Copyright © 1997, by the author(s).
All rights reserved.

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

IN SITU METROLOGY FOR DEEP ULTRAVIOLET PHOTOLITHOGRAPHY CONTROL

by

Nickhil Jakatdar

Memorandum No. UCB/ERL M97/98

31 December 1997

COVER PAGE

IN SITU METROLOGY FOR DEEP ULTRAVIOLET PHOTOLITHOGRAPHY CONTROL

by

Nickhil Jakatdar

Memorandum No. UCB/ERL M97/98

31 December 1997

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

In-situ Metrology for Deep Ultraviolet Photolithography Control

by
Nickhil Jakatdar

Electrical Engineering and Computer Sciences
University of California at Berkeley

Professor Costas J. Spanos, Advisor

In-situ metrology promises to provide effective manufacturing line operation, reduced cycle times and improved process quality for semiconductor processing. The challenges in developing this technology lie in identifying useful and relatively simple observables in the lithography sequence, relating these observables to the final quantity of interest, developing simple but effective control strategies and finally, integrating this with the production line equipment. Of particular interest is Deep Ultraviolet (DUV) Lithography (248 nm), which is becoming the key technology driver in the semiconductor industry and will remain to do so through the 0.18 μm generation.

This thesis reviews the opportunities for metrology and control in the DUV lithography sequence. It identifies various sensors and algorithms for real time in-situ control and investigates simple control schemes suitable for a typical production line setup. Studies correlating resist thinning to deprotection have been done and its application for lithography control has been proposed. This study promises to cut the wafer to wafer process variation (typically 24 nm) to below 10 nm using adaptive control strategies. A novel real-time algorithm for optical constant extraction over a broadband has been presented using neural network technology in conjunction with an Adaptive Simulated Annealing technique.

Chapter 1. Introduction 7

1.1. Background and Motivation 7
1.2. Developing In-Situ Metrology: A Step by Step Approach 9
1.2.1. When to Monitor - Sensor location in the process stream 9
1.2.2. What to Monitor - Identifying the Observables 10
1.2.3. How to Monitor - Designing the Sensor 10
1.2.4. Real Time Algorithms 10
1.3. Thesis Organization 11

Chapter 2. Chemically Amplified Photoresists 13

2.1. Introduction 13
2.2. Resist Chemistry 13
2.3. Basic Mechanism for Positive DUV Photoresists 14
2.4. Deprotection 15
2.5. Comparison with I-line Photoresists 15
2.6. Experiment to determine cause of photoresist thinning 16
2.6.1. Experimental Design 16
2.6.2. Experiment Description 17
2.7. Results 18

Chapter 3. Deprotection Induced Thickness Loss 21

3.1. Introduction 21
3.2. Fourier Transform Infrared Spectroscopy (FTIR) 21
3.3. Experiment to Correlate Resist Thinning to Deprotection 23
3.3.1. Experimental Design 23
3.3.2. Experiment Description 23
3.3.3. Results 24
3.4. DUV Lithography Control through Resist Monitoring 28
3.5. Experiment 28
3.6. Modeling 29
3.7. Implications to process controllability and implementation issues 30

Chapter 4. Artificial Neural Networks (ANN) and Adaptive Simulated Annealing (ASA) 33

4.1. Introduction 33
4.2. Simulated Annealing 33
4.2.1. Basic of Simulated Annealing 33
4.2.2. Adaptive Simulated Annealing 35
4.3. Articial Neural Networks 35
4.3.1. Multi-Layer Perceptron 36
4.3.2. Radial Basis Function Network (RBFN) 38
4.3.3. Training the RBFN 39

Chapter 5. Comparison of the ASA and the NN-ASA Algorithms 43

5.1. Introduction 43
5.2. Broadband Reflectometry 43
5.3. The Forouhi-Bloomer (F-B) Dispersion Relation 44
5.4. Experimental Setup 45
5.5. Optimization using the ASA algorithm 45
5.6. The Neural Network Enhanced ASA Optimization Algorithm 47
5.6.1. Parameter Extraction using ASA 47
5.6.2. Monte Carlo Simulation using F-B formulation & Maxwell equations. 48
5.6.3. Spectral Feature Selection 49
5.6.4. Neural Network Training and Validation 50
5.7. Results 50

Chapter 6. Conclusions 53

6.1. Summary 53
6.2. Future Work 55

References 57

A Software Code for the NN-ASA Algorithm 59

B List of Symbols 63

Acknowledgments

I would like to thank my research advisor, Professor Costas J. Spanos, for his valuable guidance, helpful suggestions and support of my research and studies. I also thank Professor Andy R. Neureuther for his comments and encouragement.

I acknowledge Ron Kovacs, Jim Dey, John Plummer, Joe Bendik, Andy Romano and Matt Hankinson for their insights, encouragement and help during my summer internship at National Semiconductor. I would also like to thank Rick Dill of IBM Corporation for sharing his advice on my research. I want to especially thank Piotr Zalicki of SC Technology for helping with the hardware problems that we encountered. I would like to thank Katalin Voros, Bob Hamilton, Kim Chan and other staff of the Microlab for their expertise and assistance.

A warm thanks goes out to present and past members of the Berkeley Computer-Aided Manufacturing (BCAM) group who helped make the experience worthwhile: Roawen Chen, Mark Hatzilambrou, Anna Ison, Herb Huang, David Mudie, Xinhui Niu, Manolis Terrovitis, Greg Luurtsema, John Musacchio, Jeff Lin, Haolin Zhang, Nikhil Vaidya, Junwei Bao, Jae-Wook Lee and Crid Yu.

I would like to thank my family in India for their love and constant support. Finally, I would like to express my deepest affection and gratitude to my wife Sudnya. Without her love and support, this thesis would not have been possible.

This work was supported by the Semiconductor Research Corporation (SRC), the state of California Micro program, and participating companies (Advanced Micro Devices, Applied Materials, Atmel Corporation, Lam Research, National Semiconductor, Silicon Valley Group, Texas Instruments).

Chapter 1 Introduction

1.1. Background and Motivation

Feature dimensions in semiconductor manufacturing are continually decreasing, while die and wafer sizes are increasing. As the critical feature size decreases below 0.3 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 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. In-line refers to processing steps or tests that "are done without moving the wafer" and are usually unobtrusive, non-contact and with little extra cost to the process. This is in contrast to off-line metrology, where the wafer needs to be removed from the processing environment to be measured. 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. An ideal in-line metrology sensor would be capable of making measurements that are sufficiently accurate, repeatable and rapid at a low cost. At present there is no single technique capable of meeting all of these demands.

1.2. Developing In-Situ Metrology: A Step by Step Approach

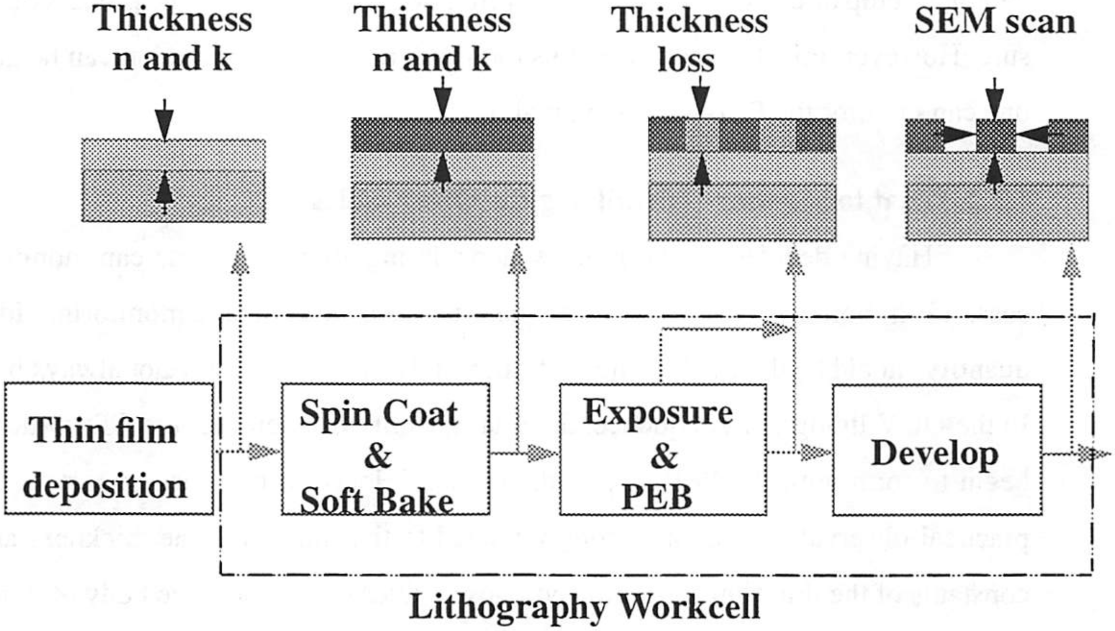


Figure 1.1. Opportunities for measurement in the DUV lithography sequence

The DUV lithography process provides the process engineer with numerous opportunities to monitor the process and wafer state (Figure 1.1). 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 in-situ sensors and metrology for a process, however, requires knowledge of the following:

1.2.1. When to Monitor - Sensor location in the process stream

The first step in designing a sensor for a specific process, is to take into account the process and equipment restrictions, while deciding on the location of the sensors. The lithography process consists of many sequential modules such as the resist spin coat, soft bake, exposure, post-exposure bake and development. Some of these modules are visited more than once, such as when an anti-reflective coating is integrated into the process. Ideally, one would like to place sensors after or during each and every module. However, this may not be practically feasible due to a variety of reasons. Due to the nature of chemically amplified resists used in DUV lithography, the delay between the exposure and the post-exposure bake (PEB) steps needs to be reduced to a minimum. This would eliminate any time consuming measurement of the wafer between these two steps. Similarly, due to the complex setup of DUV steppers, it is very inconvenient to monitor the process during exposure. However, this does not restrict us too much as lots of information can be gathered if one can monitor the PEB process in real time.

1.2.2. What to Monitor - Identifying the Observables

Having decided on the process steps during/after which one can monitor the process, it is important to decide on what quantity one is interested in monitoring. Ideally, this quantity should be the final quantity of interest. However, this may not always be possible. 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 that are strongly related to the final CD. The thickness and optical constants of the thin film stack are good observables due to the large body of work done in measuring these quantities both accurately and in real time as well as due to studies correlating the reflectivity to the CD.

1.2.3. How to Monitor - Designing the Sensor

Having identified the quantities that one would like to monitor, it is then important to either use existing sensors or 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. An in-situ implementation of the broadband reflectometer has been successfully demonstrated in the past [2]. Thus, the third step in designing a sensor for a specific process, is to identify a technique which is highly sensitive to the observable that needs to be measured.

1.2.4. Real Time Algorithms

Most sensors gather a lot of data that is indirectly related to the quantity being measured. Algorithms are required to extract relevant information from this data. To realize real time control, algorithms are needed that can extract this information in real time and feed this to other modules for either feedback or feedforward control. Algorithms for extracting the optical constants for wavelengths in the visible range did very well in this aspect. However, due to the complex behavior of the optical constants in the DUV range of wavelengths, real time algorithms have again become an active research area. This is the final step in the designing of in-situ metrology for real time applications.

1.3. Thesis Organization

This thesis begins with an overview of chemically amplified DUV photoresists, along with definitions of some of the terms and a comparison with the traditional I-line photoresists in Chapter 2. Chapter 3 begins with a primer on Fourier Transform Infrared Spectroscopy (FTIR) as an off-line tool to characterize the DUV photoresist chemistry. This is followed by a detailed description of the experiment and the setup that was used to identify an observable for the DUV lithography process (exposed area thickness loss) and correlation of this observable to the resist chemistry (deprotection). It concludes with the work done in using deprotection induced thickness loss for control of the DUV lithography sequence. Chapter 4 gives an introduction to Neural Networks (NN) and Adaptive Simulated Annealing (ASA) as tools that will be used to construct a real time algorithm for optical constants extraction. Chapter 5 describes the dispersion relation used in conjunction with the ASA method for off-line parameter extraction followed by the NN enhanced optimization algorithm (NN-ASA). It ends with a case study of Polysilicon optical constants extraction using the NN-ASA algorithm. Finally, conclusions of this research are presented in Chapter 6.

Chapter 2 Chemically Amplified Photoresists

2.1. Introduction

As the resolution requirements increase, the irradiating wavelengths must shift deeper into the UV region. In this range, the brightness of existing light sources is severely reduced and so is the efficiency of the optical elements. Since the total input energy incident on a resist is the product of light source brightness, time and transmission efficiency of the optics, a decrease in the first and third factors needs to be compensated by an increase in the exposure time, thus resulting in a lower throughput. This motivated the development of photoresists with higher *quantum yields* ϕ, defined as

$$\phi_d = \frac{P_{ind}}{P_{abs}} \tag{2.1}$$

thus resulting in higher throughput. An acid molecule generated by exposure introduces an avalanche of catalytic reactions resulting in higher sensitivity and thus giving these resists their name.

2.2. Resist Chemistry

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.

2.3. Basic Mechanism for Positive DUV Photoresists

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 by-products 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 2.1)

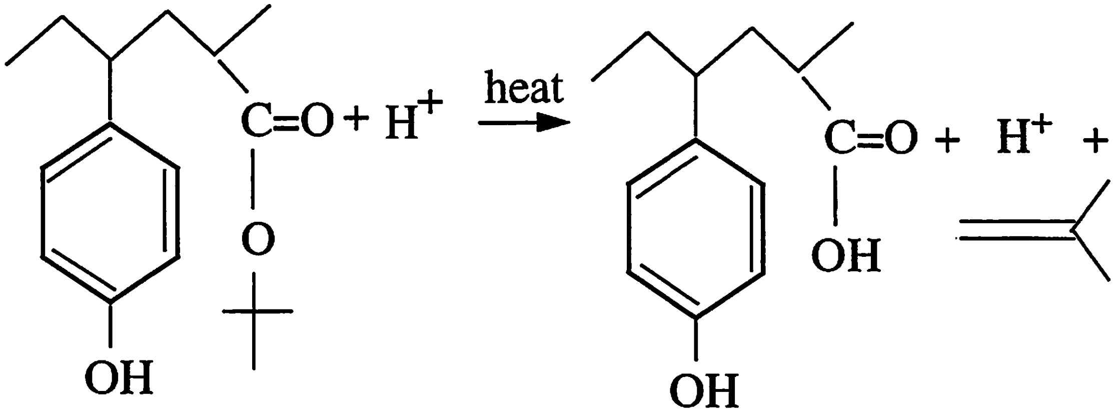


Figure 2.1. Resist Mechanism during the Exposure and Post Exposure Bake Steps for a commercial DUV photoresist.

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 use 10-30% of the hydroxyl groups blocked, with 20% a typical value [16][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.

2.4. Deprotection

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 [18][19][20]. 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. It has become common practice to employ Fourier Transform Infrared Spectroscopy (FTIR) for off-line quantification of the photoacid-induced deprotection of positive chemically amplified photoresists [21].

2.5. Comparison with I-line Photoresists

Conventional positive I-line photoresists are three component materials (i.e. matrix, sensitizer and solvent), whose properties are altered by the photochemical transformation of the photosensitive component, from that of a dissolution inhibitor to that of a dissolution enhancer. According to Dill [22], the sensitizer, also known as the photoactive compound concentration (PAC), can effectively model the exposure and development of I-line positive photoresists. In Dill's exposure model, the absorption of light decreases as M decreases, and in his development model, the reaction can be approximated as a surface-limited etching reaction, whose rate is controlled by the degree of exposure. In summary, the status of the photoresist can be effectively known by monitoring the PAC. Since PAC is related to the extinction coefficient of the photoresist through Dill's A and B parameters,

$$k = \lambda\frac{(A(\lambda) \times PAC + B(\lambda))}{4\pi} \qquad (2.2)$$

measurement of the extinction coefficient of the photoresist would yield the PAC [2].

However, while I-line photoresists bleach during exposure, this is not the case for DUV chemically amplified photoresists. Although, there seems to be a change in color during PEB, this is not due to the bleaching of the resist but rather due to resist thinning (Figure 2.2).

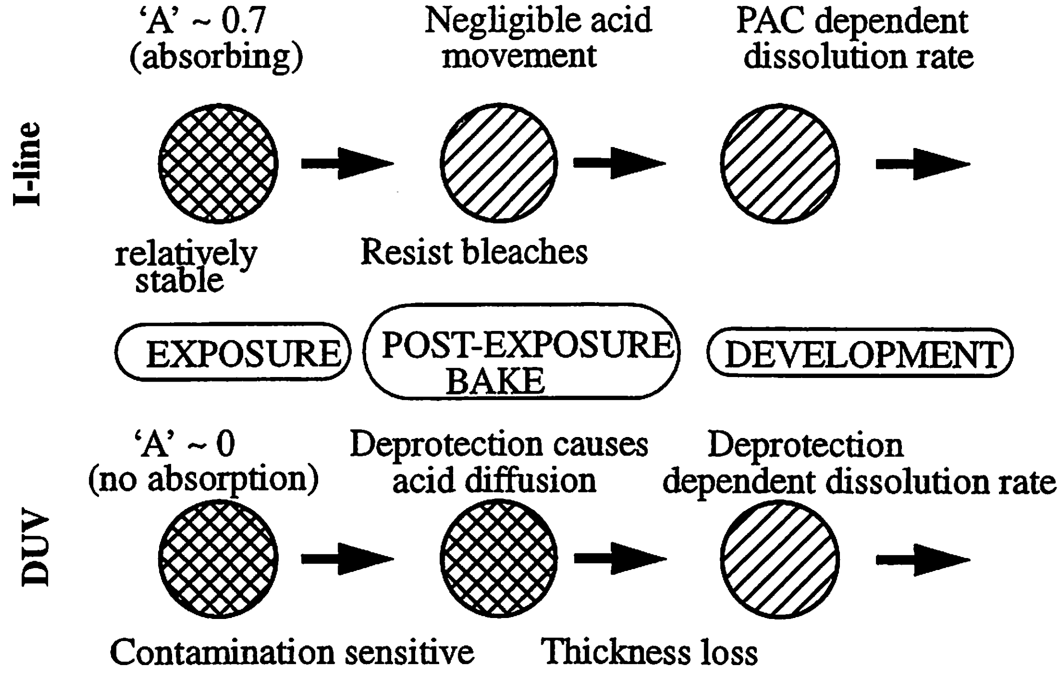

'A' ~ 0.7
(absorbing)
Negligible acid movement
PAC dependent dissolution rate
I-line
relatively stable
Resist bleaches
EXPOSURE
POST-EXPOSURE BAKE
DEVELOPMENT
'A' ~ 0
(no absorption)
Deprotection causes acid diffusion
Deprotection dependent dissolution rate
DUV
Contamination sensitive
Thickness loss

Figure 2.2. I-line photoresists bleach but DUV photoresists do not.

Since chemically amplified photoresists have a similar composition, the first choice for observables to monitor the exposure and PEB steps would be the optical constants of the DUV photoresists, if they were exposure dependent.

2.6. Experiment to determine cause of photoresist thinning

A designed experiment was carried out to determine whether the color change during the PEB was due to change in the absorbance of the resist (change in extinction coefficient *k*) or simply due to a change in the resist thickness.

2.6.1. Experimental Design

To design a statistical experiment, we must decide on the process inputs and the responses to be monitored. Since we are interested in determining the change in absorbance of the photoresist, the response is the extinction coefficient. From our experience with chemically amplified photoresists, we have seen that a change is noticed during the PEB step. Thus, the exposure dose, PEB temperature and PEB time are chosen as the process inputs. $2^3 = 8$ runs are needed for a full factorial experiment. These were augmented with replicated experiments at the operating point, giving a total of 11 runs. The experimental settings are shown in Table 2-1.

Table 2-1. Range of values for the process inputs

Input Factors	Lower Setting (-)	Std. Setting (0)	Higher Setting (+)
Exposure Dose (mJ/cm2)	2.7	3	3.3
PEB Temperature (deg.C)	130	140	150
PEB Time (seconds)	80	90	100

The runs are summarized in Table 2-2. The actual order of execution has been randomized to avoid any blocking effects [3], which occur when similar experiments are executed within a short time span..

Table 2-2. Full Factorial Design Experiment

Run #	Exposure Dose	PEB Temperature	PEB time
1	0	0	0
2	+	-	-
3	+	-	+
4	-	+	-
5	+	+	-
6	-	+	+
7	0	0	0
8	-	-	+
9	-	-	-
10	+	+	+
11	0	0	0

2.6.2. Experiment Description

The wafers were primed with HMDS after which they were coated with a commercial DUV photoresist (UV-5 from Shipley Co.) and soft baked at the standard operating conditions on the FSI wafer track. These wafers were then taken to a Tencor 1250, a single angle spectroscopic ellipsometer, where they were all measured from 235 nm to 1200 nm, with the angle of incidence being 75.87 degrees. The wafers were then taken to the ISI stepper, where they were flood exposed and then post exposure baked on the FSI, according to the DOE. The wafers were again taken to the ellipsometer to be measured.

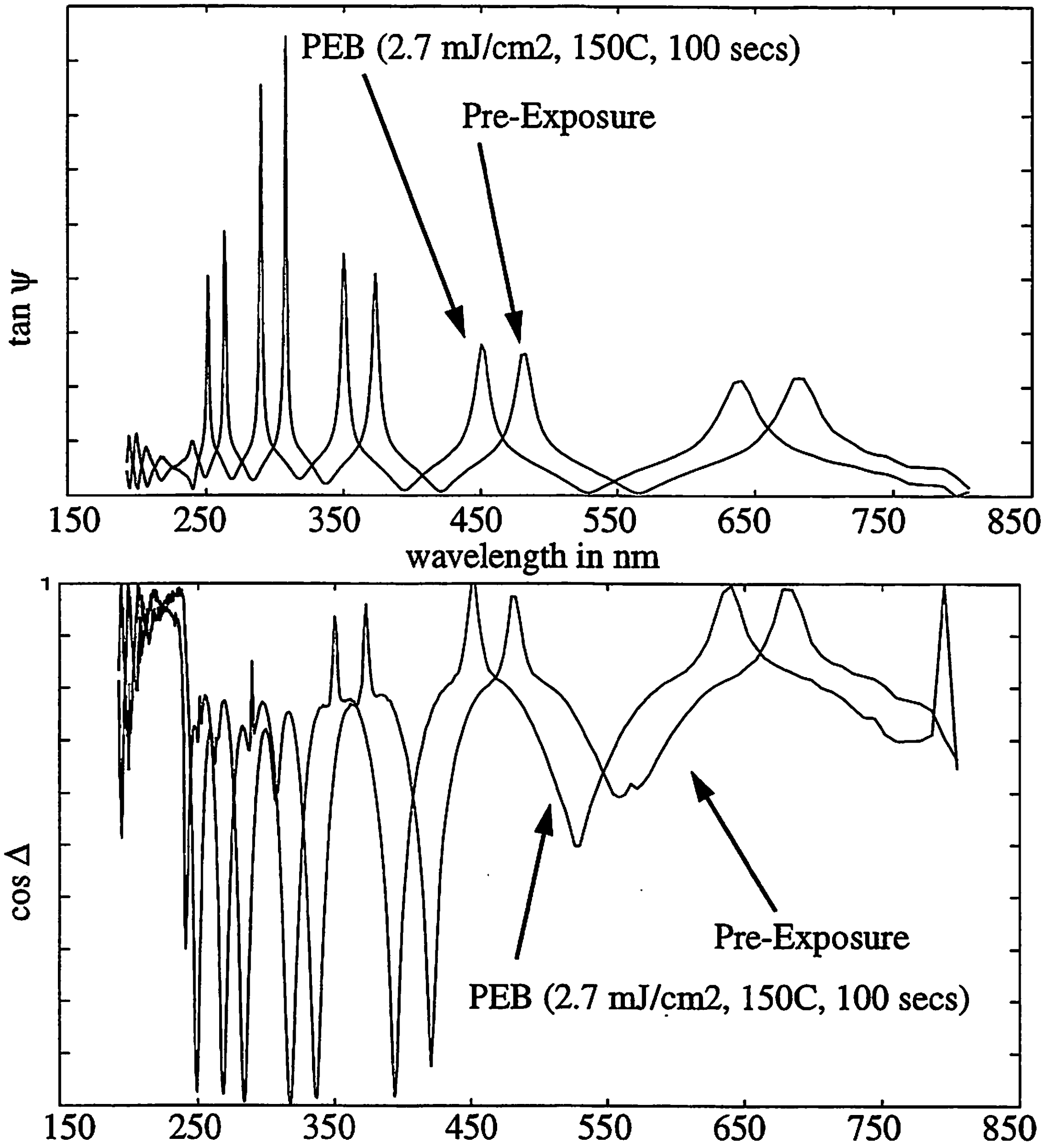


Figure 2.3. Ellipsometry readings for a wafer (run #5) before exposure and after PEB

2.7. Results

The ellipsometry signals obtained, contain information about the optical constants of the photoresist before exposure and after PEB. Curve fitting techniques have to be employed, using global optimization algorithms, to extract this information. Many ellipsometry tools have in-built software for this, but do not guarantee a global solution. Hence, we used Timbre [4], which is an Adaptive Simulated Annealing (ASA) based global optimization toolbox. We also used the Forouhi-Bloomer (F-B) dispersion relation to model the dispersion of the optical constants with wavelength. (For further details about this equation and ASA, refer to chapters 4 and 5)

Before exposure, the real part of the refractive index n and the imaginary part k of the photoresist, were the same for all the wafers to the fourth decimal place, as expected. However, the optical constants after the PEB module did not change either. This indicated that the optical constants were not a function of the processing conditions, as opposed to I-line photoresists. The deviation of the ellipsometry signals before exposure and after PEB was simply due to a photoresist thickness loss (Figure 2.3). Figure 2.4 shows an example of the fitting obtained using the ASA in conjunction with the F-B equations.

This study showed that the DUV photoresists (Shipley's resists in particular) behaved differently from I-line photoresists and hence warranted an observable, other than the absorption coefficient, to monitor the exposure and PEB steps. Moreover, this study suggested using the thickness loss as an observable. However, a more detailed investigation needed to be done, to ascertain whether this thickness loss was meaningfully related to the state of the photoresist.

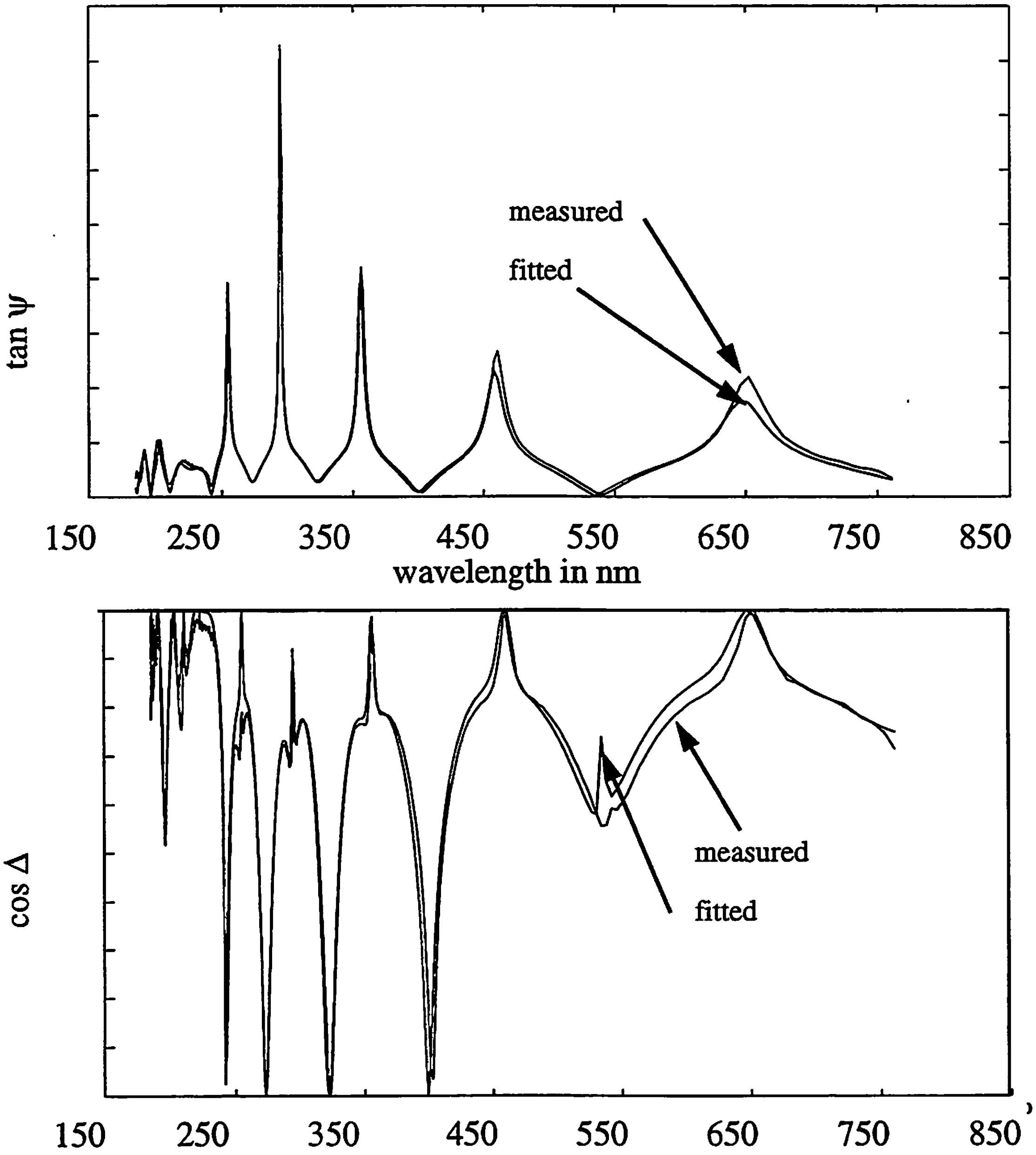


Figure 2.4. Comparison of measured and fitted ellipsometry data for run #5 of DOE

Chapter 3 Deprotection Induced Thickness Loss

3.1. Introduction

The conclusions of chapter 2 indicated that chemically amplified photoresists needed a different measurable, as compared to I-line photoresists, to facilitate control of the photolithography sequence. An observable was needed, that was not only closely related to the wafer state but was also readily measurable in-situ. The results of the DOE described earlier, pointed to the exposed area thickness loss to be a prime candidate for the measurable. In addition, due to the volatility of the cleaved side chain groups, a qualitative prediction about resist thinning can be made. However, a study needed to be done that would correlate the wafer state, which in the case of chemically amplified photoresists would be the deprotection, to the observable (thickness loss). Deprotection is a measure of the amount of deblocking of the resin and hence can be monitored using Fourier Transform Infrared Spectroscopy (FTIR).

3.2. Fourier Transform Infrared Spectroscopy (FTIR)

Fourier transform InfraRed (FTIR) spectroscopy is a powerful analytical tool for characterizing and identifying organic molecules. The absorption IR spectrum of an organic compound serves as its fingerprint and provides specific information about chemical bonding and molecular structure. Infrared light is energetic enough to excite molecular vibrations to higher energy levels. IR spectra usually have sharp features that are characteristic of specific types of molecular vibrations, making the spectra useful for sample identification. The infrared spectrum of a compound is essentially the superposition of absorption bands of specific functional groups, yet subtle interactions with the surrounding atoms of the molecule impose the stamp of individuality on the spectrum of each compound. For qualitative analysis, one of the best features of an infrared spectrum is that the absorption or the lack of absorption in specific wavelengths can be correlated with specific stretching and bending motions and, in some cases, with the relationship of these groups to the remainder of the molecule. Thus, by interpretation of the spectrum, it is possible to state that certain functional groups are present in the material and that certain others are absent. With thus one spectral frame, the possibilities for the unknown can be sometimes narrowed so sharply that comparison with a library of pure spectra permits identification.

FTIR is an interferometric spectrometer, which makes use of the Michelson Interferometer (Figure 3.1). Basically, it consists of a light source, a beamsplitter, a fixed mirror and a moving mirror. The collimated radiation strikes the beamsplitter, where about half of the light is transmitted to the fixed mirror. The remainder of the light is reflected onto the moving mirror. As the two beams are reflected off the surface of the two mirrors, they combine at the beamsplitter, where constructive and destructive interference occurs depending on the position of the moving mirror relative to the fixed mirror. The resulting beam passes through to the sample and continues to the detector.

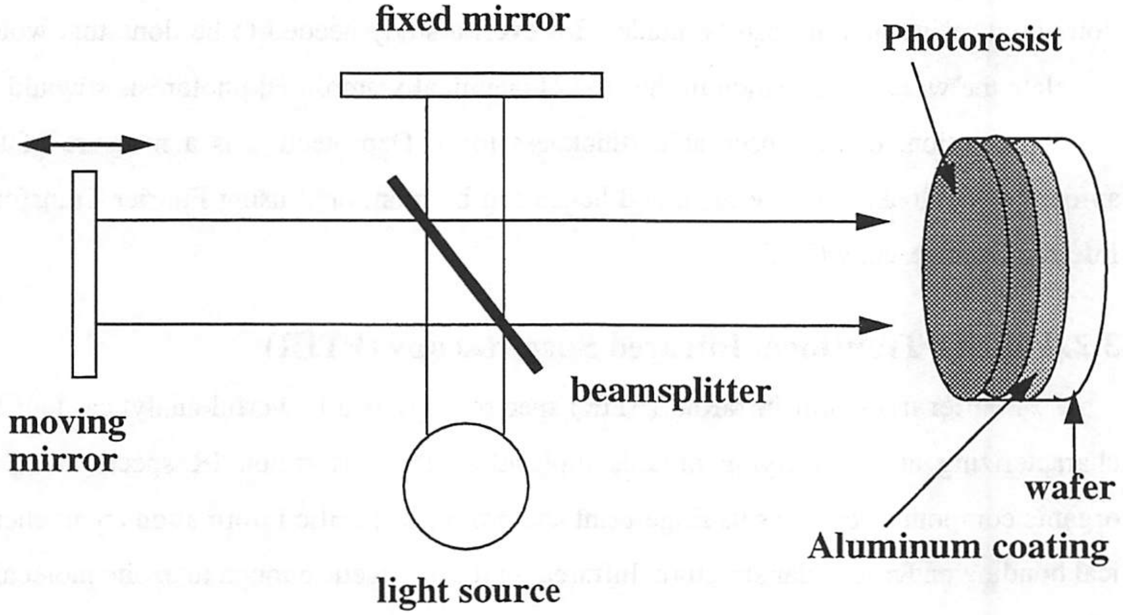


Figure 3.1. The basic principle behind the FTIR spectrometer (Michelson Interferometer)

3.3. Experiment to Correlate Resist Thinning to Deprotection

3.3.1. Experimental Design

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. Experiment Description

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 then 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 then taken to a Tencor 1250 single angle broadband ellipsometer for pre-exposure thickness measurements. The wafers were then exposed using the ISI stepper (KrF excimer laser) at 248 nm with the pattern shown in Figure 3.2 and post exposure baked.

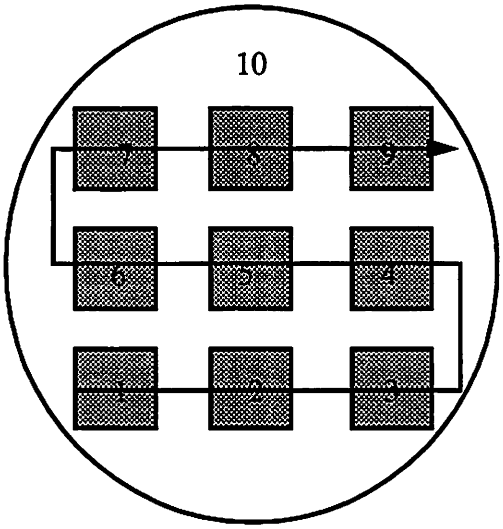


Figure 3.2. Layout of the blanket exposure areas on the wafer

The wafers were taken once again taken 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. The deprotection D was 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.

$$R_i = \frac{A_{dester}}{A_{0ester}} \tag{3.1}$$

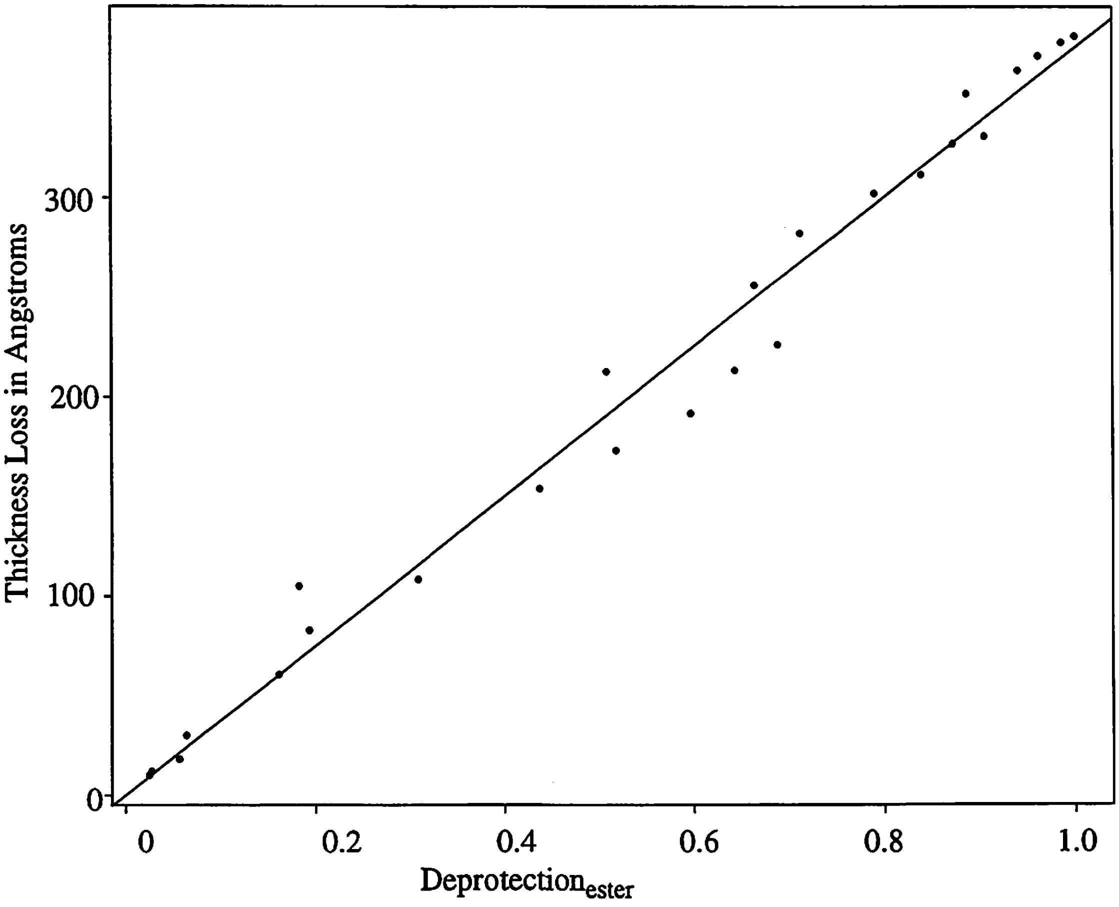
$$D_{ester} = 1 - R_i \tag{3.2}$$

All the thickness loss and deprotection measurements were made on the exposed areas (1 - 9) as well as 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.

3.3.3. Results

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 equation (3.1).

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 3.3 shows the fit.



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

Average model prediction error = 16.52 on 24 degrees of freedom

F-statistic: 5460 on 1 and 24 degrees of freedom

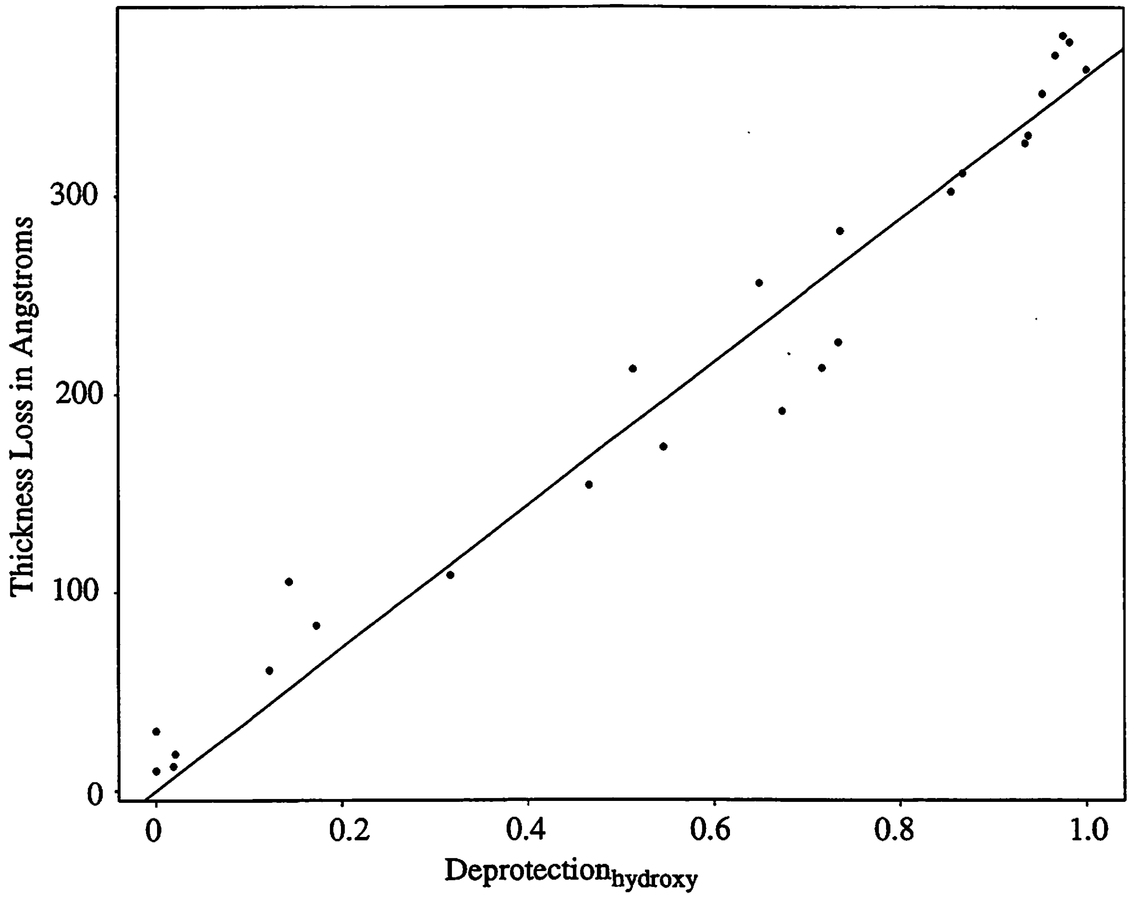
Figure 3.3. Thickness loss as a function of the deprotection measured by monitoring the normalized ester absorbance

Model	Value	Std. Error	t value	Pr(>|t|)
Slope	375.0471	5.0759	73.8884	0.0000

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

$$T_{loss} = 375.0471 \times D_{ester} \tag{3.3}$$

A similar study was done using the integrated hydroxy absorbance as a measure of deprotection and correlated to the final thickness loss. However, this signal was more noisy due to the broad hydroxy absorbance bands. However, this still yields reasonable results.



Summary of Fit: Multiple $R^2 = 0.9897$
Average model prediction error = 25.33 on 24 degrees of freedom
F-statistic: 2310 on 1 and 24 degrees of freedom

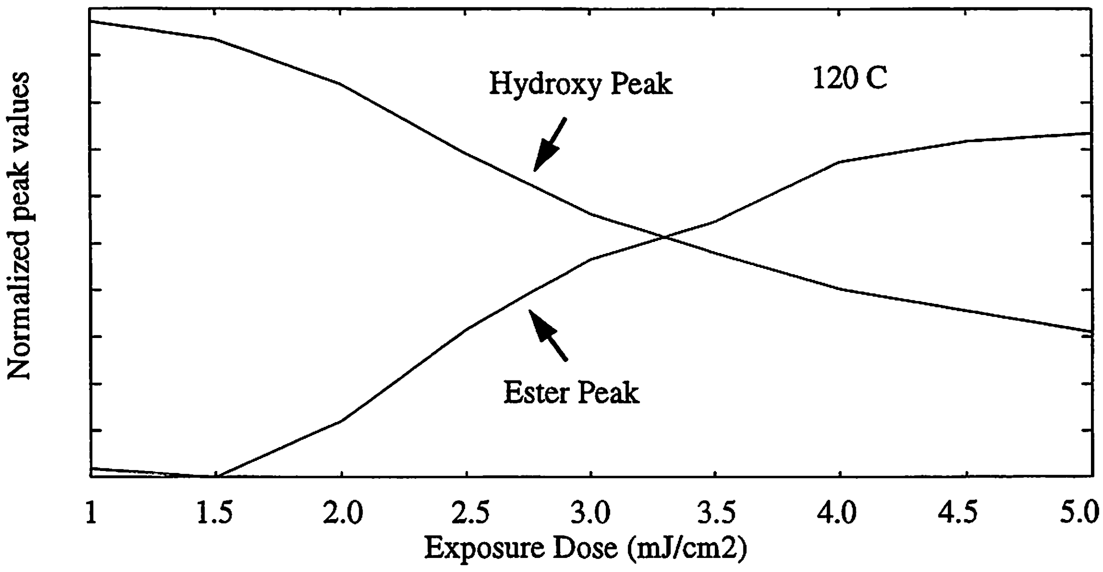
Figure 3.4. Thickness loss as a function of the deprotection measured by monitoring the normalized hydroxy absorbance

Model	**Value**	**Std. Error**	**t value**	**Pr(>|t|)**
Slope	360.2167	7.4942	48.0661	0.0000

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

$$T_{loss} = 360.2167 \times D_{hyd} \tag{3.4}$$

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 rejected the intercept term very strongly (Pr > |t| = 0.8442). Figure 3.5 shows the behavior of the normalized ester and hydroxy peaks processed at a given temperature as a function of exposure dose.



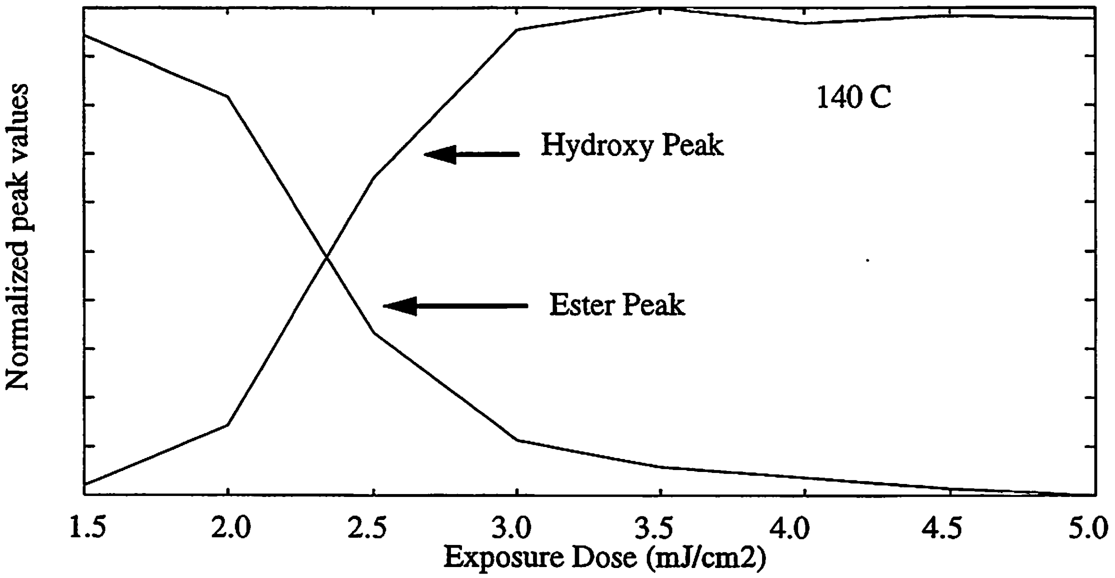


Figure 3.5. The normalized ester and hydroxy peaks as a function of exposure dose for different PEB temperatures

These results strongly suggest that a process control scheme could be based on thickness loss measurements. This is discussed in more detail in the next section.

3.4. DUV Lithography Control through Resist Monitoring

In the preceding sections, we have identified deprotection as a measure of the state of the photoresist. We have correlated the thickness loss in the exposed areas to the deprotection of the photoresist at different temperatures and have shown that this thickness loss is a result of the deprotection induced sidechain collapse. Hence, we have chosen to name this phenomenon deprotection induced thickness loss (DITL). Thus, it has been shown that we can monitor the wafer state in DUV lithography by simply monitoring the exposed area resist thickness loss. Work on using this information for feedback control and diagnosis for the exposure and PEB modules is described in the following sections. We begin with a brief description of the experimental work and the modeling techniques used to build the control models and conclude with the implications of this study in process controllability and the various control architectures that could be used.

3.5. Experiment

We used a statistically designed experiment to build models for the exposed resist thickness loss and the final CD at 0.24 μm as a function of film stack reflectivity and exposure dose. A commercial Shipley DUV photoresist (UV5) was used. The resist thickness loss was measured on blanket areas while the CD was measured in areas patterned with a standard resolution reticle (Figure 3.6). A FSI wafer track and ISI wafer stepper were used for the processing while the measurements were made on a KLA CD-SEM 8100 and a TENCOR 1250 single angle broadband ellipsometer.

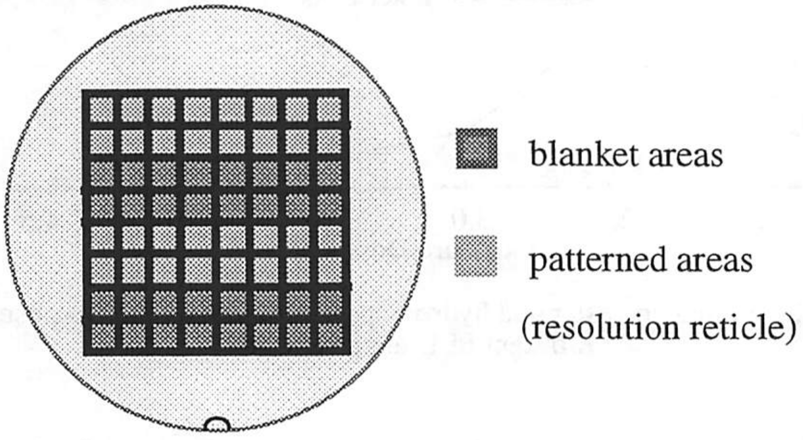


Figure 3.6. Blanket and Patterned Exposure Pattern for the CD control experiment

A total of 10 blank silicon wafers were coated with photoresist at different spin speeds yielding resist thicknesses from 7510 to 7720 Angstroms. Each wafer was soft-baked and was measured for reflectance at 248 nm. The wafers were then exposed using 16 doses for the patterned areas and 16 doses for the blanket areas (Different doses were used for the patterned and blanket areas to account for diffraction effects). The range of the doses had their centers at the optimum for the dose to size and dose to clear. Replication of the doses across the wafer provided better estimates of the error. These wafers were then baked and measured for thickness in the exposed areas to extract thickness loss. This was followed by the develop step and measurement of the CD.

3.6. Modeling

Control models were built using linear regression techniques for the CD as a function of exposure dose (patterned dose) and reflectance as well as for the thickness loss as a function of exposure dose (blanket area) and reflectance. This yielded the following models:

$$T_{loss} = -378.1 + 199.9 \times D_b - 308.9 \times R_{248} \tag{3.5}$$

$$CD(\mu m) = 0.5553 - 0.0395 \times D_p + 0.08 \times R_{248} \tag{3.6}$$

A Monte Carlo simulation was done to build models for the CD as a function of the thickness loss assuming that the blanket area doses were approximately 40% of the patterned area doses. This yielded the following model

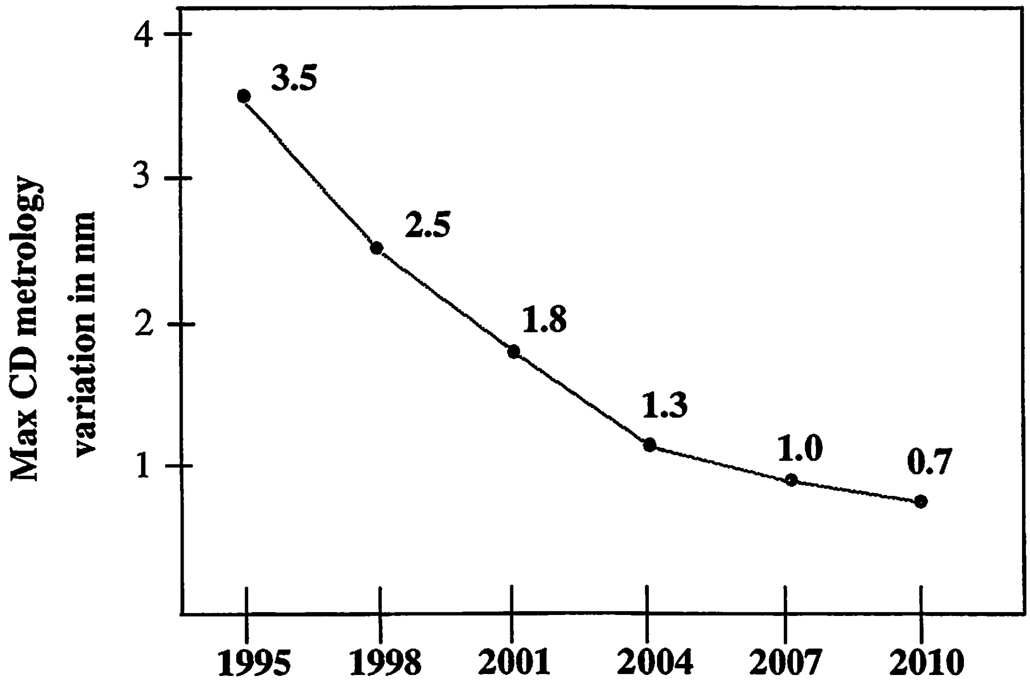
$$CD(\mu m) = 0.4522 - 0.02 \times D_p - 0.00024 \times T_{loss} \tag{3.7}$$

The R^2 for the model was 0.95.

3.7. Implications to process controllability and implementation issues

	1995 *0.35 µm*	*1998* *0.25 µm*	*2001* *0.18 µm*	*2004* *0.13 µm*	*2007* *0.10 µm*	*2010* *0.07 µm*
Gate CD Tolerance (nm)	35	25	18	13	10	7
Final CD output metrology (nm) 3σ reproducibility [atoms]	3.5 [8]	2.5 [6]	1.8 [4]	Transition to inline and in situ control required.		
Overlay (OL) tolerance (nm)	100	75	50	40	30	20
OL output metrology 3σ reproducibility (nm)	10	7.5	5	4	3	2
OL Process Control Metrology: Pre-expose alignment mark distortion estimate (nm)	10	7.5	5	4	3	Sensor/ method required

Figure 3.7. SIA Roadmap for maximum allowed Gate CD Tolerance due to Process Variation



$\sigma_{metrology} = 0.1 * (\sigma_{process})$

Figure 3.8. SEMATECH Roadmap for maximum allowed Metrology Variation

The SIA Roadmap [23] (Figure 3.7) predicts a reduction in the gate critical dimension by using advanced optical and non-optical technologies, advanced resist systems and tighter process control requirements. The SIA roadmap also dictates a maximum allowable CD metrology variation of 2.5 nm by the 0.25 μm generation and 1.8 nm by the 0.18 μm generation in the year 2001 after which it suggests a transition to in-line and in-situ control. Metrology variation specifications on current state-of-the-art systems such as the Scanning Electron Microscope (SEM) based systems is approx 5 nm (3-σ) [24]; higher than that required by the SIA/SEMATECH roadmap. This is reflected in the SIA Roadmap of Technology Needs which states that "Tighter process control requirements will accelerate the adoption of inline and in situ control methods and metrology. Process and tool sensor development will be critical in enabling the use of adaptive process control technology." [23]

A literature survey and talks with process engineers have indicated a 3-σ deviation of approx 25 to 30 nm in the CD from wafer to wafer. This study promises to cut this wafer to wafer process variation below 10 nm using adaptive process control strategies. This calculation is based on equation [3.6] along with the assumption that the typical variation of in-situ reflectometers is approximately 1%. This would improve further with advances in in-situ sensor technology. In addition, the identification of easy to measure observables would allow the use of simple broadband reflectometry as an in-situ sensor. This would reduce the time required for off-line metrology as well as the need to have send-ahead wafers. Typical control architectures that could make use of the metrology proposed would be feed-forward to the exposure step to compensate for reflectivity variation and real time or feedback control to the PEB step by observing the thickness loss and comparing it to a baseline model.

Implementing this strategy would require blanket area exposures to use the thickness loss as a control observable. However, for a given exposure dose, the effective dose coupled into blanket areas is higher than that in patterned areas, due to diffraction effects. To get a similar effective dose in both the blanket and patterned areas, the blanket areas need to be exposed at a lower dose. This requires the stepper to make two passes. In the first pass, the patterned exposure dose is dialed in while in the second pass, the reticle is unloaded and the blanket area exposure is dialed in. This would mean reduced throughput. This problem can be eliminated by using a small area on the product reticle (typically used for bonding pads) that has a line space pattern smaller than the illumination wavelength. This would cause an attenuation of the patterning dose without much modulation . The line space pattern could be designed such that it is easy to print on the reticle while at the same time achieving the required attenuation so as to keep the thickness loss in the linear region. The presence of these blanket areas on each die could also provide a measure of uniformity across the wafer using an in-situ broadband reflectometer with a moving head.

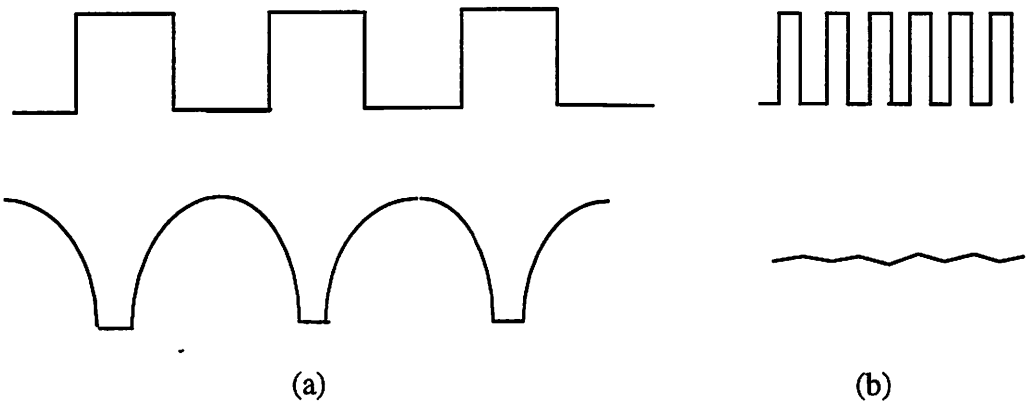


Figure 3.9. a) Aerial image for features above resolution limit of stepper b) Aerial image for features much smaller than the resolution limit of stepper

Chapter 4 Artificial Neural Networks (ANN) and Adaptive Simulated Annealing (ASA)

4.1. Introduction

One of the main issues related to the implementation of in-situ sensors in a real time or run to run control framework is the development of accurate and robust real time algorithms. This chapter gives a brief introduction to the techniques used in this study viz. Neural Networks (NN) and Adaptive Simulated Annealing (ASA).

4.2. Simulated Annealing

Simulated annealing (SA) 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, probability density function of the state space, etc.

4.2.1. Basic of 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 which 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 $\delta E = E_i - E$;
            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
        if random(0,1) < p($\delta E$) then
            return true
        else
            return false
        endif
    endif
```

Figure 4.1. Pseudo-code for the Simulated Annealing Algorithm

There are five major components in SA implementation:

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

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

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

4.2.2. Adaptive Simulated Annealing

There are numerous algorithms that attempt to overcome the disadvantages of simulated annealing. One of the most promising of these algorithms, for the constrained optimization problem, is the ASA. In Adaptive simulated annealing, a parameter x^i_k in dimension i generated at annealing time k with the range $x^i_k \in [A_i, B_i]$, calculated by

$$x^i_{k+1} = x^i_k + p^i(B_i - A_i) \tag{4.1}$$

where $p^i \in [-1, 1]$.

The generation function is

$$g_T(p) = \prod \frac{1}{2(|p^i| + T_i)\ln(1 + 1/T_i)} \tag{4.2}$$

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

$$p^i = \mathrm{sgn}(u^i - 0.5)T_i[(1 + 1/T_i)^{[2u^i - 1]} - 1] \tag{4.3}$$

A straightforward cooling schedule for T_i is

$$T_i(k) = T_{0i}\exp(-c_i k^{1/D}) \tag{4.4}$$

4.3. 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 [5], the Hopfield Network [5], Multi-Layer Perceptron [6], Radial Basis Function Network [6], etc. are some examples. Neural Networks find extensive application in industry in modeling processes which are inherently complex and hence difficult to understand. 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. However, in many cases, 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 good approximating functions for nonlinear models with high computation speeds even with large dimensionality of problem 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 application. An introduction to both these approaches is presented in the following sections.

4.3.1. 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 [6].

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_{hi} be the input to the ith node of the hth layer and y_{hi} be the corresponding outputs. Then,

$$y_{hj} = \frac{1}{1+e^{-\phi_{hj}}} \tag{4.5}$$

where

$$\Phi_{hj} = \sum_{i=1}^{S_{h-1}} w_{ji}x_{hi} + \theta_{hj} \text{ for } j = 1,2,\ldots.S_h \text{ and } h = 1, 2, \ldots L \tag{4.6}$$

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 4.2

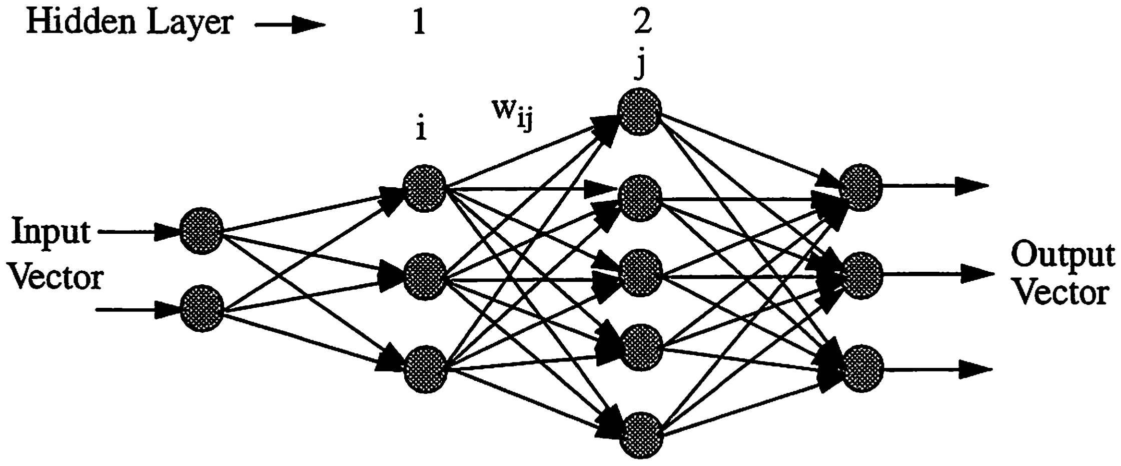


Figure 4.2. Architecture of a Multi Layer Perceptron

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' these input-output mapping by iteratively minimizing an error function. In this case, the error function, E, is the sum of the square of the difference between the calculated (y_j) and the desired output.

$$E = \sum_{j=1}^{N} (\hat{y}_j - y_j)^2 \tag{4.7}$$

where N is the number of output nodes.

MLPs have been successfully applied to solve complex problems by training 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.

4.3.2. 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 4.3). As a result, the contours of constant activation of the hidden layer are hyperspheres instead of hyperplanes as with 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.

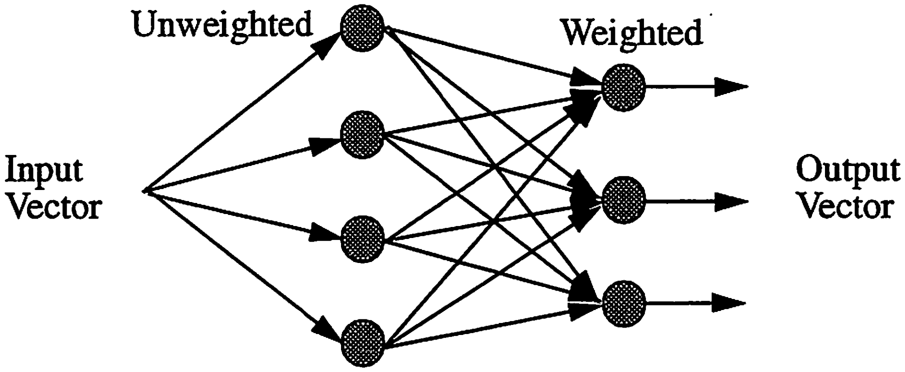


Figure 4.3. Architecture of a Radial Basis Function Network

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,

$$a_h = \phi\left(-\frac{\|x - x_h\|^2}{\sigma_h^2}\right) \tag{4.8}$$

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

$$y_i = \sum_{h=1}^{H+1} w_{ih} a_h = \sum_{h=1}^{H+1} w_{ih} \phi\left(\frac{\|x - x_h\|^2}{\sigma_h^2}\right) \tag{4.9}$$

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 $a_{H+1} = 1$.

4.3.3. 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 [4.7].

The optimality of an RBFN for a particular application is largely dependent on the number of nodes in the hidden layer. By taking 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.

$$d = \min_{1 \le i,\, j \le n_0,\, i \ne j} \{dist(x_i - x_j)\} \text{ for i,j} = 1, 2, \ldots, n_0 \tag{4.10}$$

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

$$dist(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_m - b_m)^2} \tag{4.11}$$

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

$$D_k = \max_{i \ne j} \{dist(x_i, x_j)\} \text{ for i,j} = 1, 2, \ldots N_k \tag{4.12}$$

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

$$R = \max_{1 \le k \le n_0} (D_k) \tag{4.13}$$

When $d > \alpha R$, (where α is a 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 K-Means algorithm is iterated to obtain the new cluster centers. The algorithm converges when the number of clusters do not change.

Chapter 5 Comparison of the ASA and the NN-ASA Algorithms

5.1. Introduction

The need for real time computations of complex functions has necessitated the development of algorithms that are both accurate and fast. There are numerous examples where such algorithms are required in the semiconductor process control area. The extraction of optical constants of thin films from in-situ broadband reflectometry and ellipsometry signals is a typical application. Another application of real time algorithms is the matching of single wavelength scatterometry signals with database traces for CD measurements. Most current algorithms are limited from either using a local optimizer, or for being too slow to be used for real time applications. We have developed an algorithm to solve the problem of extracting the optical constants from broadband reflectometry / ellipsometry signals.

5.2. 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 [1]. In most semiconductor thin-film reflectometry, the spectral reflectance of a sample is measured through the use of relative reflectance methods. In this method, the comparison of the reflectance from the test sample with that of a reflectance standard (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

$$\min_{\lambda} \{ \sum ((R(\lambda) - R(n, k, \lambda, T)) w_{\lambda})^2 \} \tag{5.1}$$

where w_{λ} is the optimization weight. R is a function of optical properties and thicknesses of all the thin-films in the stack. Different settings of w_{λ} yield different optimization speeds, sometimes even different results. Figure 5.1 shows the block diagram of the various modules required to tackle this problem. The two techniques that can be employed for optimization, viz. the Adaptive Simulated Annealing (ASA) and the Neural Network (NN) have been described in Chapter 4. The choice of the dispersion relation is the other critical step in this optimization problem and is discussed in section 5.3

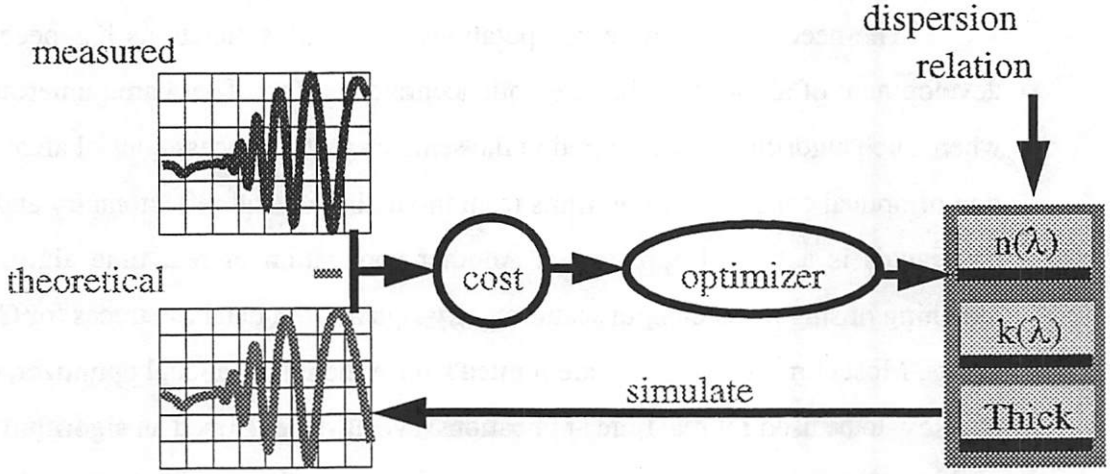


Figure 5.1. The objective is to match the simulated and measured broadband spectra by tuning the parameters of the dispersion relationship using the optimizer

5.3. 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:

$$k(E) = \sum_{i}^{q} \frac{A_i(E-E_g)^2}{E^2 - B_iE + C_i}, \qquad n(E) = n(\infty) + \sum_{i=1}^{q} \frac{B_{0i}E + C_{0i}}{E^2 - B_iE + C_i}, \quad (5.2)$$

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.

5.4. 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.

5.5. Optimization using the ASA algorithm

Since there exist local minima in the solution of Eq [5.1] 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 Figure [5.2]

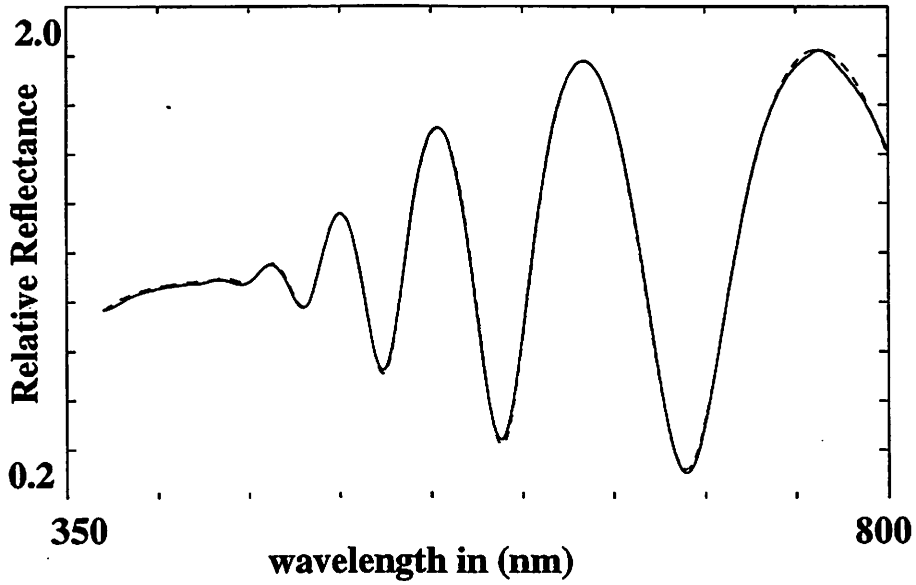


Figure 5.2. Results of the ASA Optimization Algorithm. Figure shows the simulated versus the experimental reflectance spectra

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.

5.6. 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

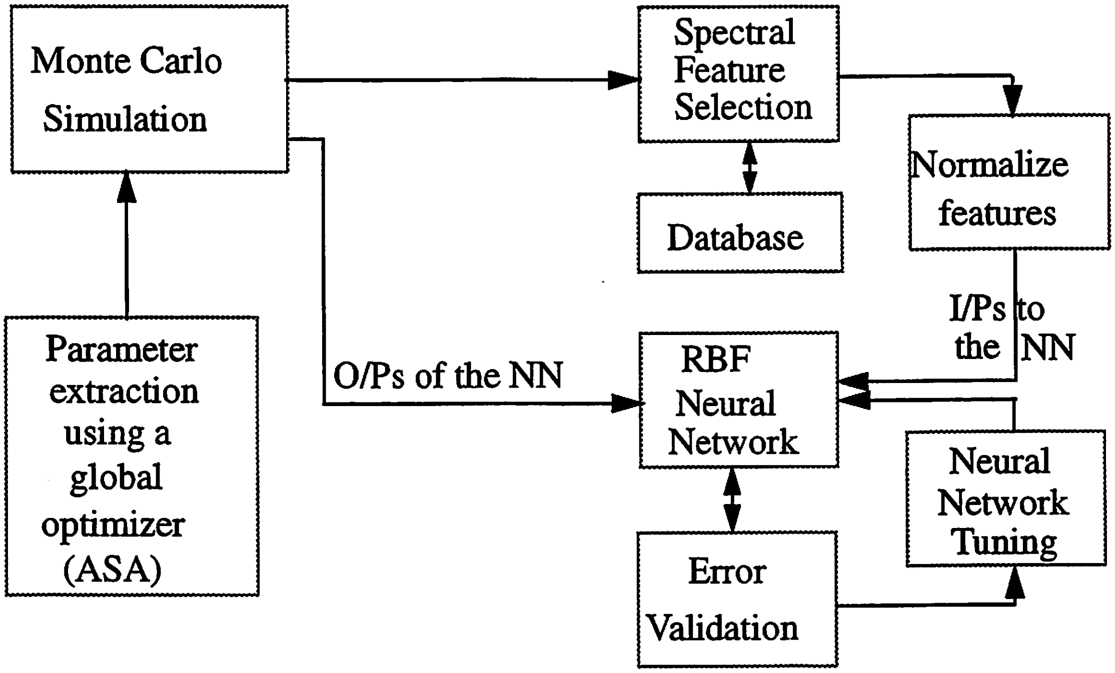

Monte Carlo Simulation
Spectral Feature Selection
Normalize features
Database
I/Ps to the NN
Parameter extraction using a global optimizer (ASA)
O/Ps of the NN
RBF Neural Network
Neural Network Tuning
Error Validation

Figure 5.3. Block Diagram of the NN-ASA Algorithm

5.6.1. Parameter Extraction using ASA

This step is the ASA optimization technique described in section 5.5. 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 section [5.5] 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.

5.6.2. 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

$$M = \begin{bmatrix} \cos(k_0 Nl) & \frac{1}{i \cdot N}\sin(k_0 Nl) \\ \frac{N}{i}\sin(k_0 Nl) & \cos(k_0 Nl) \end{bmatrix} \tag{5.3}$$

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

$$M = \prod_{j=1}^{N_T} M_j \tag{5.4}$$

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

$$R = \frac{(M_{11} + M_{12}N_{si})N_{air} - (M_{21} + M_{22}N_{si})}{(M_{11} + M_{12}N_{si})N_{air} + (M_{21} + M_{22}N_{si})} \tag{5.5}$$

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.

5.6.3. 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 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 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

$$n = \sqrt{\frac{1 \pm \sqrt{R}}{1 \mp \sqrt{R}} n_3 n_1} \tag{5.6}$$

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.

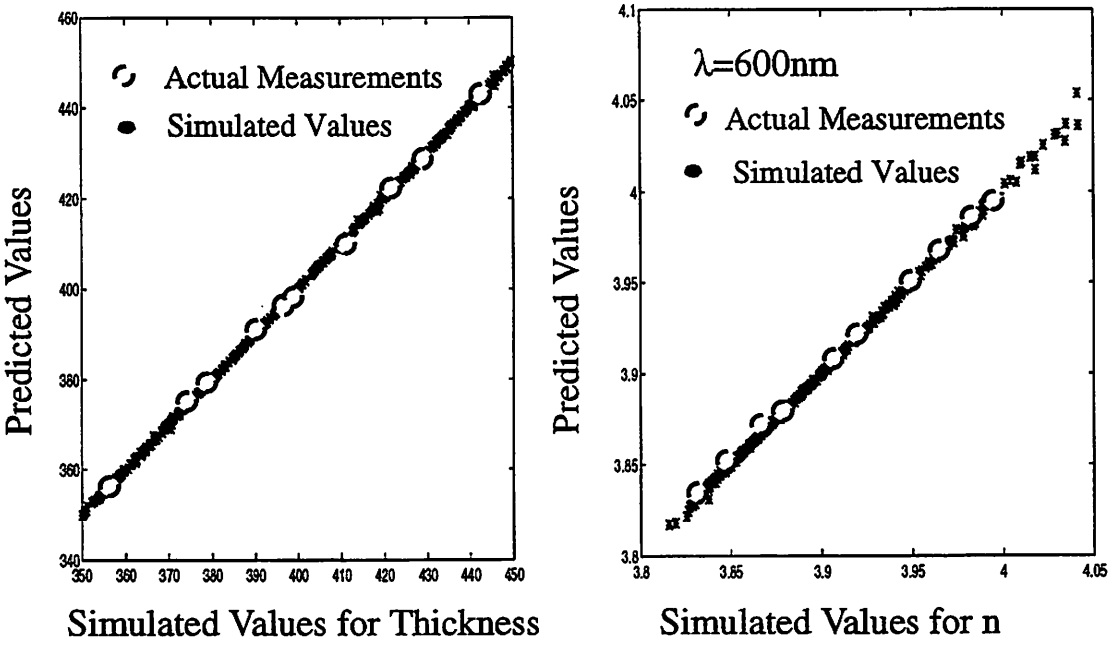
5.6.4. Neural Network Training and Validation

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.

5.7. Results

The results of this optimization are shown in Figure 5.4. 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 λ = 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 5.4. Performance of the NN-ASA algorithm for poly-Si on native oxide on Si stack.

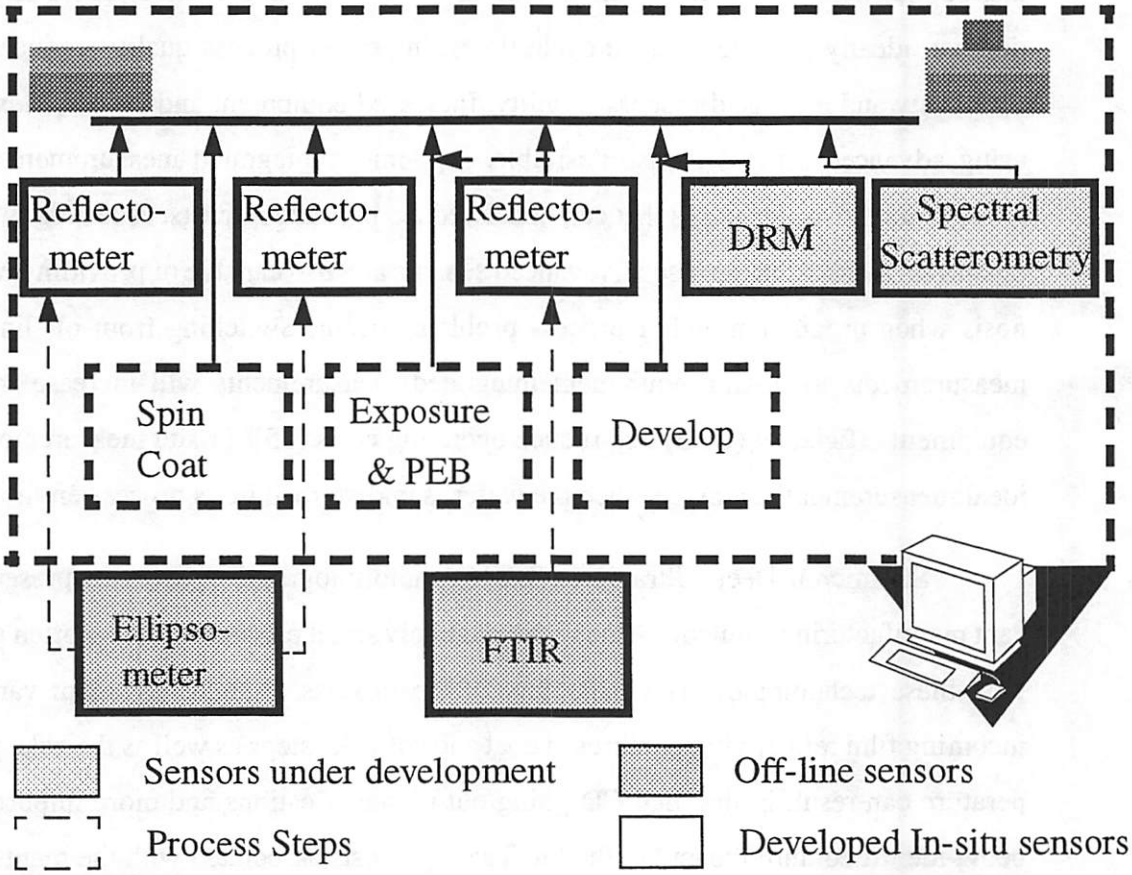
Chapter 6 Conclusions

6.1. Summary

As semiconductor manufacturing technology advances, there exists a need for process monitoring and control. Some of these improvements may be possible by investigating and integrating advanced process sensors. Sensors typically provide information to equipment controllers for effective manufacturing line operation. Effective sensor implementation can ideally provide reduced cycle times, improved process quality, reduced process variability and increased process stability. Increased equipment and process flexibility by using advanced sensors is also desirable. Equipment integrated measurements can also reduce monitor wafer usage that can hence reduce production costs, especially when eight or twelve inch wafers are used. Advanced sensors are also capable of providing rapid diagnosis when production and/or process problems occur. Switching from off-line product measurements to in-situ, equipment integrated measurements will increase the overall equipment efficiency (OEE) and reduce operating costs [15]. In-situ measurements are an ideal measurement technique since the wafer is undisturbed in its process environment.

Submicron Deep Ultraviolet (DUV) photolithographic processes present significant manufacturing challenges due to the relatively small 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 CD going out of specifications and more importantly, not being identified until the end of the lot. The high costs associated with the manufacture of Integrated Circuits (ICs) necessitates higher yields and throughput. Lithography, thus presents itself as an ideal candidate for in-situ process monitoring and control.

As was seen in Figure 1.1, there are numerous opportunities for measurements in the DUV lithography sequence. The aim of this research was to develop in-situ sensors and metrology for the DUV lithography sequence. This has been successfully completed. However, the final goal is to develop in-situ metrology for all the observables in the lithography sequence and then integrate these sensors into a supervisory control scheme. Figure 6.1 shows the various sensors developed or currently under development for the lithography sequence and one possible integration scheme for the sensors. Spectral Scatterometry looks at the patterned wafer after development and aims to predict the critical dimension using a broadband reflectance spectrum. The DRM sensor analysis is being modified to predict the percentage of patterned area as well as absolute thickness values. Both these sensors are currently in the research phase. The figure also shows the off-line metrology tools that were used to calibrate the in-situ sensors.



Reflecto- meter
Reflecto- meter
Reflecto- meter
DRM
Spectral Scatterometry
Spin Coat
Exposure & PEB
Develop
Ellipso- meter
FTIR
▨ Sensors under development
▨ Off-line sensors
Process Steps
☐ Developed In-situ sensors

Figure 6.1. Supervisory Control Scheme for the DUV Lithography Sequence

6.2. Future Work

Having developed the equipment models, the next step would be to implement feed-back control on the lithography sequence using the exposure dose and the PEB time as control knobs based on reflectance measurements made after the spin coat and soft bake steps. Adaptive modeling will be required to account for drifts in the equipment models over time. Although elementary, this study would demonstrate the efficacy of run-to-run control using in-situ sensors.

The next step would be to design more advanced in-situ sensors that will be able to measure more complicated quantities, such as the CD, reliably. This would enable a more direct control since our control decisions would be based on the measurement of a quantity that we are interested in controlling as opposed to the measurement of a quantity that is indirectly related to the final quantity of interest. This would also enable higher throughput in the process because the wafers could be measured in-situ and in real time, thus eliminating the time consuming post-process measurements. Spectral scatterometry is one such technology that is being investigated.

The final goal is to integrate the sensors and control algorithms into a supervisory control scheme for the lithography sequence and demonstrate its efficacy on a production line process.

References

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

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

[3] G. Box, W. Hunter & S. Hunter, "Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building" 1st ed., New York, John Wiley & Sons, 1978

[4] X. Niu, Timbre - An Adaptive Simulated Annealing Based Optimization Toolbox

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

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

[7] S. Mandayam, et. al. "Inverse Problems in Magnetostatic NDE", Journal of Non-Destructive Evaluation, Vol 7, Nos. 1-2, pp. 111-120, 1988

[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] X. Niu and C. J. Spanos, "In-Situ Optical Metrology of Polycrystalline Silicon", SRC TECHCON'96.

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

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

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

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

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

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

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

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

[20] 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

[21] N. K. Eib, E. Barouch, U. Hollerbach, S. Orszag, "Characterization and Simulation of Acid Catalyzed DUV Positive Photoresist", SPIE vol. 1925, 1993

[22] F. H. Dill et al, "Modeling Projection Printing of Positive Photoresist", IEEE Trans. on Electron Devices, vol. ED-22, No. 7, July 1975

[23] http://www.sematech.org/public/roadmap/doc/tbl18_gif.gif

[24] KLA-Tencor 8100 CD-SEM Product Information.

Appendix A Software Code for the NN-ASA Algorithm

```
* Monte Carlo Simulation
B2 <- runif(1000, 7.113467*0.99, 7.113467*1.01);
B3 <- runif(1000,8.750376*0.99, 8.750376*1.01);
C2 <- runif(1000, 14.15319*0.99, 14.15319*1.01);
C3 <- runif(1000, 20.24209*0.99, 20.24209*1.01);
thick <- runif(1000, 350, 450);

mat <- t(rbind(B2, B3, C2, C3, thick));
write.table(mat, file="random1.data", sep=" ")
```

```
load random1.data;

lambda = 200:900;
e = lambdatoe(lambda);


Nsi = forouhi(e, 1.06, [0.00405 0.01427 0.06830 0.17488], [6.885 7.401 8.634
10.652], [11.864 13.754 18.812 29.841], 1.95);

Nsio2 = forouhi(e, 7.00, [0.00867, 0.02948, 0.01908, 0.01711], [20.729, 23.273,
28.163, 34.301], [107.499, 136.132, 199.876, 297.062], 1.226);

ref = zeros(1000, 700);


A1 = 0.08951641;
A2 = 0.0499771;
A3 = 0.08799342;
A4 = 0.07993872;
B1 = 7.216723;
B2 = 7.113467;
B3 = 8.750376;
B4 = 10.20751;
C1 = 13.15469;
C2 = 14.15319;
C3 = 20.24209;
C4 = 34.91485;
Eg = 1.298818;
nInf = 2.452188;
```

```
fid1 = fopen(‘randomN’, ‘w’);
fid2 = fopen(‘randomK’, ‘w’);
fid3 = fopen(‘randomRf’, ‘w’);

load ref.norm
refSi = ref;
ref = zeros(1000, 700);

for i=1:1000
  Npoly = forouhi(e, Eg, [A1 A2 A3 A4], [B1 random1(i,1) random1(i,2) B4], [C1
random1(i,3) random1(i,4) C4], nInf);
  thickVec = [-1; random1(i, 5); 31.4; -1];
  for j=1:700
    thisLambda = j+199;
    nVec = [1; Npoly(j); Nsio2(j); Nsi(j)];
    [a, b, ref(i,j), d] = ReflectTEM(thisLambda, nVec, thickVec, 0);
    ref(i, j) = ref(i, j) / refSi(j, 2);
    fprintf(fid1, ‘%12.8f ’, real(Npoly(j)));
    fprintf(fid2, ‘%12.8f ’, imag(Npoly(j)));
    fprintf(fid3, ‘%12.8f ’, ref(i, j));
  end
  fprintf(fid1, ‘\n’);
  fprintf(fid2, ‘\n’);
  fprintf(fid3, ‘\n’);
%  plot(ref(i,:));

end

fclose(fid1);
fclose(fid2);
fclose(fid3);

function [new] = rbf1(begin,interval,end,testthick)

best=[];
new=[];
counter = ((end - begin)/interval);
%counter=0;
for i=1:(counter+1)
       spread = begin + ((i - 1)*interval);

%===============================================================
% parameters
%===============================================================
disp_freq=100;
max_neuron=1000;
err_goal=0.000001;
%spread=.1;

DP=[disp_freq max_neuron err_goal spread];
%===============================================================
% inputs
%===============================================================
load bpr1.in;
P=bpr1; clear bpr1;
load bpr1.out;
T=bpr1’; clear bpr1;
```

```
[W11,B11,W21,B21,epochs,TR]=solverb(P,T,DP);

%==============================================================
% outputs
%==============================================================
save Wrbf1.mat W11 B11 W21 B21
[A11,A21]=simurb(P,W11,B11,W21,B21);
load test1.dat;
[A11,A21]=simurb(test1,W11,B11,W21,B21);
        new=[new A21];
       siger1=0;
       siger2=0;
       siger3=0;
       [rt ct]=size(testthick);
       for j=1:rt
              siger1 = siger1 + (testthick(j,1) - A21(1,j))^2;
%             siger2 = siger2 + (testthick(j,2) - A21(2,j))^2;
%             siger3 = siger3 + (testthick(j,3) - A21(3,j))^2;
       end;
       siger1 = sqrt(siger1/rt)
%      siger2 = sqrt(siger2/rt)
%      siger3 = sqrt(siger3/rt)
%      best = [best; i spread siger A21];


% where siger is the sigma error for that spread

end;
%[best2 i]=sort(best(:,3))
```

Appendix B List of Symbols

a_h	activation at the output of the hth node
A_{dester}	peak area of ester bond at exposure dose d
A_{0ester}	peak area of ester bond with no exposure
D_b	dose in blanket areas
D_p	dose in patterned areas
D_{xester}	deprotection as measured by ester vibration band
D_{hyd}	deprotection as measured by hydroxy vibration band
E	photon energy
E_g	optical energy band gap
ϕ	transfer function of the hidden node
ϕ_d	quantum yield at dose d
$g(x)$	probability density of state space
$h(\delta E)$	probability of accepting new cost function
h_j	bias for the jth node
H	number of hidden nodes
k	extinction coefficient
L_n	repetition function
L	number of hidden layers
L_i	number of transitions in a simulated annealing scheme
λ	wavelength in nm
M	characteristic matrix
n	index of refraction
P_{ind}	number of photo induced events

	number of photons absorbed
R	reflectance
R_i	ratio of peak areas of absorbance band with and without exposure
R_λ	reflectance at wavelength λ nm
S_h	number of nodes in the h^{th} layer
σ_h	width of the basis function
T	control parameter in simulated annealing schemes
T_n	temperature function
T_{loss}	thickness loss in exposed areas
w_{ji}	weight connecting the i^{th} node of $(h-1)^{th}$ layer to j^{th} node of h^{th} layer MSE mean square error
w_λ	optimization weight
x_h	center of the basis function at hth hidden node
x_{hi}	input to the ith node of the hth layer
y_i	output from the j^{th} node of output layer
y_{hj}	output from j^{th} node of h^{th} layer