile data and a global optimization routine.

The process models together with the metrology scheme provide the building blocks of a process controller. The process models developed in this thesis provide insight into the observables available at every step of the lithography sequence and their correlation to the final critical dimension. The metrology schemes indicate the sensors and algorithms required to efficiently measure these observables. These building blocks are used to develop a Kalman Filter based process controller that integrates or fuses information from multiple sensors, such as the deprotection induced thickness loss system and the specular spectroscopic scatterometry system. This scheme promises a better than 50% reduction in the deep ultraviolet lithography process variability.

Professor C. J. Spanos

Committee Chairman

Table of Contents

Chapter 1. Introduction 1

1.1. Motivation 1
1.1.1. Metrology 4
1.1.2. Modeling 4
1.1.3. Control 5
1.2. Thesis Organization 5

Chapter 2. Modeling, Optimization and Control Algorithms 7

2.1. Introduction 7
2.2. Optimization through Simulated Annealing 7
2.2.1. Introduction to Simulated Annealing 7
2.2.2. Adaptive Simulated Annealing 9
2.3. Learning through Articial Neural Networks 10
2.3.1. Multi-Layer Perceptron 11
2.3.2. Radial Basis Function Network (RBFN) 13
2.3.3. Training the RBFN 14
2.4. Control Utilizing Kalman Filter 16

Chapter 3. In-line/In-situ Metrology for the Pattern Transfer Process 18

3.1. Introduction 18
3.2. Thin Film Metrology 19
3.2.1. Optical Constants 19
3.2.2. Broadband Reflectometry 20
3.2.3. Neural Network - Adaptive Simulated Annealing (NN-ASA) 22
3.2.3.1. Parameter Extraction using ASA 24
3.2.3.2. Monte Carlo Simulation using the F-B formulation and Reflectance equations 24
3.2.3.3. Spectral Feature Selection 25
3.2.3.4. Neural Network Training and Validation 26
3.2.4. Results of the NN-ASA Algorithm 27
3.3. Exposure and PEB Metrology 28
3.3.1. Deprotection Induced Thickness Loss (DITL) 28
3.3.2. Broadband Reflectometry for Thickness Loss 30
3.3.3. Correlating Thickness Loss to Deprotection 31
3.3.4. Results of DITL Metrology 32
3.4. Post-Develop and Post-Etch Profile Metrology 36
3.4.1. Patterned Profiles 36
3.4.2. Spectroscopic Ellipsometry 36
3.4.3. A Library-Based Methodology for CD Profile Extraction 39
3.4.4. Results of Specular Spectroscopic Scatterometry 41
3.4.4.1. Photoresist on ARC on Silicon 41

3.4.4.2. Metal Stack 43

Chapter 4. **Lithography Modeling** 46

4.1. Introduction 46
4.2. Volume Shrinkage in Chemically Amplified Resist Systems 47
4.3. Static Model for Thickness Loss 47
4.4. Dynamic Model for Thickness Loss 50
4.4.1. Physical Models 50
4.4.2. Boundary Conditions 52
4.4.3. Computational Approach 53
4.5. Optimization Framework for DITL Parameter Extraction 54
4.6. Results : Parameter Extraction in DITL Models 55
4.7. Summary 58

Chapter 5. **Lithography Simulation** 59

5.1. Introduction 59
5.2. Simulator Calibration Framework 60
5.2.1. Process Models 60
5.2.1.1. Thin Film Interference 60
5.2.1.2. Exposure & PEB Process Model 61
5.2.1.3. Resist Develop Model 61
5.2.2. Estimating the values of the model parameters 61
5.2.3. Experimental Methodology for Parameter Extraction through Profile Matching 63
5.2.4. Formulating an optimization problem for estimating the unknown variables 65
5.2.5. Framework for Parameter Extraction through Profile Matching 66
5.3. Experiments for Parameter Extraction through Profile Matching 69
5.3.1. Unpatterned Experiments 69
5.3.2. Patterned Experiments 70
5.4. Results of Parameter Extraction through Profile Matching 71
5.4.1. Category I parameters 71
5.4.2. Category II parameters 72
5.4.3. Category III parameters 73
5.5. Summary 76

Chapter 6. **Run-to-Run Control in the DUV Lithography Sequence** 77

6.1. Introduction 77
6.2. Sources of Variability in the Lithography Sequence 78
6.2.1. Incoming Thin Film 78
6.2.2. Spin Coat and Soft Bake 78
6.2.3. Exposure 78
6.2.4. Post-Exposure Bake (PEB) 79
6.2.5. Development 79

6.3. Metology Schemes ... 79
6.4. In-line RtR Controller Design ... 80
6.4.1. Experiments ... 80
6.4.2. Process Models ... 82
6.4.3. Drift Model ... 83
6.4.4. RtR Control Architecture : Scenario I ... 84
6.4.4.1. Simulation Results ... 86
6.4.5. RtR Control Architecture : Scenario II ... 90
6.4.5.1. Simulation Results ... 91

Chapter 7. Conclusions ... 93

7.1. Thesis Summary ... 93
7.2. Future Work ... 94

References ... 96

Appendix A ... 100

List of Figures

Figure 1.1. The reduction in the lithography exposure wavelength and minimum feature size over time. 12

Figure 1.2. The trade-off between speed and yield as a function of sLeff. 13

Figure 1.3. A Metrology, Modeling and Control Framework for the DUV Lithography Process. 13

Figure 2.1. Pseudo-code for the Simulated Annealing Algorithm 18

Figure 2.2. Architecture of a Multi Layer Perceptron 22

Figure 2.3. Architecture of a Radial Basis Function Network 23

Figure 3.1. The objective is to match the simulated and measured broadband spectra by tuning the parameters of the dispersion relationship using the optimizer. 31

Figure 3.2. Results of the ASA Optimization Algorithm. Figure shows the simulatedversustheexperimentalreflectancespectra(relativetobareSilicon). 32

Figure 3.3. Block Diagram of the NN-ASA Algorithm 33

Figure 3.4. Performance of the NN-ASA algorithm for poly-Si on native oxide on Si stack. 38

Figure 3.5. Resist Mechanism during the Exposure and Post Exposure Bake Steps for a commercial DUV photoresist. 39

Figure 3.6. Layout of the blanket exposure areas on the wafer 41

Figure 3.7. Thickness loss as a function of the deprotection measured by monitoring the normalized ester absorbance. 42

Figure 3.8. Thickness loss as a function of the deprotection measured by monitoring the normalized hydroxy absorbance 44

Figure 3.9. The normalized ester and hydroxy peaks as a function of exposure dose for different PEB temperatures. 45

Figure 3.10. Spectroscopic ellipsometry measurements. 47

Figure 3.11. A library based methodology for CD profile extraction. 49

Figure 3.12. (a) grating structures; (b) Focus-exposure matrix experiment setup for experimental verification. 52

Figure 3.13. Comparison between the extracted grating profiles and the CD-AFM profile across the focus-exposure matrix. The four AFM profiles with (dose,focus)levelof(2,-2),(3,-1),(-1,2)and(2,2)havenotbeenmeasured. 53

Figure 3.14. (a) Grating structures; (b) Focus-exposure matrix experiment setup. 54

Figure 3.15. Comparison of top CD between specular spectroscopic scatterometry and a 55

top-down CD-SEM. 55

Figure 4.1. Results of the ASA Optimization Algorithm. Figure shows the simulated versus the experimental reflectance spectra 59

Figure 4.2. Block Diagram of the NN-ASA Algorithm 60
Figure 4.3. Performance of the NN-ASA algorithm for poly-Si on native oxide on Si stack. 64
Figure 4.4. DUV Chemically Amplified Resist Mechanism during the Exposure and Post Exposure Bake Steps. 65
Figure 4.5. Thickness loss as a function of the deprotection, measured by monitoring the normalized ester absorbance. 65
Figure 4.6. Simulation Approach for the Moving Boundary Problem 72
Figure 4.7. Block diagram for resist parameter extraction using the dynamic model and experimental data. 73
Figure 4.8. Static model fitting experimental data for UV-5. Circles denote simulated results while stars denote experimental data. 75
Figure 4.9. Static Model fitting experimental data for AZ-2549. Bold line denotes simulation while stars denote experimental data. 75
Figure 4.10. Intermediate results for acid concentration, deprotection, volatile group concentration and free volume concentration as a function of depth at 5 different time steps 76
Figure 4.11. Measurements of thickness loss versus time for UV-5. Simulated loss in bold; measured loss in star 76
Figure 4.12. Measurements of thickness loss versus time for AZ 2549. 77
Figure 5.1. Experimental flow for parameter extraction. 84
Figure 5.2. Finding initial estimates for category II parameters (parameters 85
Figure 5.3. Computing the sum-squared error between simulated and digitized SEM images 86
Figure 5.4. Flowchart of unpatterned characterization experiments (Update Category 87
II and III). 87
Figure 5.5. The mask layout design for the experiment. 89
Figure 5.6. An example of a cross-sectional SEM image. 90
Figure 5.7. Experimental and Fitted Values for Deprotection vs Exposure Dose as a function of 3 different PEB temperatures (120, 130, 135 degrees C) 91
Figure 5.8. Develop rate versus the normalized concentration of unreacted sites. Figure shows the fitting of the Mack develop model to the data..... 92
Figure 5.9. Fitted versus simulated profiles across different focus-layout combinations for training data. 94
Figure 5.10. Fitted versus simulated profiles across different focus-layout combinations for testing data. 95
Figure 6.1. BlanketandPatternedExposurePatternfortheCDcontrolmodelexperiment 101
Figure 6.1. RtR Control Architecture for DUV Lithography in Scenario I 104
Figure 6.1. Simulation Architecture 105
Figure 6.2. RtR Control Architecture for DUV Lithography in Scenario II.... 107

List of Tables

Geometric Definition of the Layout in the Experiment (in mm) 89
Category I: Well Known Parameter Values 90
Category II: Parameters subject to experimental and modeling errors 93
Category III: Parameters that must be inferred indirectly 93

Chapter 1 Introduction

1.1. Motivation

The semiconductor industry is unique in having sustained such rapid technology development over so long a period, growing at an annual rate of approximately 15% over the past three and a half decades. It has been said that if other industries, such as the airline industry, had progressed at the same rate over the last 30 years, it would have been possible to fly from San Francisco to New York in less than a minute for less than a dollar. However, it now appears that the industry is rapidly approaching a formidable "100 nm barrier", consisting of an unprecedented number of distinct technical challenges which threaten continuation of its historical success formula. Two of the "Grand Challenges" identified in the National Technology for Roadmap (1999 Edition) are affordable lithography at or below 100 nm and solutions for Metrology & Test [1].

Ever since the invention of the integrated circuit, patterning has been achieved by lithographic techniques that use visible light. With the relentless decrease in feature size required for the productivity increases necessary to follow Moore's Law, lithography sources have progressed to ever shorter wavelengths, leading to today's exposure tools based on deep ultraviolet light. However, this reduction in exposure wavelength has not been pushed as aggressively as the reduction in the feature size, as plotted in Figure 1.1. While innovative technological approaches have enabled the industry to manufacture sub-wavelength feature sizes, it will get increasingly difficult as the ratio of the feature size to exposure wavelength falls below 0.5. To make matters worse, industry analysts estimate that a state-of-the-art fab for the 0.15 μm technology will cost more than $2.5 Billion and will touch the $10 Billion mark for sub 0.10 μm technology generation.

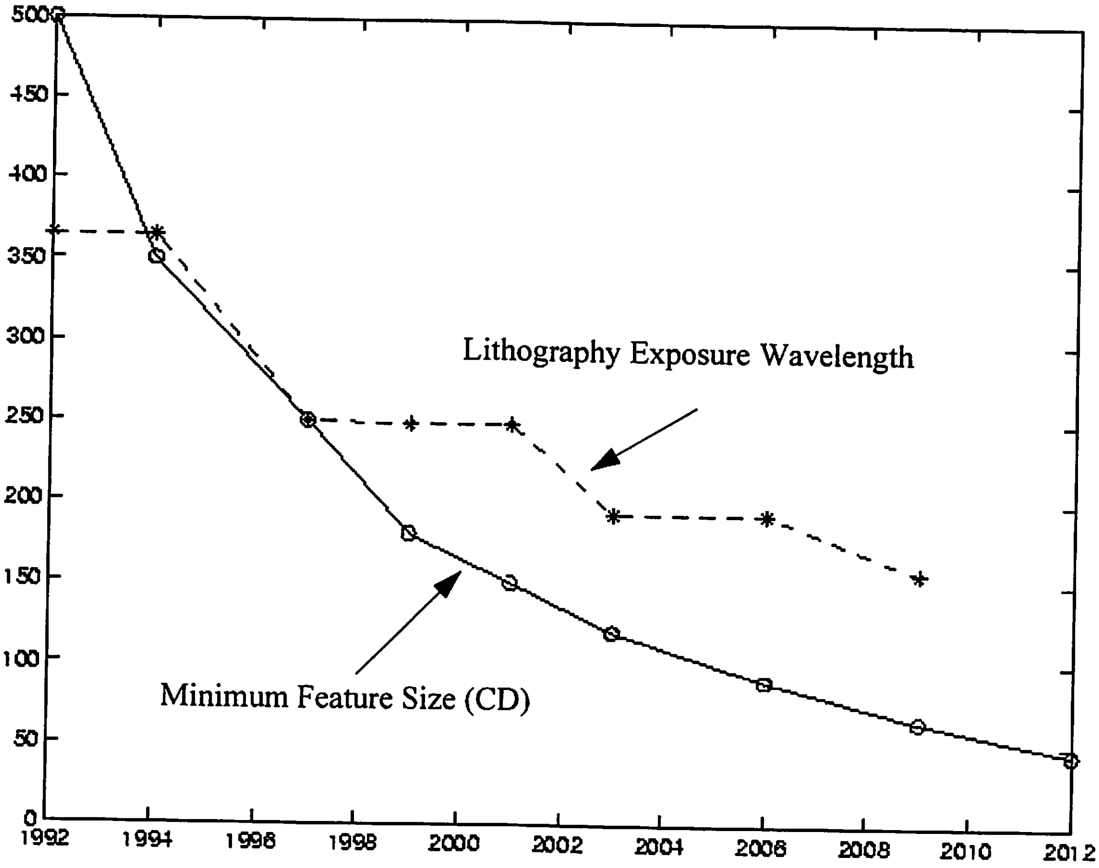


Figure 1.1. The reduction in the lithography exposure wavelength and minimum feature size over time.

To keep the industry on its historic 25-30% / year reduction in cost/function despite the escalating factory costs (20% / year), greater synergy must be developed between the areas of metrology, modeling and control. Sturtevant et.al. have demonstrated the efficacy of such a synergy in the pattern transfer sequence, where they used a simple run-to-run controller working in unison with simple process models, and a CD-SEM to reduce variability by 54% [2]. The tighter lot critical dimension (CD) distribution allowed the target to shift to shorter L_{eff} without a yield hit, allowing for a $2 Million increase in revenues per 1000 wafer starts, as shown in Figure 1.2. AMD and Honeywell used a similar framework to demonstrate a 70% reduction in process scrap and rework in the lithography sequence [3].

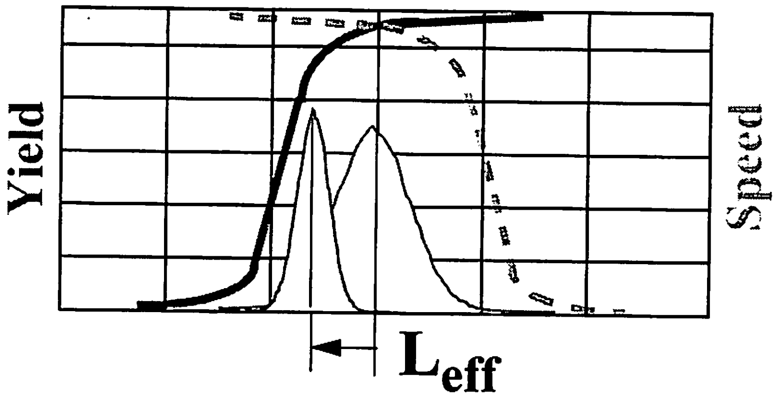


Figure 1.2. The trade-off between speed and yield as a function of σ_{Leff}.

The key components to keeping on the manufacturing cost learning curve are shown in Figure 1.3.

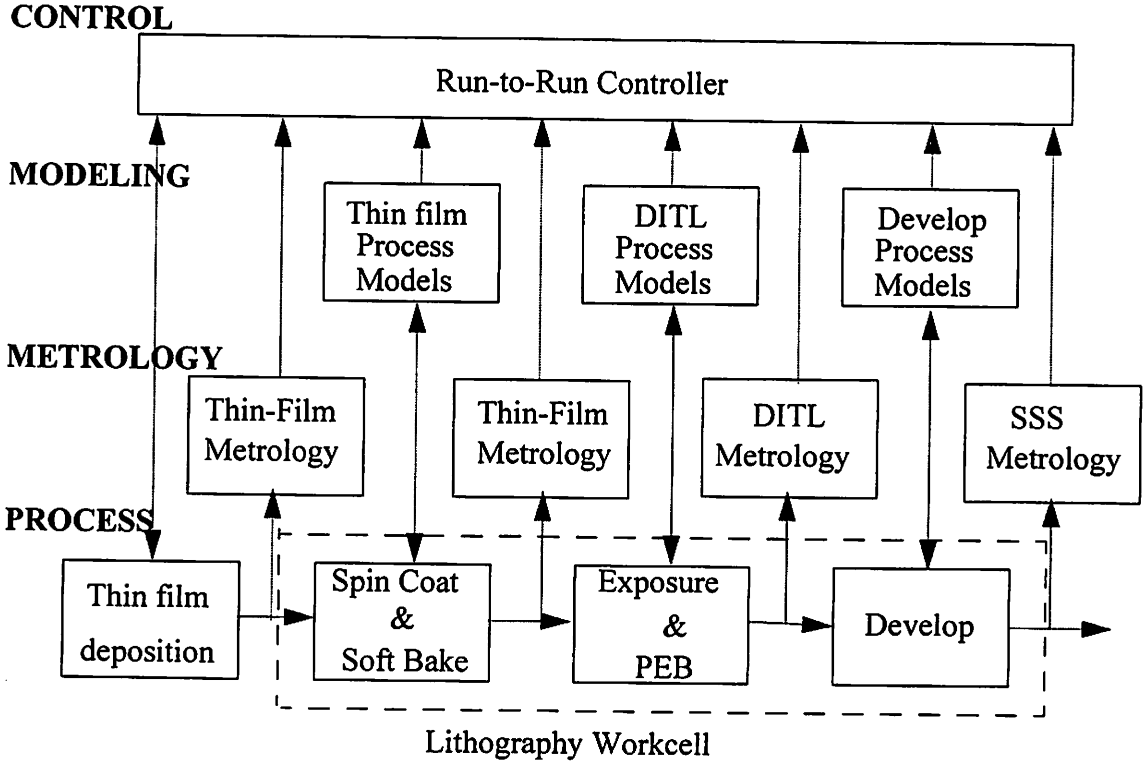


Figure 1.3. A Metrology, Modeling and Control Framework for the DUV Lithography Process.

1.1.1. Metrology

Metrology is essential to the development and improvement of new processes and tools for future technology generations. Metrology can potentially reduce the cost of manufacturing and time-to-market for new products through better characterization of process tools and processes. As device dimensions shrink, the challenge for physical metrology will be to keep pace with inline electrical testing that provides critical electrical performance data. Manufacturing sub-100 nm devices will require the availability of robust in-situ equipment, process, and wafer-state sensors in process tools. These sensors must have adequate repeatability, reproducibility and calibration capability to provide the necessary real time information for fault detection and process control. This would need the development of increasingly faster data acquisition and computational algorithms for converting sensor data into useful information. It is acknowledged that the implementation of in-situ metrology will be driven by reducing pilot wafer use, while simultaneously increasing process capability.

1.1.2. Modeling

Modeling, according to the NTRS, is like a stool that requires three legs for a stable result. The three legs are:

1. Models - a mental image of reality, formalized in a mathematical model.

2. Simulators - the computer codes that implement the models

3. Calibration and Validation - the comparison of simulated results to relevant experimental data to determine numerical values for parameters, and to demonstrate "suitability for purpose".

In early stages of development, modeling and simulation technology is often used to provide insight into technology directions and interactions between options. During later stages, it is used for quantitative analyses like optimization, sensitivity analysis and process diagnosis. Models ranging from easy-to-use to complex, from fast-executing to computationally intensive, and from high accuracy in a constrained process space to moderately predictive over wide ranges should be available to satisfy all uses. It is expected that optical lithography at 100 nm and below will require modeling approaches that integrate the analysis of the stepper/scanner optics with that of the resist materials. New phenomena, such as polymer-surface interaction, chemical amplification and its effect on deprotection and diffusion, and the interaction of local resolution enhancements with image non-idealities, will become important. Matching the simulation results to the real measurements becomes a challenge, because of the non-linearity of the optimization problem and the presence of multiple local minima.

1.1.3. Control

Process engineers have expended considerable effort in the design of "get-it-right-the-first-time" processes. However, the aggressive CD control specifications projected for future device generations are fast outpacing the inherent process or equipment variability. Given the continually reducing process windows available with the sub-150 nm technology node causing misprocessing to become prohibitively expensive, semiconductor manufacturing is slowly moving from fixed process recipe, open-loop control to closed-loop control via sensor-driven model-based integrated manufacturing (SDMBIM) [1]. In-situ process control is a critical solution for the future factory as conventional metrology becomes less reliable, more expensive, and can only identify process excursions after significant yield loss. Control algorithms are required that can efficiently fuse information from the multiple metrology schemes available to reduce process variability.

1.2. Thesis Organization

This thesis presents a framework to integrate the aspects of metrology, modeling and process control for the deep ultraviolet lithography (DUV) sequence.

This thesis begins with Chapter 2 providing a review of optimization, learning and control algorithms. Simulated Annealing (SA) is the optimization algorithm discussed due to its global optimization behavior. In this thesis, SA is applied to parameter extraction required in resist modeling and lithography recipe generation. An introduction to Neural Networks (NN) is presented as a candidate for a real time learning algorithm. The final section presents an introduction to Kalman Filters (KF) as an algorithm suitable for run-to-run control and for sensor fusion.

Chapter 3 presents metrology for the thin film processes (deposition and spin-on/soft-bake), for the exposure and post-exposure bake (PEB) processes, and for the develop and etch processes. The metrology study consists of identifying the observables and their correlation to the final quantity of interest, identifying the appropriate sensor that could be potentially integrated in-line/in-situ, developing an algorithm that could be used for real-time use, and finally presenting results from such an application. The algorithms presented in this chapter include the NN-ASA algorithm for rapid thin film optical constant extraction, the modeling of deprotection induced thickness loss, and the algorithm to reconstruct profiles using specular spectroscopic scatterometry.

Chapter 4 presents the static and dynamic models developed to explain the deprotection induced thickness loss mechanism. This includes the simulation framework used, and experimental results validating the proposed mechanism. Results are presented for commercially available chemically amplified resists from Shipley and Clariant.

Chapter 5 presents a framework for efficient lithography simulator calibration, thus reducing reliance on experimentation. Experimental results for both unpatterned and patterned characterization experiments are matched to the output of theoretical models presented earlier over a training set. This framework demonstrates excellent predictive capabilities when used with a test set of experimental data, and has the potential to improve yield ramp rates and reduce development costs when implemented commercially.

In chapter 6, the various sources of variability in the lithography sequence are enumerated followed by metrology schemes for each of the process steps. This is followed by the run-to-run controller design that includes the experiments performed, the process and drift models used and two different scenarios for the run-to-run control architecture. The first scenario assumes a less aggressive metrology integration, relying on off-line CD metrology, while the second scenario assumes a complete in-line sensor integration for both intermediate as well as CD metrology.

Chapter 7 provides concluding remarks and future work in the area of metrology, modeling and control of the lithography sequence.

Chapter 2 Modeling, Optimization and Control Algorithms

2.1. Introduction

The implementation of in-line/in-situ sensors in a real time or run-to-run control framework requires high-speed and accurate algorithms for model-building, optimization and control. In this thesis, the optimization algorithms are used for high dimensionality parameter extraction in non-linear functions, while the modeling algorithms are used to speed up the parameter extraction process so as to make it practical for real-time applications. The control algorithms are used for sensor fusion in a run-to-run control environment. Specifically, this chapter covers Adaptive Simulated Annealing (ASA), Neural Networks (NN) and Kalman Filters.

2.2. Optimization through Simulated Annealing

Simulated annealing (SA) [4] is a probabilistic optimization technique well suited to multi-modal, discrete, non-linear and non-differentiable functions. SA's main strength is its statistical guarantee of global minimization, even in the presence of many local minima. However, simulated annealing methods are notoriously slow. There are various approaches to address the speed problem in SA such as by using different annealing algorithms, including the cooling schedule and probability density function of the state space [4].

2.2.1. Introduction to Simulated Annealing

Pseudo-code for the SA algorithm is presented in Figure 4.1. The control parameter T is decreased after a number of transitions, L_i, and can, therefore, be described by a sequence of homogeneous Markov chains, each generated at a fixed value of T.

```
Procedure Homogeneous SA algorithm
Begin
    Initialize ($n$, $T_n$, $i$, $L_n$)
    while $n = 0$;
    Repeat
        Repeat
            Generate state $j$ a neighbor to $i$;
            Calculate $\delta E = E_i - E_j$;
            if Accept($\delta E$, $L_n$) == true then i = j
        until $L_n$;
        $\imath = n + 1$;
        Update $L_n$;
        Update $T_n$;
    until StoppingCriterion == true
End

Subroutine Accept($\delta E$, $T_n$)
    if $\delta E \le 0$ then return true
    else
        return true with probability h(δE)
    endif
```

Figure 2.1. Pseudo-code for the Simulated Annealing Algorithm

There are five major components in SA implementation:

1) Temperature function T_n, or cooling schedule. T_n is the "temperature" parameter, n is the number of times the temperature parameter has changed. The initial value of T_0 is generally relatively high, so that most changes are accepted and there is little chance of the algorithm been trapped in local minimum. The cooling schedule is used to reduce the temperature parameter through the process of optimization.

2) Repetition function L_n. This is to decide how many changes are to be attempted at each value of T.

3) Probability density $g(x)$ of state-space of Dm parameters.

4) Probability $h(\delta(E, T_n))$ for acceptance of new cost-function given the previous state.

5) Stopping criterion. This is to decide how to terminate the algorithm.

2.2.2. Adaptive Simulated Annealing

There are numerous algorithms that attempt to overcome the disadvantages of simulated annealing, viz. the inability to rapidly converge to the global minimum. One of the most promising of these algorithms, for the constrained optimization problem, is the Adaptive Simulated Annealing (ASA) [5].

In ASA, there are two temperature notations, namely the parameter temperature T_i, associated with the ith parameter and the cost temperature T_{cost}.

T_i controls the generation function of the ith parameter. The state of ith parameter ${x^i}_{k+1}$ at annealing time $k+1$ with the range ${x^i}_{k+1} \in [A_i, B_i]$ is calculated from the previous state x^i_k by

$$ {x^i}_{k+1} = {x^i}_k + p^i(B_i - A_i) \tag{2.1} $$

where $p^i \in [-1, 1]$. The generation function is

$$ g_T(p) = \prod_{i=1}^{D} \frac{1}{2(|p^i| + T_i)\ln(1 + 1/T_i)} , \tag{2.2} $$

and p^i is generated from value u^i drawn from the uniform distribution $u^i \in U[0, 1]$ by

$$ p^i = \operatorname{sgn}(u^i - 0.5)T_i[(1 + 1/T_i)^{[2u^i - 1]} - 1] . \tag{2.3} $$

If ${x^i}_{k+1}$ falls outside of the range $[A_i, B_i]$, p^i is re-generated until ${x^i}_{k+1}$ is in the correct range.

A cooling schedule for T_i is

$$ T_i(k) = T_{0i}\exp(-c_i {k_i}^{1/D}) \tag{2.4} $$

where T_{0i} is the initial temperature of the ith parameter, k_i is the generation number for the ith parameter, and c_i is the cooling scaling factor for T_i . $T_{0,i}$ is usually set to 1.

A cooling schedule for T_{cost} is given by

$$T_{\text{cost}}(k_{\text{cost}}) = T_{0,\text{cost}}\exp(-c_{\text{cost}}k_{\text{cos}t}^{1/D}) \tag{2.5}$$

where $T_{0,\text{cost}}$ is the initial temperature of the acceptance function, k_{cost} is the number of acceptance, and c_{cost} is the cooling scaling factor for T_{cost} . $T_{0,\text{cost}}$ is usually set to the average initial value of some initial sample runs.

There are two important tuning parameters, " Temperature Ratio Scale " s_r and " Temperature Anneal Scale " s_a to control c_i ,

$$c_i = -(\log(s_r))\exp\left(-\frac{\log(s_a)}{Dm}\right) \quad . \tag{2.6}$$

Another tuning parameter " Cost Parameter Scale Ratio " s_p is used to link c_{cost} and c_i,

$$c_{\text{cost}} = c_i s_p \ . \tag{2.7}$$

Even through c_i can be set according to the ith parameter, however, for simplicity, usually it is set to be independent of i.

2.3. Learning through Articial Neural Networks

Artificial Neural Networks are widely used in functional approximation and pattern classification applications due to their capability for modeling complex and highly non-linear functions. There are many different kinds of ANNs. Rosenblatt's Perceptron Model [6], the Hopfield Network [6], Multi-Layer Perceptron [7], Radial Basis Function Network [7], etc. are some examples. Neural Networks find extensive use in the industry in modeling processes which are inherently complex and hence difficult to formulate. In general, physical systems are characterized with the help of mathematical models. Very accurate models can be built when the physics underlying the system being modeled is known. In many cases however, the mechanism is either too complex for practical modeling, or unknown. This calls for empirical modeling techniques to develop approximate mathematical models, which are inferred from available data. ANNs have shown to provide efficient approximating functions for nonlinear models, even with large problem dimensionality, due to their highly parallel structure and powerful representational capacity.

Among all the architectures available, the Multi-Layer Perceptron and the Radial Basis Function Network (RBFN) exhibit the best performance in terms of convergence and training time for our functional approximation applications. An introduction to both these approaches is presented in the following sections.

2.3.1. Multi-Layer Perceptron

MLPs are a class of feedforward neural networks that typically consist of three types of layers, namely, the input layer, the hidden layers, and the output layer. In this sense they are a generalization of the single layer perceptrons [7].

Nodes in different layers are connected to each other via links characterized by "weights". The input to the ith node of the hth layer is the weighted sum of all the outputs from the h-1th layer. The model of each neuron in the network is associated with a continuously differentiable transfer function. The most commonly used form satisfying this condition is the sigmoidal transfer function. This is mathematically described as follows: Let x_{hij} be the input to the ith node of the hth layer from the jth node in the h-1 layer, and y_{hi} be the corresponding output. Then,

$$y_{hi} = \frac{1}{1 + e^{-\phi_{hi}}} \tag{2.8}$$

where

$$\Phi_{hi} = \sum_{j=1}^{S_{h-1}} w_{ji} x_{hij} + \theta_{hi} \text{ for i = 1,2,....} S_h \text{ and h = 1, 2, ... L} \tag{2.9}$$

The θ_{hj} term is the bias for the jth node and S_h is the number of neurons in the hth layer. A conventional MLP structure is shown in Figure 2.2

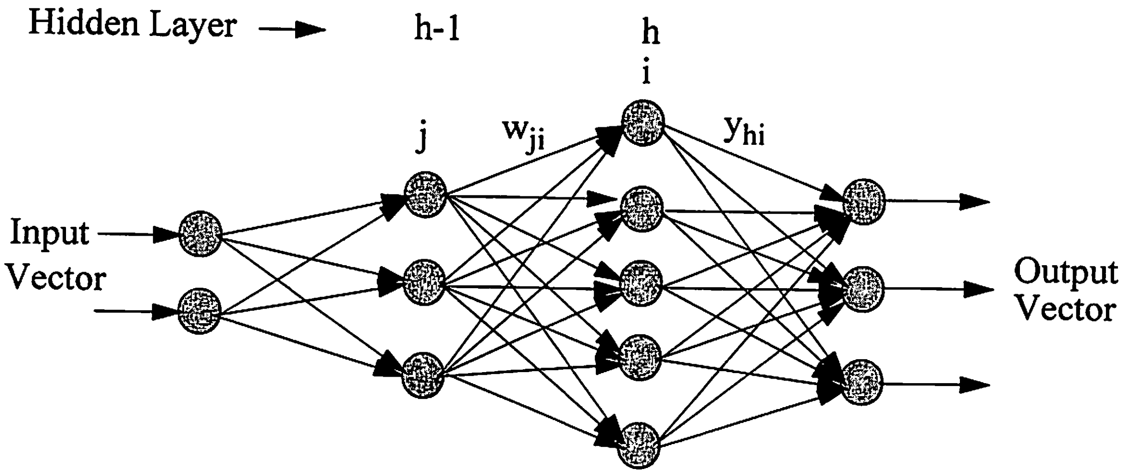


Figure 2.2. Architecture of a Multi Layer Perceptron

Typically, a neural network operates in two phases, namely training and testing. In the training phase of the MLP, the desired outputs are clamped to the output nodes for the corresponding inputs. The network 'learns' this input-output mapping by iteratively minimizing an error function. In this case, the error function, E, is the sum of squares of the difference between the calculated (y_j) and the desired output,

$$E = \sum_{j=1}^{N} (\hat{y}_j - y_j)^2 \tag{2.10}$$

where N is the number of output nodes.

MLPs have been successfully applied to solve complex problems, by adapting to them in a supervised manner using the popular back-propagation algorithm. Since this algorithm is based on the error correction rule, it can also be considered as a generalization of the Least Means Square (LMS) [8] algorithm. The back-propagation performs a stochastic gradient descent in the weight space. Basically, the error back-propagation process consists of two passes through the different layers of the network. In the forward pass, an input vector is applied to the input layer and its effect is propagated forward to the output layer to provide the response of the network to the input stimulus. The weights of the connections in the network remain fixed. In the backward pass, error is propagated backwards from the output layer, and the weights are adjusted using an error correction rule so as to make the actual response move closer to the desired response.

2.3.2. Radial Basis Function Network (RBFN)

Unlike Multi-Layer Perceptrons (MLPs), RBFNs use a distance metric in the input space to determine the hidden layer activations (Figure 2.3). As a result, the contours of constant activation of the hidden layer are hyperspheres instead of the hyperplanes used in MLPs. The contours are finite in length and form closed regions of significant activation, as opposed to MLPs where the contours are infinite in length and form semi-infinite regions of significant activation.

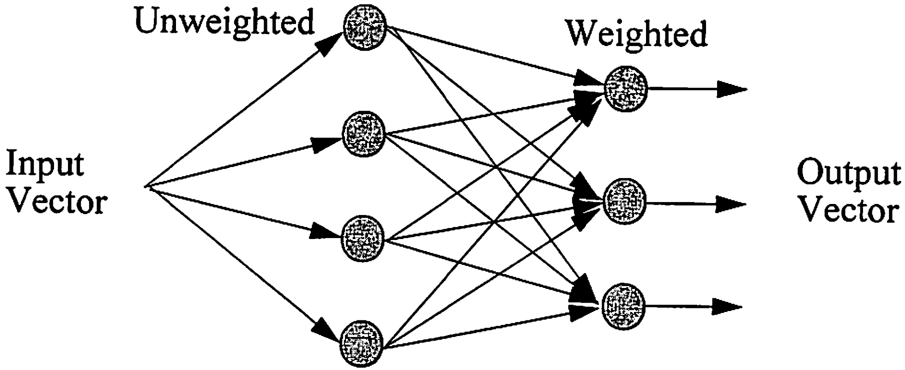


Figure 2.3. Architecture of a Radial Basis Function Network

1) The first layer is simply a fanout of the inputs to the hidden layer and are not weighted connections.

2) The hidden layer consists of H radial units plus one bias node with a constant activation of one. The transfer function of the hidden node is computed using a basis function ϕ,

$$a_h = \phi\left(\frac{\|x - x_h\|^2}{\sigma_h^2}\right) \tag{2.11}$$

where a_h is the output of the unit h in the hidden layer for a given input x.

Each RBF node is characterized by two internal parameters, namely x_h and σ_h: x_h is the position of the basis center in the N-dimensional feature space, and σ_h is a distance scaling parameter, which is the width in the input space over which the unit will have a significant influence. The connections in the second layer of the RBFN represent weights of the linear combination.

The output layer has nodes which are linear summation units. The value of the ith output node y_i is given by

$$y_i = \sum_{h=1}^{H+1} w_{ih} a_h = \sum_{h=1}^{H+1} w_{ih} \phi\left(\frac{\|x - x_h\|^2}{\sigma_h^2}\right) \tag{2.12}$$

where w_{ih} are the interconnection weights from the hidden nodes to the ith output node. The (H+1)th node is the bias node with $a_{H+1} = 1$.

2.3.3. Training the RBFN

There are several variations in the techniques for training the RBFN. The most commonly used technique is based on the algorithm suggested by Moody and Darken [9]. This method trains the RBFN in three sequential stages:

1) The first stage consists of determining the number of unit centers H and position of the unit centers x_h by the k-means clustering algorithm, an unsupervised technique that places unit centers centrally among clusters of training points.

2) Next the unit widths are determined using a nearest neighbor heuristic that ensures the smoothness and continuity of the fitted function. The width of any hidden unit is taken as the RMS (root mean square) distance to the P nearest unit centers, where P is a design parameter.

3) Finally, the weights of the second layer of connections are determined by linear regression, the objective function to be minimized being the sum of the squared error as given in Equation [2.10].

The optimality of an RBFN for a particular application is largely dependent on the number of nodes in the hidden layer. By using an excess number of nodes, we may overfit the function being approximated by a higher order function. In this case, the training points may give acceptable error, but the test points would give unsatisfactory results. Similarly, taking too few hidden nodes would result in a sub-optimal model.

The conventional k-Means algorithm is largely dependent on the number of clusters, k being the choice of the initial cluster centers and the order in which the data is presented. Linearly separable data are reasonably clustered by the k-means algorithm depending on the spatial properties of the training data. In training RBFNs, adaptive forms of the k- Means algorithms have been used to obtain optimum results.

In this algorithm, the number of clusters is automatically adjusted on the basis of spatial distribution of the samples. The k-Means algorithm is first applied by arbitrarily selecting the cluster centers, n_0. The minimum intercluster distance (*d*) is then calculated.

$$d = \min_{1 \le i,j \le n_0,\, i \ne j} \{dist(x_i - x_j)\} \quad \text{for i, j} = 1, 2, ..., n_0 \tag{2.13}$$

where X's are the n_0 cluster centers and dist is the Euclidean distance given by

$$dist(\vec{a}, \vec{b}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_m - b_m)^2} \tag{2.14}$$

in an m-dimensional space.

The diameter (D_k) of the kth cluster is defined as the maximum distance between two samples in cluster k. The largest diameter (*R*) is computed next. If x's are the points in the cluster k, the intracluster distance D_k is given by

$$D_k = \max_{i \ne j} \{dist(x_i, x_j)\} \quad \text{for i, j} = 1, 2, \ldots, n_k \tag{2.15}$$

where N_k is the number of points in cluster k and,

$$R = \max_{1 \le k \le n_0} (D_k) \tag{2.16}$$

When d > αR, (where α is an empirically preset threshold value) it means that the scatter plot of the points belonging to the largest cluster exceeds the threshold value that has been preset as a fraction of the largest diameter R. This intracluster distance can be reduced by increasing the number of clusters, n_0. Therefore, if $d > \alpha R$, the number of cluster centers is incremented. Otherwise, it is decremented. The iterative k-Means algorithm is used to obtain the new cluster centers. The algorithm converges when the number of clusters do not change.

2.4. Control Utilizing Kalman Filter

A Kalman Filter is an extension of Linear Least Squares estimation applied to stochastic processes. Kalman Filter development is done from a state space description of the desired and measured signals. Also, Kalman Filter does not assume stationarity of the desired signal, thus making it an ideal tool to handle practical process control problems. A thorough treatment of Kalman Filters can be found in [10].

The Kalman Filtering problem begins with the following signal model

$$\vec{x}_{k+1} = F_k \vec{x}_k + G_k \vec{w}_k \tag{2.17}$$

$$\vec{z}_k = H_k^* \vec{x}_k + \vec{v}_k \tag{2.18}$$

We assume $\vec{v}_k$, $\vec{w}_k$ are both zero mean, independent of each other, and have covariances $E[\vec{v}_k \vec{v}_l^*] = R_k \delta[k-l]$, $E[\vec{w}_k \vec{w}_l^*] = Q_k \delta[k-l]$. The initial state $\vec{x}_0$ is random with mean $\bar{x}_0$, covariance P_0, and is independent of $\vec{v}_k$, $\vec{w}_k$. Let $\vec{z}_k$ denote the set of observations $\{z_0, z_1, \ldots, z_k\}$, and define the estimator error covariance matrix as

$$\Sigma_{k+1|k} = E[(\vec{x}_{k+1} - \hat{x}_{k+1}(\vec{z}_k))(\vec{x}_{k+1} - \hat{x}_{k+1}(\vec{z}_k))^*] \tag{2.19}$$

where $\vec{x}_{k+1|k} = \ddot{x}_{k+1}(\vec{z}_k)$ is our estimate of $\vec{x}_{k+1}$ based on the data $\vec{z}_k$. The problem is to find the function $\ddot{x}_{k+1|k}(\vec{z}_k)$ that minimizes $\Sigma_{k+1|k}$. It is shown in [10] that the Kalman Filter, which is an affine function of $\vec{z}_k$, achieves an estimator error covariance, $\Sigma_{k+1|k}$ which is less than or equal to the estimator error covariance of any other affine estimator. The Kalman Filter is defined by the recursion relations:

$$(\hat{x}_{k+1|k} = (F_k - K_k H_k^*)\hat{x}_{k|k-1} + K_k \vec{z}_k), \hat{x}_{0|-1} = \bar{x}_0 \tag{2.20}$$

$$K_k = F_k \Sigma_{k|k-1} H_k [H_k^* \Sigma_{k|k-1} H_k + R_k]^{-1} \tag{2.21}$$

The estimation error covariance used in (2.21) to compute the Kalman gain, K_k, is given by a discrete time Riccati equation:

$$\Sigma_{k+1|k} = F_k[\Sigma_{k|k-1} - \Sigma_{k|k-1}H_k(H_k{}^*\Sigma_{k|k-1}H_k + R_k)^{-1}H_k{}^*\Sigma_{k|k-1}]F_k{}^* + G_kQ_kG_k{}^* \quad (2.22)$$

where $\Sigma_{0|-1} = P_0$

When $\vec{v}_k$, $\vec{w}_k$ and $\vec{x}_0$ are Gaussian, much more can be said about the Kalman Filter. $\hat{x}_{k+1}(\vec{z}_k)$ is actually the conditional mean $E[\vec{x}_{k+1}|\vec{z}_k]$, and $\Sigma_{k+1|k}$ is the conditional covariance $E[(\vec{x}_{k+1}-\bar{x}_{k+1})(\vec{x}_{k+1}-\bar{x}_{k+1})^*|\vec{z}_k]$. Thus the Kalman Filter equations become a mechanism for updating the entire conditional probability density of $\vec{x}_k$ [10]. To extend the use of the Kalman Filter to integrate measurements from multiple sensors at different measurement frequencies, the measurement update step is repeated as many times in a time step as the number of sensors collecting data in that time step.

Often the signal model that best represents our desired signal is time invariant, and for computational reasons we would like to estimate the signal with a time invariant filter. Although the Kalman Filter is in general a time varying filter, there are conditions under which it asymptotically becomes time invariant, and is truly time invariant with a proper choice of P_0. Our signal model is now:

$$\vec{x}_{k+1} = F\vec{x}_k + G\vec{w}_k \quad (2.23)$$

$$\vec{z}_k = H^*\vec{x}_k + \vec{v}_k \quad (2.24)$$

Again, we assume $\vec{v}_k$, $\vec{w}_k$ are both zero mean, independent of each other, and have covariances $E[\vec{v}_k \vec{v}_l{}^*] = R_k\delta[k-l]$, $E[w_kw_l{}^*] = Q_k\delta[k-l]$.

$$\bar{\Sigma} = F[\bar{\Sigma} - \bar{\Sigma}H(H^*\Sigma H + R)^{-1}H^*\bar{\Sigma}]F^* + GQG^* \quad (2.25)$$

As a result, the Kalman gain approaches a limiting value of

$$\bar{K} = F\bar{\Sigma}H[H^*\bar{\Sigma}H + R]^{-1} \quad (2.26)$$

It should also be clear that if P_0 is selected to be $\bar{\Sigma}$, the Kalman Filter will be truly time invariant, not just asymptotically time invariant.

Chapter 3 In-line/In-situ Metrology for the Pattern Transfer Process

3.1. Introduction

The DUV lithography process provides the process engineer with numerous opportunities to monitor the process and wafer state as shown in chapter 1. In-situ sensors with real time capability of analyzing data and using this information for closed loop control, are good candidates for a supervisory control scheme. Developing metrology for a process, however, requires knowledge of what to monitor in the process sequence, when to monitor it and how to monitor it. A system that answers all of the above questions adequately would constitute a practical metrology system.

The first step in designing a metrology system for a specific process is to decide on what quantity one is interested in monitoring. In the DUV lithography sequence, the final quantity of interest is the CD, which does not begin to form until the PEB step at the earliest. Hence, it becomes important to identify practical observables, available early in the process, that are strongly related to the final CD. This provides an opportunity to control the final quantity in a feed-forward sense.

Having identified the quantities that one would like to monitor, the second step is to evaluate existing sensors, and if appropriate, design new sensors that can measure these quantities. The most widely used off-line metrology tools are broadband reflectometers and ellipsometers for the measurement of the optical constants and thickness of uniform (unpatterned) thin films. An in-situ implementation of the broadband reflectometer has been successfully demonstrated in the past [11]. Thus, the third step in designing a metrology system for a specific process, is to identify a technique which is highly sensitive to the observable that needs to be measured. Examples include using reflectometry for thin film measurements, fourier transform infrared techniques (FTIR) for chemical analyses, etc.

The fourth and final step is to identify algorithms that allow rapid analysis of the sensor data that is related to the quantity being measured. Algorithms are required to extract relevant information from the data. To realize immediate control, algorithms are needed that can extract this information efficiently and feed this to other process modules for either feedback or feedforward control.

In the subsequent sections, we will describe metrology systems designed for the different process steps.

3.2. Thin Film Metrology

3.2.1. Optical Constants

Optical properties of any material can be described by the complex index of refraction, $\tilde{n} = n - jk$, where n is the refractive index and k is the extinction coefficient. Both n and k depend on the wavelength of light, λ, or, more fundamentally, to the photon energy, $E = (hc)/\lambda$. For the purpose of lithography control, the $n(\lambda)$ and $k(\lambda)$ at wavelengths in the 200-600 nm range should be determined. The reason optical constants are important is because they play a key role in defining the energy coupling during exposure. Therefore, a change in the optical constants affects the reflectivity of the thin film and hence the effective exposure energy absorbed by the resist film. This in turn affects the CD. The reason for broadband determination is to reduce the effect of sensor noise at certain wavelengths.

The Forouhi-Bloomer (F-B) equations are commonly used dispersion formulations derived from the Kramers Kronig relationship, with some simplifying assumptions that are suitable for most semiconductor materials [12]. The F-B equations are given by:

$$k(E) = \sum_{i}^{q} \frac{A_i (E - E_g)^2}{E^2 - B_i E + C_i}, \qquad n(E) = n(\infty) + \sum_{i=1}^{q} \frac{B_{0i} E + C_{0i}}{E^2 - B_i E + C_i}, \qquad (3.1)$$

where E_g represents the optical energy band-gap [12]. A characteristic of the F-B formulation is its relative simplicity. The number of terms required to approximate the dispersion behavior for different films varies according to the composition of the film. Most films require between 8-12 terms to be accurately represented from the deep ultraviolet to the near infrared wavelength range. This means that optimization must take place over a large set of parameters.

3.2.2. Broadband Reflectometry

Because of its inherent simplicity, normal incidence reflectometry is often integrated into the real-time process control paradigm for several reasons: good spatial resolution, high throughput, accuracy and ease of automation [13]. In most semiconductor thin-film reflectometry, the spectral reflectance of a sample is measured through the use of relative reflectance methods. In these methods, the comparison of the reflectance from the test sample with that of a standard sample (usually a bare wafer) is measured and analyzed. The theoretical reflectance can be calculated from the optical properties and thickness of each film. Measured and theoretical curves can be matched by fitting for the film thickness and optical properties. The problem is formulated as

$$\min_i \{\sum((R(\lambda_i) - R(n_i, k_i, \lambda_i, l))w_i)^2\} \tag{3.2}$$

where w_i is the optimization weight and the reflectance R is a function of optical properties and thicknesses of all the thin-films in the stack. Different settings of w_i yield different optimization speeds, sometimes even different results. Figure 5.1 shows the block diagram of the various modules required to tackle this problem. Two techniques that can be employed for optimization, that is the Adaptive Simulated Annealing (ASA) combined with the Neural Network (NN) modeling have been described in Chapter 2. .

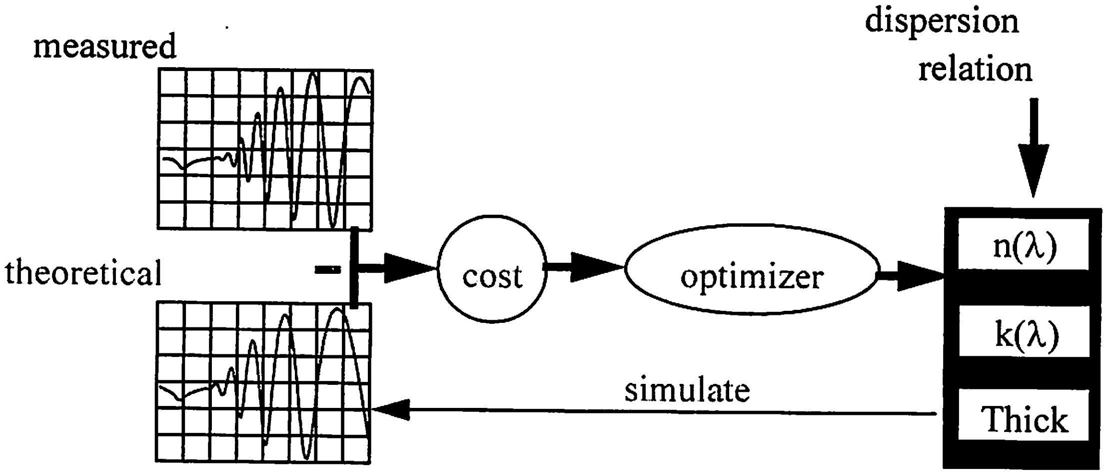


Figure 3.1. The objective is to match the simulated and measured broadband spectra by tuning the parameters of the dispersion relationship using the optimizer.

Due to the high dimensionality of the F-B dispersion relation (16 parameters) and the expensive cost function, the ASA technique takes an average of 10 minutes of SUN-SPARC 20 CPU time per run. If this is done off-line, it does not pose any problems. To increase the probability that the global minima is reached, the ASA could be run on the same wafer signal multiple times using different starting points. However, each computation, on an average, requires 10 minutes of CPU time, which would be considered impractical for any real time application. Further work has been done in reducing the metrology parameter space using a Bayesian screening technique [14]. This resulted in reducing the metrology parameter space to 4 parameters per film, and the CPU time to 1 minute on a

SUN SPARC 20 for a single film stack. A typical fit obtained with the ASA is shown in Figure [5.2].

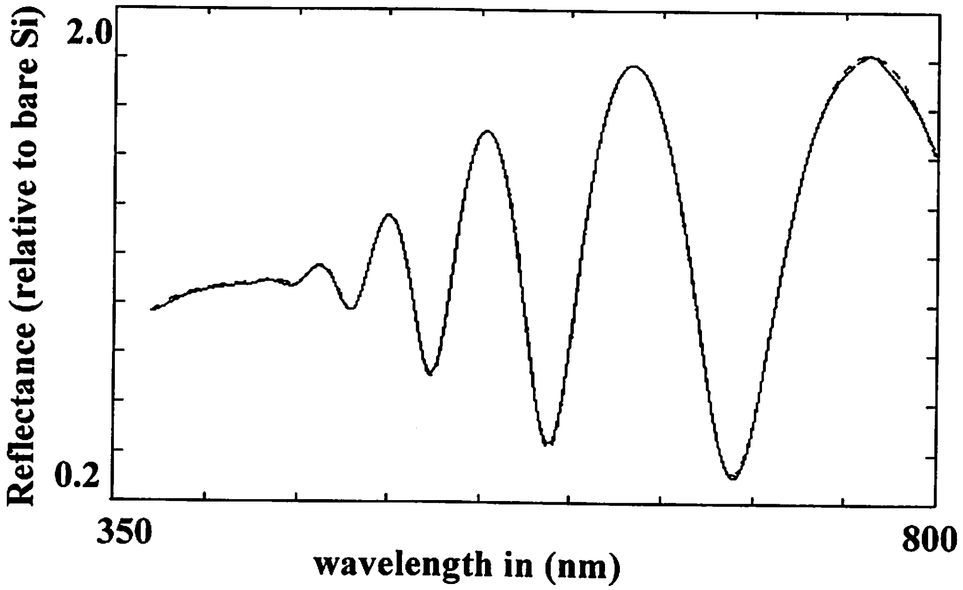


Figure 3.2. Results of the ASA Optimization Algorithm. Figure shows the simulated versus the experimental reflectance spectra (relative to bare Silicon).

The drawback of this technique lies in its computational requirements. Since our goal is to develop an algorithm that could be used in real time applications, we need to reduce the computation time down to a few seconds. This motivated the NN-ASA algorithm, described in the following section.

3.2.3. Neural Network - Adaptive Simulated Annealing (NN-ASA)

The NN-ASA algorithm combines the high speed optimization prowess of Neural Networks in a localized space with the global optimization abilities of Adaptive Simulated Annealing to provide a high speed, global optimization algorithm.

The NN-ASA algorithm is best explained through a case study. One lot of twenty 4" wafers were deposited with polysilicon with a thickness of 400 nm. Phophorus doping was used, and the LPCVD deposition time was two hours at 650 degrees Celcius. Due to the gas depletion effects intrinsic in conventional LPCVD tubes, there must be a temperature gradient along the length of the tube to compensate for the reduced reactive gas concentration. A difficulty with this process is that Poly-Si properties depend very strongly on deposition temperature, and will thus vary with wafer position along the tube [15]. The wafers were then measured off-line for reflectance using a commercial SC Technology broadband reflectometer. The data acquisition was done from 350 nm to 800 nm. There was one off-line measurement made per wafer, yielding a total of 20 measurements. These measurements were made on the center of the wafer using a footprint 1 mm in diameter for a duration of 3 seconds.

This algorithm was designed to enhance the ASA optimization routine so as to be suitable for real time applications. A block diagram for this algorithm is shown in Figure 3.3. The basic blocks of this setup are:

1) Parameter Extraction using ASA

2) Monte Carlo Simulation using the F-B formulation and Maxwell's equations

3) Spectral Feature Selection

4) Neural Network Training and Validation

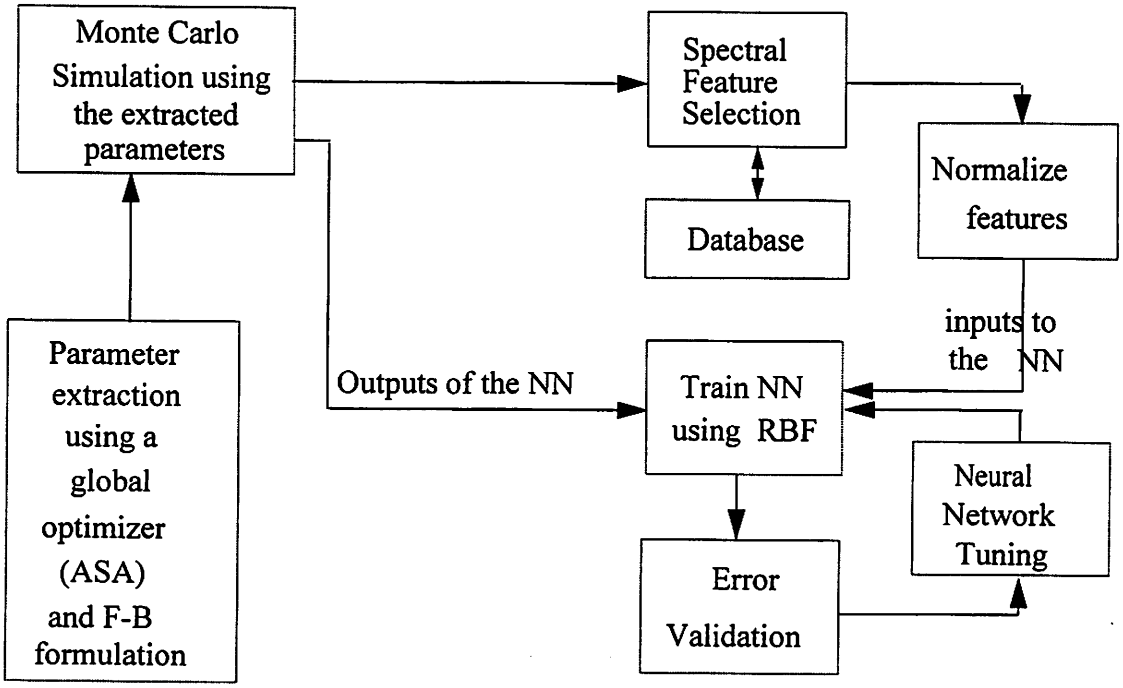


Figure 3.3. Block Diagram of the NN-ASA Algorithm

3.2.3.1. Parameter Extraction using ASA

The objective of this step was to extract the correct thickness values along with the corresponding optical constants. Due to the dependence of the optical constants of the film on the deposition conditions, it is incorrect to assume a fixed value for the optical constants and solve only for the film thickness. This necessitates the simultaneous extraction of both the optical constants and the thickness. Due to the high dimensionality and presence of local minima in the optimization problem, ASA was used as the optimization engine.

The twenty wafers were analyzed for the optical constants and the thickness. This algorithm was run three times per wafer to increase the chances of reaching the global minimum. The extraction procedure was automated and allowed to run overnight. We reached the global mimimum in two or more cases for each wafer. This provided us with the range of values over which the optical constants varied in the LPCVD process. This also provided us with the range of values over which the parameters of the F-B equations varied. The importance of this step is that it provides us with an idea of the natural variability of our LPCVD chamber.

3.2.3.2. Monte Carlo Simulation using the F-B formulation and Reflectance equations

We assumed that the typical variation in the parameters of the F-B equation were worse than those extracted from the ASA algorithm. A +/- 1% perturbation around the mean values was applied to all the statistically important parameters [14] of the F-B equation. We also used a +/- 50 nm perturbation to the mean thickness value whereas the variation in thickness, as extracted by ASA, was around +/- 30 nm. This was done to account for the fact that this particular lot may have had lower variability than the average.

A uniform distribution was used to generate values for each of the 4 parameters of the F-B equations as well as the thickness of the polysilicon (a native oxide of 25-45 Angstroms was assumed for all the wafers). 1000 vectors containing 5 elements each were generated. We thus had a poly-silicon, native oxide and silicon stack, with variable optical constants for the top-most layer. The next step was to generate the simulated broadband reflectance spectra using Maxwell's equations.

The optical properties of a layer of film are described by its characteristic matrix M. Assuming a normal incident angle, the characteristic matrix is given by

$$M = \begin{bmatrix} \cos(k_0\tilde{n}l) & \frac{1}{i \cdot N}\sin(k_0\tilde{n}l) \\ \frac{N}{i}\sin(k_0\tilde{n}l) & \cos(k_0\tilde{n}l) \end{bmatrix} \tag{3.3}$$

where $\tilde{n}$ is the complex index of refraction, l is the film thickness, $k_0 = \frac{2\pi}{\lambda}$. The characteristic matrix of a stack of N_T films is then

$$M = \prod_{j=1}^{N_T} M_j \tag{3.4}$$

Assume that the two end films are semi-infinite, in other words, the thickness values of the air and silicon substrate are ∞, the reflectivity of the entire stack is

$$R = \frac{(M_{11} + M_{12}\tilde{n}_{si})\tilde{n}_{air} - (M_{21} + M_{22}\tilde{n}_{si})}{(M_{11} + M_{12}\tilde{n}_{si})\tilde{n}_{air} + (M_{21} + M_{22}\tilde{n}_{si})} \tag{3.5}$$

where the subscripts of M refer to the row and column numbers respectively and $\tilde{n}$ denotes the complex index of refraction for the various layers. This step generates 1000 simulated broadband reflectance spectra in 2 minutes on a SUN SPARC 20.

3.2.3.3. Spectral Feature Selection

This step decides the features that should serve as the input to the neural network. This requires a physical understanding of the problem and is hence a very important step, as it lends physical intuition to the otherwise empirical neural network approach. It means looking at that part of the spectrum that carries maximum information about the optical constants of the film. This region would differ from stack to stack. When we are interested in measuring the optical constants of the polysilicon film in polysilicon-silicon stacks, we use the longer wavelengths, where poly is not absorbing and hence the reflectance spectrum contains the maximum information about the optical constants of the polysilicon. In the case of photoresist as in photoresist-polysilicon-silicon stacks, we use the shorter wavelengths since poly is opaque to the UV, and the resultant reflectance depends only on the layers deposited on poly. This step was automated by placing the part of the spectrum that needs to be used for each stack configuration in a database. We then reduce the dimensionality of the input features further by noting that the wavelengths at which the extrema occur, and the reflectance at those wavelengths, are correlated to the thickness and refractive indices of the film [16]. Stokowski's paper has shown that the film refractive index affects reflectance values away from the reflectance maxima. The larger changes in reflectance with refractive index occur at the minima. At a minimum, the reflectance value is related to the refractive indices of a non-absorbing film ($\tilde{n}$), its substrate ($\tilde{n}_3$) and the ambient medium ($\tilde{n}_1$) by the equation

$$\tilde{n} = \sqrt{\frac{1 \pm \sqrt{R}}{1 \mp \sqrt{R}}\tilde{n}_3\tilde{n}_1} \qquad (3.6)$$

Although we do not use this form of the equation, it is interesting to note that the broadband reflectance spectra can provide information on the refractive indices of the top layer in the non absorbing portion of the spectra.

The output of the physical filter is a vector of the wavelengths at which the maxima and minima occur, as well as the intensities at these extrema. It was observed that the neural network training improved when the inputs were normalized. One possible explanation for this is that we are using a k-means clustering algorithm with a single spread parameter in the Radial Basis Function. If we were to use multiple spread parameters in our design, we could avoid normalizing our inputs, but this would be at the cost of finding optimum values for a larger set of NN design parameters.

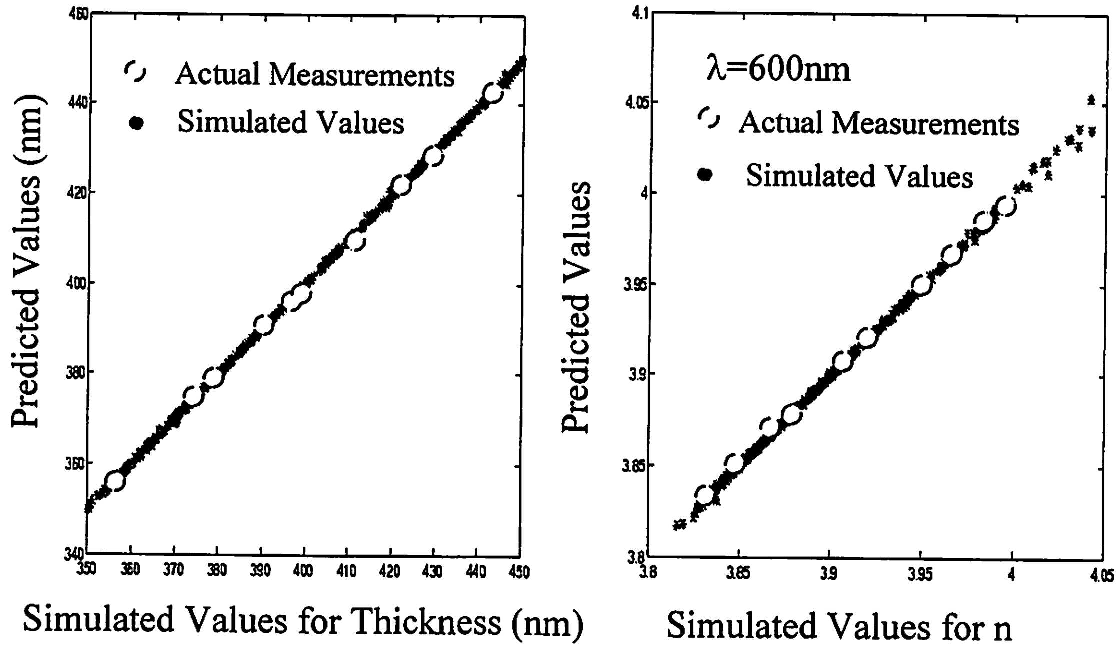
3.2.3.4. Neural Network Training and Validation

A radial basis function neural network architecture was used due to its well proven functional approximation capabilities [6]. The inputs to the network were the normalized outputs of the physical filter, while the outputs during the training stage were the optical constants used to generate the simulated reflectance spectra. The design parameter of the network was the spread of the Gaussian functions. We used a network with a single spread, so it was necessary to normalize the outputs of the physical filter.

The 1000 data points were divided into two blocks. One block of 600 was used for training, and the other block of 400 was used for testing. An automated routine was written in Matlab to pick the value of the spread that minimized the error of the testing samples. The values of the other design parameters were kept fixed at their optimum values.

3.2.4. Results of the NN-ASA Algorithm

The results of this optimization are shown in Figure 3.4. The figure shows the predicted values of thickness versus the simulated values, as well as the predicted values of the real part of the refractive index versus the simulated values at 600 nm. We chose to use this wavelength because most of the available data on polysilicon refractive indices in the literature is found at this wavelength. At $\lambda = 600$ nm, the extinction coefficient k is zero, so it was not predicted here. As can be seen from the figure, the prediction capabilities of the neural network were excellent. However, the main goal of using the neural network based optimization routine was to cut down on the computation time. This approach reduced the computation time on a SUN-SPARC 20 down from 1 minute to less than 1 second. This made it possible to use this algorithm for real time computation of the optical constants from broadband reflectance spectra. The training and testing phase took close to 1 hour on a SUN-SPARC 20. However, it is important to note that the ASA extraction and the neural network training and testing are both one-time tasks and can be done off-line.



Parameter	Range	Values	σ
Thickness	100 nm	350 - 450 nm	0.300 nm
n(600nm)	0.25	3.800 - 4.050	0.0012

Figure 3.4. Performance of the NN-ASA algorithm for poly-Si on native oxide on Si stack.

3.3. Exposure and PEB Metrology

3.3.1. Deprotection Induced Thickness Loss (DITL)

Chemically Amplified Resists (CARs) are composed of a polymer resin, which is very soluble in an aqueous base developer due to the presence of hydroxyl groups. These hydroxyl groups are "blocked" by reacting the hydroxyl group with some longer chain molecule, such as a t-BOC group, resulting in a very slowly dissolving polymer. In addition, there are possibly some dyes and additives along with the casting solvent.

The mechanism can be broken down into the initiation, the deprotection and the quenching stages. In the initiation phase, the exposure energy causes the Photo-Acid Generator (PAG) to produce acid. In the deprotection phase, these H+ ions attack the side chains (t-BOC) of the polymer and generate more H+ ions, thus making the resist even more soluble. This takes place in the presence of heat. In the quenching stage, the H+ ions are slowly quenched by anything more basic than the acid, such as the additives and the byproducts of the reaction. In short, the t-BOC blocked polymer undergoes acidolysis to generate the soluble hydroxyl group in the presence of acid and heat [17]. (Figure 3.5)

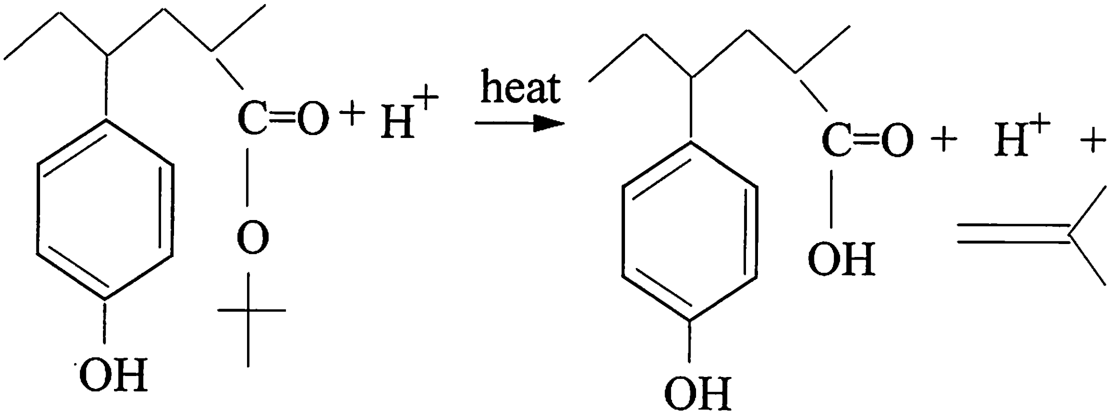


Figure 3.5. Resist Mechanism during the Exposure and Post Exposure Bake Steps for a commercial DUV photoresist.

The blocking group is such an effective inhibitor of dissolution, that nearly every blocked site on the polymer must be deblocked in order to obtain significant dissolution. Thus the photoresist is usually made more "sensitive" by only partially blocking the resin. Typical photoresists block 10-30% of the hydroxyl groups, 20% being a common value [18][17]. The cleaved t-BOC is volatile and evaporates, causing film shrinkage in the exposed areas. The extent of this exposed photoresist thinning is dependent on the molecular weight of the blocking groups.

There does not seem to be any universal definition for deprotection due to the different resist chemistries. However, in all its definitions, the term refers to the amount of deblocking of the resin. For the resist chemistry shown in figure 2.1, the deprotection reaction is quantitatively followed by monitoring the loss of the ester (C-O-C 1150 cm^{-1}) and the gain of the hydroxyl (O-H 3100 - 3400 cm^{-1}) vibrational bands [19][23][24]. The larger the exposure energy, the greater the number of H+ ions generated. Similarly, increasing the PEB temperature increases the amount of reaction between the H+ ions and the side chains, and hence the deprotection. The deprotection D is measured by taking the ratio of the integrated areas at a given exposure to the integrated area of the absorbance plot for no exposure (R_i) and subtracting from 1.

$$R_i = \frac{A_{dester}}{A_{0ester}} \tag{3.7}$$

$$D_{ester} = 1 - R_i \tag{3.8}$$

Theoretically, the CD is a stong function of the exposure and PEB process parameters. However, since the CD is formed only during the develop step, being able to predict it through measurements of other observables that are also strong functions of the exposure and PEB steps, would help in predicting the CD before it is even formed. A study was carried out to identify an observable that demonstrated such a strong dependence on the exposure and PEB process settings, while allowing a simple sensor setup for its measurement.

The process inputs for this experiment were the exposure dose and the PEB temperature, while the response variables were the amount of deprotection and the exposed area resist thickness loss. The exposure dose was varied from 1 mJ/cm^2 to 5 mJ/cm^2 in steps of 0.5 mJ/cm^2 on each wafer (nine blanket area exposures). The PEB temperature was varied from 130 degrees Celsius to 150 degrees Celsius in 10 degrees Celsius steps, thus requiring a total of 3 wafers.

3.3.2. Broadband Reflectometry for Thickness Loss

The same sensor as described in Section 3.2.2 can be used for this film thickness application.

3.3.3. Correlating Thickness Loss to Deprotection

Any FTIR experiment usually requires the use of highly reflective substrates to increase the signal to noise ratio. This is usually done by coating the wafers with either Aluminum or Tungsten. Hence, the three wafers were first coated with tungsten. The wafers were primed with HMDS on the FSI wafer track, after which UV5, a chemically amplified photoresist was spun on and soft-baked using the standard process recipe. These wafers were taken to a Tencor 1250 single angle broadband ellipsometer for pre-exposure thickness measurements. The wafers were exposed using the ISI stepper (KrF excimer laser) at 248 nm with the pattern shown in Figure 3.6 and post exposure baked.

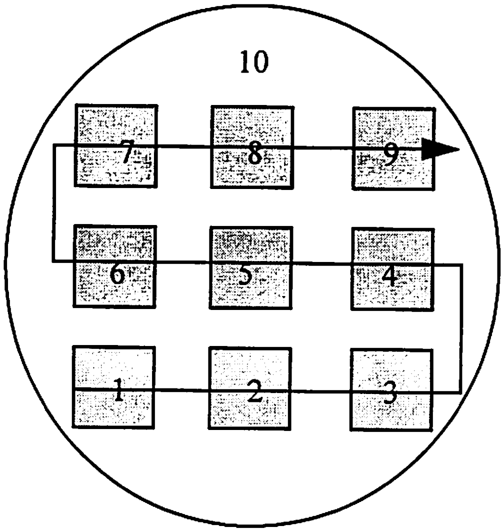


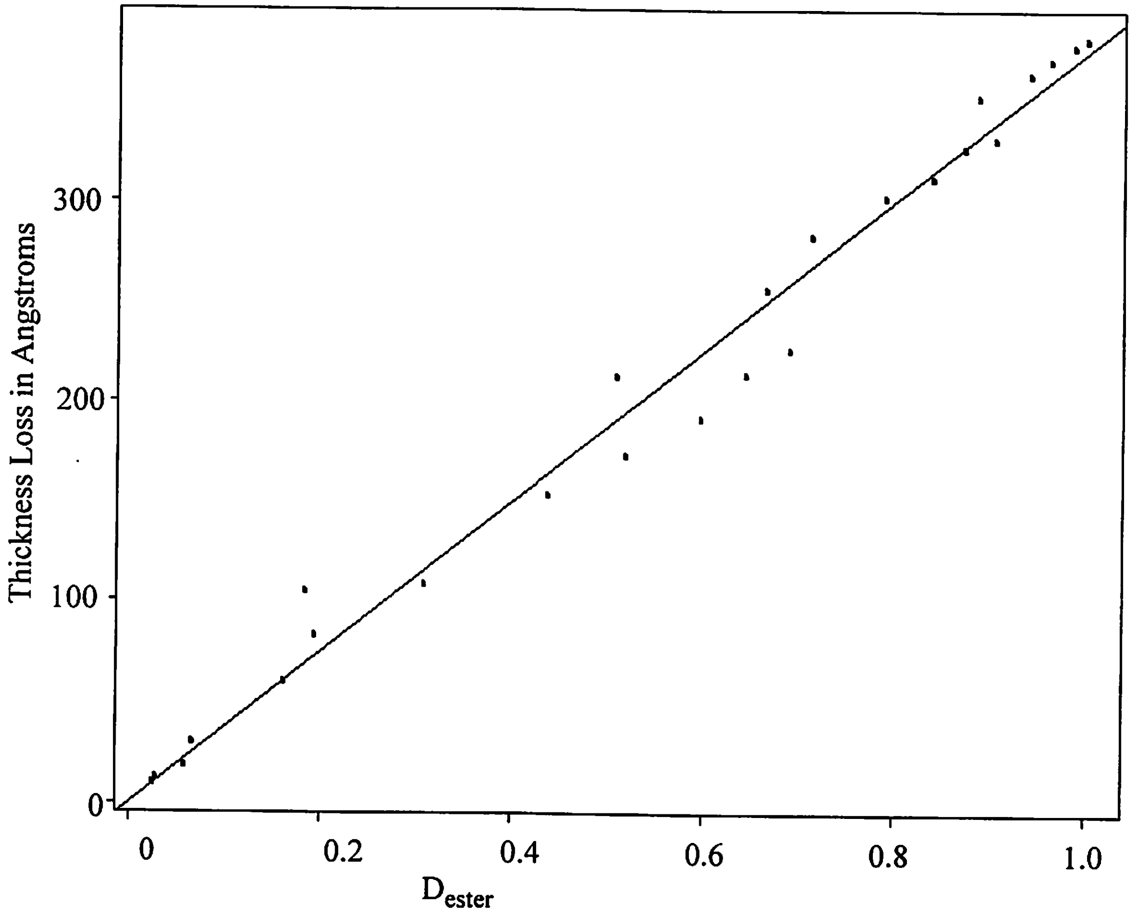
Figure 3.6. Layout of the blanket exposure areas on the wafer

The wafers were taken once again to the Tencor for post bake measurements of thickness. This provided the thickness loss as a function of the different exposure doses and PEB temperatures. Next, the wafers were taken to a FTIR tool where a Bio-Rad Spectrometer was used to measure the IR absorption of the hydroxy (O-H; 3100 - 3400 cm^{-1}) and ester (C-O-C; 1150 cm^{-1}) vibrational bands. All the thickness loss and deprotection measurements were made on the exposed areas (1 - 9) as well as on one unexposed area (10) on the wafer. The wafers were fractured to facilitate measurement. Thirty two scans were used with a resolution of 1 cm^{-1}. The time required for a single measurement was about 1 minute. The integration of the spectra to yield the deprotection was done using a computer macro.

3.3.4. Results of DITL Metrology

The thickness loss measured in the unexposed areas of the wafers was assumed to be due to solvent evaporation. This value was subtracted from all the thickness loss measurements of the corresponding wafer. The aim was to correlate this resultant thickness loss to the amount of deprotection. The deprotection was extracted using Eq. (3.7).

Linear Regression was used to build a model for the thickness loss in the exposed areas as a function of the amount of deprotection. Figure 3.7 shows the fit.



Summary of Fit: Multiple $R^2 = 0.9956$

Average model prediction error = 16.52 on 24 degrees of freedom

$F_{1,24}=5460$; model is highly significant

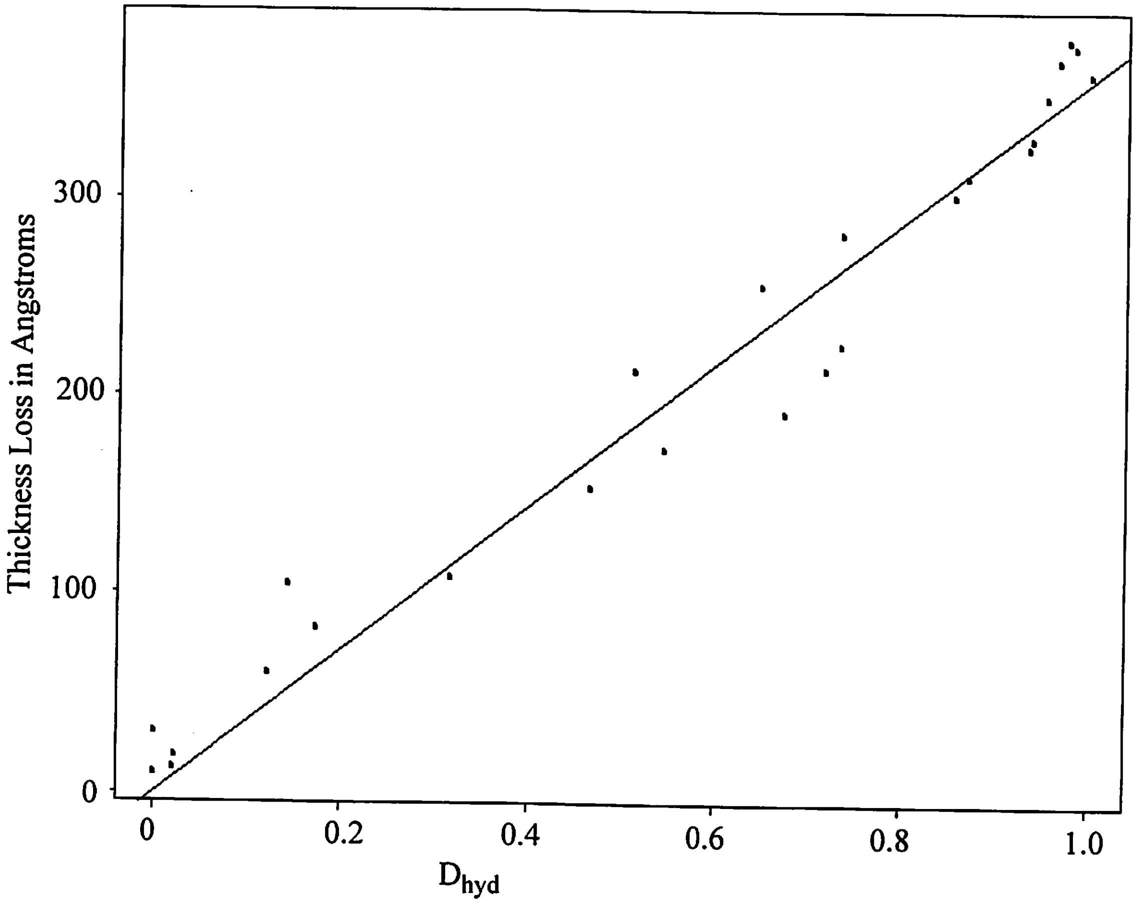
Figure 3.7. Thickness loss as a function of the deprotection measured by monitoring the normalized ester absorbance.

Model	**Value**	**Std. Error**	**t value**	**Pr(>|t|)**
Slope	375.04	5.08	73.89	0.0000

The final model for thickness loss as a function of deprotection is

$$T_{loss} = 375.04 \times D_{ester} \tag{3.9}$$

A similar study was done using the integrated hydroxy absorbance as a measure of deprotection, correlated to the final thickness loss. Even though this signal was more noisy due to the broad hydroxy absorbance bands, the model is significant. .



Summary of Fit: Multiple $R^2 = 0.9897$

Average model prediction error = 25.33 on 24 degrees of freedom

$F_{1,24} = 2310$; model is highly significant

Figure 3.8. Thickness loss as a function of the deprotection measured by monitoring the normalized hydroxy absorbance

Model	**Value**	**Std. Error**	**t value**	**Pr(>|t|)**
Slope	360.22	7.49	48.07	0.0000

The final model for thickness loss as a function of deprotection is

$$T_{loss} = 360.22 \times D_{hyd} \tag{3.10}$$

Note the absence of the intercept term in the two models. This is because we have subtracted the thickness loss in the unexposed regions, and have hence accounted for the solvent loss. The regression model showed no evidence of an intercept term (Pr > |t| = 0.84). Figure 3.9 shows the behavior of the normalized ester and hydroxy peaks processed at a given temperature as a function of exposure dose.

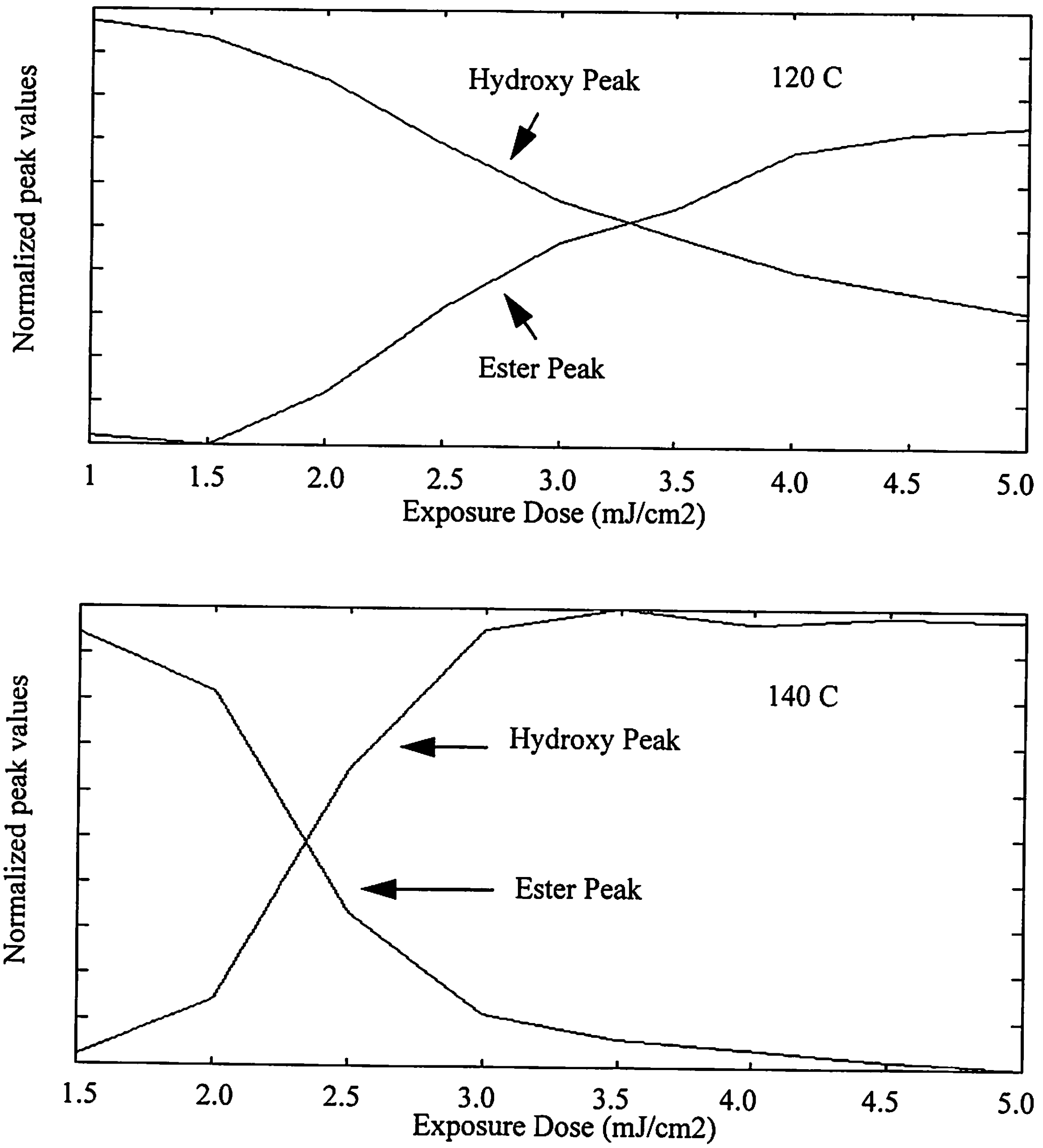


Figure 3.9. The normalized ester and hydroxy peaks as a function of exposure dose for different PEB temperatures.

3.4. Post-Develop and Post-Etch Profile Metrology

3.4.1. Patterned Profiles

With the progress of deep sub-micron technology, the accurate and efficient measurement of parameters such as the line width, height/depth and side-wall profile are becoming increasingly critical in developing and characterizing lithography and plasma etch processes. Scanning electron microscopes (SEM) and atomic force microscopes (AFM) can deliver direct images of small structures, but they are expensive, and can be time-consuming or even destructive. Electrical measurements can provide information on final effective CD linewidths, but they cannot be used *in-situ*, and cannot deliver reliable profile information. Niu, et. al. have shown the use of specular spectroscopic scatterometry, an optical metrology technique that uses a spectroscopic ellipsometer to extract patterned profiles [22]. The advantage of using an optical technique is that it lends itself to inexpensive, in-line measurement schemes.

3.4.2. Spectroscopic Ellipsometry

Spectroscopic Ellipsometry (SE) has become an essential metrology tool for the semiconductor industry [23]. The basic principle of SE is based on the fact that linearly polarized incident light has reflection coefficients that depend on the direction of polarization. The two polarization directions of interest are the p or TM (electric field parallel to the plane of incidence) and the s or TE (electric field perpendicular to the plane of incidence). Any linearly polarized light can be decomposed into the p and s components. These component waves experience different amplitude attenuations and different absolute phase shifts upon reflection; hence, the state of polarization is changed. Ellipsometry refers to the measurement of the state of polarization before and after reflection for the purpose of studying the properties of the reflecting boundary. The measurement is usually expressed as

$$\rho = \tan\Psi e^{j\Delta} = \frac{\tilde{r}_p}{\tilde{r}_s} \tag{3.11}$$

where $\tilde{r}_p$ and $\tilde{r}_s$ are the complex reflection coefficient for TM and TE waves.

Ellipsometry derives its sensitivity from the fact that the polarization-altering properties of the reflecting boundary are modified significantly even when ultra-thin films are present. Consequently, ellipsometry has become a major means of characterizing thin films. The basics of ellipsometry is illustrated in Figure 3.10.

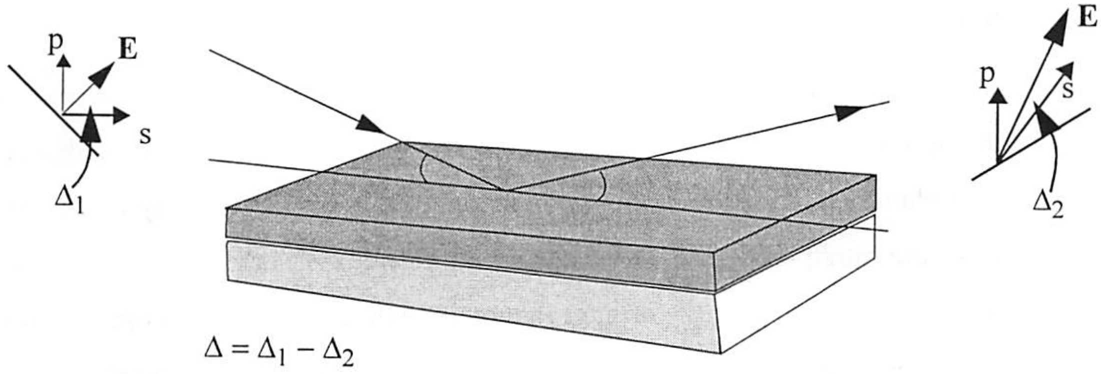


Figure 3.10. Spectroscopic ellipsometry measurements.

Ellipsometry measures polarization-state-in vs. polarization-state-out. Although visible light is commonly used, propagation of virtually any transverse wave can lead to an ellipsometric measurement.

The advantage of ellipsometry over reflectometry is its accuracy. First, ellipsometry measures the polarization state of light by looking at the ratio of values, rather than the absolute intensity of the reflected light. This property is especially useful in the DUV wavelength range, where very little light is typically available. Second, ellipsometry can gather the phase information (Δ) in addition to reflectivity ratio (ψ) information. Phase information provides more sensitivity to the thin-film variation. SE provides reading of $\Delta(\lambda)$ and $\psi(\lambda)$, which are then used to accurately characterize thickness and refractive index of thin films.

For lithography process control, semiconductor materials, anti-reflective coatings and photoresists need to be characterized by spectroscopic ellipsometry. In the DUV range, most of the above materials are absorbing. We now consider the case of an absorbing film which has a complex index of refraction at a given wavelength. The real and imaginary parts of the refractive index, and the thickness of the film cannot all be determined from a single set of ellipsometer readings. Since a set of ellipsometer readings consists of only two values, Ψ and Δ, it cannot determine all three quantities for the film. However, using a dispersion relation to describe the behavior of the optical constants over a broadband ensures that the refractive index does not vary randomly but follows a certain function.

There are several options for dispersion relationships, such as the Cauchy for the visible wavelength range, the F-B for the wavelength range from the deep ultraviolet to the near infrared.

Variable Angle Spectroscopic Ellipsometry (VASE) tools have been developed for the purpose [24][25]. The data analysis techniques depend on the used dispersion relation formulation. The method which commercial tools employ usually consists of two steps. First, the film thickness and the real part of the index n in the transparent region ($k = 0$, usually in the red or near infrared range) are extracted by a local optimization algorithm. Then, the film thickness is used for both n and k extraction at shorter wavelengths, where the film is absorbing. The advantage of this approach is that it achieves unique solutions. In each step, there are two unknowns and two measured parameters. The disadvantage is the inherent lack of accuracy. Because the determination of the wavelength range where $k = 0$ is quite arbitrary, small errors of the film thickness extracted in the transparent region can be propagated and magnified in the shorter wavelength region. In this thesis, we address this problem by using dispersion models derived from Kramers-Kronig relation and a global optimization technique.

The novel idea introduced in this work is the use of SE equipment for specular spectroscopic scatterometry (SSS). SSS measures the 0th order diffraction responses of a grating at multiple wavelengths. Given the 0th order diffraction responses, one can then attempt to reconstruct the grating profile. Conventional spectroscopic ellipsometry equipment can be directly used in this type of metrology. In other words, we do not need special equipment for specular spectroscopic scatterometry, as the cost of hardware is shifted to software. Compared to single-wavelength, variable-angle scatterometry, specular spectroscopic scatterometry has the advantage of the additional information contained in the spectral component.

A spectroscopic ellipsometer is used in this work for 1D gratings. With this configuration, the ratio of the 0th order complex transverse electric (TE) and transverse magnetic (TM) reflectivity $\rho = \tilde{r}_{p,0}/\tilde{r}_{s,0} = \tan\Psi e^{i\Delta}$ is measured, where $\tilde{r}_{p,0}$ is the 0th order TM reflectance coefficient and $\tilde{r}_{s,0}$ is the 0th order TE reflectance coefficient. Using a spectroscopic ellipsometer has two advantages. First, the measurement of the ratio of TE *and* TM provides more sensitivity than just using the measurement of TE *or* TM. Second, we can make direct use of a commercial ellipsometer, hence avoiding the additional hardware expense.

3.4.3. A Library-Based Methodology for CD Profile Extraction

The extraction of a CD profile can be viewed as an optimization problem. The objective is to find a profile whose simulated diffraction responses match the measured responses. Optimization techniques, such as simulated annealing and gradient based optimization techniques, can be applied. However, it is not feasible to exhaustively search for the optimum match when analyzing complicated profiles. A practical way is to generate the simulation responses before the measurement [26]. A library-based methodology for CD profile extraction has been used in this work.

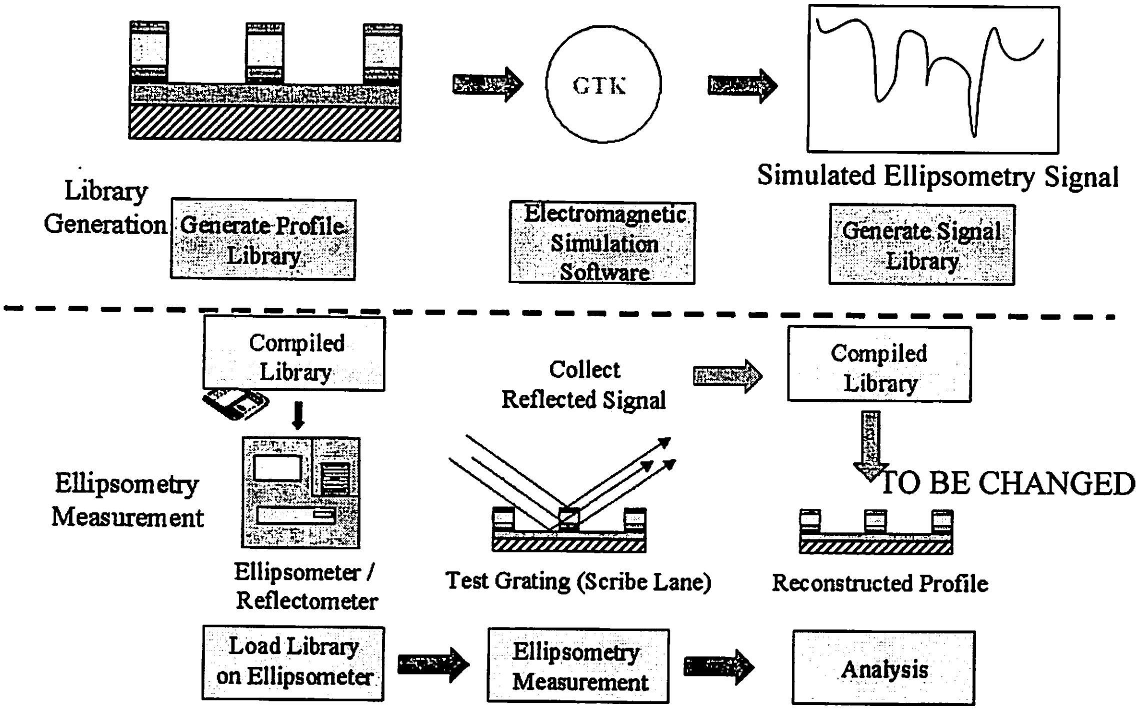


Figure 3.11. A library based methodology for CD profile extraction.

Figure 3.11. describes the extraction flow, including the following 3 steps:

[1.] Mask information, technology characteristics, thin-film information (optical properties n, k and thickness values), are used to obtain a collection of profiles. The profile information includes the information of the grating layer and of all the underneath layers. The profiles can be obtained by a tuned TCAD simulator, or by a random profile generator. The set of pre-simulated profiles should be sufficient and efficient. *Sufficient* means that there should be enough profiles in the library for most possible process results, while *efficient* means that there should not be too many unnecessary profiles in the library.

[2.] The profiles are used as inputs to a diffraction grating simulator, such as *gtk*, to generate the simulated diffraction responses. Usually the diffraction efficiencies, $\tan\Psi$ and $\cos\Delta$ are simulated over a wide range of wavelengths.

[3.] Specular spectroscopic scatterometry diffraction responses are measured and compared with the library. If the library is sufficient, there will be one or more profiles whose simulated responses will match those of the measured sample.

The key of the success for this library-based extraction methodology is the sufficiency of the library, combined with efficient simulation and search methods. The uniqueness of solution is an issue in this approach. In other words, it is possible that different profiles may lead to similar diffraction responses. There are several theoretical studies on this issue [27][28]. This problem has been addressed experimentally in [26][49]. Due to the larger number of degrees of freedom available in this approach (considering both phase and magnitude), we have not encountered non-uniqueness issues so far.

To extract the CD profile from a measurement, the measured $\tan\Psi$ and $\cos\Delta$ are compared with each simulated $\tan\Psi$ and $\cos\Delta$ in the library. If there is a "good" match between the measured and simulated signals, the corresponding CD profile is considered as the extracted profile. Mathematically, the matching is done by minimizing the cost function

$$\min\left\{\sum_{\lambda}((\log(\tan\Psi_{\text{measured},\lambda})-\log(\tan\Psi_{\text{theoretical},\lambda}))^2 w_{\tan\Psi,\lambda}+ \right. \quad (3.12)$$

$$\left.(\cos\Delta_{\text{measured},\lambda}-\cos\Delta_{\text{theoretical},\lambda})^2 w_{\cos\Delta,\lambda})\right\}$$

while searching the library. In our approach we set $w_{\tan\Psi,\lambda} = w_{\cos\Delta,\lambda} = 1$.

3.4.4. Results of Specular Spectroscopic Scatterometry

The experimental verification of specular spectroscopic scatterometry consists of two parts: the verification of the forward diffraction grating simulation from given CD profiles, and the verification of the inverse CD profile extraction from the scatterometry measurement. In this section, we will focus on the experimental work done in the inverse problem for three different case studies.

3.4.4.1. Photoresist on ARC on Silicon

A focus-exposure matrix experiment was done using UV-5, a commercial chemically amplified photoresist on a bottom anti-reflection coating (ARC) on Silicon stack. As shown in Figure 3.12, we have 5 focus settings and 7 dose settings. The dose settings are coded as -3, -2, -1, 0, 1, 2, and 3, indicating the values from 11.5 mJ/cm2 to 14.5 mJ/cm2 in steps of 0.5 mJ/cm2. The focus settings are coded as -2, -1, 0, 1, and 2, indicating the values from -0.2 μm to 0.2 μm in steps of 0.1 μm. In total we have 31 settings on each wafer. The mask has 0.28 μm/0.28 μm line/space gratings. After exposure and PEB, UV5 was developed to form 200 μm-by-200 μm grating regions. A KLA-Tencor Prometrix® UV-1280SE was used to measure the ratio of 0th order TE and TM fields. The incident angle is 70.5 degrees. The light beam was focused on a 30 μm-by-70 μm region.

The library-based CD profile extraction methodology described in Section 3.4.3 is implemented. A profile library is randomly generated from a set of profile primitives [26]. About 180,000 profiles are generated, and the corresponding diffraction responses are simulated by *gtk*. Because of the computational cost, each grating profile is simulated every 10nm from 240nm to 780nm. The number of retained orders for TE and TM are 31 and 41, respectively. Using *gtk*, the simulation for one profile, which includes both TE and TM on 53 wavelengths, takes approximately 2 minutes on a Sun UltraSparc I 167MHz workstation.

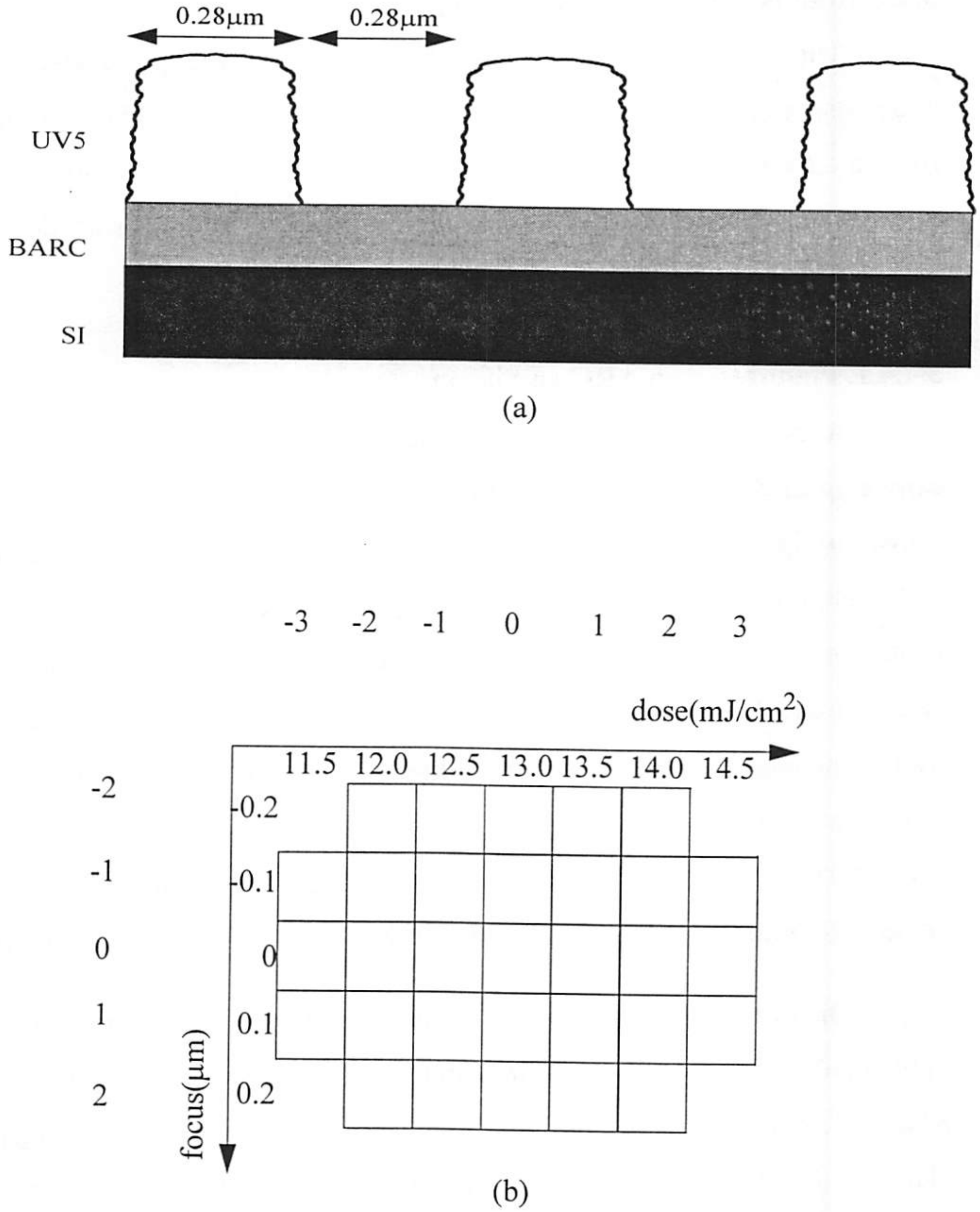


Figure 3.12. (a) grating structures; (b) Focus-exposure matrix experiment setup for experimental verification.

Figure 3.13 shows the comparison between the extracted profiles and AFM measurements across the entire focus-exposure matrix. The AFM measurements agree very closely with the extracted profiles.

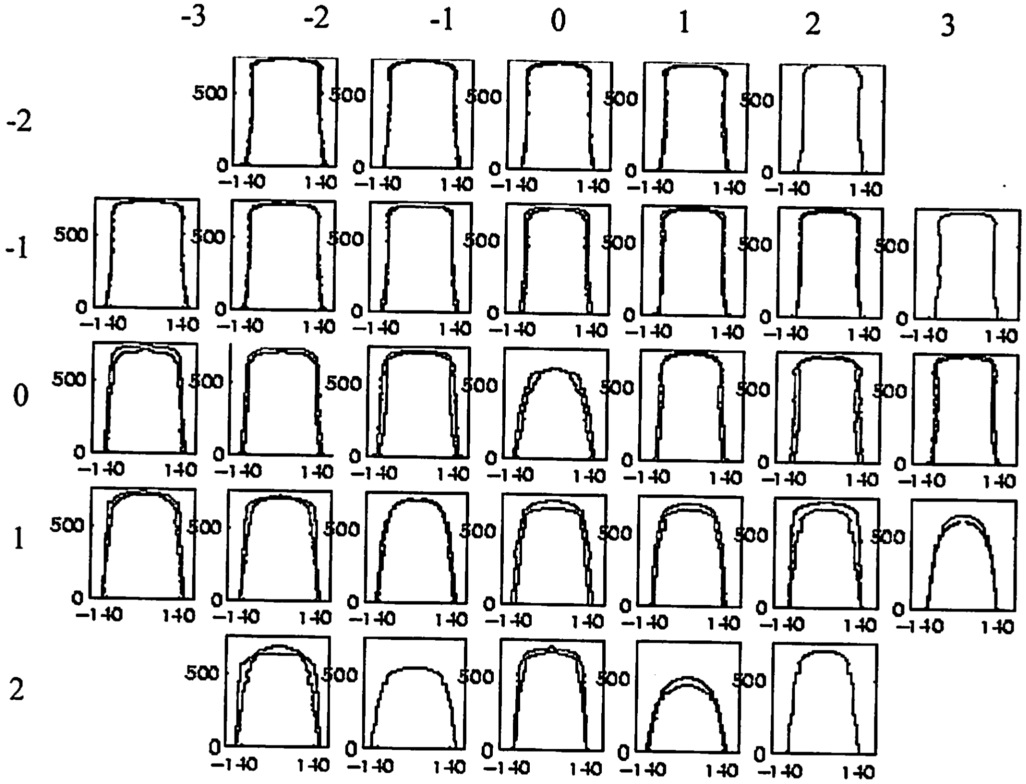
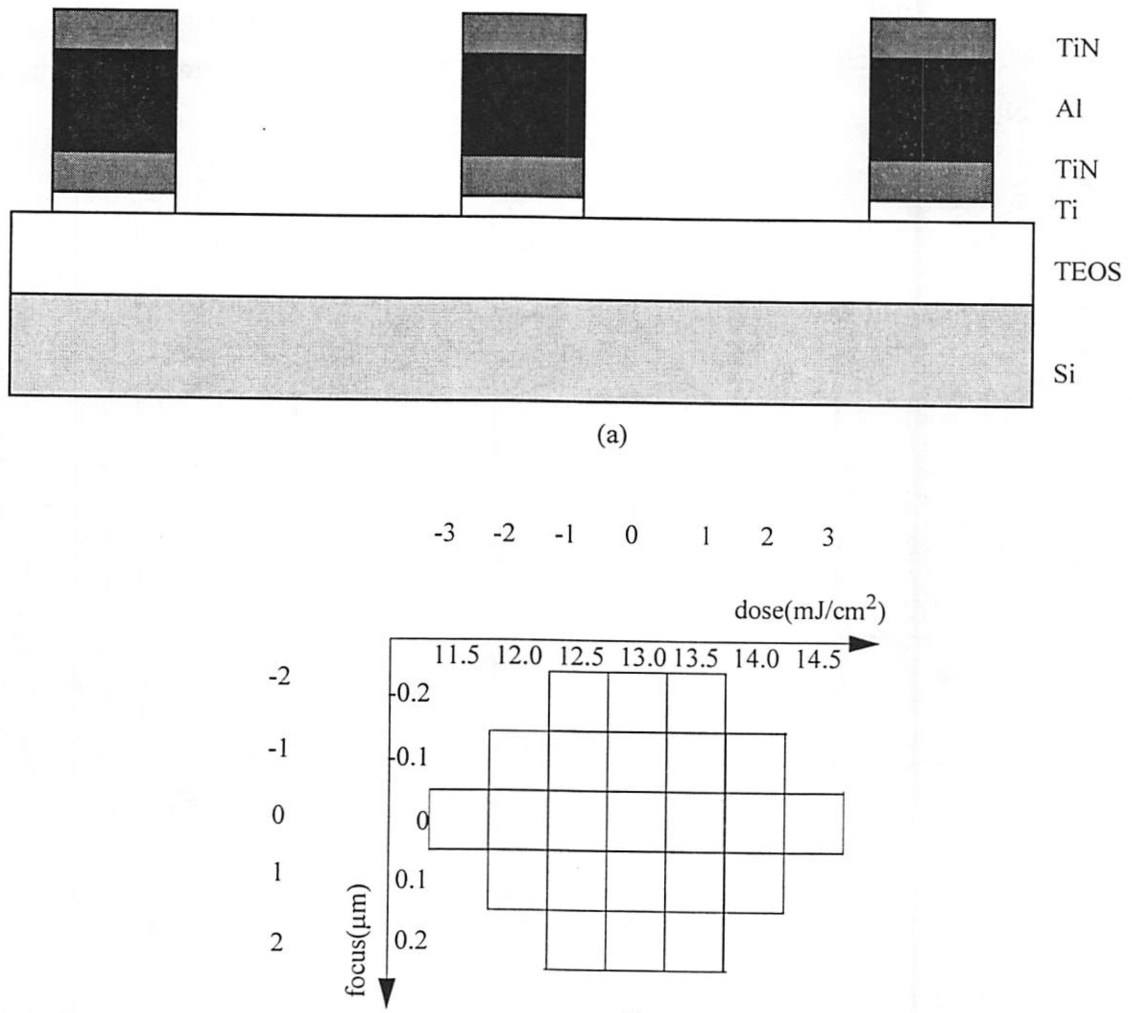


Figure 3.13. Comparison between the extracted grating profiles and the CD-AFM profile across the focus-exposure matrix. The four AFM profiles with (dose, focus) level of (2,-2), (3,-1), (-1,2) and (2,2) have not been measured.

3.4.4.2. Metal Stack

A focus-exposure matrix experiment was done on a stack involving TiN/Al/TiN/Ti/TEOS/Silicon with UV6, a commercially available DUV resist on an anti-reflective coating. As shown in Figure 3.12, we have 5 focus settings and 7 dose settings. The dose settings are coded as -3, -2, -1, 0, 1, 2, and 3, indicating the values from 11.5 mJ/cm2 to 14.5 mJ/cm2 in steps of 0.5 mJ/cm2. The focus settings are coded as -2, -1, 0, 1, and 2, indicating the values from -0.2 μm to 0.2 μm in steps of 0.1 μm. In total we have 31 settings on each wafer. The mask has 0.22 μm/0.44 μm line/space gratings. A KLA-Tencor Prometrix® UV-1280SE was used to measure the ratio of 0th order TE and TM fields. The incident angle is 70.5 degrees. The light beam was focused on a 30μm-by-70μm region.



(a)

(b)

Figure 3.14. (a) Grating structures; (b) Focus-exposure matrix experiment setup.

The library-based CD profile extraction methodology described in Section 3.4.3 is implemented. A profile library is randomly generated from a set of profile primitives [26]. About 200,000 profiles are generated, and the corresponding diffraction responses are simulated by *gtk*. Because of the computational cost, each grating profile is simulated every 10nm from 240nm to 780nm. The number of retained orders for TE and TM are 31 and 41, respectively. Using *gtk*, the simulation for one profile, which includes both TE and TM on 53 wavelengths, takes approximately 2 minutes on a Sun UltraSparc I 167MHz workstation.

Figure 3.15 shows the comparison of "top" CD measurement from CD-SEM and specular spectroscopic scatterometry across the focus-exposure matrix. The correlation coefficient is 0.9225. On the average, the CD-SEM "top" CD measurement is about 12.1nm larger than the "top" CD value extracted from specular spectroscopic scatterometry.

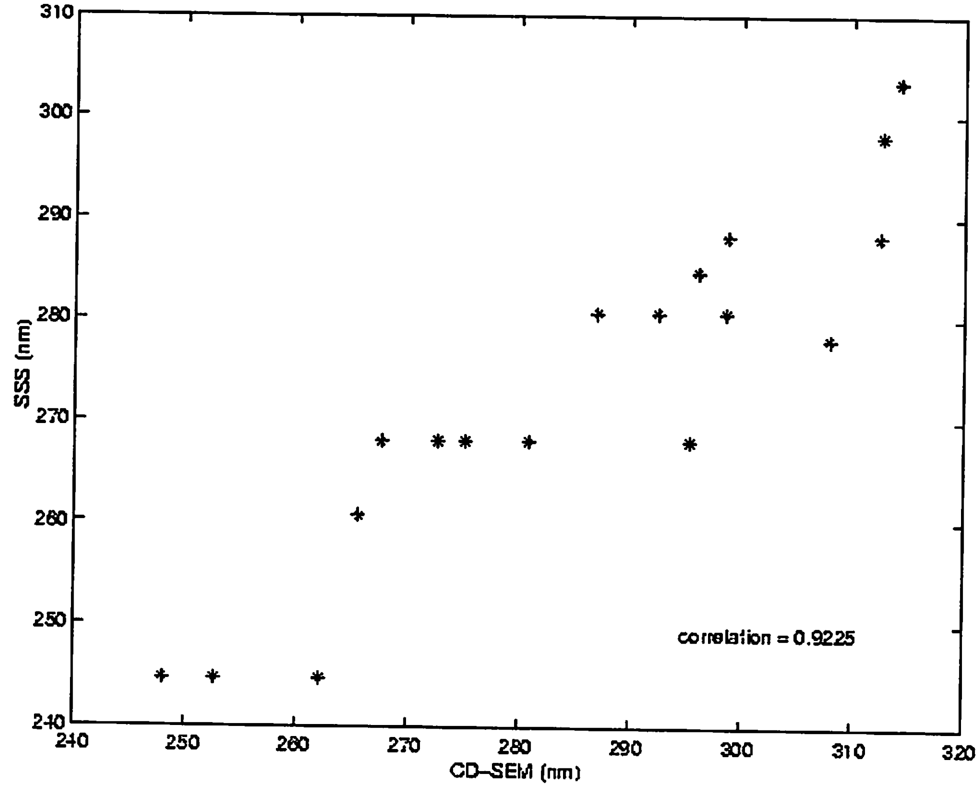


Figure 3.15. Comparison of top CD between specular spectroscopic scatterometry and a top-down CD-SEM.

While there were no cross-section SEM measurements made on this particular wafer, there were cross-section SEM measurements made on an identical "sister" wafer which yielded results within 7 nm (1-σ) of the SSS predicted results but more than 15 nm (1-σ) of the CD-SEM results. This provides considerable confidence in the accuracy of the SSS approach.

Chapter 4 Lithography Modeling

4.1. Introduction

As we enter the DUV lithography generation, the developmental phase of the photolithography process is becoming crucial due to the high costs associated with equipment and materials, and the continually reducing time-to-market. Improvements in the modeling of chemically amplified resists are necessary to extract the maximum possible information from the minimum amount of experimentation. This includes modeling of both the thin film as well as the pattern transfer sequence.

The extraction of optical constants of thin films from in-situ broadband reflectometry and ellipsometry signals is an application that requires algorithms that are both accurate and fast. An algorithm to solve the problem of extracting the optical constants from broadband reflectometry / ellipsometry signals is necessitated for the real time computations of complex functions.

All high activation, chemically amplified resist systems (CARS) exhibit a significant volume shrinkage during the post-exposure bake (PEB) step (typically 4% to 15% in current resist systems) [1]. Current models for PEB and development do not take into consideration this shrinkage for calculating line-widths. At present, workers at AMD [2] are developing a new methodology for characterizing PEB. Adding shrinkage characterization to the understanding of bake and development will significantly improve our understanding of lithography resist processing. This volume shrinkage manifests itself in the form of thickness loss, and it has been shown that this shrinkage is directly proportional to the deprotection of the resist in flood exposed films [3].

This work aims at describing the kinetics of the post-exposure bake process by tracking the volume shrinkage observed in high activation energy resists. We begin with a brief introduction to the physical mechanism underlying the volume shrinkage followed by the proposed mechanism for both the static as well as the dynamic models. An optimization framework is then presented, and it is used to extract the parameters of the models. Finally, results of modeling two different resists are discussed.

4.2. The Forouhi-Bloomer (F-B) Dispersion Relation

Optical properties of any material can be described by the complex index of refraction, $N = n - jk$, where n is the refractive index and k is the extinction coefficient. Both n and k depend on the wavelength of light, λ, as well as the photon energy, E, according to $E = (hc)/\lambda$. For the purpose of lithography control, the $n(\lambda)$ and $k(\lambda)$ at wavelengths in the range of the exposure wavelengths should be determined. The reason that the optical constants are important is because they are strongly correlated to the processing conditions and the reason for determination over a broadband is to reduce the effect of sensor noise at certain wavelengths. The Forouhi-Bloomer (F-B) equations are derived from the Kramers Kronig relationship with some simplifying assumptions that are suitable for most semiconductor materials [11]. The F-B equations are given by:

$$k(E) = \sum_{i}^{q} \frac{A_i(E - E_g)^2}{E^2 - B_iE + C_i}, \qquad n(E) = n(\infty) + \sum_{i=1}^{q} \frac{B_{0i}E + C_{0i}}{E^2 - B_iE + C_i}, \tag{4.1}$$

where E_g represents the optical energy band-gap [11]. A characteristic of the F-B equations is its relative simplicity. The number of terms required to approximate the dispersion relations for different films varies according to the composition of the film. Most films require between 2-4 terms to be represented with the required amount of accuracy. This means that optimization must take place over a large set of parameters.

4.3. Experimental Setup

A lot of twenty four inch wafers were deposited with polysilicon with a thickness of 400 nm. Phophorus doping was used and the time of deposition in the LPCVD chamber was two hours at 650 degrees Celcius. Due to the gas depletion effects intrinsic in conventional LPCVD chambers, the temperature needs to be increased along the length of the tube to compensate for the reduced deposition rate. A difficulty with this process is that Poly-Si properties depend very strongly on deposition temperature, and will thus vary with wafer position along the tube [12]. The wafers were then measured off-line for reflectance using a commercial SC Technology broadband reflectometer. The data acquisition was done from 350 nm to 800 nm. There was one measurement made per wafer yielding a total of 20 measurements. These measurements were made on the center of the wafer using a footprint 1 mm in diameter for a duration of 3 seconds. These measurements were made offline.

4.4. Optimization using the ASA algorithm

Since there exist local minima in the solution of Eq. (4.1) for a multiple-layer thin-film system, traditional optimization algorithms are not appropriate here. The major advantage of simulated annealing over other methods, as mentioned in the earlier chapter, is its ability to avoid becoming trapped at local minima. The algorithm employs a random search, which not only accepts changes that decrease the objective function, but also some changes that increase it, at least temporarily.

We used the ASA technique to extract the optical constants and the thickness from the reflectance spectra for all the 20 wafers. Due to the high dimensionality of the problem (16 parameters) and the expensive cost function, this technique took an average of 10 minutes of SUN-SPARC 20 CPU time per run. Since this was done off-line, it did not pose any problems. To increase the probability that the global minima was reached, the ASA was run on the same wafer signal three times using different starting points. The convergence probability of ASA algorithms from past experience was around 0.9. Using a binomial distribution, we estimated the probability of reaching the global minimum two or more times to be 0.97. However, each computation, on an average, required 10 minutes of CPU time which would be considered impractical for any real time application. Further work was done in reducing the metrology parameter space using a Bayesian screening technique [13]. This resulted in reducing the metrology parameter space to 4 parameters and the CPU time to 1 minute. The fit obtained with the ASA is shown in Fig 4.1

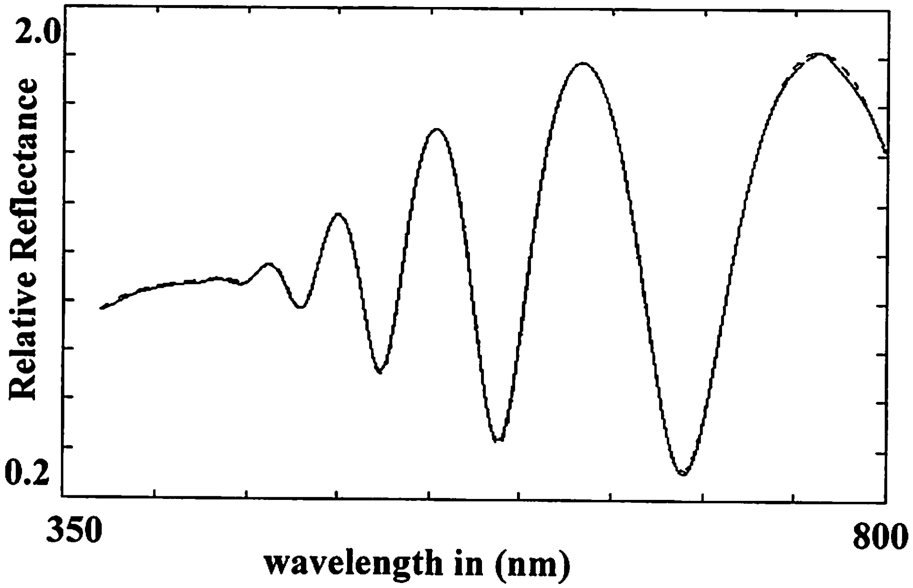


Figure 4.1. Results of the ASA Optimization Algorithm. Figure shows the simulated versus the experimental reflectance spectra

The drawback of this technique lay in its speed. Since our goal was to develop an algorithm that could be used in real time applications, we needed to reduce the computation time down to a few seconds. This motivated the NN-ASA algorithm, described in the following sections.

4.5. The Neural Network Enhanced ASA Optimization Algorithm

This algorithm was designed to enhance the ASA optimization routine so as to be suitable for real time applications. A block diagram for this algorithm is shown in Figure 5.3. The basic blocks of this setup are

1) Parameter Extraction using ASA

2) Monte Carlo Simulation using the F-B formulation and Maxwell's equations

3) Spectral Feature Selection

4) Neural Network Training and Validation

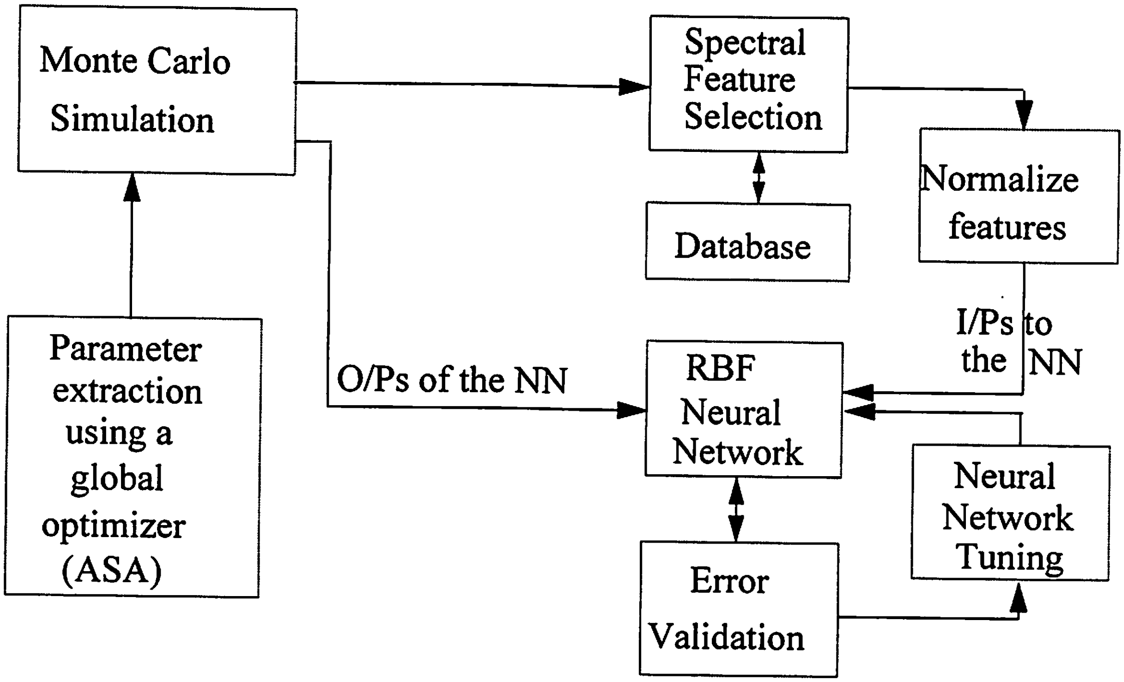


Figure 4.2. Block Diagram of the NN-ASA Algorithm

4.5.1. Parameter Extraction using ASA

This step is the ASA optimization technique described in the earlier chapter. The twenty wafers were analyzed for the optical constants and the thickness. This process was run three times per wafer to increase the chances of reaching the global minimum. The extraction procedure was automated and allowed to run overnight. We reached the global mimimum in two or more cases all the times, as was predicted in chapter 3 using the binomial distribution. This provided us with the range of values over which the optical constants varied in the LPCVD chamber. This also provided us with the range of values over which the parameters of the F-B equations varied. The importance of this step is that it provides us with an idea of the natural variability of our LPCVD chamber.

4.5.2. Monte Carlo Simulation using F-B formulation and Maxwell's equations

We assumed that the typical variation in the parameters of the F-B equation were worse than those extracted from the ASA algorithm. A +/- 1% perturbation around the mean values, was applied to all the statistically important parameters [?13?] of the F-B equation. We also used a +/- 50 nm perturbation to the mean thickness value whereas the typical variation in thickness was around +/- 30 nm. This was done to account for the fact that this particular lot may have had lower variability than the average.

A uniform distribution was used to generate values for each of the 4 parameters of the F-B equations as well as the thickness of the polysilicon (a native oxide of 25-45 Angstroms was assumed for all the wafers). 1000 vectors containing 5 elements each were generated. We thus had a Poly-Silicon on native oxide on Silicon stack with variable optical constants for the topmost layer. The next step was to generate the simulated broadband reflectance spectra using Maxwell's equations.

The optical properties of a layer of film are described by its characteristic matrix M. Assuming a normal incident angle, the characteristic matrix is given by

$$M = \begin{bmatrix} \cos(k_0 Nl) & \frac{1}{i \cdot N}\sin(k_0 Nl) \\ \frac{N}{i}\sin(k_0 Nl) & \cos(k_0 Nl) \end{bmatrix} \tag{4.2}$$

where N is the index of refraction, l is the film thickness, $k_0 = \frac{2\pi}{\lambda}$. The characteristic matrix of a stack of N_T films is then

$$M = \prod_{j=1}^{N_T} M_j \tag{4.3}$$

Assume that the two end films are semi-infinite, in other words, the thickness values of the air and silicon substrate are ∞, the reflectivity of the entire stack is

$$R = \frac{(M_{11} + M_{12}N_{si})N_{air} - (M_{21} + M_{22}N_{si})}{(M_{11} + M_{12}N_{si})N_{air} + (M_{21} + M_{22}N_{si})} \tag{4.4}$$

where the subscripts of M refer to the row and column numbers respectively and N denotes the complex index of refraction for the various layers. This step generates 1000 simulated broadband reflectance spectra.

4.5.3. Spectral Feature Selection

This step decides the features that should serve as the input to the neural network. This requires a physical understanding of the problem and is hence a very important step as it lends a physical intuition to the otherwise empirical neural network approach. It would mean looking at that part of the spectrum that carries maximum information about the opti cal constants of the film. This region would differ from stack to stack. When we are interested in measuring the optical constants of the polysilicon film in polysilicon-silicon stacks, we use the higher wavelengths where poly is not absorbing and hence the reflectance spectrum contains the maximum information about the optical constants of the polysilicon. In the case of photoresist as in photoresist-polysilicon-silicon stacks, we use the lower wavelengths since poly is opaque to the UV and the resultant reflectance depends only on the layers deposited on poly. This step was automated by placing the part of the spectrum that needs to be used for each stack configuration in a database. We then reduce the input features further by noting that the wavelengths at which the extrema occur and the intensities at those wavelengths are correlated to the thickness and refractive indices of the film [?14?]. Stokowski's paper has shown that the film refractive index affects reflectance values away from the reflectance maxima. The larger changes in reflectance with refractive index occur at the minima. At a minimum, the reflectance value is related to the refractive indices of a non-absorbing film (n), its substrate (n_3) and the ambient medium (n_1) by the equation

$$n = \sqrt{\frac{1 \pm \sqrt{R}}{1 \mp \sqrt{R}} n_3 n_1} \tag{4.5}$$

Although we do not use this form of the equation, it is interesting to note that the broadband reflectance spectra can provide information on the refractive indices of the top layer in the non absorbing portion of the spectra.

The output of the physical filter is a vector of the wavelengths at which the maxima and minima occur as well as the intensities at these extrema. It was observed that the neural network training improved when the inputs were normalized. One possible explanation for this is that we are using a K-means clustering algorithm with a single spread parameter in the Radial Basis Function. If we were to use multiple spread parameters in our design, we could avoid normalizing our inputs but this would be at the cost of finding optimum values for a larger set of NN design parameters.

4.5.4. Neural Network Training and Validation

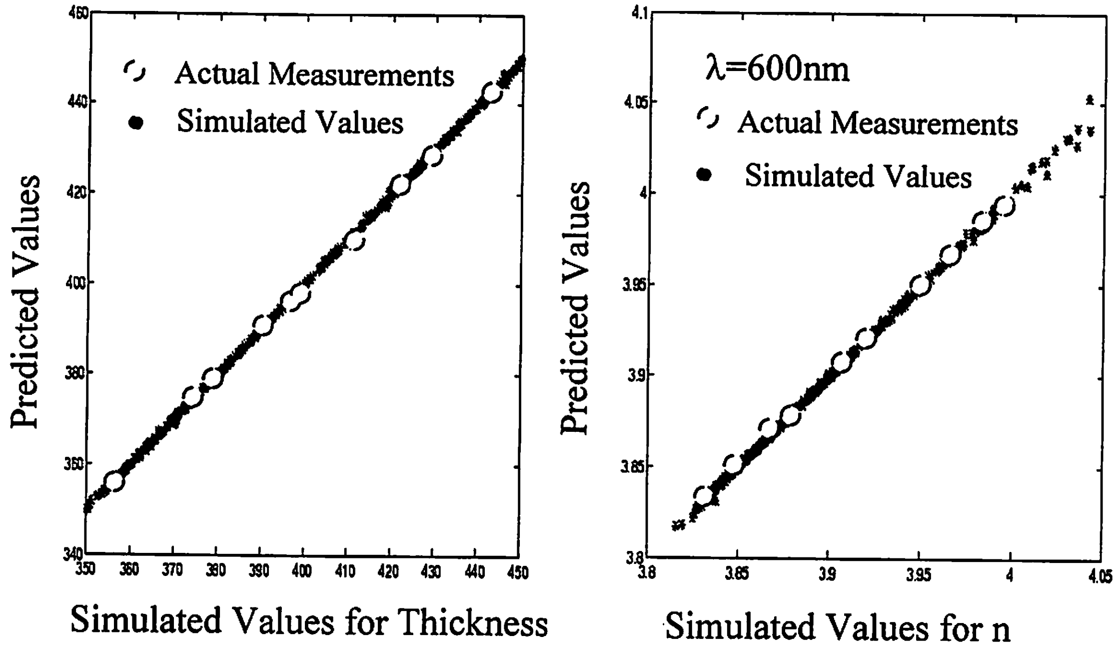
A radial basis function neural network architecture was used due to its well proven functional approximation prowess [?5?]. The inputs to the network were the normalized outputs of the physical filter while the outputs, during the training stage were the optical constants used to generate the simulated reflectance spectra. The design parameter of the network was the spread of the Gaussian functions. We used a network that used a single spread and hence the need to normalize the outputs of the physical filter.

The 1000 inputs were divided into two blocks. One block of 600 was used for training and the other block of 400 was used for testing. An automated routine was written in Matlab [Appendix A] to pick the value of the spread that minimized the error of the testing samples. The values of the other design parameters were kept fixed at their optimum values.

4.6. Results

The results of this optimization are shown in Fig 4.3. The figure shows the predicted values of thickness versus the simulated values as well as the predicted values of the real part of the refractive index versus the simulated values at 600 nm. We chose to use this wavelength because most of the available data on polysilicon refractive indices in the literature is found at this wavelength. At $\lambda = 600$ nm, the extinction coefficient k is zero and was hence not predicted here. As can be seen from the figure, the prediction capabilities of the neural network were excellent. However, the main goal of using the neural network based optimization routine was to cut down on the computation time. This approach reduced the computation time on a SUN-SPARC 20 down from 1 minute to less than 1 second. This now made it possible to use this algorithm for real time computation of the optical constants from broadband reflectance spectra. The training and testing phase took close to

1 hour on a SUN-SPARC 20. However, it is important to note that the ASA extraction and the neural network training and testing are both one time tasks and can be done off-line.



Parameter	Range	Values	σ
Thickness	100 nm	350 - 450 nm	0.300 nm
n(600nm)	0.25	3.800 - 4.050	0.0012

Figure 4.3. Performance of the NN-ASA algorithm for poly-Si on native oxide on Si stack.

4.7. Volume Shrinkage in Chemically Amplified Resist Systems

Chemically Amplified Resists (CARs) are typically composed of a polymer resin which is very soluble in an aqueous base developer, a protecting t-BOC group causing a very slowly dissolving polymer, photo-acid generators and possibly some dyes and additives along with the casting solvent. The deprotection mechanism can be broken down into the initiation, the deprotection and the quenching stages. In the initiation phase, the exposure energy causes the Photo-Acid Generator (PAG) to produce acid. In the deprotection phase, these H+ ions attack the side chains (t-BOC) of the polymer and generate more H+ ions, thus making the resist even more soluble (Fig 4.4). This takes place in the presence of heat. In the quenching stage, the H+ ions are slowly quenched by anything more basic than the acid, such as the additives and the by-products of the reaction. The cleaved t-BOC is volatile and evaporates, causing film shrinkage in the exposed areas. The extent of this exposed photoresist thinning is dependent on the molecular weight of the blocking groups.

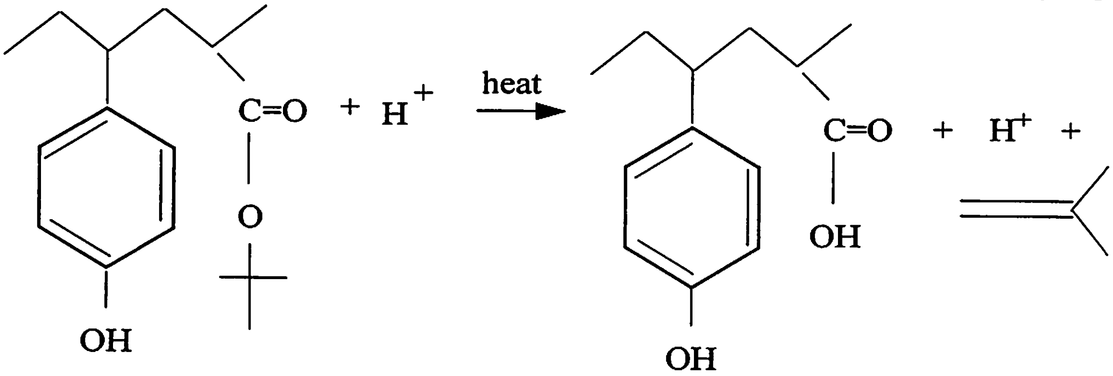


Figure 4.4. DUV Chemically Amplified Resist Mechanism during the Exposure and Post Exposure Bake Steps.

An experiment was performed to correlate the deprotection of the resist as measured by a Fourier Transform Infrared Spectroscopy tool to the observed volume shrinkage (Fig 4.5). The results showed that the intrinsic reaction mechanism occurring in the resist during the bake could be observed through the volume shrinkage.

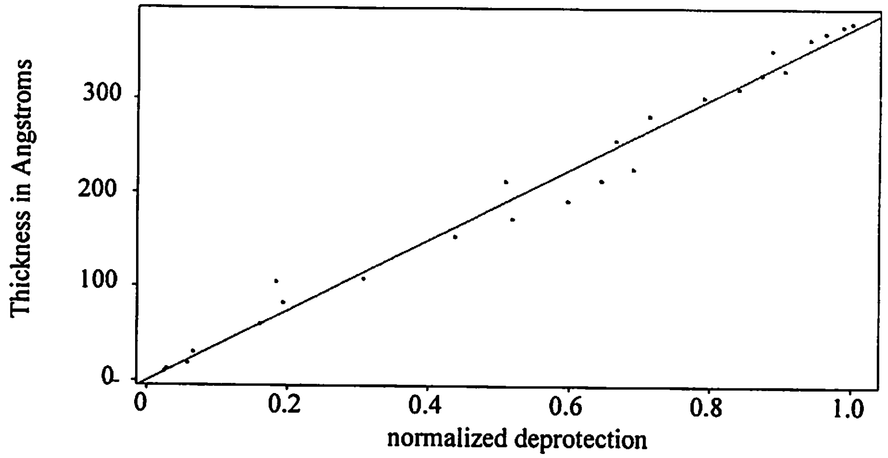


Figure 4.5. Thickness loss as a function of the deprotection, measured by monitoring the normalized ester absorbance.

Next we study this shrinkage and attempt to understand the reaction mechanism occurring in the resist. This leads to the static and dynamic models described in the subsequent sections.

4.8. Static Model for Thickness Loss

Static models are useful in cases when only a single macroscopic output is measurable practically, rather than detailed 1-, 2- or 3-dimensional measurements of a phenomenon. In this example, deprotection is easily observed through thickness loss measurements, whereas 1-d measurements of the deprotection through depth into the resist would be practically impossible. The deprotection induced thickness loss at different doses can be used in conjunction with a static model of the PEB process to extract relevant simulation parameters such as the Dill's *C* parameter, the relative quencher concentration [*Q*], the amplification reaction rate (E_{amp}, A_{amp}), etc.

Currently there exists no model for the bake process that can account for the commonly observed initial delay in the increase of the deprotection vs. dose at different temperatures. In this section, a novel model for the PEB process is derived.

The effective exposure dose is first calculated by accounting for the reflectivity at the air-resist interface and this is converted into acid as in Eq. (4.6).

$$[Acid]_{dose} = [PAG]_0\left(1 - e^{-C \times Dose}\right) \qquad (4.6)$$

where *Dose* is the effective exposure dose, calculated by accounting for the reflectivity at the air-resist interface, $[PAG]_0$ is the initial concentration of the photoacid generator and *C* is the rate of photoacid formation in cm^2/mJ.

During the PEB process, the t-BOC blocked polymer undergoes acidolysis to generate the soluble hydroxyl group in the presence of acid and heat. The conventional modeling of the PEB process [3] is given as

$$m = e^{-k_{amp}\left(\frac{1 - e^{-k_{loss}t}}{k_{loss}}\right)[Acid]_{dose}} \qquad (4.7)$$

where m is the normalized concentration of unreacted blocking sites, k_{amp} is the acid amplification rate in sec^{-1} and k_{loss} is the acid loss factor in sec^{-1}. Both these factors are modeled using a temperature-dependent Arrhenius relationship.

However, the combined exposure and PEB models Eq. (4.7) do not account for the commonly observed initial delay in the increase of the deprotection vs. dose at different temperatures. We think that this is due to the quencher designed into most chemically amplified resists. Thus, we propose a corrected model for the PEB to account for the relative quencher concentration $[Q]$. In this model, we assume that during the PEB process, acid is lost in neutralization reactions with bases that are either designed into the resist, or exist as unreacted portions of the polymer. This indicates that the bases will also correspondingly reduce with time, and the difference between the acid and base concentrations will remain constant throughout the PEB process. We model the above mechanism through Eq. (4.8) and Eq. (4.9).

$$\frac{\partial}{\partial t}[Acid] = -\alpha[Acid][Q] \qquad (4.8)$$

$$[Acid]_t - [Q]_t = [Acid]_0 - [Q]_0 \qquad (4.9)$$

where α is the neutralization reaction coefficient modeled by an Arrhenius function of temperature, $[Q]$ is the relative quencher concentration (relative to $[PAG]_0$) and $[Acid]$ is the acid concentration as defined in (2). The initial value for quencher $[Q]_0$ is a parameter that can be extracted from the fitting procedure described in the following section, while the initial acid concentration $[Acid]_0$ is obtained from Eq. (4.6). Solving the above equations yields the following analytical solution Eq. (4.10) for the acid concentration as a function of the PEB time, t.

$$[Acid]_t = \frac{([Acid]_0 - [Q]_0)}{1 - \frac{[Q]_0}{[Acid]_0}\exp(-\alpha([Acid]_0 - [Q]_0)t)} \qquad (4.10)$$

Meanwhile, the deprotection reaction is typically modeled by Eq. (4.11), where k_{amp} is the reaction amplification rate in sec^{-1}, modeled by an Arrhenius temperature relationship, and $[M]$ is the concentration of protected sites remaining at time t.

$$\frac{\partial}{\partial t}[M] = -k_{amp}[Acid]_t[M] \tag{4.11}$$

Substituting Eq. (4.10) in Eq. (4.11), and solving for the normalized m (normalized concentration of unreacted blocking sites), we get

$$\frac{[M]_t}{[M]_0} = m = \left[\frac{[Acid]_0 e^{(\alpha([Acid]_0 - [Q]_0)t)} - [Q]_0}{[Acid]_0 - [Q]_0}\right]^{-\frac{k_{amp}}{\alpha}} \tag{4.12}$$

The model depicted in Eq. (4.12) differs from previous work in that it accounts for the fact that the quenchers (both parasitic and designed) are consumed in the neutralization reaction. This allows better modeling of the initial delay in deprotection increase with exposure dose than existing models, and hence provides an estimate of the relative quencher concentration $[Q]_0$.

4.9. Dynamic Model for Thickness Loss

4.9.1. Physical Models

One of the underlying assumptions in modeling the latent image through continuity equations has been that the resist volume remains constant. Ignoring the volume shrinkage obviously affects the accuracy of existing models. The goal of this work is to describe the physical processes occurring in the resist during the PEB step for 1-dimensional (flood) exposures, in the presence of the volume shrinkage. We begin with a description of the physical mechanisms occurring in the process, and then provide mathematical equations to represent these physical processes.

We propose the following mechanism: During the exposure step, the photo-acid generators produce acid on reaction with photons. This is represented as a normalized acid concentration (u), normalized to the initial photoacid generator concentration. The initial 1-d distribution of the acid within the resist depends on the optical constants of the resist and of the underlying film at the exposure wavelength. With sufficient energy (in the form of temperature), the acid molecules begin to diffuse around (D_u) and attack the polymer side-chains at a certain rate (k_2), which has an Arrhenius relationship to temperature, thus causing deprotection (v). The acid diffusion is widely believed to be free-volume dependent, and is modeled using an exponential (D_{u0},a) [4]. The deprotected molecule (w) is volatile and begins to diffuse (D_w) through the resist till it escapes the resist bulk from the top. The amount of molecules that escape would depend on the partial pressure in the wafer track. In the meanwhile, the escaped volatile molecule leaves behind a vacancy or hole (h) that begins to collapse at a rate specific to the particular polymer (k_3). This is the polymer relaxation process that eventually causes the volume shrinkage in the resist (ϖ). Deprotection refers to a normalized quantity between 0 and 1, and hence scaling factors are needed to convert the deprotection into corresponding volatile group concentrations (k_1). Similarly, a scaling factor is needed to convert the hole concentration into a corresponding volume shrinkage factor (k_4) to take into account the area of the flood-exposed site. In the case of low activation energy resist systems, this mechanism begins during the exposure step itself, while in the case of high activation energy systems, this process begins to occur only during the PEB step.

The mechanism described above can be described with the following six equations:

$$\frac{\partial}{\partial t}(u\varpi) = \nabla \bullet (D_u \nabla u)\varpi \tag{4.13}$$

$$D_u = D_{u0}\exp(ah) \tag{4.14}$$

$$\frac{\partial v}{\partial t} = k_2 u(1-v) \tag{4.15}$$

$$\frac{\partial}{\partial t}(w\varpi) = [D_w \nabla^2 w + k_1 k_2 u(1-v)]\varpi \tag{4.16}$$

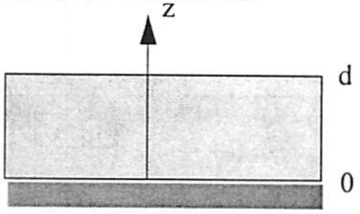
$$\frac{\partial}{\partial t}(h\varpi) = [-D_w \nabla^2 w - k_3 h]\varpi \tag{4.17}$$

$$\frac{\partial \varpi}{\partial t} = -k_4 k_3 h \varpi \tag{4.18}$$

Eq. (4.13) suggests that the rate of change of acid at any point in the resist is governed by the non-linear acid diffusion within the resist caused by the acid gradient that exists at that point. This gradient in turn exists due to changing exposure conditions in the film caused by internal reflectance of light and absorbance of light by the resist film during exposure. The acid diffusion is modeled by Eq. (4.14). Eq. (4.15) represents the rate of normalized deprotection reaction and is proportional to the amount of acid and the amount of unreacted sites. Eq. (4.16) indicates that the rate of change of volatile groups within an infinitesimal volume element is proportional to the number of volatile groups diffusing through the volume element and the generation rate is proportional to the normalized deprotection rate. Eq. (4.17) represents the rate of change of holes (or free volume). The generation rate of holes within any infinitesimal volume element depends on the number of volatile groups diffusing out of that volume element while the destruction rate of holes within that same volume element is dependent on the polymer relaxation rate constant. Eq. (4.18) models the volume shrinkage within each volume element and is proportional to the free volume relaxation rate. The ϖ term is included in all the equations that deal with concentrations, to account for the changing volume at each time instant.

4.9.2. Boundary Conditions

The equations defined above do not have a closed form analytical solution due to the inter-dependencies of the various differential equations on one another. Hence they must be solved numerically, subject to the boundary conditions for this problem. They would be as follows:



$$\left.\frac{\partial u}{\partial z}\right|_{z=d} = 0 \tag{4.19}$$

$$\left.\frac{\partial w}{\partial z}\right|_{z=0} = 0 \ ; \quad w|_{z=d} = 0 \tag{4.20}$$

$$u|_{t=0} = u_0(z) \tag{4.21}$$

$$u_0(z) = 1 - \exp(-CD(z)) \tag{4.22}$$

$$D(z) = D_0\left[\exp(-\alpha z) + |r|^2\exp(-\alpha(2d-z)) - 2|r|\exp(-\alpha d)\cos\left(\frac{4\pi n}{\lambda}(d-z)\right)\right] \quad (4.23)$$

where D_0 is the applied dose in mJ/cm2, corrected by the reflectivity at the air-resist interface, α is the linear absorbance of the resist film in μm^{-1}, d is the film thickness in μm, n is the real part of the refractive index, λ is the exposure wavelength in μm, C is the acid production rate in cm^2/mJ and r is the reflectivity coefficient of the resist/substrate interface. This exposure process in chemically amplified resists has been modeled by Byers, et. al. [3] using a simplified version of the full wave equation solution.

Eq. (4.19) indicates that there is no acid loss due to evaporation at the resist surface. However, this condition can easily be modified to model T-topping or environmental contamination. Eq. (4.20) indicates that the volatile group escapes from the resist only at the resist-air interface, and this gradient is facilitated by maintaining the volatile group concentration at the resist-air interface at zero. Eq. (4.21) states that the initial acid distribution during the exposure is determined by the aerial image and the optical properties of the resist.

$$h|_{t=0} = h_0 \quad (4.24)$$

$$v|_{t=0} = w|_{t=0} = 0 \quad (4.25)$$

Eq. (4.24) refers to the presence of an initial free volume concentration in the resist which is a function of the spin-on and soft bake process [5]. Eq. (4.25) states that the deprotection and volatile group concentrations at the beginning of the PEB process are zero (for low activation energy resist systems, this would be a value greater than zero).

4.9.3. Computational Approach

A finite difference system was set up for the equations and boundary conditions described above. A simple forward difference technique (explicit method) was used to progress the equations in time with sufficient time and space steps to avoid numerical instability problems [6].

We divide the one dimensional space domain (z direction, which is perpendicular to the surface of the resist) and the time domain into many discrete steps, and use the following formulae to approximate the divergence in one dimensional case and the time derivative of a function numerically

$$\nabla u(z,t) = \frac{\partial}{\partial z}u(z,t) \approx \frac{1}{2h}[u(z+h,t) - u(z-h,t)] \tag{4.26}$$

$$\nabla 2u(z,t) = \frac{\partial^2}{\partial z^2}u(z,t) \approx \frac{1}{h^2}[u(z+h,t) - 2u(z,t) + u(z-h,t)] \tag{4.27}$$

$$\frac{\partial}{\partial t}u(z,t+dt) \approx \frac{u(z,t+dt) - u(z,t)}{dt} \tag{4.28}$$

After doing calculations for each time step, the size of each of the grids decreases due to the volume shrinkage in each element. At the end of each time step, we adjust the grid sizes to their original sizes h, and interpolate with the spline method [6] to get the function values at the new grid points. A indication variable is used to record the change of the total thickness as simulation progresses, and adjust the total number of grids accordingly, as shown in Fig 4.6

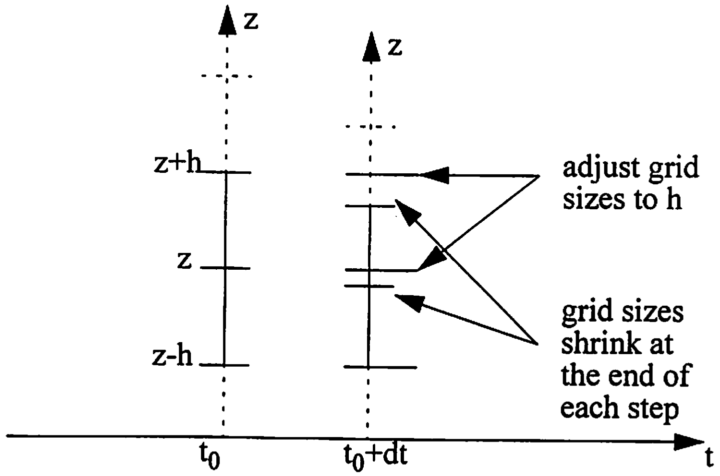


Figure 4.6. Simulation Approach for the Moving Boundary Problem.

We used 200 space steps (1 step = 3.25 nm) at the beginning of the simulation and 5000 time steps (1 step = 16 ms) in our computation, yielding 10 seconds for simulating one thickness loss versus time set for one dose, on a 350 MHz Pentium-II processor. An implicit method such as the Crank-Nicholson method Eq.(6) would have allowed for fewer time and space steps but would require matrix computations to solve simultaneous equations.

4.10. Optimization Framework

The block diagram for the optimization process is shown in Fig 4.7. The optimization was carried out for both the static as well as the dynamic models. The effective acid in the resist is calculated using Eq. (4.23). This initial acid distribution is fed along with the first set of resist parameters generated by the optimization engine, to the 1-d volume shrinkage simulator. The output of the simulator is compared with the experimental data, the resulting error is fed to SA and a new set of parameters is generated. This process continues until the sum squared error between the model prediction and the experimental data reduces below a pre-determined threshold value.

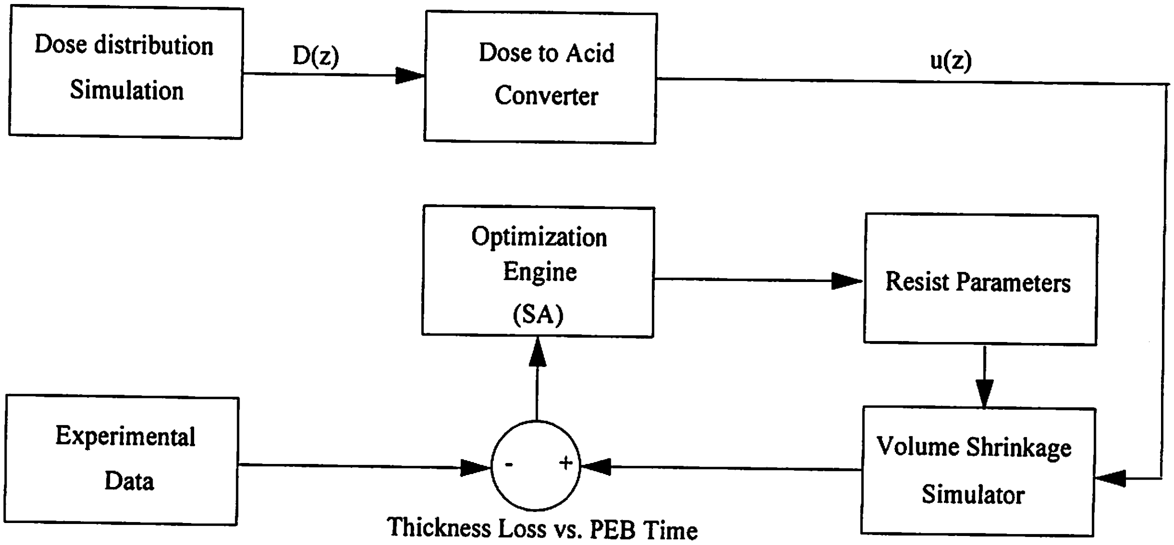


Figure 4.7. Block diagram for resist parameter extraction using the dynamic model and experimental data.

4.11. Results and Discussion

We begin with results of the static model fitting to the experimental data for UV-5, AZ 2549. Fig 4.8 shows the results of fitting the static model to experimental data for UV-5 while Fig 4.9 shows the results of fitting the static model to experimental data for AZ-2549. The results for the static cases had an error less than 0.01 deprotection units, 1-σ.

The results for the dynamic model are presented next. Fig 4.10 shows intermediate results for the acid concentration, deprotection, volatile group concentration and free volume concentration as a function of depth at 5 different time steps. Fig 4.11 shows the results for UV-5 while Fig 4.12 shows the results for AZ 2549. The results of fitting to the dynamic model seem better for UV-5 than for AZ 2549. One possible reason is that in the case of UV-5, the measurements were made off-line, thus allowing for a more accurate measurement. In the case of AZ 2549, the setup did not allow for very accurate placement of the reflectometer spot size within the exposed areas. This caused measurements that were made on the boundary of unexposed and exposed areas, thus providing inaccurate results for the thickness loss. In some cases, it took between 4-5 seconds to correct this misalignment of the spot, thus causing an offset between the measured and simulated results (as seen in Fig 4.12).

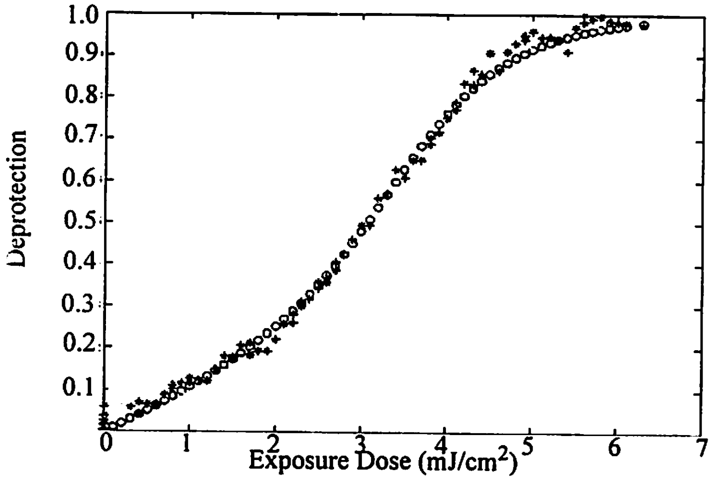


Figure 4.8. Static model fitting experimental data for UV-5. Circles denote simulated results while stars denote experimental data.

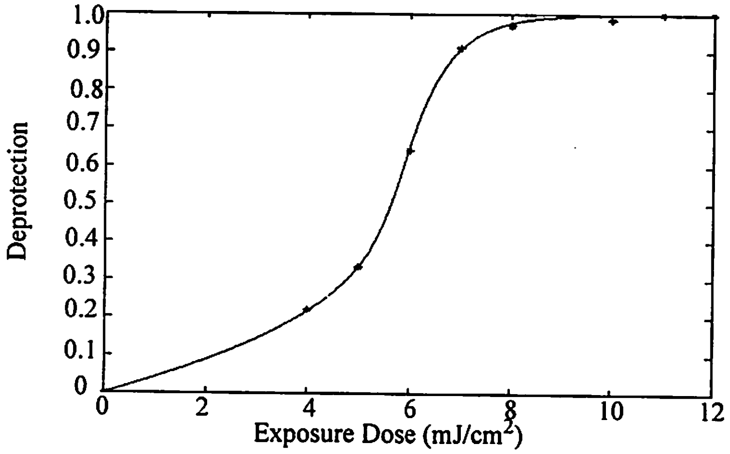


Figure 4.9. Static Model fitting experimental data for AZ-2549. Bold line denotes simulation while stars denote experimental data.

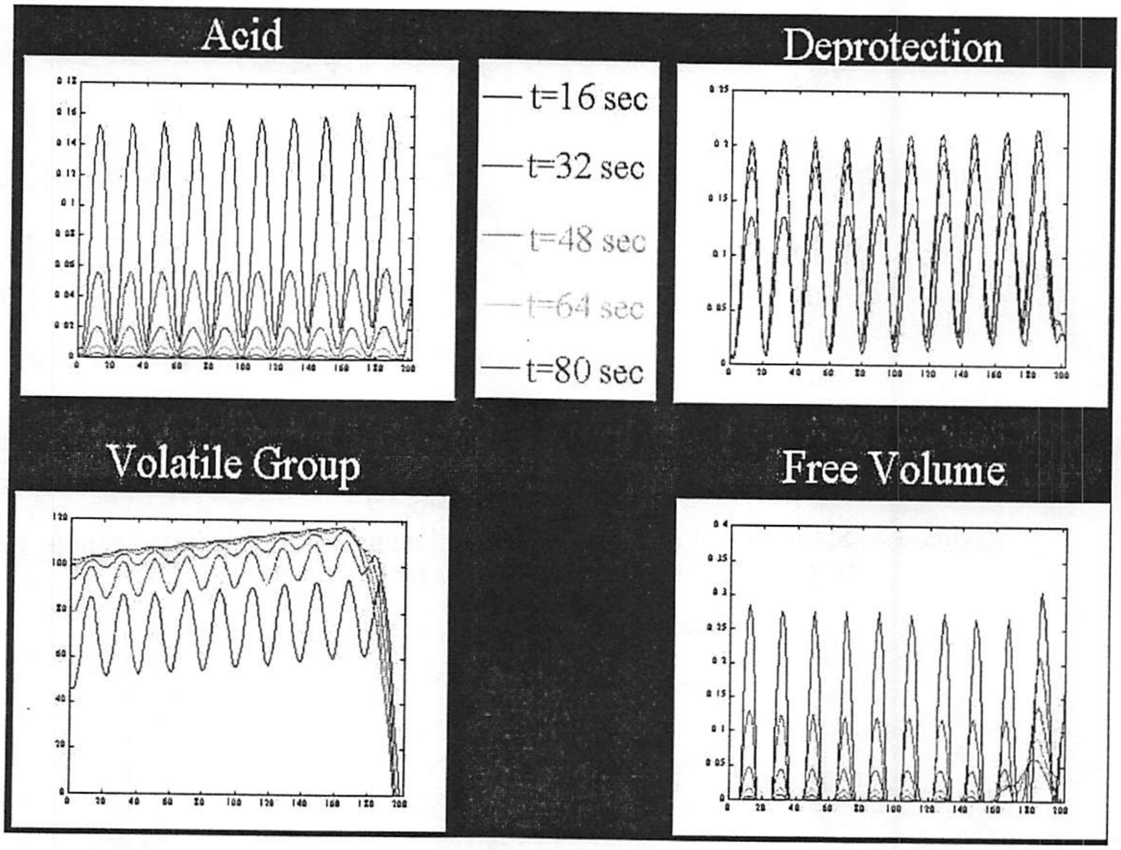


Figure 4.10. Intermediate results for acid concentration, deprotection, volatile group concentration and free volume concentration as a function of depth at 5 different time steps

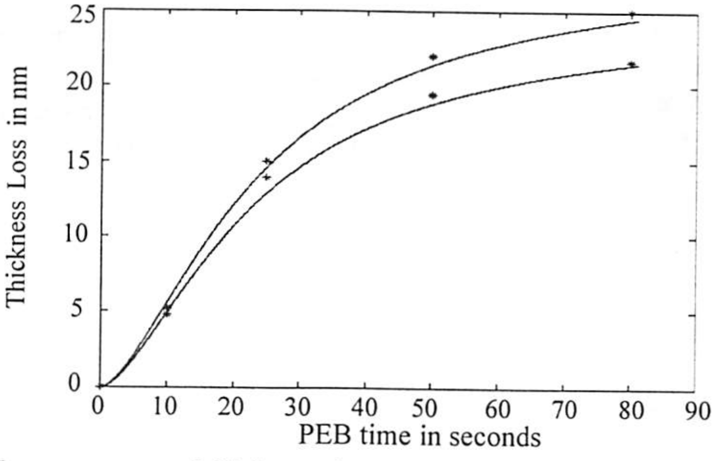


Figure 4.11. Measurements of thickness loss versus time for UV-5. Simulated loss in bold; measured loss in star.

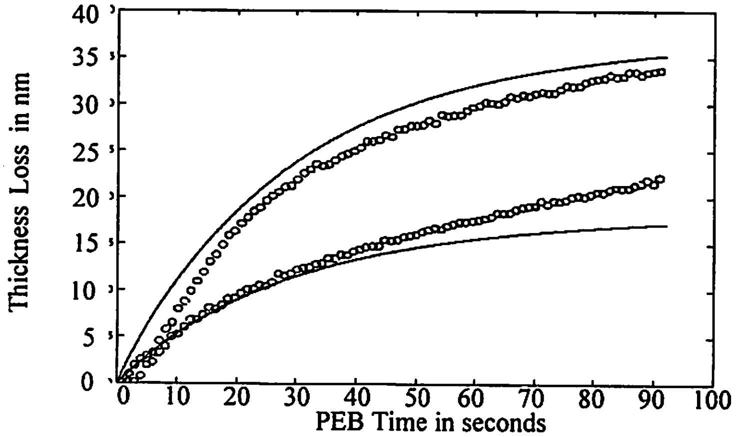


Figure 4.12. Measurements of thickness loss versus time for AZ 2549.

We tabulate the results for UV-5 and AZ-2549 below.

TABLE 1. Results of Parameter Extraction

Parameter	UV-5	AZ-2549
K_{amp}	0.3406	0.3673
K_{α}	1.16	2.06
C	.059	0.0798
Q	.177	0.367

4.12. Summary

In this paper we have proposed both dynamic and static physical models for volume shrinkage in chemically amplified resists. The proposed dynamic model successfully predicts the volume shrinkage observed in resists and could be used to gain insight into the resist mechanism. The static model successfully models the quenching action in the resist across the complete exposure dose spectrum and in the process, extracts critical resist parameters useful for "what-if" analysis in lithography simulation.

Chapter 5 Lithography Simulation

5.1. Introduction

Traditionally, lithography process development has relied on short-loop and pilot-lot experiments to understand the effects of particular process control factors. However, high experimental costs and increasing complexity of lithographic patterns and processes is such that one must resort to simulation. Technology Computer-Aided Design (TCAD), focusing on predictive simulation, is becoming very important for lithography process development and control. An efficient development process would reduce the number of characterization experiments devoted to developing a new recipe, will reduce time-to-market and will drastically cut development costs.

Technology Computer-Aided Design (TCAD) tools are playing an important role in the design and manufacturing of ICs. As the cost of computation decreases and the cost of experimentation and equipment increases, TCAD tools are becoming essential cost effective alternatives. Sophisticated TCAD packages that simulate the entire lithography process include SAMPLE [35], PROLITH[1], SOLID-C[2], etc.

Developing an effective TCAD simulation environment depends on accurate process models, as well as on correct model parameters. Simulators contain a set of differential equations that attempt to model real systems. These equations are physical or chemical in nature, and are usually derived from first principles. While most of the DUV lithography process steps are fairly accurately modeled, certain critical process steps, such as post-

---

[1] FINLE Technologies, P.O. Box 162712, Austin, TX 78716

[2] Sigma-C GmbH, 901 Campisi Way, #248, Campbell, CA 95008

exposure bake (PEB), still suffer from modeling inadequacies. This problem has been addressed in the earlier chapter.

All the models require the use of constants. These may be parameters that represent either the chemical or physical constants of the phenomena involved. There are also constants that are empirical in nature. While some parameters are well known, many of them are not known accurately, so a lot of experimentation is devoted to extracting their values. However, the large number of the unknown parameters, and the non-linear nature of the models, renders traditional optimization techniques, such as steepest descent, useless for parameter extraction. Manual optimization procedures, wherein the parameters of the model are changed one at a time to fit experimental data, are erroneous, because they neglect interaction effects between the different parameters. Thus, process simulation engineers spend a very large amount of time to calibrate a model and after that, only use the model to study general process trends due to the lack of confidence in its results.

In this chapter, an efficient methodology is proposed for extraction of information from standard unpatterned and patterned resist characterization experiments, to be ultimately used for the calibration of lithography simulation tools.

5.2. Simulator Calibration Framework

5.2.1. Process Models

Accurate process models for the lithography sequence have been developed in chapter 4. Here, the updated models for the exposure, PEB and develop step are presented together for completeness.

5.2.1.1. Thin Film Interference

$$Dose(z) = Dose(0) \times \left(e^{-\gamma z} + |r|^2 e^{-\gamma(2d-z)} - 2|r| e^{-\gamma d} \cos\left(\frac{4\pi n(d-z)}{\lambda}\right)\right) \qquad (5.1)$$

where the $Dose(0)$ is the applied dose in mJ/cm^2, corrected by the reflectivity at the air-resist interface, z is the depth into the resist in µm, γ is the attenuation within the resist film in µm^{-1}, d is the film thickness in µm, n is the real part of the refractive index, λ is the exposure wavelength in µm, and r is the reflectivity coefficient of the resist/substrate interface.

5.2.1.2. Exposure & PEB Process Model

$$m = \left[\frac{[Acid]_0 e^{(\alpha([Acid]_0 - [Q]_0)t)} - [Q]_0}{[Acid]_0 - [Q]_0}\right]^{-\frac{k_{amp}}{\alpha}} \tag{5.2}$$

where m is the normalized concentration of unreacted blocking sites, α is the neutralization reaction coefficient modeled by an Arrhenius function of temperature, $[Q]_0$ is the relative quencher concentration (relative to $[PAG]_0$), k_{amp} is the reaction amplification rate in sec^{-1}, modeled by an Arrhenius temperature relationship and $[Acid]_0$ is the acid concentration calculated as:

$$[Acid]_0 = \left(1 - e^{-C \times Dose(z)}\right) \tag{5.3}$$

5.2.1.3. Develop Process Model

$$R(m) = R_{max}\frac{(A+1)(1-m)^n}{A+(1-m)^n} + R_{min} \tag{5.4}$$

$$A = \frac{(n+1)}{(n-1)}(1-m_{th})^n \tag{5.5}$$

where R_{max} is the maximum development rate, R_{min} is the minimum development rate, m_{th} is the value of m at the inflection point of the data, called the threshold PAC concentration, and n is the dissolution selectivity parameter, which controls the contrast of the photoresist.

5.2.2. Estimating the values of the model parameters

Some of the coefficients in the process models represent either the chemical or physical characteristics of the materials and equipment involved. Others are empirical in nature. One needs accurate estimates of the values of these process and equipment parameters in order to do predictive simulation. In this study, we classify these parameters into three categories, depending on the accuracy with which they are initially known.

The first category (Category I) consists of parameters that are obtained through direct, well recognized measurements, such as the numerical aperture and partial coherence of a projection system or the refractive index and thickness of thin films. The second category (Category II) consists of parameters whose values are extracted by running unpatterned wafer experiments, such as the DRM for development related parameters (maximum and minimum development rate, resist contrast, etc.) and fourier transform infrared spectroscopy (FTIR) for PEB related parameters [36] (amplification rate coefficient, acid loss rate, etc.). These parameters are extracted by fitting the physical models described in the preceding section to the raw data. This approach provides parameter values that are subject to both experimental error as well as model-bias errors, due to model inaccuracies. Hence, the values of the parameters in this category are not as well known as those in the first category.

Finally, the third category (Category III) consists of empirical parameters such as the focus offset between the stepper focal plane and the simulator focal plane. In the same category are physical parameters that are very difficult to obtain experimentally, such as the diffusion coefficient and the amplification reaction order in chemically amplified resist systems.

While the parameters in the first category are fixed at known values, those in the second and third categories are obtained through optimization. The category II, or the parameters subject to experimental and modeling errors, go through a two step optimization process. In the first step, rough estimates of these parameters are obtained by fitting physical models to experimental data from unpatterned resist experiments. The second step is to fine-tune these parameters by fitting to cross-sectional resist profiles available from patterned resist experiments. This second step is also used to tune the category III parameters that must be inferred indirectly.

The following are examples of important DUV lithography process and equipment parameters that are listed by category, as defined above.

Category I: Well Known Parameter Values.

Resist Refractive Index (Real part)
Resist Refractive Index (Imaginary part)
ARC Refractive Index (Real part)
ARC Refractive Index (Imaginary part)

Category II: Parameters subject to experimental and modeling errors

Amplification Rate (Pre-exponent)
Amplification Rate (Activation Energy)
Neutralization Rate (Pre-exponent)
Neutralization Rate (Activation Energy)
Dill's A parameter
Dill's B parameter
Dill's C Parameter
Relative Quencher Concentration
Maximum Develop Rate
Minimum Develop Rate
Developer Selectivity
Developer threshold PAC

Category III: Parameters that must be inferred indirectly

PEB Diffusivity (Pre-exponent)
PEB Diffusivity (Activation Energy)
Focus offset between stepper and simulator

5.2.3. Experimental Methodology for Parameter Extraction through Profile Matching

As mentioned earlier, there is a need to cut time to market and to reduce high development costs. This calls for a more efficient development process that reduces the number of unpatterned and patterned characterization experiments devoted to extracting process parameters, so as to gain maximum possible information from minimum amount of experimentation. There have been numerous experimental procedures proposed in the literature for this purpose, as shown in the previous section.

Values of thin film optical constants, such as refractive indices and thicknesses of organic and inorganic materials, are obtained very accurately by using spectroscopic ellipsometry. This technique has the ability to provide the optical constants for each individual film in a multi-layer stack. Measurement of resist parameters, such as the photospeed, amplification rate, acid loss rate and normalized acid quencher concentration are typically obtained through either FTIR or Deprotection Induced Thickness Loss (DITL) [36] studies. In these techniques, wafers are subjected to multiple exposure and thermal doses, and the deprotection or volume shrinkage is observed in flood-exposed areas. While an exposure meander (a serpentine trail of incrementally increasing exposure doses) is done on every wafer, the number of wafers required is the same as the number of PEB temperatures used. Measurement of the develop rate parameters, such as the maximum and minimum develop rates, developer selectivity and developer threshold, is achieved using either the standard DRM or the Poor Man's DRM [37]. In the former technique each wafer is exposed with a different exposure dose, and the resist thickness is monitored in real time during the development process. This data is then converted into develop rate vs. deprotection and is used for extracting develop rate parameters, such as the minimum and maximum development rate, developer selectivity and the deprotection threshold. In the latter technique, each wafer is processed with an exposure meander and is developed at different development times. The remaining resist thickness is measured at each exposure dose and development time after the development process, and is converted into develop rate vs. deprotection and used for extracting development parameters.

The standard patterned characterization experiment involves processing a focus-exposure matrix (FEM) on each wafer. In this technique, wafers are exposed with a FEM and the profile is then measured in each die with either a cross-section scanning electron microscope (SEM) or an atomic force microscope (AFM). Figure 5.1 illustrates the experimental flow described in this section.

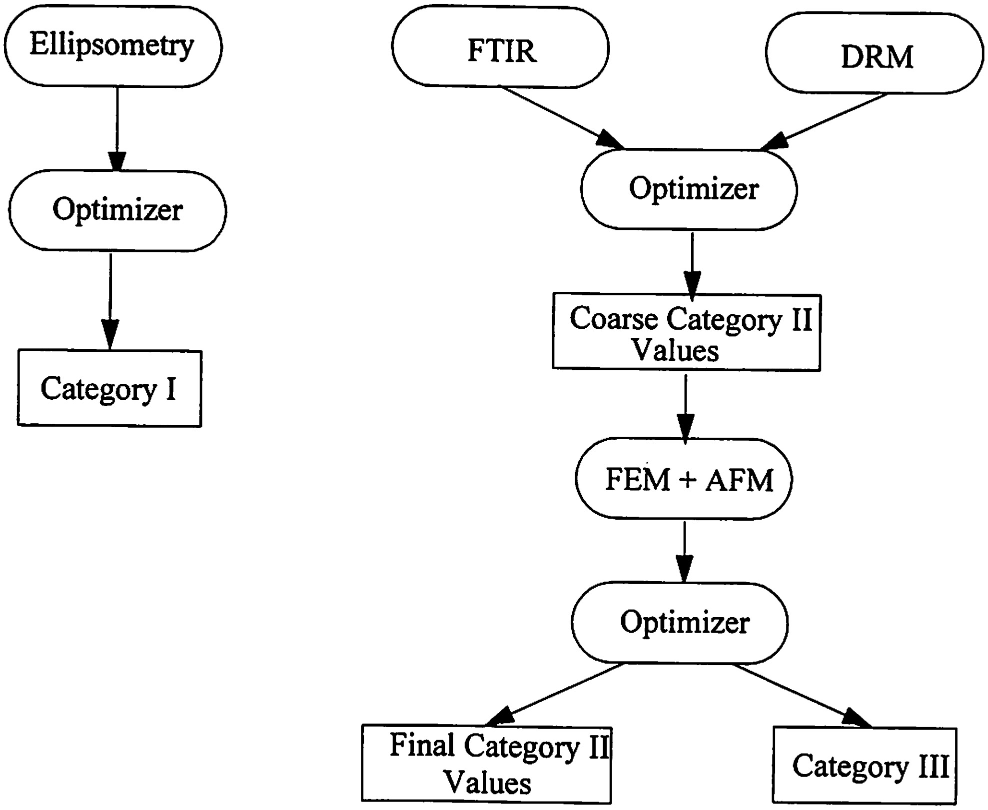


Figure 5.1. Experimental flow for parameter extraction.

5.2.4. Formulating an optimization problem for estimating the unknown variables

Simulated Annealing was used as the optimization engine in this problem. For continuous variables, the continuity has to be taken into account by suitably selected discretization steps: very small steps result in an incomplete exploration of the variation domain with small and frequent function improvements; very large steps may produce too many unacceptable function variations. This particular problem optimized a continuous variable space. Hence, we chose to use step sizes that were between 0.001% and 1% of the initial parameter values, depending on how sensitive the simulation was to the particular parameters.

Extensive work has been done in extending simulated annealing ideas from combinatorial problems to continuous functions [38], [39], etc. In this work, we use a commercial package as the optimization kernel[3].

5.2.5. Framework for Parameter Extraction through Profile Matching

In this section, a methodology is proposed that ties the process models, parameter categories, experimental data and the optimization technique together to form an efficient simulator calibration framework. The framework employs two steps. In the first step, the unpatterned characterization experiments are carried out. The process engineer then uses the optimization routine to tune the parameters in the unpatterned process models based on the experimental results. This step yields the value of the parameters subject to experimental and modeling errors (Figure 5.2).

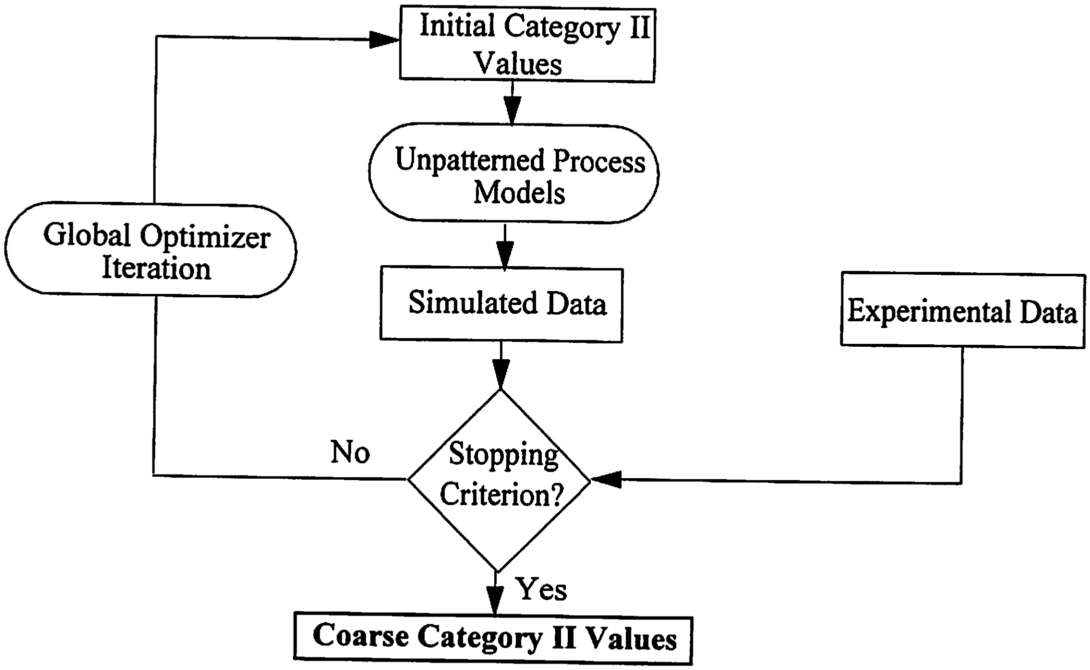


Figure 5.2. Finding initial estimates for category II parameters (parameters subject to experimental and modeling errors)

In the second step, the process engineer runs a FEM and measures the cross-sectional profiles using either an AFM or a cross-section SEM. In the former case, the profile is already in ascii format and can be compared directly to the simulator output, while in the

---

[3] Timbre Technology, Inc., 2000 Walnut Ave., #H-103, Fremont, CA 94538.

latter case the SEM profile is in JPEG format and thus needs to be converted to ascii format before it can be compared to the simulator output.

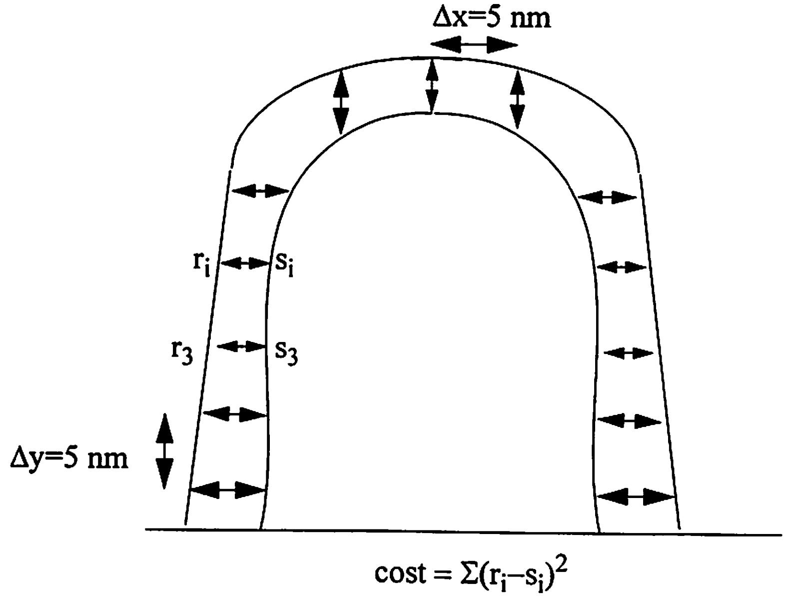


Figure 5.3. Computing the sum-squared error between simulated and digitized SEM images

This conversion process is done using a standard graphics package, such as Adobe Photoshop. The process engineer assigns ranges for the parameters subject to experimental and modeling errors (category II) and the parameters that must be inferred indirectly (category III), while fixing the values for the well known parameter values (category I). The ranges for the parameters depend on the confidence the process engineer has in their accuracy; this would typically depend on the accuracy of the metrology associated with the unpatterned experiments. The optimization routine is now run, varying the parameters subject to experimental and modeling errors and the parameters that must be inferred indirectly, until the simulated profile sufficiently matches the experimental profile (Figure 5.4).

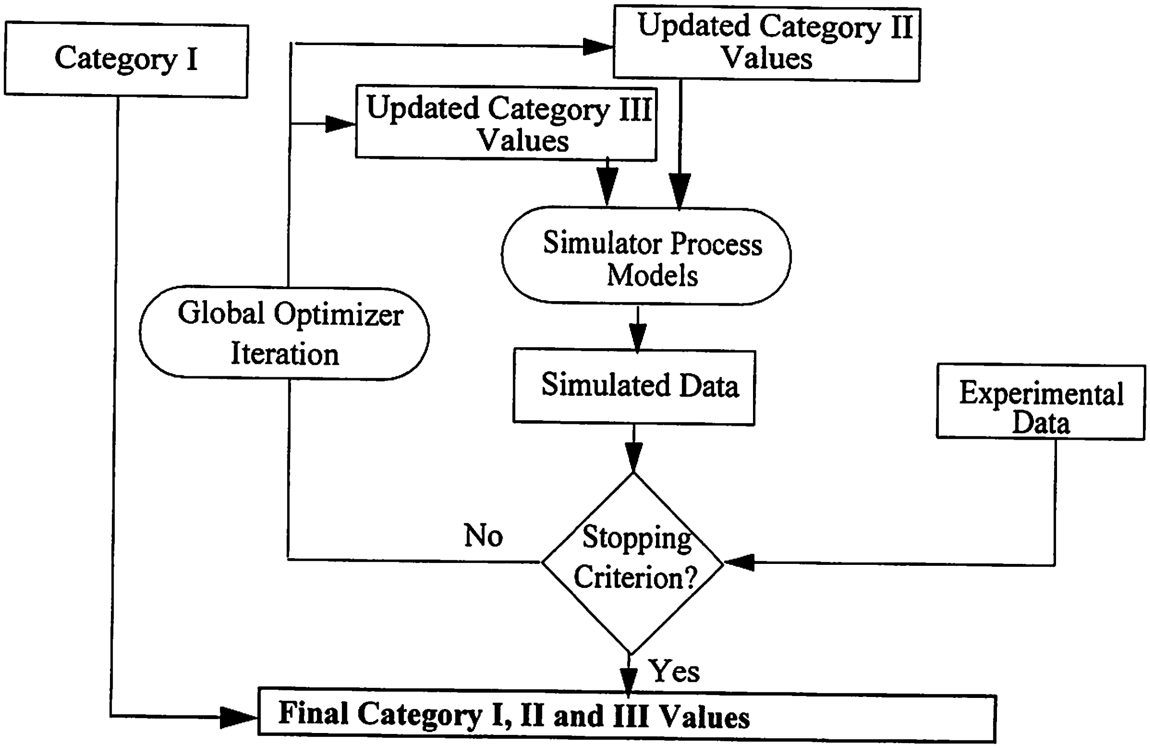


Figure 5.4. Flowchart of unpatterned characterization experiments (Update Category II and III).

The stopping criterion used is the R^2 value.

$$R^2 = 1 - \sum_i \frac{(y_{meas,i} - y_{theoretical,i})^2}{(y_{meas,i} - \bar{y}_{meas})^2} \tag{5.6}$$

where i is the number of steps the height of the resist profile is divided into, $y_{meas,i}$ is the measured CD at height step i, $y_{theoretical,i}$ is the simulated profile CD at the same height step i and $\bar{y}_{meas}$ is the average of the measured CDs over all height steps.

The parameter values obtained at the end of this phase are the final calibrated parameter values for the simulator and given materials. The larger the range of settings covered by the experimental data, the more global and hence predictive the simulation calibration procedure is. Thus, we now have an efficient method for feeding back experimental profile information to calibrate simulators.

5.3. Experiments for Parameter Extraction through Profile Matching

The experiments utilized flood exposures as well as patterned wafers, done using a FEM at standard resolution patterns. All the wafers were coated with Shipley's UV-5 resist at the standard processing conditions. An FTIR tool and a spectroscopic ellipsometer were used for all the thin film measurements, while a cross-section SEM and an AFM were used for all the profile measurements.

5.3.1. Unpatterned Experiments

We employed Ellipsometry and FTIR spectroscopy to extract the final values for category I parameters and initial values for the category II parameters using the exposure and PEB models shown in (5.2) and (5.4). This involved three wafers that were each flood-exposed with twenty-five different exposure doses in a meander ranging from 0 to 7.6 mJ/cm^2. These wafers were then subjected to 3 different PEB temperatures from 120 to 135 degrees Celsius. Before the exposure step, the anti-reflective coating and the resist were each measured for thickness and optical constants (real and imaginary part of the refractive indices) using a spectroscopic ellipsometer, such as the KLA-Tencor 1250 SE. The thicknesses on all the 25 sites was measured again after the PEB step, thus yielding the DITL [40]. These wafers were then measured with a QS-1200 FT-IR System from Bio-Rad. The deprotection was measured by tracking the ester bond (1150 cm^{-1} peak). This experiment yielded deprotection vs. dose and temperature, as well as raw ellipsometry signals containing information on the thin film optical constants.

The Poor Man's DRM [37] experiment was performed using ten wafers at different exposure doses and development times. This yielded develop induced thickness loss versus exposure doses, which was converted into develop rates versus concentration of unreacted sites using the DITL model [37].

5.3.2. Patterned Experiments

An experiment was carried out on a wafer with different layout designs. There were a total of 10 layouts, each exposed at 2 focus settings that were 0.3 μm apart. The layout geometry is shown in Figure 5.5. There were 7 geometric parameters, $d_0, \ldots, d_6$, defining each layout. The geometric parameters of the 10 layouts are shown in Table 5.1. The wafers were then cleaved and measured with a cross-section SEM. Figure 5.6 shows one of the cross-section SEM images. The outer line profiles were digitized and compared to the simulation output using Prolith v5.07

.

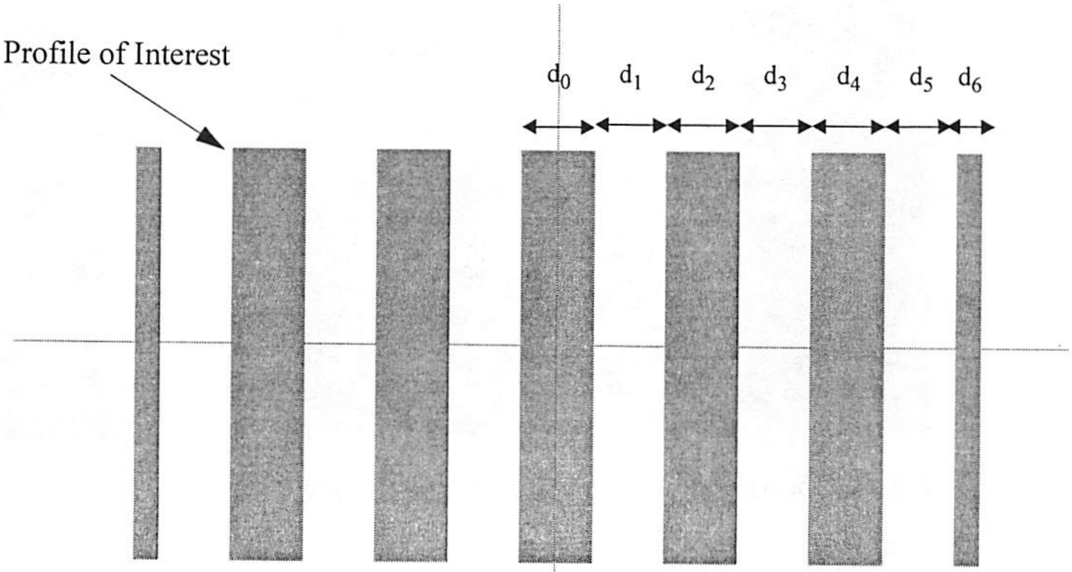


Figure 5.5. The mask layout design for the experiment.

	d_0	d_1	d_2	d_3	d_4	d_5	d_6
layout 0	0.22	0.26	0.22	0.26	0.22	0.26	0.08
layout 1	0.24	0.24	0.24	0.24	0.23	0.26	0.08
layout 2	0.26	0.22	0.26	0.22	0.24	0.26	0.08
layout 3	0.20	0.28	0.20	0.28	0.21	0.26	0.08
layout 4	0.18	0.30	0.18	0.30	0.20	0.26	0.08
layout 5	0.22	0.28	0.22	0.28	0.22	0.28	0.08
layout 6	0.24	0.26	0.24	0.26	0.23	0.28	0.08
layout 7	0.26	0.24	0.26	0.24	0.24	0.28	0.08
layout 8	0.20	0.30	0.20	0.30	0.21	0.28	0.08
layout 9	0.18	0.32	0.18	0.32	0.22	0.28	0.08

Table 5.1. **Geometric Definition of the Layout in the Experiment (in μm)**

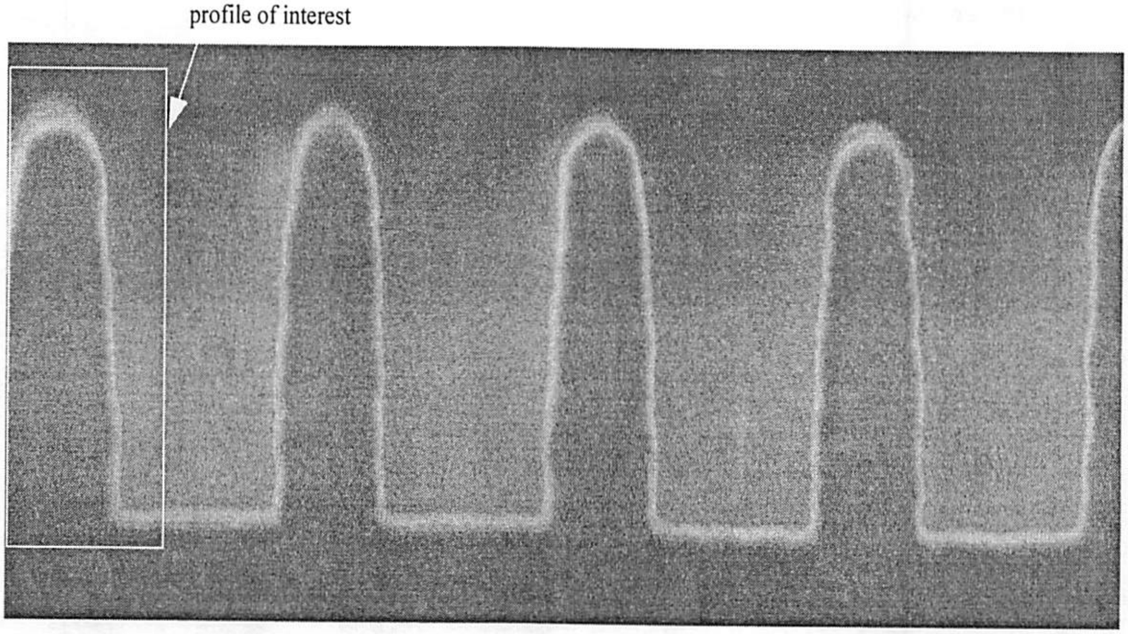


Figure 5.6. An example of a cross-sectional SEM image.

5.4. Results of Parameter Extraction through Profile Matching

5.4.1. Category I parameters

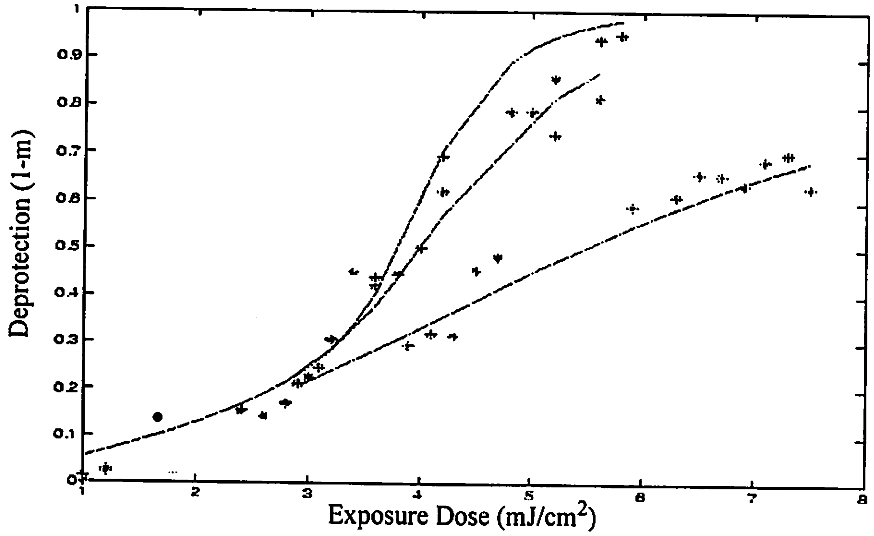
Spectroscopic ellipsometry measurements made on photoresist on an anti-reflective coating on silicon, provided values for the refractive index n, the extinction coefficient k and the thickness for each of the films. These values were measured with very high confidence (intervals <0.1%, as specified by the commercial metrology tool), and were hence kept fixed for the rest of the study. Their values are shown in Table 5.2.

Parameter	Value
Resist Refractive Index (Real part)	1.8038
Resist Refractive Index (Imaginary part)	0.01036
ARC Refractive Index (Real part)	1.4525
ARC Refractive Index (Imaginary part)	0.4028

Table 5.2. Category I: Well Known Parameter Values

5.4.2. Category II parameters

The PEB model was used to extract the Dill's C parameter, the relative quencher concentration [Q], the pre-exponent for acid amplification A_{amp}, the exponent for acid amplification E_{amp}, the pre-exponent for the neutralization reaction A_{α} and the exponent for the neutralization reaction E_{α}. However, unlike existing models, this model explains the data over the entire range of doses, and hence provides good estimates of all of the above parameter values. Using (5.2), the SA optimization procedure provided the fit to the experimental data obtained from the FTIR experiment, and the corresponding parameter values are shown in Figure 5.7. The values of all the extracted parameters were very close to the values obtained using alternate chemical tests, such as the colorimetric titration procedure [41], and activation energy studies commonly performed to study chemical reactions. The FTIR experimental data was optimized over 6 parameters and took less than one minute of CPU time on a 450 MHz P-II processor to converge to the error limit specified (R^2 of 0.95).



$\ln(A_{amp}) = 48.063$ 1/s

$E_{amp} = 39.629$ cal/mole

$\ln(A_{\alpha}) = 49.809$ 1/s

$E_{\alpha} = 39.605$ cal/mole

$C = 0.0563$ cm^2/mJ

$Q = 0.191$

Figure 5.7. Experimental and Fitted Values for Deprotection vs Exposure Dose as a function of 3 different PEB temperatures (120, 130, 135 degrees C)

The Mack develop model was used to extract the dissolution parameters. Figure 5.8 shows the fit to the experimental data, and the corresponding parameter values are shown to the right of the figure. Once again, the optimized parameter values were compared to those obtained from alternate tests, such as monitoring the dissolution rate in unexposed areas and high exposure areas in order to determine the minimum and maximum develop rates, respectively. The values obtained were similar in all cases. The DRM experimental data was optimized over 4 parameters and took less than one minute of CPU time on a 450 MHz P-II processor to converge to the error limit specified (R^2 of 0.95).

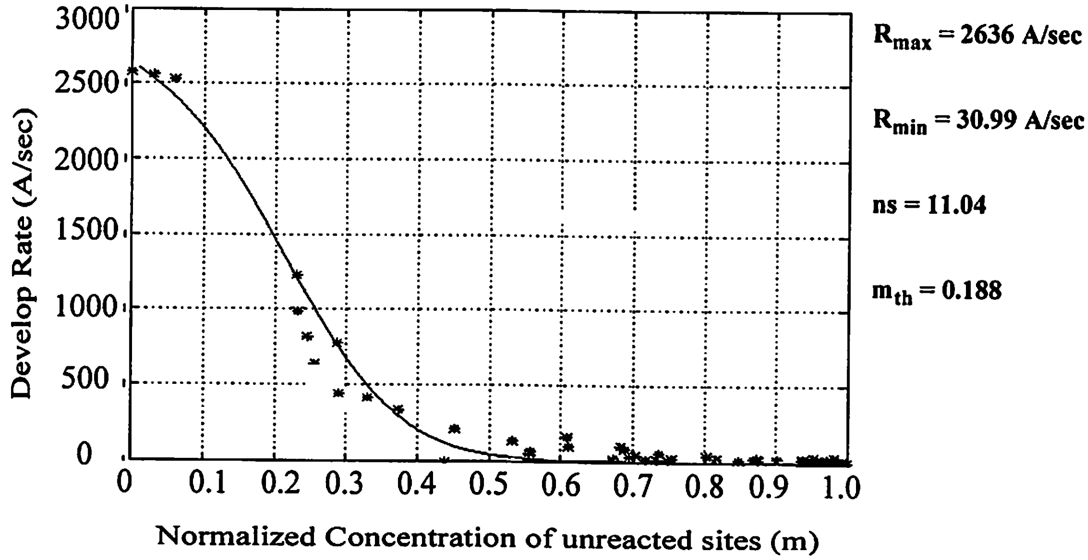


Figure 5.8. Develop rate versus the normalized concentration of unreacted sites. Figure shows the fitting of the Mack develop model to the data.

It should be noted that, while the initial values for the parameters in this category were extracted using the framework shown in Figure 5.2, the final optimized parameter values, shown in Table 5.3, were extracted using the framework shown in Figure 5.4.

5.4.3. Category III parameters

The framework in Figure 5.4 was used to extract the parameters that must be inferred indirectly (Category III), and to also refine estimates of the parameters (category II). It was difficult to validate the estimates in this category with any alternate methods due to the nature of the parameters. The final Category III parameter values are shown in Table 5.4.

In this particular problem, since all the patterned experiments were done at a single PEB temperature, we could not extract the pre-exponent and the activation energy simultaneously. Hence, we assumed the pre-exponents to be the values obtained from the unpatterned resist experiments that were done at multiple PEB temperatures, and we optimized

Parameter	Value
Amplification Rate (Pre-exponent)	47.65 1/s
Amplification Rate (Activation Energy)	39.623 Kcal/mole
Neutralization Rate (Pre-exponent)	49.809 1/s
Neutralization Rate (Activation Energy)	39.605 Kcal/mole
Dill's A parameter	0.00 /µm
Dill's B parameter	0.4683 /µm
Dill's C Parameter	0.065 cm^2/mJ
Relative Quencher Concentration	0.15065
Maximum Develop Rate	3664.5 A/sec
Minimum Develop Rate	6.06 A/sec
Developer Selectivity	9.8164
Developer threshold PAC	0.1433

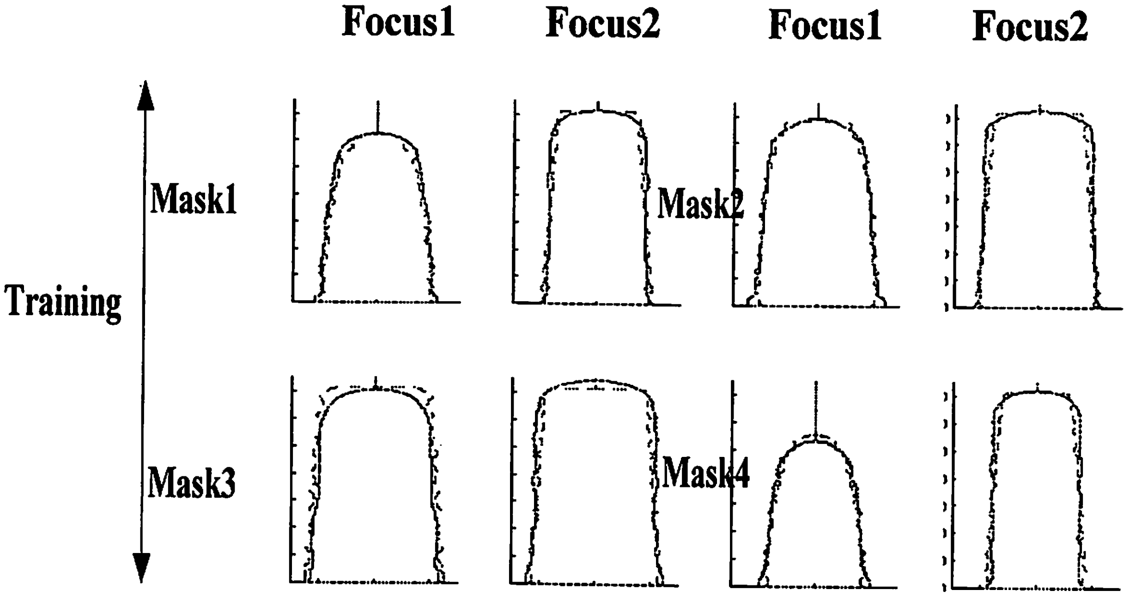
Table 5.3. Category II: Parameters subject to experimental and modeling errors

Parameter	Value
PEB Diffusivity (Pre-exponent)	50.453 nm^2/s
PEB Diffusivity (Activation Energy)	39.623 Kcal/mole
Relative Focus	-0.55µm

Table 5.4. Category III: Parameters that must be inferred indirectly

only for the activation energies. If there were more than one temperature used in the patterned experiments, one could have extracted both parameters simultaneously.

Overall, the patterned data optimization was done over 12 parameters, and the package used for the simulations was Prolith v5.07. The optimization procedure took approximately 20 hours of CPU time on a 400 MHz P-II processor to converge to the error limit specified (R^2 of 0.95). There were 4 different layouts at 2 different focus settings used as training data for the simulator calibration procedure. Having calibrated the simulator, the next step was to test its predictive capabilities. The cross-sectional profile information from the other 6 layouts at 2 different focus settings were used as testing data. Figure 5.9 shows the fit between profiles measured by the cross-section SEM and those derived from simulation for the training data while Figure 5.10 shows the same for the testing data..



PEB Temperature = 130°C
PEB time = 90 secs
Exposure Dose = 13.5 mJ/cm^2

Focus1 = 0.0 μm
Focus2 = 0.3 μm
Resist Thickness = 655 nm

Develop Time = 60 secs
NA = 0.57
σ = 0.5

Figure 5.9. Fitted versus simulated profiles across different focus-layout combinations for training data.

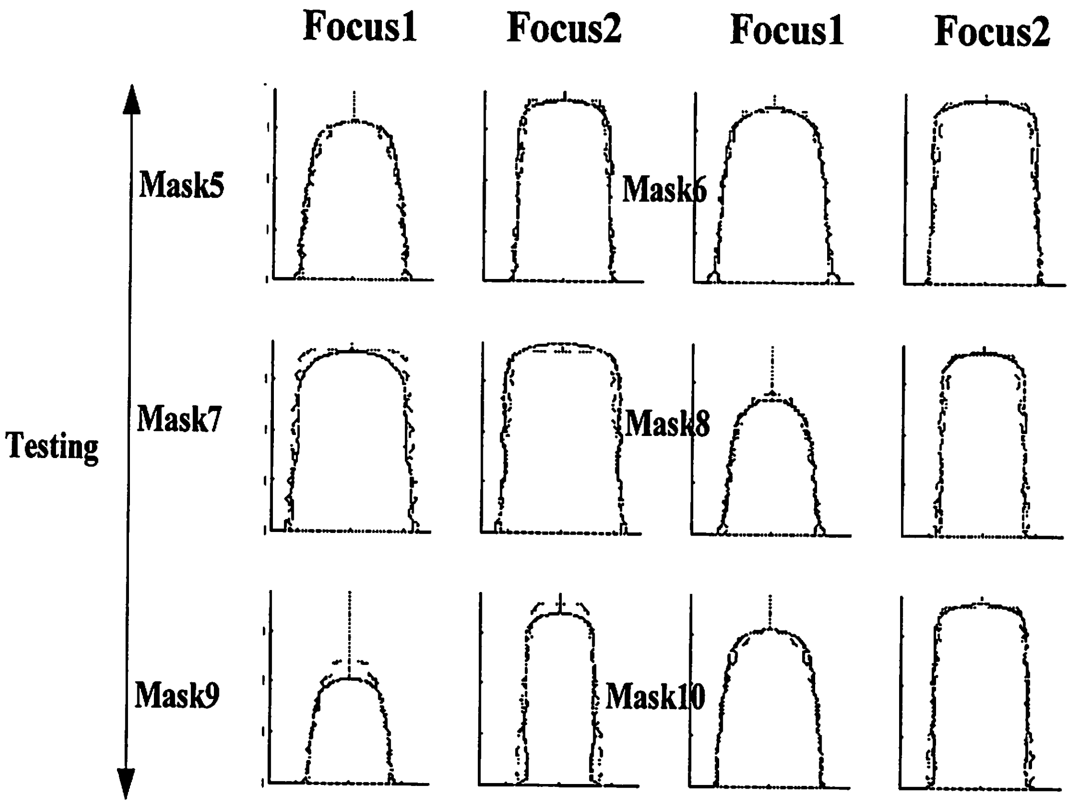


Figure 5.10. Fitted versus simulated profiles across different focus-layout combinations for testing data.

5.5. Summary

The main goal of this work was to develop an efficient simulator calibration framework that increases the confidence of the process engineer in the simulation tool, and subsequently reduces reliance on experimentation. By transferring the burden of process development from hardware (process equipment) into software (simulation), the process engineer will benefit from both reduced development costs and faster time to yield.

We presented the experimental work done, both in unpatterned and patterned characterization experiments, along with the mask layouts and dimensions used for a state-of-the-art process. The results demonstrate excellent predictive capabilities. We believe that this framework can potentially improve yield ramp rates and hence reduce development costs.

Chapter 6 Run-to-Run Control in the DUV Lithography Sequence

6.1. Introduction

Feature dimensions in semiconductor manufacturing are continually decreasing, while die and wafer sizes are increasing. As the critical feature size decreases below 0.25 μm, Deep Ultraviolet (DUV) lithography remains the key technology driver in the semiconductor industry, accounting for approximately 35% of processed wafer cost. However, submicron DUV photolithographic processes present significant manufacturing challenges due to the relatively narrow process windows often associated with these technologies. The sensitivity of the process to small upstream variations in incoming film reflectivity, photoresist coat and softbake steps as well as the bake plate temperature can result in the final critical dimension (CD) going out of specifications. Further, CD problems are usually not identified until the end of the lot. The high costs associated with the manufacture of Integrated Circuits necessitates higher yields and throughput, requiring a reduction in process variability. One approach to reducing process variability is to use a supervisory system that controls the process on a real time or run-to-run basis [1]. Real time control involves the collection of sensor signals during the processing of a wafer and adapting the process recipe during the course of the wafer. Run-to-run control involves adapting the process recipe between wafers. Real time control is more aggressive and more involved than run-to-run control in general.

High end devices such as microprocessors require a considerable number of process steps. Therefore, it is becoming increasingly important to have an accurate, quantitative description of the submicron structure after each step. Currently the lithography process is monitored before photoresist spin on (index, thickness and uniformity measurement of incoming stack) and after development (linewidth and profile measurement). Inspection at the initial and final stages of the process, however, provides only a measurement of the cumulative effects of all the upstream process steps. To isolate the effect of each process step, monitoring at each step is necessary. This need for wafer process monitoring requires in-line sensors and real time algorithms to facilitate real time analysis of sensor signals. Algorithms and sensors for this purpose have been described in the preceding chapters. In-line metrology is preferred to off-line metrology due to increased throughput and possibly yield.

The need for in-situ and/or in-line process monitoring must however be balanced with critical manufacturing issues such as possible adverse effects on throughput, cost, sensor integration into an overall control strategy, possibly limited sensor reliability, etc. Most commercial metrology equipment is either too slow or too complex to be implemented in an in-line arrangement. The sensors and algorithms described in the preceding chapters can satisfy the requirement of making measurements that are sufficiently accurate, repeatable and rapid at a low cost.

Run-to-run control on a lot to lot time scale has already been successfully implemented at Motorola, implementation and results of which are discussed in [24 from John]. This study showed that a very simplistic control algorithm and CD-SEM measurements on a lot-to-lot scale, reduced the pre-etch CD standard deviation from 9.4 nm to 6 nm and the post-etch CD standard deviation from 11.1 nm to 7.1 nm. While this study reduced the variability occuring on a lot-to-lot time scale, we believe that the use of the more sophisticated control algorithms working in unison with in-line/in-situ sensors will reduce this process variability even further, by reducing both the wafer-to-wafer variability in addition to the lot-to-lot variability. Musacchio [Thesis] showed that slightly more sophisticated algorithms, such as the EWMA with robust drift cancellation, provided a further 40% reduction in lot-to-lot CD variability.

6.2. Sources of Variability in the Lithography Sequence

As described in detail in chapter 1, the DUV lithography sequence consists of the spin-coat and soft-bake steps followed by the exposure and post-exposure bake steps and finally the development step.There have been quite a few studies discussing the sources of variability in the lithography sequence [Sturtevant][Crid]. These studies, along with discussions with fab managers, provide us with the most likely sources of variation. The sources of variation at each process step are given below:

6.2.1. Incoming Thin Film

Variations in the incoming thin film stack optical properties, such as the complex index of refraction and thickness, change the reflectance as seen by the stepper/scanner. This changes the effective exposure energy that gets coupled into the resist. The changes in the optical properties of the underlying thin film could be due to a change in the deposition conditions, such as temperature, gas flow, etc.

6.2.2. Spin Coat and Soft Bake

Typically, there is a 5-10% variation from batch-to-batch in chemically amplified resist systems. This variation could occur in the viscosity or the quencher concentration. A variation in the former leads to different resist thicknesses at the same spin speed, thus causing a change in the effective exposure energy coupled into the resist. A variation in the latter acts to neutralize less or more of the generated acid, effectively changing the effective exposure energy coupled into the resist. Variation in the soft bake time or temperature affects the decomposition rate and the amount of solvent left in the resist, both of which change the effective exposure dose and diffusivity of the acid during the post-exposure bake step.

6.2.3. Exposure

The light source in a stepper/scanner is an excimer laser, which provides the required exposure energy through a discrete number of laser pulses. Each laser pulse has random variations in its energy and hence the final exposure energy is typically controlled to within +/- 2%. This difference in the set vs. actual dose plays contributes significantly due to the chemically amplified resist systems.

6.2.4. Post-Exposure Bake (PEB)

The PEB is probably the single biggest contributor to CD variability due to the chemically amplified resists being very sensitive to the thermal dose. Variations in the bake plate temperature affect both the deprotection rate as well as the acid diffusion, causing variation in the CDs.

6.2.5. Development

Variation in the develop step is found to be minimal if a correct recipe is used that optimizes the combination between the spray step and the puddle step.

6.3. Metrology Schemes

Having outlined the possible sources of CD variation, the next step is to identify the metrology schemes that could be used to measure the intermediate steps. The different metrology systems for the incoming reflectance, deprotection induced thickness loss (DITL) and the CD have been described in chapter 2. We will devote this section to determining possible in-line/in-situ configurations for the different metrology systems as well as studying the measurement frequency that each system is capable of.

Wafer reflectance measurements, both before and after the spin-coat and soft-bake step, can be made with an normal incidence reflectometer as described in chapter 2. There are a number of commercial reflectometers, both single wavelength and spectroscopic, such as the SC Technology and the Ocean Optics systems, that can be integrated on the chill plates of some wafer track systems that have enough open space above the chill plate. In wafer tracks that do not have the luxury of the open space above the chill plate, reflectometry systems have been successfully installed on the robotic arm that moves the wafer between the different modules. It is the simplicity and the low cost of the reflectometry system that allows it to be relatively easily integrated in-line and this measurement could therefore possibly be made on every wafer.

The DITL measurements, as described in chapter 2, can be made using a standard spectroscopic reflectometer. In an in-line configuration, DITL measurements would be made during the chill step (after the PEB step). Hence, the reflectometer could be mounted in the same way as described in the preceding paragraph. In an in-situ configuration, for end-pointing the bake step, a more compact version of the reflectometer would need to be set up since the bake plates are usually very tightly sealed in order to maintain tight temperature control. Commercial reflectometers that allow precise positioning and pattern recognition are now being introduced into the market through companies like Nanometric, Inc. and Nova. Once again, these measurements could possibly be made on every wafer due to the above stated advantages of reflectometry.

The use of optical tools, such as spectroscopic ellipsometers, for in-line CD metrology have been described earlier. The footprint required for ellipsometers is typically more than that required for reflectometry, due to the larger number of optical elements required to polarize the light. One concept for in-line CD metrology that is being investigated is a metrology end-station. This could be placed between the lithography and the etch equipment, thus allowing CD measurements to be made on every wafer both before and after etch.

In this chapter, we discuss an algorithm that allows information from one or more of the in-line sensors to be integrated at different time scales. This is known as sensor fusion and has been used in a plethora of fields [References]. Sensor fusion allows better estimation of the process noise and hence a better controller performance.

6.4. In-line RtR Controller Design

6.4.1. Experiments

The first step is to determine the relationships between the different critical process parameters, such as the exposure energy, PEB time (interchangeable with the PEB temperature), and the process observables, such as the DITL and the CD. This involved carrying out a designed experiment and was performed at National Semiconductor Corporation's class 1 fab in Sunnyvale, CA.

The experiment was carried out as follows. Four bare silicon wafers were spun on with 660A of an anti-reflective coating and then baked at the standard soft-bake conditions. After chilling the wafers, they were spun on with Shipley UV5, a commercially available chemically amplified resist, to a thickness of 6550A. After spin on, the wafers were soft-baked at the standard conditions to remove solvent and chilled again. All the wafers were then measured for thickness using a KLA-Tencor UV1250SE, a single angle spectroscopic ellipsometer. The wafers were then sent to a DUV stepper, where two exposure passes were made. In the first pass, the standard product mask was used with exposure doses ranging from 12.2 mJ/cm^2 to 13.8 mJ/cm^2 in steps of 0.2 mJ/cm^2 , centered around the nominal exposure dose to size of 13 mJ/cm^2. The focus was set to the best focus, obtained by running a send ahead test wafer. The goal of the first pass was to print CDs that could be measured with a CD-SEM. The second pass was used to print blanket exposures in order to measure the DITL. This was done by exposing the wafers using doses ranging from 4.6 mJ/cm^2 to 5.4 mJ/cm^2 in steps of 0.1 mJ/cm^2. The nominal dose was centered around the dose to clear, i.e. 5 mJ/cm^2. The exposure doses for the blanket areas were chosen so as to maintain the ratio of dose to size to dose to clear for all the die. This accounted for their different diffraction efficiencies thus ensuring that the blanket and adjacent patterned areas received the same effective exposure dose. Figure shows the details of how each die was exposed.

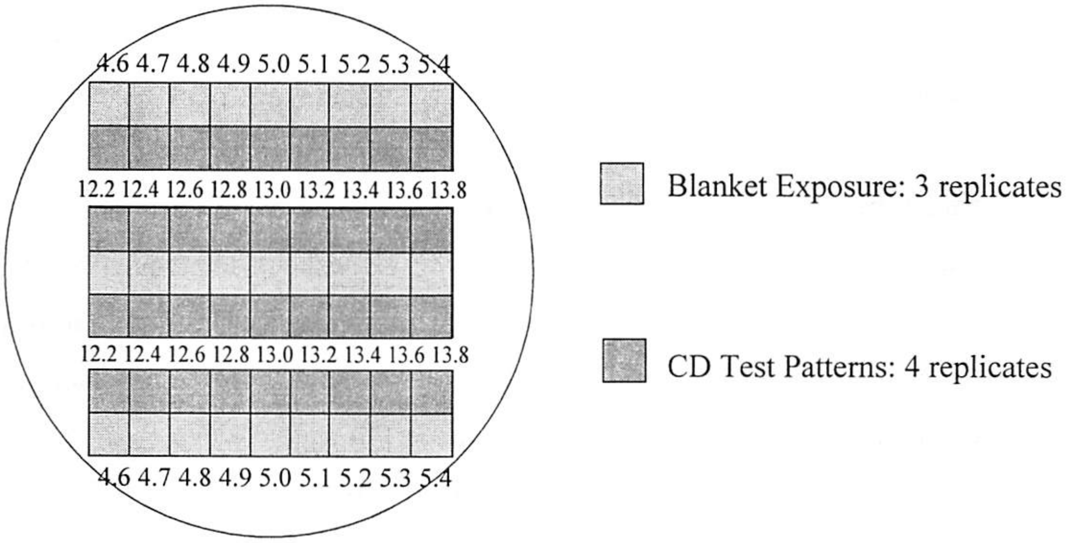


Figure 6.1. Blanket and Patterned Exposure Pattern for the CD control model experiment

The wafers were then post-exposure baked at 135 degrees Celsius at 75, 90, 90 and 105 seconds. After being cooled on the chill plate, the four wafers were measured on the KLA-Tencor spectroscopic ellipsometer for remaining resist thickness in the flood exposed areas, thus providing the DITL as a function of exposure dose and PEB time. The wafers were then developed for 60 seconds and taken to KLA 8100 CD-SEM for CD measurements in the patterned die. The details of the experiment are listed in Table 6-1.

Process Recipe	Value / Range of Values
Resist Thickness	6550 (Angstroms)
ARC Thickness	660 (Angstroms)
Exposure Dose (Blanket)	4.6 : 0.1 : 5.4 (mJ/cm^2)
Exposure Dose (Patterned)	12.2 : 0.2 : 13.8 (mJ/cm^2)
PEB Time	75, 90, 90, 105 (seconds)
PEB Temperature	130 (degrees Celsius)
Develop Time	60 (seconds)

Table 6-1. Experimental Conditions used in Process Model Building

6.4.2. Process Models

Regression models were built to model the DITL, the CDs and the process inputs. Transformations were introduced in some of the terms in order to improve the fit of linear models to the experimental data. The results are summarized in Table 6-2 [22 of John]. The first and the second models relate DITL and CD (respectively) to the process inputs i.e. bake time and exposure dose. The last model relates CD to DITL and the PEB time. The rationale for having both DITL and PEB time in the model to predict CD is that while DITL captures the deprotection behavior, it does not effectively capture the diffusion of acid within the resist during the PEB. Studies have shown that the diffusion is a fairly linear function of the PEB time [Neureuthers paper]. It should be pointed out that although the CDs in this experiment were measured with a CD-SEM, they would ideally have been measured with the specular spectroscopic scatterometry technique since that would allow both the observables (CD and DITL) to be measured in-line. .

#	Model	Adj. R^2	Std. Error
I	$\sqrt{DITL} = -23.4615 + 2.5176 \times Dose + 0.0334 \times Time$	0.9802	$0.22A^{0.5}$
II	$CD = -0.016 \times Dose + 0.0906 \times \sqrt{Time} - 0.0052 \times Time$	0.9996	3.837 nm
III	$CD = -0.0029 \times Time + 0.0508 \times \sqrt{Time} - 0.000255 \times DITL$	0.9995	4.216 nm

Table 6-2. Modeling Experiment Results

6.4.3. Drift Model

Having found static models relating the inputs and outputs of the DUV lithography process, it is now necessary to develop a model of how the process drifts. The issue addressed in this section is where in our process model should we account for this variation.

In section 6.2 we surveyed the likely sources of variability in the DUV lithography sequence. We attributed the CD variability to incoming wafer reflectivity variation, batch to batch resist variation and exposure and thermal dose variation. These sources of variability as well as others have been modeled to first order with additive noise on the process inputs - exposure dose and PEB time [John thesis].

$$Dose_{eff} = Dose_{tuned} + Noise_{dose} \tag{6.1}$$

$$Time_{eff} = Time_{tuned} + Noise_{Time} \tag{6.2}$$

Here, we have introduced the terms $\mathrm{Dose_{eff}}$ and $\mathrm{Time_{eff}}$ to denote the sum of each input setting and noise term. Our final models for DITL and CD will be those summarized in Table 6-2, but with $\mathrm{Dose_{eff}}$ and $\mathrm{Time_{eff}}$ substituted wherever Dose and Time appear.

$$\sqrt{DITL} = -23.4615 + 2.517 \times Dose_{eff} + 0.0334 \times Time_{eff} \tag{6.3}$$

$$CD = -0.016 \times Dose_{eff} + 0.0906 \times \sqrt{Time_{eff}} - 0.0052 \times Time_{eff} \tag{6.4}$$

Finally, we assume that some measurement noise occurs in measuring both the DITL as well as the CD.

$$DITL_{meas} = DITL_{true} + noise \tag{6.5}$$

$$CD_{meas} = CD_{true} + noise \tag{6.6}$$

6.4.4. RtR Control Architecture

6.4.4.1. Scenario I

In this scenario, the in-line reflectometer measures the resist thickness before and after the exposure and PEB steps, in order to calculate the DITL. This DITL value is used to estimate the post-develop CD which is then used in conjunction with a standard RtR control algorithm, to prescribe a recipe for the subsequent wafer. A schematic of the control architecture is shown in Figure 6.1, and a summary of the notation used in the controller development is presented in Table 6-3 [Johns thesis]. .

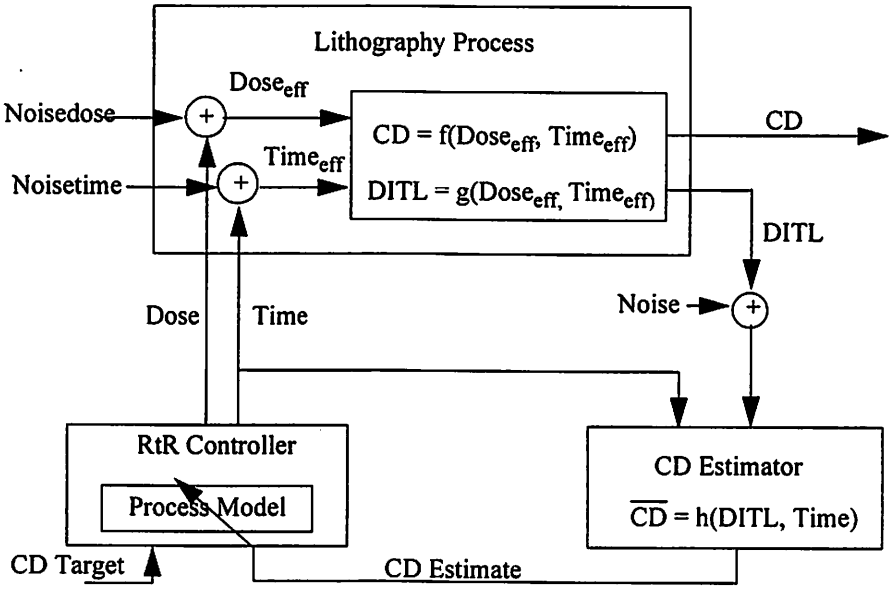


Figure 6.1. RtR Control Architecture for DUV Lithography in Scenario I

Definition	Explanation
T=CD Target	Desired CD
$y_k^0 = CD_k$	True CD of wafer k
$y_k = \hat{CD}_k$	CD estimate of wafer k
$e_k = y_k - y_k^0$	CD estimation error of wafer k
$u_k = \begin{bmatrix} Dose_k \\ Time_k \end{bmatrix}$	Input vector for wafer k

Table 6-3. Control Architecture Notation

The CD estimation block of Figure 6.1 takes a presumably noisy DITL measurement and the wafer's PEB time and applies Model III of Table 6-2 to arrive at an estimate for CD.

$$y_k = \hat{CD}_k = -0.0029 \times Time_k + 0.0508 \times \sqrt{Time_k} - 0.000255 \times DITL_k \qquad (6.7)$$

The target CD and CD estimate of the previous wafer are fed into the RtR Controller block of Figure 6.1. Musacchio [thesis] has shown that the use of the additive noise terms to the exposure dose and PEB time in the drift model makes the sensitivities to the process inputs fixed to first order. He uses an EWMA technique to handle this offset drift problem. Jakatdar, et.al. [src98report]

6.4.4.2. Simulation

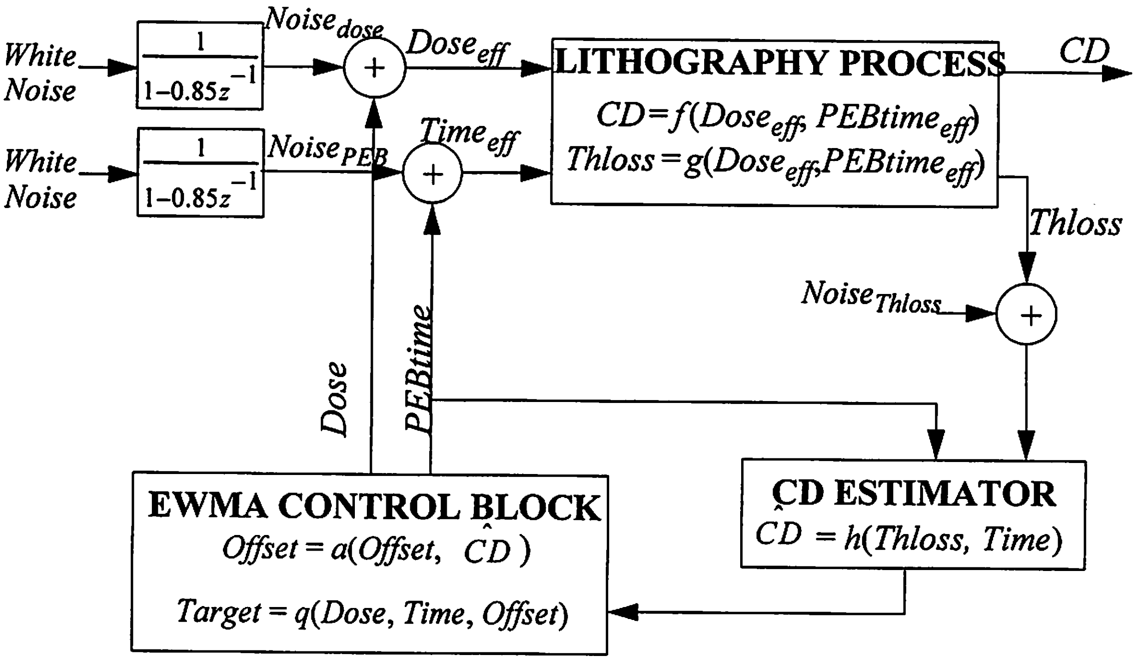


Figure 6.1. Simulation Architecture

Figure 6.1 outlines the simulation architecture. The Simulator simulates three blocks: the drifting lithography process, the CD estimator, and the EWMA controller. The CD estimator and EWMA Control blocks are implemented with the equations given in Section 5.4.4. Simulating the lithography process block also requires specifying a stochastic drift model, which is the purpose of this section.

The lithography process block takes exposure dose and PEB time as inputs, and outputs CD and Thickness Loss. The model for the lithography process is that described by Section 5.4.3. Thus, *f*(,) and *g*(,) of Figure 5.9 are the models for CD and thickness loss found from the modeling experiment, with the substitution of a $Dose_{eff}$ and $PEBtime_{eff}$ for $Dose$ and $PEBtime$.

We must assume that $Noise_{dose}$ and $Noise_{PEB}$ has some autocorrelation structure between runs. Because we do not know the exact autocorrelation structure we will encounter in the real lithography process, by default we assume the simplest model possible, a first order AR. The modeling technique is that illustrated in Section 2.3.1**. We define

$$x_k = \begin{bmatrix} \{Noise_{dose}\}_k \\ \{Noise_{PEB}\}_k \end{bmatrix} \tag{6.8}$$

Now suppose we expect the variance of $Noise_{dose}$ to be σ^2_{dose} and the variance of $Noise_{PEB}$ to be σ^2_{PEB}. Also suppose we expect the autocorrelation coefficient between successive runs to be $P_{d,1}$ and $P_{d,2}$ for $Noise_{dose}$ and $Noise_{PEB}$ respectively. With the following definitions:

$$K = \begin{bmatrix} \sigma^2_{Dose} & 0 \\ 0 & \sigma^2_{PEB} \end{bmatrix}, F = \begin{bmatrix} f_1 & 0 \\ 0 & f_2 \end{bmatrix}, G = \sqrt{K - FKF*} \tag{6.9}$$

the stochastic process for x_k is

$$x_{k+1} = Fx_k + Gw_k \tag{6.10}$$

Where w_k is a 2 dimensional gaussian random vector with identity covariance. As indicated by Figure 5.9, the random process generated by (6.10) can be generated by passing white noise through two filters, each with a pole at f_i and a zero at 0.

The thickness loss measurement error, $Noise_{Thloss}$, is assumed to be white and normally distributed with variance σ^2_{Thloss}.

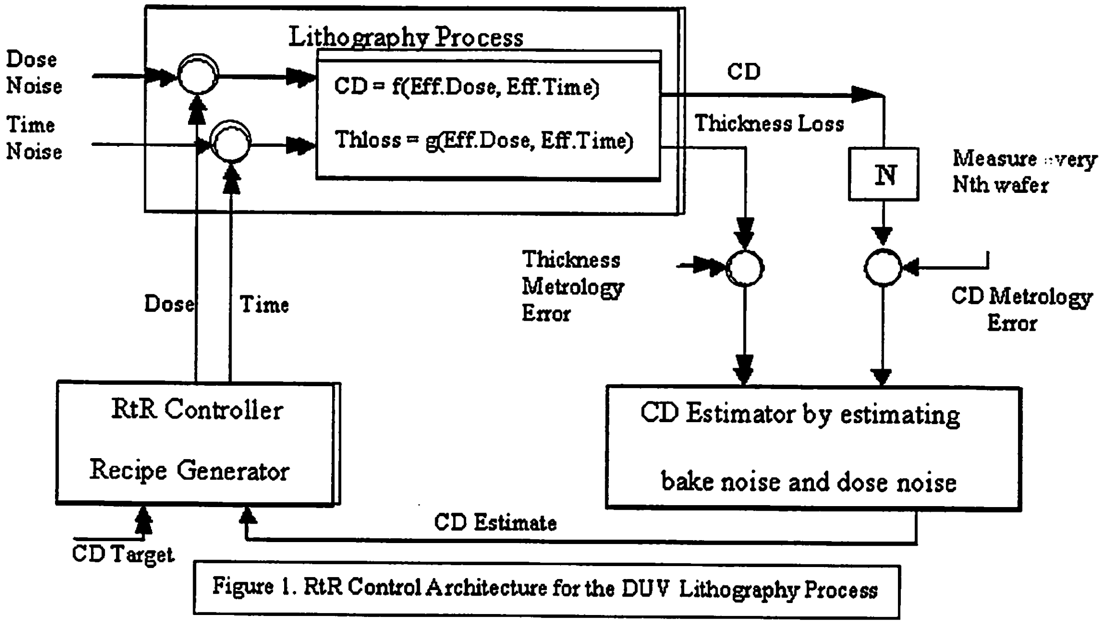


Figure 6.2. RtR Control Architecture for DUV Lithography in Scenario II

Chapter 7 Conclusions

7.1. Thesis Summary

The history of the semiconductor industry has been characterized by Moore's Law that predicts the exponential scaling factor for integrated circuit density. To keep the industry on its historic 25-30% / year reduction in cost/function despite the escalating factory costs (20% / year), greater synergy must be developed between the areas of metrology, modeling, simulation and control. This thesis has presented a framework that integrates the metrology of wafer level observables with physical models for the same, in order to achieve enhanced predictive simulation as well as to facilitate run-to-run control of the lithography process.

Metrology was developed for the intermediate steps of the pattern transfer sequence, from unpatterned thin films to the latent image in the photoresist during exposure and PEB to the final patterned CD. The development of a metrology scheme for each step was characterized by first identifying the observable at that step and its correlation to the final CD, followed by identifying the appropriate sensor and then presenting an algorithm that rapidly analyzes the sensor data. While some of the algorithms used in this thesis are in their original incarnation, some others, such as the NN-ASA, have been modified to overcome the drawbacks of the original algorithms. Finally, experimental results for each metrology scheme were presented.

This thesis also introduced the concept of deprotection induced thickness loss or DITL, as a means of providing information on the state of the resist through the measurement of a simple observable. Both static and dynamic models were presented that help to get a better estimate of important model coefficients, such as the diffusion coefficient and the acid quencher concentration. The computation of the DITL was a moving boundary problem that needed to be solved.The experimental results show the validity and importance of using such a concept to extract the model coefficients.

A lithography simulation framework is presented that provides a hierarchical method for determining the model coefficients of lithography simulation models. These coefficients are determined using experimental data from both unpatterned and patterned wafers, undergoing a coarse optimization loop followed by a finer optimization loop. The results demonstrate that this technique could dramatically reduce the time to develop a process through increased predictive simulation capabilities.

Finally, a run-to-run controller is presented that attempts to correct for process drift at the different process steps in the lithography sequence. The controller uses a Kalman Filter to provide estimates of the noise and uses process models based on a statistical design of experiments technique. Two scenarios are considered, differing in the type of metrology as well as the frequency of measurements available. The simulation results indicate the efficacy of using such a scheme for a real-world lithography sequence.

7.2. Future Work

While this thesis has focused on the building blocks for a completely automated process development and manufacturing sequence, there has been more emphasis placed on the experimental verification of the metrology, modeling and simulation blocks with only simulation results being presented for the control block. The experimental verification of the run-to-run control scheme presented here could be an important area of future research, with the emphasis being on the integration of sensors with the processing equipment as well as studying the efficacy of the different control algorithms in actual drift conditions.

Also, while the reduction of variability on a wafer-to-wafer and lot-to-lot time scale has been the goal of this run-to-run controller, the spatial components of variability, viz. die-to-die and within die components have not been addressed. The spatial component of variability is becoming an increasingly bigger source of variability in the lithography process. Dealing with this problem would require schemes for spatially resolved actuation, such as field-varying exposure doses and a spatially resolved bake plates [47]. This multiple input, multiple output would be a complex control problem.

In the area of metrology, the specular spectroscopic scatterometry (SSS) technique can be extended to other processing steps as well, such as multi-material gratings as in Chemical Mechanical Polishing (CMP). The migration of SSS to CMP would require some novel theoretical work since previous grating work has attacked the problem of single material gratings with air. In addition to being used for multiple applications, not enough is understood about the sensitivity of this technique to the different spectroscopic ellipsometry configurations and incident angles. This could prove to be an important area of research in the near future.

An interesting area of future research in simulation would be to identify the sensitivity of the photoresist profile to the various lithography model coefficients. Determination of their sensitivities would provide a means of model coefficient reduction in the optimization problem, thus significantly enhancing the speed of simulator calibration.

References

[1] "The National Technology Roadmap for Semiconductors: Technology Needs", 1999 Edition, Semiconductor Industry Association (SIA)

[2] D. Gerold, et.al., "Run-to-Run Control Benefits to Photolithography", SEMATECH AEC/APC Workshop IX, Septemeber 21-24, 1997.

[3] M. Yelverton, "An Application of Advanced Process Control in a High-Volume Manufacturing Operation", SEMATECH AEC/APC Workshop XI, September 11-16, 1999.

[4] L. Ingber, "Simulated Annealing: Practice versus Theory", Mathematical and Computer Modeling, vol. 18, no. 11, 29-57, 1993.

[5] L. Ingber, "Adaptive Simulated Annealing (ASA)", ftp://alumni.caltech.edu/pub/ingber/, 1995.

[6] R. P. Lippmann, "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, Vol.4, pp. 4-22, April 1987

[7] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994

[8] S. Haykin, Adaptive Filter Theory, Prentice-Hall, New Jersey, 1991

[9] J. Moody & C. J. Darken, "Fast Learning in Networks of Locally Tuned Processing Units", Neural Computing, Vol. 1, pp. 281-294, 1989

[10] B. Anderson, J. Moore, Optimal Filtering, New Jersey, Prentice Hall 1979

[11] S. Leang, C. J. Spanos, "A Novel In-line Automated Metrology for Photolithography" IEEE Transactions on Semiconductor Manufacturing. Feb 1996

[12] A. R. Forouhi and I. Bloomer, "Optical Properties of Crystalline Semiconductors and Dielectrics", Phys. Rev. B, vol. 38, 1865, 1988

[13] S. Leang, "Supervisory Control System for a Photolithographic Workcell" M.S. Thesis, July 1992

[14] X. Niu and C. J. Spanos, "Statistical Enhancement of a Reflectometry Metrology System" First International Workshop on Statistical Metrology. June 1997.

[15] S. Wolf, R.N. Tauber, "Silicon Processing for the VLSI era" Vol 1, Lattice Press, 1986

[16] S. E. Stokowski, "Measuring refractive indices of films on semiconductors by micro-reflectometry", SPIE vol. 1261, 1990

[17] H. Ito and C. G, Willson, "Applications of Photoinitiators to the Design of Resists for Semiconductor Manufacturing", in Polymers in Electronics, ACS Symposium Series 242 (1984) pp. 11-23.

[18] C. Mack, "Inside Prolith - A Comprehensive Guide to Optical Lithography Simulation", February 1997.

[19] R. Carpio, J. D. Byers, J. S. Petersen, W. Theiss, "Advanced FTIR Techniques for Photoresist Process Characterization", SPIE vol. 3050, 1997

[20] C. Chen, J. Lee, M. Blackwell, "Fourier Transform Infrared Analysis of Resist Process and its Application", SPIE vol. 1086, 1989

[21] A. Krasnoperova, et. al., "Modeling and Simulations of a Positive Chemically Amplified Photoresist for X-Ray Lithography", J. Vac. Sci. Technology B 12(6) Nov. - Dec. 1994

[22] X. Niu, N. Jakatdar, J. Bao, C. Spanos, "Specular Spectroscopic Scatterometry in DUV Lithography", SRC Techcon, 1998.

[23] E. A. Irene, "Applications of Spectroscopic Ellipsometry to Microelectronics", Thin Solid Films, vol. 233, no.1-2, pp. 96-111, 1993.

[24] R. A. Synowicki, J. N. Hilfiker, C.L. Henderson, "Refractive Index measurements of Photoresist and Anti-Reflective Coatings with Variable Angle Spectroscopic Ellipsometry", Proc. of SPIE, vol. 3332, pp. 384-390, 1998

[25] P. Boher, J. L. Stehle, J. P. Piel and C. Defranoux, "Precise Measurement of ARC Optical Indices in the Deep UV Range by Variable Angle Spectroscopic Ellipsometry", Proc. of SPIE, vol. 3050, pp. 205-214, 1997

[26] X. Niu, "An Integrated System of Optical Metrology for Deep Sub-Micron Lithography", Ph.D. thesis, 1999.

[27] A. Roger and M. Breidne, "Grating Profile Reconstruction by an Inverse Scattering Method", Optics Communications, vol. 35, no. 3, pp. 299-302, 1980

[28] G. Bao, "On the Relation Between the Coefficients and Solutions for a Diffraction Problem", Inverse Problems, vol. 14, no. 4, pp. 787-798, August 1998

[29] L. Capodieci, et. al.,"A novel methodology for post-exposure bake calibration and optimization based on electrical linewidth measurements and process metamodeling", EIPB Conference, May 1998

[30] N. Jakatdar, et.al. "Characterization of a Positive Chemically Amplified Photoresist from the Viewpoint of Process Control for the Photolithography Sequence", SPIE vol. 3332, pp.586-593, 1998

[31] M. Zuniga, A.R. Neureuther, "Reaction-Diffusion Modeling and Simulations in Positive DUV Resists", JVST-B, Nov-Dec 1995, vol. 13, pp. 2957-62

[32] J. Byers, et.al., "Characterization and Modeling of a Positive Chemically Amplified Resist", SPIE Vol. 2438, pp.153-166, 1995

[33] L. Pain, et.al., "Free Volume Effects in Chemically Amplified DUV Positive Resists", Microelectronic Engineering 30 (1996) pp. 271-274

[34] W. Vetterling, et. al.,"Numerical Recipes for C", 2nd edition, 1992

[35] W. Oldham, S. Nandgaonkar, A.R. Neureuther, M. O'Toole, "A General Simulator for VLSI Lithography and Etching Processes: Part I - Application to Projection Lithography", IEEE Transactions on Electron Devices, Vol. 26, No. 4, April 1979

[36] N. Jakatdar, et. al., "Characterization of a Positive Chemically Amplified Photoresist from the Viewpoint of Process Control for the Photolithography Sequence", SPIE Vol. 3332, pp.586-593, 1998

[37] N. Jakatdar, et. al., "Characterization of a Positive Chemically Amplified Photoresist for Simulation Using a Modified Poor Man's DRM Methodology", SPIE Vol. 3332, pp.578-585, 1998

[38] H. H. Szu and R. L. Hartley, Proceedings of the IEEE, vol. 75, 1538-1540, Nov. 1987.

[39] C, Tsallis and D. A. Stariolo, Physica A, vol. 233, no. 1-2, 395-406, 15 Nov. 1996.

[40] D. Marquardt, "An algorithm for least-squares estimation of non-linear parameters", SIAM J. Appl. Math., 11, pp. 431-441, 1963

[41] S. Thornton, C. Mack, "Lithography Model Tuning: Matching Simulation to Experiment", SPIE Vol. 2726, pp.223-234, 1996

[42] J. Sturtevant, M. Weilemann, K. Green, J. Dwyer, E. Robertson, R. Hershey, "Implementation of a Closed Loop CD and Overlay Controller for Sub-0.25 μm Patterning", SPIE vol. 3332, 1998

[43] J. Musacchio, "Run to Run Control in Semiconductor Manufacturing", M.S. Thesis, November 1998.

[44] C. Yu, "Integrated Circuit Process Design for Manufacturability using Statistical Metrology", Ph.D. thesis, August 1996

[45] A. Iturralde, "A Review of Sensing Technologies for Semiconductor Process Applications", ISSM, 1995.

[46] N. Jakatdar, J. Musacchio, C. Spanos, "Control Models for the DUV Workcell", SRC Deliverable Report, April 1998.

[47] M. Zuniga, A.R. Neureuther, "Reaction Diffusion Modeling and SImulations in Positive DUV Resists", JVST B (Microelectronics and Nanometer Structures), vol. 13, no. 6, pp. 2957-62, Nov.-Dec. 1995

[48] K. El-Awady, C. Schaper, T. Kailath, "Thermal Cycle Module for Photoresist Processing", SEMATECH AEC/APC Symposium X, Vail, CO, Oct 1998

[49] S. Bushman and S. Farrer, "Scatterometry Measurements for Process Monitoring of Poly-Silicon Gate Etch", Proc. of SPIE, vol. 3213, 79-90, 1997.

Appendix A Symbols Used

Parameter	**Description**	**Units**
α	Neutralization Reaction Coefficient	-
Φ_{hi}	Linear Combination of the inputs to the ith node in the hth layer used in Logistic Sigmoid Functions	-
λ	Wavelength of light	nm
σ	Partial Coherence	-
θ_{hi}	Bias term for the ith neuron in the hth layer	-
ϖ	Volume Shrinkage Element	μm^3
a_h	Output of the hth layer for a given input x	-
A_i	Fitting Parameter in the Forouhi-Bloomer Equation	-
B_i	Fitting Parameter in the Forouhi-Bloomer Equation	-
c_{cost}	Cooling Scaling Factor for T_{cost}	-
C	Rate of Photoacid Formation	cm^2/mJ
C_i	Fitting Parameter in the Forouhi-Bloomer Equation	-
d	Minimum Inter-cluster Distance	-

D	Deprotection	-
$DITL$	Deprotection Induced Thickness Loss	nm
Dm	Dimensionality of parameters	-
D_k	Diameter of the kth cluster	-
D_u	Acid Diffusivity	$\mu m^2/sec$
D_w	Volatile Molecule Diffusivity	$\mu m^2/sec$
E	Sum squared error	-
E_g	Optical Energy Band-Gap	eV
$g(x)$	Probability Density of state space of D parameters	-
$h(\delta(E,T_n))$	Acceptance Probability of a new cost-function	-
k	Extinction Coefficient	-
k_1	Scaling Factor to convert deprotection into corresponding volatile group concentration	-
k_3	Hole Collapse Rate	sec^{-1}
k_4	Scaling Factor to convert hole concentration into volume shrinkage	-
k_{amp}	Acid Amplification Rate	sec^{-1}
k_{cost}	Number of acceptances	-
k_{loss}	Acid Loss Factor	sec^{-1}
l	Thickness	nm

L_i	Number of Transitions in SA	-
L_n	Repetition Function	-
m	normalized concentration of unreacted blocking sites	-
mth	Value of m at the inflection point of the DRM curve	-
M	Characteristic Matrix for Reflectance Calculations	-
n	Refractive Index	-
n_0	Number of cluster centers	-
ns	Developer Selectivity	-
N	Complex Index of Refraction	-
N_k	Number of points in cluster k	-
$[PAG]_0$	Initial Concentration of Photoacid Generator	mol/cm^3
Q	Normalized Acid Quencher Concentration	-
r_p	Reflection Coefficient for TM wave	-
r_s	Reflection Coefficient for TE wave	-
R	Relative Reflectance	-
R_{max}	Maximum Deveop Rate	A/sec
R_{min}	Minimum Develop Rate	A/sec
s_α	Temperature Anneal Scale	-
s_D	Cost Parameter Scale Ratio	-
S_h	Number of neurons in the hth layer	-

s_T	Temperature Ratio Scale	-
t	PEB time	sec
T	Cost Temperature	-
T_{loss}	Thickness Loss or Volume Shrinkage	nm
T_n	Temperature Function or Cooling Schedule	-
T_{cost}	Initial Temperature of the Acceptance Function	-
u	Normalized Acid Concentration	-
v	Deprotection	-
w	Normalized Deprotected Concentration	-
w_i	Optimization Weight	-
w_{ij}	Optimization weight from the ith node to the jth node	-
$x^i{}_k$	State of the ith parameter at annealing time k	-
z	Depth into the resist	μm

Appendix A